Assurons un monde plus ouvert

CNP assurances

Recherche et Prospective stratégique

Septembre 2025

Le risque algorithmique en finance : enjeux et perspectives à l'horizon 2040

Table des matières

Résumé exécutif **2**

Qu'est-ce que le risque de biais algorithmique ? **2**

Définition du risque de biais algorithmique 2

Un risque aux effets accrus à l'ère de l'IA et de l'IA générative 3

La confiance dans l'IA et ses algorithmes abaisse l'esprit critique de ses utilisateurs 3

Derrière le voile des algorithmes, l'angle mort du risque 4

Les développeurs et les utilisateurs dépendants des travaux des Etats-Unis et de la Chine 4

Qui est impacté par ce risque ? **6**

Les métiers les plus concernés par ce risque 6

Focus sectoriel Finance et bancassurance 8

Que faire face à ce risque ? **9**

Les réponses des organisations face à ce risque 9

Face à généralisation de l'IA, quels futurs possibles pour le capital humain ? 9

Scenario 1 – Business as usual 9

Scenario 2 – Confiance aveugle en l'IA 10

Scenario 3 – Renforcement de l'esprit critique 10

Résumé exécutif

Les systèmes d'IA envahissent et transforment les métiers de la finance et de l'assurance. Ils s'appuient sur des données historiques souvent biaisées, ou incomplètes qui amplifient des biais socioéconomiques existants. Or, la confiance, l'équité et la conformité réglementaire sont essentielles pour les assureurs. S'ils prennent le risque que leurs clients soient victimes de discriminations systémiques, ils mettent en gage leur crédibilité et encourent la peine de voir leur réputation faiblir. Les défis sont donc de taille pour rester compétitifs et se donner les capacités de se réinventer : transparence accrue et responsabilisation des équipes, éthique et inclusion financière, performance et confiance client. Face à ces défis, les acteurs de la finance développent le cadre de contrôle nécessaire. Il doit garantir la qualité des technologies exploitées et l'explicabilité des décisions qui sont prises pour un processus d'innovation responsable et durable.

Qu'est-ce que le risque de biais algorithmique ?

Définition du risque de biais algorithmique

Le biais algorithmique est une erreur systématique dans le fonctionnement d'un algorithme, qui conduit à des décisions injustes, discriminatoires ou inéquitables envers certains individus ou groupes. Les biais aggravent ou perpétuent des discriminations socioéconomiques, ethniques et sexistes préexistantes.

Ils peuvent provenir :

- De données historiques biaisées : biais de sélection,
- Soit de préjugés humains intégrés lors de la conception[1] , soit du fait d'une interprétation erronée ou partielle des résultats : biais de confirmation,
- De défauts dans la structure et l'optimisation des modèles : biais d'algorithmie.

Les études scientifiques citent les deux premières sources comme étant celles que les entreprises doivent fréquemment affronter.

En effet, l'amplification par l'IA de biais sociaux existant dans les données historiques est un phénomène récurrent et bien connu. A titre d'illustration : les services de modération de contenu sur les réseaux sociaux, ayant conduit à une censure abusive de communautés[2]. Les automates programmés pour lutter contre la prolifération de contenus malveillants suppriment mécaniquement des mots associés à des propos haineux, collectés parmi les données sur internet, pour entrainer les modèles.

Le biais d'algorithmie est quant à lui lié à la structure et à l'optimisation des modèles. Le cas du logiciel COMPAS est bien connu. Plusieurs Etats américains l'ont adopté pour évaluer le risque de récidive des personnes arrêtées. L'algorithme sous-estimait le risque pour les individus « blancs » et surestimait celui des autres groupes, « noirs » notamment. COMPAS avait été optimisé pour maximiser la précision globale, au détriment de la précision pour chacun des groupes « raciaux »[3].

Les biais algorithmiques sont à l'origine de situations inopportunes pour les organisations et en particulier pour les institutions financières pouvant entrainer des conséquences sérieuses :

- réputationnelle : dégradation de l'image de l'entreprise.
- réglementaire : sanctions de la part des régulateurs. En France, par exemple, la discrimination commise à l'égard de personnes physiques ou morales, est punie. Les sanctions peuvent aller jusqu'à trois ans d'emprisonnement et 45 000 euros d'amende[4].
- Opérationnelle : dégradation des performances de l'entreprises.

La mise en place de mesures comme l'amélioration de la qualité des données d'entrainement est un enjeu de taille pour les entreprises. En 2020, une étude de Gartner estimait à 12,9 millions d'euros la perte annuelle moyenne mondiale due à l'utilisation de données de mauvaises qualité[5].

[1] Biais algorithmiques - Data Bird
[2] IA et discrimination : un cas d'école de « discrimination algorithmique »
[3] Machine Bias : by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
[4] Article 225-2 du Code pénal
[5] Data quality: Best practice for accurate insights, Gartner

Un risque aux effets accrus à l'ère de l'IA et de l'IA générative

Le risque de biais algorithmiques est lié à :

- la généralisation des systèmes algorithmiques dans les processus des entreprises,
- la diffusion d'information, de contenus en libre accès,
- et l'adoption massive des IA génératives par des utilisateurs non-initiés.

Cette révolution n'est pas près de s'arrêter. Ces dernières années, le secteur a attiré des capitaux importants. Rien qu'en 2024, la part des capitaux privés s'élève à 33,94 milliards de dollars dans le monde, soient 8,5 fois de plus qu'en 2022[6].

Compte tenu de la rapidité d'expansion des activités en la matière, trois grands facteurs aggravent l'ampleur du risque considéré : humain, technologique, géoéconomique.

La confiance dans l'IA et ses algorithmes abaisse l'esprit critique de ses utilisateurs

Le premier facteur tient à la perception des usagers de l'IA. Un excès de confiance en la technologie et un faible esprit critique[7] accroissent la propension aux biais. Des chercheurs de l'Université Carnergie Mellon et de Microsoft ont montré que les personnes accordant une grande confiance aux IA ont tendance à déprioriser l'analyse critique. La confiance dans ses propres compétences est en revanche associée à davantage d'exercice de l'esprit critique[8]. Or, selon une étude de KPMG et l'université de Melbourne, 58 % des usagers considèrent les IA comme dignes de confiance[9]. Ces derniers ont même plus confiance en l'IA qu'aux humains : selon un sondage mondial Ipsos de 2024, ils sont 54 % à davantage croire en les capacités de l'IA qu'en celles des humains pour éviter les discriminations et les préjugés[10].

Cette croyance s'observe dès l'enfance. Une étude du journal *Computers in Human Behavior* a mis en évidence le fait que les enfants de 3 à 6 ans ont tendance à accorder leur confiance aux robots sociaux[11] plutôt qu'aux humains. L'étude montre même que lorsque les humains se trompent, les enfants considèrent que c'est délibéré[12]. Même constat auprès de la génération Z, nés entre 1990 et 2010 : 41% des jeunes Américains ont plus confiance en l'IA selon une étude conduite par Censuswide, cabinet anglais spécialisé dans les études de marché[13]. Ils préfèrent suivre les conseils des IA, comme Chat GPT, plutôt que d'avoir recours à une expertise humaine, sur des sujets liés à :

- la santé mentale,
- la carrière,
- les relations interpersonnelles,
- le planning financier, soit une « approche structurée pour organiser et optimiser [des] finances à long terme »[14].

La confiance dans les algorithmes et les modèles peut :

- minimiser les résultats qui contredisent les hypothèses de départ,
- limiter la remise en question de leurs décisions ou propositions,
- et donc la détection des biais présents[15].

[6] Hai Stanford, "The 2025 AI Index Report", 2025
[7] Amanda Sturgill, Unlocking the Link Between Generative AI Confidence and Critical Thinking Skills Mars 2025
[8] Amanda Sturgill, Unlocking the Link Between Generative AI Confidence and Critical Thinking Skills, 25/05/2025
[9] Gillespie, N., Lockey, S., Ward, T., Macdade, A., & Hassed, G. (2025). Trust, attitudes and use of artificial intelligence: A global study 2025. The University of Melbourne and KPMG. DOI 10.26188/28822919.
[10] Ipsos, « L'IA peut-elle être plus juste que les humains ? Perspectives d'une enquête Ipsos mondiale », Août 2024
[11] conçus pour interagir de manière significative avec les humains, en utilisant des comportements sociaux et des émotions pour établir des relations et fournir un accompagnement
[12] Rebecca Stower, Arvid Kappas, Kristyn Sommer, When is it right for a robot to be wrong? Children trust a robot over a human in a selective trust task, Computers in Human Behavior, Volume 157, Août 2024
[13] Forbes Gen Z Trust AI More Than Humans in Their Careers, New Study Shows
[14] Bit panda, Comment créer un plan financier personnel, n.d
[15] Charlotte Barot, [2/3] Croire ou douter : la question des biais de confiance dans la prise de décision, Laboratoire d'Innovation Numérique de la CNIL, 16/10/2024

Derrière le voile des algorithmes, l'angle mort du risque

Le second facteur résulte de l'opacité des algorithmes et modèles d'Intelligence artificielle. Une majorité des modèles récents, notamment ceux basés sur le deep learning, sont considérés comme des « boites noires »[16]. La dernière vague d'IA a produit des modèles plus complexes, tels que les réseaux neuronaux ou les IA génératives. Mais leur transparence et leurs capacités à être audités restent à développer[17].

Le recours à l'IA pour produire des modèles de manière autonome s'accroît rapidement. D'après le cabinet de conseil McKinsey, en 2024, 64 % des entreprises dans le monde utilisaient régulièrement l'IA générative, contre 33 % en 2023[18]. En outre, le rapport AI Index 2025 de Stanford indique que les assistants de codage IA surpassent désormais les programmeurs humains dans certaines tâches.

Les développeurs et les utilisateurs dépendants des travaux des Etats-Unis et de la Chine

Le dernier facteur tient au fait que la collecte des données d'entrainements et les développements des modèles se font majoritairement dans des pays en proie aux discriminations : identité personnelle, situation sociale ou économique, santé, pour ne citer que ceux-là. En Chine par exemple, des études montrent que le marché du travail favorise les hommes Han[19] par rapport aux femmes et aux hommes d'ethnies minoritaires. Des écarts de salaires non négligeables illustrent cette discrimination. De plus, Le sex ratio était de 103 hommes pour 100 femmes en 2024 avec une préférence historique des parents pour les garçons. [20]. Cette surreprésentation de certaines catégories de la population[21], et les singularités culturelles perpétuées sont sources de biais. La croissance exponentielle de ces données rend la détection et la correction des biais d'autant plus complexe[22][23].

En matière de recherche et développement, 73 % des modèles sont aujourd'hui conçus en Chine et aux Etats-Unis. Plus précisement, 15 modèles sont développés en Chine, 40 aux Etats-Unis et seulement trois à l'échelle européenne[24]. Leaders en matière d'innovation technologique, ces deux Etats créent une forte dépendance à leurs travaux des développeurs et usagers à travers le monde.

Le marché des modèles de langage ou LLM est globalement concentré autour d'un nombre réduit d'acteurs : OpenAI, Google, Anthropic, Meta, xAI, Mistral, etc. Cette concentration limite la diversité des perspectives et visions du monde que les LLM/modèles restituent à travers leurs réponses. Cela renforce les biais existants, faute de contrepoids méthodologiques ou culturels suffisants.

Cette position influence non seulement les normes et les standards de l'IA mais également les pratiques professionnelles et les capacités cognitives des utilisateurs. [25].

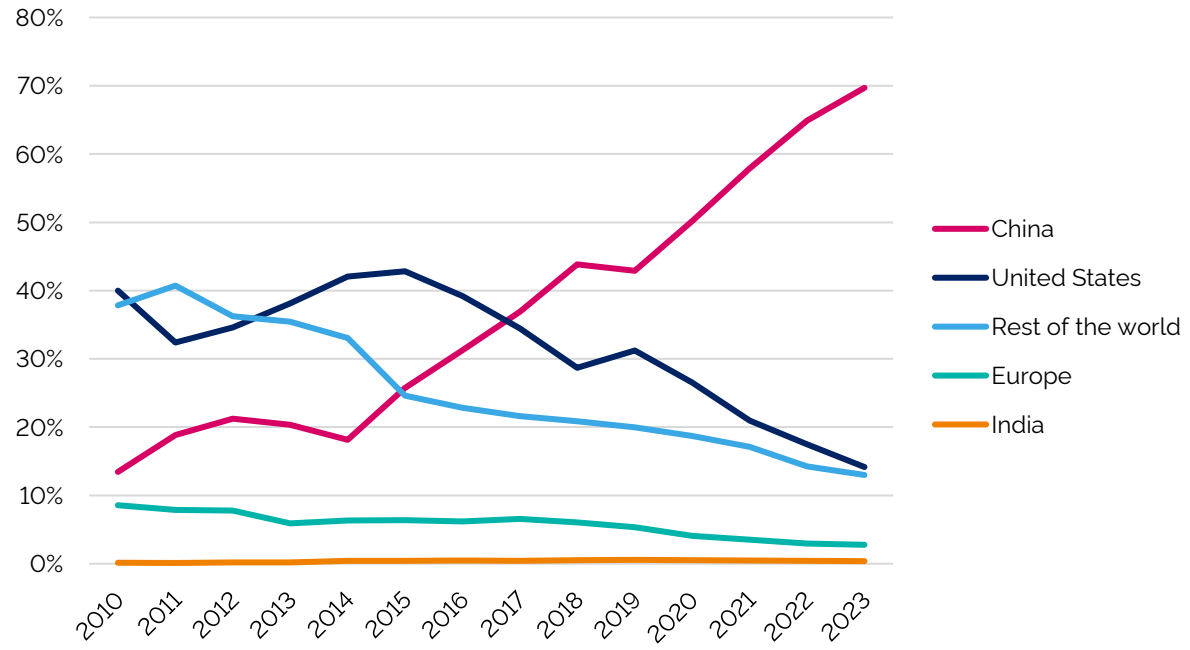
[16] *Steven de Simseo, « Black Box IA : Comprendre le mystère des algorithmes d'IA », Simseo, 30/09/2024*
[17] *Midhat Tilawat, « Rapport sur les biais de l'IA 2025 : la discrimination des LLM est pire que vous ne le pensez ! », 1 mai 2025*
[18] *Klaxoon, « L'IA générative peut-elle vraiment améliorer la collaboration en entreprise ? », 03/09/2024*
[19] *Thomas Chia, Daniel Hruschka - Educational and income disparities among minorities in China*
[20] *Thomas Chia, Daniel Hruschka - Educational and income disparities among minorities in China*
[21] *Jiafu An, Difang Huang, Chen Lin, Mingzhu Tai, "Measuring Gender and Racial Biases in Large Language Models", Cornell University, Mars 2024*
[22] *Institut Montaigne, « Algorithmes : contrôle des biais S.V.P », Mars 2020*
[23] *Notable AI Models, Epoch AI, 30/06/2025*
[24] *Hai Stanford, "The 2025 AI Index Report", 2025*
[25] *Institut Montaigne, « Algorithmes : contrôle des biais S.V.P », Mars 2020*

L'IA, un secteur dominé par la Chine et les Etats-Unis

La Chine et les Etats-Unis dominent la recherche mondiale en IA, comme l'illustre l'évolution des brevets, ce qui peut orienter la recherche.

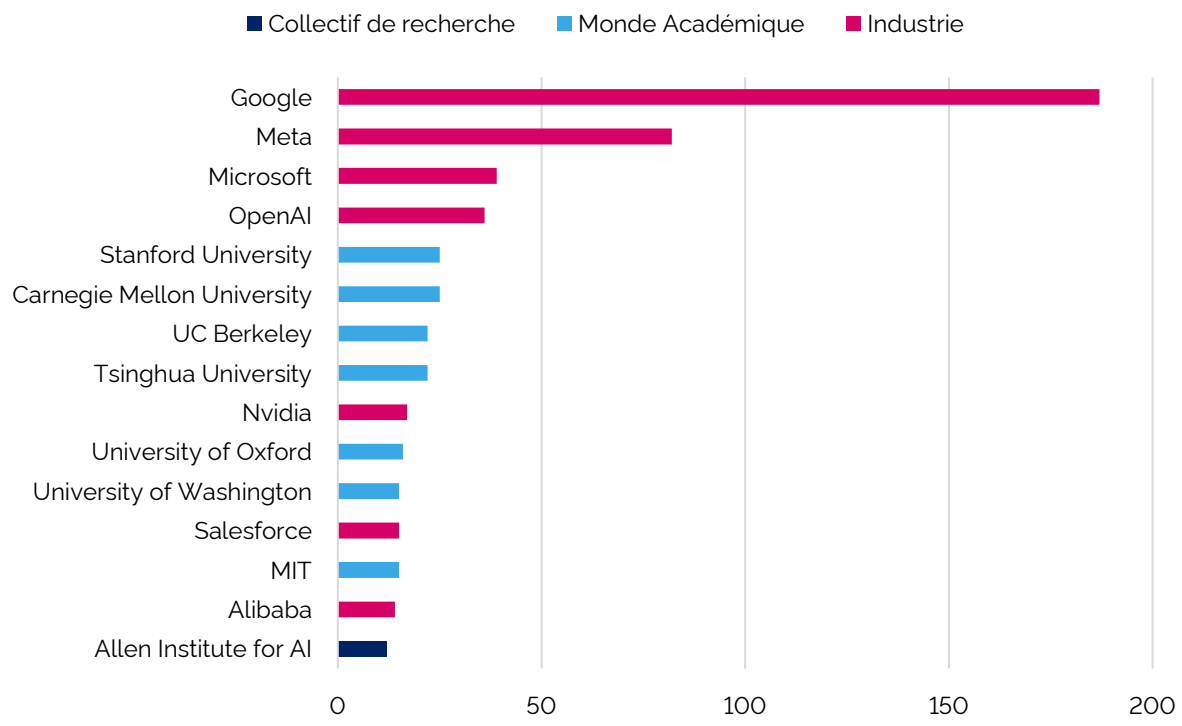
.

Brevets d'IA délivrés (% du total mondial) par zones géographiques sélectionnées, (2010-2023) - Source : Source: AI Index, 2025 | Chart: 2025 AI Index report



A ce biais géographique et culturel s'ajoute l'influence démesurée de quelques acteurs comme Google, Meta, Microsoft et OpenAI, responsables de la plupart des conceptions de modèles d'IA notables.

Nombre de modèles d'IA notables par organisation entre 2014 et 2024
Source : Epoch AI, 2025 | Chart: 2025 AI Index report



Qui est impacté par ce risque ?

Les métiers les plus concernés par ce risque[26]

Les métiers les plus concernés par l'essor de l'IA sont les plus impactés par le risque des biais algorithmiques. Ils mobilisent des compétences cognitives avancées, notamment dans les domaines :

- du développement logiciel,
- de la rédaction technique,
- de l'analyse de données
- et de la création de contenu.

Selon l'étude d'Anthropic[27], les tâches associées aux professions informatiques - développeurs, data scientists, bio-informaticiens - représentent à elles seules 37,2 % des interactions avec l'IA. Les métiers de la communication, du marketing et des médias suivent avec 10,3 %, traduisant une forte adoption des outils génératifs pour la production de textes, d'images ou de vidéos.

Ces métiers se situent majoritairement dans la tranche supérieure des salaires intermédiaires, notamment ceux nécessitant un diplôme de niveau licence. Les professions très peu qualifiées ou très spécialisées, requérant un doctorat ou une formation médicale, sont moins concernées. Cela s'explique soit par la nature physique des tâches, soit par des contraintes réglementaires.

Les métiers de la santé, de la construction ou de la logistique par exemple restent peu concernés à ce stade. En revanche, les fonctions éducatives, scientifiques et administratives optent significativement pour l'IA.

L'adoption ne se traduit pas à ce stade par une automatisation massive des métiers dans leur globalité. L'IA s'intègre progressivement tel un levier pour améliorer l'efficacité des processus. Environ 36 % des professions l'utilisent pour au moins

[26] https://www.allaboutai.com/fr-fr/ressources/statistiques-ia/biais-intelligence-artificielle/
[27] The Anthropic Economic Index, Anthropic, 10/02/2025

un quart de leurs tâches. Elles sont seulement 4 % à l'exploiter pour plus de 75 % de leurs activités. Cette intégration suggère une transformation progressive des métiers plutôt qu'une substitution brutale.

A mesure que l'IA se démocratise dans la société, et que les métiers l'adoptent, les biais que peuvent véhiculer les modèles utilisés se répandent également.

Nous proposons ce panorama des métiers les plus exposés et ayant un impact direct et immédiat sur la vie des individus :

Métiers	Conséquences potentielles du risque de biais algorithmique pour les individus	Conséquences potentielles du risque de biais algorithmique pour la société
Santé	Prise en charge des patients inadéquate augmentant les risques de mortalité et de bonne santé[28]	Réduction de l'espérance de vie et de l'espérance de vie en bonne santé[29]
Ressources Humaines	Inégalité d'accès à l'emploi dans le cas de filtrage[30]	Réduction de la productivité en raison d'une allocation sous-optimale des ressources humaines[31]
Services publics	Condamnation injuste ou sanction disproportionnée, prestation sociale non accordée, traitement de faveur[32]	Perte de confiance dans les institutions, alimentation des tensions sociales[33]
Education	Inégalités des chances, dégradation de l'estime de soi et de la santé mentale[34]	Réduction de la productivité en raison d'un sous-développement des compétences[35]

D'autres métiers victimes de ces biais peuvent également entraîner des conséquences délétères :

Métiers	Conséquences potentielles du risque de biais algorithmique pour les individus	Conséquences potentielles du risque de biais algorithmique pour la société
Médias	Dégradation de la santé mentale par la recommandation de contenus négatifs chargés émotionnellement	Amplification des biais dans la société, influence sur les décisions politiques[36]
Vente	Dégradation des finances de certains consommateurs du fait des discriminations, réduction des opportunités[37]	Accroissement des inégalités dans la société[38]
Métiers Intellectuels et créatifs	Réduction de l'esprit critique[39], réduction de la créativité[40]	Homogénéisation des idées, perte de la productivité, réduction de l'innovation[41]

[28] *Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: A call for open science. Patterns (N Y). 2021 Oct 8;2(10):100347. doi: 10.1016/j.patter.2021.100347. PMID: 34693373; PMCID: PMC8515002.*

[29] *Sécurité des patients, OMS, 11/09/2025*

[30] *Thierry Kirat, Algorithmes et discrimination dans l'emploi, Statistique & Société 13/01/2025*

[31] *Les inégalités et le monde du travail, Organisation Internationale du Travail, 2021*

[32] *Melissa Hamilton, Pamela Ugwudike, A 'black box' AI system has been influencing criminal justice decisions for over two decades – it's time to open it up, The Conversation, 26/07/2023*

[33] *Dubé, Richard et Margarida Garcia. « Enjeux politiques et juridiques autour du principe de la confiance du public dans l'administration de la justice pénale. » Les Cahiers de droit, volume 61, numéro 1, mars 2020, p. 7–34. https://doi.org/10.7202/1068779ar*

[34] *Prabha Kannan, How Harmful Are AI's Biases on Diverse Student Populations?, Stanford HAI, 03/10/2024*

[35] *Rapport mondial de suivi sur l'éducation, 2020: Inclusion et éducation : tous, sans exception, UNESCO, 09/12/2020*

[36] *Salsa Della Guitara Putri, Eko Priyo Purnomo, Tiara Khairunissa, Echo Chambers and Algorithmic Bias: The Homogenization of Online Culture in a Smart Society, SHS Web of Conferences, 2024*

[37] *Kristin Johnson, Frank Pasquale, Jennifer Chapman, Artificial Intelligence, Machine Learning, and Bias in Finance: Toward Responsible Innovation, Fordham Law Review, 2019*

[38] *Kristin Johnson, Frank Pasquale, Jennifer Chapman, Artificial Intelligence, Machine Learning, and Bias in Finance: Toward Responsible Innovation, Fordham Law Review, 2019*

[39] *Gerlich, M. AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. Societies 2025, 15, 6. https://doi.org/10.3390/soc15010006*

[40] *Anil R. Doshi, Oliver P. Hauser, Generative AI enhances individual creativity but reduces the collective diversity of novel content.Sci. Adv.10, eadn5290(2024). DOI:10.1126/sciadv.adn5290*

[41] *Anil R. Doshi, Oliver P. Hauser, Generative AI enhances individual creativity but reduces the collective diversity of novel content.Sci. Adv.10, eadn5290(2024).DOI:10.1126/sciadv.adn5290*

Focus sectoriel Finance et bancassurance

Le secteur financier est particulièrement touché par le risque des biais algorithmiques en raison de processus de plus en plus automatisés. La Banque centrale européenne identifie l'IA comme un facteur de risque systémique émergent. Les modèles déployés sans supervision humaine engendrent des biais algorithmiques qui eux-mêmes amplifient les inégalités d'accès au crédit et à l'assurance[42]. Les institutions financières présentent trois types de processus que ces biais peuvent mettre à mal :

Type de processus	Description
Evaluation du risque	Les acteurs de l'assurance sont tenus d'évaluer le risque porté par leurs clients et associés. • Décisions d'investissement faussées : les modèles prédictifs utilisés pour anticiper les tendances de marché ou les comportements clients peuvent être biaisés, entraînant des erreurs stratégiques et/ou des pertes financières (estimation erronée du profil de risque d'une cible potentielle d'investissement). • Discrimination dans l'octroi de crédit ou la couverture d'un assuré : des algorithmes peuvent refuser un prêt à des personnes en se basant sur des critères comme le lieu de résidence, (ex : quartiers jugés « à risque »), même si elles sont solvables. Ils peuvent aussi entraîner une tarification d'assurance biaisée. Les biais algorithmiques peuvent aboutir à sur ou sous-évaluer certains risques. Une institution financière peut ainsi voir sa sécurité entamée, ou perdre en compétitivité par rapport aux autres.
Evaluation de l'authenticité d'une opération ou d'une déclaration	Les institutions financières doivent faire face au risque de fraudes pouvant provenir de clients ou autres acteurs malveillants : • la fraude à l'assurance, • la fraude des moyens de paiements. Des biais algorithmiques peuvent limiter la capacité des institutions financières à détecter des fraudes, ou dans l'analyse de risque signaler injustement certaines catégories de personnes plus souvent que d'autres.
Evaluation des besoins du clients	De plus en plus de chargés de relation client ont recours à des IA pour gérer les bases d'information de l'entreprise. L'analyse des besoins clients permet de filtrer les bases de manières à identifier les offres et services adaptés. Des biais algorithmiques peuvent conduire à des recommandations inappropriées ou en décalage avec les attentes.

Ces différentes situations peuvent entrainer des conséquences sérieuses pour les entreprises :

- **Atteinte à la réputation**
  Une entreprise utilisant des algorithmes discriminatoires peut subir une perte de confiance de la part des clients et partenaires, voire des poursuites judiciaires.
- **Sanctions prudentielles**
  Les régulateurs financiers imposent de plus en plus de normes éthiques et de transparence. Un algorithme biaisé peut exposer l'entreprise à des sanctions ou à des obligations de mise en conformité coûteuses.

[42] *The rise of artificial intelligence: benefits and risks for financial stability*, ECB

Que faire face à ce risque ?

Les réponses des organisations face à ce risque

Face à ces risques, les organisations peuvent s'appuyer sur des leviers de gouvernance, technologiques, partenariaux, organisationnels et enfin assurantiels.

En matière de gouvernance, la rédaction de chartes éthiques et de règlements internes permet de guider l'élaboration des modèles et des algorithmes. Des audits et évaluations des algorithmes viennent vérifier régulièrement ce qui est fait dans l'entreprise. Enfin, les grandes organisations se dotent d'un comité d'éthique qui évalue les systèmes d'IA. Le rôle de ce comité est d'analyser les cas d'usage et d'informer les dirigeants en cas d'écart.

Dans le domaine technologique, le premier pas des organisations pour lutter contre les biais est de s'assurer de la qualité des données d'entraînement et de leur représentativité. Elles peuvent s'efforcer de rendre les IA et algorithmes plus transparents, auditables et explicables. Il est essentiel qu'elles intègrent le retour des utilisateurs pour corriger les biais au fur et à mesure qu'ils sont remontés. L'enjeu d'explicabilité est particulièrement nécessaire pour les IA issues du *machine learning* . Appelé XAI pour « Explicable AI », cela regroupe un ensemble de méthodes et de techniques visant à rendre les décisions des IA traçables et explicables[43].

Les départements des ressources humaines ont également un rôle à jouer :

- en promouvant de la diversité au sein des équipes,
- en sensibilisant et en formant les salariés afin qu'ils soient en mesure de prendre de meilleures décisions lors de la conception et l'utilisation des IA et algorithmes.

Des partenariats avec les parties prenantes - utilisateurs finaux, ONG, etc. - et le régulateur (à travers l'usage de bac à sable réglementaire) peuvent venir compléter les mesures et la vision interne de l'entreprise. La pratique du bac à sable réglementaire permet à une organisation d'expérimenter une technologie émergente. Elle bénéficie d'exemption temporaire en échange d'une collaboration étroite avec le régulateur[44].

Enfin, les organisations peuvent transférer une partie de ce risque en souscrivant une assurance qui inclut dans ses garanties le biais algorithmique et les discriminations. Le marché est relativement peu développé mais il devrait devenir plus mature au fur et à mesure que les assureurs auront davantage de recul pour quantifier le risque.

CNP Assurances applique une charte rigoureuse de gestion des risques de modèles, adaptée à leur niveau de matérialité et à leur usage, avec validation humaine systématique pour les modèles de machine learning. Les biais, la performance et l'équilibre des données sont vérifiés et documentés dès la conception, via des tests approfondis sur les bases de données d'apprentissage et de test. L'entreprise s'efforce de sensibiliser et former ses collaborateurs à ces questions dès leur intégration.

Face à généralisation de l'IA, quels futurs possibles pour le capital humain ?

Scenario 1 – Business as usual

L'IA est largement adoptée par les entreprises sur tous leurs métiers : elle participe de plus en plus à la prise de décisions.

Un évènement va marquer un tournant : le scandale autour d'un système de notation de crédit qu'utilise une néo-banque européenne. L'algorithme, entraîné sur des données historiques biaisées, défavorise systématiquement certaines classes sociales. L'affaire ne débouche sur aucune régulation, mais elle pousse plusieurs entreprises à auditer leurs modèles en interne. Les grandes entreprises, devenues conscientes des impacts des risques liés aux biais algorithmiques, prennent des mesures de gouvernance et se dotent de règlements internes.

En 2030, dans de nombreux secteurs, les opérateurs humains ne prennent plus les décisions eux-mêmes et jouent progressivement un rôle de superviseur dans les processus. Ce rôle devient la norme, sans réellement garantir une réelle capacité de contrôle.

Le marché des données se transforme. Elles sont échangées à prix d'or entre entreprises, ce qui accélère leur privatisation. Les directives européennes se voulant favorables à l'innovation freinent la tendance sans y mettre un terme. Les modèles

[43] *Tech Dispatch on Explainable Artificial Intelligence, European Data Protection Supervisor, 2023*
[44] *The role of sandboxes in promoting flexibility and innovation in the digital age, OCDE, 13/07/2023*

d'IA sont entraînés sur des corpus propriétaires, souvent peu diversifiés. Cette pratique permet de réduire les influences des lobbys et des groupes malveillants. Mais elle ralentit le renforcement du cadre réglementaire autour du risque. La domination d'un nombre réduit de créateurs de modèles et l'origine des données perpétuent encore certains biais dans les sociétés.

À l'approche de 2040, l'IA est omniprésente. L'équilibre entre performance et équité reste fragile, maintenu par une vigilance diffuse, plus culturelle que réglementaire.

Scenario 2 – Confiance aveugle en l'IA

L'enthousiasme autour de l'IA atteint un nouveau sommet. Les performances sont telles que les humains doutent peu à peu de leur propre légitimité à remettre en question les décisions des machines.

L'Union européenne adopte une loi visant à en encadrer l'usage dans la rédaction d'ouvrages et la production de l'information. Le texte ne parvient néanmoins pas à freiner le phénomène de désinformation. La frontière entre fiction, opinion et vérité devient floue. Paradoxalement, l'adhésion du public reste intacte.

En 2030, plusieurs États, séduits par les gains des systèmes automatisés, choisissent de déréguler partiellement le secteur de l'IA. Les garde-fous sont levés au nom de l'innovation. Des lobbys et des acteurs étatiques exploitent l'IA comme levier de propagande pour faire avancer leurs intérêts. L'opinion publique est orientée et le jeu démocratique faussé. Les débats s'appauvrissent. Les idées s'homogénéisent sous le poids d'une pensée algorithmique dominante.

Des études indépendantes révèleront en 2035 que les systèmes d'IA utilisés dans les services publics de plusieurs pays discriminent systématiquement certaines minorités. La « neutralité technique » des algorithmes, invoquée par les autorités, suffit à étouffer l'indignation. Les victimes sont réduites au silence face à des décisions opaques et automatisées.

En 2040, l'IA est devenue l'interface principale entre les individus et le monde. Elle filtre l'information, oriente les choix, structure les récits. Les humains, convaincus de l'objectivité des machines, ont cessé de les interroger. La créativité s'est raréfiée, les imaginaires se sont alignés. Les biais, eux, sont toujours là — invisibles, intégrés, normalisés.

Scenario 3 – Renforcement de l'esprit critique

Les révélations sur les biais algorithmiques se succèdent. D'abord relayées par des médias spécialisés, elles finissent par toucher le grand public. Plusieurs États européens lancent une initiative pour encadrer l'usage des systèmes d'IA dans les services publics. Ce cadre impose des audits réguliers et la publication d'indicateurs de fiabilité.

Dès 2030, les grandes organisations intègrent des mécanismes permettant de contester une décision algorithmique, d'en obtenir une explication, voire de la faire réviser par un humain. Une règle perçue comme une contrainte, devient rapidement un gage de sérieux. Les programmes scolaires introduisent un enseignement dédié à l'esprit critique numérique. Les élèves apprennent à remettre en question les résultats des algorithmes pour en identifier les limites. Une nouvelle génération plus lucide et exigeante émerge.

Dans le même temps, les institutions internationales prennent position avec une charte mondiale sur l'usage éthique de l'IA signée par plus de 80 pays.

À l'approche de 2040, les biais n'ont pas disparu, mais ils sont visibles, discutés, corrigés. L'IA est devenue un outil puissant, mais encadré. L'humain n'est plus un simple utilisateur passif : il est acteur, superviseur, et parfois contradicteur. La société a appris à vivre avec l'IA sans s'y soumettre, en cultivant un équilibre fragile mais précieux entre confiance et vigilance.

Protégez l'environnement. N'imprimez ce document que si nécessaire

Crédit photo première page : Getty / nortonrsx