

Adhoc du cep

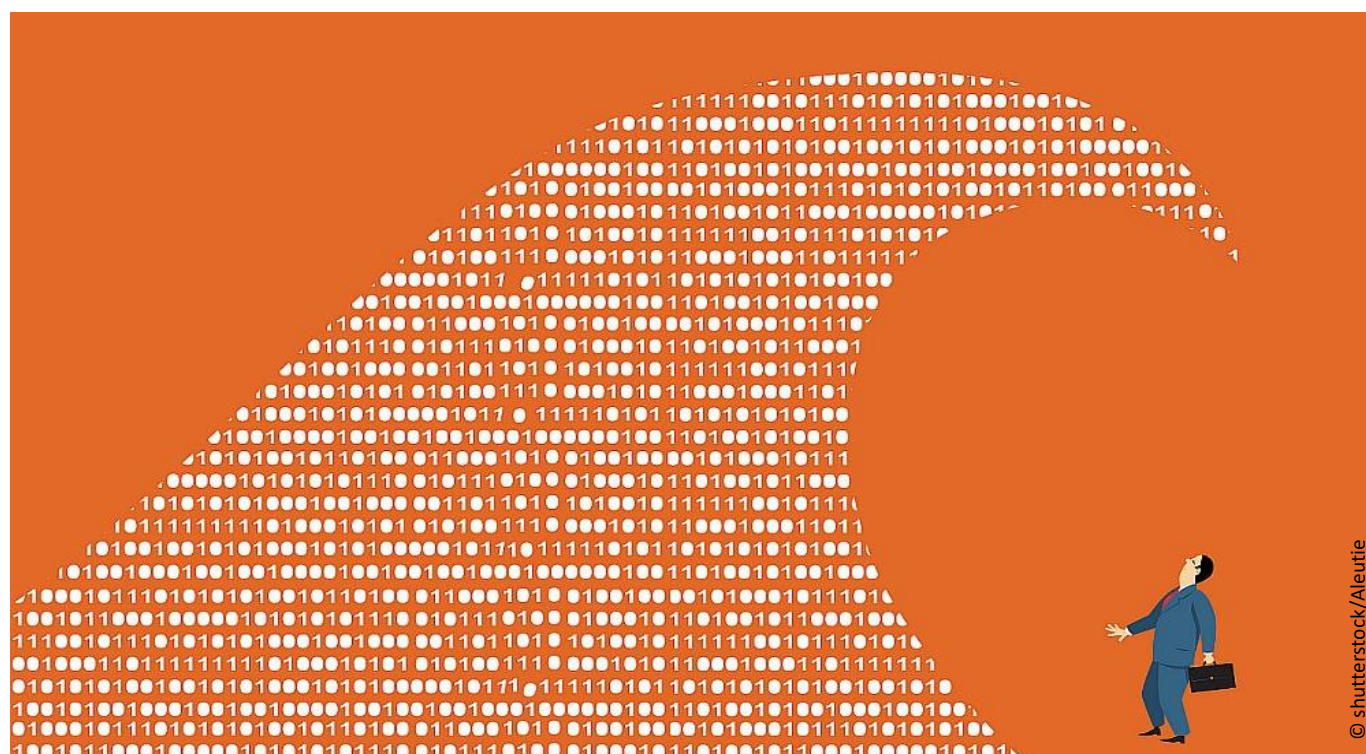
No 15 | 2022

13 décembre 2022

L'IA comme risque systémique dans une poly-crise

Le danger de la prédiction algorithmique dans les environnements inconnus

Anselm Küsters



Le pouvoir tant vanté de l'intelligence artificielle (IA) dépend, dans de nombreux domaines d'application, des données avec lesquelles elle est formée. Comme notre monde de plus en plus dynamique et interdépendant s'écarte des observations passées utilisées pour la conception et la formation, les systèmes d'IA deviendront moins fiables. Pour renforcer la résilience européenne, la proposition de loi européenne sur l'IA devrait être modifiée pour tenir compte de la faillibilité accrue des systèmes d'IA en période de poly-crise, principalement lorsque le secteur public les utilise.

- ▶ En ces temps de désordre mondial, certains espèrent que les systèmes d'IA prédictifs pourront se frayer un chemin à travers le bruit et aider, par exemple, à prévenir la prochaine pandémie ou à économiser des ressources pour lutter contre le changement climatique.
- ▶ Cependant, comme les événements du monde réel s'écartent de plus en plus des points de données utilisés pour faire des prédictions algorithmiques, les systèmes d'IA deviendront moins fiables et nécessiteront une surveillance accrue. Plusieurs exemples récents illustrent cette tendance inquiétante.
- ▶ Plutôt que de peaufiner les algorithmes pour les crises actuelles, les décideurs politiques devraient établir des conditions cadres solides qui, sans maximiser l'efficacité, permettront aux économies européennes axées sur la technologie de fonctionner raisonnablement lorsque la prochaine crise frappera. C'est pourquoi le Center for European Policy Berlin/Fribourg (cep Berlin/Fribourg) appelle à une classification des systèmes d'IA en fonction du contexte dans la proposition de loi sur l'IA et à la divulgation des résultats d'audits spécifiques sur l'IA pour une évaluation par les pairs.

Table des matières

1	Introduction : L'IA à la rescousse ?	3
2	Fond : le cygne noir rencontre le rhinocéros gris	3
3	Exemples : pourquoi les outils d'IA fonctionnent mal en période anormale.....	5
4	Perspectives : comment atténuer les risques de fausses prédictions algorithmiques	7
5	Conclusion : la nécessité de règles solides	10

1 Introduction : L'IA à la rescousse ?

Le pouvoir de l'intelligence artificielle (IA) est indubitablement séduisant, il n'est donc pas surprenant qu'en période de désordre mondial, certains expriment l'espoir que les algorithmes puissent aider à prévenir la prochaine pandémie,¹ économiser des ressources pour lutter contre le changement climatique,² ou même faire émerger une future société plus équitable et inclusive.³ Ce que l'on oublie souvent, c'est que tout système d'IA est fondamentalement limité par le type de données avec lesquelles il est entraîné. En règle générale, les algorithmes modernes d'apprentissage automatique fonctionnent mieux lorsqu'ils sont entraînés sur de grandes quantités de données de haute qualité et sont moins performants lorsque les données sous-jacentes sont inexactes, incomplètes, non pertinentes, invalides, périmées ou incohérentes.⁴ Qu'implique cette relation fondamentale pour notre époque actuelle de désordre mondial, caractérisée par de multiples crises allant de la guerre de la Russie contre l'Ukraine aux pénuries d'énergie, à l'inflation galopante et au chaos climatique ?

Comme les événements du monde réel s'écartent de plus en plus des points de données utilisés pour faire des prédictions algorithmiques, les systèmes d'IA deviendront moins fiables et nécessiteront une surveillance accrue. Il est crucial de reconnaître les limites de ces systèmes, qui découlent des frontières épistémiques. Par conséquent, les politiciens et les régulateurs doivent faire face à une réalité qui donne à réfléchir : cette technologie en soi ne nous sauvera pas des récessions, des guerres ou du changement climatique. Ce dont nous avons besoin, c'est d'un cadre réglementaire solide, à commencer par la loi sur l'IA actuellement en projet, et d'une réflexion fondée sur le bon sens, que seuls les humains possèdent jusqu'à présent.

2 Arrière-plan : le cygne noir rencontre le rhinocéros gris

Prendre conscience de la viabilité limitée des modèles quantifiés en cas de changement soudain est un point de départ précieux pour réfléchir au rôle des systèmes d'IA dans le monde chaotique d'aujourd'hui. Les limites apparentes des modèles économiques classiques au cours des premiers mois de la crise financière de 2007-2008 en sont un excellent exemple : « Face à la crise », a déploré Jean-Claude Trichet, alors président de la Banque centrale européenne, en novembre 2010, « nous nous sommes sentis abandonnés par les outils conventionnels ». Ces outils avaient été formulés et affinés au cours d'une période de volatilité macroéconomique remarquablement faible, connue sous le nom de « Grande Modération »⁵ - et étaient désormais surchargés par la propagation des turbulences sur les marchés. En l'absence d'orientations claires dans les cadres analytiques existants, les banquiers centraux du monde entier se sont tournés vers les épisodes de crise historiques pour en tirer des leçons qualitatives.⁶

¹ Makin, S. (2022), [Could an algorithm predict the next pandemic ?](#), Nature Outlook (26.10.2022),

² Miller, K. (2022), [Building Intelligent Agents to Reach Net Zero 2050](#), HAI Stanford University (3.10.2022),

³ Lobel, O. (2022), *The Equality Machine: Harnessing Digital Technology for a Brighter, More Inclusive Future*, New York: Public Affairs.

⁴ Pogrebivsky, S. (2021), [The 6 Attributes of High-Quality Data](#), DataGroomr (8.7.2021),

⁵ Bean, C. (2010), *The Great Moderation, the Great Panic, and the Great Contraction*, Journal of the European Economic Association 8 (2-3), pp. 289-325.

⁶ Eichengreen, B. (2015), *Hall of mirrors: the Great Depression, the Great Recession, and the uses-and misuses-of history*, Oxford: Oxford University Press.

L'observation selon laquelle la plupart de ces économistes ont fini par utiliser des analogies avec la Grande Dépression des années 30,⁷ illustre le fait que les repères utiles pour s'orienter sont rares lorsqu'on cherche à tirer des leçons du passé en période de crise. Les statistiques de l'histoire économique montrent que les grandes crises financières sont des événements rares : En moyenne, les crises se produisent tous les 25 ans, et les nouvelles récessions commencent généralement tous les huit ans.⁸ Cela limite considérablement toute tentative empirique, y compris l'entraînement de systèmes d'IA, par exemple, pour prédire de tels événements. En outre, même si des données sur ces crises existent, elles sont presque par définition « sales », comme les appellent les spécialistes des données, ce qui signifie qu'elles manquent d'informations ou ne mesurent pas la réalité aussi précisément que dans un contexte normal. Les modèles conçus ou entraînés en vue d'une « période normale » peuvent donc commencer à échouer lorsque des chocs externes importants se produisent et marquent le début d'une « période anormale ».

En raison de la nature inhérente du risque, ce problème ne peut être entièrement anticipé ou contourné, un phénomène qui a conduit l'ancien trader de Wall Street Nassim Nicholas Taleb à inventer le terme « cygnes noirs ».⁹ Les cygnes noirs sont des événements imprévus et imprévisibles aux conséquences extrêmes et, en tant que tels, ils ne peuvent jamais être entièrement couverts par nos modèles. Outre la crise financière de 2007-2008, des événements plus récents, comme la pandémie de Covid et l'attaque de la Russie contre l'Ukraine, entrent également dans cette catégorie. L'utilisation de systèmes d'IA pour aider les décideurs dans de telles conditions anormales implique généralement des « critères contradictoires et incomparables », par exemple le coût par rapport au bien-être humain.¹⁰ En outre, les économistes ont utilisé l'expression « rhinocéros gris » pour désigner des risques bien connus et à évolution lente qui peuvent amplifier les chocs externes, comme le niveau élevé actuel d'endettement des ménages ou le changement climatique mondial.¹¹ Si les cygnes noirs et les rhinocéros gris compliquent toute prédiction, y compris le raisonnement humain, les applications d'IA alimentées par les mathématiques qui circulent aujourd'hui sont beaucoup plus opaques que les systèmes de prédiction traditionnels, généralement incontestables ou sans appel, et peuvent amplifier les boucles de rétroaction négative sur de grandes distances en peu de temps.¹² L'utilisation accrue de ces applications crée donc de nouveaux risques systémiques.¹³

Les choses se compliquent encore lorsque plusieurs chocs se produisent simultanément et établissent des interconnexions qui ne sont généralement pas prévues. L'historien de l'économie Adam Tooze a récemment décrit l'état actuel du monde comme une « poly-crise », désignant une interaction de chocs disparates qui, dans son ensemble, est pire que la somme de ses parties.¹⁴ L'image qui se dégage de ce bref tour d'horizon est claire : la présence de multiples cygnes noirs et rhinocéros gris dans la polycrise actuelle risque fort d'avoir un impact sur l'utilité de nos algorithmes qui sont, par nécessité, formés sur des données passées, qui pourraient maintenant être rapidement dépassées. Ce problème

⁷ Küsters, A. (2022). Applying Lessons from the Past ? Exploring Historical Analogies in ECB Speeches through Text Mining, 1997-2019. *International Journal of Central Banking* 18 (1), pp. 277-329.

⁸ Paul, P. (2019), Modeling Financial Crises, FRBSF Economic Letter (2019-08).

⁹ Taleb, N. (2007), *The Black Swan : The Impact of the Highly Improbable*, New York : Random House.

¹⁰ Mostaghim, S. (2020), [AI to the Rescue : Life-and-Death Decision-Making under Conflicting Criteria](#), Site web du projet (non daté),

¹¹ Marja Nykänen, M. (2022), [Black swans and grey rhinos - lessons of crises on macroprudential policy, discours d'ouverture de la conférence sur l'analyse du risque systémique](#), Helsinki (5.5.2022),

¹² O'Neil, C. (2016), *Weapons of Math Destruction : Comment le Big Data accroît les inégalités et menace la démocratie*, New York : Crown Publishers.

¹³ Galaz, V. et al. (2021), Artificial intelligence, systemic risks, and sustainability, *Technology in Society* 67, 101741.

¹⁴ Tooze, A. (2022), [Welcome to the world of the polycrisis](#), FT (28.10.2022),

d'entraînement des modèles sur des ensembles de données plus restreints que la population qu'ils sont censés refléter en fin de compte est une forme de « fuite de données », qui menace la fiabilité de l'apprentissage automatique dans toutes les disciplines.¹⁵ Pire encore, il semble que l'utilisation de systèmes d'IA prédictifs pour éclaircir une situation en temps de crise pourrait même exacerber une situation déjà mauvaise.

3 Exemples : pourquoi les outils d'IA fonctionnent mal en période anormale

Les dangers peuvent être démontrés en remontant brièvement le temps jusqu'aux premiers mois de la pandémie de Covid. Dès que les premiers blocages ont commencé, les économistes ont fait des pieds et des mains pour se faire une idée de l'état actuel de l'économie mondiale en s'appuyant sur des données inédites en temps réel, car leurs mesures habituelles étaient soit trop lentes, soit peu fiables. Au lieu d'attendre les estimations officielles de l'inflation et du chômage, ils se sont de plus en plus appuyés sur des indicateurs auparavant obscurs, tels que les statistiques de mobilité d'Apple ou de Google, ou les données sur les réservations dans les restaurants, pour se faire une idée de l'activité économique globale.¹⁶ Si ces mesures ont permis de saisir les changements de comportement des consommateurs pendant la pandémie, les systèmes d'IA formés sur des données correspondant aux anciens schémas de dépenses ont immédiatement rencontré des problèmes.

Prenons le cas de Fair Isaac Corp (FICO), un développeur de logiciels basé aux États-Unis dont les outils d'IA réputés pour la détection des fraudes par carte de crédit et de débit sont utilisés par les grandes banques pour prendre leurs décisions de prêt. Au début de la pandémie, sur la base de l'expérience passée, ces outils s'attendaient à beaucoup plus d'achats en personne que virtuels, une situation qui a finalement conduit à ce qu'un grand nombre de transactions virtuelles soient signalées comme problématiques. Cela signifie que l'algorithme sous-jacent a recommandé de refuser des millions d'achats légitimes, tandis que les consommateurs enfermés tentaient avec empressement de se procurer des biens essentiels en ligne.¹⁷ À peu près à la même époque, en Chine, les outils d'évaluation du crédit automatisés par l'IA de grands acteurs tels que Ant Group, qui avaient continuellement maintenu les taux de défaillance à un pour cent, ont également beaucoup moins bien fonctionné.¹⁸ Cela s'explique en partie par l'augmentation du stress financier au cours de la première pandémie. Toutefois, les nouveaux problèmes sont probablement aussi dus au fait que les scores d'Ant ne se fondaient pas seulement sur des facteurs traditionnels comme l'historique de crédit, mais aussi sur des critères beaucoup plus larges, tels que les loisirs et les préférences d'achat d'un utilisateur - en d'autres termes, des mesures comportementales qui ont connu des changements spectaculaires en raison de l'auto-isollement.

Dans d'autres contextes de crise, comme la santé ou le changement climatique, qui tendent à ressembler davantage à des risques de « rhinocéros gris » qu'à des « cygnes noirs », le problème n'est pas le changement soudain des indicateurs critiques, mais le manque de données de haute qualité qui affaiblit la prédiction algorithmique. En raison des défis relatifs à la taille et à la composition des données utilisées pour entraîner les modèles d'IA pour les soins de santé, ces derniers ne sont pas aussi précis

¹⁵ Gibney, E. (2022), [Could machine learning fuel a reproducibility crisis in science ?](#), Nature Outlook (26.7.2022),

¹⁶ The Economist (2020), [Why real-time economic data need to be treated with caution](#) (23.7.2020),

¹⁷ Dave, P. (2022), [When the AI goes haywire, bring on the humans](#), Reuters (13.10.2022),

¹⁸ Chorzempa, M. (2022), The Cashless Revolution : China's Reinvention of Money and the End of America's Domination of Finance and Technology, New York : Public Affairs, pp. 98 et suivantes, 103.

pour prédire les maladies que les rapports le suggèrent généralement.¹⁹ De même, comme la qualité et la quantité des données météorologiques et autres données géographiques recueillies varient d'un pays à l'autre, les modèles de prévision des risques climatiques pour certaines zones peuvent être erronés.²⁰ En général, les petites communautés à faible revenu qui sont confrontées aux plus grandes pertes financières et aux plus grands risques dans un monde de changement climatique²¹ ne disposent pas du type de données climatiques de haute qualité nécessaire à la formation de systèmes d'alerte précoce capables de réagir rapidement. Par exemple, les chercheurs en énergie critiquent le fait que les plans actuels de lutte contre le changement climatique fondés sur les données échouent en raison de leur manque de données et de modèles spécifiques à l'Afrique.²² Ce fossé entre ceux qui ont des données et ceux qui n'en ont pas entrave la collaboration²³ et peut renforcer les inégalités entre les communautés, car un système d'IA adaptera ses recommandations aux zones où il peut s'appuyer sur une plus grande quantité de données.

De fait, les experts se rendent de plus en plus compte que l'utilisation de l'analyse prédictive dans des situations de complexité, de désordre ou de données insuffisantes peut même renforcer les tendances négatives sous-jacentes. Par exemple, pour gérer les retombées de la crise des opioïdes, les États-Unis s'appuient sur un important algorithme de risque de toxicomanie qui, comme les chercheurs l'ont maintenant documenté,²⁴ semble aggraver la situation. Les outils sous-jacents de dépistage des patients basés sur l'IA sont intrinsèquement défectueux car ils ne peuvent pas gérer des constructions complexes comme la santé humaine et, par conséquent, refusent souvent le traitement aux patients les plus vulnérables ou représentant des cas médicalement complexes. La prise de conscience du potentiel de ces systèmes automatisés à se transformer en « armes de destruction mathématique »²⁵ est lourde de conséquences lorsqu'on réfléchit à la réglementation : « Si les analyses prédictives créent en partie la réalité qu'elles prétendent prédire », comme l'a récemment conclu Carissa Véliz, chercheuse à l'université d'Oxford, « alors elles sont en partie responsables des tendances négatives que nous connaissons à l'ère numérique, de l'augmentation des inégalités à la polarisation, en passant par la désinformation et le préjudice causé aux enfants et aux adolescents ».²⁶ La polycrise actuelle va exacerber ces problèmes fondamentaux qui se cachent derrière les prédictions algorithmiques.

Un dernier exemple tiré du domaine de la sécurité illustre ce constat inquiétant. Une étude récente publiée par l'Agence des droits fondamentaux de l'Union européenne critique l'utilisation accrue des systèmes d'IA pour le maintien de l'ordre prédictif.²⁷ Là encore, le problème des outils basés sur l'IA – dans ce cas pour les forces de l'ordre – est qu'ils s'appuient généralement sur des données historiques et donc potentiellement périmées et biaisées, l'IA apprenant les modèles émergents. Les « boucles de rétroaction » qui en résultent sont particulièrement dommageables en cas de poly-crise. Par exemple, une présence policière réduite dans une zone se traduit généralement par une diminution du nombre

¹⁹ Berisha, V. / Julie Liss, J. (2022), [AI in Medicine Is Overhyped](#), Scientific American (19.10.2022),

²⁰ Center for Data Innovation (2022), [How Does the Data Divide Impact Global Policy Challenges](#) ? discussion en ligne (7.12.2022),

²¹ Naddaf, M. (2022), [Climate change is costing trillions - and low-income countries are paying the price](#), Nature News (7.11.2022),

²² Mutiso, R. (2022), [Net-zero plans exclude Africa](#), Nature World View (2.11.2022),

²³ Diebold, G. (2022), [Data Divide or Digital Divide ? Ou les deux ?](#), ISE Magazine (7.11.2022),

²⁴ Szalavitz, M. (2021), [La douleur était insupportable. Alors pourquoi les médecins l'ont-ils repoussée ?](#), Wired (11.8.2021),

²⁵ O'Neil, C. (2016), Weapons of Math Destruction : Comment le Big Data accroît les inégalités et menace la démocratie, New York : Crown Publishers.

²⁶ Véliz, C. (2021), [If AI Is Predicting Your Future, Are You Still Free ?](#), Wired (27.12.2021),

²⁷ FRA (2022), [Bias in algorithms - Artificial intelligence and discrimination](#), Luxembourg : Office des publications de l'Union européenne,

d'inculpations, ce qui conduit à une diminution du nombre de policiers dans cette zone à l'avenir, car l'IA recommande d'allouer des ressources à d'autres sites qu'elle juge plus critiques. Cette situation rend à son tour la zone en question plus vulnérable aux chocs externes, tels que les attaques contre les infrastructures critiques ou le terrorisme, puisque le modèle d'IA n'intègre que des données sur les accusations criminelles passées et ne peut pas penser stratégiquement à d'autres aspects pertinents, tels que l'emplacement d'une gare fréquemment visitée ou d'un fournisseur d'énergie critique. Le même problème peut être détecté pour les systèmes d'IA prédictifs utilisés dans les domaines de la migration, de l'asile et du contrôle des frontières, qui deviennent de plus en plus pertinents en période de mobilité humaine induite par le changement climatique, mais dont les évaluations incorrectes « ont des conséquences importantes sur l'état de préparation des États membres, mais aussi sur la probabilité que les individus puissent accéder à une protection internationale ».²⁸ En d'autres termes, l'utilité des outils de prédiction basés sur l'IA est limitée à des domaines étroits - axés, par exemple, sur la finance, l'environnement, la médecine ou la sécurité - à l'exclusion de tous les autres. Comme l'a noté Martin Wolf, penser le monde en silos intellectuels peut être efficace dans un monde raisonnablement stable, mais cela échouera inévitablement dans une poly-crise.²⁹

4 Perspectives : comment atténuer les risques de fausses prédictions algorithmiques

Dans une certaine mesure, la collecte de nouvelles données, l'augmentation du partage des données au niveau mondial et l'adoption de normes communes en matière de données peuvent améliorer la couverture géographique et chronologique, ce qui permet d'obtenir de meilleurs modèles quantitatifs. Une autre piste pourrait être l'apprentissage par renforcement, qui ne dépend pas d'ensembles de données externes mais d'informations créées pendant la formation. Certains chercheurs espèrent développer des systèmes d'IA plus résilients basés sur un type de raisonnement statistique connu sous le nom de « planification séquentielle dans l'incertitude ».³⁰ Toutefois, ces solutions n'offrent pas de panacée, car tous les ensembles de données sont imparfaits d'une manière ou d'une autre et des « cygnes noirs » peuvent toujours se produire à tout moment. En outre, la rapidité et l'interconnexion croissantes des crises font qu'il est pratiquement impossible de mettre à jour les données à une fréquence suffisante. Il est révélateur que même les experts qui espèrent que l'apprentissage automatique pourra aider à identifier les virus les plus dangereux concèdent que les outils prédictifs ne pourront pas empêcher la prochaine pandémie.³¹ De même, les experts suggèrent qu'il n'existe « aucune solution miracle » pour produire des modèles d'IA fiables et cliniques.³²

En résumé, la réponse à la question de savoir si nos modèles peuvent espérer agir sur un monde de plus en plus complexe se résume à savoir si nous apprenons simplement aux algorithmes à répéter des choses qu'ils connaissent déjà ou s'ils sont capables d'apprendre des principes entièrement nouveaux - une question qui fait encore l'objet d'un vif débat parmi les informaticiens et les philosophes, et qui ne sera probablement pas résolue de sitôt. Compte tenu de cette insécurité et de cette ambivalence, il ne faut pas confier aux algorithmes tant vantés le soin de résoudre les crises actuelles ou d'assurer la résilience. Il faut au contraire adopter une approche plus nuancée. La meilleure défense contre les

²⁸ Access Now et al. (2022), [Lettre ouverte : La loi européenne sur l'IA doit protéger les personnes en déplacement](#), p. 3.

²⁹ Wolf, M. (2022), [How to think about policy in a polycrisis](#), FT (29.11.2022),

³⁰ Miller, K. (2022), [Building Intelligent Agents to Reach Net Zero 2050](#), HAI Stanford University (3.10.2022),

³¹ Makin, S. (2022), [Could an algorithm predict the next pandemic ?](#), Nature Outlook (26.10.2022),

³² Berisha, V. / Julie Liss, J. (2022), [AI in Medicine Is Overhyped](#), Scientific American (19.10.2022),

risques de la prédiction algorithmique en période de désordre mondial est de maintenir un paysage réglementaire simple et robuste. Les règles ne doivent pas être si détaillées et englobantes que les start-ups innovantes ne puissent pas émerger en premier lieu. Elles doivent plutôt définir des conditions-cadres permettant au marché de distinguer à long terme les outils de prédiction basés sur l'IA qui sont sensés de ceux qui ne le sont pas, sans que ces applications puissent causer des dommages à la société à court terme.

Ce cadre réglementaire robuste devrait notamment obliger davantage d'équipes d'entreprise à examiner les performances de l'IA afin d'éviter des problèmes tels que ceux signalés par FICO. Alors que la nouvelle loi européenne sur l'IA exigera un certain contrôle, il reste encore beaucoup à faire : une enquête de McKinsey auprès de 1 843 entreprises en 2021 suggère que la plupart ne contrôlent pas régulièrement les programmes basés sur l'IA après leur lancement.³³ Même les entreprises disposant d'équipes d'IA responsables n'investissent pas assez, ce qui a pour conséquence de faire souffrir les personnes travaillant dans ce domaine.³⁴ Les entreprises devraient être légalement obligées de publier leurs audits de l'IA, ce qui inciterait davantage à dépenser plus de ressources pour enquêter sur les méfaits de l'IA. Si ces audits ont été suggérés pour accroître la responsabilité des algorithmes, leur mise en œuvre correcte nécessite des orientations réglementaires sur des normes suffisantes et des pratiques communes.³⁵ Étant donné que la loi sur l'IA et la loi sur la responsabilité en matière d'IA actuellement prévues par l'UE obligeront les entreprises à documenter la manière dont elles atténuent les préjudices, des normes clairement définies et davantage de ressources humaines et techniques sont nécessaires pour éviter un nouvel « épuisement » dans le secteur de l'audit, qui pourrait exacerber les préjudices causés par l'IA au lieu de les atténuer. Une solution possible serait de rendre obligatoire la divulgation des éléments critiques des conclusions d'un audit pour un examen par les pairs.³⁶

Un point de départ crucial pour l'adaptation sera les négociations en trilogue de l'année prochaine entre la Commission européenne, le Conseil et le Parlement sur la loi sur l'IA, qui est l'initiative phare européenne pour réglementer les systèmes d'apprentissage automatique à l'ère numérique. Ses règles devraient être modifiées pour tenir compte de la faillibilité accrue des systèmes d'IA en période de polycrise. En général, l'imprévisibilité des événements de type « cygne noir » et « rhinocéros gris » suggère qu'une approche purement fondée sur le risque pourrait ne pas être suffisante, car nous ne pouvons pas connaître le risque global d'un système donné. Toutefois, si l'on accepte comme acquise l'approche fondée sur le risque du projet actuel, on pourrait intégrer les dangers qui surviennent en période de poly-crise en classant une proportion plus élevée de systèmes pilotés par l'IA comme « à haut risque » chaque fois que le climat économique ou politique actuel suggère que leurs données d'apprentissage pourraient être déconnectées de la réalité. Les systèmes à haut risque au sens de la loi sur l'IA sont ceux qui peuvent avoir un impact significatif sur les chances de vie d'un utilisateur, et sont donc tenus de certifier, entre autres, des données de formation de haute qualité, une supervision humaine adéquate, et des tests de précision et de robustesse. Le texte de compromis le plus récent pour la loi sur l'IA énumère huit types concrets de systèmes qui entrent dans cette catégorie, tels que

³³ McKinsey Analytics (2021), The state of AI in 2021 (décembre).

³⁴ Heikkilä, M. (2022), [Responsible AI has a burnout problem](#), MIT Technology Review (28.10.2022),

³⁵ Costanza-Chock, S. / Raji, I. / Buolamwini, J. (2022), Who Audits the Auditors ? Recommendations from a Field Scan of the Algorithmic Auditing Ecosystem, ACM Conference on Fairness, Accountability, and Transparency, pp. 1571-1583.

³⁶ Costanza-Chock, S. / Raji, I. / Buolamwini, J. (2022), Who Audits the Auditors ? Recommendations from a Field Scan of the Algorithmic Auditing Ecosystem, ACM Conference on Fairness, Accountability, and Transparency, p. 1579.

les systèmes automatisés pour la formation professionnelle ou l'application de la loi.³⁷ Toutefois, un système de classification sensible au contexte permettrait d'inclure davantage de systèmes d'IA dans ce régime réglementaire en période de polycrise, garantissant ainsi que ces systèmes doivent répondre à des normes plus élevées en matière de données et de robustesse. Par exemple, on pourrait exiger que toutes les applications d'IA entrant dans la catégorie des « systèmes à risque limité », qui ne nécessitent en temps normal que des obligations de transparence rudimentaires, soient soumises aux obligations supplémentaires applicables aux systèmes à « risque élevé » lorsque des chocs externes actuels, tels qu'une guerre, une pandémie ou la perte d'une infrastructure critique, augmentent la probabilité de décisions erronées de la part des modèles préformés. Alors que les grandes entreprises ou le secteur public pourraient comparativement facilement engager plus de juristes ou d'informaticiens pour mettre en œuvre les exigences relatives aux systèmes à haut risque, les petites entreprises ou les start-ups n'ont pas ces possibilités, ce qui doit être pris en compte.

La position de négociation du Conseil européen sur la loi sur l'IA, adoptée le 6 décembre 2022, est donc un pas dans la mauvaise direction, car elle cherche à assouplir les exigences pour les systèmes à haut risque.³⁸ En ce qui concerne la classification des systèmes d'IA en tant que systèmes à haut risque, la proposition de compromis inclut désormais une couche horizontale supplémentaire pour garantir que les systèmes d'IA qui ne sont pas susceptibles de provoquer des violations graves de la santé, de la sécurité ou des droits fondamentaux (parce que leur sortie n'est qu'accessoire à une action ou une décision) sont exemptés. Si cette mesure est en principe bienvenue car elle soulagera les start-up innovantes, le règlement horizontal devrait en outre tenir compte de l'évolution des risques attendus en cas de chocs externes ou de crises multiples. L'art. 6(3) du projet de loi sur l'IA devrait donc être formulé de manière à exempter les systèmes dont le résultat est totalement immatériel par rapport à l'action ou à la décision à prendre et, par conséquent, peu susceptible d'entraîner un risque significatif pour la santé, la sécurité ou les droits fondamentaux, *même dans le contexte d'une poly-crise*, c'est-à-dire de chocs externes multiples et simultanés. Pour les start-ups, il est particulièrement problématique que les exigences horizontales envisagées ne soient émises qu'au moins un an après l'entrée en vigueur de la loi sur l'AI.³⁹ Par conséquent, le cep demande que le Parlement européen, dans son positionnement attendu début 2023, se concentre sur les risques supplémentaires de la prédiction algorithmique survenant en cas de crise et demande une mise en œuvre plus rapide de la règle horizontale.

Une autre possibilité de prendre en compte les effets de la poly-crise dans la loi sur l'IA est offerte par l'enregistrement envisagé des systèmes d'IA. Selon le projet initial de la Commission, les fournisseurs de systèmes d'IA à haut risque devraient être tenus d'enregistrer leurs systèmes dans une base de données de l'UE lorsqu'ils entrent sur le marché. Plusieurs organisations de la société civile ont fait remarquer qu'un régime de transparence significatif devrait en outre fournir des informations sur l'utilisation réelle de ces systèmes dans la pratique.⁴⁰ Cela est particulièrement nécessaire en période de poly-crise, car les exemples décrits ci-dessus montrent que les risques associés à la prédiction algorithmique peuvent être nettement plus élevés dans des situations anormales et hautement dynamiques qu'une évaluation statique des risques basée sur la simple description de l'activité d'une entreprise. Encore une fois, des obligations de transparence particulièrement strictes ne devraient pas entraver

³⁷ La version du 25 novembre 2022 peut être consultée [ici](#). Siehe auch : Kullas, M. / Harta, L. (2021), Europäisches Gesetz über Künstliche Intelligenz Kurzfassung, [cepAnalyse](#) zu COM2021 206 (13.12.2021),

³⁸ Stierle, S. (2022), [Schwierige Trilog-Verhandlungen im neuen Jahr](#), Tagesspiegel Background (7.12.2022),

³⁹ Gorzala, J. (2022), [AI Act der EU : KI-Regulierung im Anmarsch](#), Der Brutkasten (7.12.2022),

⁴⁰ Aszódi, N. (2022), [Wie die Regierung bei den Risiken von KI wegschaut](#), Tagesspiegel Background (6.12.2022),

les start-ups innovantes, mais devraient être formulées en particulier pour les systèmes d'IA utilisés par les entreprises dominantes ou les autorités publiques, car ceux-ci ont potentiellement l'impact le plus important en cas de « cygne noir » ayant des dimensions sociétales, comme le déclenchement d'une guerre. Par conséquent, le cep se félicite que la loi sur l'IA ait été révisée pour indiquer que certains utilisateurs de systèmes d'IA à haut risque qui sont des autorités publiques, des institutions ou d'autres entités seront également tenus de s'enregistrer dans la base de données européenne des systèmes d'IA à haut risque.

Dans l'ensemble, il est important de souligner que les changements réglementaires proposés concernant les audits de l'IA et la loi européenne sur l'IA ne visent pas à restreindre l'application de l'IA à des environnements bien définis et contrôlés, car cela pourrait diminuer les effets d'apprentissage cruciaux. L'objectif est plutôt de trouver des « règles robustes » qui permettent exactement cela : une application sûre à un système complexe, comme l'est sans aucun doute la poly-crise actuelle. Une fois que les législateurs auront défini les conditions du cadre réglementaire de manière à ce que l'utilisation de systèmes basés sur l'IA pour la prise de décisions économiques ou politiques ne puisse pas causer de dommages à l'ensemble de la société à court terme, le marché pourra séparer les outils de prédiction basés sur l'IA sensibles des outils non sensibles à long terme.

5 Conclusion : la nécessité de règles robustes

Dans l'ensemble, plutôt que de peaufiner les algorithmes pour les crises actuelles, les décideurs et les régulateurs européens devraient établir des règles du jeu générales qui, sans maximiser l'efficacité, permettront aux économies de plus en plus axées sur la technologie de fonctionner raisonnablement si et quand le monde entre dans la prochaine crise. De manière générale, l'imprévisibilité des événements de type « cygne noir » et « rhinocéros gris » suggère qu'une approche purement fondée sur le risque, telle que proposée dans la loi européenne sur l'IA, pourrait ne pas être suffisante, car nous ne pouvons pas connaître le risque global d'un système donné. Toutefois, si l'on accepte l'approche fondée sur le risque proposée dans le projet actuel, il est possible d'intégrer les dangers qui surviennent en période de poly-crise en classant une proportion plus élevée de systèmes d'IA comme « à haut risque » lorsque le climat économique ou politique actuel suggère que leurs données d'apprentissage pourraient être déconnectées de la réalité. Plus important encore, les dirigeants politiques, les entrepreneurs et les journalistes, qui embrassent avec enthousiasme le potentiel des algorithmes modernes, doivent mieux comprendre et communiquer sur le risque de préjudice qu'ils encourent en cas de dégradation de leurs performances en période de poly-crise.



Auteur :

Anselm Küsters, LL.M., chef de la division Digitalisation et nouvelles technologies.

kuesters@cep.eu

Traduction:

Mathilde Baudouin

cepfrance@cep.eu

Centrum für Europäische Politik FREIBURG | BERLIN

Kaiser-Joseph-Straße 266 | D-79098 Freiburg

Schiffbauerdamm 40 Raum 4315 | D-10117 Berlin

Tel. + 49 761 38693-0

Le **Centrum für Europäische Politik** FREIBURG | BERLIN, le **Centre de Politique Européenne** PARIS, et le **Centro Politiche Europee** ROMA forment le **réseau des Centres de Politique Européenne** FREIBURG | BERLIN | PARIS | ROMA.

Libre d'intérêts particuliers et neutre sur le plan politique, le réseau des centres de politique européenne fournit une analyse et une évaluation de la politique de l'Union européenne, visant à soutenir l'intégration européenne et à défendre les principes d'un système économique de marché libre, de soutenir l'intégration européenne et de défendre les principes d'un système économique de marché libre.