

L'intelligence artificielle (IA) : une solution à la rationalité limitée des acteurs économiques ?

Quotmane Ahmed¹

¹Maitre-assistant Université Mouloud Mammeri (Tizi-Ouzou),
Algérie

ouotmaneahmed@gmail.com

Résumé :

Ce travail exploratoire sur les effets de l'intelligence artificielle (IA) sur la rationalité humaine, en particulier la rationalité des acteurs économiques, a révélé une complexité notable. Les résultats mettent en lumière une ambivalence et une nuance significatives. L'usage de l'IA peut, dans certaines situations, compenser les limites de la rationalité humaine telles qu'évoquées par Herbert Simon. Cependant, dans d'autres cas, elle pourrait affaiblir cette rationalité. Ces conclusions soulignent la nécessité d'une approche équilibrée et contextuelle de l'application de l'IA dans les décisions économiques, en tenant compte à la fois de ses avantages potentiels et de ses implications sur la rationalité humaine.

Mots clés : Intelligence artificielle, rationalité limitée, optimisation décisionnelle, apprentissage automatique

Abstract:

This exploratory study on the effects of artificial intelligence (AI) on human rationality, particularly the rationality of economic actors, has revealed notable complexity. The results highlight significant ambivalence and nuance. The use of AI can, in certain situations, compensate for the limitations of human rationality as discussed by Herbert Simon. However, in other cases, it may weaken this rationality. These findings underscore the need for a balanced and contextual approach to the application of AI in economic decision-making, considering both its potential advantages and its implications for human rationality.

Keywords: Artificial intelligence, bounded rationality, decision optimization, machine learning

Introduction

Dans le contexte économique mondial, la rationalité des acteurs économiques demeure une préoccupation de premier plan, tant pour les théoriciens que pour les praticiens. Au cœur de cette analyse réside le problème de la « rationalité limitée », conceptualisée par Herbert Simon, qui expose les contraintes cognitives et informationnelles auxquelles sont confrontés les individus lors de leurs processus décisionnels économiques. Dans ce contexte, l'émergence de l'IA suscite un intérêt croissant, en tant que moyen potentiel de compenser ces lacunes et d'améliorer la rationalité des acteurs économiques.

Dans cette optique, notre investigation se concentre sur la mesure dans laquelle l'intelligence artificielle peut constituer une solution viable pour surmonter la rationalité limitée des acteurs économiques. Pour ce faire, nous débutons le processus de recherche par une interrogation fondamentale : dans quelle mesure l'IA peut-elle être considérée comme une solution efficace pour transcender la rationalité limitée des acteurs économiques. À travers cette problématique, nous essayerons d'explorer l'impact de l'intelligence artificielle sur la rationalité des décideurs économiques. Nous envisageons d'examiner attentivement sa capacité à fournir des analyses approfondies, des prédictions précises et une gestion améliorée des informations complexes, dans le but de mieux appréhender son potentiel à améliorer la prise de décision économique.

De façon générale, l'objectif de cette recherche est d'analyser les interactions entre l'IA et la rationalité économique, en mettant en évidence les défis et les opportunités liés à l'intégration croissante de l'IA dans les processus décisionnels économiques. Elle s'inscrit par ailleurs dans une approche interdisciplinaire, conjuguant des concepts de l'économie comportementale, de l'IA et de la gestion des données. En explorant les implications de l'IA sur la rationalité économique, elle vise à enrichir notre compréhension des défis associés à l'évolution technologique dans le domaine économique.

Notre postulat initial suggère que l'intégration de l'IA dans les processus décisionnels économiques ouvre des perspectives prometteuses pour l'amélioration de la rationalité des acteurs économiques. Nous anticipons que l'IA, en exploitant des algorithmes sophistiqués et une capacité de traitement des données inégalée, permettra une prise de décision plus éclairée et efficiente. En outre, en réduisant les biais cognitifs et les erreurs humaines, elle pourrait contribuer à une meilleure gestion des risques et à des décisions plus stratégiques. En intégrant des modèles prédictifs et des analyses avancées, l'IA pourrait également offrir une compréhension plus approfondie des tendances du marché et des opportunités émergentes, permettant ainsi aux acteurs économiques de mieux anticiper et de réagir aux changements de l'environnement économique. En somme, notre hypothèse suggère que l'IA a le potentiel de transformer radicalement la manière dont les décisions économiques sont prises, offrant ainsi des avantages significatifs en termes d'efficacité, de précision et de rationalité dans le processus décisionnel.

I Le concept de rationalité limitée : Origines, définition et implications

Le concept de rationalité limitée, introduit par Herbert Simon dans les années 1950, représente une approche alternative à la rationalité parfaite traditionnellement supposée dans les modèles économiques classiques et néoclassiques. L'auteur remet en question l'idée selon laquelle les individus agissent de manière parfaitement rationnelle en maximisant leurs intérêts dans un environnement d'information parfaite. Au lieu de cela, il soutient que les êtres humains sont confrontés à des limitations cognitives, temporelles, financières et informationnelles qui les empêchent de traiter toutes les informations disponibles et de prendre des décisions optimales.

- Contraintes informationnelles : elles désignent les limites rencontrées par les individus dans l'acquisition, l'accès et le traitement des données pertinentes pour la

prise de décision. Dans le contexte financier par exemple, ces contraintes peuvent comprendre des lacunes dans la disponibilité des informations sur les actifs financiers, des retards dans la divulgation des données financières, ainsi que des coûts associés à la collecte et à l'analyse des informations. Ces contraintes informationnelles entravent la capacité des individus à prendre des décisions de placement complètement informées. En raison de ces limitations, les investisseurs peuvent être amenés à prendre des décisions basées sur des informations incomplètes ou erronées, ce qui peut compromettre leur rationalité dans le processus de prise de décision financière.

- Contraintes financières : elles signalent les restrictions liées aux ressources monétaires disponibles pour les individus. Cela peut inclure des limitations budgétaires, des restrictions d'accès au crédit, ou des exigences de liquidité qui limitent la capacité des individus à investir ou à diversifier leur portefeuille de manière optimale. En cela, elles peuvent contraindre les individus à prendre des décisions de placement sous-optimales en raison de leur incapacité à accéder à certains actifs financiers ou à diversifier leur portefeuille de manière appropriée. Ceci peut entraîner une sous-optimisation des rendements ou une exposition accrue au risque.

- Contraintes cognitives : elles font référence aux limitations liées aux capacités cognitives des individus à traiter l'information et à prendre des décisions rationnelles. Cela peut inclure des biais cognitifs, des limitations de traitement de l'information, ou des erreurs de jugement qui peuvent influencer les décisions financières des individus. Les contraintes cognitives peuvent aussi mener les acteurs à des décisions irrationnelles ou sub-optimales en raison de l'influence des biais cognitifs sur le processus de prise de décision. Par exemple, la sur-confiance au niveau des marchés financiers peut produire une surestimation des rendements attendus, tandis que l'aversion à la perte peut se traduire par des décisions excessivement prudentes.

- Contraintes temporelles : elles décrivent les limites de temps auxquelles les individus sont confrontés lorsqu'ils prennent des décisions. Ces limites peuvent être dues à des échéances strictes, des pressions de calendrier ou des contraintes de délai dans un environnement dynamique. Face à des délais stricts et des pressions de calendrier, les individus doivent souvent se contenter d'informations partielles, adoptant des heuristiques et des raccourcis cognitifs pour parvenir à une décision dans les délais impartis. Cette précipitation peut accroître le risque de décisions impulsives ou basées sur des intuitions plutôt que sur une analyse rationnelle complète. Ainsi, les contraintes temporelles obligent les individus à trouver un compromis entre l'exhaustivité de l'analyse et la nécessité d'agir rapidement, ce qui peut avoir des implications sur la qualité et la rationalité des décisions prises.

De ce point de vue, la rationalité limitée peut être définie comme un processus de prise de décision dans lequel les individus adoptent des heuristiques simples pour résoudre des problèmes complexes, plutôt que de chercher activement la solution optimale. Pour Simon H (1956) : « La rationalité limitée est la notion selon laquelle les individus, lorsqu'ils prennent des décisions, opèrent sous des contraintes de connaissance, de capacité cognitive et de temps. Ils font usage d'heuristiques plutôt que de procédures optimales pour atteindre un résultat satisfaisant, mais pas nécessairement optimal, dans un délai limité »¹. Cette approche reconnaît les contraintes auxquelles sont confrontés les individus dans leur environnement décisionnel et met en avant la notion d'approximation satisfaisante plutôt que d'optimisation parfaite. En d'autres termes, les individus cherchent des solutions qui sont "suffisamment bonnes" compte tenu de leurs ressources limitées.

¹ Simon, H. A. (1956). Rational Choice and the Structure of the Environment. *Psychological Review*, 63(2), 129-138.

Le concept de rationalité limitée a profondément influencé divers domaines, notamment l'économie, la psychologie, la sociologie et la théorie organisationnelle. En économie, ce concept remet en question les fondements de la théorie néoclassique en mettant en évidence les limites de la rationalité humaine dans la prise de décision économique. Contrairement à l'hypothèse néoclassique selon laquelle les individus agissent de manière parfaitement rationnelle pour maximiser leur utilité, la rationalité limitée reconnaît que les individus sont confrontés à des contraintes cognitives et informationnelles qui limitent leur capacité à traiter toutes les informations disponibles et à prendre des décisions optimales. Ainsi, ce concept souligne l'importance des comportements heuristiques, c'est-à-dire des raccourcis cognitifs et des règles simples, dans le processus de prise de décision économique.

En psychologie, la reconnaissance de la rationalité limitée a conduit au développement de modèles comportementaux qui prennent en compte les biais cognitifs et les limitations de traitement de l'information chez les individus. Ces modèles reconnaissent que les individus sont sujets à des biais tels que la surconfiance, l'aversion à la perte et l'ancrage, qui peuvent influencer leurs décisions. Par conséquent, la psychologie économique et comportementale s'intéresse à comprendre comment ces biais affectent les choix individuels et collectifs et comment ils peuvent être atténués ou exploités dans différents contextes.

Sur le plan pratique, la prise en compte du concept implique des conséquences importantes pour la conception des politiques publiques, la gestion des organisations et la prise de décision stratégique. En intégrant la compréhension de la rationalité limitée dans ces domaines, il est possible de promouvoir des résultats plus efficaces et équitables. Cela suppose la simplification des choix pour rendre les décisions plus accessibles, la fourniture d'informations claires pour aider les individus à prendre des décisions éclairées, et l'élaboration de mécanismes de prise de décision qui tiennent compte des heuristiques et des biais cognitifs.

II Méthodologie de travail :

Pour collecter les données nécessaires à notre recherche, nous avons adopté une approche méthodique en entreprenant une recherche documentaire exhaustive. Celle-ci s'est déroulée dans plusieurs bases de données académiques et professionnelles renommées, ainsi que dans des revues spécialisées, garantissant ainsi une couverture large et approfondie du domaine. Nous avons pris soin de sélectionner des articles récents et pertinents qui se concentrent spécifiquement sur le rôle et l'impact de l'intelligence artificielle dans le processus décisionnel économique. Ces articles ont été soumis à un examen minutieux et à une synthèse méthodique afin d'identifier les tendances émergentes, les défis cruciaux et les opportunités prometteuses associées à l'utilisation de l'IA dans ce contexte spécifique. En consolidant ces informations, notre objectif était d'obtenir une compréhension approfondie et nuancée des implications de l'IA dans le processus décisionnel économique, fournissant ainsi une base solide pour notre analyse et nos conclusions.

En parallèle, nous avons entrepris l'identification et la sélection d'études empiriques récentes portant spécifiquement sur les impacts de l'intelligence artificielle sur les décisions prises par les individus et les organisations dans une variété de contextes. Cette démarche nous a permis de recueillir des données concrètes issues du terrain, offrant ainsi un éclairage complémentaire aux informations théoriques obtenues dans la littérature académique. Après avoir identifié ces études, nous avons procédé à une analyse méticuleuse de leurs résultats, en mettant en évidence les nuances et les implications des effets de l'IA sur les processus décisionnels. Nous avons examiné en détail les méthodologies utilisées dans ces études, notamment les techniques d'analyse de données et les modèles statistiques employés, ainsi que les mesures de performance et les indicateurs de succès utilisés pour évaluer l'impact de l'IA. Cette approche rigoureuse nous a permis de compléter et de valider les informations

recueillies dans la littérature académique, renforçant ainsi la robustesse de notre analyse et la fiabilité de nos conclusions.

Concernant le protocole de maîtrise de la qualité des données, notre approche a été rigoureuse afin de garantir la fiabilité et la validité de nos résultats. Dans un premier temps, nous avons accordé une attention particulière à la sélection de sources de données reconnues et fiables dans le domaine de l'intelligence artificielle et de la prise de décision économique. Cela nous a permis de nous appuyer sur des travaux académiques et des recherches professionnelles de qualité, soutenant la crédibilité de nos analyses. De plus, nous avons établi des critères stricts pour la sélection des articles et des études empiriques, en accordant une préférence aux travaux présentant des méthodologies robustes et des résultats vérifiables. Cette approche a permis de garantir la pertinence et la solidité des données utilisées dans notre recherche. Enfin, nous avons procédé à une analyse approfondie des données collectées, en employant des techniques statistiques avancées et des outils d'analyse qualitative pour identifier les tendances et les patterns significatifs. Cette démarche analytique rigoureuse a permis d'assurer la précision et la fiabilité de nos conclusions, nécessaire à la validité de notre étude dans le domaine de l'intelligence artificielle et de la prise de décision économique.

III Dualité complexe de l'intelligence artificielle : entre le rationnel et l'éthique :

L'intelligence artificielle (IA) suscite un débat passionné entre ses possibilités rationnelles et les préoccupations éthiques qu'elle soulève. Cette dualité complexe interpelle tant les scientifiques que les philosophes, et elle est au cœur des discussions contemporaines sur l'avenir de la technologie et de la société. En examinant de près les progrès de l'IA et les questions éthiques qui l'accompagnent, il devient évident que ces deux aspects sont profondément entrelacés et nécessitent une réflexion approfondie.

L'IA représente un domaine interdisciplinaire qui conjugue des avancées en informatique, en statistiques et en apprentissage automatique pour créer des systèmes capables de résoudre des tâches traditionnellement associées à l'intelligence humaine. Du traitement du langage naturel à la vision par ordinateur en passant par la planification et la prise de décision, l'IA offre des possibilités infinies d'automatiser et de rationaliser divers processus.

Un agent d'intelligence artificielle est en fait : « un système informatique qui est capable de percevoir son environnement grâce à des capteurs, d'analyser ces perceptions en utilisant des algorithmes et des modèles d'apprentissage automatique, et d'agir de manière autonome pour atteindre des objectifs spécifiques dans cet environnement »². Même si la définition d'un agent d'intelligence artificielle peut varier en fonction du contexte spécifique dans lequel il est utilisé, mais de manière générale, il peut être décrit comme un système informatique conçu pour agir de manière autonome afin d'accomplir des tâches spécifiques en fonction de son environnement et des objectifs qui lui sont assignés. En cela, la distinction avec l'intelligence humaine réside principalement dans leur fonctionnement et leurs capacités :

- Capacités cognitives : Les agents d'IA sont généralement conçus pour exceller dans des tâches spécifiques en utilisant des algorithmes et des modèles d'apprentissage automatique. Leur performance dépend souvent de la qualité et de la quantité des données sur lesquelles ils sont formés. En revanche, l'intelligence humaine est souvent caractérisée par sa capacité à généraliser, à apprendre de nouvelles informations de manière flexible et à s'adapter à des situations variées avec une compréhension contextuelle.

² Negnevitsky, M. (2011). Artificial Intelligence: A Guide to Intelligent Systems. Pearson.

- **Conscience et subjectivité** : Les agents d'IA ne possèdent pas de conscience ou de subjectivité comme les êtres humains. Ils ne ressentent pas d'émotions, de désirs ou de motivations intrinsèques. Leur comportement est déterminé par des règles programmées et des objectifs prédéfinis, et ils ne sont pas capables de prendre des décisions en fonction de leur propre compréhension subjective de la situation.
- **Limites et spécificité des tâches** : Les agents d'IA sont généralement conçus pour exécuter des tâches spécifiques et limitées pour lesquelles ils sont formés. Leur performance peut être exceptionnelle dans leur domaine d'expertise, mais ils peuvent être inefficaces voire incompétents dans des domaines où ils n'ont pas été formés. En revanche, l'intelligence humaine est souvent caractérisée par sa capacité à s'adapter à une variété de tâches et de situations, et à tirer parti de l'expérience et de la connaissance pour résoudre de nouveaux problèmes.

En conséquence, du point de vue de la rationalité, la capacité de l'intelligence artificielle à analyser et à interpréter de vastes ensembles de données évoque un progrès significatif par rapport aux méthodes traditionnelles de traitement de l'information. Les algorithmes d'apprentissage automatique, en particulier les réseaux de neurones profonds, sont capables d'apprendre à reconnaître des modèles complexes et de faire des prédictions précises à partir de données brutes. Contrairement aux humains, qui peuvent être limités par des biais cognitifs ou des limitations de traitement de l'information, les systèmes d'IA peuvent analyser les données de manière objective et exhaustive, en exploitant chaque élément d'information disponible pour tirer des conclusions pertinentes.

Dans le domaine médical, par exemple, l'intelligence artificielle a révolutionné la manière dont les diagnostics sont établis et les traitements sont planifiés. Les systèmes d'IA peuvent analyser des milliers, voire des millions d'images médicales avec une précision et une rapidité inégalées, permettant ainsi aux médecins de détecter des signes précoce de maladies telles que le cancer ou les maladies cardiaques. Cette capacité à identifier des anomalies subtiles qui pourraient échapper à l'œil humain donnent lieu à des diagnostics plus précis et à des interventions plus rapides, améliorant ainsi les résultats pour les patients. De surcroit, elle offre la possibilité d'optimiser les processus médicaux et d'améliorer l'efficacité des soins de santé, en analysant les dossiers médicaux électroniques pour repérer des tendances et des corrélations entre les symptômes, les traitements et les résultats des patients, permettant ainsi aux professionnels de la santé d'adapter les protocoles de traitement en fonction des besoins individuels de chaque patient, ce qui peut dans une large mesure concourir à la réduction des erreurs médicales, à la minimisation des coûts de santé et l'amélioration de l'expérience globale des patients.

Cependant, cette rationalité apparente n'est pas sans failles. L'utilisation généralisée et abusive de l'intelligence artificielle dans divers domaines soulève des préoccupations importantes concernant les biais algorithmiques et les conséquences néfastes qu'ils peuvent générer. Pour cause, les modèles d'IA sont souvent formés sur des ensembles de données historiques qui peuvent refléter et amplifier les préjugés sociaux existants, perpétuant ainsi les inégalités et les discriminations dans la prise de décision automatisée (Barocas & Selbst, 2016). Comme l'ont démontré certaines études dans le domaine de gestion des ressources humaines, les algorithmes de sélection des candidats lors de la phase de sélection peuvent par exemple être influencé par des modèles de recrutement passés qui ont favorisé certains groupes démographiques au détriment d'autres (Dastin, 2018), de la même façon que dans le secteur des prêts, d'autres études indiquent que les algorithmes utilisés pour évaluer la solvabilité des emprunteurs peuvent aussi involontairement perpétuer des pratiques discriminatoires en accordant des prêts de manière inéquitable en fonction de la race, du sexe ou d'autres caractéristiques démographiques (Mullainathan & Spiess, 2017).

Ces exemples mettent en lumière la manière dont les biais dans les données historiques peuvent se propager dans les décisions automatisées prises par les algorithmes d'IA. L'émergence de ces biais soulève des questions éthiques importantes en plus des risques d'erreurs potentielles dans la prise de décision. Autrement dit, lorsque des décisions critiques sont automatisées par des algorithmes, il est essentiel de garantir qu'elles ne favorisent ni ne pénalisent injustement certains groupes de personnes. Les biais algorithmiques introduisent potentiellement des injustices et des discriminations systémiques, ce qui soulève des préoccupations éthiques majeures quant à l'équité et à la justice dans notre société. Pour ces raisons, l'équité algorithmique est devenue une préoccupation majeure et vise déjà à garantir que les systèmes d'IA prennent des décisions qui sont justes et impartiales pour tous les individus, indépendamment de leur origine ou de leur statut socio-économique (Chouldechova, 2017). Cela pourrait se traduire par un examen minutieux des données d'entraînement pour identifier et corriger les biais existants, ainsi que la conception d'algorithmes qui sont transparents et explicables, permettant ainsi aux utilisateurs de comprendre comment les décisions sont prises et de remédier à tout comportement discriminatoire (Diakopoulos, 2016).

En ce sens, la nécessité de transparence dans les systèmes d'intelligence artificielle s'inscrit dans le contexte plus large de l'équité algorithmique, qui vise à garantir des décisions justes et impartiales pour tous les individus, en dépit des préoccupations liées aux biais algorithmiques et aux discriminations potentielles. La transparence sur la manière dont les systèmes d'IA fonctionnent et sur les données utilisées dans leur formation permet non seulement de renforcer la confiance du public dans ces systèmes, mais aussi d'encourager une surveillance et un examen indépendants pour détecter et corriger les éventuels biais ou injustices (Goodman & Flaxman, 2016). En rendant publics le fonctionnement et les données usagées par les systèmes d'IA, les organisations permettent de la sorte aux parties prenantes externes, telles que les chercheurs, les régulateurs et le grand public, de mieux comprendre comment ces systèmes prennent des décisions et quelles sont les implications de ces décisions.

Cet aspect renforce premièrement la confiance du public dans les systèmes d'IA, particulièrement encore dans les domaines sensibles tels que la santé, la finance et la justice, où les décisions prises par des algorithmes peuvent avoir un impact significatif sur la vie des individus ; deuxièmement il facilite la surveillance et l'examen indépendants des systèmes d'IA en mettant à disposition des données sur lesquelles les algorithmes sont formés, ainsi que sur les caractéristiques et les poids de ces algorithmes, pour permettre aux régulateurs d'évaluer la justice et l'équité des décisions prises; et enfin il encourage la responsabilité des organisations qui développent et déploient des systèmes d'IA en les incitant à adopter des pratiques éthiques et responsables, à investir dans la détection et la correction des biais, ainsi qu'à concevoir des mécanismes de recours pour les individus affectés par les décisions des systèmes d'IA.

Il en résulte de ces effets contradictoires et nuancés la conclusion suivante : bien que l'intelligence artificielle offre un potentiel immense pour résoudre des problèmes complexes et accroître l'efficacité dans divers domaines, son développement actuel est entaché par des ambiguïtés et des défis persistants. À ce stade, l'intelligence artificielle ne garantit pas nécessairement une efficacité incontestable. En fait, les biais cognitifs humains, qui sont inévitablement intégrés dans la conception et le déploiement des systèmes d'IA, peuvent altérer la précision et l'impartialité des résultats produits. Ces biais cognitifs, souvent inconscients, peuvent se manifester à travers les décisions prises lors de la collecte et du traitement des données, ainsi que dans la conception des algorithmes. Par conséquent, même les systèmes d'IA les plus sophistiqués peuvent être imprégnés de partialité, compromettant ainsi leur efficacité perçue. De surcroit, les biais algorithmiques, qui résultent de l'utilisation de données historiques souvent biaisées pour former les modèles d'IA, constituent un défi

majeur, en ce sens qu'ils peuvent perpétuer et même amplifier les injustices et les discriminations existantes, impactant ainsi négativement les décisions prises par les systèmes d'IA dans des domaines critiques tels que le recrutement, les prêts, ou encore le système judiciaire et de santé.

En conclusion, la dualité de l'intelligence artificielle, combinant des capacités prometteuses et des limites significatives, lève des questions complexes sur son rôle dans les sociétés contemporaines : d'une part, elle offre des opportunités passionnantes pour améliorer les processus décisionnels, accroître l'efficacité opérationnelle et résoudre des problèmes complexes à une échelle sans précédent; d'autre part, elle pose des défis éthiques et sociaux considérables, notamment en ce qui concerne l'équité, la transparence et la responsabilité dans la prise de décision automatisée.

IV Effets ambivalents de l'intelligence artificielle sur la rationalité humaine : une exploration des recherches documentaires

L'intelligence artificielle (IA) se situe à la croisée de deux paradigmes décisionnels fondamentaux : la rationalité limitée, inhérente à la cognition humaine, et l'idéal de rationalité parfaite, souvent sous-tendu par les modèles conventionnels de prise de décision. En amalgamant les aspects de la rationalité limitée dans des systèmes informatiques capables d'analyser rapidement d'énormes volumes de données et de fournir des recommandations étayées par des algorithmes sophistiqués, elle demeure toujours confrontée aux contraintes propres à la rationalité humaine, telles que la compréhension contextuelle, la prise en compte des valeurs éthiques et la gestion de l'incertitude.

La recherche académique n'a pas tardé à mettre en évidence les effets ambivalents de l'usage de l'intelligence artificielle sur le fonctionnement des sociétés, tantôt certaines études soulignent ses contributions significatives à l'efficacité opérationnelle, à l'innovation technologique et à la qualité des services, tantôt d'autres mettent en garde contre les risques potentiels tels que la perte d'emplois, la déshumanisation des interactions, et l'accentuation des inégalités sociales.

IV 1 Mécanismes de renforcement de la rationalité décisionnelle :

L'intelligence artificielle (IA) révolutionne la rationalité décisionnelle en exploitant des mécanismes sophistiqués pour traiter des quantités massives de données, identifier des schémas complexes et fournir des insights utiles. En automatisant les tâches répétitives et en proposant des recommandations personnalisées basées sur des modèles prédictifs, elle optimise les processus décisionnels, libérant ainsi les décideurs des contraintes de temps et de ressources. Elle offre par ailleurs la capacité d'apprendre et de s'adapter en continu, améliorant la pertinence et la précision des décisions prises. Ces mécanismes positionnent l'IA comme un outil indispensable pour renforcer la rationalité décisionnelle et relever les défis complexes du monde moderne marqué par de l'incertitude et de l'inconstance

Eléments	Rationalité Limitée Humaine	Rationalité Avancée de l'IA
Traitement des Environnements Incertains	Difficulté à prendre des décisions dans des contextes complexes et incertains	Capacité à analyser et à répondre de manière adaptative à des environnements dynamiques et incertains grâce à des algorithmes d'apprentissage et à des modèles prédictifs
Nature des Décisions	Souvent influencée par des biais cognitifs et émotionnels	Basée sur des algorithmes sophistiqués et une analyse objective des données
Processus d'Apprentissage	Limité par la capacité cognitive et l'expérience individuelle	Utilisation d'algorithmes d'apprentissage automatique pour améliorer les performances au fil du temps et tirer des enseignements des données
Optimalité des Décisions	Souvent sub-optimales en raison de contraintes cognitives et de biais perceptuels	Recherche de solutions optimales basées sur des analyses approfondies et une évaluation systématique des

		alternatives
Contraintes	Confinement à la capacité cognitive humaine et à l'expérience limitée	Dépendance aux ressources informatiques et aux algorithmes disponibles, avec des limitations potentielles de données ou d'infrastructures
Adaptabilité aux Changements	Résistance aux changements et aux mises à jour une fois les décisions prises	Capacité à s'adapter rapidement aux nouveaux défis et à modifier les recommandations en fonction des nouvelles données et des évolutions de l'environnement
Prise en compte du contexte	Souvent influencée par des facteurs externes difficiles à intégrer de manière holistique	Capacité à tenir compte de multiples variables contextuelles et à ajuster les décisions en fonction des circonstances spécifiques

Source : auteur

a) Complémentarité cognitive :

L'intelligence artificielle (IA) joue un rôle crucial dans la promotion de la complémentarité cognitive en fournissant des capacités de traitement de l'information et d'analyse des données qui enrichissent considérablement les capacités cognitives humaines. Cette complémentarité découle directement de la capacité inégalée de l'IA à traiter rapidement et efficacement de vastes quantités de données, ainsi qu'à identifier des schémas complexes, offrant ainsi une perspective élargie et des insights irremplaçables qui surpassent souvent l'observation humaine seule.

Les travaux de recherche menés par David Autor et Frank Levy (2003) ont clairement démontré comment l'automatisation, largement facilitée par l'intelligence artificielle, libère les humains des tâches répétitives et les oriente vers des activités plus cognitivement exigeantes. Cette transition vers des activités plus complexes est bénéfique à plusieurs égards. D'abord, elle permet aux individus de consacrer leur temps et leurs efforts à des activités qui requièrent une expertise humaine, comme la créativité, l'innovation et la résolution de problèmes complexes. Secondelement, elle favorise le développement de compétences cognitives et socio-émotionnelles essentielles, telles que la pensée critique, la collaboration et l'empathie, qui sont intrinsèquement humaines et difficiles à reproduire par des systèmes d'IA.

La complémentarité cognitive entre l'IA et l'humain est en ce sens indéniablement bénéfique pour la prise de décision. En combinant les forces de l'IA - précision, rapidité et capacité à traiter de grandes quantités de données - avec les capacités humaines - jugement, intuition et expertise contextuelle -, les décideurs peuvent profiter d'une perspective plus complète et équilibrée lorsqu'ils sont confrontés à des situations complexes. Par exemple, dans le domaine médical, les médecins utilisent couramment des systèmes d'IA pour analyser des données médicales et proposer des diagnostics préliminaires, tout en appliquant leur expertise clinique et leur jugement pour prendre des décisions éclairées sur les options de traitement.

En gros, la complémentarité cognitive entre l'IA et l'humain offre des opportunités sans précédent pour repenser la manière dont nous abordons les tâches cognitives et les processus décisionnels. Plutôt que de remplacer les capacités humaines, l'IA renforce et enrichit notre intelligence collective, nous permettant ainsi de relever des défis plus complexes et de réaliser des avancées significatives dans un large éventail de domaines.

b) Optimisation des décisions et personnalisation intelligente :

L'optimisation des décisions grâce à l'intelligence artificielle représente une avancée majeure dans de nombreux domaines, car elle permet d'améliorer la qualité et l'efficacité des processus décisionnels. L'intégration d'algorithmes sophistiqués dans les systèmes d'IA permet d'analyser des ensembles de données massives et complexes, extrayant ainsi des informations précieuses qui seraient difficilement accessibles par les méthodes traditionnelles. Cette capacité d'analyse approfondie permet de prendre des décisions plus érudites, basées sur des données empiriques plutôt que sur des intuitions subjectives.

Toujours dans le domaine de la santé, les systèmes d'IA sont dans certaines circonstances utilisés pour analyser les dossiers médicaux des patients, les résultats d'examens

et d'analyses, ainsi que les données génétiques, afin de recommander des traitements personnalisés adaptés à chaque individu. Cette approche, basée sur l'apprentissage automatique et l'analyse prédictive, permet d'identifier les meilleures options de traitement en fonction des caractéristiques spécifiques de chaque patient, améliorant ainsi les résultats cliniques et réduisant les risques d'erreurs médicales. Une étude menée d'ailleurs par Ribeiro et al. (2016) a clairement démontré la façon dont les techniques d'apprentissage automatique sont utilisées pour analyser des données médicales massives, telles que les dossiers médicaux des patients, et recommander des traitements personnalisés en fonction des caractéristiques individuelles de chaque patient.

Dans le même sillage, dans le domaine financier cette fois-ci, l'intelligence artificielle est abondamment exploitée pour optimiser les décisions d'investissement. En analysant en temps réel les données du marché, les tendances économiques, les performances passées des actifs et les préférences des investisseurs, les systèmes d'IA parviennent à recommander dans plusieurs cas des portefeuilles d'investissement personnalisés qui maximisent les rendements tout en minimisant les risques. Cette approche permet aux investisseurs de prendre des décisions plus nuancées et mieux informées, en tenant compte de multiples facteurs et en anticipant les fluctuations du marché. A cet effet, les travaux de Lhabitant et al. (2019) concluent que l'exploitation des systèmes d'IA est très souvent optimal car cela permet d'analyser en temps réel les données du marché, les tendances économiques et les préférences des investisseurs pour prôner des portefeuilles d'investissement personnalisés qui maximisent les rendements tout en minimisant les risques.

c) Apprentissage automatique

L'apprentissage automatique, en tant que domaine clé de l'intelligence artificielle, repose sur des algorithmes sophistiqués qui permettent aux systèmes informatiques d'apprendre à partir de données et d'améliorer leurs performances sans intervention humaine directe. Cette approche révolutionnaire permet aux machines de détecter des modèles complexes dans les données et de prendre des décisions éclairées en se basant sur ces modèles, ouvrant ainsi la voie à une multitude d'applications innovantes et impactantes dans divers domaines.

Silver et ses collègues (2016) ont mis en lumière l'efficacité de l'apprentissage automatique à travers une série d'exemples convaincants. Parmi ceux-ci, l'exploit dans le domaine des jeux, où des algorithmes d'apprentissage automatique ont été développés pour rivaliser avec des joueurs humains dans des jeux aussi complexes que le Go et le Poker. Ces algorithmes ont démontré une capacité remarquable à apprendre des stratégies sophistiquées et à s'adapter de manière dynamique aux changements dans l'environnement du jeu, illustrant ainsi la puissance de l'apprentissage automatique dans la résolution de problèmes complexes.

Parallèlement, les progrès réalisés dans le domaine de la reconnaissance d'images grâce à l'apprentissage automatique sont également impressionnantes. D'autres études ont montré comment ces algorithmes peuvent être entraînés à identifier et à classer automatiquement des objets dans des images avec une précision remarquable, comparable voire supérieure à celle des humains. Cette capacité a des implications importantes dans des domaines tels que la médecine, où l'analyse automatisée d'images médicales peut aider les professionnels de la santé à diagnostiquer des maladies avec une précision accrue et une rapidité améliorée, améliorant ainsi les résultats pour les patients.

Par ailleurs, l'utilisation généralisée de l'apprentissage automatique dans le développement de systèmes de recommandation personnalisés a considérablement amélioré l'expérience utilisateur dans des domaines comme le commerce électronique et le streaming de contenu. Ces systèmes exploitent les données sur les préférences et le comportement des utilisateurs pour fournir des recommandations sur mesure, augmentant ainsi l'engagement des utilisateurs et stimulant la croissance des plateformes.

En somme, il faut conclure que l'apprentissage automatique favorise l'innovation dans de nombreux domaines en permettant aux systèmes informatiques d'analyser des données massives et de découvrir des informations précieuses et des modèles cachés. Cette capacité à extraire des connaissances à partir de données crée de nouvelles opportunités pour développer

des produits et des services innovants, ainsi que pour améliorer les processus existants. Pour revenir vers les exemples précédents, l'utilisation de techniques d'apprentissage automatique pour analyser les données médicales permet aux chercheurs de découvrir de nouveaux traitements, de prédire les épidémies de maladies, et d'optimiser les protocoles de traitement. De même, dans le secteur financier, l'application de l'apprentissage automatique dans l'analyse des données financières conduit à de nouvelles stratégies d'investissement, à une détection plus rapide de fraudes et à une gestion plus efficace des risques.

IV 2 Mécanismes d'affaiblissement de la rationalité décisionnelle :

L'intelligence artificielle, malgré ses innombrables contributions à l'optimisation des processus décisionnels et à la complémentarité cognitive, révèle également des facettes problématiques susceptibles d'altérer la rationalité des individus. Les études académiques et les recherches empiriques mettent en lumière la propension des systèmes d'IA à induire des biais cognitifs et algorithmiques, impactant ainsi la capacité des utilisateurs à prendre des décisions de manière autonome et rationnelle. Cette dépendance croissante à l'égard des recommandations automatisées pourrait compromettre l'exercice d'un discernement critique et l'autonomie décisionnelle des individus. De plus, la complexité intrinsèque des algorithmes d'IA peut obscurcir leur fonctionnement interne, engendrant ainsi un manque de transparence qui nuit à la compréhension des processus décisionnels et à la confiance des utilisateurs. Cette perspective scientifique souligne l'impérieuse nécessité d'approfondir notre compréhension des mécanismes par lesquels l'intelligence artificielle peut altérer la rationalité humaine, afin de mieux appréhender ses implications et d'orienter les politiques de régulation et d'éthique de manière informée.

Erreurs et risques de l'IA	Habilités de la cognition humaine
Erreurs de calcul dues à une mauvaise programmation informatique.	Capacité à raisonner de manière abstraite et à prendre en compte le contexte.
Biais algorithmiques causés par des données d'entraînement non représentatives.	Capacité à comprendre et à intégrer des valeurs éthiques et morales dans la prise de décision.
Surconfiance dans les recommandations des systèmes d'IA, conduisant à une négligence des informations contradictoires.	Capacité à remettre en question les informations et à prendre des décisions de manière critique.
Opacité des algorithmes d'IA, rendant difficile la compréhension de leurs décisions et de leur fonctionnement.	Capacité à expliquer et à justifier les décisions prises, favorisant ainsi la confiance et la transparence.
Déresponsabilisation des individus, transférant la responsabilité des décisions aux systèmes d'IA.	Sens de la responsabilité personnelle et capacité à assumer les conséquences de ses actions.
Perte de compétences cognitives dues à une dépendance excessive à l'égard des systèmes d'IA.	Capacité à apprendre et à s'adapter à de nouveaux environnements ou situations.
Erreurs dues à des changements de dernière minute ou à des imprévus non pris en compte par les algorithmes.	Capacité à s'adapter rapidement aux changements et à prendre des décisions en fonction du contexte actuel.

Source : auteur

a) Biais algorithmiques :

Les biais algorithmiques, résultant des distorsions inhérentes aux données utilisées pour former les algorithmes d'intelligence artificielle, peuvent considérablement affecter la rationalité décisionnelle. Les biais cognitifs, qui sont des tendances systématiques à penser de manière irrationnelle ou à prendre des décisions fondées sur des heuristiques simplifiées, peuvent être reflétés dans les données utilisées pour former les algorithmes d'intelligence artificielle. Lorsque ces biais cognitifs sont incorporés dans les données d'entraînement, ils deviennent des biais algorithmiques qui peuvent influencer les décisions prises par les systèmes automatisés. Ces derniers peuvent résulter de diverses sources, notamment des données d'entraînement déséquilibrées, des corrélations accidentnelles entre certaines caractéristiques et les résultats attendus, ou encore des lacunes dans la conception des algorithmes. Leur impact sur la rationalité décisionnelle est complexe et multifactoriel.

Une illustration pertinente de ce phénomène réside dans le domaine du recrutement, où les systèmes d'IA chargés de trier les CV peuvent reproduire des biais historiques liés à l'origine ethnique, au genre ou à d'autres caractéristiques protégées. Dans le secteur de la santé, une étude menée par Obermeyer et al. (2019) a révélé par exemple que les algorithmes de santé utilisés pour identifier les patients nécessitant un suivi médical étaient biaisés contre les patients noirs, les sous-estimant par rapport à leurs homologues blancs. Cette distorsion dans les résultats est attribuable à des données d'entraînement déséquilibrées ou à des corrélations accidentnelles entre certaines caractéristiques et le résultat attendu, ce qui compromet la validité et l'équité des décisions prises par ces systèmes.

Par ailleurs, ces biais sont relativement exacerbés par des facteurs socio-économiques ou culturels, comme l'ont démontré Buolamwini et Gebru (2018) d'après qui les algorithmes de reconnaissance faciale avaient une précision moindre lorsqu'ils étaient confrontés à des personnes de couleur ou de genre non binaire, ce qui souligne les préjugés inhérents aux données d'entraînement et les limites des technologies d'IA actuelles. Cette distorsion dans les résultats est favorable à la formation de décisions erronées ou discriminatoires, compromettant ainsi la rationalité décisionnelle et exacerbant les inégalités existantes.

L'opacité des algorithmes d'IA, souvent protégée par des secrets commerciaux ou des restrictions de propriété intellectuelle, rend difficile la détection et la correction des biais algorithmiques. Cette absence de transparence entrave la compréhension des mécanismes sous-jacents des systèmes d'IA et limite la capacité des individus à évaluer la justesse et l'équité des décisions qui les concernent. En conséquence, cela soulève des préoccupations supplémentaires en matière de responsabilité et de confiance dans l'utilisation croissante de l'IA dans divers domaines, de la finance à la santé en passant par la justice.

b) Sur-confiance :

La sur-confiance, phénomène bien documenté en psychologie cognitive, se réfère à une tendance humaine à surestimer ses propres capacités ou les performances d'un système, souvent sans prendre en compte les informations contradictoires ou les incertitudes. Lorsqu'il s'agit de systèmes d'intelligence artificielle (IA), cette sur-confiance peut être exacerbée par la perception de l'IA comme étant objective et impartiale, ce qui incite les utilisateurs à accorder une confiance excessive aux conseils ou aux décisions fournies par ces systèmes.

Ici aussi, plusieurs études, telles que celles menées par Dietvorst, Simmons et Massey (2014), ont montré que les individus ont tendance à accorder une confiance excessive aux prédictions des algorithmes, même quand'ils sont informés de leur performance limitée. Toujours dans le domaine médical, des recherches ont révélé que les médecins peuvent surestimer la capacité des algorithmes de diagnostic à détecter certaines maladies, ce qui entraîne une confiance excessive dans les résultats et une négligence des signes cliniques ou des données supplémentaires qui pourraient contredire les recommandations de l'IA³.

Cette sur-confiance se manifeste surtout dans le domaine de la prise de décision financière, où les investisseurs accordent parfois trop de poids aux recommandations des systèmes d'IA en matière d'investissement, sans tenir compte de leurs limites ou erreurs potentielles⁴. Dans ce contexte, les investisseurs suivent dans la plus grande partie des cas aveuglément les avis des algorithmes de trading, formant des décisions financières risquées ou inappropriées.

c) Déresponsabilisation :

La déresponsabilisation se définit comme le processus par lequel les individus ou les organisations transfèrent la responsabilité de leurs actions, décisions ou conséquences à d'autres entités, souvent des systèmes ou des processus automatisés, tels que les algorithmes d'intelligence artificielle. Cela se traduit par une diminution de l'engagement personnel dans le processus décisionnel et une perte de sens des responsabilités individuelles. La déresponsabilisation peut découler de divers facteurs, y compris une confiance excessive dans les technologies, une dépendance accrue aux systèmes automatisés et une réduction de la prise en compte des conséquences éthiques ou morales de ses propres actions.

³ Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.

⁴ Barber, B. M., & Odean, T. (2016). The brave new world of hedge fund investing: Implications of high-frequency and algorithmic trading. *Review of Financial Studies*, 30(3), 917-946.

- Transfert de responsabilité : Les individus transfèrent souvent la responsabilité de leurs décisions aux systèmes d'IA, considérant les décisions des algorithmes comme infaillibles, déclinant ainsi toute responsabilité personnelle.
- Perte de compétences décisionnelles : En s'appuyant excessivement sur les systèmes d'IA pour prendre des décisions à leur place, les individus perdent progressivement leurs compétences décisionnelles et leur capacité à évaluer de manière critique les informations disponibles.
- Absence de remise en question : Face aux recommandations des algorithmes d'IA, les individus ne remettent généralement pas en question les résultats ou les décisions proposés, supposant qu'ils sont automatiquement corrects ou optimaux.
- Biais de confirmation : Les individus ont tendance à accepter les résultats des systèmes d'IA qui confirment leurs propres croyances ou préférences, même lorsqu'il existe des preuves contradictoires.
- Perte de contrôle : L'utilisation de l'intelligence artificielle donne souvent l'impression aux individus qu'ils ont perdu le contrôle sur le processus décisionnel, devenant ainsi dépendants des recommandations automatisées des systèmes d'IA.

Dans l'exemple de modèle de langage, à l'exemple de ChatGPT, la surconsommation de l'intelligence artificielle implique une déresponsabilisation implique de plusieurs façons :

- Attribution de la fiabilité : Les utilisateurs attribuent une fiabilité excessive aux réponses générées par le modèle, en supposant qu'elles sont toujours correctes et précises. Cela induit une déresponsabilisation de leur propre jugement critique et de leur prise de décision.
- Absence de validation : En raison de la facilité avec laquelle les réponses sont générées, les utilisateurs sont habituellement moins enclins à vérifier l'exactitude des informations fournies par le modèle. Ceci aboutit à une déresponsabilisation quant à la vérification des faits et à la validation des données.
- Transfert de responsabilité : Les utilisateurs sont fréquemment tentés de transférer la responsabilité de leurs décisions ou de leurs actions au modèle de langage, en justifiant leurs choix par les réponses fournies par celui-ci. Ce qui traduire pareillement une forme de déresponsabilisation personnelle et à une perte de conscience des conséquences de leurs actions.

Il est tout de même essentiel de rappeler que la déresponsabilisation n'est pas intrinsèquement liée à l'intelligence artificielle, mais plutôt à la manière dont elle est utilisée dans divers contextes. L'usage excessif de l'intelligence artificielle procède souvent de contraintes temporelles ou de pressions liées à la productivité dans un environnement marqué par le rythme effréné du monde moderne. Dans ces conditions, les individus sont susceptibles de céder la responsabilité de leurs décisions à des systèmes automatisés pour gagner du temps ou éviter des tâches jugées fastidieuses. Ce phénomène reflète une dépendance accrue aux technologies et une diminution de la réflexion critique sur les implications des actions humaines. Ainsi, la déresponsabilisation dans le contexte de l'intelligence artificielle résulte davantage de choix individuels et organisationnels que d'une caractéristique inhérente à la technologie elle-même.

d) Perte de compétences :

La perte de compétences découle de divers processus, y compris la sélection adverse, qui se produit lorsque l'utilisation excessive de l'intelligence artificielle diminue l'exercice des compétences cognitives. La forte prévalence de l'automatisation des tâches cognitives réduit forcément les occasions pour les individus de pratiquer leur raisonnement et leur prise de

décision. Cette diminution de l'exercice cognitif induit à son tour une atrophie des compétences cognitives, conduisant ainsi à une perte de compétences au fil du temps.

D'autres mécanismes aboutissent vers la destruction et perte de compétences : la sur-confiance et déresponsabilisation entre autres. Dans le cas où les individus surestiment la fiabilité des calculs algorithmiques et négligent les informations contradictoires ou incertaines, ils deviennent moins enclins à remettre en question leurs propres décisions ou à rechercher activement des alternatives. Cette sur-confiance entraîne une diminution de l'exercice des compétences cognitives et de la réflexion critique, car cela altère les capacités intellectuelles humaines à analyser de manière approfondie les informations disponibles et à évaluer les effets des différents scénarios possibles. De la même façon, l'externalisation des responsabilités vers les systèmes d'IA, affaiblit les compétences cognitives à travers l'effet induit d'une pensée passive et une prise de décision subordonnée à la machine.

Prenons l'exemple d'une application de navigation GPS. Lorsque les individus utilisent régulièrement cette application pour se déplacer, ils peuvent devenir dépendants de ses calculs et prédictions pour planifier leurs trajets. Au fil du temps, ils peuvent perdre l'habitude de consulter des cartes traditionnelles, d'observer leur environnement ou même de se fier à leur propre sens de l'orientation. Cette dépendance excessive à l'égard de l'application GPS entraîne conséquemment une diminution de leurs compétences en matière de navigation, les rendant moins capables de s'orienter de manière autonome dans des situations où l'application est indisponible ou peu fiable.

En outre, si l'application GPS fournit une conjecture erronée ou imprécise, les utilisateurs peuvent être captivés de suivre aveuglément cette orientation sans remettre en question sa validité. Cette sur-confiance dans les capacités de l'application les empêche ainsi d'exercer leur propre jugement et de rechercher des signes contradictoires dans leur environnement. En conséquence, ils seront moins prédisposés à développer et à maintenir leurs compétences en matière de navigation, ce qui les rend plus dépendants de l'application à l'avenir.

Subséquemment, cet exemple illustre comment la dépendance excessive à l'égard de l'intelligence artificielle, comme une application GPS, concourt à la perte de compétences en matière de navigation chez les individus, en affaiblissant leur capacité à se conduire de manière autonome et à exercer un jugement critique quand ils planifient leurs déplacements. Ce que cet exemple rappelle par ailleurs, est qu'au niveau de l'intelligence artificielle, les erreurs de calcul ou de traitement des données existent toujours, et peuvent survenir en raison de divers facteurs tels qu'une mauvaise programmation informatique, des données d'entrée incorrectes ou incomplètes, ou encore des changements de dernière minute dans l'environnement. Ces erreurs produisent parfois des résultats imprévus ou des décisions erronées de la part des systèmes d'IA, compromettant ainsi leur fiabilité et leur efficacité.

V Synergie entre cognition humaine et intelligence artificielle :

Les effets nuancés démontrés jusqu'ici à travers la comparaison entre la cognition humaine et la cognition artificielle mettent en évidence la nécessité d'une approche complémentaire plutôt que substitutive dans leur exploitation. Ces deux intelligences ne peuvent pas être utilisées de manière interchangeable, car elles possèdent des avantages distincts qui peuvent être exploités de manière complémentaire. Les limitations de l'une peuvent être compensées par les forces de l'autre, et vice versa.

La cognition humaine excelle dans la capacité à comprendre les contextes culturels, les nuances linguistiques et les situations imprévues. Elle est également dotée d'une capacité à raisonner de manière flexible, à prendre en compte les valeurs éthiques et à exercer un jugement moral. En revanche, l'intelligence artificielle brille dans le traitement rapide de grandes quantités de données, la détection de modèles complexes et la prise de décision

algorithmique. Elle peut automatiser des tâches répétitives et fastidieuses, permettant ainsi aux humains de se concentrer sur des tâches plus complexes et cognitivement exigeantes.

Ensemble, ces deux composantes forment un écosystème cognitif où les interactions sont continues et bidirectionnelles. Les humains alimentent l'intelligence artificielle en données, en contexte et en expertise, tandis que l'intelligence artificielle fournit aux humains des informations précises, des recommandations pertinentes et des outils pour améliorer leur productivité et leur efficacité.

Cependant, comme dans n'importe quel autre système, l'écosystème cognitif comporte pareillement des conflits et des tensions. Ces antagonismes émergent à différents niveaux, notamment dans la compétition pour les compétences, les questions éthiques liées à la prise de décision, la responsabilité, la transparence et la diversité culturelle. Alors que l'IA offre des opportunités d'innovation et d'efficacité, elle soulève également des préoccupations quant à son impact sur l'identité professionnelle, les valeurs éthiques et la confiance sociale.

L'un des enjeux majeurs de la société contemporaine réside alors dans la montée en puissance et l'expansion des calculs algorithmiques au sein de l'écosystème cognitif. Cette évolution suscite des interrogations profondes quant à l'impact de l'IA sur les vies, les interactions sociales, et les modes de pensée. Ce phénomène dépasse largement le cadre d'un simple débat philosophique pour devenir une inquiétude centrale en matière de gouvernance, d'éthique et de politique. Les questions émergentes concernant la confidentialité des données, la justice algorithmique, et la répartition des pouvoirs entre les acteurs humains et non humains soulignent la nécessité d'une réflexion approfondie et d'une action concertée pour façonner un avenir où l'IA serve véritablement le bien commun tout en préservant les valeurs fondamentales de l'humanité.

Dans ce cadre, les distorsions et les défis rencontrés dans l'écosystème cognitif actuel alimentent déjà les discussions et les réflexions institutionnelles sur la nécessité de repenser les cadres réglementaires et les structures institutionnelles existantes. Face aux avancées rapides de l'intelligence artificielle et à ses implications profondes, les gouvernements, les institutions et les organismes de réglementation se trouvent confrontés à la nécessité de développer de nouvelles politiques, normes et lois pour encadrer l'utilisation de ces technologies de manière éthique, responsable et juste. Cela nécessite une approche holistique et multidisciplinaire, impliquant non seulement des experts en technologie, mais également des spécialistes en éthique, en droit, en sciences sociales et en politique, afin de garantir que l'IA soit déployée de manière à servir l'intérêt public et à préserver les droits et les valeurs fondamentales de la société.

Dans cette dynamique d'adaptation et de régulation, il est crucial de reconnaître que l'essor de l'intelligence artificielle ne se limite pas à une simple évolution technologique, mais représente également une transformation sociétale majeure. L'utilisation généralisée de systèmes d'IA dans les domaines de la santé pourrait par exemple révolutionner les soins de santé, en permettant des diagnostics plus précoces et des traitements plus efficaces ; cependant, cela soulève des questions sur la confidentialité des données médicales et l'accès équitable aux technologies de pointe. De la même façon, dans le domaine de la justice, l'IA peut être utilisée pour prédire les tendances criminelles ou aider à prendre des décisions judiciaires, mais cela nécessite par ailleurs une vigilance accrue pour éviter les biais et les injustices systémiques. En gros, l'intégration croissante de l'intelligence artificielle (IA) dans les sociétés modernes suscite des questions fondamentales qui traversent divers domaines, de la créativité à l'éducation en passant par la gouvernance : dans le domaine de la créativité, l'IA devient un partenaire dans le processus artistique, remettant en question nos conceptions traditionnelles de l'originalité ; en éducation, elle offre un potentiel transformateur en adaptant les méthodes d'enseignement aux besoins individuels des élèves, mais suscite des préoccupations quant à la perte d'interaction humaine ; dans la gouvernance, l'IA permet des

capacités d'analyse de données sans précédent, mais dégage une réflexion approfondie sur les questions de responsabilité et de transparence.

Toutes ces préoccupations et ambivalences indiquent une seule chose : que la cohabitation entre l'intelligence artificielle et la cognition humaine devient un sujet d'une importance capitale qui mène inéluctablement vers un bouleversement radical des structures sociales, de nos politiques éducatives et de nos mécanismes de gouvernance. En cela, l'intelligence artificielle va au-delà de la simple disruption économique pour impacter profondément notre tissu social. Pour naviguer avec succès dans ce nouveau paysage technologique, plusieurs mesures sont prises dans plusieurs pays et d'autres en réflexion, fondées sur des principes scientifiques et illustrées par des exemples concrets :

- Développement éthique: cela requiert une approche méthodique et rigoureuse, où les chercheurs et les développeurs sont tenus d'incorporer des principes éthiques fondamentaux tout au long du processus de conception et de programmation des systèmes d'IA. Il est impératif que des valeurs telles que la transparence, la responsabilité et l'équité soient intrinsèquement intégrées dans les algorithmes de prise de décision. Par exemple, en favorisant la transparence des modèles d'IA et en rendant explicites les critères de décision, les développeurs peuvent atténuer les risques de biais discriminatoires inhérents aux données et aux processus algorithmiques. Une illustration de cette approche est fournie par le Projet Algorithm Watch, fondé en 2016 en Allemagne, avec pour but la surveillance et à l'évaluation des systèmes d'IA. Ce projet se consacre à une évaluation critique de l'utilisation des algorithmes dans divers domaines, mettant en lumière les implications sociales et éthiques et plaident inlassablement pour une transparence accrue et une responsabilité renforcée dans l'utilisation de ces technologies. En effet, ce projet sert de référence incontournable dans la promotion d'une démarche éthique et réfléchie dans le développement de l'IA, en mettant en exergue l'importance cruciale de l'analyse critique et de la surveillance continue des applications algorithmiques afin de garantir leur alignement avec des normes éthiques et sociales rigoureuses.
- Renforcement de l'éducation et de la formation: cela est essentiel pour permettre aux individus de comprendre et de tirer parti des possibilités offertes par cette technologie émergente, tout en relevant les défis qu'elle pose. Les institutions académiques doivent intégrer des modules spécifiques sur l'IA dans leurs programmes afin de fournir aux étudiants une compréhension approfondie de ses concepts, de ses applications et de ses implications dans divers domaines. Par exemple, en étudiant les algorithmes d'apprentissage automatique, les étudiants peuvent acquérir les compétences nécessaires pour concevoir et développer des systèmes d'IA. Aussi, une sensibilisation à l'éthique de l'IA est cruciale pour garantir que ces technologies sont utilisées de manière responsable et respectueuse des droits humains. Le programme "AI for Good" de l'UNESCO illustre parfaitement cette approche en encourageant les jeunes à explorer comment l'IA peut être utilisée pour résoudre des problèmes sociaux majeurs tels que la santé, l'éducation et l'environnement. En formant la prochaine génération de leaders et de professionnels à exploiter l'IA de manière éthique et responsable, ces initiatives contribuent à façonner un avenir où cette technologie est mise au service du bien commun et du progrès social.
- Renforcement de la collaboration interdisciplinaire : cela revêt une importance cruciale pour aborder les enjeux complexes et multifacettes liés à cette technologie émergente. Les initiatives de recherche et de développement en IA doivent impérativement encourager la synergie entre différentes disciplines telles que les sciences, l'éthique, le droit, la sociologie et la politique publique. Cette collaboration

permet de prendre en compte les divers aspects éthiques, sociaux, juridiques et politiques de l'IA, et d'élaborer des solutions holistiques et équilibrées. Par exemple, le "Partenariat mondial sur l'IA" constitue une illustration de cette approche collaborative, en réunissant des chercheurs de renommée mondiale, des décideurs politiques et des experts en éthique pour élaborer des lignes directrices sur l'utilisation responsable de l'IA. En favorisant le dialogue et la coopération entre les différents acteurs impliqués dans le développement et la réglementation de l'IA, de telles initiatives permettent de promouvoir une utilisation éthique, équitable et bénéfique de cette technologie, dans le respect des valeurs fondamentales de la société.

- Promotion de la gouvernance et de la réglementation : c'est aussi une étape indispensable pour garantir une utilisation éthique et responsable de cette technologie émergente. Les gouvernements ont un rôle primordial à jouer en élaborant des politiques et des réglementations qui encadrent de manière rigoureuse l'utilisation de l'IA, en veillant à ce qu'elle soit conforme aux principes éthiques et respectueuse des droits de l'homme. L'Union européenne a démontré son leadership en adoptant le Règlement général sur la protection des données (RGPD), une législation révolutionnaire qui établit des normes strictes pour la collecte, le traitement et la gestion des données personnelles. Cette initiative est une preuve tangible de l'engagement des gouvernements à protéger les droits des individus dans un monde de plus en plus numérisé. A travers la promulgation des lois et des réglementations solides, les gouvernements peuvent créer un cadre juridique propice à l'innovation tout en garantissant la protection des valeurs fondamentales de la société.

- Investissement dans la recherche sur l'intelligence artificielle (IA) éthique et responsable : c'est une démarche impérative pour façonner un avenir où cette technologie est mise au service du bien commun. Les organismes de financement de la recherche ont le devoir de soutenir activement des projets axés sur l'IA éthique, en accordant des subventions à des initiatives qui visent à garantir que le développement et l'utilisation de l'IA respectent les principes éthiques fondamentaux. Le Future of Life Institute, en tant qu'acteur majeur dans ce domaine, joue un rôle crucial en finançant des projets de recherche qui mettent l'accent sur l'élaboration de directives et de frameworks pour une utilisation responsable de l'IA. Cette initiative illustre l'engagement ferme de la communauté scientifique à veiller à ce que les avancées en matière d'IA bénéficient à l'humanité dans son ensemble. En investissant dans la recherche sur l'IA éthique, les organismes de financement contribuent à orienter le développement technologique dans une direction conforme aux valeurs éthiques et sociales, assurant ainsi un impact positif sur la société tout en préservant les droits et la dignité des individus.

L'ensemble de ces initiatives reflète une reconnaissance croissante du besoin d'intégrer l'intelligence artificielle (IA) de manière éthique et responsable dans le système cognitif global. Cela implique une transition vers une approche où les progrès technologiques sont guidés par des considérations éthiques, sociales et juridiques, en plus de la simple innovation technologique. En adoptant des pratiques de développement éthique de l'IA, en renforçant l'éducation dans ce domaine et en promouvant la collaboration interdisciplinaire, la société reconnaît que l'IA ne peut être exploitée pleinement que lorsqu'elle est alignée avec les valeurs humaines fondamentales. La promotion d'une gouvernance et d'une réglementation appropriées, ainsi que l'investissement dans la recherche sur l'IA éthique, indiquent un désir de façonner un avenir où l'IA contribue positivement au système cognitif en respectant les droits, les valeurs et la dignité des individus. En somme, cela signifie que l'intégration de l'IA dans le système cognitif doit se faire de manière réfléchie et éthique, en reconnaissant à la fois ses avantages potentiels et les défis éthiques et sociaux qu'elle pose.

Conclusion :

En conclusion, l'impact de l'intelligence artificielle sur la rationalité des acteurs humains, notamment dans le contexte économique, est à la fois ambivalent et complexe. D'une part, l'IA offre des opportunités d'amélioration de la prise de décision grâce à ses capacités d'analyse de données avancées et de modélisation de scénarios complexes. Cependant, elle comporte également des risques, tels que la dépendance excessive aux algorithmes et la potentialisation des biais existants dans les données et les modèles. Ces effets ambivalents soulignent la nécessité de mesures et de décisions institutionnelles pour atténuer les éventuels affaiblissements de la rationalité humaine.

Pour contrer ces risques, des mesures institutionnelles doivent être mises en place pour réguler l'utilisation de l'IA dans les processus décisionnels économiques. Cela pourrait inclure la mise en œuvre de cadres réglementaires stricts pour garantir la transparence et la responsabilité dans l'utilisation de l'IA, ainsi que des normes éthiques pour guider le développement et le déploiement de ces technologies. Il est essentiel d'investir dans la sensibilisation et la formation des acteurs économiques aux enjeux liés à l'utilisation de l'IA, afin de renforcer leur capacité à prendre des décisions éclairées et à comprendre les limites de ces technologies.

En parallèle, des mécanismes de surveillance et d'évaluation continus sont nécessaires pour surveiller l'impact de l'IA sur la rationalité des acteurs économiques et pour ajuster les politiques et les réglementations en conséquence. Cela pourrait impliquer la création d'organismes de régulation dédiés chargés de surveiller l'application des normes et des directives en matière d'IA dans le domaine économique.

En substance, la gestion des effets ambivalents de l'intelligence artificielle sur la rationalité des acteurs économiques requiert une approche holistique et proactive, impliquant à la fois des mesures réglementaires et des actions de sensibilisation et de formation. Seules des décisions institutionnelles et des initiatives concertées permettront de garantir que l'IA contribue de manière positive à l'amélioration des processus décisionnels économiques tout en préservant la rationalité et les valeurs fondamentales des acteurs humains.

Bibliographie :

- Autor, D. H., & Levy, F. (2003). *The economics of work and pay*: Seventh edition. Prentice Hall.
- Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review*, 104(3), 671-732.
- Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5(2), 153-163.
- Diakopoulos, N. (2016). Accountable Algorithmic Journalism. *Digital Journalism*, 4(7), 937-958.
- Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. Retrieved from <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>
- Goodman, B., & Flaxman, S. (2016). European Union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine*, 38(3), 50-57.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
- Lee, S., Chung, M., & Lee, J. (2015). Personalized recommendation based on users' personality and emotion in social networking services. *International Journal of Distributed Sensor Networks*, 11(6), 435713.
- Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, 31(2), 87-106.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.

<https://doi.org/10.1126/science.aax2342>

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135-1144).

Simon, H. A. (1956). Rational Choice and the Structure of the Environment. *Psychological Review*, 63(2), 129-138.

Negnevitsky, M. (2011). *Artificial Intelligence: A Guide to Intelligent Systems*. Pearson.