



**RÉPUBLIQUE
FRANÇAISE**

*Liberté
Égalité
Fraternité*

Vie publique
Au cœur du débat public

Comment faire de l'intelligence artificielle une technologie vertueuse pour l'emploi et le travail ?

Publié le 11 octobre 2024

Temps de lecture 19 minutes

Par : Flore Barcellini ; Nathalie Greenan et Moustafa Zouinar

Les opportunités comme les risques de l'intelligence artificielle (IA) pour le monde du travail sont nombreux. Pour l'instant, leur évaluation scientifique reste limitée. Que suggèrent les résultats empiriques en économie et en ergonomie et comment les pouvoirs publics peuvent-ils favoriser une approche nuancée et vertueuse des usages de l'IA ?

L'IA : une technologie en pleine mutation qui devient une affaire politique et sociale

L'Intelligence artificielle (IA) connaît actuellement un regain d'intérêt sans précédent dans les sphères politiques et socio-économiques. Les discours qui en émanent prônent une nouvelle ère d'innovations technologiques et de services, à même de transformer le monde du travail. Face à cette évolution, plusieurs pays ont mis en place des plans stratégiques visant à soutenir et encadrer le développement de l'IA. En France, le rapport Villani de 2018 a conduit au lancement du Programme national de recherche en IA (PNRIA). La Commission de l'intelligence artificielle a publié un rapport en mars 2024, contenant 25 recommandations tandis qu'au niveau européen, le Règlement sur l'IA, qui établit des règles harmonisées fondées sur les risques, a été adopté par le Conseil de l'Europe en mai 2024.

Les interrogations sur l'IA au travail ne sont pas pour autant nouvelles. Les premiers systèmes d'intelligence artificielle (SIA) (années 1960-70), basés sur l'IA symbolique qui repose sur l'utilisation de modèles de processus cognitifs (raisonnement, perception, compréhension, etc.) prédéfinis par les concepteurs à l'aide de règles et de symboles, étaient envisagés comme des outils d'assistance dans divers domaines professionnels, mais leurs usages sont restés assez limités. Les SIA actuels, basés sur une approche "connexionniste" et des algorithmes d'apprentissage dit profond qui consiste à faire apprendre à des réseaux de neurones artificiels à réaliser des tâches cognitives sans modèles prédéfinis, sont devenus beaucoup plus performants.

Certains de ces systèmes restent circonscrits à des domaines spécifiques, tels que l'analyse d'images ou le traitement automatique du langage avec des utilisations dans différents secteurs (par exemple, le transport avec les véhicules (semi)-autonomes, l'imagerie médicale, la relation client, la maintenance prédictive) sur des tâches bien définies (Zouinar, 2020), d'autres comme ChatGPT, lancé en novembre 2022, ont ouvert de nouvelles perspectives sur l'évolution potentielle des métiers et du travail. Ces systèmes se distinguent des précédentes applications d'IA "mono-tâche" par leur polyvalence, offrant de multiples champs d'application : ils peuvent traduire, calculer, générer des textes ou des programmes informatiques, des images, des sons et des vidéos. Grâce à son accessibilité et sa large diffusion, l'IA générative s'approche aujourd'hui des applications web grand public, comme les moteurs de recherche et pourrait devenir un outil professionnel comme un autre.

Type d'IA	Description	Exemples
-----------	-------------	----------

Type d'IA	Description	Exemples
Symbolique	Utilise des symboles et des règles logiques pour modéliser des processus cognitifs (raisonnement, compréhension, résolution de problèmes, perception)	Systèmes experts (ex: MYCIN pour le diagnostic médical)

Type d'IA	Description	Exemples
Connexionniste dont...	Basée sur des réseaux de neurones artificiels qui apprennent et développent des modèles numériques à partir de données	

Type d'IA	Description	Exemples
...faible	Conçue pour des tâches spécifiques, sans conscience, ni compréhension du sens des données manipulées	<ul style="list-style-type: none"> - Reconnaissance d'objets, de visage ou de scènes (ex: classification d'objets dans Google Photos) - Traduction automatique (ex: DeepL, Google Translate)

Type d'IA	Description	Exemples
...boîte noire	Modèles d'apprentissage complexes dont les processus de production de résultats sont difficiles à déterminer	<ul style="list-style-type: none"> - Réseaux de neurones profonds utilisés dans le trading algorithmique, certains systèmes de recommandation (ex: algorithme de recommandation de Netflix), les grands modèles de langage comme Chat-GPT ou la reconnaissance faciale - certains algorithmes à base d'arbres de décisions (exemple : algorithmes Random Forest)

Type d'IA	Description	Exemples
...génération	Capable de créer du contenu basé sur des données d'apprentissage	- GPT-3 pour la génération de texte - DALL-E pour la création d'images - MuseNet pour la composition musicale

Type d'IA	Description	Exemples
... forte ou générale (hypothétique)	IA avec une intelligence comparable ou supérieure à l'humain, avec conscience et compréhension du sens des données	Actuellement théorique

Les opportunités et les risques pour l'entreprise et le salarié

L'histoire de l'IA a connu des cycles d'enthousiasme et de désillusion. Actuellement, enthousiasmés par les récentes avancées de l'IA, les acteurs économiques y voient une opportunité d'automatisation et de transformation des systèmes de production, notamment avec des objectifs de performance. Les promesses associées à l'IA reposent essentiellement sur trois hypothèses :

- du côté du travail, l'IA est supposée augmenter les capacités humaines dans la réalisation de certaines tâches, contribuer à réduire les erreurs en fiabilisant des tâches ou soulager les humains de tâches perçues comme "sans valeur ajoutée", notamment lorsqu'elles sont répétitives ou pénibles. Dans le domaine médical par exemple, l'IA a montré des résultats prometteurs pour fiabiliser des tâches de diagnostic liés à l'analyse d'images. Dans la finance, elle facilite la détection des fraudes ou des anomalies de transaction. Dans l'industrie, les robots guidés par l'IA pourraient effectuer des tâches dangereuses ou pénibles ;

- l'IA ouvre également des opportunités en termes d'emploi et de structure de production. Le développement de la technologie, sa mise en œuvre et l'innovation de stimulus devrait générer des nouveaux emplois. Certains voient même dans l'industrie 4.0 incluant l'IA une opportunité de réindustrialisation pour l'économie européenne ;

- l'IA répondrait également en partie à des enjeux de pénurie de main d'œuvre et de vieillissement de la population. Cette vision se retrouve dans le projet de société 5.0 promue par les pouvoirs publics japonais.

Néanmoins, il convient de noter que ces hypothèses n'ont fait l'objet que de peu d'évaluations scientifiques solides en contexte réel. Par ailleurs, l'IA actuelle relève principalement de l'IA dite "faible", spécialisée dans des tâches précises, et non d'une IA "forte" ou "générale" capable de remplacer entièrement l'intelligence humaine dans tous les domaines.

Au-delà de ces promesses, l'intégration des SIA actuels dans le monde du travail présente également des risques majeurs :

- la substitution partielle ou totale de l'humain, impliquant la perte de son expertise et de ce qui donne du sens à son travail : ceci se manifeste à travers l'automatisation croissante des tâches, y compris celles considérées comme "cognitives" et basées sur l'expérience et l'expertise des professionnels (par exemple les tâches de diagnostic médical). Cette évolution touche désormais des métiers dits qualifiés (informaticien, comptable, traducteur, etc.), entraînant une possible déqualification de certains travailleurs et une polarisation entre les tâches considérées comme à forte valeur ajoutée et celles jugées routinières. Cette approche néglige souvent l'expérience réelle des professionnels, l'importance des dimensions collectives du travail, et la valeur réelle de ces tâches pour les processus de production. De plus, les modèles d'IA se caractérisent par de nombreuses limites de nature "cognitive" réduisant leur capacité réelle à remplacer totalement l'humain dans des tâches nécessitant une compréhension fine du contexte ou une prise de décision complexe. À ceci s'ajoute l'opacité des modèles d'IA qualifiés de "boîtes noires", dont le fonctionnement interne est difficile ou impossible à interpréter, même pour ses concepteurs. Si ces imperfections des SIA sont sous-estimées par les organisations qui imposent leur utilisation, les risques se multiplieront. Cela pourrait compromettre la fiabilité des décisions prises par ou avec une IA, la crédibilité des professionnels, le sens et la qualité de leur travail, et à terme leur santé ;

- la subordination croissante aux systèmes algorithmiques et aux contraintes des organisations du travail : elle se manifeste à travers le développement du management algorithmique ou la rigidification des règles organisationnelles du travail liée à l'introduction de l'IA. Dans ces contextes, l'IA est utilisée pour attribuer, optimiser, superviser et évaluer les tâches humaines, ouvrant la porte à une surveillance accrue des performances et des comportements des travailleurs. La perte d'autonomie pour les professionnels pourrait advenir via des usages de l'IA qui entraînent la diminution de leur capacité décisionnelle et de l'exercice de leur jugement professionnel. De tels systèmes tendraient à optimiser les performances individuelles plutôt que collectives, fragmentant les équipes et limitant la coopération entre travailleurs ;

- le mirage de la collaboration entre humains et SIA : la collaboration humain-IA renvoie à l'idée qu'humains et SIA ne seraient pas substituables mais pourraient travailler en collaboration, augmentant leurs capacités réciproquement les SIA augmentant potentiellement les humains "formant" les SIA avec des données massives structurées, en fournissant des explications sur ses actions et en surveillant son bon fonctionnement. Cependant, le terme de collaboration paraît excessif car les SIA sont loin de disposer de capacités cognitives et sociales nécessaires à la

collaboration, par exemple : manque de capacités de communication, de compréhension des intentions et des buts poursuivis par les humains, difficulté à interpréter le contexte d'action et à s'ajuster (Zouinar, 2020).

Apports de l’IA	Risques et conséquences négatives
------------------------	--

Apports de l'IA	Risques et conséquences négatives
<p>Détection de fractures : - "Filet de sécurité" en cas de fatigue ou de situations de travail "dégradées" (fatigue, travail de nuit, forte charge de travail) propices à des erreurs de diagnostic; - Gain de temps</p> <p>Dictée à reconnaissance vocale : - Gain de temps dans la production des comptes-rendus radiologique</p>	<p>Détection de fractures : - Confiance excessive dans le système ; - Diminution de l'acquisition de compétences pour les jeunes radiologues- Perte de temps dans le traitement des faux positifs</p> <p>Dictée à reconnaissance vocale : - Élimination du travail des secrétaires dans la production des comptes-rendus ; - Perte pour les secrétaires d'une activité source de développement ; - Moins d'opportunités d'interactions entre radiologues et secrétaire ; - Perte du contrôle de la qualité des comptes-rendus finaux ; - Augmentation de la charge de travail du fait de la nécessité de corriger les erreurs du système</p>

Au-delà de l'IA : les vrais moteurs de la révolution du travail

Si les SIA offrent des opportunités pour améliorer la productivité et transformer le monde du travail, leur intégration dans les organisations soulève des défis importants en termes d'autonomie des travailleurs, de travail collectif et de préservation de l'expertise humaine. Il apparaît nécessaire d'adopter une approche équilibrée, reconnaissant à la fois le potentiel de l'IA pour améliorer certains aspects du travail tout en évitant les écueils de la substitution et de la subordination (Zouinar, 2020 ; Acemoglu et al., 2023). L'analyse critique des résultats empiriques récents obtenus par la recherche en économie et en ergonomie invite à questionner une forme de "solutionnisme", qui conduit à voir les technologies comme des "remèdes simples et sur étagère" à des problèmes complexes comme des enjeux de santé, de performance ou de réduction des coûts.

Ainsi, l'approche ergonomique nous montre que l'introduction d'une technologie comme l'IA dans une situation de travail ne détermine pas à elle seule les usages sociaux et organisationnels qui se déploieront (Gamkrelidze, 2022). Les conséquences de l'introduction notamment au niveau organisationnel - et l'emploi résultent de multiples décisions managériales dans les entreprises et les institutions qui sont le fruit d'arbitrage entre différentes "rationalités", par exemple financière et technique et qui s'inscrivent dans une histoire et un contexte (manque de personnel, augmentation de la charge de travail, fort turn-over, etc.). Ces décisions ne prennent souvent pas en compte le travail réel et ses contraintes et misent essentiellement sur la technique comme seule réponse à des problèmes dont les causes sont en fait systémiques.

Du côté des approches économiques, de nombreux travaux considèrent qu'il y a transformation technologique dans une entreprise dès lors qu'elle introduit une technologie émergente dans son système de production. Or l'entreprise peut tout à fait sous-utiliser cette technologie, continuer à traiter ses affaires courantes et à produire "comme d'habitude" ou avoir sous-documentées les conséquences néfastes de cette technologie au moment où la décision de l'introduire est prise. Dans ce cas, la transformation technologique et l'innovation amorcées restent incomplètes, voire non advenues.

Les travaux du projet européen Beyond 4.0 soulignent que la technologie, aussi bien conçue soit-elle, n'est pas le seul facteur qui intervient dans ce processus. La transformation technologique dépend ainsi des investissements conjoints des entreprises dans les technologies émergentes, la R&D et la capacité d'apprentissage de l'organisation, essentielle pour développer des usages "vertueux" des technologies et minimiser les risques.

L'enjeu des politiques publiques : promouvoir des "scénarios vertueux" d'usages de l'IA

L'introduction d'une technologie émergente telle que l'IA implique plusieurs scénarios, certains seront vertueux pour les humains et l'environnement (redéploiement des compétences au sein des métiers, préservation de la santé des travailleurs, développement des capacités à préserver l'environnement...), d'autres seront susceptibles d'accroître l'épuisement des ressources (fortes restructurations des métiers, pertes de savoirs et d'expertise, atteintes à la santé des travailleurs, moindre maîtrise des coûts environnementaux). L'enjeu des recherches et des politiques publiques est alors d'agir en amont du processus de développement de la technologie et de définition de ses usages pour favoriser des scénarii vertueux plutôt que de n'intervenir qu'en aval pour corriger des problèmes lorsqu'ils se manifestent, alors qu'ils auraient pu être anticipés en investissant dans la capacité d'apprentissage des organisations

et dans les capacités des professionnels. Pour favoriser l'émergence des scénarios vertueux d'usages des IA, c'est-à-dire des usages soutenables, deux éléments sont centraux :

- d'une part, les usages des SIA doivent s'inscrire dans des formes de travail soutenables (Gaudart et al., 2019), qui conjuguent la performance des organisations avec la préservation de la santé des travailleurs et le développement de leurs compétences tout au long de la vie professionnelle ;
- d'autre part, l'introduction de l'IA doit améliorer la qualité des connaissances produites et faciliter l'exploration de savoirs nouveaux sur le processus de production des entreprises et le travail des professionnels.

Le développement de la capacité d'apprentissage des organisations joue positivement sur ces deux éléments. Pourtant, cette capacité a plutôt stagné sur les deux dernières décennies. Face à la globalisation des économies, puis la grande récession, les entreprises auraient en effet privilégié un investissement technologique pour rationaliser et optimiser leurs processus de production tout au long des chaînes de valeur (Bodrogi et Adler, 2018) et négligé le développement de la capacité d'apprentissage des organisations. Pourtant ce sont les conduites de projets réellement participatives, c'est-à-dire ouvertes à l'inversion des rapports de force dans les entreprises et institutions et ancrées dans des modèles du travail "tel qu'ils se font", qui permettront au tissu économique et à la société de tirer collectivement les meilleurs bénéfices des SIA (Barcellini, 2019 ; Gamkrelidze, 2022 ; Barcellini et Lefeuvre, 2023). Les dispositifs existants de dialogue social pourraient être mobilisés pour favoriser les investissements dans la capacité d'apprentissage des organisations.

Inversement, si l'on ne promeut pas cette capacité d'apprentissage des organisations, les scénarios d'épuisement des humains et de l'environnement pourraient se développer dans la course à la réduction des coûts. Une concentration trop forte des entreprises sur le marché des produits (De Loeker et al. 2020), notamment celui des IA génératives et de ses applications dérivées (Korinek et Vipra, 2023), mais aussi une concentration trop importante des employeurs sur le marché du travail (Araki et al. 2023) sont des éléments de contexte défavorables aux scénarios vertueux d'usage et de déploiement de l'IA.

Une attention particulière doit donc être portée aux formes d'innovation qui favorisent la concentration sur les marchés des produits et du travail et à la façon dont elles sont soutenues par les autorités publiques au travers de dispositifs comme le crédit d'impôt recherche. Ainsi, les dépenses de R&D destinées à la standardisation et à l'innovation marketing devraient être identifiées et suivies dans les budgets de R&D. Poser des limites aux formes de management algorithmiques et de surveillance automatisée en prenant appui sur le cadre législatif existant (RGPD par exemple) et sur la négociation collective (Doellgast et al., 2022) est un autre élément important pour favoriser une innovation centrée sur l'humain. Une attention particulière doit également être portée aux inégalités de genre liées aux usages des IA car les impacts identifiés sur ces inégalités sont plus souvent négatifs (Bianchini et al., 2023). Ensuite, cette littérature identifie aussi des tensions entre le développement de l'IA et celui des technologies vertes.

Ces constats appellent donc à une analyse fine des motifs et stratégies des entreprises et institutions quant à l'introduction de l'IA, aux modèles de conduites de projets, de conduite des changements organisationnels et de positionnement sur les marchés qu'elles mettent en œuvre. Ces constats appellent également à une réflexion sur les discours et éléments de langage énoncés quant aux objectifs de développement durable Développement qui répond aux besoins du présent sans compromettre la capacité des générations futures à répondre aux leurs poursuivis.