L'IA générative : repères, enjeux et contextualisation

Ghislaine Chartron

L'IA générative : repères, enjeux et contextualisation

Ghislaine CHARTRON
Professeur du CNAM, Chaire d'ingénierie documentaire, INTD et DICEN-Idf
ghislaine.chartron@lecnam.net

revue Mediadoc, décembre 2023, n°31

Depuis le lancement de ChatGPT par la société OpenAI en novembre 2022, l'IA et surtout l'IA générative (IAG) se sont installées quotidiennement dans les débats sur la transformation de nombreuses activités humaines, et notamment les activités liées au traitement de l'information et à l'enseignement. L'objectif de cet article est de rappeler quelques repères fondamentaux sur ces technologies, d'en apprécier quelques enjeux et limites dans le contexte de l'exercice des professeurs documentalistes.
Nous aborderons successivement des repères technologiques, des exemples de services, le rôle central de la qualité des données, des usages dans le contexte qui est le nôtre, et enfin les enjeux actuels de la régulation dans une vision de *co-design* en confiance avec ces technologies.

Repères sur les technologies

L'Intelligence artificielle prend ses racines dans le sillage de la cybernétique au cours des années 1940-1960. Son développement est étroitement lié aux développement des ordinateurs, de leurs capacités de calcul et l'objectif poursuivi est de simuler le raisonnement humain par une machine dans des tâches de plus en plus complexes. Les travaux de John Von Neumann et d'Alan Turing ont conduit à l'architecture de nos ordinateurs contemporains. L'éventuelle intelligence des machines a été formulée par Turing dans son célèbre test qui continue à challenger les équipes de développeurs: si un humain dialoguant avec une IA ne se rend pas compte que c'est une machine qui lui répond, alors la frontière humain-machine est atteinte et le progrès technologique couronné de succès[1].
Herbert Simon a prophétisé en 1957 que l'IA arriverait à battre un humain aux échecs, ce que fit Deep Blue (système expert d'IBM) en 1997 au jeu d'échec contre Garry Kasparov.
Les années 1980-1990 furent l'âge d'or des systèmes experts avec l'avènement des premiers microprocesseurs en 1972. Les technologies s'appuyaient alors sur des bases de connaissances et des ensembles de règles qu'il fallait établir au préalable dans chacun des domaines d'application. Le système expert Mycin spécialisé dans le diagnostic des maladies du sang et la prescription de médicaments fut l'un des premiers déployés en médecine. Mais les limites furent vite atteintes : l'élaboration des bases de connaissances et des règles est une lourde tâche, spécifique à chaque domaine et en évolution constante. Un désenchantement de l'IA s'en est suivi dans les années 90. Le rebond marqué depuis les années 2010 fut conjoncturel à de nouvelles puissances de calcul à des prix compétitifs, à la disponibilité croissante de données massives. Le changement de paradigme fut amorcé, remettant les calculs statistiques au premier plan. Il ne s'agit plus de codage de règles et de connaissances mais de laisser les machines découvrir les corrélations et les classifications par des calculs statistiques massifs. L'apprentissage des machines s'appuie désormais sur des modélisations statistiques et calculatoires variées. L'apprentissage profond (le *deep learning*) fondé sur des réseaux de neurones artificiels est aujourd'hui le plus prometteur, au cœur des IA génératives de contenu.

[1] Histoire de l'intelligence artificielle, https://www.coe.int/fr/web/artificial-intelligence/history-of-ai

Il faut donc retenir deux approches du développement de l'IA au cours des dernières décennies :

- L'IA symbolique repose sur des systèmes de règles et de bases de connaissances élaborées par des humains et intégrées aux logiciels. Leur périmètre est limité à un domaine particulier ; ce type d'IA est souvent attribué du qualificatif d'*IA faible*.
- L'apprentissage machine est fondé sur des modèles statistiques et des algorithmes auto-apprenants. Les algorithmes d'apprentissage sont nombreux : forêt d'arbres de décision, régression, clusterisation, K-moyennes … Chaque algorithme est conçu pour traiter un type particulier de problème d'apprentissage automatique. Le deep-learning est, quant à lui, fondé sur des réseaux de neurones, nécessitant de gros volumes de données d'apprentissage. Le périmètre d'application est infini, ce type d'IA est souvent attribué du qualificatif d'*IA forte*.

Les développements actuels peuvent adopter une méthode combinatoire de ces deux approches.

Les grands modèles linguistiques (LLM) sont des systèmes d'intelligence artificielle formés sur des ensembles de données textuels massifs. Les LLM sont souvent construits avec un type de réseau neuronal appelé transformateur. Les transformateurs peuvent apprendre des dépendances statistiques à longue portée entre les mots, essentielles à la compréhension et à la génération d'un langage naturel. Les modèles de transformateur comprennent plusieurs couches, chacune effectuant une tâche différente. Une fois que le modèle a appris les relations entre les mots, il peut générer un nouveau texte similaire à celui sur lequel il a été formé. Il n'y a donc, dans ce contexte, aucune représentation formelle du langage par la machine, que des statistiques combinatoires…
La concurrence entre LLM est forte et l'accélération des développements est impressionnante en cette fin 2023 notamment entre les géants du numérique. La différenciation se fait sur la complexité des tâches résolues, avec des modèles dont la performance dépend à la fois des données d'apprentissage et des milliards de paramètres possibles. Citons quelques LLM :

- GPT (avec les versions de 1 à 5), modèle d'OpenAI (texte et image),
- ChatGPT : LLM pour les chatbots, OpenAI,
- LLaMA : LLM développé par META,
- PalM : LLM de Google,
- FLAN UL2 de Hugging Face…

Il faut enfin souligner que dans cette course folle qui est lancée, la dépendance aux LLM devient un sujet stratégique prioritaire. En France et en Europe les acteurs sont de plus en plus dépendants des LLM pré-entrainés d'origine nord-américaine, comme l'API d'OpenAI. En riposte, les initiatives européennes et nationales se multiplient. Mistral AI[2], une start-up française fondée en avril 2023, spécialisée dans le développement de l'intelligence artificielle, est présentée comme une alternative à OpenAI. Le laboratoire KYUTAI[3] a été lancé en novembre 2023 par Xavier Niel et une équipe dont plusieurs membres ont eu une expérience chez les Gafam.

[2] https://mistral.ai/
[3] https://www.lemondeinformatique.fr/actualites/lire-ia-iliad-lance-le-laboratoire-de-recherche-kyutai-92163.html

Les services et leurs caractéristiques majeures

On peut distinguer pour le moment :

- Les IA génératives de textes, des Chatbots très souvent (ChatGPT d'OpenAI, Bard de Google, Claude d'Anthropic, Jasper.ai d'AIexplorer, …),
- Les IA génératives d'images (Midjourney, Stable-Diffusion, Dall-E 2 d'OpenAI…),
- Les IA génératives de vidéos (D-ID, Synthesia, Pictory, …).

Mais la convergence est amorcée. Fin 2023, Google vient d'annoncer son modèle Gemini qui serait multimodal, donc capable de traiter tout type de contenu : image, texte, musique, code informatique, langage oral…

Pour exemple, ChatGPT (Chat Generative Pre-trained Transformer) est un agent conversationnel basé sur les LLM d'OpenAI, il est capable de répondre à des questions, de compléter des phrases, de traduire des textes, de résumer des textes, de générer des textes selon certaines recommandations de style, de tenir des conversations avec des humains (Linc-Cnil, 2023).

Si les premières IA génératives ont commencé par être testées de façon isolée, la tendance est désormais l'intégration dans les environnements de travail, au plus proche des activités, ce qui est révélateur d'une maturité croissante. Ainsi les moteurs de recherche intègrent progressivement ces technologies comme le service de conversation de Bing (cf. figure 1), l'assistant *Copilot* est associé à la suite Microsoft, la plateforme IBM-Watsonx est déployée au plus près des besoins des métiers.

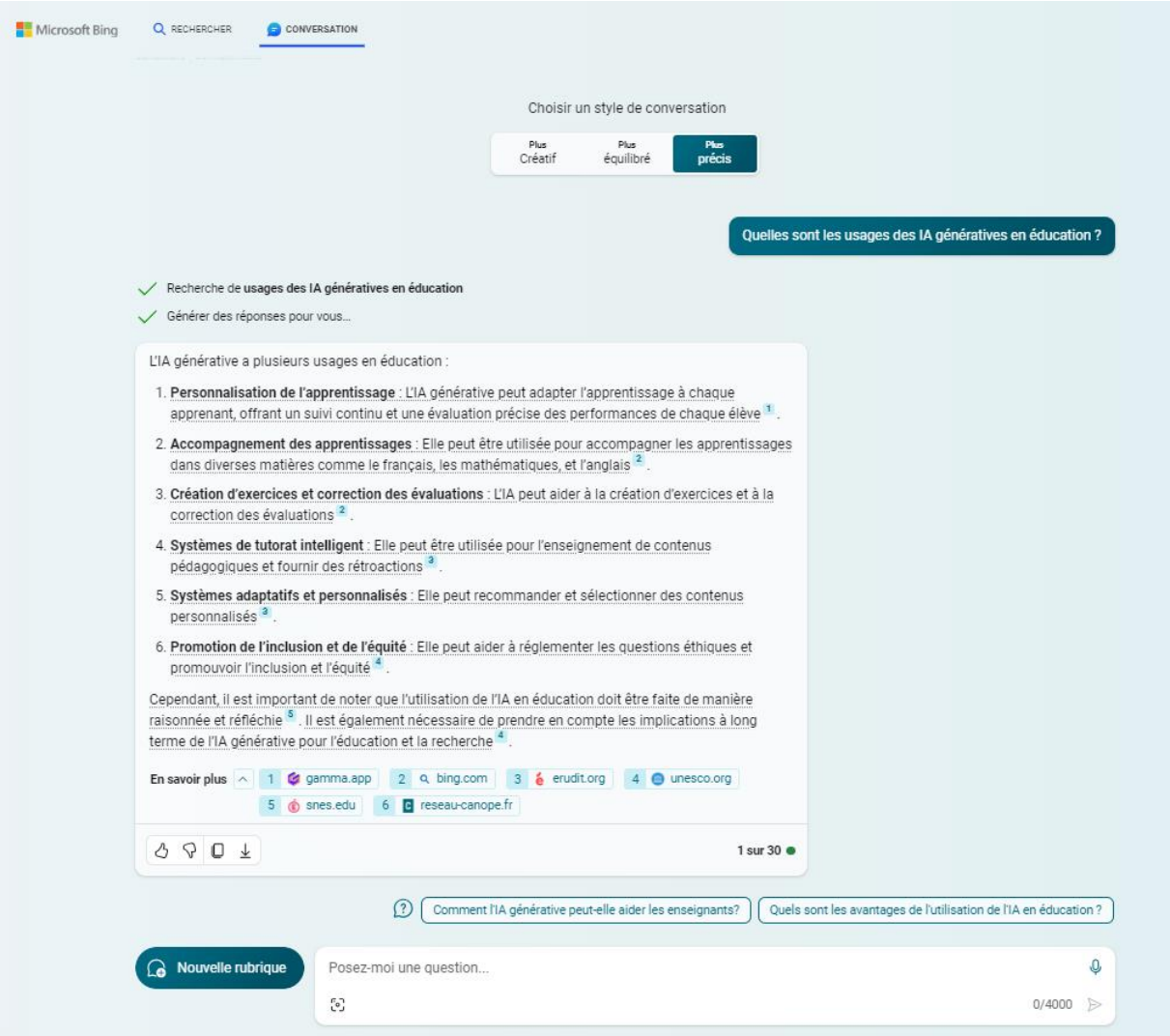

Figure 1 : Intégration de chatGPT dans l'interface du moteur Bing

Les modes d'interaction humaine avec ces technologies sont par ailleurs renouvelés, le *prompt* et le *fine-tuning* ont fait leur apparition. Un prompt est une instruction que l'on donne à une IA qui va l'interpréter et proposer un résultat. On pourrait le comparer à une requête posée à un moteur de recherche, mais le dialogue se fait désormais en langage naturel ; la compréhension pragmatique du fonctionnement des algorithmes conduit souvent à formuler les mêmes recommandations :

- Utiliser un langage simple,
- Intégrer des exemples de réponse ou de format attendu,
- Intégrer des éléments de contexte,
- Reformuler, affiner progressivement les requêtes.

Le fine-tuning (réglage fin) permet, quant à lui, d'optimiser les performances d'un modèle pré-entraîné existant en ré-entraînant le modèle sur des données spécifiques de manière à ajuster ses poids et ses paramètres. Donc c'est un apprentissage supervisé sur des données labellisées. L'objectif est d'adapter un système d'IA à des tâches ou à des domaines spécifiques comme classer des documents juridiques, générer des résumés d'articles médicaux…

Autre point majeur, les données d'apprentissage sont au centre du fonctionnement des IAG et leur qualité impacte prioritairement la pertinence des services fournis. La majorité des IAG grand public « se nourrissent » des données du web en accès libre, Wikipédia est ainsi une source importante ainsi que toutes les ressources en *open access* avec des licences permissives de réutilisation. Des pratiques illicites ou non consenties ont déjà été au cœur des débats comme le cas de la plateforme Books3[4], une gigantesque base de données contenant près de 200 000 livres piratés, dont s'est servi notamment Meta pour l'apprentissage profond. Le *scrapping* des sites de media est aussi au cœur de certaines tensions. L'analogie peut être faite avec les conflits qui ont opposé Google et les media pour le service Google news, le moteur ayant tiré profit, dès ses débuts, des fils d'actualités des sites de presse sans compensation pour les producteurs des contenus. Le partage de la valeur est la question centrale dans l'innovation de ces services, la création en 2019 d'un droit voisin pour les agences de presse fut la régulation trouvée au niveau européen[5] concernant Google news.

Protéger ses données ou les ouvrir pour les IAG grand-public ? C'est aussi un nouveau dilemme pour les producteurs de contenus, éditeurs, agences de presse… La directive européenne du droit d'auteurs de 2019 a donné la possibilité d'activer une clause d'*opt-out* de scrapping concernant les usages « non recherche » et de nombreux éditeurs l'ont activée principalement face à Google. La stratégie poursuivie est de réserver ses données pour des IAG développées sur ses propres plateformes, comme l'on fait Getty Image, et les éditeurs juridiques par exemple. Mais le risque plus politique et peut-être plus fondamental pour nos sociétés est de laisser se développer des usages grand-public d'IAG nourries avec des données de faible qualité. Désinformation, manipulation d'opinions en seront un risque majeur. Nous pourrions assister à des négociations similaires à celles déployées pour Google news à court terme.
Les jeux de données sont aussi porteurs de nombreux biais, comme le biais culturel que voudrait contrer la récente initiative d'organismes publics **«** Villers-Cotterêt » lançant une base de données expérimentale destinée à combattre les biais culturels des IA majoritairement anglo-saxonnes[6].

Les usages

Ce sont tous les secteurs qui expérimentent aujourd'hui l'intégration de l'IA dans leurs activités. Des percées notables sont à remarquer en médecine, en droit et en finance à en croire la veille que nous propose la presse spécialisée sur l'IA[7].

Concernant plus spécifiquement les IAG, les plateformes de contenu sont de plus en plus nombreuses à co-construire désormais leurs services avec ces technologies (Chartron, Raulin, 2022). C'est le cas de la documentation et de l'édition juridique qui proposent une diversité de nouvelles fonctionnalités aux professionnels du droit :

- Des interactions en langage naturel,
- Des recommandations contextualisées,
- Des arguments jurisprudentiels,

[4] https://www.usine-digitale.fr/article/des-maisons-d-edition-obtiennent-le-retrait-de-books3-une-gigantesque-base-de-donnees-utilisee-pour-entrainer-des-modeles-d-ia.N2161982
[5] https://eur-lex.europa.eu/FR/legal-content/summary/copyright-and-related-rights-in-the-digital-single-market.html
[6] https://presse.economie.gouv.fr/intelligence-artificielle-letat-sengage-pour-rendre-laction-publique-plus-simple-plus-efficace-au-benefice-des-francais/
[7] https://www.actuia.com/

- Des réponses précises, raisonnées, étayées, immédiatement utiles,
- Des arguments de doctrine administrative “Pour et Contre”,
- Des conseils et des explications juridiques détaillés…

Par exemple, Lefebvre Sarrut, leader européen de la connaissance juridique et fiscale, a lancé le déploiement de GenIA-L sur ses plate-formes en matière d'édition juridique.

Les institutions culturelles expérimentent également ces innovations à différents niveaux : aide à l'indexation, catalogage, reconnaissance d'écriture manuscrite, exploration et analyse des collections, aide à la décision pour la gestion des collections[8]. Plus largement, les expériences dans le domaine de la documentation se multiplient[9]. Les auteurs expérimentent, de leur côté, le travail d'écriture avec des IA comme co-auteur[10].

Concernant l'enseignement, les élèves, étudiants et enseignants ont bien entendu testé l'intégration des IAG dans leur quotidien : produire une dissertation, un état de l'art sur un sujet, résoudre un problème mathématique, générer un programme informatique, faire un plan de cours, … Des assistants personnalisés sont testés comme Khanmigo[11] de la Khan Academy, des services spécialisés par matière émergent comme MathGPT[12] pour les mathématiques.
Le guide de l'Unesco, *Guidance for Generative AI in education and research* (Unesco, 2023), identifie deux usages majeurs pour l'enseignement : le co-design des programmes de cours et le chatbot comme assistant personnel de l'apprenant. Des manuels sont également initiés pour accompagner les enseignants (Colin de la Higuera et Jotsna Iyer, 2023).

Les accueils sont aussi controversés : interdire ou intégrer ces nouvelles technologies ?
L'interdiction présente l'écueil d'écarter du processus d'apprentissage des technologies qui deviennent communes dans les activités quotidiennes et, de fait, isoler les institutions éducatives en conséquence. Par ailleurs, l'intégration non contrôlée peut nuire à la formation du raisonnement, de l'esprit critique et conduire à une perte de capacités rédactionnelles chez l'apprenant. Le critère d'efficience qui sous-tend l'usage de ces technologies (individualisation de l'apprentissage, coaching individuel) doit être confronté à des valeurs plus éducatives du collectif et du vivre ensemble.

Dans l'enseignement supérieur, l'enquête réalisée par Compilatio et l'institut de sondage Le Sphinx dans les universités de France du 21 juin au 15 août 2023 auprès de 1242 enseignants et 4443 étudiants (Compilatio, 2023), montre que déjà 1 étudiant sur 2 (55%) déclare utiliser un outil d'IA générative au moins occasionnellement. Les usages déclarés concernent essentiellement l'appréhension d'un sujet et l'aide à la rédaction (cf. figure 2). Par contre, les outils d'IA générative ne remplacent pas les moteurs de recherche : les étudiants déclarent encore à 77% utiliser les moteurs de recherche comme principale source de documentation.

[8] https://www.bnf.fr/fr/une-feuille-de-route-pour-lintelligence-artificielle-la-bnf, https://francearchives.gouv.fr/fr/actualite/491300912
[9] https://www.archimag.com/tags/intelligence-artificielle
[10] Raphaël DOAN travaille avec une IA comme co-auteur dans son roman: Si Rome n'avait pas chuté, 2023, https://academiesciencesmoralesetpolitiques.fr/2023/06/12/raphael-doan-si-rome-navait-pas-chute-2023/
[11] https://www.khanacademy.org/khan-labs
[12] https://www.mathgpt.com/

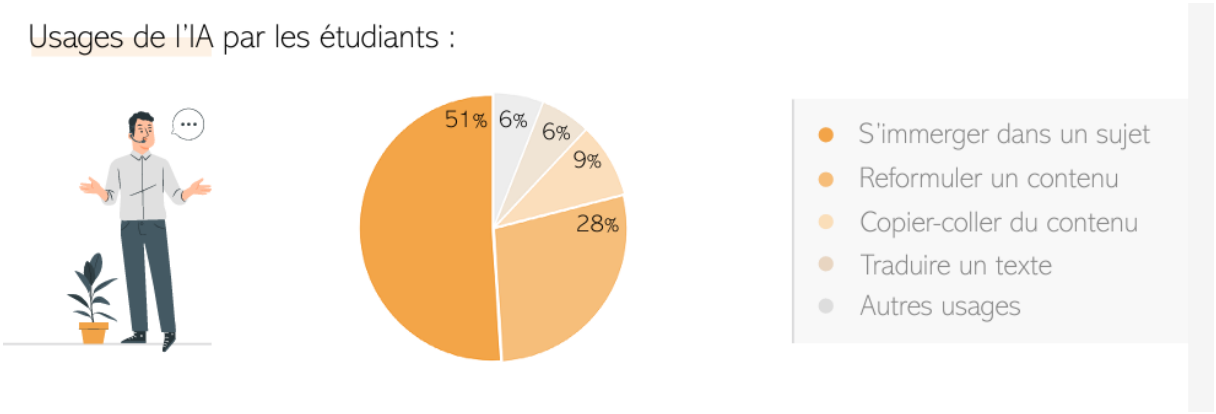

Figure 2 : enquête Compilatio (Compilatio, 2023)

Près de 2/3 des enseignants et des étudiants considèrent qu'il ne faut pas interdire l'usage des IA qui peuvent permettre d'introduire de nouvelles méthodes d'apprentissage, mais par contre qu'il faut encadrer ces usages.

Ce que nous avons, par exemple, commencer à faire cette année au sein du master MEDAS (mégadonnées et analyse sociale) du CNAM[13] avec l'équipe enseignante. Les mesures prises visent à intégrer ces technologies tout en préservant les compétences fondamentales à acquérir en terme de capacité d'analyse, de production et de recul critique.
Au niveau d'un cadrage général, les mesures suivantes ont été décidées :

- Informer les élèves des usages autorisés des IA génératives,
- Allonger la durée des soutenances de fin d'études afin que les élèves, au-delà du contenu du mémoire, produisent une analyse distanciée sur des aspects plus généraux de la formation,
- Citer l'usage de l'IA dans les travaux d'élèves incluant l'outil utilisé, les étapes du prompt et la date de saisie,
- Plus largement, concernant la prevue, les élèves sont responsables de l'archivage de leurs interactions avec l'IA générative et peuvent être amenés sur demande de l'équipe pédagogique à reproduire le déroulé des interactions.

Au niveau de chaque module d'enseignement :
Possibilité d'intégrer de manière transverse des compétences liées à l'IA générative dans les enseignements. Mais c'est à chaque enseignant d'apprécier l'intégration de ces technologies dans son enseignement.

Pour conclure sur les usages, les enjeux de transparence vont fortement conditionner les usages et la confiance en ces technologies, transparence sur les données d'apprentissage (Longpre, 2023), sur les biais potentiels, sur les types d'algorithmes utilisés pour réduire l'effet boite noire même s'il restera souvent très difficile d'apprécier cette dimension algorithmique. Les dispositifs en open source (par exemple Hugging-Face, https://huggingface.co/) sont aussi la

[13]https://formation.cnam.fr/rechercher-par-discipline/master-mega-donnees-et-analyse-sociale-medas--1085595.kjsp

garantie d'un contrôle, d'un accès aux données et aux modèles pour le plus grand nombre de développeurs, contrairement aux technologies fermées comme celles de Google et d'OpenAI.

La régulation des IA, IAG :

Au regard de la primauté des valeurs humaines et des dérives potentielles que peut engendrer ces technologies sur la vie privée, l'équité de traitement, la surveillance de masse, la discrimination, la désinformation… Des mesures sont muries à différents niveaux pour encadrer et réguler ces technologies, elles sont d'ordre juridique, financière, de contrôle, éthique et éducatif.

- Au niveau individuel, l'éducation des usagers reste le premier garde-fou,
- Les cadres réglementaires (règlement RGPD, IA Act…) et ses instances médiatrices (CNIL notamment en France) encadrent le développement du marché et de l'offre de services avec des sanctions financières associées,
- La responsabilisation des acteurs et les principes par défaut (*privacy by design; responsability by design*) renvoient à des valeurs éthiques et d'auto-régulation,
- Les processus d'audits et de certifications des services numériques, autant que cela soit possible, sont également des garanties qui devraient se développer. La certification de logiciels n'est pas une démarche nouvelle. Elle existe depuis près de 40 ans notamment pour l'aviation (norme DO 178C notamment pour les pilotes automatiques). Pour des secteurs comme le ferroviaire (norme EN 50128), le nucléaire (norme IEC 60880), l'automobile (ISO 26262) et plus généralement tous les systèmes industriels faisant appel à l'électronique et aux systèmes informatiques.

Pour la Commission européenne, le règlement IA[14] en cours d'adoption vise à prévenir les violations possibles de certains droits fondamentaux (droit à la dignité humaine, au respect de la vie privée et à la protection des données à caractère personnel, à la non-discrimination et à l'égalité entre les femmes et les hommes). Le système de prévention prévu par la Commission européenne repose sur le contrôle de la mise en place de dispositifs de conformité par les entreprises selon des niveaux de « risques IA » identifiés (inacceptable, élevé, moyen, faible). Le règlement IA est donc une réglementation de mise en conformité.

Pour conclure

L'objectif de cet article était d'éclairer les fondamentaux, les enjeux et quelques premiers usages significatifs des IAG dans le contexte de travail des professeurs documentalistes. Les institutions éducatives ne peuvent pas adopter une posture purement critique face à la déferlante des nouveaux services ouverts par ces technologies dans l'ensemble de la société.
Il s'agit donc de démystifier ces technologies, d'en décrypter les opportunités et les risques, de les encadrer. Pour ce faire, il convient de s'immerger dans ce nouvel environnement, de tester notamment les modèles spécialisés pour l'éducation, d'impliquer les élèves dans cet exercice et de développer avec eux à la fois des compétences nouvelles comme l'art du prompt mais

[14] https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A52021PC0206

aussi la réflexivité nécessaire pour apprécier la créativité mais aussi les limites de cette nouvelle dimension du numérique.

Bibliographie

Chartron, G., Raulin, A. (2022). L'intelligence artificielle dans le secteur de l'information et de la documentation : défis, impacts et perspectives. *I2D - Information, données & documents*, 1, 8-12. https://doi-org.proxybib-pp.cnam.fr/10.3917/i2d.221.0008

Colin de la Higuera, Jotsna Iyer (2023). *L'IA pour les enseignants, un manuel ouvert*, https://www.ai4t.eu/book/ia-pour-les-enseignants--un-manuel-ouvert-1/about-this-book?path=index

Compilatio, Sphinx (2023). *Résultats d'enquête : enseignants et étudiants confrontent leurs regards sur l'IA*, 7/11/23, https://www.compilatio.net/blog/communique-presse-enquete-ia-2023

LINC-CNIL (2023). [Dossier IA générative] - *ChatGPT : un beau parleur bien entraîné*, https://linc.cnil.fr/dossier-ia-generative-chatgpt-un-beau-parleur-bien-entraine

Longpre, S. (2023). *The Data Provenance Initiative: A Large Scale Audit of Dataset Licensing & Attribution in AI*, arXiv e-prints, doi:10.48550/arXiv.2310.16787

Unesco (2023). *Guidance for generative AI in education and research*, 44p, https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

Veille dans le domaine :

Actus IA, https://www.actuia.com/

Archimag, https://www.archimag.com/tags/intelligence-artificielle