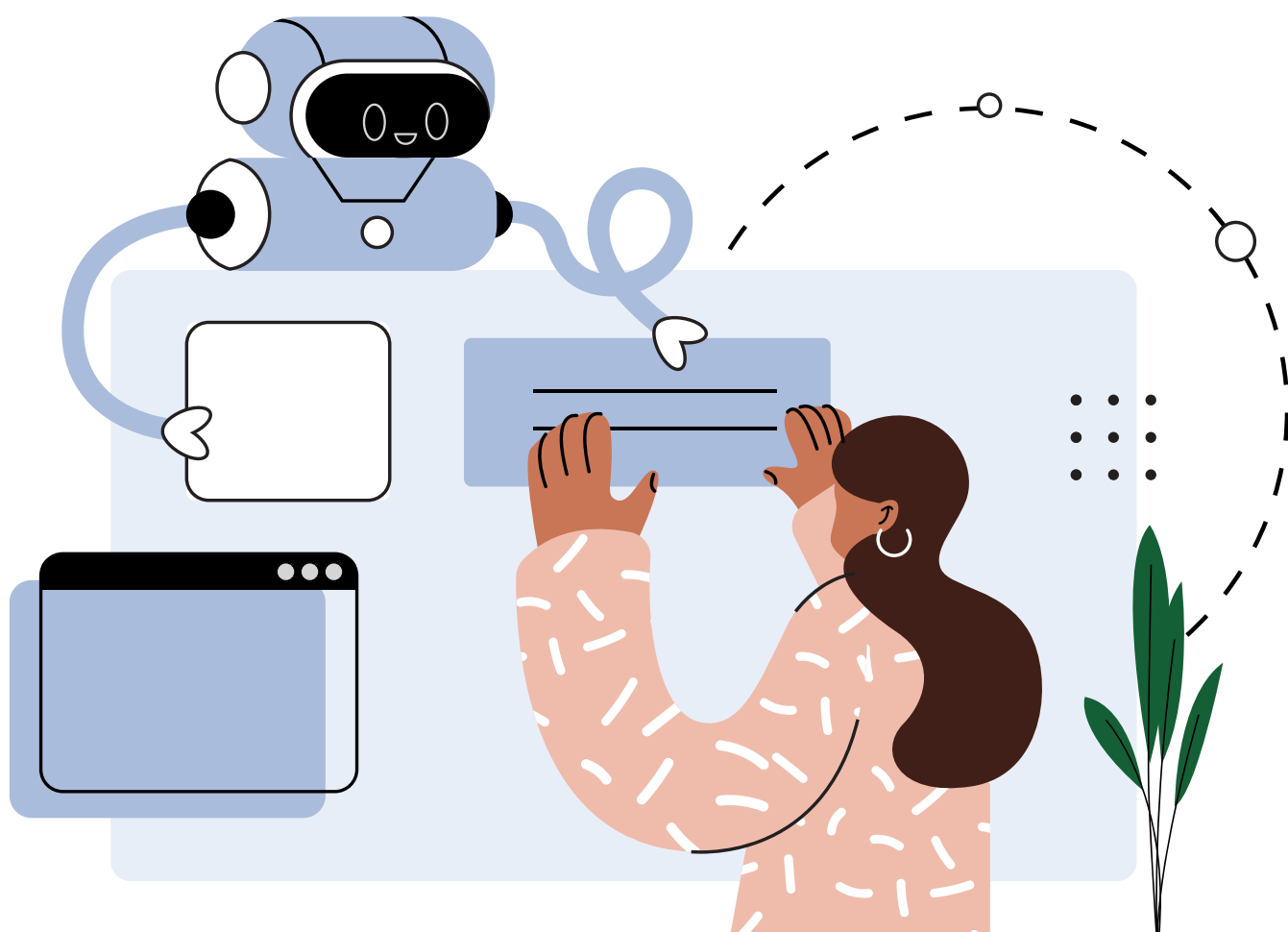


Orientations pour l'intelligence artificielle générative dans l'éducation et la recherche



L'UNESCO : chef de file pour l'éducation

L'éducation est la priorité absolue de l'UNESCO car c'est un droit humain fondamental qui constitue la pierre angulaire de la paix et du développement durable. L'UNESCO est l'agence des Nations Unies spécialisée pour l'éducation. Elle assure un rôle moteur aux niveaux mondial et régional pour renforcer le développement, la résilience et la capacité des systèmes nationaux d'éducation au service de tous les apprenants. L'UNESCO dirige également les efforts pour répondre aux défis mondiaux actuels par le biais de l'apprentissage transformateur, en mettant particulièrement l'accent dans toutes ses actions sur l'égalité des genres et l'Afrique.



L'agenda mondial Éducation 2030

En tant qu'institution des Nations Unies spécialisée pour l'éducation, l'UNESCO est chargée de diriger et de coordonner l'agenda Éducation 2030, qui fait partie d'un mouvement mondial visant à éradiquer la pauvreté, d'ici à 2030, à travers 17 Objectifs de développement durable. Essentielle pour atteindre chacun de ces objectifs, l'éducation est au cœur de l'Objectif 4 qui vise à « **assurer l'accès de tous à une éducation de qualité, sur un pied d'égalité, et promouvoir les possibilités d'apprentissage tout au long de la vie** ». Le Cadre d'action Éducation 2030 définit des orientations pour la mise en œuvre de cet objectif et de ces engagements ambitieux.



Publié en 2024 par l'Organisation des Nations Unies pour l'éducation, la science et la culture
7, place de Fontenoy, 75352 Paris 07 SP, France
© UNESCO 2024

ISBN 978-92-3-200312-6

<https://doi.org/10.54675/HBCX3851>



Œuvre publiée en libre accès sous la licence Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) (<http://creativecommons.org/licenses/by-sa/3.0/igo/>). Les utilisateurs du contenu de la présente publication acceptent les termes d'utilisation de l'Archive ouverte de libre accès UNESCO (<https://www.unesco.org/fr/open-access/cc-sa>).

Les images marquées d'un astérisque (*) ne sont pas couvertes par la licence CC-BY-SA et ne peuvent en aucune façon être reproduites sans l'autorisation expresse des détenteurs des droits de reproduction.

Titre original : *Guidance for generative AI in education and research*

Publié en 2023 par l'Organisation des Nations Unies pour l'éducation, la science et la culture

Les désignations employées dans cette publication et la présentation des données qui y figurent n'impliquent de la part de l'UNESCO aucune prise de position quant au statut juridique des pays, territoires, villes ou zones, ou de leurs autorités, ni quant au tracé de leurs frontières ou limites.

Les idées et les opinions exprimées dans cette publication sont celles des auteurs ; elles ne reflètent pas nécessairement les points de vue de l'UNESCO et n'engagent en aucune façon l'Organisation.

Traduction : Catherine Dhaussy

La traduction de la publication en français a fait l'objet d'une relecture technique par François Bocquet, jeune retraité, Consultant chez ASKA-SE (SaaS, IA et éducation et RGPD) et ancien chef de projet veille et prospective à la Direction du numérique pour l'éducation Ministère de l'Éducation.

Crédit de la couverture : Olexandra Simkina/Shutterstock.com*

Mise en page et impression dans les ateliers de l'UNESCO

Imprimé en France

R É S U M É

Vers une approche de l'utilisation de l'IA générative centrée sur l'homme

Les outils d'IA générative (*GenAI* en anglais) accessibles au public apparaissent ces derniers temps de manière très rapide, et la publication de versions itératives va plus vite que l'adaptation des cadres réglementaires nationaux. L'absence de réglementation nationale relative à l'IA générative dans la plupart des pays laisse la confidentialité des données des utilisateurs sans protection et les établissements d'enseignement passablement dépourvus face à ces outils.

Ces premières recommandations de l'UNESCO sur l'IA générative dans l'éducation, à l'échelle internationale, visent à aider les pays à mettre en œuvre des actions immédiates, à planifier des politiques à long terme et à développer les capacités humaines afin de garantir une vision de ces nouvelles technologies qui soit centrée sur l'humain.

Le présent document offre une évaluation des risques potentiels que la *GenAI* pourrait poser aux valeurs humanistes fondamentales qui promeuvent l'action humaine, l'inclusion, l'équité, l'égalité des sexes, les diversités linguistiques et culturelles, mais aussi la pluralité des opinions et des expressions.

Il propose aux agences gouvernementales des mesures-clefs pour réglementer l'utilisation des outils de l'IA générative, notamment en rendant obligatoire la protection de la confidentialité des données et en envisageant une limite d'âge pour leur utilisation. Il décrit les exigences imposées aux fournisseurs de *GenAI* pour permettre leur utilisation éthique et efficace dans l'éducation.

Les orientations proposées ici soulignent la nécessité pour les établissements d'enseignement de valider le déploiement des systèmes d'IA générative en fonction de leur pertinence pour l'éducation sur le plan éthique et pédagogique. Elles constituent une invitation à l'adresse de la communauté internationale pour réfléchir à leurs implications à long terme dans le domaine de la connaissance, de l'enseignement, de l'apprentissage et de l'évaluation.

La publication propose des recommandations concrètes aux décideurs politiques et aux établissements d'enseignement sur la manière dont les utilisations des outils de l'IA générative peuvent être conçues pour protéger les individus et bénéficier réellement aux élèves, aux étudiants, aux apprenants et aux chercheurs.

Alors que **ChatGPT** atteignait

100 millions

d'utilisateurs actifs mensuels en janvier 2023, en juillet de la même année seul **un pays** avait publié une réglementation sur **l'IA générative**.



unesco

« Les guerres prenant naissance dans l'esprit des hommes et des femmes, c'est dans l'esprit des hommes et des femmes que doivent être élevées les défenses de la paix. »

Orientations pour l'intelligence
artificielle générative dans
l'éducation et la recherche

Avant-propos



© UNESCO

L'intelligence artificielle générative (IA générative ou *GenAI*) a fait irruption sur la scène publique à la fin de 2022 avec le lancement de ChatGPT, qui est devenue l'application à la croissance la plus rapide de toute l'histoire. Ces applications de *GenAI* ont le pouvoir d'imiter les capacités humaines pour produire des résultats de toutes sortes, du texte, des images, des vidéos, de la musique ou des codes logiciels, ce qui a fait sensation. Des millions de personnes utilisent désormais l'IA générative dans leur vie quotidienne et le potentiel d'adaptation des modèles à des applications d'IA spécifiques à un domaine semble illimité.

Ces capacités étendues de traitement de l'information et de production de connaissances présentent un énorme potentiel d'implications pour l'éducation, car elles imitent la pensée complexe, qui constitue le fondement de l'apprentissage humain. Alors que les outils d'IA générative sont de plus en plus capables d'automatiser certains niveaux de base de l'écriture et de la création d'œuvres d'art, ils obligent les décideurs et les institutions du secteur de l'éducation à réexaminer pourquoi nous apprenons, ce que nous apprenons et comment nous apprenons. Il s'agit là de considérations essentielles pour l'éducation dans cette nouvelle phase de l'ère numérique.

Cette publication vise à aider la planification de réglementations, de politiques et de développement des capacités humaines nécessaires pour garantir que l'IA générative devienne un outil qui profite réellement aux enseignants, aux apprenants et aux chercheurs et participe à leur autonomisation.

Le document propose aux agences gouvernementales des étapes-clefs pour réglementer l'utilisation de l'IA générative. Il présente également des cadres et des exemples concrets pour la formulation de politiques et la conception d'enseignements qui permettent une utilisation éthique et efficace de cette technologie dans l'éducation. Il invite, enfin, la communauté internationale à examiner les implications profondes, à long terme, de l'IA générative sur la manière dont nous comprenons la connaissance et définissons le contenu, les méthodes et les résultats de l'apprentissage, ainsi que sur la manière dont nous évaluons et validons l'apprentissage.

Les présentes orientations s'appuient sur la *Recommandation sur l'éthique de l'intelligence artificielle* publiée par l'UNESCO en 2021 et sont ancrées dans une approche humaniste de l'éducation qui promeut l'action humaine, l'inclusion, l'équité, l'égalité des genres et la diversité culturelle et linguistique, ainsi que la pluralité des opinions et des expressions. Elles répondent également à l'appel du rapport 2021 de la Commission internationale sur l'avenir de l'éducation, *Repenser nos futurs ensemble : un nouveau contrat social pour l'éducation*, qui vise à redéfinir notre relation avec la technologie, comme partie intégrante de nos efforts pour renouveler le contrat social pour l'éducation.

L'IA ne doit pas se substituer à l'intelligence humaine. Elle nous invite plutôt à reconsidérer nos positions établies au sujet de la connaissance et de l'apprentissage humain. J'espère que ces orientations nous aideront à redéfinir de nouveaux horizons pour l'éducation et qu'elles éclaireront notre réflexion collective et nos actions de collaboration, de manière à conduire à des perspectives d'un apprentissage numérique pour tous résolument centré sur l'humain.

Stefania Giannini

Sous-Directrice générale pour l'éducation, UNESCO

Remerciements

Sous la direction de Stefania Giannini, sous-directrice pour l'éducation, et avec l'appui de Sobhi Tawil, directeur de l'équipe pour l'avenir de l'apprentissage et de l'innovation à l'UNESCO, la rédaction de la publication a été dirigée par Fengchun Miao, chef de l'unité pour la technologie et l'IA dans l'éducation.

Des remerciements particuliers reviennent à Wayne Holmes, professeur associé à l'University College London, qui a participé à la rédaction de certaines sections de la publication.

Cette publication est le fruit d'un effort collectif de responsables de l'éducation et d'experts dans le domaine de l'IA et de l'éducation.

Elle a bénéficié des réflexions et des contributions de nombreux experts, dont : Mutlu Cukurova, professeur à l'University College London ; Colin de la Higuera, titulaire la chaire de l'UNESCO Ressources éducatives libres et intelligence artificielle à Nantes Universités ; Shafika Isaacs, associée de recherche à l'université de Johannesburg ; Natalie Lao, directrice exécutive de l'App Inventor Foundation ; Qin Ni, professeur associé à Shanghai Normal University ; Catalina Nicolin, experte en TIC dans l'éducation au European Digital Education Hub en Roumanie ; John Shaw-Taylor, titulaire de la chaire UNESCO en IA et professeur de statistiques informatiques et d'apprentissage automatique à l'University College London ; Kelly Shirohira, directrice exécutive à Jet Education Services ; Ki-Sang Song, professeur à la Korea National University of Education ; et Ilkka Tuomi, scientifique en chef chez Meaning Processing Ltd, en Finlande.

De nombreux collègues de l'UNESCO ont également apporté leur contribution sous diverses formes, notamment : Dafna Feinholz, chef de la Section pour la bioéthique et l'éthique des sciences et des technologies ; Francesc Pedró, directeur de l'Institut international pour l'enseignement supérieur en Amérique latine et dans les Caraïbes ; Prateek Sibal, spécialiste de programme, Section pour les politiques numériques et la transformation numérique ; Saurabh Roy, chargé de projet principal à la Section pour le développement des enseignants, Division des politiques et des systèmes d'éducation et de formation tout au long de la vie ; Benjamin Vergel De Dios, spécialiste de programme en TIC dans l'éducation, Section pour l'innovation éducative et le développement des compétences au Bureau de Bangkok ; les collègues de l'entité Diversité des expressions culturelles du Secteur de la culture ; et Mark West, spécialiste de programme, Division de l'avenir de l'apprentissage et de l'innovation.

Nous remercions également Glen Hertelendy, Luisa Ferrara et Xianglei Zheng, de l'Unité pour la technologie et l'IA dans l'éducation, équipe pour l'avenir de l'apprentissage et de l'innovation, pour avoir coordonné la production de cette publication.

Nous remercions enfin Jenny Webster pour la révision et la relecture du texte, ainsi que Ngoc-Thuy Tran pour la conception de la mise en page.

Table des matières

Avant-propos	2
Remerciements	3
Acronymes et sigles	6
Introduction	7
1. Qu'est-ce que l'IA générative et comment fonctionne-t-elle ?	8
1.1 Qu'est-ce que l'IA générative ?	8
1.2 Comment l'IA générative fonctionne-t-elle ?	8
1.2.1 Comment les modèles d'IA générative de texte fonctionnent-ils ?	9
1.2.2 Comment les modèles d'IA générative d'image fonctionnent-ils ?	11
1.3 Développer l'ingénierie des invites pour générer des résultats pertinents	11
1.4 L'émergence de l'EdGPT et ses implications	13
2. Controverses au sujet de l'IA générative et de ses implications pour l'éducation	14
2.1 L'aggravation de la pauvreté numérique	14
2.2 Adaptation des réglementations nationales	14
2.3 Le consentement en matière d'utilisation des contenus	15
2.4 Non-traçabilité des modèles	15
2.5 Les contenus générés par l'IA polluent Internet	16
2.6 Manque de compréhension du monde réel	16
2.7 Réduire la diversité des opinions et marginaliser encore davantage les voix déjà marginalisées	17
2.8 Générer des truccages plus importants	17
3. Réglementer l'utilisation de l'IA générative dans l'éducation	18
3.1 Une approche de l'IA centrée sur l'humain	18
3.2 Les étapes de la régulation de l'IA générative dans l'éducation	18
3.3 Règlement sur l'IA générative : quelques éléments-clefs	20
3.3.1 Agences gouvernementales de régulation	20
3.3.2 Fournisseurs d'outils de <i>GenAI</i>	22
3.3.3 Usagers institutionnels	23
3.3.4 Les usagers à titre individuel	23
4. Vers un cadre politique pour l'utilisation de l'IA générative dans l'éducation et la recherche	24
4.1 Encourager l'inclusion, l'équité et la diversité linguistique et culturelle	24
4.2 Protéger l'humain	25
4.3 Surveiller et valider les systèmes d'IA générative pour l'éducation	25
4.4 Développer les compétences des apprenants en IA, y compris des compétences liées à l'IA générative	26

4.5 Renforcer la capacité des enseignants et des chercheurs à utiliser correctement l'IA générative	26
4.6 Promouvoir le pluralisme des opinions et l'expression plurielle des idées.....	27
4.7 Tester des modèles d'application pertinents au niveau local et constituer une base de données cumulative.....	27
4.8 Examiner les implications à long terme d'une manière intersectorielle et interdisciplinaire.....	28
5. Favoriser l'utilisation créative de l'IA générative dans l'éducation et la recherche	29
5.1 Stratégies institutionnelles pour faciliter l'utilisation responsable et créative de l'IA générative.....	29
5.2 Une approche d'interaction « centrée sur l'humain et pertinente sur le plan pédagogique ».....	30
5.3 Co-conception de l'utilisation de l'IA générative dans l'éducation et la recherche	30
5.3.1 L'IA générative au service de la recherche	30
5.3.2 L'IA générative comme facilitateur de l'enseignement	31
5.3.3 L'IA générative en tant que « coach particulier » pour l'acquisition à son rythme de compétences fondamentales.....	33
5.3.4 L'IA générative pour faciliter l'apprentissage par enquête ou par projet	34
5.3.5 L'IA générative pour soutenir les apprenants ayant des besoins particuliers	35
6. La place de l'IA générative dans l'avenir de l'éducation et de la recherche.....	38
6.1 Des questions éthiques inédites	38
6.2 Droits d'auteur et propriété intellectuelle	38
6.3 Sources de contenu et d'apprentissage.....	38
6.4 Réponses homogénéisées versus résultats diversifiés et créatif.....	39
6.5 Repenser l'évaluation et les résultats de l'apprentissage.....	39
6.6 Processus de réflexion.....	40
Remarques finales.....	40
Références	41
 Liste des tableaux	
Tableau 1. Techniques utilisées dans l'IA générative	8
Tableau 2. Les GPT d'OpenAI	9
Tableau 3. Co-conception des utilisations de la <i>GenAI</i> pour la recherche.....	31
Tableau 4. Co-conception des utilisations de la <i>GenAI</i> pour accompagner les enseignants et l'enseignement.....	32
Tableau 5. Co-conception de l'utilisation de la <i>GenAI</i> en tant que coach particulier pour l'acquisition à son rythme de compétences fondamentales dans les langues et les arts.....	33
Tableau 6. Co-conception des utilisations de la <i>GenAI</i> pour faciliter l'apprentissage par enquête ou par projet.....	34
Tableau 7. Co-conception des utilisations de la <i>GenAI</i> pour soutenir les apprenants ayant des besoins particuliers.....	36

Acronymes et sigles

Concepts et technologies

AEV	Auto-encodeurs variationnels (anglais : VAE)
AM	Apprentissage machine/Apprentissage automatique (anglais : ML)
API	Interface de programmation d'applications
GenAI	Intelligence artificielle générative – IA générative
Go	Gigaoctets (anglais : GB)
GPT	Transformateur génératif pré-entraîné
IA	Intelligence artificielle (anglais : AI)
IAD	Intelligence artificielle distribuée (anglais : DAI)
IAG	Intelligence artificielle générative (anglais : AGI)
ICT	Technologies de l'information et de la communication (anglais : ICT)
LaMDA	Modèle linguistique pour les applications de dialogue
LLM	Grand modèle linguistique (anglais : LLM)
RAG	Réseaux antagonistes génératifs (anglais : GAN)
RGPD	Règlement général sur la protection des données (anglais : GDPR)
RNA	Réseau de neurones artificiels/Réseau neuronal artificiel (anglais : ANN)

Instances et organismes

ACC	Administration du cyberspace de la Chine (anglais : CAC)
AGCC	Nuage gouvernemental d'IA (Singapour) (anglais : <i>AI Government Cloud Cluster</i>)
CNUCED	Conférence des Nations Unies sur le commerce et le développement
OCDE	Organisation de coopération et de développement économiques
UE	Union européenne
UNESCO	Organisation des Nations Unies pour l'éducation, la science et la culture

Introduction

Le lancement, fin 2022, de ChatGPT, premier outil d'intelligence artificielle générative (IA générative ou *GenAI*) facile à utiliser et largement accessible au grand public¹, suivi de versions de plus en plus sophistiquées, a provoqué une onde de choc dans le monde entier et alimente, entre les grandes entreprises technologiques, la course au positionnement dans le domaine du développement de modèles de *GenAI*.

Partout dans le monde, la préoccupation initiale, dans le champ de l'éducation, était que ChatGPT et des outils d'IA générative similaires seraient utilisés par les élèves et les étudiants pour tricher dans le cadre de leurs devoirs, menaçant ainsi la valeur de l'évaluation de l'apprentissage, de la certification et des qualifications (Anders, 2023). Alors que certains établissements d'enseignement ont interdit l'utilisation de ChatGPT, d'autres ont accueilli avec prudence l'arrivée des *GenAI* (Tilli, 2023). De nombreuses écoles et universités, par exemple, ont adopté une approche plus progressiste, estimant que « plutôt que de chercher à interdire leur utilisation des *GenAI*, les [élèves et] étudiants et le personnel doivent être soutenus dans l'utilisation efficace, éthique et transparente de ces outils » (Russell Group, 2023). Cette approche reconnaît que l'IA générative est largement disponible, qu'elle est susceptible de devenir de plus en plus sophistiquée et qu'elle possède un potentiel négatif spécifique et un potentiel positif exceptionnel dans le domaine de l'éducation.

De fait, il existe une multitude d'utilisations possibles à l'IA générative. Elle peut automatiser le traitement de l'information et la présentation des résultats dans toutes les représentations-clefs symboliques de la pensée humaine. Elle permet d'obtenir des résultats finaux à partir d'éléments partiels. Parce qu'elle libère les humains de certaines formes de réflexions assez basiques, cette nouvelle génération d'outils d'IA pourrait avoir de profondes implications sur la manière dont nous comprenons l'intelligence humaine et l'apprentissage.

Mais la *GenAI* soulève également de nombreuses préoccupations fondamentales liées à des questions telles que la sécurité, la confidentialité des données, les droits d'auteur et les manœuvres relevant de la manipulation. Certains de ces risques plus généraux liés à l'intelligence artificielle ont été exacerbés par l'IA générative, d'autres sont apparus avec cette dernière génération d'outils. Il est désormais urgent que chacune de ces questions et préoccupations soit pleinement comprise et prise en compte.

Cette publication est conçue pour répondre à ce besoin urgent. Un ensemble thématique de lignes directrices sur l'IA générative pour l'éducation ne devrait cependant pas être compris comme l'affirmation selon laquelle cette IA constitue la solution aux défis fondamentaux de l'éducation. Malgré le battage médiatique, il est peu probable que l'IA générative résolve à elle seule les problèmes auxquels sont confrontés les systèmes éducatifs de par le monde. Pour relever les défis existant de longue date dans le domaine éducatif, il est essentiel de garder présente à l'esprit l'idée selon laquelle ce sont la capacité humaine et l'action collective, et non la technologie, qui constituent les facteurs déterminants pour trouver des solutions efficaces aux défis fondamentaux auxquels sont confrontées les sociétés.

Cette publication vise donc à soutenir la juste élaboration de réglementations, de politiques et de programmes de développement des capacités humaines, afin de garantir que la *GenAI* devienne un outil qui profite véritablement aux enseignants, aux apprenants et aux chercheurs et leur donne les moyens d'agir. S'appuyant sur la *Recommandation sur l'éthique de l'intelligence artificielle* de l'UNESCO, les présentes orientations sont ancrées dans une approche centrée sur l'humain, qui promeut l'action humaine, l'inclusion, l'équité, l'égalité des genres et la diversité culturelle et linguistique, ainsi que la pluralité des opinions et des expressions.

Le guide examine d'abord ce qu'est l'IA générative et comment elle fonctionne, en présentant les diverses technologies et modèles disponibles (chapitre 1), avant d'identifier une série de questions éthiques et politiques controversées autour de l'IA en général et de l'IA générative en particulier (chapitre 2). Vient ensuite une discussion sur les étapes et les éléments-clefs à examiner lorsque l'on cherche à réglementer l'IA générative sur la base d'une approche centrée sur l'humain – une approche garantissant une utilisation éthique, sûre, équitable et qui a du sens (chapitre 3). Le chapitre 4 propose quant à lui des mesures qui peuvent être prises pour développer des cadres politiques cohérents et complets afin de réglementer l'utilisation de l'IA générative dans l'éducation et la recherche, tandis que le chapitre 5 examine les possibilités d'utiliser l'IA générative de manière créative dans la conception des programmes, l'enseignement, l'apprentissage et les activités de recherche. Le chapitre 6 conclut ce document d'orientation par des considérations sur les implications à long terme de l'IA générative pour l'éducation et la recherche.

1. Qu'est-ce que l'IA générative et comment fonctionne-t-elle ?

1.1 Qu'est-ce que l'IA générative ?

L'IA générative (*GenAI*) est une technologie d'intelligence artificielle (IA) qui génère automatiquement du contenu en réponse à des messages rédigés dans des interfaces de conversation en langage naturel. Plutôt que de simplement traiter des pages web existantes, en s'appuyant sur le contenu existant, l'IA générative produit un nouveau contenu. Le contenu peut apparaître dans des formats qui comprennent toutes les représentations symboliques de la pensée humaine : textes écrits en langage naturel, images (y compris photographies, peintures numériques et dessins animés), vidéos, musique et code logiciel. La *GenAI* est entraînée à l'aide de données collectées sur des pages web, des conversations sur les réseaux sociaux et d'autres médias en ligne. Elle génère son contenu en analysant, à l'aide de statistiques, la distribution des mots, des pixels ou d'autres éléments dans les données qu'elle a intégrées et en identifiant et répétant des modèles communs (par exemple, quels mots suivent, en règle générale, quels autres mots).

Si l'IA générative peut créer de nouveaux contenus, elle ne peut en revanche pas produire de nouvelles idées ou de nouvelles solutions aux défis du monde réel, car elle ne comprend pas les objets du monde réel ni les relations sociales qui sous-tendent le langage. En outre, malgré la fluidité et les résultats impressionnants de la *GenAI*, on ne peut pas lui faire confiance pour ce qui est de l'exactitude. Même les créateurs de ChatGPT reconnaissent que « si des outils comme ChatGPT peuvent souvent générer des réponses qui semblent raisonnables, on ne peut pas compter sur eux pour ce qui est de l'exactitude ». (OpenAI, 2023). Le plus souvent, les erreurs passent inaperçues si l'utilisateur n'a pas de solides connaissances sur le sujet traité.

1.2 Comment l'IA générative fonctionne-t-elle ?

Les technologies spécifiques qui sous-tendent l'IA générative font partie de la famille des technologies d'IA dites « apprentissage machine » (AM, *ML* pour *machine learning* en anglais), qui utilisent des algorithmes pour leur permettre d'améliorer continuellement et automatiquement leurs performances à partir de données. Le type d'apprentissage automatique qui a conduit à de nombreuses avancées dans le domaine de l'IA ces dernières années, par exemple l'utilisation de l'IA pour la reconnaissance faciale, est connu sous le nom de réseaux de neurones artificiels (RNA), qui s'inspirent du fonctionnement du cerveau humain et de ses connexions synaptiques entre les neurones. Il existe de nombreux types de RNA.

Les technologies d'IA qui génèrent du texte ou de l'image sont toutes deux basées sur un ensemble de technologies d'IA qui sont depuis des années à la disposition des chercheurs¹. ChatGPT, par exemple, utilise un transformateur génératif pré-entraîné (*GPT*, en anglais), tandis que l'IA générative d'image utilise généralement ce que l'on appelle des réseaux antagonistes génératifs (RAG, *GAN* en anglais) (voir **Tableau 1**)³.

Tableau 1. Techniques utilisées dans l'IA générative

Apprentissage machine (AM)		Ce type d'IA utilise des données pour améliorer automatiquement ses performances.
Réseaux antagonistes génératifs (RAG)		Ce type d'AM s'inspire de la structure et du fonctionnement du cerveau humain (par exemple, les connexions synaptiques entre les neurones).
IA générant du texte	Transformateurs à usage général	Ce type de RNA est capable de se concentrer sur différents ensembles de données pour déterminer comment ils sont liés les uns aux autres.
	Grands modèles linguistiques (GML, <i>LLM</i> en anglais)	Ce type de transformateur à usage général est entraîné à partir de grandes quantités de données textuelles.
	Transformateur génératif pré-entraîné (en anglais <i>Generative pre-trained transformer, GPT</i>) ⁴	Ce type de LLM est pré-entraîné sur des quantités de données encore plus importantes, ce qui permet au modèle de repérer les nuances de la langue et de générer un texte cohérent prenant en compte le contexte.
IA générant des images	Réseaux antagonistes génératifs (RAG)	Les réseaux neuronaux de ce type sont utilisés pour générer des images.
	Autoencodeurs variationnels (<i>VAE</i> en anglais)	

1.2.1 Comment les modèles d'IA générative de texte fonctionnent-ils ?

L'IA générant du texte utilise un type de RNA connu sous le nom de transformateur à usage général, et un type de transformateur à usage général appelé grand modèle de langage ou grand modèle linguistique. C'est pourquoi les systèmes d'IA générative de texte sont souvent appelés grands modèles de langage (GML, en anglais *LLM*). Le type de GML utilisé par la *GenAI* textuelle est connu sous le nom de transformateur génératif pré-entraîné ou, en anglais, GPT (d'où le « GPT » dans « ChatGPT »).

ChatGPT est basé sur GPT-3, développé par les équipes d'OpenAI. Il s'agit de la troisième itération de leur GPT, la première ayant été lancée en 2018 et la plus récente, GPT-4, en mars 2023 (voir **Tableau 2**). Chaque version de GPT développée par OpenAI a amélioré de manière itérative la précédente, grâce à des avancées dans les architectures d'IA, les méthodes d'entraînement et les techniques d'optimisation. Un aspect bien connu de ces progrès continus

est l'utilisation de quantités croissantes de données pour entraîner les « paramètres », dont le nombre augmente de façon exponentielle. Les paramètres peuvent être considérés, sur le plan métaphorique, comme des boutons qui peuvent être ajustés pour affiner les performances du GPT. Ils comprennent les « poids » du modèle, c'est-à-dire les paramètres numériques qui déterminent la manière dont le modèle traite les données d'entrée et produit les données de sortie.

Outre les progrès réalisés dans l'optimisation des architectures d'IA et des méthodes d'entraînement, cette itération rapide a également été rendue possible par les quantités massives de données⁵ et par les améliorations des capacités informatiques dont disposent les grandes entreprises. Depuis 2012, ces capacités informatiques utilisées pour l'entraînement des modèles d'IA générative doublent tous les trois ou quatre mois. À titre de comparaison, la loi de Moore avait une période de doublement de deux ans (OpenAI, 2018 ; Stanford University, 2019).

Tableau 2. Les GPT d'OpenAI

Version	Date de lancement	Quantité de données d'apprentissage	Nombre de paramètres	Caractéristiques
GPT-1	2018	40 Go	117 millions	Capable d'effectuer des tâches de traitement du langage naturel telles que compléter des textes et répondre à des questions.
GPT-2	2019	40 Go	1 500 millions	Capable d'effectuer des tâches plus complexes de traitement du langage naturel, telles que la traduction automatique et le résumé.
GPT-3	2020	17 000 Go	175 000 millions	Capable d'effectuer des tâches avancées de traitement du langage naturel, telles que la rédaction de paragraphes cohérents et la génération d'articles entiers. Capable également de s'adapter à de nouvelles tâches avec seulement quelques exemples.
GPT-4 ⁶	2023	1 000 000 Go (chiffre non confirmé)	170 000 000 millions (chiffre non confirmé)	Fiabilité accrue et capacité à traiter des instructions plus complexes.

Une fois que le GPT a été entraîné, la création d'une réponse textuelle à une instruction comporte les étapes suivantes :

1. L'invite (en anglais, *prompt*) est décomposée en unités plus petites (appelées jetons, *token* en anglais) qui sont introduites dans le GPT.
2. Le GPT utilise des modèles statistiques pour prédire les mots ou les phrases susceptibles de répondre de manière cohérente à l'invite.
 - le GPT identifie des modèles de mots et de phrases qui coïncident couramment dans son modèle de données volumineuses préconstruit (qui comprend des textes extraits d'Internet ou provenant d'autres sources) ;
 - à partir de ces modèles, le GPT estime la probabilité d'apparition de mots ou d'expressions spécifiques dans un contexte donné ;
 - à partir d'une prédiction aléatoire, le GPT utilise cette estimation de probabilités pour prédire le mot suivant ou la phrase suivante dans sa réponse.
3. Les mots ou les phrases qui ont été prédits sont convertis en texte lisible.
4. Le texte qu'on peut lire est filtré à travers ce que l'on appelle des « garde-fous », afin de supprimer tout contenu offensant.

5. Les étapes 2 à 4 sont répétées jusqu'à ce qu'une réponse soit terminée. La réponse est considérée comme finie lorsqu'elle atteint une limite maximale de jetons ou répond à des critères d'arrêt prédéfinis.
6. La réponse est traitée *a posteriori* pour améliorer la lisibilité : elle est mise en forme, la ponctuation ajoutée et d'autres améliorations apportées (par exemple, la réponse commence par des mots qu'un être humain pourrait utiliser, tels que « Bien sûr », « Sans doute » ou « Je suis désolé »).

Alors que les GPT et leur capacité à générer automatiquement du texte sont à la disposition des chercheurs depuis 2018, ce qui a rendu le lancement de ChatGPT si novateur, c'est son accès gratuit via une interface facile à utiliser, ce qui signifie que toute personne disposant d'un accès à Internet peut essayer l'outil et s'en servir. Le lancement de ChatGPT a provoqué une onde de choc dans le monde entier et a rapidement incité d'autres entreprises internationales à rattraper leur retard, ainsi que de nombreuses start-ups, soit par le lancement de leurs propres systèmes similaires, soit par la création de nouveaux outils à partir de ceux-ci.

En juillet 2023, on comptait parmi les options alternatives à ChatGPT :

- **Alpaca**⁷ : version améliorée du Llama de Meta, de l'université de Stanford, qui vise à lutter contre les fausses informations, les stéréotypes sociaux et le langage toxique des GML.
- **Bard**⁸ : LLM de Google, basé sur les systèmes LaMDA et PaLM 2, qui a accès à Internet en temps réel et peut donc fournir des informations actualisées.
- **Chatsonic**⁹ : réalisé par Writesonic, il s'appuie sur ChatGPT tout en extrayant des données directement de Google. Il a donc moins de chances de produire des réponses factuellement erronées.
- **Ernie** (également connu sous le nom de **Wenxin Yiyao** 文心一言)¹⁰ : GML bilingue de Baidu, toujours en cours de développement, qui intègre de larges connaissances à des ensembles de données massives pour générer du texte et des images.
- **Hugging Chat**¹¹ : par Hugging Face, qui a mis l'accent sur l'éthique et la transparence tout au long de son développement, de son entraînement et de son déploiement. En outre, toutes les données utilisées pour entraîner les modèles sont en libre accès.
- **Jasper**¹² : suite d'outils et d'interfaces de programmation d'applications qui, par exemple, peuvent être formés pour écrire dans le style

préférée d'un utilisateur. Il est également possible de générer des images.

- **Llama**¹³ : GML en source ouverte de Meta, qui requiert moins de puissance de calcul et moins de ressources pour tester de nouvelles approches, valider le travail des autres et explorer de nouveaux cas d'utilisation.
- **Open Assistant**¹⁴ : approche en *open-source* conçue pour permettre à toute personne disposant d'une expertise suffisante de développer son propre GML ; outil conçu à partir de données de formation collectées par des bénévoles.
- **Tongyi Qianwen** (通义千问)¹⁵ : GML d'Alibaba qui peut répondre à des questions en anglais ou en chinois ; en cours d'intégration dans la suite d'outils commerciaux d'Alibaba.
- **YouChat**¹⁶ : GML qui intègre des capacités de recherche en temps réel pour fournir davantage de contexte et des informations complémentaires, afin de générer des résultats plus précis et plus fiables.

La plupart de ces outils sont gratuits (dans certaines limites), d'autres sont à code source ouvert (*open source*). De nombreux autres produits basés sur l'un de ces modules d'apprentissage tout au long de la vie sont en cours de lancement. Quelques exemples :

- **ChatPDF**¹⁷ : rédige des résumés et répond aux questions sur les documents PDF qui lui sont fournis
- **Elicit - l'assistant de recherche en IA**¹⁸ : vise à automatiser une partie du travail des chercheurs, en identifiant les articles pertinents et en résumant les informations-clefs.
- **Perplexity**¹⁹ : fournit un « centre de connaissances » (*knowledge hub*) aux personnes qui recherchent des réponses rapides et précises, adaptées à leurs besoins.

Les outils basés sur le GML sont également intégrés dans d'autres produits, tels que les navigateurs web. Certaines extensions du navigateur Chrome sont basées sur ChatGPT :

- **WebChatGPT**²⁰ : permet à ChatGPT d'accéder à Internet afin de permettre des conversations plus précises et plus à jour
- **Compose AI**²¹ : complète automatiquement les phrases dans le courriel, par exemple
- **TeamSmart AI**²² : fournit « une équipe d'assistants virtuels »
- **Wiseone**²³ : simplifie les informations en ligne.

En outre, ChatGPT a été intégré à certains moteurs de recherche²⁴ et il est installé dans de vastes portefeuilles d'outils (par exemple Microsoft Word et Excel), ce qui le rend encore plus présent dans les bureaux des entreprises et les établissements d'enseignement du monde entier (Murphy Kelly, 2023).

Enfin, et cela constitue une transition intéressante vers l'IA générative d'image, l'outil le plus récent d'OpenAI, GPT-4, est capable d'accepter des images ainsi que du texte dans les instructions qu'il traite. En ce sens, il est multimodal. Sur cette base, certains affirment que l'appellation de « grand modèle linguistique » (GML) est de moins en moins appropriée. C'est la raison pour laquelle des chercheurs de l'université de Stanford ont suggéré le terme de « modèle de fondation » (Bommasani et al., 2021). Cette nouvelle proposition n'est cependant pas encore adoptée à grande échelle.

1.2.2 Comment les modèles d'IA générative d'image fonctionnent-ils ?

La *GenAI* pour l'image et la *GenAI* pour la musique utilisent généralement un type différent de réseaux de neurones artificiels, connu sous le nom de réseaux antagonistes génératifs (GAN) qui peuvent également être combinés avec des autoencodeurs variationnels. Les GAN comportent deux parties (deux « adversaires ») : le « générateur » et le « discriminateur ». Dans le cas des GAN d'images, le générateur crée une image aléatoire en réponse à une invite, et le discriminateur tente de faire la distinction entre cette image générée et des images réelles. Le générateur utilise ensuite le résultat du discriminateur pour réajuster ses paramètres, afin de créer une autre image. Le processus est répété, éventuellement des milliers de fois, le générateur produisant des images de plus en plus réalistes que le discriminateur est de moins en moins capable de distinguer des images réelles. Par exemple, un GAN efficace, entraîné à partir d'un ensemble de milliers de photographies de paysages, peut générer de nouvelles images de paysages, images de synthèse qu'il est presque impossible de distinguer de photographies réelles. De la même manière, un GAN entraîné à partir d'un ensemble de données de musique populaire (ou même de la musique d'un seul artiste) peut générer de nouveaux morceaux de musique qui suivent la structure et la complexité de la musique originale.

En juillet 2023, plusieurs modèles d'IA générative d'image sont disponibles. Ils génèrent tous des images à partir d'invites textuelles. La plupart sont libres d'utilisation, dans certaines limites :

- **Craiyon**²⁵ : précédemment connu sous le nom de DALL-E mini
- **DALL-E 2**²⁶ : l'outil de *GenAI* d'OpenAI
- **DreamStudio**²⁷ : outil de Stable Diffusion pour les images

- **Fotor**²⁸ : incorpore la *GenAI* dans une gamme d'outils d'édition d'images
- **Midjourney**²⁹ : outil indépendant de *GenAI* pour les images
- **NightCafe**³⁰ : interface pour la Stable Diffusion et DALL-E 2
- **Photosonic**³¹ : générateur d'art AI de WriteSonic

Quelques exemples faciles d'accès d'IA générative pour la vidéo :

- **Elai**³² : permet de convertir en vidéos des présentations, des sites web et du texte.
- **GliaCloud**³³ : peut générer des vidéos à partir de contenus d'actualité, de messages sur les médias sociaux, d'événements sportifs en direct et de données statistiques.
- **Pictory**³⁴ : permet de créer automatiquement des vidéos courtes à partir de longs contenus.
- **Runway**³⁵ : offre une gamme d'outils de création et d'édition de vidéos (et d'images).

Enfin, quelques exemples de *GenAI* musicales faciles d'accès :

- **Aiva**³⁶ : permet de créer automatiquement des bandes sonores personnalisées.
- **Boomy**³⁷, **Soundraw**³⁸, et **Voicemod**³⁹ : peuvent générer des chansons à partir de n'importe quel texte et ne nécessitent aucune connaissance en matière de composition musicale.

1.3 Développer l'ingénierie des invites pour générer des résultats pertinents

Si l'utilisation de l'IA générative peut être aussi simple que la saisie d'une question ou d'une autre invite, il faut reconnaître qu'en réalité, il n'est toujours pas évident pour l'utilisateur d'obtenir exactement le résultat qu'il souhaite. Ainsi, l'image de synthèse révolutionnaire de *Théâtre D'opéra Spatial*, qui a remporté un prix à la Colorado State Fair aux États-Unis, a nécessité des semaines de rédaction d'invites et d'ajustement de centaines d'images afin de générer le résultat final (Roose, 2022). De manière similaire, le défi que représente la rédaction d'invites efficaces pour l'IA générative de texte a conduit à un nombre croissant d'emplois d'ingénierie des invites apparaissant sur les sites de recrutement (Popli, 2023). « L'ingénierie des invites » fait référence aux processus et aux techniques de composition des entrées dans le but de produire des réponses de la part de l'IA générative qui ressemblent au mieux à l'intention souhaitée par l'utilisateur.

L'ingénierie de l'invite est plus efficace lorsque l'invite articule une chaîne de raisonnement cohérente centrée sur un problème particulier ou une chaîne de pensée dans un ordre logique. Il existe quelques recommandations spécifiques.

- Utiliser un langage **simple**, clair et direct, qui peut être facilement compris, en évitant les formulations complexes ou ambiguës.
- Inclure des **exemples** pour illustrer la réponse souhaitée ou le format des réponses générées.
- Mentionner le **contexte**, qui est essentiel pour générer des réponses pertinentes et significatives
- **Affiner** et itérer si nécessaire, en essayant différentes variations.
- Faire preuve d'**éthique**, en évitant les invites susceptibles de générer des contenus inappropriés, tendancieux ou nuisibles.

Il faut commencer par reconnaître que les résultats de l'IA générative ne sont pas fiables d'emblée, sans évaluation critique. Comme l'écrit OpenAI à propos de sa version de GPT la plus sophistiquée⁴⁰.

” En dépit de son potentiel, GPT-4 présente les mêmes limites que les modèles GPT précédents. La plus importante est qu'il n'est pas encore totalement fiable (il « hallucine » à propos de certains faits et commet des erreurs de raisonnement). Il convient de se montrer très prudent lors de l'utilisation des résultats d'un modèle linguistique, en particulier dans les contextes à fort enjeu, et de veiller à ce que le protocole exact (comme l'évaluation humaine, les éléments de contexte supplémentaires ou l'évitement total des utilisations à fort enjeu) corresponde aux besoins d'une utilisation spécifique.

Compte tenu de la qualité des résultats de la *GenAI*, des tests utilisateurs et des évaluations de performance rigoureuses sont nécessaires avant de valider les outils en vue d'une adoption à grande échelle ou à enjeux d'importance. Ces exercices doivent être conçus en intégrant une mesure de

performance qui soit la plus pertinente possible pour le type de tâche pour laquelle les utilisateurs demandent à l'IA générative de fournir des résultats. Par exemple, pour la résolution de problèmes mathématiques, la « précision » pourrait être utilisée comme mesure principale pour quantifier la fréquence à laquelle un outil de *GenAI* fournit la bonne réponse ; pour répondre à des questions délicates, la mesure principale pour mesurer la performance pourrait être le « taux de réponse » (la fréquence à laquelle la *GenAI* répond immédiatement à une question) ; pour le code, la mesure peut être « la part des codes générés qui sont directement exécutables » (c'est-à-dire si le code généré peut être directement exécuté dans un environnement de programmation et réussir les tests unitaires) ; et pour le raisonnement visuel, la mesure peut être la « correspondance exacte » (les objets visuels générés correspondent-ils exactement à la vérité de terrain ?) (Chen et al. , 2023).

En résumé, l'IA générative est facile à utiliser à un niveau superficiel ; les résultats plus sophistiqués nécessitent quant à eux une intervention humaine qualifiée et doivent être évalués de manière critique avant d'être utilisés.



Implications pour l'éducation et la recherche

Bien que l'IA générative puisse aider les enseignants et les chercheurs à générer des textes utiles ou d'autres résultats pour les aider dans leur travail, ce n'est pas nécessairement un processus simple. Il peut falloir plusieurs itérations d'une invite avant d'obtenir le résultat souhaité. On peut craindre que les jeunes apprenants, parce qu'ils sont par définition moins expérimentés que les enseignants, n'acceptent sans le savoir et sans réflexion critique des résultats obtenus par la *GenAI* qui sont superficiels, inexacts ou même nuisibles.

1.4 L'émergence de l'EdGPT et ses implications

Étant donné que les modèles d'IA générative peuvent servir de fondation, de point de départ, pour développer des modèles plus spécialisés ou spécifiques à un domaine, certains chercheurs ont suggéré que les GPT soient renommés « modèles de base » (Bommasani et al., 2021). Dans le domaine de l'éducation, les développeurs et les chercheurs ont commencé à affiner un modèle de base pour développer des « EdGPT »⁴¹. Les modèles d'EdGPT sont entraînés avec des données spécifiques à des fins éducatives. En d'autres termes, EdGPT vise à affiner le modèle qui a été dérivé de quantités massives de données d'entraînement générales, en utilisant de plus petites quantités de données éducatives de haute qualité et spécifiques à un domaine.

Cela fournit potentiellement à l'EdGPT une plus grande marge de manœuvre pour accompagner efficacement la réalisation des transformations citées dans la section 4.3. Par exemple, les modèles d'EdGPT ciblant la co-conception de programmes d'études peuvent permettre aux éducateurs et aux apprenants de créer du matériel éducatif utile, tel que des plans de cours, des quiz et des activités interactives qui suivent étroitement une approche pédagogique efficace, y compris pour des objectifs spécifiques de programmes d'études et avec niveaux de défi donnés pour des apprenants avec des besoins particuliers. De même, dans le contexte d'un tutorat personnalisé en compétences linguistiques, un modèle de base ajusté aux besoins, avec des textes appropriés pour une langue particulière, pourrait être utilisé pour générer des phrases, des paragraphes ou des exemples de conversations. Lorsque les apprenants interagissent avec le modèle, celui-ci peut leur répondre par des textes pertinents et grammaticalement corrects, au niveau qui leur convient. En théorie, les résultats des modèles EdGPT devraient aussi contenir moins de préjugés ordinaires ou de contenu répréhensible que les modèles GPT standard, mais le risque de générer des erreurs existe tout de même. Il est

donc essentiel de rappeler que, sauf évolution significative des modèles et de l'approche *GenAI*, l'EdGPT peut toujours générer des erreurs et se heurter à d'autres limites. Par conséquent, il reste indispensable que les principaux utilisateurs de l'EdGPT, en particulier les enseignants et les apprenants, adoptent une perspective critique à l'égard de chaque résultat.

À l'heure actuelle, les efforts déployés pour affiner les modèles de base en vue d'une utilisation plus ciblée des GPT dans l'éducation n'en sont qu'à leurs débuts. Parmi les projets existants, citons EduChat, un modèle de base développé par l'East China Normal University pour fournir des services d'enseignement aussi bien que d'apprentissage, et dont les codes, les données et les paramètres sont partagés en *open source*⁴². MathGPT constitue un autre exemple ; il est développé par le TAL Education Group, un GML qui se concentre sur la résolution de problèmes liés aux mathématiques et à l'organisation de conférences pour des utilisateurs du monde entier⁴³.

Toutefois, avant que des progrès significatifs puissent être envisagés, il est essentiel que des efforts soient déployés pour affiner les modèles de base, non seulement en injectant des connaissances relatives aux matières concernées et en éliminant les biais, mais aussi en ajoutant des connaissances sur les méthodes d'apprentissage les plus pertinentes et sur la manière dont on peut refléter cela dans la conception des algorithmes et des modèles. Le défi consiste à déterminer dans quelle mesure les modèles d'EdGPT peuvent aller au-delà de la connaissance des matières concernées, pour cibler également la pédagogie centrée sur l'élève et les interactions positives entre l'enseignant et l'élève. Un autre défi serait, aussi, de déterminer dans quelle mesure les données relatives à l'apprenant et à l'enseignant peuvent être collectées et utilisées de manière éthique afin d'enrichir un EdGPT. Enfin, il est également nécessaire de mener des recherches assez poussées pour s'assurer que l'EdGPT ne porte pas atteinte aux droits humains des apprenants de tous niveaux et ne déresponsabilise pas les enseignants.

2. Controverses au sujet de l'IA générative et de ses implications pour l'éducation

Après un premier chapitre sur la nature de l'IA générative et de son fonctionnement, cette deuxième partie examine les controverses et les risques éthiques soulevés par les systèmes de *GenAI* et étudie certaines de leurs implications dans le domaine de l'éducation.

2.1 L'aggravation de la pauvreté numérique

On le répète, l'IA générative s'appuie sur d'énormes quantités de données et sur une puissance de calcul considérable, en plus de ses innovations itératives dans les architectures d'IA et les méthodes d'apprentissage, qui ne sont pour la plupart accessibles qu'aux plus grandes entreprises technologiques internationales et à quelques économies nationales (principalement les États-Unis, la République populaire de Chine et, dans une moindre mesure, l'Europe). Cela signifie que la possibilité de créer et de contrôler l'IA générative est hors de portée de la plupart des entreprises et de la plupart des pays, en particulier ceux du Sud.

L'accès aux données devenant de plus en plus nécessaire pour le développement économique des États et pour les possibilités offertes aux populations, les pays et les individus qui n'ont pas accès à suffisamment de données ou qui n'ont pas les moyens de garantir cet accès se retrouvent dans une situation de « pauvreté en données » (Marwala, 2023). La situation est similaire pour l'accès à la puissance de calcul. La diffusion rapide de la *GenAI* dans les pays et les régions technologiquement avancées a accéléré de manière exponentielle la production et le traitement des données, en même temps qu'elle intensifiait dans le Nord global la concentration de la richesse liée à l'IA. En conséquence immédiate de quoi, les régions pauvres en données se sont trouvées encore plus exclues et risquent à long terme d'être colonisées par les normes sur lesquelles reposent les modèles de GPT. Les modèles ChatGPT actuels sont entraînés à partir de données provenant d'utilisateurs en ligne qui reflètent les valeurs et les normes du Nord. Cela ne correspond pas aux algorithmes d'IA pertinents à l'échelle locale dans les communautés pauvres en données, dans de nombreuses régions du Sud ou dans les communautés défavorisées du Nord.



Implications pour l'éducation et la recherche

Il est nécessaire que les chercheurs, les enseignants et les apprenants adoptent un point de vue critique sur les valeurs, les normes culturelles et les coutumes sociales intégrées dans les modèles de formation fondés sur la *GenAI*. Les décideurs politiques doivent être conscients de l'aggravation des inégalités causée par le fossé grandissant dans l'entraînement et le contrôle des modèles de *GenAI*, et prendre des mesures pour y remédier.

2.2 Adaptation des réglementations nationales

Les principaux développeurs de *GenAI* ont été critiqués parce qu'ils ne permettent pas à leurs systèmes de faire l'objet d'un examen académique indépendant et rigoureux (Dwivedi et al., 2023)⁴⁴. Les technologies fondamentales de l'IA d'une entreprise ont tendance à être protégées en tant que propriété intellectuelle de l'entreprise. Dans le même temps, de nombreuses entreprises qui commencent à utiliser l'IA générative trouvent qu'il est de plus en plus difficile de garantir la sécurité de leurs systèmes (Lin, 2023). En outre, malgré les appels à la réglementation lancés par l'industrie de l'IA elle-même⁴⁵, l'élaboration d'une législation sur la création et l'utilisation de toutes les IA, y compris les *GenAI*, accuse souvent un retard par rapport au rythme du développement, qui est très rapide. Cela explique en partie les difficultés rencontrées par les agences nationales ou locales pour comprendre les questions juridiques et éthiques et légiférer à ce sujet⁴⁶.

Alors que l'IA générative peut augmenter les capacités humaines pour accomplir certaines tâches, le contrôle des entreprises qui développent l'IA générative est limité. Cela soulève la question de la réglementation, en particulier en ce qui concerne l'accès et l'utilisation des données à l'échelle nationale, y compris les données concernant les institutions et les individus, ainsi que les données générées sur le territoire des pays. Il est nécessaire de posséder une législation adéquate : les agences gouvernementales locales pourront ainsi exercer un certain contrôle sur les différentes vagues de *GenAI*, afin d'assurer la gouvernance de ce bien public.



Implications pour l'éducation et la recherche

Les chercheurs, les enseignants et les apprenants doivent être conscients du manque de réglementations destinées d'une part à protéger de manière adéquate la propriété des institutions et des individus à l'échelle du pays, mais aussi les droits des utilisateurs nationaux de la *GenAI*, d'autre part à répondre aux questions législatives soulevées par ce type d'IA.

2.3 Le consentement en matière d'utilisation des contenus

On l'a déjà vu, les modèles d'IA générative sont construits à partir de grandes quantités de données (du texte, des sons, du code, des images...) souvent récupérées sur Internet et généralement sans l'autorisation du propriétaire. De nombreux systèmes d'IA générative par l'image et certains systèmes par le code ont par conséquent été accusés de violation de la propriété intellectuelle. À l'heure actuelle, plusieurs affaires juridiques sont en cours à ce sujet à l'échelle internationale.

Il a également été souligné que les GPT pourraient contrevenir à des lois telles que le règlement général sur la protection des données de l'Union européenne (RGPD, 2016), en particulier le droit à l'oubli, car il est actuellement impossible de supprimer les données d'une personne (ou les résultats de ces données) d'un modèle de GPT une fois que l'outil a été entraîné en utilisant ces données.



Implications pour l'éducation et la recherche

- Les chercheurs, les enseignants et les apprenants doivent connaître la législation en matière de propriété de données et vérifier que les outils d'IA générative qu'ils utilisent ne contreviennent pas à des réglementations existantes.
- Ces mêmes chercheurs, enseignants et apprenants doivent également être conscients que les images ou les codes créés avec *GenAI* peuvent violer les droits de propriété intellectuelle d'une personne, et que les images, les sons ou les codes qu'ils créent et partagent sur Internet peuvent être exploités par d'autres IA génératives.

2.4 Non-traçabilité des modèles

On a depuis longtemps admis que les réseaux de neurones artificiels (RNA) sont généralement des « boîtes noires », c'est-à-dire que leur fonctionnement interne n'est pas ouvert à des fins d'inspection. Par conséquent, les RNA ne sont pas « transparents » ou « explicables et traçables », et il n'est pas possible de savoir comment leurs résultats ont été atteints.

Si l'approche globale, y compris les algorithmes utilisés, peut généralement être expliquée, les modèles particuliers et leurs paramètres, y compris les poids du modèle, ne peuvent pas être vérifiés, ce qui explique pourquoi tel ou tel résultat généré ne peut être expliqué dans sa spécificité. Un modèle comme GPT-4 comporte des milliards de paramètres et poids (voir le **Tableau 2**) et ce sont les poids qui, pris dans leur ensemble, contiennent les structures apprises que le modèle utilise pour générer ses résultats. Comme les paramètres/poids ne sont pas transparents dans les RNA (**Tableau 1**), il est impossible d'expliquer de manière précise comment un résultat spécifique est créé par ces modèles.

Le manque de transparence et d'explicabilité de l'IA générative se révèle de plus en plus problématique à mesure que cette IA se complexifie (voir **Tableau 2**), produisant souvent des résultats inattendus ou indésirables. Qui plus est, les modèles de *GenAI* héritent des biais présents dans leurs données d'apprentissage qui, étant donné la nature non transparente des modèles, sont difficiles à détecter et à traiter ; ils répètent et perpétuent ensuite ces biais.

Enfin, cette opacité est également une cause majeure des problèmes de confiance qui se posent autour de l'IA générative (Nazaretsky et al., 2022a). Si les utilisateurs ne comprennent pas comment un système d'IA générative est parvenu à un résultat spécifique, ils sont moins susceptibles de vouloir l'utiliser et l'adopter (Nazaretsky et al., 2022b).



Implications pour l'éducation et la recherche

Les chercheurs, les enseignants et les apprenants doivent être conscients du fait que les systèmes de *GenAI* fonctionnent comme des boîtes noires et qu'il est par conséquent difficile, voire impossible, de savoir pourquoi ou comment un contenu particulier a été créé. L'absence d'explications sur la manière dont les résultats sont générés tend à enfermer les utilisateurs dans la logique définie par les paramètres conçus dans les systèmes de *GenAI*. Ces paramètres peuvent refléter des valeurs et des normes culturelles ou commerciales spécifiques qui biaisent implicitement le contenu produit.

2.5 Les contenus générés par l'IA polluent Internet

Les données à partir desquelles GPT est entraîné sont généralement tirées d'Internet, qui contient trop souvent un langage discriminatoire et inacceptable : les développeurs ont donc dû mettre en place ce qu'ils appellent des « garde-fous » pour éviter que les résultats GPT ne soient offensants et/ou contraires à l'éthique. Cependant, en raison de l'absence de réglementations strictes et de mécanismes de contrôle efficaces, le matériel tendancieux généré par l'IA se répand de plus en plus sur Internet, polluant l'une des principales sources de contenus ou de connaissances pour la plupart des apprenants à travers le monde. Cela est d'autant plus important que le matériau brut généré par *GenAI* peut sembler assez précis et convaincant, alors qu'il recèle souvent des erreurs et des idées biaisées. Cela représente un risque élevé pour les jeunes apprenants qui ne possèdent pas de connaissances préalables solides sur le sujet abordé. Cela pose également un risque récurrent pour les futurs modèles GPT qui seront entraînés à partir de textes extraits du Web que les modèles GPT ont eux-mêmes créés et qui contiennent également leurs préjugés et leurs erreurs.



Implications pour l'éducation et la recherche

- Les chercheurs, les enseignants et les apprenants doivent être conscients que les systèmes de *GenAI* sont capables de produire des contenus offensants et contraires à l'éthique.
- Ils doivent également connaître les problèmes qui risquent de se poser à long terme relativement à la fiabilité des connaissances lorsque les futurs modèles GPT seront basés sur des textes générés par des modèles GPT antérieurs.

2.6 Manque de compréhension du monde réel

Les textes générés par GPT sont parfois appelés de manière péjorative « perroquets stochastiques » car, comme mentionné précédemment, bien qu'ils puissent produire des écrits qui semblent convaincants, ces productions contiennent souvent des erreurs et peuvent inclure des déclarations préjudiciables (Bender et al., 2021). Ce phénomène se produit parce que les GPT se contentent de répéter des modèles de langage trouvés dans leurs

données d'apprentissage (généralement des textes tirés d'Internet), en commençant par des modèles aléatoires (ou « stochastiques ») et sans en comprendre le sens – tout comme un perroquet peut imiter des sons sans comprendre réellement ce qu'il dit.

Le décalage entre les modèles de *GenAI* qui *semblent* comprendre le texte qu'ils utilisent et génèrent et « la réalité », à savoir qu'ils ne comprennent pas la langue qu'ils utilisent et le monde réel, peut conduire les enseignants et les apprenants à accorder au résultat un niveau de confiance qui ne présente aucune garantie. Cela présente de sérieux risques dans le domaine de l'éducation. En effet, l'IA générative n'est pas fondée sur une observation du monde réel ou sur d'autres aspects fondamentaux de la méthode scientifique, pas plus qu'elle n'est alignée sur des valeurs humaines ou sociales. La *GenAI* est donc incapable de générer des contenus véritablement originaux sur le monde réel, les objets et les relations entre eux, les personnes et les relations sociales, les relations homme-objet ou les relations homme-technologie. La question de savoir si le contenu apparemment nouveau généré par les modèles de *GenAI* peut être reconnu, en tant que tel, comme relevant du savoir scientifique, constitue ainsi un sujet controversé.

Nous l'avons déjà indiqué, les GPT peuvent fréquemment produire des textes inexacts ou peu fiables. De fait, on sait bien que les GPT inventent des choses qui n'existent pas dans la vie réelle. Certains appellent cela une « hallucination », alors que d'autres critiquent l'utilisation d'un terme aussi anthropomorphique et par là même trompeur. Les entreprises qui développent de l'IA générative reconnaissent ce risque. En bas de page, l'interface publique de ChatGPT indique, par exemple : « ChatGPT peut produire des informations inexactes sur des personnes, des lieux ou des faits »².

Quelques-uns de ses défenseurs ont également suggéré que la *GenAI* représente une étape importante dans l'évolution vers l'intelligence générale artificielle (IGA, en anglais AGI), un terme suggérant une classe d'IA plus intelligente que les humains. Toutefois, cette idée est depuis longtemps critiquée, le principal argument étant que l'IA ne progressera jamais au point d'atteindre le stade de l'IGA, du moins tant qu'elle ne procédera pas à l'union, d'une manière ou d'une autre, en symbiose, de l'IA basée sur les connaissances (également connue sous le nom d'IA symbolique ou basée sur les règles) et de l'IA basée sur les données (ce qu'on appelle l'apprentissage automatique) (Marcus, 2022). Notons enfin que ces revendications relatives à l'IGA ou à la sensibilité nous détournent également d'un examen plus poussé de certains problèmes actuellement causés par l'IA, tels que la discrimination cachée à l'encontre de groupes déjà victimes de discrimination (Metz, 2021).



Implications pour l'éducation et la recherche

- Un texte créé par l'IA générative peut ressembler de manière troublante à un texte produit par un humain, comme si l'outil comprenait le texte qu'il a généré. Et pourtant, la *GenAI* ne comprend rien : au lieu de comprendre, ces outils assemblent des mots d'une manière couramment observée sur le Web. Le texte généré peut également être incorrect.
- Les chercheurs, les enseignants et les apprenants doivent être conscients non seulement du fait qu'un GPT ne comprend pas le texte qu'il génère, mais aussi qu'il peut générer, et génère souvent, des déclarations incorrectes. Les individus doivent donc adopter une approche critique à l'égard de tout ce que l'outil d'IA génère.

2.7 Réduire la diversité des opinions et marginaliser encore davantage les voix déjà marginalisées

ChatGPT et les outils similaires ont tendance à ne produire que des réponses répondant à des normes qui adoptent les valeurs des propriétaires/créateurs des données utilisées pour entraîner les modèles. De fait, si une séquence de mots apparaît fréquemment dans les données d'apprentissage – comme c'est le cas pour les sujets courants et non controversés ou pour les croyances dominantes – elle est susceptible d'être répétée par le GPT dans ses résultats.

Cela risque d'entraver et de compromettre le développement d'opinions nuancées et l'expression d'idées plurielles. Les populations pauvres en données, y compris les communautés marginalisées du Nord, ont une présence numérique minimale ou limitée en ligne. Leurs voix ne sont ainsi pas entendues et leurs préoccupations ne sont pas représentées dans les données utilisées pour former les GPT, et apparaissent donc rarement dans les contenus créés. Pour cet ensemble de raisons, compte tenu de la méthodologie d'entraînement basée sur des données provenant de pages web et de conversations sur les médias sociaux, les modèles de GPT risquent de marginaliser davantage encore des personnes déjà désavantagées.



Implications pour l'éducation et la recherche

- Alors que les développeurs et les fournisseurs de modèles de *GenAI* ont la responsabilité fondamentale de traiter de manière systématique les biais dans les ensembles de données et les résultats de ces modèles, du côté utilisateur les chercheurs, les enseignants et les apprenants doivent savoir que le résultat du texte de *GenAI* ne représente que la vision la plus répandue ou la représentation dominante du monde au moment où ses données d'entraînement ont été produites et que certaines d'entre elles sont problématiques ou biaisées (par exemple, la représentation stéréotypée du rôle respectif des hommes et des femmes).
- Les apprenants, les enseignants et les chercheurs ne devraient jamais prendre pour argent comptant les informations fournies par la *GenAI* et devraient toujours les analyser de manière critique.
- Ces chercheurs, enseignants ou apprenants doivent également être conscients de la façon dont les voix minoritaires peuvent être laissées de côté, parce que les voix minoritaires sont par définition moins courantes dans les données utilisées pour l'entraînement des modèles.

2.8 Générer des trucages plus importants

Outre les controverses communes à tous les types d'IA générative, les réseaux antagonistes génératifs peuvent être utilisés pour modifier ou manipuler des images ou des vidéos existantes afin d'en générer de fausses, qui s'avèrent difficiles à distinguer des vraies. La *GenAI* facilite de plus en plus la création de ces « hypertrucages » (le terme anglais de *deepfakes* est plus fréquemment utilisé) et de ce que l'on appelle les informations fallacieuses ou infox (*fake news* en anglais). Autrement dit, la *GenAI* permet à certains individus de commettre assez facilement des actes contraires à la morale ou à l'éthique, voire criminels, tels que la diffusion de fausses informations, la promotion de discours haineux et la diffusion de visages, à l'insu des personnes concernées ou sans leur consentement, dans des films entièrement truqués et parfois compromettants.



Implications pour l'éducation et la recherche

Si les fournisseurs d'IA générative ont l'obligation de protéger les droits d'auteur et les droits à l'image des utilisateurs, les chercheurs, les enseignants et les apprenants doivent quant à eux être conscients que toute image qu'ils partagent sur Internet peut être incorporée dans les données d'entraînement de *GenAI* et peut être manipulée et utilisée de manière contraire à l'éthique.

3. Réglementer l'utilisation de l'IA générative dans l'éducation

Pour répondre aux controverses autour de l'IA générative et exploiter les avantages potentiels de l'IA générative dans l'éducation, il faut commencer par la réglementer. La réglementation de l'IA générative à des fins éducatives implique un certain nombre d'étapes et de mesures politiques basées sur une approche centrée sur l'humain, afin de garantir son utilisation éthique, sûre, équitable et non dénuée de sens.

3.1 Une approche de l'IA centrée sur l'humain

La *Recommandation sur l'éthique de l'intelligence artificielle* adoptée par l'UNESCO en 2021 fournit le cadre normatif nécessaire pour commencer à aborder les multiples controverses autour de l'IA générative, y compris celles qui touchent à l'éducation et à la recherche. Elle se fonde sur une approche de l'IA centrée sur les individus, selon laquelle l'utilisation de l'IA devrait se trouver au service du développement des capacités humaines en vue d'un avenir inclusif, juste et durable. Cette approche doit être guidée par les principes des droits de l'Homme ainsi que par la nécessité de protéger la dignité humaine et la diversité culturelle, lesquelles définissent les biens communs de la connaissance. En termes de gouvernance, une approche centrée sur l'humain nécessite une réglementation adéquate qui puisse garantir son action en tant qu'individu, mais aussi la transparence et la responsabilité publique.

Le *Consensus de Beijing sur l'intelligence artificielle et l'éducation* (2019) précise ce qu'une approche centrée sur l'humain implique en matière d'utilisation de l'IA dans le contexte de l'éducation. Ce Consensus pose que l'utilisation des technologies de l'IA dans l'éducation devrait renforcer les capacités humaines pour un développement durable et une collaboration efficace entre l'humain et la machine dans la vie courante, l'apprentissage et le travail. Il appelle également à des actions complémentaires pour garantir un accès équitable à l'IA dans le but de soutenir les personnes marginalisées et de lutter contre les inégalités, tout en promouvant la diversité linguistique et culturelle. Le document suggère d'adopter des approches pangouvernementales, intersectorielles et multipartites dans l'élaboration des politiques sur l'IA dans l'éducation.

La publication *IA et éducation : guide pour les décideurs politiques* (UNESCO, 2022b) précise ce que signifie une approche centrée sur l'humain lorsqu'on examine les bénéfices et les risques de l'IA dans l'éducation et le rôle de l'éducation comme moyen de développer les compétences en matière d'IA. Elle propose des recommandations concrètes pour l'élaboration de politiques visant à piloter l'utilisation de l'IA afin de (i) permettre un accès inclusif aux programmes d'apprentissage, en particulier pour les populations vulnérables, par exemple les apprenants en situation de handicap ; (ii) soutenir les options d'apprentissage ouvertes et personnalisées ; (iii) améliorer le traitement des données pour élargir l'accès et améliorer la qualité de l'apprentissage ; (iv) observer les processus d'apprentissage et alerter les enseignants sur les risques d'échec ; (v) améliorer la compréhension et développer les compétences pour une utilisation éthique et signifiante de l'IA.

3.2 Les étapes de la régulation de l'IA générative dans l'éducation

Avant l'apparition de ChatGPT, les gouvernements avaient élaboré ou adapté des cadres destinés à réglementer la collecte et l'utilisation des données et l'adoption de systèmes d'IA dans l'ensemble des secteurs, y compris celui de l'éducation. Cela a fourni un contexte législatif et politique en vue de la réglementation des applications d'IA nouvellement émergentes. À la suite de la publication de plusieurs modèles compétitifs de *GenAI* à partir de novembre 2022, ces gouvernements ont apporté différentes réponses politiques, allant de l'interdiction de l'IA générative jusqu'à l'évaluation des besoins d'adaptation des cadres existants, en passant par la formulation en urgence de nouvelles réglementations.

Les stratégies gouvernementales visant à réglementer et à encourager l'utilisation créative de l'IA générative ont été cartographiées et examinées en avril 2023 (UNESCO, 2023b)⁴⁷. Cet examen propose une série de sept mesures que les agences d'État peuvent prendre pour réglementer l'IA générative et réaffirmer le contrôle public, avec pour objectif de tirer parti de son potentiel dans tous les secteurs, y compris celui de l'éducation.

Étape 1. Valider les réglementations générales internationales ou régionales en matière de protection des données ou en élaborer de nouvelles au niveau national

L'entraînement des modèles de d'IA générative a nécessité la collecte et le traitement de données en ligne provenant de citoyens de nombreux pays. L'utilisation sans consentement préalable de données et de contenus par ces modèles de *GenAI* pose un nouveau défi en matière de protection des données.

Les règlements généraux sur la protection des données, dont le RGPD entré en vigueur dans l'UE en 2018 est l'un des exemples précurseurs, fournissent le cadre juridique nécessaire pour réglementer la collecte et le traitement des données personnelles par les fournisseurs de *GenAI*. Selon le portail *Data Protection and Privacy Legislation Worldline* de la Conférence des Nations unies sur le commerce et le développement (CNUCED), 137 pays sur 194 ont mis en place une législation visant à assurer la protection des données et de la vie privée⁴⁸.

Dans quelle mesure ces cadres sont-ils mis en œuvre dans les pays, la réponse reste toutefois incertaine. Il est donc particulièrement important de veiller à ce qu'ils soient implémentés de manière correcte, notamment par un contrôle régulier du fonctionnement des systèmes de *GenAI*, que ce soit leur entraînement ou leur utilisation. Il est également urgent que les pays qui n'ont pas encore de lois de portée générale sur la protection des données s'en dotent.

Étape 2. Adopter/Adapter et financer des stratégies pangouvernementales en matière d'IA

La réglementation de l'IA générative doit faire partie intégrante de stratégies nationales plus globales en matière d'IA, stratégies destinées à garantir une utilisation sûre et équitable de l'IA dans tous les secteurs du développement, y compris celui de l'éducation. La formulation, la validation, le financement et la mise en œuvre des stratégies nationales en matière d'IA requièrent une approche pangouvernementale. Seule une telle approche peut garantir la coordination des actions intersectorielles nécessaires pour apporter des réponses cohérentes aux nouveaux défis.

Au début de l'année 2023, quelque 67 pays⁴⁹ avaient élaboré ou prévu d'élaborer des stratégies nationales sur l'IA, 61 d'entre elles prenant la forme d'une stratégie d'IA autonome et 7 étant des chapitres sur l'IA intégrés dans des stratégies nationales plus larges en matière de TIC ou de numérisation. Sans surprise, compte tenu de sa nouveauté, aucune de ces stratégies nationales n'avait encore abordé l'IA générative en tant que question spécifique au moment de la rédaction du présent rapport.

Il est essentiel que les pays revoient leurs stratégies nationales existantes en matière d'IA, ou en élaborent de nouvelles, en veillant à ce que des dispositions soient prises pour réglementer l'utilisation éthique de l'IA dans tous les secteurs, y compris celui de l'éducation.

Étape 3. Renforcer et mettre en œuvre des réglementations spécifiques sur l'éthique de l'IA

Pour aborder les dimensions éthiques posées par l'utilisation de l'IA, des réglementations spécifiques sont nécessaires.

L'étude réalisée par l'UNESCO en 2023 sur les stratégies nationales en matière d'IA indique que l'identification de ces questions éthiques et la formulation de principes directeurs ne sont communes qu'à une quarantaine de pays⁵⁰. Et même alors, les principes éthiques devront être traduits en lois et réglementations applicables ; or c'est rarement le cas. En effet, seule une vingtaine d'États ont défini des réglementations claires relatives à l'éthique de l'IA, notamment en ce qui concerne l'éducation, que ce soit dans le cadre des stratégies nationales en matière d'IA ou par un autre biais. Il est d'ailleurs intéressant de noter que si l'éducation est considérée comme un domaine d'action dans quelque 45 stratégies nationales en matière d'IA⁵¹, les références à l'éducation sont le plus souvent formulées en termes de développement des compétences et des talents en matière d'IA nécessaires pour soutenir la compétitivité nationale, bien plutôt qu'en termes de préoccupations éthiques.

En ce qui concerne les pays qui ne disposent pas encore de réglementations sur l'éthique de l'IA, ils doivent les formuler et les mettre en œuvre de toute urgence.

Étape 4. Adapter ou appliquer les lois existantes sur le droit d'auteur pour réglementer le contenu généré par l'IA

L'utilisation de plus en plus répandue de l'IA générative a fait apparaître de nouveaux défis pour le droit d'auteur. Ils concernent à la fois le contenu ou le travail protégé par le droit d'auteur sur lequel les modèles sont entraînés et le statut des connaissances « non humaines » qu'ils produisent.

À l'heure actuelle, seuls la Chine, les pays de l'UE et les États-Unis ont adapté leurs lois sur le droit d'auteur pour prendre en compte certaines implications de l'IA générative. Le Bureau américain du droit d'auteur, par exemple, a décidé que les résultats des systèmes d'IA générative tels que ChatGPT n'étaient pas protégeables en vertu de la loi américaine sur le droit d'auteur, avançant que « le droit d'auteur ne peut protéger que le matériel qui est le produit de la créativité humaine » (Bureau américain du droit d'auteur, 2023). Au même moment, dans l'UE, la proposition de loi sur l'IA exige des développeurs d'outils d'IA qu'ils mentionnent les documents protégés par le droit

d'auteur qu'ils ont utilisés pour construire leurs systèmes (Commission européenne, 2021). La Chine, par le biais de sa réglementation sur l'IA générative publiée en juillet 2023, exige que les résultats de la *GenAI* soient étiquetés en tant que contenu généré par l'IA et ne les reconnaît que comme des résultats de la synthèse numérique.

La réglementation de l'utilisation de matériel protégé par le droit d'auteur dans l'entraînement des modèles de *GenAI* et la définition du statut du droit d'auteur issu des résultats de la *GenAI* apparaissent comme de nouvelles responsabilités des lois sur le droit d'auteur. Il est indispensable d'adapter de manière urgente les lois existantes pour en tenir compte.

Étape 5. Élaborer des cadres réglementaires relatifs à l'IA générative

Le rythme rapide du développement des technologies de l'IA contraint les agences de gouvernance compétentes à accélérer le renouvellement de leurs réglementations. En juillet 2023, seul un pays, la Chine, avait publié des réglementations officielles spécifiques sur l'IA générative. Le Règlement provisoire sur le service d'IA générative publié le 13 juillet 2023 (Administration du cyberspace de la Chine, 2023a) exige des fournisseurs de systèmes d'IA générative qu'ils identifient et marquent le contenu, les images et les vidéos générés par l'IA de manière adéquate et dans un cadre légal, conformément au règlement en vigueur sur la synthèse profonde pour les services d'information en ligne (*Regulation on Deep Synthesis in the Framework of Online Information Services*). Il est nécessaire que d'autres cadres nationaux spécifiques à la *GenAI* soient élaborés sur la base d'une évaluation des lacunes existant dans les réglementations et les lois locales en vigueur.

Étape 6. Renforcer les capacités en vue d'une bonne utilisation de la *GenAI* dans l'éducation et la recherche

Les écoles, ainsi que les autres établissements d'enseignement, doivent développer leurs capacités à comprendre les bénéfices et les risques potentiels de l'IA, y compris de l'IA générative, dans le domaine de l'éducation. Ce n'est que sur la base d'une telle compréhension que ces établissements pourront valider l'utilisation d'outils d'IA. Les enseignants et les chercheurs doivent en outre être soutenus pour renforcer leurs capacités à utiliser correctement l'IA générative, notamment par le biais d'une formation et d'un accompagnement continu. Un certain nombre de pays ont lancé de tels programmes de renforcement des capacités, notamment Singapour, qui offre une plateforme dédiée au développement des capacités d'IA des établissements d'enseignement par le biais de son *AI Government Cloud Cluster*, qui comprend un référentiel dédié de modèles de GPT (Ocampo, 2023).

Étape 7. Réfléchir aux implications à long terme de la *GenAI* dans le domaine de l'éducation et de la recherche

L'impact des versions actuelles de l'IA générative commence à peine à se manifester, et leurs effets sur l'éducation doivent encore être pleinement explorés et compris. Entre-temps, des versions plus puissantes de l'IA générative et d'autres classes d'IA continuent d'être développées et déployées. Des questions cruciales continuent cependant de se poser, quant aux implications de la *GenAI* pour la création, la transmission et la validation des connaissances – pour l'enseignement et l'apprentissage, pour la conception et l'évaluation des programmes d'études, ainsi que pour la recherche et les droits d'auteur. La plupart des pays commencent à peine à intégrer la *GenAI* dans l'éducation, alors même que les conséquences à long terme restent à étudier. Pour garantir que l'IA voie son utilisation centrée sur l'être humain, il est urgent de mener un débat public ouvert et des dialogues politiques sur les implications à long terme. Un débat global, impliquant le gouvernement, le secteur privé et d'autres partenaires concernés, devrait servir à fournir les informations et les contributions nécessaires pour le renouvellement continu des réglementations et des politiques.

3.3 Règlement sur l'IA générative : quelques éléments-clefs

Ce sont tous les pays qui doivent légiférer sur l'IA générative en vue de s'assurer qu'elle soit bénéfique au développement dans le domaine de l'éducation et dans d'autres contextes. Cette section propose différentes actions, liées à des éléments-clefs et qui peuvent être menées par : (1) les organismes gouvernementaux de réglementation, (2) les développeurs d'outils basés sur l'IA, (3) les utilisateurs institutionnels et (4) les utilisateurs individuels. Bien que de nombreux éléments dépassent le cadre des États, ils doivent tous être considérés à la lumière du contexte local, c'est-à-dire des systèmes éducatifs et des cadres réglementaires généraux déjà en place dans le pays concerné.

3.3.1 Agences gouvernementales de régulation

Une approche pangouvernementale est nécessaire pour coordonner la conception, l'harmonisation et la mise en œuvre des réglementations sur l'IA générative. Sept éléments et actions-clefs constituent des préconisations.

- **Assurer une coordination intersectorielle.** Mettre en place un organisme national chargé de diriger l'approche pangouvernementale de la *GenAI* et de coordonner la coopération entre les différents secteurs.

- **Harmoniser la législation.** Aligner le cadre sur les contextes législatifs et réglementaires pertinents de chaque pays – avec, par exemple, les lois d'ensemble sur la protection des données, les réglementations sur la sécurité d'Internet, les lois sur la sécurité des données recueillies auprès des citoyens ou utilisées à leur bénéfice, ou d'autres législations et pratiques habituelles pertinentes. Évaluer l'adéquation des réglementations existantes et toute adaptation nécessaire en réponse aux nouvelles questions posées par la *GenAI*.
- **Trouver un équilibre entre la réglementation de la *GenAI* et la promotion de l'innovation en matière d'IA.** Encourager la coopération intersectorielle entre les entreprises, les différents organismes, les établissements d'enseignement et de recherche, ainsi que les agences publiques concernées, afin de développer conjointement des modèles dignes de confiance ; encourager la création d'écosystèmes à source ouverte (*open-source*) dans le but de promouvoir le partage des ressources de supercalculateurs et des ensembles de données de pré-entraînement de qualité ; favoriser, aussi, l'application pratique de la *GenAI* dans tous les secteurs et la création de contenus de qualité pour l'intérêt général.
- **Évaluer et classer les risques potentiels de l'IA.** Instaurer des principes et mettre en place un processus pour l'évaluation et la catégorisation de l'efficacité, de la sûreté et de la sécurité des services de l'IA générative, avant leur déploiement et tout au long du cycle de vie du système. Envisager des mécanismes de catégorisation basés sur les niveaux de risque que la *GenAI* peut impliquer à l'échelle des citoyens. Les classer selon une réglementation stricte (c'est-à-dire en interdisant les applications ou systèmes basés sur l'IA qui présentent des risques inacceptables), avec une réglementation spéciale pour les applications à haut risque et une réglementation générale s'appliquant aux applications qui ne sont pas répertoriées comme étant à haut risque. Le projet de loi sur l'IA rédigé par l'UE représente un bon exemple de cette approche.
- **Protéger la confidentialité des données.** Tenir compte du fait que l'utilisation de la *GenAI* implique presque toujours que les utilisateurs partagent leurs données avec le fournisseur de la *GenAI* ; missionner les instances en charge de la rédaction et de la mise en œuvre des lois pour la protection des données à caractère personnel des utilisateurs ; identifier et combattre, lorsqu'ils ne sont pas légaux, le stockage, le profilage et le partage des données.

- **Définir et appliquer une limite d'âge pour l'utilisation de la *GenAI*.** La plupart des applications d'IA générative sont d'abord conçues pour des utilisateurs adultes. Ces applications comportent souvent des risques non négligeables pour les mineurs, notamment l'exposition à des contenus inappropriés et le risque de manipulation. À la lumière de tous ces risques et compte tenu de l'incertitude considérable qui continue d'entourer les applications itératives de l'IA générative, il est fortement recommandé d'imposer des limites d'âge pour les technologies d'IA à usage général afin de protéger les droits et le bien-être des enfants.

À l'heure actuelle, les conditions d'utilisation de ChatGPT exigent que les utilisateurs soient âgés d'au moins 13 ans et que les utilisateurs de moins de 18 ans aient l'autorisation de leurs parents ou de leur tuteur légal pour utiliser les services⁵².

Ces restrictions et autres seuils d'âge découlent de la loi sur la protection de la vie privée en ligne des enfants (*Children's Online Privacy Protection Act*) en vigueur aux États-Unis (Federal Trade Commission, 1998). Adoptée en 1998, avant la généralisation des médias sociaux et bien avant la création d'applications *GenAI* puissantes et faciles à utiliser telles que ChatGPT, la loi américaine précise que les organisations ou les fournisseurs individuels de médias sociaux ne sont pas autorisés à procurer des services aux enfants de moins de 13 ans sans l'autorisation de leurs parents. De nombreux commentateurs estiment que ce seuil est trop bas et plaident en faveur d'une législation qui porterait l'âge à 16 ans. Le RGPD de l'Union européenne (2016) précise quant à lui que les utilisateurs doivent être âgés d'au moins 16 ans pour utiliser les services des médias sociaux sans autorisation parentale.

L'émergence de divers chatbots de la *GenAI* rend indispensable, dans chaque pays, un examen attentif – et un débat public – pour déterminer la juste limite d'âge pour les échanges conduits de manière autonome avec les plateformes de la *GenAI*. Le seuil minimum devrait être fixé à 13 ans. Les États devront également décider si l'autodéclaration de l'âge reste un moyen de vérification approprié. Ils devront définir, enfin, les responsabilités des fournisseurs d'IA générative en ce qui concerne le contrôle de l'âge et les responsabilités des parents ou des tuteurs en ce qui concerne la surveillance des conversations menées en autonomie par les mineurs.

- **Assurer la propriété nationale des données et combattre le risque de pauvreté en données.** Prendre des mesures législatives pour protéger à l'échelle nationale la propriété des données et réglementer les fournisseurs de *GenAI* qui opèrent à l'intérieur de ses frontières. Pour les ensembles de données générés par les citoyens et qui sont utilisés à des fins commerciales, établir des règlements pour favoriser une coopération mutuellement bénéfique afin que cette catégorie de données ne sorte pas du pays en vue d'être exploitée exclusivement par les grandes entreprises technologiques.

3.3.2 Fournisseurs d'outils de *GenAI*

Parmi les fournisseurs d'IA générative, on trouve des organisations et des individus qui sont responsables du développement et de la mise à disposition des outils de *GenAI*, et/ou qui utilisent les technologies de *GenAI* pour fournir des services, notamment par le biais d'interfaces de programmation d'applications (IPA, *API* en anglais) programmables. La plupart des fournisseurs influents d'outils de l'IA générative sont des entreprises disposant de capitaux extrêmement importants. Il convient de préciser aux fournisseurs d'outils de *GenAI* qu'ils sont responsables de l'éthique dès le stade de la conception ; ils sont également responsables de la mise en œuvre des principes éthiques stipulés dans les règlements. Dix catégories de responsabilités sont concernées.

- **Responsabilités humaines.** Les fournisseurs de *GenAI* devraient être tenus responsables de l'adhésion aux valeurs fondamentales et aux finalités légitimes concernant le respect de la propriété intellectuelle et des pratiques éthiques, tout en empêchant la propagation de la désinformation et des discours haineux.
- **Des données et des modèles fiables.** Les fournisseurs de *GenAI* devraient être tenus de prouver la fiabilité et le caractère éthique des sources de données et des méthodes utilisées par leurs modèles et leurs résultats. Ils doivent adopter des données et des modèles de fondation aux sources légales reconnues et respecter les lois applicables en matière de propriété intellectuelle (par exemple dans le cas de données protégées par des droits de propriété intellectuelle). En outre, lorsque des données à caractère personnel sont requises par les modèles, la collecte de ces informations ne doit s'effectuer qu'avec le consentement éclairé et explicite des personnes concernées.
- **Production de contenu non discriminatoire.** Les fournisseurs d'IA générative doivent interdire la conception et le déploiement de systèmes de *GenAI* qui génèrent un contenu tendancieux ou discriminatoire fondé sur la race, la nationalité, le sexe ou d'autres données protégées. Ils doivent veiller à ce que des « garde-fous » solides soient en place pour empêcher la *GenAI* de produire des contenus offensants, biaisés ou faux, tout en s'assurant que les individus impliqués sont protégés et ne sont pas exploités.
- **Explicabilité et transparence des modèles de *GenAI*.** Les fournisseurs devraient soumettre aux agences d'État leurs explications sur : les sources, l'échelle et les types de données utilisées pour entraîner les modèles ; les règles d'étiquetage des données utilisées lors de la pré-formation ; les méthodes ou algorithmes que leurs modèles utilisent pour générer du contenu ou des réponses ; les services que leurs outils de *GenAI* fournissent. Si nécessaire, ils devraient aider les agences de l'État à comprendre la technologie et les données. La prédisposition de la *GenAI* à générer des contenus porteurs d'erreurs et de réponses contestables devrait également être rendue transparente pour les utilisateurs.
- **Marquage du contenu produit par de l'IA générative.** Conformément aux lois et réglementations applicables à la synthèse d'informations en ligne assistée par l'IA, les fournisseurs doivent marquer, ou « annoter », correctement et légalement les documents, rapports, images et vidéos générés par l'IA générative. Par exemple, des résultats issus de cette IA doivent être clairement identifiables comme ayant été produits par un algorithme génératif.
- **Principes de sécurité et de sûreté.** Il appartient aux fournisseurs d'IA générative d'assurer un service sûr, robuste et durable tout au long du cycle de vie d'un système de *GenAI*.
- **Spécifications sur l'adéquation de l'accès et de l'utilisation.** Les fournisseurs d'IA générative devraient fournir des spécifications claires sur le public concerné, les scénarios d'utilisation et les objectifs de leurs services, et aider les utilisateurs des outils de *GenAI* à prendre des décisions rationnelles et responsables.
- **Reconnaître les limites et éviter les risques prévisibles.** Les fournisseurs de *GenAI* devraient clairement annoncer les limites des méthodes utilisées par les systèmes et leurs résultats. Ils

doivent mettre au point des technologies pour garantir que les données d'entrée, les méthodes et les résultats n'entraînent pas de dommages prévisibles auprès des utilisateurs, ainsi que des protocoles pour atténuer les dommages non prévisibles lorsque ceux-ci se produisent. Ils doivent également fournir des conseils pour aider les utilisateurs à comprendre le contenu généré par l'IA sur la base de principes éthiques, mais aussi pour éviter qu'ils ne fassent démesurément confiance au contenu généré et qu'ils n'en deviennent dépendants.

- **Mécanismes de réclamations et de recours.** Les fournisseurs d'IA générative doivent établir des mécanismes et des canaux pour recueillir les réclamations des utilisateurs et du grand public, et prendre des mesures adaptées pour accepter et traiter ces réclamations.
- **Contrôle et signalement de l'utilisation illicite.** Les fournisseurs coopéreront avec les agences d'État pour faciliter la surveillance et le signalement des utilisations illicites. Cela inclut les personnes qui utilisent les contenus de *GenAI* de manière illégale ou en violation des valeurs éthiques ou sociales – promotion de la désinformation ou du discours haineux, production de spam ou création de logiciels malveillants.

3.3.3 Usagers institutionnels

Les utilisateurs – ou usagers – institutionnels comprennent les autorités et les institutions relevant du domaine de l'éducation, telles que les universités et les écoles qui ont la responsabilité de déterminer si l'IA générative doit être utilisée et quels types d'outils *GenAI* doivent être achetés et déployés au sein de l'institution.

- **Audit institutionnel des algorithmes, des données et des résultats de la *GenAI*.** Instaurer des mécanismes pour contrôler le mieux possible les algorithmes et les données utilisés par les outils de *GenAI* et les résultats qu'ils génèrent. Cela devrait inclure des évaluations et des audits réguliers, la protection des données des utilisateurs et le filtrage automatique des contenus inappropriés.
- **Valider la proportionnalité et protéger le bien-être des utilisateurs.** Mettre en œuvre des mécanismes de classification nationaux ou élaborer une politique institutionnelle pour classer et valider les systèmes et les applications de *GenAI*. Veiller à ce que les systèmes d'IA générative adoptés par l'institution soient conformes aux cadres éthiques validés à l'échelle locale et ne causent

aucun préjudice prévisible aux utilisateurs-cibles de l'institution, en particulier aux mineurs et aux groupes vulnérables.

- **Étudier et évaluer les impacts à long terme.** Au fil du temps, l'utilisation d'outils ou de contenus de *GenAI* dans l'éducation peut avoir des effets marqués sur le développement de capacités humaines telles que la pensée critique et la créativité. Il convient d'évaluer ces effets potentiels et d'en tenir compte.
- **Adéquation avec l'âge.** Envisager de faire respecter un âge minimum pour l'utilisation autonome de l'IA générative dans l'établissement.

3.3.4 Les usagers à titre individuel

Les utilisateurs à titre individuel comprennent potentiellement toutes les personnes dans le monde qui ont accès à Internet et à au moins un type d'outil de *GenAI*. Le terme « usagers à titre individuel », tel qu'il est employé ici, se réfère principalement aux enseignants, chercheurs et apprenants relevant de l'apprentissage formel ou à ceux qui participent à des programmes d'études non formels.

- **Sensibilisation aux termes de référence sur l'utilisation de la *GenAI*.** Lorsqu'ils signent ou expriment leur consentement à des accords de service, les utilisateurs devraient être conscients de l'obligation qu'ils ont de respecter les termes de référence stipulés dans l'accord, ainsi que les lois ou réglementations qui sous-tendent l'accord.
- **Utilisation éthique des applications de la *GenAI*.** Les utilisateurs doivent se servir de l'IA générative de manière responsable et éviter de le faire d'une manière qui pourrait nuire à la réputation et aux droits légitimes d'autres personnes.
- **Contrôler et signaler les applications de *GenAI* illégales.** Lorsqu'ils découvrent des applications d'IA générative qui enfreignent une ou plusieurs réglementations, les utilisateurs sont tenus d'en informer les instances gouvernementales de réglementation.

4. Vers un cadre politique pour l'utilisation de l'IA générative dans l'éducation et la recherche

La régulation de l'IA générative en vue d'en exploiter les avantages potentiels pour l'éducation et la recherche nécessite l'élaboration de politiques appropriées. Les données de l'enquête 2023 citées ci-dessus indiquent que seuls quelques pays ont adopté des politiques ou des plans spécifiques pour l'utilisation de l'IA dans l'éducation. Le chapitre précédent a offert une perspective, identifié les étapes nécessaires et les éléments-clefs ainsi que les actions qui peuvent être entreprises par les différentes parties prenantes. Ce chapitre-ci présente les mesures qui peuvent être prises pour élaborer des cadres politiques cohérents et exhaustifs dans le but de réglementer l'utilisation de l'IA générative dans l'éducation et la recherche.

La publication de 2022 *IA et éducation : guide pour les décideurs politiques* (UNESCO, 2022b) constitue à cet égard un point de départ. Ce document propose un ensemble complet de recommandations pour conseiller les gouvernements dans l'élaboration et la mise en œuvre de politiques sectorielles sur l'IA et l'éducation, en mettant l'accent sur la promotion d'une éducation de qualité, l'équité sociale et l'inclusion. La plupart des recommandations restent pertinentes et peuvent être ajustées pour guider au mieux la formulation de politiques spécifiques sur l'IA générative dans l'éducation. Huit mesures spécifiques pour la planification des politiques sur la *GenAI* dans l'éducation et la recherche visent à compléter cette orientation existante.

4.1 Encourager l'inclusion, l'équité et la diversité linguistique et culturelle

L'importance cruciale de l'inclusion doit être reconnue et abordée tout au long du cycle de vie de l'IA générative. Plus précisément, les outils de la *GenAI* ne pourront pas contribuer à relever les défis fondamentaux de l'éducation ou à réaliser les engagements de l'ODD 4 si ces outils ne sont pas rendus accessibles à tous (indépendamment du sexe, de l'appartenance ethnique, des besoins éducatifs spécifiques, du statut socio-économique, de la situation géographique, du statut de déplacé, etc.) et si, de par leur conception même, ils ne font pas progresser l'équité, la diversité linguistique et le pluralisme culturel. Pour y parvenir, trois mesures politiques sont préconisées.

- Identifier ceux qui n'ont pas ou ne peuvent pas se permettre d'avoir une connexion internet ou des données numériques, et prendre les mesures nécessaires pour promouvoir la connectivité universelle et les compétences numériques, afin de réduire les obstacles à un accès équitable et inclusif aux applications de l'IA. Activer des mécanismes de financement durables pour le développement et la mise à disposition d'outils basés sur l'IA auprès des apprenants porteurs de handicap ou qui ont des besoins particuliers. Promouvoir l'utilisation de l'IA générative pour soutenir l'apprentissage tout au long de la vie, quels que soient l'âge, le lieu de résidence ou l'origine des apprenants.
- Définir des critères pour la validation des systèmes de *GenAI* afin de s'assurer qu'il n'existe pas de préjugés sexistes, de discrimination à l'encontre des groupes marginalisés ou de discours haineux intégrés dans les données ou les algorithmes.
- Développer et mettre en œuvre des spécifications inclusives pour les systèmes de *GenAI* et appliquer des mesures institutionnelles pour protéger la diversité linguistique et culturelle lors du déploiement de l'IA générative dans l'éducation et la recherche à grande échelle. Pour être pertinentes, les spécifications devraient exiger des fournisseurs de l'IA générative qu'ils incluent dans l'entraînement des modèles GPT des données en plusieurs langues, en particulier des langues locales ou indigènes, afin d'améliorer la capacité de l'IA générative à répondre à des textes multilingues et à en produire. Les spécifications et les mesures institutionnelles devraient empêcher, de manière très stricte, les fournisseurs d'IA de supprimer, intentionnellement ou non, les langues minoritaires ou de discriminer les locuteurs de langues indigènes, et exiger de ces fournisseurs qu'ils cessent d'exploiter des systèmes qui promeuvent les langues ou les normes culturelles dominantes.

4.2 Protéger l'humain

À mesure que l'IA générative devient de plus en plus sophistiquée, un danger majeur réside dans sa capacité à saper l'action humaine. Comme de plus en plus d'individus s'en servent comme d'un support pour leur écriture ou pour d'autres activités de création, ils pourraient involontairement en venir à s'y fier. Cela peut compromettre le développement des compétences intellectuelles. Si l'IA générative peut être utilisée pour défier et élargir la pensée humaine, on ne doit cependant pas la laisser usurper cette pensée humaine. La protection et l'amélioration de l'esprit humain devraient toujours être au cœur de la réflexion lors de la conception et du déploiement de l'IA générative, selon sept perspectives.

- Informer les apprenants des types de données que la *GenAI* peut collecter auprès d'eux, de la manière dont ces données sont utilisées et de l'impact qu'elles peuvent avoir sur leur éducation et plus généralement sur leur vie.
- Protéger la motivation intrinsèque des apprenants à grandir et à apprendre en tant qu'individus. Renforcer l'autonomie des humains sur leurs propres approches de la recherche, de l'enseignement et de l'apprentissage dans un contexte d'utilisation de systèmes de *GenAI* de plus en plus sophistiqués.
- Empêcher l'utilisation de la *GenAI* lorsqu'elle priverait les apprenants de la possibilité de développer leurs capacités cognitives et leurs compétences sociales par l'observation du monde réel, par des pratiques empiriques pouvant être des expériences, des discussions avec d'autres humains, ou par un raisonnement logique indépendant.
- Garantir une interaction sociale suffisante et une exposition adéquate aux créations humaines, éviter que les apprenants ne deviennent dépendants de la *GenAI*.
- Utiliser les outils *GenAI* pour réduire la pression des devoirs et des examens, plutôt que de l'accentuer.
- Consulter les chercheurs, les enseignants et les apprenants pour connaître leur point de vue sur l'IA générative et utiliser le retour d'information pour décider si des outils spécifiques de l'IA générative doivent être déployés à l'échelle institutionnelle, et le cas échéant comment ils doivent l'être. Encourager les apprenants, les enseignants et les chercheurs à critiquer et à remettre en question les méthodologies qui sous-tendent les systèmes d'IA, l'exactitude des contenus produits et les normes ou les pédagogies qu'ils peuvent instituer.
- Éviter de déléguer la responsabilité humaine aux systèmes de *GenAI* lorsqu'il s'agit de prendre des décisions aux enjeux cruciaux.

4.3 Surveiller et valider les systèmes d'IA générative pour l'éducation

Nous l'avons déjà indiqué, le développement et le déploiement des *GenAI* devraient être éthiques dès leur conception. Par la suite, une fois que l'IA générative est utilisée, et tout au long de son cycle de vie, elle doit être soigneusement contrôlée et validée – risques éthiques, pertinence et rigueur pédagogiques, impact sur les élèves, les étudiants, les enseignants et sur les relations entre la classe et l'école. À cet égard, cinq actions sont recommandées.

- Mettre en place des mécanismes de validation pour vérifier si les systèmes d'IA générative utilisés dans l'enseignement et la recherche ne comportent pas de préjugés, en particulier de préjugés sexistes, et s'ils sont entraînés à partir de données représentatives de la diversité (en termes de genre, de handicap, de statut social et économique, d'origine ethnique et culturelle, de localisation géographique).
- Aborder la question complexe du consentement éclairé, en particulier dans les contextes où les mineurs ou d'autres apprenants vulnérables ne sont pas en mesure de fournir un consentement véritablement éclairé.
- Vérifier si les contenus générés par la *GenAI* comprennent des images truquées (*deepfake*), des infox (inexactes ou fausses) ou des discours haineux. S'il s'avère que la *GenAI* génère des contenus inappropriés, les institutions et les éducateurs doivent être prêts à prendre des mesures rapides et énergiques pour éliminer le problème, à tout le moins en atténuer les effets négatifs, et être capables de le faire.
- Procéder à une validation stricte des applications de *GenAI* au point de vue éthique avant qu'elles ne soient officiellement mises en place dans les établissements d'enseignement et de recherche (autrement dit, adopter une approche éthique dès la conception).
- Avant de prendre des décisions sur le déploiement institutionnel, s'assurer que les applications de *GenAI* en question ne causent pas de dommages prévisibles aux élèves et aux étudiants, qu'elles sont efficaces et pertinentes sur le plan éducatif en fonction de l'âge et des capacités des apprenants concernés, et qu'elles sont en phase avec des principes pédagogiques sains (c'est-à-dire fondés et pertinents relativement aux domaines de connaissance, aux résultats d'apprentissage attendus et au développement des valeurs visés).

4.4 Développer les compétences des apprenants en IA, y compris des compétences liées à l'IA générative

Pour une utilisation sécurisée, éthique et utile de l'IA dans le domaine de l'éducation, et plus généralement dans la vie, il est indispensable que les apprenants développent leurs compétences en matière d'IA. Toutefois, selon les données de l'UNESCO, au début de 2022 (UNESCO, 2022c), seuls une quinzaine de pays avaient élaboré et mis en œuvre, ou étaient en train de concevoir, dans les établissements scolaires, des programmes d'enseignement de l'IA validés à l'échelle gouvernementale. Les derniers développements de l'IA générative ont encore renforcé le besoin pressant pour chacun d'atteindre un niveau d'alphabétisation suffisant dans les dimensions humaines et technologiques de l'IA, de comprendre son fonctionnement en termes généraux, ainsi que l'impact spécifique de la *GenAI*. Pour ce faire, cinq actions sont nécessaires et désormais urgentes.

- S'engager à fournir des programmes d'enseignement de l'IA validés par les instances gouvernementales pour l'éducation scolaire, l'enseignement et la formation techniques et professionnels, ainsi que pour l'apprentissage tout au long de la vie. Il est attendu que les programmes d'enseignement de l'IA couvrent l'impact de l'IA sur nos vies, y compris les questions éthiques qu'elle soulève, mais également qu'ils abordent la compréhension des algorithmes et des données, de manière adaptée à l'âge des apprenants, et qu'ils traitent aussi des compétences nécessaires pour l'utilisation pertinente et créative des outils de l'IA, y compris les applications de la *GenAI*.
- Soutenir les établissements d'enseignement supérieur et de recherche pour qu'ils améliorent les programmes visant à renforcer et développer les talents locaux en matière d'IA.
- Promouvoir l'égalité des sexes dans le développement des compétences avancées en matière d'IA et créer un vivier de professionnels respectant un équilibre entre hommes et femmes.
- Étudier les prévisions intersectorielles concernant les changements d'emplois provoqués, à l'échelle nationale et internationale, par la dernière automatisation de la *GenAI*, et, sur la base des changements prospectifs de la demande, améliorer les compétences durables à long terme pour l'éducation à tous les niveaux, ainsi que pour les systèmes d'apprentissage tout au long de la vie.
- Proposer des programmes spéciaux pour les travailleurs et les citoyens seniors qui pourraient avoir besoin d'acquérir de nouvelles compétences et de s'adapter à de nouveaux environnements.

4.5 Renforcer la capacité des enseignants et des chercheurs à utiliser correctement l'IA générative

Selon les données de l'enquête 2023 sur l'utilisation de l'IA pour l'éducation par les gouvernements (UNESCO, 2023c), seuls sept pays (Chine, Espagne, Finlande, Géorgie, Qatar, Thaïlande et Turquie) ont déclaré avoir élaboré ou être en train d'élaborer des cadres ou des programmes de formation sur l'IA à l'intention des enseignants. Seul le ministère de l'Éducation de Singapour a indiqué avoir créé un référentiel en ligne axé sur l'utilisation de ChatGPT dans l'enseignement et l'apprentissage. Cela montre clairement que dans la plupart des pays, les enseignants n'ont pas accès à une formation bien structurée sur l'utilisation de l'IA dans l'éducation, en particulier sur l'utilisation de la *GenAI*.

Pour préparer les enseignants à l'utilisation responsable et efficace de l'IA générative, quatre mesures sont proposées aux différents pays.

- Formuler ou adapter des orientations basées sur des études locales pour aider les chercheurs et les enseignants à naviguer dans les outils de *GenAI* facilement accessibles, et orienter la conception de nouvelles applications d'IA spécifiques à un domaine.
- Protéger les droits des enseignants et des chercheurs lorsqu'ils utilisent la *GenAI*, mais aussi la valeur de leurs pratiques. Plus précisément, analyser le rôle unique que jouent ces enseignants, qui facilitent la réflexion complexe, organisent les interactions des individus et promeuvent les valeurs humaines.
- Définir des orientations en matière de valeurs, de connaissances et de compétences dont les enseignants ont besoin pour comprendre et utiliser les systèmes *GenAI* de manière efficace et éthique. Permettre à ces enseignants de créer des outils spécifiques basés sur la *GenAI* pour faciliter l'apprentissage à la fois en classe et dans leur propre parcours professionnel.
- Examiner avec attention les compétences dont les enseignants ont besoin pour comprendre et utiliser l'IA pour l'enseignement, l'apprentissage et leur parcours professionnel, et intégrer de nouveaux ensembles de valeurs, de connaissances et de compétences sur l'IA dans les cadres de compétences et dans les programmes de formation des enseignants en formation initiale et continue.

4.6 Promouvoir le pluralisme des opinions et l'expression plurielle des idées

Comme indiqué précédemment, l'IA générative n'a de véritable compréhension ni de l'invite ni de la réponse apportée. Ses réponses sont basées sur les probabilités des modèles linguistiques entraînés à partir des données du Web qu'elle a intégrées. Pour résoudre certains problèmes fondamentaux liés à ses résultats, de nouvelles méthodes sont actuellement à l'étude, comme la connexion de la *GenAI* à des bases de données de connaissances et à des moteurs de raisonnement. En tout état de cause, en raison de son mode de fonctionnement, de ses sources et des points de vue implicites de ses développeurs, l'IA générative, par définition, reproduit les visions dominantes du monde dans ses résultats, au détriment des opinions minoritaires et plurielles. Par conséquent, pour que la civilisation puisse s'épanouir dans toutes ses dimensions, il est essentiel que nous reconnaissons que la *GenAI* ne pourra jamais constituer une source de connaissances faisant autorité, quel que soit le sujet qu'elle aborde.

Par conséquent, les utilisateurs doivent regarder les résultats de *GenAI* avec un œil critique et suivre quelques recommandations particulières.

- Comprendre le rôle de l'IA générative en tant que source d'information instantanée mais souvent peu fiable. Bien que certains plugins et outils basés sur le GML mentionnés précédemment soient conçus pour répondre à la nécessité d'accéder à des informations validées et à jour, il n'existe pas encore de preuves solides de leur efficacité.
- Encourager les apprenants et les chercheurs à se montrer critiques vis-à-vis des réponses fournies par l'IA générative. Reconnaître qu'elle ne fait généralement que répéter des opinions établies ou standard, masquant ainsi les opinions plurielles et minoritaires et la pluralité d'idées.
- Offrir aux apprenants suffisamment d'occasions d'apprendre par essais et erreurs, par des expériences empiriques et par l'observation du monde réel.

4.7 Tester des modèles d'application pertinents au niveau local et constituer une base de données cumulative

Les modèles de *GenAI* sont jusqu'à présent dominés par des informations provenant du Nord global et sous-représentent les voix du Sud global et des communautés indigènes. Ce n'est qu'au prix d'efforts résolus, par exemple en exploitant des données synthétiques (Marwala, 2023), que les outils de la *GenAI* seront adaptés au contexte et aux besoins des communautés locales, en particulier celles du Sud. Afin d'explorer des approches adaptées aux besoins locaux, tout en collaborant plus largement, on peut recommander huit actions.

- Veiller à ce que la conception et l'adoption des *GenAI* fassent l'objet d'une planification stratégique plutôt que privilégier un processus d'acquisition passif et non critique.
- Inciter les concepteurs des *GenAI* à cibler des options d'apprentissage ouvertes et variées, permettant une démarche d'exploration.
- Tester et développer des cas où l'IA est entraînée à partir de données probantes, conformément aux priorités éducatives, plutôt que sur la nouveauté, le mythe ou le battage médiatique.
- Accompagner l'utilisation de l'IA générative dans le but de stimuler l'innovation dans la recherche, notamment en exploitant les capacités de calcul, les données à grande échelle et les résultats de la *GenAI* pour façonner et inspirer l'amélioration des méthodologies de recherche.
- Examiner les implications sociales et éthiques introduites par l'intégration de la *GenAI* dans les processus de recherche.
- Établir des critères spécifiques basés sur des recherches et des méthodologies pédagogiques dûment éprouvées et construire une base de preuves de l'efficacité de l'IA générative en termes de soutien à l'offre d'éducation inclusive, de réalisation des objectifs d'apprentissage et de recherche, mais aussi de promotion des diversités linguistiques et culturelles.
- Adopter des mesures itératives pour renforcer les preuves de l'impact social et éthique de la *GenAI*.
- Analyser les coûts environnementaux de l'exploitation des technologies d'IA à grande échelle (par exemple, l'énergie et les ressources nécessaires à l'entraînement des modèles de GPT) et définir des objectifs durables à atteindre par les fournisseurs d'IA afin d'éviter d'aggraver le changement climatique.

4.8 Examiner les implications à long terme d'une manière intersectorielle et interdisciplinaire

Les approches intersectorielles et interdisciplinaires sont essentielles à l'utilisation efficace et éthique de l'IA générative dans l'éducation et la recherche. Ce n'est qu'en s'appuyant sur un éventail d'expertises, tout en réunissant de multiples parties prenantes, que les principaux défis seront identifiés rapidement et traités efficacement, dans le but de minimiser les implications négatives à long terme tout en tirant parti des avantages dans la continuité et selon une logique cumulative. Dans ce cadre, trois actions sont recommandées.

- Collaborer avec les développeurs d'IA, les éducateurs, les chercheurs et les représentants des parents, des élèves et des étudiants pour planifier des ajustements à l'échelle systémique dans les programmes d'études et les méthodologies d'évaluation, afin d'exploiter pleinement le potentiel de la *GenAI* pour l'éducation et la recherche et d'en atténuer les risques.
- Constituer une expertise intersectorielle et interdisciplinaire comprenant des éducateurs, des chercheurs, des spécialistes de l'apprentissage, des ingénieurs en IA et des représentants d'autres parties prenantes, afin d'examiner les implications à long terme de la *GenAI* pour l'apprentissage et la production de connaissances, la recherche et les droits d'auteur, les programmes d'études et l'évaluation, ainsi que la collaboration humaine et la dynamique sociale.
- Fournir des conseils pertinents pour informer les mises à jour itératives des réglementations et des politiques.

5. Favoriser l'utilisation créative de l'IA générative dans l'éducation et la recherche

Dès la sortie de ChatGPT, des inquiétudes ont été exprimées par les enseignants du monde entier quant au potentiel qu'il offrait pour générer des écrits et à la manière dont il pourrait aider les apprenants de tous niveaux à tricher. Plus récemment, de nombreuses personnes et institutions, y compris certaines des plus grandes universités à l'échelle internationale, ont affirmé que « le génie était sorti de la bouteille », et que des outils comme ChatGPT sont là pour durer et peuvent être utilisés de manière productive dans des contextes éducatifs. Dans le même temps, les internets regorgent de suggestions sur l'usage de l'IA générative dans l'éducation et la recherche. Il s'agit notamment d'utiliser cet outil pour inspirer de nouvelles idées, créer des exemples multi-perspectives, développer des plans de cours et des présentations, résumer des documents existants et motiver la création d'images. Bien que de nouvelles idées apparaissent sur les réseaux sociaux presque tous les jours, les chercheurs et les éducateurs sont encore en train de déterminer exactement ce que la *GenAI* signifie pour l'enseignement, l'apprentissage et la recherche. Plus particulièrement, les personnes à l'origine de nombreuses propositions d'utilisation peuvent ne pas avoir considéré les principes éthiques à leur juste mesure, tandis que d'autres sont motivées par le potentiel technologique de la *GenAI* plutôt que par les besoins des chercheurs, des enseignants ou des apprenants. Ce chapitre présente des moyens de faciliter l'utilisation créative de l'IA générative dans l'éducation.

5.1 Stratégies institutionnelles pour faciliter l'utilisation responsable et créative de l'IA générative

Comme indiqué précédemment, les établissements d'enseignement et de recherche devraient élaborer, mettre en œuvre et valider des stratégies et des cadres éthiques appropriés pour guider l'utilisation responsable et éthique des systèmes et des applications de la *GenAI*, le tout avec l'objectif de répondre aux besoins de l'enseignement, de l'apprentissage et de la recherche. Quatre stratégies sont proposées afin d'atteindre cet objectif.

- Mise en œuvre institutionnelle des principes éthiques.** Veiller à ce que les chercheurs, les enseignants et les apprenants utilisent les outils de *GenAI* de manière responsable et éthique, et qu'ils aient une approche critique quant à l'exactitude et la validité des résultats.
- Orientation et formation.** Fournir aux chercheurs, aux enseignants et aux apprenants des conseils et une formation sur les outils de *GenAI* pour s'assurer qu'ils comprennent les questions éthiques, par exemple les biais dans l'étiquetage des données et les algorithmes, et qu'ils se conforment aux réglementations en vigueur sur la confidentialité des données à caractère personnel et la propriété intellectuelle.
- Renforcer les capacités d'ingénierie pour les invites de la *GenAI*.** En plus des connaissances spécifiques au sujet, les chercheurs et les enseignants auront également besoin d'une expertise dans l'ingénierie et l'évaluation critique des messages d'instruction soumis à l'IA générative. Vu la complexité des défis posés par la *GenAI*, les chercheurs et les enseignants doivent recevoir une formation et un soutien de qualité pour y parvenir.
- Détecter le plagiat basé sur la *GenAI* dans les travaux écrits.** La *GenAI* pourrait permettre aux élèves et aux étudiants de faire passer pour leur propre travail un texte qu'ils n'ont pas écrit, ce qui constituerait un nouveau type de « plagiat ». Les fournisseurs d'IA générative sont tenus d'apposer sur leurs produits un filigrane « généré par l'IA » ; par ailleurs, des outils sont en cours d'élaboration pour permettre l'identification des contenus produits par l'IA. Notons cependant que ces mesures ne semblent guère efficaces. La stratégie institutionnelle immédiate consiste plutôt à faire respecter la déontologie académique et à renforcer la responsabilité de chacun grâce à une détection rigoureuse effectuée par des êtres humains. La stratégie sur le long terme consiste, pour les établissements et les éducateurs, à repenser la conception des devoirs écrits pour éviter qu'ils ne soient utilisés à évaluer des tâches que les outils de *GenAI* peuvent accomplir mieux que les apprenants humains. Ces productions écrites devraient plutôt évaluer ce que les humains savent faire et que la *GenAI* et les autres outils d'IA ne peuvent pas faire, y compris l'application de valeurs humaines, telles que la compassion et la créativité, à des défis complexes du monde réel.

5.2 Une approche d'interaction « centrée sur l'humain et pertinente sur le plan pédagogique »

Lorsqu'ils décident de l'opportunité et de la manière d'utiliser les IA génératives, les chercheurs et les éducateurs devraient donner la priorité à l'humain et à l'interaction responsable et adéquate, sur le plan pédagogique, entre les humains et les outils d'IA. Cela inclut notamment cinq aspects.

- l'utilisation de l'outil ou des outils doit contribuer à répondre aux besoins des êtres humains et à rendre l'apprentissage ou la recherche plus efficace qu'une approche sans technologie ou une autre approche alternative ;
- l'utilisation du ou des outils par les éducateurs et les apprenants doit être fondée sur leur motivation intrinsèque ;
- le processus d'utilisation du ou des outils doit être contrôlé par les éducateurs, les apprenants ou les chercheurs ;
- le choix et l'organisation du ou des outils et du contenu généré devraient répondre à un certain équilibre, en fonction de la tranche d'âge des apprenants, des résultats attendus et du type de connaissance visée (par exemple, factuelle, conceptuelle, procédurale ou métacognitive) ou du problème concerné (par exemple, bien structuré ou mal structuré) ;
- les processus d'utilisation devraient garantir une implication engagée des humains avec la *GenAI* et la pensée élaborée, ainsi que la responsabilité humaine pour les décisions liées à l'exactitude du contenu généré par l'IA, les stratégies d'enseignement ou de recherche, et leur impact sur les comportements humains.

5.3 Co-conception de l'utilisation de l'IA générative dans l'éducation et la recherche

L'utilisation de la *GenAI* dans l'éducation et la recherche ne devrait être ni imposée par une approche descendante, ni motivée par une agitation commerciale. Au contraire, son utilisation sûre et efficace devrait être conçue conjointement par les enseignants, les apprenants et les chercheurs. Il est également nécessaire de disposer d'un processus de pilotage et d'évaluation solide pour examiner l'efficacité et l'impact à long terme des différentes utilisations.

Pour faciliter cette co-conception, ce document d'orientation propose un cadre composé de six perspectives afin de consolider les interactions appropriées sur le plan pédagogique et de donner la priorité à l'action humaine.

- la pertinence des domaines de connaissances ou des problèmes ;
- les résultats attendus ;
- les outils appropriés en matière de *GenAI* et leurs avantages comparatifs ;
- les exigences de la part des utilisateurs ;
- les méthodes pédagogiques humaines requises et les exemples d'invites ;
- les risques éthiques.

Ce chapitre donne des exemples de la façon dont un processus de co-conception dans l'utilisation de l'IA générative peut éclairer les pratiques de recherche, aider l'enseignement, fournir un encadrement pour l'acquisition à son propre rythme de compétences fondamentales, faciliter la pensée complexe et soutenir les apprenants ayant des besoins particuliers. Ces exemples ne représentent que la partie émergée de l'iceberg au vu du nombre croissant de domaines concernés.

5.3.1 L'IA générative au service de la recherche

Les modèles d'IA générative ont démontré leur potentiel pour élargir les perspectives de recherche et enrichir non seulement l'exploration des données mais aussi les analyses documentaires (voir le **Tableau 3**). Bien qu'un éventail plus large d'utilisations puisse émerger, de nouvelles recherches sont nécessaires pour définir le domaine potentiel des problématiques de recherche et les résultats attendus, pour démontrer efficacité et précision, et pour garantir que l'action humaine dans la compréhension du monde réel par le biais de la recherche ne sera pas menacée par l'utilisation d'outils d'IA.

Tableau 3. Co-conception des utilisations de la GenAI pour la recherche

Utilisations potentielles mais non vérifiées	Domaines de connaissances ou problèmes concernés	Résultats attendus	Outils de GenAI et avantages comparatifs	Exigences pour les utilisateurs	Méthodes pédagogiques et exemples d'invites	Risques
Conseiller en IA pour des projets de recherche	Peut être utile dans des domaines bien structurés relatifs aux problèmes de recherche	Développer et répondre à des questions de recherche, suggérer des méthodologies adéquates. Transformation potentielle : Coach particulier pour la planification de la recherche	En commençant par la liste de la section 1.2, évaluer si les outils de GenAI sont accessibles localement, à source ouverte, rigoureusement testés ou validés par les autorités. Examiner plus en détail les avantages et les défis de tout outil de GenAI particulier et s'assurer qu'il répond correctement aux besoins humains spécifiques.	Le chercheur doit avoir une connaissance de base du ou des sujets. Le chercheur doit développer sa capacité à vérifier l'information et être particulièrement capable de détecter les citations de documents de recherche inexistantes.	Idées de base pour la définition des problèmes de recherche (par exemple, public cible, questions, contexte), ainsi que des méthodologies, des résultats attendus et des formats. Exemple d'invite : <i>Rédigez 10 questions de recherche potentielles pour [le sujet x] et classez-les par ordre d'importance pour [le domaine de recherche y].</i>	Être attentif au risque, bien réel, que la GenAI invente des informations (telles que des publications de recherche inexistantes) et que les utilisateurs soient tentés de copier et coller des plans de recherche générés par l'IA, ce qui peut réduire les possibilités pour les jeunes chercheurs d'apprendre par essais et erreurs.
Explorateur de données génératives et analyse documentaire	Peut être utile dans les domaines mal structurés des problèmes de recherche.	Collecte automatique d'informations, exploration d'un large éventail de données, proposition de projets de revues de la littérature et automatisation de certaines parties de l'interprétation des données. Transformation potentielle : Des formateurs en IA pour l'exploration des données et l'analyse de la littérature	En commençant par la liste de la section 1.2, évaluer si les outils de GenAI sont accessibles localement, à source ouverte, rigoureusement testés ou validés par les autorités. Examiner plus en détail les avantages et les défis de tout outil de GenAI particulier et s'assurer qu'il répond correctement aux besoins humains spécifiques.	Les chercheurs doivent avoir de solides connaissances quant aux méthodologies et aux techniques d'analyse des données.	Définition progressive des problèmes, étendue des données et sources de la littérature, méthodologies utilisées pour l'exploration des données et l'analyse de la littérature, et aussi résultats attendus et format de ces résultats.	Se méfier des informations fabriquées par la GenAI, du traitement inapproprié des données, des éventuelles violations de la vie privée, du profilage non autorisé et des préjugés sexistes. Nécessité d'être attentif à la propagation des normes dominantes et à la menace qu'elles font peser sur les normes alternatives et les opinions plurielles.

5.3.2 L'IA générative comme facilitateur de l'enseignement

L'utilisation de plateformes d'IA générative générales et d'outils de GenAI éducatifs spécifiques devrait être conçue en vue d'améliorer la compréhension qu'ont les enseignants de leur matière, mais aussi la connaissance qu'ils ont des méthodes d'enseignement, par exemple la co-conception par l'enseignant et l'IA de plans de cours, d'ensembles de cours ou de programmes complets. Les assistants pédagogiques conversationnels assistés par la GenAI, ou « jumeaux

génératifs des assistants pédagogiques »⁵³, qui sont pré-entraînés sur la base de données provenant d'enseignants et de bibliothèques expérimentés, ont été testés dans certains établissements d'enseignement et peuvent présenter un potentiel jusqu'ici inconnu ainsi que des risques à ce jour inédits sur le plan éthique. Les processus d'application pratique et les itérations ultérieures de ces modèles doivent encore faire l'objet d'un audit minutieux au moyen du cadre recommandé dans le présent document d'orientation et être protégés par un contrôle exercé par un humain, comme le montre le **Tableau 4**.

Tableau 4. Co-conception des utilisations de la *GenAI* pour accompagner les enseignants et l'enseignement

Utilisations potentielles mais non vérifiées	Domaines de connaissances ou problèmes concernés	Résultats attendus	Outils de <i>GenAI</i> et avantages comparatifs	Exigences pour les utilisateurs	Méthodes pédagogiques et exemples d'invites	Risques
Co-concepteur de programme ou de cours	Connaissances conceptuelles sur certains sujets d'enseignement et connaissances des procédures concernant les méthodologies d'enseignement	Aider au processus de conception des programmes et des cours, notamment en exposant ou en développant des points de vue sur les domaines-clefs du sujet cible et en définissant la structure du programme. Il peut également aider les enseignants à concevoir les tests et les examens en proposant des exemples de questions et de grilles d'évaluation. Transformation potentielle : Programme d'études généré par l'IA	En commençant par la liste de la section 1.2, évaluer si les outils de <i>GenAI</i> sont accessibles localement, en source ouverte, rigoureusement testés et validés par les autorités. Examiner plus en détail les avantages et les défis de tout outil de <i>GenAI</i> spécifique et s'assurer qu'il répond correctement à des besoins humains particuliers.	Les enseignants doivent comprendre et préciser ce qu'ils veulent que le programme, les cours, les leçons ou les tests traitent et en vue de quel objectif, s'ils veulent aborder les connaissances procédurales (comment faire quelque chose) ou conceptuelles, et quelle théorie de l'enseignement ils souhaitent appliquer.	Questions à <i>GenAI</i> sur les propositions de structure et d'exemples de connaissances factuelles sur le(s) sujet(s) concerné(s), sur les propositions de méthodes et de processus d'enseignement pour des sujets ou des problèmes définis, ou sur la création d'ensembles de cours ou de plans de leçons basés sur le(s) sujet(s) et sur le formatage. Les individus concepteurs de programmes d'études doivent vérifier les connaissances factuelles et s'assurer de la pertinence des programmes de cours proposés.	Le risque que la <i>GenAI</i> impose des normes et des méthodes pédagogiques dominantes est élevé. Elle peut involontairement perpétuer des pratiques d'exclusion en faveur des groupes déjà riches en données et renforcer les inégalités dans l'accès à des opportunités éducatives pertinentes et de qualité, en désavantageant les groupes pauvres en données.
Chatbot génératif jouant le rôle d'assistant d'enseignement	Connaissance conceptuelle de plusieurs domaines dans des problèmes structurés.	Fournir un soutien personnalisé, répondre aux questions et identifier les ressources. Transformation potentielle : Les jumeaux génératifs des assistants d'enseignement	En commençant par la liste de la section 1.2, évaluer si les outils de <i>GenAI</i> sont accessibles localement, en source ouverte, rigoureusement testés et validés par les autorités. Examiner plus en détail les avantages et les défis de tout outil de <i>GenAI</i> spécifique et s'assurer qu'il répond correctement à des besoins humains particuliers.	Il aide les enseignants mais cible directement les apprenants, ce qui exige de ces derniers qu'ils aient suffisamment de connaissances préalables, d'aptitudes et de compétences métacognitives pour contrôler les résultats de la <i>GenAI</i> et identifier les informations erronées. Elle pourrait donc être plus pertinente pour les étudiants de l'enseignement supérieur.	Exige des enseignants qu'ils comprennent clairement les problèmes, qu'ils surveillent la conversation et qu'ils aident les apprenants à vérifier les réponses sujettes à caution fournies par la <i>GenAI</i> .	Compte tenu des capacités actuelles des modèles de <i>GenAI</i> , les établissements d'enseignement doivent garantir une supervision humaine des réponses fournies par les outils de <i>GenAI</i> , en portant une attention particulière aux risques de désinformation. Il peut également limiter l'accès des apprenants à des conseils et à un soutien humains, en empêchant que se crée et se développe une relation solide entre l'enseignant et l'élève, ce qui est particulièrement préoccupant concernant les enfants et les adolescents.

5.3.3 L'IA générative en tant que « coach particulier » pour l'acquisition à son rythme de compétences fondamentales

Alors que la réflexion complexe et la créativité font l'objet d'une attention croissante dans la définition des résultats de l'apprentissage, l'importance des compétences fondamentales dans le développement psychologique des enfants et des adolescents et la progression de leurs compétences ne fait aucun doute. Parmi de nombreuses aptitudes, ces compétences fondamentales comprennent

l'écoute, la prononciation et l'écriture d'une langue maternelle ou étrangère, ainsi que les notions élémentaires de calcul, d'art et de codage. La méthode pédagogique de « l'exercice et la pratique » ne devrait pas être considérée comme obsolète ; au contraire, elle devrait être redynamisée et améliorée par les technologies de la *GenAI* afin de favoriser la répétition par les apprenants, chacun à son propre rythme, des compétences fondamentales. S'ils sont guidés par des principes éthiques et pédagogiques, les outils de la *GenAI* ont le potentiel de devenir des tuteurs, mentors, coachs particuliers pour une telle pratique, comme l'illustre le **Tableau 5**.

Tableau 5. Co-conception de l'utilisation de la *GenAI* en tant que coach particulier pour l'acquisition à son rythme de compétences fondamentales dans les langues et les arts

Utilisations potentielles mais non vérifiées	Domaines de connaissances ou problèmes concernés	Résultats attendus	Outils de <i>GenAI</i> et avantages comparatifs	Exigences pour les utilisateurs	Méthodes pédagogiques et exemples d'invites	Risques
Tuteur linguistique particulier	Apprentissage des langues, y compris pratique de la conversation	Engager les apprenants dans une pratique conversationnelle pour les aider à améliorer leurs compétences en matière de compréhension orale, d'expression orale et d'écriture, en leur proposant un retour d'information, des corrections et une modélisation de leur langue maternelle ou de la langue étrangère concernée. Aider les apprenants à améliorer leurs compétences relatives à l'écriture. Transformation potentielle : Cours particulier de langue au niveau débutant	En commençant par la liste de la section 1.2, évaluer si les outils de <i>GenAI</i> sont accessibles localement, en source ouverte, rigoureusement testés et validés par les autorités. Examiner plus en détail les avantages et les défis de tout outil de <i>GenAI</i> spécifique et s'assurer qu'il répond correctement à des besoins humains particuliers.	Une limite d'âge peut être fixée pour les conversations en autonomie en raison des résultats ne prenant pas en compte le contexte culturel ou inadaptés à l'âge qui sont fournis par les systèmes <i>GenAI</i> . L'apprenant doit avoir la motivation intrinsèque pour engager une conversation avec un système d'IA. L'apprenant doit être capable d'adopter une approche critique vis-à-vis des contenus de la <i>GenAI</i> et de s'assurer qu'ils sont exacts.	Lorsqu'ils utilisent les plateformes générales de <i>GenAI</i> , les enseignants (humains) peuvent guider les apprenants pour qu'ils engagent un échange avec les outils de <i>GenAI</i> afin de demander un retour d'information pour s'améliorer, corriger leur prononciation ou fournir des exemples d'écriture. Par exemple : <i>Démarrez avec moi une conversation dans la langue [x] pour m'aider dans ma progression.</i> <i>Suggérez-moi des idées pour m'aider à écrire sur [sujet x].</i>	Nécessité d'être attentif à un langage ne prenant pas en compte le contexte culturel ou inexact au vu du contexte, ainsi qu'à la perpétuation involontaire de stéréotypes ou de préjugés culturels. En l'absence de stratégies pédagogiques pertinentes pour simuler les motivations intrinsèques des apprenants, la créativité et l'originalité des enfants risquent d'être limitées, ce qui les conduira à écrire de manière standardisée, avec des formules toutes faites. Le risque existe également de limiter les possibilités d'interactions dans la vie réelle, d'opinions et d'expressions plurielles et de freiner la pensée critique.
Professeur particulier pour les matières artistiques	Compétences techniques dans des domaines artistiques tels que la musique ou le dessin	Proposer des suggestions de techniques artistiques (par exemple, des conseils sur la perspective et la couleur) ou de composition musicale (par exemple, la mélodie et la progression des accords). Transformation potentielle : Professeur particulier d'art au niveau débutant	En commençant par la liste de la section 1.2, évaluer si les outils de <i>GenAI</i> sont accessibles localement, en source ouverte, rigoureusement testés et validés par les autorités. Examiner plus en détail les avantages et les défis de tout outil de <i>GenAI</i> spécifique et s'assurer qu'il répond correctement à des besoins humains particuliers.	Dès le début, les apprenants doivent avoir des objectifs initiaux de création artistique ou musicale, une compréhension au moins minimale des éléments-clés du domaine de l'art ou de la musique, et les capacités de base pour analyser les œuvres d'art ou les compositions musicales.	Les enseignants (humains) peuvent demander aux apprenants de comparer les techniques artistiques des outils d'IA avec leurs propres œuvres d'art. Les enseignants ou accompagnateurs (humains) doivent encourager les apprenants à développer et à appliquer leur imagination et leur créativité, que la <i>GenAI</i> ne peut pas remplacer. Exemple d'invite : <i>Suggérez-moi des idées qui pourraient m'inspirer pour créer une image sur [sujets/ idées]</i>	Peut exposer les jeunes apprenants à des contenus inappropriés ou choquants, ce qui peut porter atteinte à leur droit à la sécurité et au bien-être. Les outils de <i>GenAI</i> risquent d'empêcher les apprenants de développer leur imagination et leur créativité

Tableau 5. Co-conception de l'utilisation de la GenAI en tant que coach particulier pour l'acquisition à son rythme de compétences fondamentales dans les langues et les arts

Utilisations potentielles mais non vérifiées	Domaines de connaissances ou problèmes concernés	Résultats attendus	Outils de GenAI et avantages comparatifs	Exigences pour les utilisateurs	Méthodes pédagogiques et exemples d'invites	Risques
Tuteur personnel pour le codage ou les mathématiques	Connaissances et compétences conceptuelles en matière de programmation pour débutants. Également applicable à l'apprentissage des mathématiques de base.	Soutien à l'apprentissage à son propre rythme des connaissances et compétences de base en matière de codage, détection des bogues dans le codage des apprenants et retour d'information immédiat, adaptation des réponses aux questions. Transformation potentielle : Enseignant personnel de codage au niveau débutant	En commençant par la liste de la section 1.3, évaluer si les outils de GenAI sont accessibles localement, en source ouverte, rigoureusement testés et validés par les autorités. Examiner plus en détail les avantages et les défis de tout outil de GenAI spécifique et s'assurer qu'il répond correctement à des besoins humains particuliers.	La recherche et la définition d'un problème, ainsi que la conception d'algorithmes pour résoudre le problème, restent les aspects essentiels de l'apprentissage du codage et de la programmation. Les apprenants doivent avoir une motivation intrinsèque pour utiliser le codage, ainsi que des connaissances de base dans l'utilisation du langage de programmation.	Les enseignants et accompagnateurs (humains) devraient enseigner les connaissances et compétences de base, et inciter les apprenants à utiliser la pensée informatique et la programmation pour résoudre des problèmes, notamment par le biais du codage collaboratif. Exemple d'invite : <i>Proposez quelques idées inhabituelles de codage.</i>	L'exactitude du retour d'information et des suggestions reste un problème : la GenAI n'a pas toujours raison. Le risque est grand que les outils de GenAI empêchent les apprenants de développer des compétences de pensée computationnelle et des capacités à trouver et à définir des problèmes significatifs pour le codage.

5.3.4 L'IA générative pour faciliter l'apprentissage par enquête ou par projet

S'ils ne sont pas utilisés de manière volontaire pour faciliter la réflexion ou la créativité, les outils de l'IA générative ont tendance à encourager le plagiat ou les résultats superficiels de la « répétition stochastique ». Cependant, étant donné

que les modèles de GenAI ont été nourris sur la base de données à grande échelle, ils sont potentiellement capables d'agir comme un partenaire dans les échanges socratiques ou comme un assistant de recherche dans l'apprentissage basé sur des projets. Ces potentiels ne peuvent toutefois être exploités que par le biais de processus de conception d'enseignement ou d'apprentissage visant à déclencher une réflexion complexe, comme l'illustre le **Tableau 6**.

Tableau 6. Co-conception des utilisations de la GenAI pour faciliter l'apprentissage par enquête ou par projet

Utilisations potentielles mais non vérifiées	Domaines de connaissances ou problèmes concernés	Résultats attendus	Outils de GenAI et avantages comparatifs	Exigences pour les utilisateurs	Méthodes pédagogiques et exemples d'invites	Risques
Partenaire pour un échange socratique	Problèmes mal structurés.	Engager les apprenants dans un dialogue rappelant le questionnement socratique sur les connaissances antérieures, conduisant à la découverte de nouvelles connaissances ou à une compréhension plus approfondie. Transformation potentielle : Opposant socratique personnalisé	En commençant par la liste de la section 1.3, évaluer si les outils de GenAI sont accessibles localement, en source ouverte, rigoureusement testés et validés par les autorités. Examiner plus en détail les avantages et les défis de tout outil de GenAI spécifique et s'assurer qu'il répond correctement à des besoins humains particuliers.	L'apprenant doit avoir atteint l'âge qui lui permet de mener des conversations en autonomie avec les outils de la GenAI. Les apprenants doivent avoir des connaissances préalables et la capacité de vérifier l'exactitude des arguments et des informations fournis.	Les enseignants (humains) peuvent aider à préparer une liste de questions progressivement plus approfondies servant d'exemples que les apprenants peuvent adapter pour en faire des invites. Les apprenants peuvent également commencer par une demande générale telle que « <i>Engagez avec moi un dialogue socratique afin de m'aider à adopter une perspective critique à l'égard du [sujet x]</i> », puis approfondir progressivement le dialogue grâce à des demandes de plus en plus précises.	Les outils de GenAI peuvent générer des réponses similaires ou standard qui limitent la confrontation des apprenants à des points de vue divers et à des perspectives variées, ce qui conduit à un effet d'écho et entrave le développement d'une pensée indépendante.

Tableau 6. Co-conception des utilisations de la *GenAI* pour faciliter l'apprentissage par enquête ou par projet

Utilisations potentielles mais non vérifiées	Domaines de connaissances ou problèmes concernés	Résultats attendus	Outils de <i>GenAI</i> et avantages comparatifs	Exigences pour les utilisateurs	Méthodes pédagogiques et exemples d'invites	Risques
Conseiller pour l'apprentissage par projet	Problèmes de recherche mal structurés en sciences dures ou en sciences sociales.	Soutenir la création de connaissances en aidant les apprenants à mener des projets d'apprentissage. Cela implique que <i>GenAI</i> joue un rôle similaire à celui du conseiller en recherche décrit dans le tableau 3. Transformation potentielle : Conseiller personnel pour l'apprentissage par projet	En commençant par la liste de la section 1.3, évaluer si les outils de <i>GenAI</i> sont accessibles localement, en source ouverte, rigoureusement testés et validés par les autorités. Examiner plus en détail les avantages et les défis de tout outil de <i>GenAI</i> spécifique et s'assurer qu'il répond correctement à des besoins humains particuliers.	Les apprenants pourraient agir en tant que chercheurs juniors dans la conception et la mise en œuvre d'apprentissages basés sur des projets. Les apprenants doivent être suffisamment âgés pour pouvoir utiliser les plateformes <i>GenAI</i> de manière autonome. Les apprenants doivent avoir la motivation et la capacité de s'engager dans des activités d'apprentissage autodirigées basées sur des projets, pour ne pas être tentés de copier-coller passivement les réponses fournies par les outils de <i>GenAI</i> .	Les enseignants (humains) guident les apprenants pour qu'ils demandent à <i>GenAI</i> de fournir des idées de base pour la définition des problèmes de recherche, comme suggéré au point 5.3.1. Les apprenants, individuellement ou en groupe, utilisent les outils de <i>GenAI</i> pour effectuer des analyses documentaires, collecter et traiter des données, et rédiger des rapports.	Les apprenants qui n'ont pas de solides connaissances préalables et la capacité nécessaire pour vérifier l'exactitude des réponses peuvent être induits en erreur par les informations fournies par les outils de <i>GenAI</i> . Cela peut également limiter les discussions et les interactions des apprenants avec leurs pairs et réduire les possibilités d'apprentissage collaboratif, ce qui pourrait nuire à leur développement social.

5.3.5 L'IA générative pour soutenir les apprenants ayant des besoins particuliers

En théorie, les modèles d'IA générative peuvent être mobilisés pour aider les apprenants souffrant de déficiences auditives ou visuelles. Les pratiques émergentes prévoient des sous-titres ou des sous-titres activés par la *GenAI* pour les apprenants sourds et malentendants, et une description audio-générée par la *GenAI* pour les apprenants malvoyants. Les modèles d'IA générative peuvent également convertir le texte en parole et la parole en texte pour permettre aux personnes souffrant de déficiences visuelles, auditives ou vocales d'accéder au contenu, de poser des questions et de communiquer avec leurs pairs. Ces fonctionnalités n'ont toutefois pas encore été exploitées à grande échelle. Selon l'enquête précédemment mentionnée qui a été menée par l'UNESCO en 2023 sur l'utilisation de l'IA par les gouvernements dans l'éducation, seuls quatre pays (Chine, Jordanie, Malaisie et Qatar) ont indiqué que leurs autorités avaient validé et recommandé des outils assistés par l'IA pour soutenir l'accès inclusif des apprenants handicapés (UNESCO, 2023c).

On observe également une tendance à entraîner des itérations de modèles de *GenAI* pour soutenir les apprenants dans l'utilisation de leur(s) propre(s) langue(s), langues

minoritaires et indigènes comprises, pour apprendre et communiquer. Par exemple, PaLM 2, le GML de nouvelle génération de Google, est entraîné sur des données parallèles couvrant des centaines de langues sous la forme de paires de textes sources/cibles. L'intégration de données parallèles multilingues est pensée pour continuer d'améliorer la capacité du modèle à comprendre et à générer des textes multilingues (Google, 2023b).

En fournissant des traductions en temps réel, des paraphrases, des corrections automatiques, les outils de *GenAI* peuvent aider les apprenants qui utilisent des langues minoritaires à communiquer leurs idées et à améliorer leur collaboration avec des camarades issus de milieux linguistiques différents. Cela ne se produit toutefois pas à grande échelle de manière spontanée. Ce n'est qu'avec un projet ciblé que ce potentiel peut être exploité pour donner de la voix aux groupes marginalisés.

Enfin, on a aussi fait l'hypothèse que les systèmes d'IA générative peuvent effectuer des diagnostics à partir des conversations avec les utilisateurs, en identifiant les problèmes psychologiques ou socio-émotionnels, mais aussi les difficultés d'apprentissage. Cependant, il existe peu de preuves du caractère efficace ou sécurisé de cette approche, et tout diagnostic exigerait une interprétation par des professionnels qualifiés.

Tableau 7. Co-conception des utilisations de la GenAI pour soutenir les apprenants ayant des besoins particuliers

Utilisations potentielles mais non vérifiées	Domaines de connaissances ou problèmes concernés	Résultats attendus	Outils de GenAI et avantages comparatifs	Exigences pour les utilisateurs	Méthodes pédagogiques et exemples d'invites	Risques
Diagnostic des difficultés d'apprentissage à partir de la conversation	Peut être utile pour les apprenants qui rencontrent des difficultés d'apprentissage dues à des problèmes psychologiques, sociaux ou émotionnels.	Utiliser la langue naturelle pour identifier les besoins des apprenants qui ont des problèmes psychologiques, sociaux ou émotionnels ou des difficultés d'apprentissage, afin de leur fournir un soutien ou un enseignement adéquat. Transformation potentielle : Conseiller principal personnalisé pour les apprenants ayant des problèmes sociaux ou émotionnels ou des difficultés d'apprentissage	En plus des principaux outils de GenAI, rechercher des chatbots alimentés par GenAI. Déterminer s'ils sont accessibles localement, s'ils sont en <i>open source</i> , s'ils ont été rigoureusement testés ou validés par les instances compétentes. Examiner plus en détail les avantages et les défis de tout outil spécifique de GenAI et s'assurer qu'il répond correctement aux besoins humains particuliers.	Les enseignants ou les spécialistes qui travaillent avec ces groupes d'apprenants devront s'assurer que les principaux conseils suggérés par le système de GenAI sont rigoureux.	Les enseignants ou les animateurs doivent créer des environnements confortables pour engager l'apprenant dans une conversation destinée à diagnostiquer les problèmes psychologiques, sociaux ou émotionnels, ou les difficultés d'apprentissage.	Peut, par inadvertance, mal diagnostiquer les difficultés spécifiques de l'apprenant, ce qui conduit à lui apporter un soutien qui n'est pas pertinent.
Outils pour l'accessibilité fonctionnant grâce à l'IA	Permet aux apprenants malentendants ou malvoyants d'accéder à un plus large éventail de contenus, ce qui améliore la qualité de leur apprentissage.	Répondre aux besoins d'accès des apprenants et soutenir l'acquisition, de leur part, de connaissances spécifiques à un sujet en fournissant des sous-titres et/ou une interprétation en langue des signes pour les contenus audio ou vidéo, ainsi que des descriptions audio pour les textes ou autres matériels visuels. Transformation potentielle : Aides linguistiques personnalisées alimentées par l'IA.	En plus des principaux outils de GenAI, rechercher des chatbots alimentés par GenAI. Déterminer s'ils sont accessibles localement, s'ils sont en <i>open source</i> , s'ils ont été rigoureusement testés ou validés par les instances compétentes. Examiner plus en détail les avantages et les défis de tout outil spécifique de GenAI et s'assurer qu'il répond correctement aux besoins humains particuliers.	Les éducateurs ou les facilitateurs doivent aider les apprenants à accéder aux outils de GenAI et à apprendre à les utiliser. Ils doivent également s'assurer que les résultats des outils aident réellement ces apprenants et ne renforcent pas les défis et les préjugés auxquels ils sont confrontés.	Nécessité de tester l'accessibilité des plateformes ou des outils afin d'identifier et de résoudre les problèmes d'accessibilité avant leur utilisation. Les outils de l'IA générative ne peuvent que fournir un accès au contenu : les éducateurs et les facilitateurs devraient plutôt se concentrer sur l'amélioration de la qualité de l'apprentissage et le bien-être social. Les éducateurs et les facilitateurs doivent apprendre aux apprenants à créer des messages vocaux ou textuels en fonction de leurs capacités propres.	Les sous-titres ou les descriptions audio produits par les plateformes GenAI qui ne sont pas conçues spécifiquement pour aider à la vision ou à l'audition sont souvent inexacts et peuvent induire en erreur les apprenants ayant des besoins particuliers. Ces outils peuvent, par inadvertance, renforcer les préjugés existants.

Tableau 7. Co-conception des utilisations de la GenAI pour soutenir les apprenants ayant des besoins particuliers

Utilisations potentielles mais non vérifiées	Domaines de connaissances ou problèmes concernés	Résultats attendus	Outils de GenAI et avantages comparatifs	Exigences pour les utilisateurs	Méthodes pédagogiques et exemples d'invites	Risques
Amplificateur génératif pour les apprenants marginalisés	Il peut être utile pour les apprenants issus de minorités linguistiques ou culturelles d'exprimer et d'amplifier leur voix, de participer en ligne et de mener des études sociales en groupes.	Fournir des traductions en temps réel, des paraphrases et des corrections automatiques pour aider les apprenants issus de groupes marginalisés à utiliser leur propre langue pour communiquer avec des pairs issus de milieux linguistiques différents. Transformation potentielle : Des GML inclusifs pour les apprenants marginalisés	On peut prendre en exemple plus spécifiquement PaLM2 Déterminer si les outils de GenAI sont accessibles localement, s'ils sont en <i>open source</i> , s'ils ont été rigoureusement testés ou validés par les instances compétentes. Examiner plus en détail les avantages et les défis de tout outil spécifique de GenAI et s'assurer qu'il répond correctement aux besoins humains particuliers.	Les apprenants doivent avoir des connaissances ou des opinions constructives sur le sujet de la conversation ou de l'étude collaborative. Ils doivent être capables d'apporter des contributions non discriminatoires, faisant preuve du sens des responsabilités et d'éviter les discours emplies de haine.	Les enseignants ou les éducateurs doivent concevoir des sujets d'étude et des tâches d'écriture pour les apprenants sur des thèmes sociaux ou culturels, ou organiser des séminaires en ligne ou des collaborations interculturelles pour stimuler les apprenants dans la formulation des idées et le partage des opinions.	Nécessité d'identifier et de corriger les erreurs de traduction et de paraphrase de l'IA qui peuvent être à l'origine de malentendus interculturels. Cette utilisation peut permettre aux apprenants marginalisés de faire entendre leur voix, mais elle ne s'attaque pas à la cause première de la pauvreté en données et ne peut donc pas décoloniser les outils d'IA.

6. La place de l'IA générative dans l'avenir de l'éducation et de la recherche

Les technologies de l'IA générative évoluent encore rapidement et sont susceptibles d'avoir un impact marqué sur l'éducation et la recherche. Par conséquent, leurs implications potentielles à long terme pour l'éducation et la recherche doivent faire l'objet d'une attention immédiate et d'un examen approfondi.

de nouvelles formes de discrimination basés sur les données ayant servi à l'entraînement et sur les méthodes utilisées par les modèles, ce qui peut donner lieu à des résultats à ce jour inconnus et potentiellement dangereux.

6.1 Des questions éthiques inédites

Les outils de plus en plus sophistiqués de la *GenAI* ne vont pas manquer de poser de nouvelles questions éthiques, qui doivent être examinées en détail. En complément des sections 2 et 3, des analyses plus approfondies et plus prospectives sont nécessaires pour révéler et traiter les questions éthiques inexplorées, sous au moins cinq angles.

- **Accès et équité.** Les systèmes d'IA générative dans l'éducation peuvent accroître les disparités existantes dans l'accès à la technologie et aux ressources éducatives, aggravant encore les inégalités.
- **Connexion humaine.** Les systèmes d'IA générative dans l'éducation peuvent réduire l'interaction entre humains et les aspects socio-émotionnels de l'apprentissage, qui sont pourtant essentiels.
- **Le développement intellectuel humain.** Les systèmes d'IA générative dans l'éducation sont susceptibles de limiter l'autonomie et l'action des apprenants en fournissant des solutions préconçues ou en réduisant l'éventail des expériences d'apprentissage possibles. Leur impact à long terme sur le développement intellectuel des jeunes apprenants doit être étudié.
- **Impact psychologique.** Les systèmes d'IA générative qui imitent les interactions humaines sont susceptibles de produire des effets psychologiques inconnus sur les apprenants, ce qui soulève des inquiétudes concernant leur développement cognitif et leur bien-être émotionnel, mais aussi quant aux risques de manipulation.
- **Préjugés et discriminations cachés.** À mesure que des systèmes de *GenAI* plus sophistiqués sont développés et appliqués dans l'éducation, ils sont susceptibles de générer de nouveaux préjugés et

6.2 Droits d'auteur et propriété intellectuelle

L'émergence de l'IA générative modifie à un rythme très rapide la manière dont les œuvres scientifiques, artistiques et littéraires sont créées, distribuées et consommées. La copie, la diffusion ou l'utilisation non autorisée d'œuvres protégées par des droits d'auteur sans l'autorisation du détenteur de ces droits constitue une violation des droits de ce dernier et peut avoir des conséquences juridiques. L'entraînement des modèles de *GenAI* a ainsi été accusé d'enfreindre le droit d'auteur. Dans une affaire récente, la chanson générée par l'IA et mettant en scène « Drake » et « The Weeknd » (Abel Tesfaye) a atteint des millions d'auditeurs avant d'être retirée en raison d'un litige sur les droits d'auteur (Coscarelli, 2023). Alors que les cadres réglementaires émergents entendent exiger des fournisseurs d'IA générative qu'ils reconnaissent et protègent la propriété intellectuelle des propriétaires du contenu utilisé par le modèle, il devient de plus en plus difficile de déterminer la propriété et le caractère original de l'immense quantité d'œuvres générées. Ce manque de traçabilité est préoccupant : non seulement il soulève des inquiétudes quant à la protection des droits des créateurs et à la juste rémunération de leurs contributions intellectuelles, mais il pose également des problèmes dans les contextes éducatifs quant à la manière dont les résultats des outils de *GenAI* peuvent être utilisés de manière responsable. Cela peut avoir de profondes implications pour le système de recherche.

6.3 Sources de contenu et d'apprentissage

Les outils d'IA générative modifient la manière dont les contenus d'enseignement et d'apprentissage peuvent être générés et mis à disposition. À l'avenir, le contenu généré par des conversations entre humains et IA pourrait devenir l'une des principales sources de production de connaissances. Cela risque de compromettre encore davantage le contact direct des apprenants avec des contenus éducatifs basés sur des

ressources, des manuels et des programmes créés et validés par des humains. L'aspect « digne de foi » des textes de la *GenAI* peut induire en erreur les jeunes apprenants qui n'ont pas suffisamment de connaissances préalables pour pouvoir identifier les inexactitudes ou les remettre en question de manière efficace. Le travail des apprenants avec des contenus non validés doit-il être reconnu comme un « apprentissage », cette controverse reste ouverte.

Il en résulte une attention portée sur des informations agrégées de seconde main, ce qui peut également réduire les possibilités pour les apprenants de construire leurs connaissances par des méthodes éprouvées telles que la perception directe et l'expérience du monde réel, l'apprentissage par essais et erreurs, la réalisation d'expériences empiriques et l'utilisation du bon sens. Cette situation peut aussi menacer la dimension sociale de la construction de la connaissance et l'attention portée aux valeurs sociales via les pratiques collaboratives en classe.

6.4 Réponses homogénéisées versus résultats diversifiés et créatif

L'IA générative restreint la pluralité des récits car les résultats générés tendent à représenter et à renforcer les points de vue dominants. L'homogénéisation des connaissances qui en résulte limite la pensée pluraliste et créative. La dépendance accrue des enseignants, des élèves et des étudiants aux outils de la *GenAI* pour trouver des suggestions peut conduire à la standardisation et au caractère normalisé des réponses, ce qui affaiblit la valeur de la pensée critique et de l'interrogation effectuée de manière autonome. L'homogénéisation potentielle de l'expression dans les documents écrits et les œuvres d'art est susceptible de limiter l'imagination, la créativité et les perspectives d'expression originales des apprenants.

Les fournisseurs de *GenAI* et les éducateurs doivent examiner dans quelle mesure l'EdGPT pourrait être développée et utilisée pour favoriser la créativité, la collaboration, la pensée critique et d'autres compétences de réflexion à un niveau avancé.

6.5 Repenser l'évaluation et les résultats de l'apprentissage

Les implications de l'IA générative relatives à l'évaluation vont bien au-delà des préoccupations immédiates concernant les apprenants qui trichent dans les devoirs écrits. Nous devons faire face au fait que la *GenAI* peut produire des documents et des essais relativement bien organisés et des œuvres d'art impressionnantes, et qu'elle peut réussir certains examens testant les connaissances dans des matières précises données. Nous devons donc repenser ce

qui doit être réellement appris et dans quel but, et comment l'apprentissage doit être évalué et validé.

Le débat critique entre les éducateurs, les décideurs politiques, les apprenants et les autres parties prenantes doit prendre en compte quatre catégories de résultats d'apprentissage.

Valeurs. Les valeurs requises pour garantir une conception et une utilisation de la technologie centrées sur l'être humain sont essentielles pour repenser les résultats de l'apprentissage et leur évaluation en cette ère du numérique. En réexaminant l'objectif de l'éducation, on doit expliciter les valeurs qui sous-tendent la manière dont la technologie est liée à l'éducation. C'est dans cette optique normative que les acquis de l'apprentissage, leur évaluation et leur validation doivent être constamment réévalués, pour répondre à l'utilisation de plus en plus omniprésente de la technologie, intelligence artificielle comprise, dans la société.

Connaissances et compétences fondamentales.

Même dans les domaines de compétences où les outils de la *GenAI* peuvent faire mieux que les humains, les apprenants auront toujours besoin de connaissances et de compétences fondamentales solides. Les compétences fondamentales en matière de lecture et d'écriture, de calcul, de culture scientifique de base, resteront dans le futur essentielles pour l'éducation. La portée et la nature de ces compétences fondamentales devront être régulièrement revues pour refléter les environnements de plus en plus riches en IA dans lesquels nous vivons.

Compétences de réflexion complexe. Les résultats d'apprentissage devront prendre en compte les compétences requises pour soutenir une réflexion de qualité ou résoudre des problèmes complexes en prenant pour base la collaboration entre l'homme et l'IA et l'utilisation des résultats générés par la *GenAI*. Il s'agit entre autres de comprendre le rôle des connaissances factuelles et conceptuelles dans l'élaboration d'une réflexion avancée, et de l'évaluation critique du contenu généré par l'IA.

Compétences professionnelles nécessaires pour travailler avec l'IA.

Dans les domaines où l'IA peut faire mieux que les humains et automatiser les tâches, les apprenants humains doivent acquérir de nouvelles compétences qui leur permettent de développer, d'exploiter et de travailler avec les outils de la *GenAI*. La refonte des résultats d'apprentissage et de l'évaluation de l'enseignement devra refléter les compétences professionnelles nécessaires pour les nouveaux emplois créés par l'IA.

6.6 Processus de réflexion

La dimension la plus fondamentale des implications à long terme de l'IA générative pour l'éducation et la recherche concerne, une fois encore, la relation complémentaire entre l'humain et la machine. L'une des questions-clefs est de savoir si les humains peuvent, éventuellement, céder à l'IA les niveaux de base de la pensée et les processus d'acquisition de compétences pour se concentrer plutôt sur les compétences de pensée plus complexes, sur la base des résultats fournis par l'IA.

L'écriture, par exemple, est souvent associée à la structuration de la pensée. Avec l'IA générative, plutôt que de partir de zéro pour planifier les objectifs, la portée et les grandes lignes d'un ensemble d'idées, les humains peuvent désormais commencer avec une ébauche correctement structurée fournie par l'IA générative. Certains experts ont qualifié l'utilisation de la *GenAI* pour générer de cette manière du texte d'« écriture sans réflexion » (Chayka, 2023). Au fur et à mesure que ces nouvelles pratiques assistées par la *GenAI* sont adoptées, les méthodes établies pour l'acquisition et l'évaluation des compétences en écriture doivent s'adapter. Une option envisageable est que l'apprentissage de l'écriture se concentre sur l'acquisition de compétences en matière de planification et de composition d'invites, d'évaluation critique des résultats de l'IA générative et de réflexion complexe, ainsi que sur la co-écriture basée sur les grandes lignes de l'IA générative.

Remarques finales

Dans la perspective d'une approche centrée sur l'homme, les outils d'IA devraient être conçus pour étendre ou accroître les capacités intellectuelles et les compétences sociales de l'homme, et non pour les affaiblir, entrer en conflit avec elles ou les usurper. On attend depuis longtemps que les outils d'IA puissent être davantage intégrés en tant qu'éléments constitutifs des outils à disposition des êtres humains afin de soutenir l'analyse et l'action pour des avenir plus inclusifs et plus durables.

Pour que l'IA constitue un élément fiable de la collaboration entre l'homme et la machine – aux niveaux individuel, institutionnel et systémique, une approche centrée sur l'humain, qui s'appuie sur la *Recommandation de l'UNESCO sur l'éthique de l'IA* publiée en 2021, doit être précisée et mise en œuvre en fonction des caractéristiques spécifiques des technologies émergentes telles que l'IA générative. Ce n'est qu'ainsi que nous pourrions garantir que la *GenAI* deviendra un outil fiable pour les chercheurs, les enseignants et les apprenants.

La *GenAI* devrait être utilisée au service de l'éducation et de la recherche ; nous devons cependant tous être conscients que l'IA générative est également susceptible de modifier les systèmes établis et leurs fondements dans ces domaines. La transformation de l'éducation et de la recherche qui sera déclenchée par la *GenAI*, si elle se produit, doit être rigoureusement examinée et dirigée par une approche centrée sur l'Homme. Ce n'est qu'ainsi que nous pourrions garantir que les potentiels des technologies utilisées dans l'éducation, et en particulier de l'IA, renforcent les capacités humaines à construire pour tous des avenir numériques inclusifs.

Références

- Administration du cyberspace de la Chine. 2023a. 国家互联网信息办公室关于《生成式人工智能服务管理办法（征求意见稿）》公开征求意见的通知 [Notice of the Cyberspace Administration of China on Public Comments on the 'Administrative Measures for Generative Artificial Intelligence Services (Draft for Comment)']. Administration du cyberspace de la Chine (ACC), Beijing (en chinois). Disponible à : http://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm (dernière consultation : 19 juillet 2023).
- . 2023b. 生成式人工智能服务管理暂行办法 [Interim Measures for the Management of Generative Artificial Intelligence Services]. Administration du cyberspace de la Chine (ACC), Beijing (en chinois). Disponible à : http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm (dernière consultation : 19 juillet 2023).
- Anders, B. A. 2023. *Is using ChatGPT cheating, plagiarism, both, neither, or forward thinking?* Cambridge, Cell Press. Disponible à : <https://doi.org/10.1016/j.patter.2023.100694> (dernière consultation : 23 juin 2023).
- Bass, D. et Metz, R. 2023. *OpenAI's Sam Altman Urges Congress to Regulate Powerful New Technology*. New York, Bloomberg. Disponible à : <https://www.bloomberg.com/news/newsletters/2023-05-17/openai-s-sam-altman-urges-congress-to-regulate-powerful-new-ai-technology> (dernière consultation : 23 juin 2023).
- Bender, E. M., Gebru, T., McMillan-Major, A. et Shmitchell, S. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *FACCT'21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. New York, Association for Computing Machinery. Disponible à : <https://doi.org/10.1145/3442188.3445922> (dernière consultation : 23 juin 2023).
- Bommasani, R. et al. 2021. *On the Opportunities and Risks of Foundation Models*. Stanford, Stanford University. Disponible à : <https://crfm.stanford.edu/report.html> (dernière consultation : 23 juin 2023).
- Bove, T. 2023. *Big tech is making big AI promises in earnings calls as ChatGPT disrupts the industry: 'You're going to see a lot from us in the coming few months'*. New York, Fortune. Disponible à : <https://fortune.com/2023/02/03/google-meta-apple-ai-promises-chatgpt-earnings> (dernière consultation : 3 juillet 2023).
- Chayka, K. 2023. *My A.I. Writing Report*. New York, The New Yorker. Disponible à : <https://www.newyorker.com/culture/infinite-scroll/my-ai-writing-robot> (dernière consultation : 1er août 2023).
- Chen, L., Zaharia, M. et Zou, J. 2023. *How Is ChatGPT's Behavior Changing over Time?* Ithaca, arXiv. Disponible à : <https://arxiv.org/pdf/2307.09009> (dernière consultation : 31 juillet 2023).
- Commission européenne. 2021. *La loi européenne sur l'intelligence artificielle*. Bruxelles, Commission européenne. Disponible à : <https://artificialintelligenceact.eu/fr> (dernière consultation : 23 juin 2023).
- Coscarelli, J. 2023. *An A.I. Hit of Fake 'Drake' and 'The Weeknd' Rattles the Music World*. New York, New York Times. Disponible à : <https://www.nytimes.com/2023/04/19/arts/music/ai-drake-the-weeknd-fake.html> (dernière consultation : 30 août 2023).
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., Carter, L., Chowdhury, S., Crick, T., Cunningham, S. W., Davies, G. H., Davison, R. M., Dé, R., Dennehy, D., Duan, Y., Dubey, R., Dwivedi, R., Edwards, J. S., Flavián, C., Gauld, R., Grover, V., Hu, M.-C., Janssen, M., Jones, P., Junglas, I., Khorana, S., Kraus, S., Larsen, K. R., Latreille, P., Laumer, S., Malik, F. T., Mardani, A., Mariani, M., Mithas, S., Mogaji, E., Horn Nord, J., O'Connor, S., Okumus, F., Pagani, M., Pandey, N., Papagiannidis, S., Pappas, I. O., Pathak, N., Pries-Heje, J., Raman, R., Rana, N. P., Rehm, S.-V., Ribeiro-Navarrete, S., Richter, A., Rowe, F., Sarker, S., Stahl, B. C., Tiwari, M. K., van der Aalst, W., Venkatesh, V., Viglia, G., Wade, M., Walton, P., Wirtz, J. et Wright, R. 2023. Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, Vol. 71. Amsterdam, Elsevier, p. 102642. Disponible à : <https://doi.org/10.1016/j.ijinfomgt.2023.102642> (dernière consultation : 25 août 2023).
- E2Analyst. 2023. *GPT-4: Everything you want to know about OpenAI's new AI model*. San Francisco, Medium. Disponible à : <https://medium.com/predict/gpt-4-everything-you-want-to-know-about-openais-new-ai-model-a5977b42e495> (dernière consultation : 1er août 2023).

Federal Trade Commission. 1998. *Children's Online Privacy Protection Act of 1998*. Washington DC, Federal Trade Commission. Disponible à : <https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa> (dernière consultation : 4 septembre 2023).

Giannini, S. 2023. *L'IA générative et le futur de l'éducation*. Paris, UNESCO. Disponible à : https://unesdoc.unesco.org/ark:/48223/pf0000385877_fre (dernière consultation : 29 août 2023).

Google. 2023a. *Recommendations for Regulating AI*. Mountain View, Google. Disponible à : <https://ai.google/static/documents/recommendations-for-regulating-ai.pdf> (dernière consultation : 23 juin 2023).

———. 2023b. *PaLM 2 Technical Report*. Mountain View, Google. Disponible à : <https://doi.org/10.48550/arXiv.2305.10403> (dernière consultation : 20 juillet 2023).

Lin, B. 2023. *AI Is Generating Security Risks Faster Than Companies Can Keep Up*. New York, The Wall Street Journal. Disponible à : <https://www.wsj.com/articles/ai-is-generating-security-risks-faster-than-companies-can-keep-up-a2bdeedd4> (dernière consultation : 25 août 2023).

Marcus, G. 2022. Hoping for the Best as AI Evolves. *Communications of the ACM*, Vol. 66, No. 4. New York, Association for Computing Machinery. Disponible à : <https://doi.org/10.1145/3583078> (dernière consultation : 23 juin 2023).

Marwala, T. 2023. *Algorithm Bias - Synthetic Data Should Be Option of Last Resort When Training AI Systems*. Université des Nations Unies. Disponible à : <https://unu.edu/article/algorithm-bias-synthetic-data-should-be-option-last-resort-when-training-ai-systems> (dernière consultation : 31 juillet 2023).

Metz, C. 2021. *Who Is Making Sure the A.I. Machines Aren't Racist?* New York, The New York Times. Disponible à : <https://www.nytimes.com/2021/03/15/technology/artificial-intelligence-google-bias.html> (dernière consultation : 23 juin 2023).

Murphy Kelly, S. 2023. *Microsoft is bringing ChatGPT technology to Word, Excel and Outlook*. Atlanta, CNN. Disponible à : <https://edition.cnn.com/2023/03/16/tech/openai-gpt-microsoft-365/index.html> (dernière consultation : 25 août 2023).

Nazaretsky, T., Cukurova, M. et Alexandron, G. 2022a. An Instrument for Measuring Teachers' Trust in AI-Based Educational Technology. *LAK22: 12th International Learning Analytics and Knowledge Conference*. Vancouver, Association for Computing Machinery, pp. 55-66.

Nazaretsky, T., Ariely, M., Cukurova, M. et Alexandron, G. 2022b. Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology*, Vol. 53, No. 4. Hoboken, NJ, Wiley, pp. 914-931. Disponible à : <https://doi.org/10.1111/bjet.13232> (dernière consultation : 1er août 2023).

Ocampo, Y. 2023. *Singapore Unveils AI Government Cloud Cluster*. Singapour, OpenGov Asia. Disponible à : <https://opengovasia.com/singapore-unveils-ai-government-cloud-cluster> (dernière consultation : 25 août 2023).

OpenAI. 2018. *AI and compute*. San Francisco, OpenAI. Disponible à : <https://openai.com/research/ai-and-compute> (dernière consultation : 23 juin 2023).

———. 2023. *Educator considerations for ChatGPT*. San Francisco, OpenAI. Disponible à : <https://platform.openai.com/docs/chatgpt-education> (dernière consultation : 23 juin 2023).

Popli, N. 2023. *The AI Job That Pays Up to \$335K—and You Don't Need a Computer Engineering Background*. New York, TIME USA. Disponible à : <https://time.com/6272103/ai-prompt-engineer-job> (dernière consultation : 23 juin 2023).

Roose, K. 2022. *An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy*. New York, The New York Times. Disponible à : <https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html> (dernière consultation : 23 juin 2023).

Russell Group, 2023. *Russell Group principles on the use of generative AI tools in education*. Cambridge, Russell Group. Disponible à : https://russellgroup.ac.uk/media/6137/rg_ai_principles-final.pdf (dernière consultation : 25 août 2023).

Stanford University. 2019. *Artificial Intelligence Index Report*. Stanford, Stanford University. Disponible à : <https://hai.stanford.edu/ai-index-2019> (dernière consultation : 23 juin 2023).

———. 2023. *Artificial Intelligence Index Report*. Stanford, Stanford University. Disponible à : <https://hai.stanford.edu/research/ai-index-2023> (dernière consultation : 23 juin 2023).

The Verge. 2023a. *OpenAI co-founder on company's past approach to openly sharing research: 'We were wrong'*. Washington DC, Vox Media. Disponible à : <https://www.theverge.com/2023/3/15/23640180/openai-gpt-4-launch-closed-research-ilya-sutskever-interview> (dernière consultation : 1er août 2023).

-----, 2023b. *OpenAI CEO Sam Altman on GPT-4: 'people are begging to be disappointed and they will be'*. Washington DC, Vox Media. Disponible à : <https://www.theverge.com/23560328/openai-gpt-4-rumor-release-date-sam-altman-interview> (dernière consultation : 1er août 2023).

Tlili, A., Shehata, B., Agyemang Adarkwah, M., Bozkurt, A., Hickey, D. T., Huang, R. et Agyemang, B. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, Vol. 10, No. 15. Berlin, Springer. Disponible à : <https://doi.org/10.1186/s40561-023-00237-x> (dernière consultation : 23 juin 2023).

UNESCO. 2019. *Consensus de Beijing sur l'intelligence artificielle et l'éducation*. Paris, UNESCO. Disponible à : <https://unesdoc.unesco.org/ark:/48223/pf0000368303> (dernière consultation : 3 juillet 2023).

-----, 2022a. *Recommandation sur l'éthique de l'intelligence artificielle*. Paris, UNESCO. Disponible à : https://unesdoc.unesco.org/ark:/48223/pf0000381137_fre (dernière consultation : 3 juillet 2023).

-----, 2022b. *IA et éducation: guide pour les décideurs politiques*. Paris, UNESCO. Disponible à : <https://unesdoc.unesco.org/ark:/48223/pf0000380006> (dernière consultation : 23 juin 2023).

-----, 2022c. *Enseigner l'intelligence artificielle au primaire et au secondaire : une cartographie des programmes validés par les gouvernements*. Paris, UNESCO. Disponible à : https://unesdoc.unesco.org/ark:/48223/pf0000380602_fre (dernière consultation : 20 juillet 2023).

-----, 2022d. *Recommandations pour l'intégration des TIC dans les politiques éducatives*. Paris, UNESCO. Disponible à : <https://unesdoc.unesco.org/ark:/48223/pf0000386039> (dernière consultation : 31 juillet 2023).

-----, 2023a. *Intelligence artificielle : l'UNESCO appelle les gouvernements à mettre en œuvre sans délai le cadre éthique mondial*. Paris, UNESCO. Disponible à : <https://www.unesco.org/fr/articles/intelligence-artificielle-lunesco-appelle-les-gouvernements-mettre-en-oeuvre-sans-delai-le-cadre> (dernière consultation : 3 juillet 2023).

-----, 2023b. *Mapping and analysis of governmental strategies for regulating and facilitating the creative use of GenAI*. Non publié.

-----, 2023c. *Survey for the governmental use of AI as a public good for education*. Non publié (Document remis à l'UNESCO).

-----, 2023. *Rapport mondial de suivi sur l'éducation, 2023: les technologies dans l'éducation : qui est aux commandes?* Paris, Global Education Monitoring Report Team. Disponible à : <https://unesdoc.unesco.org/ark:/48223/pf0000388893> (dernière consultation : 25 août 2023).

-----, 2023. *ChatGPT and Artificial Intelligence in Higher Education: Quick start guide*. Caracas, UNESCO International Institute for Higher Education in Latin America and the Caribbean. Disponible à : <https://unesdoc.unesco.org/ark:/48223/pf0000385146> (dernière consultation : 25 août 2023).

Union européenne. 2016. *Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données)*. Bruxelles, Journal officiel de l'Union européenne. Disponible à : <http://data.europa.eu/eli/reg/2016/679/oj> (dernière consultation : 23 juin 2023).

US Copyright Office. 2023. *Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence*. *Federal Register*, Vol. 88, No. 51. Washington DC, United States (U.S.) Copyright Office, Library of Congress, pp. 16190-16194. Disponible à : <https://www.federalregister.gov/d/2023-05321> (dernière consultation : 3 juillet 2023).

Notes

- 1 Les modèles de *GenAI* ont été mis à la disposition des chercheurs et d'autres protagonistes bien avant la sortie de ChatGPT. Par exemple, en 2015, Google a publié ce qu'il a appelé "DeepDream" (<https://en.wikipedia.org/wiki/DeepDream>).
- 2 Voir <https://chat.openai.com>
- 3 Pour une explication des techniques et technologies de l'IA et de leur relation, voir UNESCO, 2022*b*, pp. 8-10.
- 4 Notons que, dans la mesure où la *GenAI* est encore relativement nouvelle, les entreprises utilisent souvent ces termes de manière différente et emploient parfois des mots différents pour désigner la même chose.
- 5 Il est à craindre que les données utilisées pour entraîner les futures itérations de l'OpenAI GPT comprennent des quantités substantielles de texte généré par des versions antérieures de GPT. Cette boucle autoréférentielle pourrait contaminer les données utilisées pour l'entraînement et compromettre ainsi les capacités des futurs modèles GPT.
- 6 NB OpenAI, l'entreprise qui a développé les GPT de ce tableau, n'a pas publié d'informations détaillées sur GPT-4 (The Verge, 2023*a*). Le nombre de paramètres a été démenti, de fait, par le PDG d'OpenAI (The Verge, 2023*b*), mais les chiffres fournis ici ont été rapportés par un certain nombre d'organismes (par exemple, voir E2Analyst, 2023). En tout état de cause, le principal enseignement à tirer est que GPT-4 est construit sur un ensemble de données massivement plus important et utilise un nombre de paramètres bien supérieur à GPT-3.
- 7 Voir <https://crfm.stanford.edu/2023/03/13/alpaca.html>
- 8 Voir <https://bard.google.com>
- 9 Voir <https://writesonic.com/chat>
- 10 Voir <https://yiyan.baidu.com/welcome>
- 11 Voir <https://huggingface.co/chat>
- 12 Voir <https://www.jasper.ai>
- 13 Voir <https://ai.facebook.com/blog/large-language-model-llama-meta-ai>
- 14 Voir <https://open-assistant.io>
- 15 Voir <https://www.alizila.com/alibaba-cloud-debuts-generative-ai-model-for-corporate-users>
- 16 Voir <https://you.com>
- 17 Voir <https://www.chatpdf.com>
- 18 Voir <https://elicit.org>
- 19 Voir <https://www.perplexity.ai>
- 20 Voir <https://tools.zmo.ai/webchatgpt>
- 21 Voir <https://www.compose.ai>
- 22 Voir <https://www.teamsmart.ai>
- 23 Voir <https://wiseone.io>
- 24 <https://www.microsoft.com/en-us/bing>
- 25 Voir <https://www.crayon.com>
- 26 Voir <https://openai.com/product/dall-e-2>
- 27 Voir <https://dream.ai/create>
- 28 Voir <https://www.fotor.com/features/ai-image-generator>
- 29 Voir <https://www.midjourney.com>
- 30 Voir <https://creator.nightcafe.studio>
- 31 Voir <https://writesonic.com/photosonic-ai-art-generator>
- 32 Voir <https://elai.io>
- 33 Voir <https://www.gliacloud.com>
- 34 Voir <https://pictory.ai>

- 35 Voir <https://runwayml.com>
- 36 Voir <https://www.aiva.ai>
- 37 Voir <https://boomy.com>
- 38 Voir <https://soundraw.io>
- 39 Voir <https://www.voicemod.net/text-to-song>
- 40 Voir <https://openai.com/research/gpt-4>
- 41 Voir <https://www.educhat.top> et <https://www.mathgpt.com>
- 42 Voir <https://www.educhat.top>
- 43 Voir <https://www.mathgpt.com>
- 44 Il existe quelques exceptions, comme Hugging Face, groupe qui se consacre au développement de l'IA en code source ouvert.
- 45 Voir, par exemple, les appels de Google (2023a) et d'OpenAI (Bass et Metz, 2023).
- 46 Pour un projet de réglementation de l'IA, voir le projet de loi sur l'IA de la Commission européenne (2021).
- 47 L'étude s'appuie sur des données recueillies dans le cadre d'une enquête de l'UNESCO distribuée à ses 193 États membres au sujet de l'utilisation de l'IA dans l'éducation par les gouvernements (UNESCO, 2023c), de l'Observatoire des politiques d'IA de l'OCDE et du rapport AI Index de l'Université de Stanford (Université de Stanford, 2023), ainsi que sur des informations de première main obtenues auprès d'un groupe d'experts internationaux.
- 48 Voir <https://unctad.org/page/data-protection-and-privacy-legislation-worldwide>
- 49 D'après la cartographie, en avril 2023, les pays suivants avaient publié des stratégies nationales sur l'IA : Allemagne, Argentine, Australie, Autriche, Belgique, Bénin, Brésil, Canada, Bulgarie, Chili, Chine, Chypre, Colombie, Danemark, Égypte, Estonie, Finlande, France, Hongrie, Inde, Indonésie, Irlande, Islande, Italie, Japon, Jordanie, Lettonie, Lituanie, Luxembourg, Malaisie, Malte, Maurice, Mexique, Norvège, Nouvelle-Zélande, Oman, Pays-Bas, Pérou, Pologne, Portugal, Philippines, Qatar, République de Corée, République tchèque, Roumanie, Fédération de Russie, Arabie Saoudite, Serbie, Singapour, Slovénie, Espagne, Suède, Thaïlande, Tunisie, Turquie, Émirats arabes unis, Royaume-Uni, États-Unis, Uruguay et Vietnam. Par ailleurs, certains pays ont intégré des stratégies en matière d'IA dans des stratégies plus larges en matière de TIC ou de numérique, notamment l'Algérie, le Botswana, le Kazakhstan, le Kenya, la Sierra Leone, la Slovaquie, la Suisse et l'Ouganda.
- 50 Selon un examen rapide de toutes les stratégies nationales en matière d'IA (UNESCO, 2023b), plus de 40 stratégies comportent des sections consacrées à la question de l'éthique.
- 51 Selon un examen rapide de toutes les stratégies nationales en matière d'IA (UNESCO, 2023b), environ 45 stratégies comportent des sections consacrées à la question de l'éducation.
- 52 Voir <https://openai.com/policies/terms-of-use>
- 53 Dans certains pays, les enseignants ont un assistant dont le rôle est de consacrer du temps à répondre aux questions des élèves et étudiants portant sur les supports de cours. La *GenAI* pourrait être utilisée pour créer des « jumeaux génératifs » à ces assistants, pour apporter un soutien aux apprenants et aux autres enseignants, ce qui cependant peut aussi présenter des aspects négatifs (par exemple en ce qui concerne les relations sociales en classe).



unesco

Organisation
des Nations Unies
pour l'éducation,
la science et la culture

Orientations pour l'intelligence artificielle généralive dans l'éducation et la recherche

Ces recommandations ont pour objectif de soutenir l'élaboration de réglementations, de politiques et de programmes visant à développer les capacités humaines, dans le but de garantir que l'intelligence artificielle généralive (*GenAI*) devienne un outil qui profite réellement aux enseignants, aux apprenants et aux chercheurs et leur donne les moyens d'agir. Il explique les techniques d'IA utilisées par la *GenAI* et dresse une liste des modèles de GPT qui sont mis à la disposition auprès du public, en particulier ceux qui font l'objet d'une licence à source ouverte. Il ouvre également une discussion sur l'émergence de l'EdGPT - les modèles *GenAI* qui sont entraînés avec des données spécifiques pour se mettre au service des objectifs dans le domaine de l'éducation. En outre, il résume certaines des principales controverses formées autour de l'IA généralive, de l'aggravation de la pauvreté numérique à l'homogénéisation des opinions, en passant par les trucages (deepfakes) et les questions liées aux droits d'auteur. Fondé sur une vision humaniste, le présent document propose des étapes-clefs pour la réglementation des outils de la *GenAI*, notamment en rendant obligatoire la protection de la confidentialité des données et en fixant une limite d'âge pour les conversations autonomes avec les plateformes de la *GenAI*. Pour accompagner l'utilisation pertinente des outils dans l'éducation et la recherche, ce guide propose une approche holistique et adaptée à tout âge pour les processus de validation éthique et de conception pédagogique.



9 789232 003126

