



MINISTÈRE
DE L'ÉDUCATION
NATIONALE
ET DE LA JEUNESSE

*Liberté
Égalité
Fraternité*

Intelligence artificielle et éducation

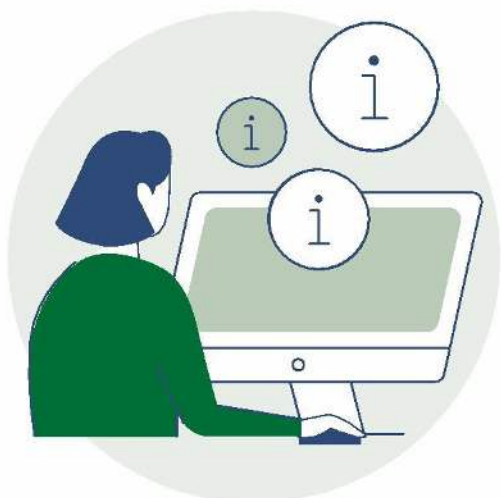
Apports de la recherche
et enjeux pour les politiques
publiques

Avril 2023



État des lieux sur les apports de la recherche et les lignes directrices des institutions internationales sur l'intelligence artificielle et l'éducation :

définitions et approches, politiques publiques, enjeux éthiques, applications, tournant des IA génératives.



Current situation on the Research Contributions Guidelines of international institutions on artificial intelligence and education:

definitions and proposals, public policies, ethical issues, applications, turning point of Generative AI.



Table des matières

Résumé.....	4
À retenir.....	5
1. L'intelligence artificielle (IA) : une diversité de définitions et d'approches.....	6
2. Quels enjeux pour les politiques publiques ?...11	11
3. Quels enjeux éthiques ?	17
4. Quels domaines d'application ?	19
5. Quelles pistes de travail pour former à l'IA et l'enseigner ?.....	23
6. Le tournant des IA génératives et des grands modèles de langage.....	26
Références.....	35

Résumé

Ce document, réalisé par la Direction du numérique pour l'éducation/ ministère de l'Éducation nationale et de la Jeunesse, propose un état des lieux (Avril 2023) sur les apports de la recherche et les lignes directrices des institutions internationales sur l'intelligence artificielle (IA) et l'éducation : diversité des définitions et des approches, enjeux pour les politiques publiques, enjeux éthiques, domaines d'application, pistes de travail pour former et enseigner, perspectives avec le tournant actuel des IA génératives et des grands modèles de langage.

Abstract

This document, produced by the Digital Department for Education (Ministry of National Education and Youth), offers an inventory (April 2023) of the contributions of research and the guidelines of international institutions on the artificial intelligence (AI) and education: diversity of definitions and approaches, issues for public policies, ethical issues, fields of application, avenues for training and teaching, perspectives with the current turn of generative AI and large language models.

Resumo

Este documento, producido por el Departamento de Educación Digital (Ministerio de Educación Nacional y Juventud), ofrece un inventario (Abril de 2023) de las contribuciones de la investigación y las directrices de las instituciones internacionales sobre la inteligencia artificial (IA) y la educación: diversidad de definiciones y enfoques, cuestiones para las políticas públicas, cuestiones éticas, campos de aplicación, vías de formación y enseñanza, perspectivas con el giro actual de las IA generativas y los grandes modelos de lenguaje.

À retenir

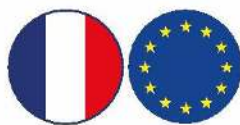
L'intelligence artificielle : une diversité de définitions et d'approches

« De façon surprenante, le terme récemment retenu pour qualifier ces stupéfiantes prouesses calculatoires est celui d'intelligence artificielle (IA). Le retour sur le devant de la scène de ce vocable forgé en 1956 par John McCarthy constitue une intéressante énigme pour l'histoire des sciences et des techniques.»
(Cardon et al., 2018)



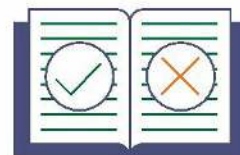
Quels enjeux pour les politiques publiques ?

« La stratégie de l'IA en éducation doit proposer des solutions aux professionnels de l'enseignement et de la formation, aider aux apprentissages et aux décisions sans jamais se substituer aux personnes.»
(Gouvernement - Direction générale des entreprises, 2023)



Quels enjeux éthiques ?

Quatre considérations clés à intégrer proposées par la Commission européenne : action humaine et contrôle humain, équité, humanité, justification des choix, transparence et explicabilité.
D'après (Directorate-General for Education, Youth, Sport and Culture & European Commission, 2022)



Les domaines d'application pour l'éducation

« L'intelligence artificielle joue un rôle important et croissant dans l'éducation. Un cas important est celui des systèmes d'enseignement personnalisés qui sont déjà bien établis, avec des preuves croissantes de leur efficacité pour améliorer l'apprentissage.»
(Khosravi et al., 2022)



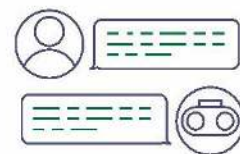
Quelles pistes de travail pour former à l'IA et l'enseigner ?

Une proposition de parcours en cinq piliers
1. Incertitude et caractère aléatoire
2. Codage et calcul
3. Sensibilisation aux données
4. Pensée critique
5. Humanisme post-IA
D'après (de la Higuera & Bocquet, 2020)



Le tournant des IA génératives et des grands modèles de langage

« Au cours des deux dernières années, une pléthore de grands modèles génératifs tels que ChatGPT ou Stable Diffusion ont été publiés. Ils sont capables d'effectuer des tâches telles qu'un système général de questions-réponses ou la création automatique d'images artistiques qui révolutionnent plusieurs secteurs.»
(Gozalo-Brizuela & Garrido-Merchan, 2023)



1 L'intelligence artificielle (IA) : une diversité de définitions et d'approches

1.1 Origines et actes fondateurs

En remontant à ses origines, en tant que nouvelle science, l'IA «reposait sur la conjecture selon laquelle toutes les **facultés cognitives**, en particulier le raisonnement, le calcul, la perception, la mémorisation, voire la découverte scientifique ou la créativité artistique, pourraient être décrites avec une précision telle qu'il devrait être **possible de les reproduire à l'aide d'un ordinateur**.» D'après (Ganascia, 2022)

« Outre un article majeur sur la morphogénèse en 1952, [les] deux papiers les plus fameux [de Turing] dans le monde de l'informatique sont « *On Computable Numbers, With an Application to the Entscheidungsproblem* » en 1937, où il pose les bases de la **machine de Turing** et de la calculabilité, et « *Computing Machinery and Intelligence* » en 1950, où il pose la **question de l'intelligence pour les machines** et énonce le **test de Turing**. » (Le Blanc, 2014)

« Une définition classique précise que l'IA est un domaine de recherche qui développe des technologies capables de faire des choses qui exigeraient de l'intelligence si elles étaient faites par des humains (Minsky, 1969). Cette approche trouve son origine chez **Turing**, ayant proposé que si une **simulation d'être humain intelligent** ne peut être distinguée d'une personne réelle, les questions relatives à l'intelligence deviennent sans objet (Turing, 1950). De nombreux spécialistes des sciences cognitives et certains chercheurs et philosophes spécialisés dans l'IA ont adopté un point de vue plus ferme, affirmant que la recherche sur l'IA peut révéler comment fonctionne l'esprit humain (Gardner, 1985). » (Holmes et al., 2022) traduction par (Bocquet, 2023)

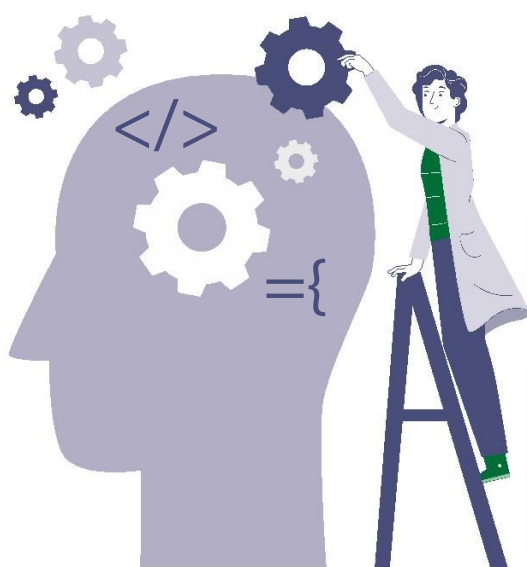
« S'il faut un début à l'histoire, commençons par la **conférence de Dartmouth** où surgit le terme désormais consacré d'« intelligence artificielle ». Elle est organisée durant l'été **1956** au Dartmouth College, près de Hanover (New Hampshire), par **Marvin Minsky et John McCarthy**, les deux pionniers. » (Le Cun et al., 2019)

« De façon surprenante, le terme récemment retenu pour qualifier ces stupéfiantes prouesses calculatoires est celui d'intelligence artificielle (IA). Le retour sur le devant de la scène de ce **vocabulaire forgé en 1956 par John McCarthy** constitue une intéressante énigme pour l'histoire des sciences et des techniques. La plupart des observateurs rigoureux soulignent en effet que c'est dans le seul domaine des méthodes d'apprentissage et, notamment, de l'apprentissage profond (*deep learning*), que des progrès sensibles de la prédiction calculée ont lieu actuellement. Or l'appartenance de ces techniques au champ de l'IA n'est pas toujours allée de soi. (...)

Les **réseaux de neurones** trouvent leur origine dans l'histoire pionnière de l'informatique et de la première cybernétique. Bien que l'étiquette soit postérieure, celle-ci peut en effet être dite « connexionniste » et ne cessera de se référer à la proposition de modéliser mathématiquement un réseau de neurones faite par le neurophysiologiste **Warren McCulloch** et le logicien **Walter Pitts en 1943**. Cet article fondateur continue, jusque dans les citations actuelles des articles de *deep learning*, à être donné comme le point de départ de l'aventure connexionniste.» (Cardon et al., 2018)

1.2 Choix humains, modélisation et automatisatisation de tâches cognitives

« L'expression « intelligence artificielle », créée dans les années 1950, désigne le **domaine de recherche qui étudie les mécanismes de l'intelligence** en les modélisant avec des algorithmes et en les expérimentant avec des machines. Ces mécanismes incluent par exemple la faculté de trouver automatiquement des solutions à des problèmes, qui peut mettre en œuvre des capacités de planification, de prédiction, de contrôle, de mémorisation ou d'apprentissage. Par extension, le terme « intelligence artificielle » est souvent utilisé pour désigner des **algorithmes** [enchaînement d'actions élémentaires pour résoudre un problème] **simulant ou ayant des points communs avec certaines des capacités d'intelligence des êtres vivants.** » (Inria, 2020)



L'IA désigne « un **système automatisé** qui, pour un ensemble donné d'objectifs définis par l'humain, est **en mesure d'établir des prévisions, de formuler des recommandations, ou de prendre des décisions** influant sur des environnements réels ou virtuels. Les systèmes d'IA interagissent avec nous et influent de façon directe ou indirecte sur notre environnement. Ils semblent souvent fonctionner de façon autonome et peuvent adapter leur comportement en fonction du contexte. » D'après (UNICEF, 2021) repris par (Holmes & Tuomi, 2022) traduction par (Bocquet, 2023)

C'est aussi « un **domaine interdisciplinaire à la fois théorique et pratique** qui allie les technologies numériques, mathématiques, statistiques et algorithmiques. » (Grand lexique français de l'Intelligence artificielle, 2021)

« Nous pouvons définir l'intelligence artificielle (IA) comme des **systèmes informatiques capables de s'engager dans des processus de type humain** tels que l'apprentissage, l'adaptation, la synthèse, l'autocorrection et l'utilisation de données pour des tâches de traitement complexe. » (Popenici & Kerr, 2017) « [Cette définition] permet de **dépasser l'opposition entre l'intelligence humaine et l'IA et d'envisager des interactions complexes entre les deux.** » (Romero et al., 2023)

1.3 Les types d'IA

On distingue plusieurs types d'IA :

Selon leur technologie :

« Une **IA basée sur les données**, ou apprentissage automatique (machine learning en anglais) et une autre IA basée sur les connaissances ou **IA symbolique**. Un troisième modèle conceptuel, l'**IA hybride**, combine les approches basées sur les données et les connaissances avec la cognition humaine. » [Traduction] D'après (Holmes et al., 2022) traduction par (Bocquet, 2023)

Selon leur portée :

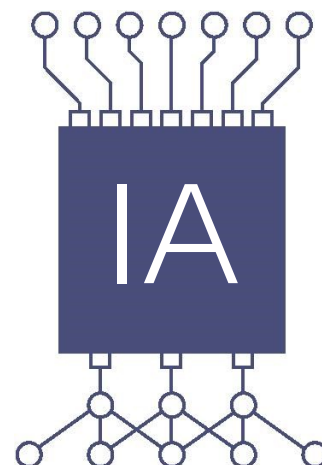
« On distingue l'**IA faible** de l'**IA forte** selon la portée d'application de ces programmes informatiques. L'intelligence artificielle faible (*weak AI*) est une intelligence artificielle qui se concentre sur une tâche précise alors que l'IA forte (*strong AI*) ou l'intelligence artificielle générale (AGI: *artificial general intelligence*) s'applique à tout problème. Tous les systèmes d'IA existants sont considérés actuellement comme des intelligences artificielles faibles. » (*Grand lexique français de l'Intelligence artificielle*, 2021)

« Intelligence artificielle complète ou forte :

Une intelligence artificielle qui serait capable de copier les aptitudes humaines (apprendre, comprendre, appréhender, raisonner, prendre des décisions, avoir une conscience, des émotions...).

À ce jour, l'intelligence artificielle forte n'existe pas, c'est une croyance.

Intelligence artificielle faible : C'est l'intelligence artificielle qu'on connaît aujourd'hui : c'est un algorithme "apprenant", en adaptant ses paramètres à des données d'apprentissage, non pourvu de capacités mentales et cognitives, mais capable d'effectuer avec beaucoup plus d'efficacité, parfois plus que l'être humain, une tâche précise. » (Inria, 2020)



IA classique et apprentissage automatique :

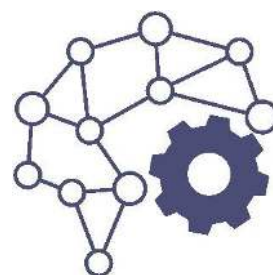
IA classique : « Une grande partie de l'IA des débuts ou « classique », connue sous le nom d'« **IA symbolique** », d'« IA à base de règles » ou de « bonne vieille IA » (en anglais *GOFAI*), implique l'écriture de séquences de SI... ALORS... et d'autres règles de logique conditionnelle, c'est-à-dire des étapes que l'ordinateur suit pour accomplir une tâche. Au fil du temps, des « **systèmes experts** » d'IA basés sur des règles ont été créés et développés pour un large éventail d'applications, telles que les diagnostics médicaux, les notations de crédit et la production. Les systèmes experts sont basés sur une approche connue sous le nom d'« ingénierie des connaissances », qui consiste à obtenir et à modéliser les connaissances des experts dans un domaine donné, ce qui représente une tâche exigeante en ressources et qui n'est pas sans complications. (...) »

Apprentissage automatique (AA) : De nombreuses avancées récentes en matière d'IA – notamment le traitement du langage naturel, la reconnaissance faciale et les voitures autonomes – ont été rendues possibles par les progrès des approches

informatiques basées sur l'apprentissage automatique. Plutôt que d'utiliser des règles, l'apprentissage **automatique (AA, machine learning, ML, en anglais) analyse de grandes quantités de données pour identifier des modèles** et construire un modèle qui est ensuite utilisé pour extrapoler les données futures : c'est en ce sens que les algorithmes, plutôt que d'être préprogrammés, sont dits « apprenants ».

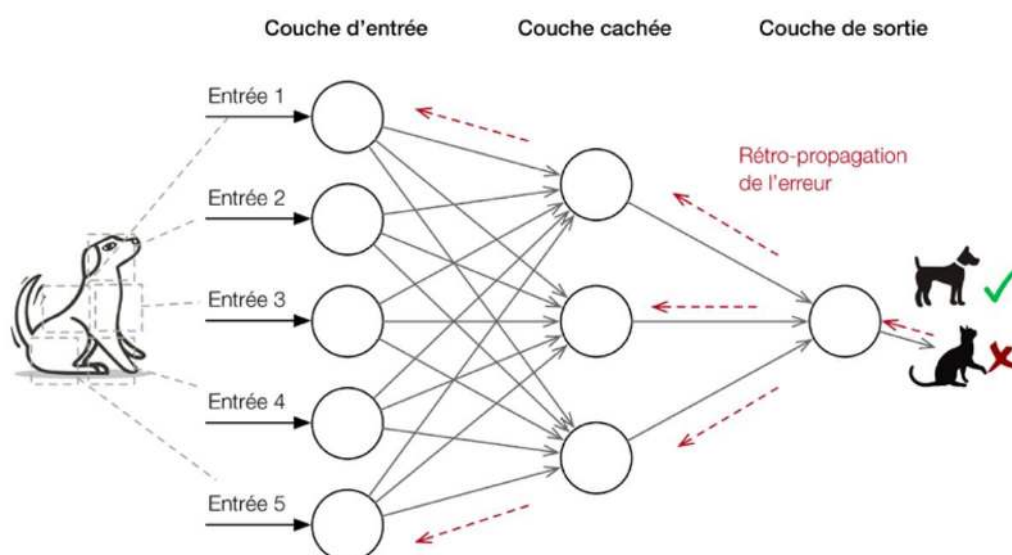
Il existe **trois approches principales de l'AA : supervisée, non supervisée, et par renforcement**. L'apprentissage supervisé implique que les données ont déjà été identifiées et décrites, par exemple des milliers de photographies de personnes préalablement « étiquetées » par des humains. L'apprentissage supervisé relie les données aux étiquettes, afin de construire un modèle qui peut être appliqué à des données similaires, par exemple pour identifier automatiquement des personnes sur de nouvelles photographies. Dans l'apprentissage non supervisé, l'IA dispose d'une quantité de données encore plus importante, mais les données n'ont alors pas été préalablement identifiées, catégorisées ou étiquetées. (...)

Un **réseau neuronal artificiel, ou réseau de neurones artificiel (RNA)**, est une approche de l'IA qui s'inspire de la structure des réseaux neuronaux biologiques – le cerveau des animaux. Les RNA comprennent chacun trois types de couches interconnectées de neurones artificiels : une couche d'entrée, une ou plusieurs couches de calcul intermédiaires cachées, enfin une couche de sortie qui fournit le résultat. (...)



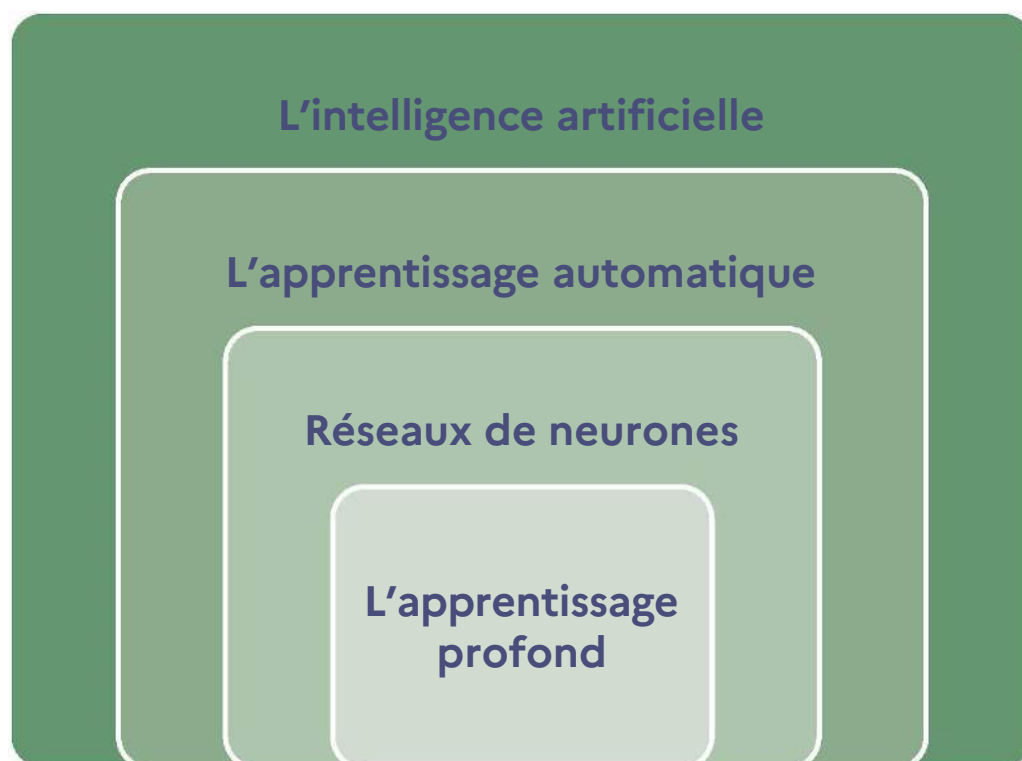
L'apprentissage profond [deep learning] fait référence aux RNA qui comprennent plusieurs couches intermédiaires. C'est cette approche qui est à l'origine de nombreuses applications de l'IA récentes et prometteuses (par exemple dans le traitement du langage naturel, la reconnaissance vocale, la vision par ordinateur, la création d'images, la découverte de médicaments ou la génomique). » (UNESCO, 2021)

Fonctionnement d'un réseau de neurones simple



(Cardon et al., 2018)

**IA, apprentissage automatique, réseau de neurones
et apprentissage profond**



(UNESCO, 2021)

2. Quels enjeux pour les politiques publiques ?

2.1 Le consensus de Beijing

« Le Consensus de Beijing recommande aux gouvernements et aux autres parties prenantes des États membres de l'UNESCO de :

- **Planifier l'IA dans les politiques de l'éducation** pour tirer parti des possibilités et relever les défis qu'apportent les technologies d'IA, adopter des approches engageant le gouvernement tout entier, intersectorielles et multipartites qui permettent aussi de mettre en place des priorités stratégiques locales pour atteindre les objectifs de l'ODD 4 [Objectif de développement durable 4 : 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].
- **Soutenir le développement de nouveaux modèles** rendus possibles par les technologies d'IA pour la fourniture de services d'éducation et de formation là où les avantages l'emportent clairement sur les risques, et utiliser des outils d'IA pour proposer des systèmes d'apprentissage tout au long de la vie qui permettent un apprentissage personnalisé à tout moment, en tout lieu, pour tous.
- **Envisager l'utilisation de données pertinentes**, le cas échéant, afin d'encourager la planification de politiques fondées sur des données probantes.
- Veiller à ce que les technologies d'IA servent à **autonomiser les enseignants au lieu de les remplacer** et développer des programmes appropriés pour le renforcement des capacités afin que les enseignants travaillent aux côtés des systèmes d'IA.
- Préparer la prochaine génération de travailleurs en **l'équipant des valeurs et des compétences nécessaires pour la vie et au travail**, les plus pertinentes à l'ère de l'IA.
- Promouvoir une **utilisation équitable et inclusive** de l'IA, indépendamment du handicap, du statut social ou économique, de l'origine ethnique ou culturelle ou de la situation géographique, en insistant sur l'égalité des sexes, tout en assurant des usages éthiques, transparents et vérifiables des données éducatives.»

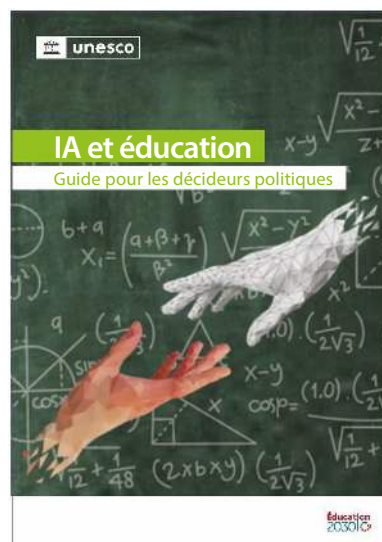
(UNESCO, 2019)



2.2 Les recommandations de l'UNESCO

1. Une vision et des priorités stratégiques à l'échelle systémique

- Définir à l'échelle du système une vision des politiques de l'IA et de l'éducation
 - garantir l'utilisation inclusive et équitable de l'IA dans l'éducation ;
 - tirer au mieux parti de l'IA pour améliorer l'éducation et l'apprentissage ;
 - promouvoir le développement de compétences nécessaires à la vie à l'ère de l'IA, notamment en enseignant le fonctionnement de l'IA et ses implications pour l'humanité ;
 - garantir l'utilisation transparente et contrôlable des données relatives à l'éducation.
- Évaluer l'état de préparation de l'ensemble du système et choisir les priorités stratégiques



2. Principe directeur en matière de politiques d'IA et d'éducation

- Adopter une approche humaniste comme principe général des politiques de l'IA et de l'éducation

3. Planification interdisciplinaire et gouvernance intersectorielle

- Mobiliser l'expertise interdisciplinaire et multipartite pour éclairer la planification des politiques et renforcer la capacité des décideurs
- Mettre en place des mécanismes de gouvernance et de coordination intersectoriels

4. Politiques et réglementations pour une utilisation équitable, inclusive et éthique de l'IA

- Définir des objectifs stratégiques transversaux et planifier des règlements et des programmes, afin de garantir l'utilisation équitable et inclusive de l'IA dans le champ éducatif

5. Plans directeurs pour l'utilisation de l'IA dans la gestion de l'éducation, l'enseignement, l'apprentissage et l'évaluation

- L'utilisation de l'IA pour stimuler et améliorer la gestion de l'éducation et les prestations
- Cultiver une utilisation de l'IA centrée sur l'apprenant pour améliorer l'apprentissage et l'évaluation
- Veiller à ce que l'IA soit utilisée pour renforcer les dynamiques des enseignants
- Planifier l'utilisation de l'IA pour soutenir l'apprentissage tout au long de la vie, quels que soient l'âge, le lieu et le contexte
- Développer des valeurs et des compétences pour la vie et le travail à l'ère de l'IA

6. Essais pilotes, suivi et évaluation, et constitution d'une base de données fiable

- Construire une base de données fiable pour soutenir l'utilisation de l'IA dans l'éducation
- Renforcer la recherche et l'évaluation dans le domaine de l'IA et de l'éducation

7. Encourager les innovations locales en matière d'IA pour l'éducation

- Promouvoir le développement local des technologies de l'information dans le champ éducatif

D'après (UNESCO, 2021) (Extraits)

2.3 Une stratégie nationale

« La stratégie de l'IA en éducation doit **proposer des solutions aux professionnels de l'enseignement et de la formation, aider aux apprentissages et aux décisions sans jamais se substituer aux personnes.** Elle doit explorer des **partenariats d'innovation entre l'École, la Recherche et les technologies de l'éducation**, en particulier pour l'innovation pédagogique, ainsi que l'utilisation contrôlée des données de scolarité ou de formation pour améliorer les politiques publiques. **Elle doit construire des connaissances et des compétences de base chez tous les élèves, les enseignants et les cadres** pour démystifier, comprendre et utiliser de manière raisonnée et éthique les potentialités de l'IA, et susciter des talents en proposant des mentions, spécialités ou parcours IA aux lycéens en amont de l'enseignement supérieur. » (...)

QUELQUES ACTIONS CLÉS

- Des partenariats d'innovation et des challenges IA en éducation pour développer la recherche de solutions grâce aux professionnels de l'éducation nationale, de la filière EdTech et de la recherche
- Des formations à l'IA pour les élèves, les enseignants et des cadres
- Des projets et partenariats IA et l'utilisation massive des données avec la recherche»

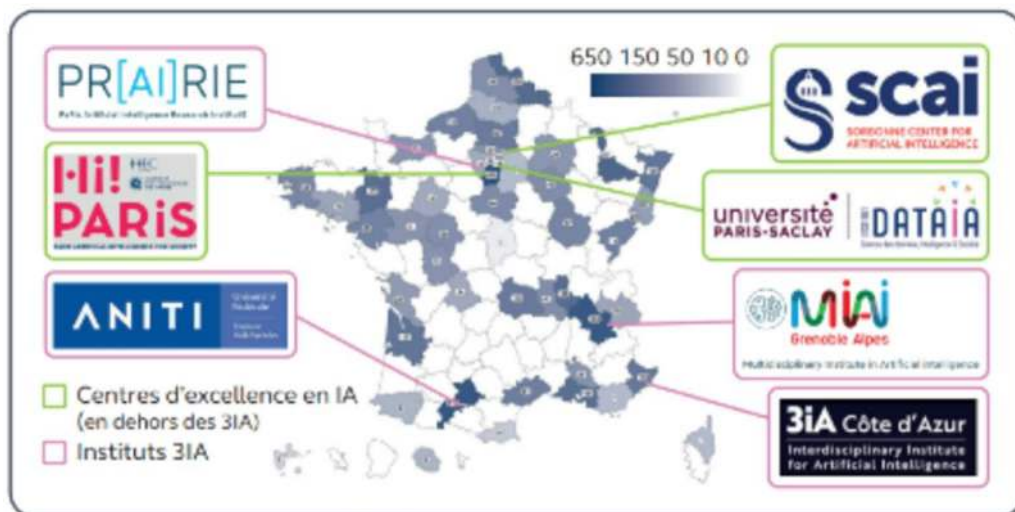
(Gouvernement - Direction générale des Entreprises, 2023)

2.4 Chronologie et implantation géographique

Les étapes d'élaboration de la stratégie nationale en IA



Implantation des instituts thématiques en IA (3IA PR[AI]RIE, MIAI, 3IA Côte d'Azur et ANITI) et des centres d'excellence (SCAI, DATAIA et Hi! PARIS) en regard des zones historiquement actives dans ce domaine



Source : Cour des comptes

Note de lecture : L'activité historique en IA est mesurée à travers la répartition départementale des thèses en IA soutenues entre 1989 et 2019 dans les établissements français d'enseignement supérieur à partir des données ouvertes de theses.fr (ABES). Les thèses sont données par année de soutenance. La détection des thèses en IA est réalisée suivant la méthode sémantique développée par la Cour. Les logos des 3IA et des centres d'excellence en IA sont tirés de leurs sites web officiels.

(Cour des comptes, 2023)

2.5 IA et éducation ouverte : les actions d'un groupe thématique numérique



(DNE-TN2, 2023)

2.6 Les performances de l'IA en lecture et en mathématiques

« [Cette] étude fait partie d'un projet global d'évaluation des capacités de l'IA et de leurs implications pour le travail et l'éducation, dirigé par le Centre de recherche et d'innovation dans l'enseignement (CERI) de l'OCDE. Le **projet sur l'IA et l'avenir des compétences (AIFS)** vise à développer des mesures des capacités de l'IA qui soient compréhensibles, complètes, reproductibles et pertinentes pour les politiques. En utilisant diverses sources d'information sur l'IA, y compris des évaluations d'experts, le projet vise à fournir aux décideurs politiques les connaissances dont ils ont besoin pour élaborer des politiques de l'éducation et du marché du travail tournées vers l'avenir. Le rapport montre des améliorations significatives des capacités de l'IA en lecture depuis 2016, qui reflètent les **progrès du traitement du langage naturel (NLP)** au cours des dernières années. Les capacités de l'IA à résoudre des tâches mathématiques n'ont pas progressé au même rythme. Toutefois, **les experts prévoient que l'augmentation des investissements dans la recherche et le développement de l'IA conduira à des progrès significatifs de l'IA à la fois en lecture et en mathématiques dans les années à venir.** Le rapport démontre également que l'IA pourrait surpasser une grande partie de la population en lecture et en mathématiques. Cela a des conséquences importantes pour l'emploi et l'éducation, car les travailleurs risquent d'être de plus en plus concurrencés par les machines dans ces domaines de compétences à l'avenir. Il souligne également la nécessité de **renforcer les compétences de base de la main-d'œuvre et de la préparer à collaborer avec l'IA dans des domaines clés.** En fournissant un exemple de la manière dont les capacités de l'IA s'améliorent par rapport à deux compétences cognitives clés des humains, cette étude souligne **l'importance d'un suivi périodique et systématique de l'évolution des capacités de l'IA et de leur comparaison avec les compétences humaines. Cela sera utile aux décideurs politiques, aux éducateurs et aux chercheurs qui cherchent à comprendre les implications des avancées technologiques pour l'avenir du travail et de l'éducation.** »

(OCDE, 2023) [Traduction]

3. Quels enjeux éthiques ?

3.1 Des lignes directrices issues des institutions internationales

« Il est essentiel de comprendre que les systèmes d'IA ne relèvent pas de la magie.

Des personnes conçoivent, entraînent et guident l'IA, qu'il s'agisse de celles qui définissent les politiques et stratégies en matière d'IA, des programmeurs de logiciels qui construisent les systèmes d'IA, des individus qui collectent et classent les données utilisées par ces systèmes ou de ceux qui interagissent avec eux. Cela signifie que **tous les acteurs de l'écosystème de développement des technologies de l'IA doivent comprendre les principaux enjeux** qui leur imposent de contribuer à une utilisation responsable de l'IA. Il peut notamment s'agir de **savoir pourquoi et comment un système d'IA a été conçu, par qui et dans quel but.** » (UNICEF, 2021)

Pour une utilisation éthique de l'IA et des données dans l'enseignement, l'apprentissage et l'évaluation, **la commission européenne propose quatre considérations clés** à intégrer :

- **action humaine et contrôle humain** : capacité d'action des acteurs de l'éducation (enseignants, apprenants, chefs d'établissement) ;
- **équité** : impartialité, inclusion, non-discrimination et juste répartition des droits et des responsabilités ;
- **humanité** : considération pour les personnes, leur identité, leur intégrité et leur dignité ;
- **justification des choix, transparence et explicabilité** : utilisation de connaissances, de faits et de données pour justifier les choix collectifs dans l'environnement scolaire, via des modèles participatifs et collaboratifs de prise de décision.

D'après (Directorate-General for Education, Youth, Sport and Culture & European Commission, 2022)

Le Conseil de l'Europe met en avant plusieurs principes pour une utilisation de l'IA en éducation respectueuse des **droits fondamentaux** :

- « *Droit à la dignité humaine.* L'enseignement, l'évaluation et l'accréditation ne doivent pas être délégués à un système d'IA.
- *Droit à l'autonomie.* Les enfants devraient avoir le droit d'éviter d'être profilés individuellement, d'éviter les parcours d'apprentissage dictés, et de protéger leur développement et leur vie future.

- *Le droit d'être entendu.* Les enfants devraient avoir le droit de ne pas s'engager dans l'utilisation d'un système d'IA, sans que cela n'affecte négativement leur éducation.
- *Droit de ne pas souffrir de discrimination.* Tous les enfants devraient avoir la possibilité de bénéficier de l'utilisation des technologies, et pas seulement ceux issus des groupes socio-économiques qui peuvent se le permettre.
- *Droit à la confidentialité et à la protection des données.* Les enfants doivent avoir le droit que leurs données ne soient pas agrégées et utilisées à des fins commerciales sans leur bénéfice direct.
- *Droit à la transparence et à l'explicabilité.* Les enfants et leurs parents doivent pouvoir comprendre et contester toute décision prise par un système IA».

D'après (Holmes et al., 2022)

3.2 La diversité des enjeux éthiques vue par la recherche

« Les enjeux éthiques et critiques que soulève l'IA en éducation sont multiples et ont des origines diverses : utilisation de **données massives**, pouvant avoir un impact sur le **respect de la vie privée** des élèves et du personnel scolaire, production par des entreprises privées plutôt que par des instances scolaires, par la recherche en informatique ou en sciences, technologies, ingénierie et mathématiques plutôt que par les sciences de l'éducation, **biais de représentativité** par manque de diversité au sein des équipes de conception.»

D'après (Collin & Marceau, 2021)

3.3 L'enjeu de l'explicabilité

« Dans le domaine éducatif, deux catégories d'utilisateurs doivent s'approprier l'environnement numérique : les apprenants et les enseignants. **La question de l'appropriation des EIAH [Environnements informatiques pour l'apprentissage humain] par les enseignants est essentielle pour que ces outils soient davantage utilisés dans l'enseignement.** Il faut donc concevoir des systèmes que les enseignants pourront adapter afin qu'ils répondent à leurs besoins. Pour cela, **l'enseignant doit pouvoir comprendre les décisions ou recommandations du système d'IA concernant ses élèves**, pour avoir confiance dans le système et, d'une certaine façon, «faire corps» avec lui pour être capable d'expliquer le comportement conjugué du dispositif tel qu'il a été mis en place. Cela nécessite une représentation explicite des connaissances et des processus de décision. **La question de l'explicabilité doit être prise en compte dès la conception.**»

(Association Française pour l'Intelligence Artificielle (AFIA), 2022)

4. Quels domaines d'application ?

4.1 Catégorisation et taxonomie

Catégorisation proposée par la Commission européenne

Enseignement	Soutien à l'apprenant	Soutien à l'enseignant	Soutien au système éducatif
<ul style="list-style-type: none">• Environnement informatique pour l'apprentissage humain• Systèmes tutoriels fondés sur le dialogue• Applications d'aide à l'apprentissage des langues	<ul style="list-style-type: none">• Environnements d'apprentissage exploratoire• Évaluation formative des écrits• Apprentissage collaboratif assisté par l'IA	<ul style="list-style-type: none">• Évaluation sommative des écrits, notation• Repérage de mots clés dans les forums d'apprenants• Assistants pédagogiques basés sur l'IA• Recommandation de ressource pédagogique	<ul style="list-style-type: none">• Exploration des données éducatives pour l'allocation des ressources• Diagnostic des difficultés d'apprentissage• Services d'orientation

D'après (European Commission. Directorate General for Education, Youth, Sport and Culture., 2022)

Autre catégorisation issue des groupes thématiques numériques :

- Systèmes adaptatifs et personnalisés (recommandation et sélection de contenus, de ressources), évaluation adaptative et correction automatique
- Robots conversationnels et systèmes de tutorat intelligent, par ex. pour l'enseignement de la démarche d'investigation et de la démarche scientifique (Cisel & Baron, 2019)
- Anticipation précoce, prédiction et prévention de l'abandon scolaire
- Détection et remédiation aux difficultés d'apprentissage
- Utilisation de savoirs produits par la recherche avec l'IA (analyse d'image, annotation, classification, résumé/synthèse automatiques, traitement de données massives sur des corpus textuels ou iconographiques)
- Usages administratifs

D'après (DNE-TN2, 2021) (DNE-TN2, 2022) (Lepage Alexandre, 2022) (Romero et al., 2021)

Taxonomie d'après la revue de littérature de (Holmes et al., 2022)

IA au service des élèves

Systèmes de tutorat intelligents

Applications assistées par l'IA (par exemple, mathématiques, synthèse vocale, apprentissage des langues)

Simulations assistées par l'IA (par exemple, apprentissage par le jeu, réalité virtuelle, réalité augmentée)

IA pour aider les apprenants à besoins éducatifs particuliers

Rédaction automatique d'essais

Agents conversationnels

Évaluation formative automatique

Orchestrations de réseaux d'apprentissage

Systèmes de tutorat basés sur le dialogue

Environnements d'apprentissage exploratoire

Assistant d'apprentissage tout au long de la vie assisté par l'IA

IA au service de l'enseignant

Détection de plagiat

Curation intelligente du matériel d'apprentissage

Surveillance de la classe

Évaluation sommative automatique

IA d'assistance de l'enseignant (y compris assistant d'évaluation)

Orchestration de la salle de classe

IA au service des institutions

Admissions (par exemple, sélection des élèves)

Planification des cours, Planification des horaires, Programmation horaire

Sécurité des écoles

Identification précoce des *décrocheurs* et les élèves à risque

e-Proctoring (surveillance des examens à distance)

D'après (Holmes et al., 2022) traduction par (Bocquet, 2023)

Enfin, (Ouyang & Jiao, 2021) proposent une catégorisation selon trois paradigmes :

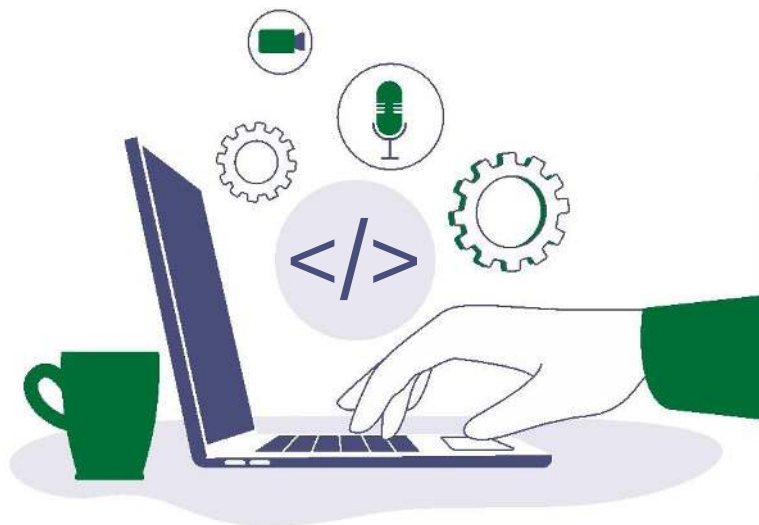
	Fondements théoriques	Implémentations	Techniques d'IA	Exemples
Paradigme 1 : dirigé par l'IA, l'apprenant en tant que destinataire	Behaviorisme [L'IA représente les connaissances du domaine et dirige les processus d'apprentissage]	Travaux antérieurs sur les systèmes de tutorat intelligents	IA basée sur des techniques relationnelles statistiques	Tuteur de programmation ACT (Anderson et al., 1990); Stat Lady (Shute, 1995)
Paradigme 2 : soutenu par l'IA, l'apprenant en tant que collaborateur	Constructivisme cognitif et social [Le système d'IA comme outil de soutien, tandis que l'apprenant collabore pour se concentrer sur le processus d'apprentissage.]	<ul style="list-style-type: none"> • Systèmes de tutorat basés sur le dialogue • Environnements d'apprentissage exploratoire 	Réseau bayésien, traitement du langage naturel, arbres de décision de Markov	Environnement exploratoire QUE (Metzler & Martincic, 1998)
Paradigme 3 : l'IA partie d'un système où l'apprenant est pilote	Connectivisme, Système adaptatif complexe [L'IA pour augmenter l'intelligence humaine, l'apprenant est placé au cœur du dispositif.]	<ul style="list-style-type: none"> • Coopération homme-ordinateur • Apprentissage personnalisé/adaptatif 	Interface cerveau-ordinateur - Apprentissage automatique, l'apprentissage profond	Modélisation prédictive MOOC en temps réel (Le et al., 2018)

Tableau d'après (Ouyang & Jiao, 2021)

4.2 Quels apports pour l'expérience d'apprentissage ?

« L'intelligence artificielle joue un rôle important et croissant dans l'éducation. Un cas important est celui des **systèmes d'enseignement personnalisés** qui sont déjà bien établis, avec des **preuves croissantes de leur efficacité pour améliorer l'apprentissage** (VanLehn, 2011 ; Kulik & Fletcher, 2016 ; Steenbergen-Hu & Cooper, 2014, 2013 ; Ma, Adesope, Nesbit, & Liu, 2014 ; du Boulay, 2016). Les systèmes d'IA dans l'éducation (AIED) peuvent également faire un **usage diversifié et sophistiqué de l'IA pour créer l'interface si importante pour l'expérience d'apprentissage**. Par exemple, l'interface peut utiliser le **traitement et la génération du langage naturel**, les **interfaces vocales**, les **avatars**, **l'analyse vidéo de l'apprenant** pour juger de son attention et de son émotion. » [Traduction] (Khosravi et al., 2022)

« En utilisant des algorithmes, **l'apprentissage peut être adaptatif**. En analysant les traces d'apprentissage de l'élève, comme des résultats à des questionnaires ou des données d'utilisation d'un logiciel, le système peut modifier son fonctionnement pour s'adapter à la personne, notamment à travers la **sélection de contenus et du niveau de difficulté**. Il commence à être possible d'analyser son comportement grâce à des capteurs, certains externes comme une caméra, et d'autres plus intrusifs comme une interface cerveau-ordinateur. Ce principe d'adaptation est au cœur de la pédagogie numérique, et se rencontre le plus souvent dans un contexte où sont aussi poursuivis des **objectifs de ludification** ; l'apprenant s'inscrivant alors dans un jeu pédagogique avec la machine, parfois en collaboration avec d'autres apprenants (Giraudon et al., 2020). Le projet KidLearn (...) propose une activité d'apprentissage dont les multiples variantes impliquant additions ou soustractions de nombres entiers ou décimaux ont



été conçues et mises en place par des didacticiens. Ces variantes sont organisées sous forme d'un graphe de difficultés croissantes, en respectant le concept de zone proximale de développement (Vygotsky, 1978). Ce concept est fondé sur l'idée que, entre un exercice trop difficile qui décourage et un exercice trop facile qui lasse, il existe une zone optimale qui maximise le progrès d'apprentissage, mesuré ici en observant les performances de l'élève au fil du jeu. Ce sont ces éléments qui sont intégrés à l'algorithme, qui va s'adapter automatiquement à la personne apprenante (Oudeyer et al., 2020).»

(Romero et al., 2023)

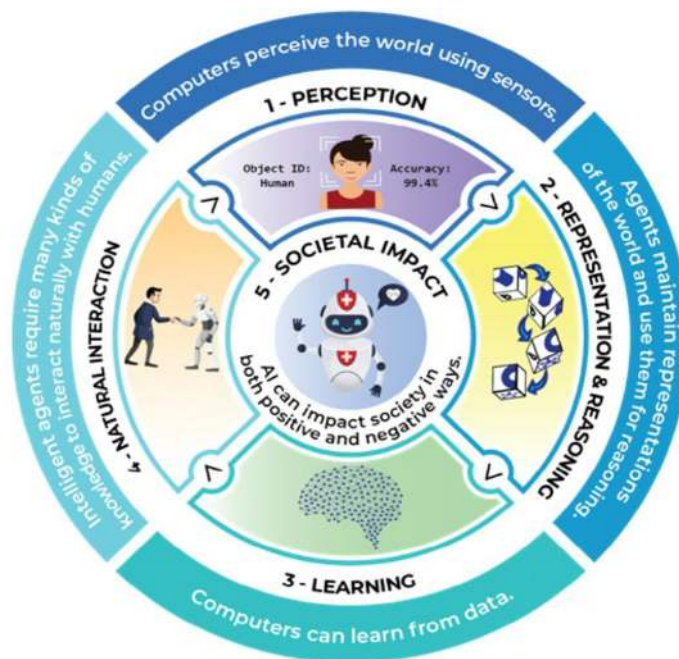
4.3 IA et production de connaissances : l'exemple des humanités numériques

«...La linguistique de corpus et la sémantique se sont considérablement développées grâce aux analyses d'IA sur la représentation des connaissances et sur l'apprentissage automatique. De même, de nombreux développements ont ouvert des **perspectives sur les études littéraires**. Au début du millénaire, sous la pression d'éditeurs (et non de scientifiques) le terme anglais *Humanities Computing* qui caractérisait les recherches à la frontière des humanités au sens anglo-saxon, c'est-à-dire des **sciences de la culture**, se transforma en *Digital Humanities* (« humanités numériques »). Ce changement de dénomination signifiait que les humanités n'étaient plus simplement instrumentées par l'utilisation d'ordinateurs, mais qu'elles se transformaient en devenant numériques. On a parlé de « **tournant calculatoire** » pour caractériser cette évolution vue comme fondamentale. Le renouvellement qui s'en est suivi a été très fructueux. Dans tous les pays, de nombreuses équipes de recherche travaillant dans le champ des humanités classiques, que ce soit en histoire, en archéologie ou en littérature, ont renouvelé leurs approches en collaboration avec des informaticiens.»

(Ganascia, 2022)

5. Quelles pistes de travail pour former à l'IA et l'enseigner ?

5.1 Les pistes proposées par AI4K12



1. Perception : Les ordinateurs (et les terminaux numériques) perçoivent le monde via des capteurs.

2. Représentation et raisonnement : Les agents d'IA entretiennent des représentations du monde, via des structures de données, et les utilisent pour raisonner, via des algorithmes, mais sans penser comme un humain.

3. Apprentissage : Les ordinateurs (et les terminaux numériques) peuvent apprendre à partir de données via des algorithmes d'apprentissage.

4. Interaction naturelle : Les agents d'IA nécessitent de nombreux types de connaissances pour interagir «naturellement» avec les humains (conversation, reconnaissance faciale, etc.) mais ne peuvent raisonner.

5. Impact sociétal : L'IA peut avoir un impact à la fois positif et négatif sur la société et nous devons en être conscient car ces technologies changent nos façons de travailler, de voyager, de communiquer, etc.

D'après (Association for the Advancement of Artificial Intelligence & Computer Science Teachers Association, 2020) [Traduction]

5.2 Pour une formation citoyenne à l'IA

« L'IA concerne tous les citoyens d'abord parce qu'elle est de plus en plus présente dans tous les secteurs de la société et ensuite parce qu'il est nécessaire de la comprendre pour participer aux débats éthiques qu'elle engendre. Il est normal de se questionner sur la pertinence de confier à des algorithmes des tâches qui mènent à des décisions cruciales, par exemple en matière de justice, d'embauche, ou d'autres situations à fortes conséquences humaines. (...) Le MOOC [« intelligence artificielle avec intelligence » de l'Inria] adopte une approche pédagogique ludique et expérientielle. (...) Les participants sont mis en action dans des activités concrètes : ils manipulent des réseaux de neurones, essaient de faire reconnaître leurs dessins par une IA, sont invités à entraîner des modèles d'apprentissage automatique eux-mêmes. Ils sont invités à réfléchir collectivement, via les forums de discussion, à des questions soulevées par le développement de l'IA. » (Romero et al., 2023)

5.3 Une proposition de parcours en cinq piliers

« L'intelligence artificielle n'a pas atteint sa maturité. Le sujet tel qu'il a été défini en 1956, étudié pendant 40 ans, a atteint ses résultats spectaculaires depuis 2012 et est encore difficile à comprendre. Il est encore plus difficile de prévoir comment les technologies vont évoluer, même dans un avenir proche. Si la construction d'un programme d'études complet est hors de portée de ce document, il est possible de proposer cinq piliers et de s'appuyer sur ceux-ci. (...)

1. Incertitude et caractère aléatoire

Les données sont incohérentes. Elles ne démontrent pas une stricte nature causale.

2. Codage et calcul

Le codage et la pensée informatique font aujourd'hui partie des programmes d'études de nombreux pays, conformément aux recommandations des experts.

3. Sensibilisation aux données

Une éducation aux données (sciences de la donnée ou data science) reposera sur des activités où les données sont collectées et visualisées, manipulées et analysées.

4. Pensée critique

Les sciences sociales peuvent et doivent apporter leur contribution à de nombreuses questions éthiques soulevées par l'IA. La pensée critique est un aspect important, mais il est essentiel qu'elle repose sur une réelle compréhension du fonctionnement de la technologie.

5. Humanisme post-IA

L'idée clé est que les progrès de l'IA nous font reconsidérer, en tant qu'êtres humains, certaines vérités de base.»

D'après (de la Higuera & Bocquet, 2020)

Production d'un module d'initiation à l'IA. Exemple d'instruction et de réponse produite par ChatGPT (11/04/23)

Instruction

Dans le contexte de l'enseignement scolaire en France. Proposer, sous la forme d'un tableau, un module d'environ une heure d'initiation à l'intelligence artificielle pour les professeurs (enseignement primaire et secondaire). Ce module doit présenter les origines de l'IA, les définitions de base, les enjeux pour l'éducation et des exemples d'application pédagogique.

Réponse

Objectif	Contenu	Durée
Introduction	Présentation du module et des objectifs	5 min
Origines de l'IA	Historique et évolution de l'IA	10 min
Définitions de base	Explication des termes clés : intelligence artificielle, apprentissage automatique, réseau de neurones, etc.	15 min
Les enjeux pour l'éducation	Discussion sur les avantages et les défis de l'IA dans l'éducation	15 min
Exemples d'applications pédagogiques de l'IA	Présentation de projets d'IA appliqués à l'éducation, tels que les chatbots d'assistance aux élèves, les systèmes de recommandation de contenus éducatifs, etc.	15 min
Les limites de l'IA	Discussion sur les limites de l'IA en tant qu'outil pédagogique et sur les préoccupations éthiques liées à son utilisation	10 min
Conclusion	Résumé des points clés et ouverture sur les perspectives d'avenir de l'IA dans l'éducation	10 min

« Ce module d'initiation à l'IA pour les professeurs peut être adapté en fonction de leur niveau de connaissance et d'expérience. Il peut également être suivi d'un atelier pratique où les enseignants pourront expérimenter eux-mêmes l'utilisation de l'IA dans leur classe. »

Source : <https://chat.openai.com/chat>

5.4 Formation à l'IA et prévention du décrochage scolaire

« La **Scientothèque** est une association sans but lucratif installée au sein de l'Expérimentarium, musée de Physique de l'Université Libre de Bruxelles. Depuis sa fondation en 2001, sa mission principale est de contribuer à la diminution des inégalités sociales. Deux réalités ont mené à la mise en place de l'association : d'une part, le risque accru de décrochage scolaire et la difficulté d'accès aux études supérieures pour les jeunes issus de milieux précarisés (Coslin, 2012) et, d'autre part, la désaffection pour les études scientifiques de la part du public féminin. Il a d'ailleurs été démontré qu'il était moins évident d'attirer les filles vers les projets à caractère scientifique ou technologique (Blanchard, 2021). (...) Une sélection d'activités a été proposée à deux groupes de jeunes de 10-12 ans durant l'année scolaire 2020-2021 lors d'**ateliers extrascolaires hebdomadaires** à l'Institut des Ursulines à Koekelberg. Ce projet, financé par la région de Bruxelles-Capitale, visait à diminuer le décrochage scolaire pour des élèves vivant dans des quartiers précarisés. Tout au long des séances, les jeunes ont pu, grâce à une **approche ludique et collaborative**, découvrir l'histoire des machines, établir des liens avec l'intelligence biologique, comprendre ce qui se cache derrière le mot « algorithme », observer et programmer des robots Thymio, s'initier à la programmation sur la plateforme Scratch, et débattre des conséquences de l'IA sur la société. » (Romero et al., 2023)

6 Le tournant des IA génératives et des grands modèles de langage

6.1 Le développement des IA génératives

La catégorie des **IA génératives** désigne les technologies d'IA entraînées sur des données massives (ensemble de Wikipédia, GitHub, réseaux sociaux, Google Images, etc.) et **pouvant générer de nouveaux contenus** (texte, image, audio, vidéo) de façon aléatoire à partir d'une information d'entrée («prompt») saisie par l'utilisateur.

Les **modèles d'IA générative** (réseaux de neurones profonds, transformers et autres modèles, tels que les réseaux antagonistes génératifs), qui se multiplient surtout depuis 2022, peuvent être répartis selon la taxonomie suivante :

- Texte à image (ex. DALL-E 2 d'OpenAI)
- Texte à représentations 3D (ex. Dreamfusion de Google Research)
- Image à texte (ex. Flamingo de Deepmind)
- Texte à vidéo (ex. Phenaki de Google Research)
- Texte à audio (Jukebox d'OpenAI)
- Texte à texte (ex. ChatGPT d'OpenAI)
- Texte à code (ex. Codex d'OpenAI)
- Texte à formule scientifique (ex. Galactica de Meta AI)
- Autres modèles

D'après (Gozalo-Brizuela & Garrido-Merchan, 2023)

La récente multiplication des IA génératives

«Au cours des deux dernières années, une **pléthore de grands modèles génératifs** tels que ChatGPT ou Stable Diffusion ont été publiés. Concrètement, ces modèles sont capables d'effectuer des tâches telles qu'un système général de questions-réponses ou la création automatique d'images artistiques **qui révolutionnent plusieurs secteurs**. Par conséquent, les implications que ces modèles génératifs ont dans l'industrie et la société sont considérables, car plusieurs tâches professionnelles peuvent être transformées. Par exemple, l'IA générative est capable de transformer de manière efficace et créative des textes en images, comme le modèle DALL-E-2 ; des textes en images 3D, comme le modèle Dreamfusion ; des images en texte, comme le modèle Flamingo ; des textes en vidéo, comme le modèle Phenaki ; des textes en audio, comme le modèle AudioLM ; des textes en d'autres textes, comme ChatGPT ; des textes en code, comme le modèle Codex ; des textes en textes scientifiques, comme le modèle Galactica ou même créer des algorithmes comme AlphaTensor.» (Gozalo-Brizuela & Garrido-Merchan, 2023)

6.2 Grands modèles de langage et nouvelle génération d'agent conversationnel

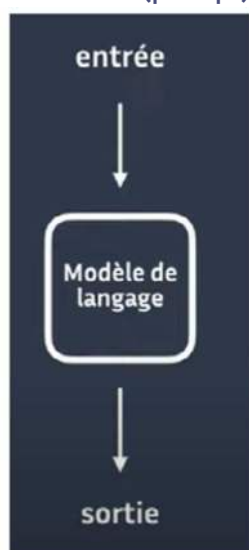
6.2.1 Comment fonctionnent les grands modèles de langage ?

Le principe des modèles de langage (ou LLM : *Large Language Models*) n'est **pas nouveau** : ce type d'algorithme est déjà présent quotidiennement avec les **assistants de rédaction des messages** instantanés. Son fonctionnement consiste à **prédire le mot suivant** à partir d'une instruction («prompt») saisie par l'utilisateur. La **phase d'entraînement** permet d'ajuster les paramètres internes à partir de données textuelles issues du web (non vérifiées). Une quantité massive de données et de paramètres permet d'améliorer les performances de prédiction. On ne peut pas pour autant parler de «compréhension» ou de «personnalité» de l'algorithme, il s'agit d'abord de **réponses produites à partir de régularités statistiques**. Par ailleurs ces modèles peuvent produire des **«hallucinations»**, en inventant de fausses informations, ce qui nécessite la plus grande vigilance de la part de l'utilisateur. Actuellement les modèles de langage les plus puissants sont détenus par de grandes entreprises, qui communiquent très peu sur leurs données d'entraînement. D'après (Inria Flowers, 2023a)

Voir aussi (Inria Flowers, 2023b) sur le «*prompting*» : «faire réaliser une certaine tâche à un modèle de langage (...) en lui expliquant cette tâche au moyen de phrases en langage naturel.» Deux méthodes sont présentées : «expliquer la tâche en langage naturel, et donner des exemples de réalisation de la tâche».

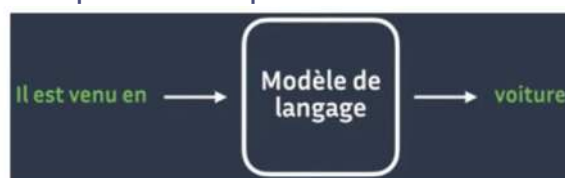
Fonctionnement d'un grand modèle de langage et prompting

Instruction (prompt)



Réponse

Principe de base : la prédiction de mots



D'après (Inria Flowers, 2023)

6.2.2 Le tournant ChatGPT : origines et socle technologique

Dans cette catégorie des IA génératives, l'application ChatGPT, ouverte au public par société OpenAI depuis le 30 novembre 2022, représente une **nouvelle génération d'agent conversationnel** (*chatbot* en anglais), «agent qui dialogue avec un utilisateur dans le but de faire une recherche, de trouver une réponse à sa question ou encore de s'exercer dans une discipline donnée» (CSEN & DNE-TN2, 2022). En raison des performances affichées, elle constitue manifestement un **«saut qualitatif»** dans les technologies de traitement automatique du langage (TAL) - ou **analyse du langage naturel (Natural Language Processing - NLP)** (Boisson & DNE-TN2, 2023).

Ce **«prototype d'agent conversationnel (...), modèle de langage affiné à l'aide de techniques d'apprentissage supervisé et d'apprentissage par renforcement»** («ChatGPT», 2023), basé sur une intelligence artificielle générative, capable de générer des textes à la demande ou du code informatique à partir de questions en langage naturel, est une **application propriétaire développée** à partir d'une optimisation du modèle algorithmique GPT 3.5.

« Le GPT-3 est plus important que le GPT-2, avec 175 milliards de paramètres. Pour avoir une idée plus précise de la dimension de cet algorithme, les données de Wikipédia, qui comprennent environ 6 millions d'articles, ne représentent que 0,06% des données entraînées par le GPT-3.» (Boa Sorte et al., 2021)

Les technologies qui fondent ChatGPT peuvent être résumées ainsi :

- il s'agit d'un **«modèle de langage étendu**, un type d'intelligence artificielle qui utilise l'apprentissage profond (une forme d'apprentissage automatique) pour traiter et générer des textes en langage naturel (...) [Ce type de modèle est] formé sur des **quantités massives de données textuelles, lui permettant d'apprendre les nuances et les complexités du langage humain»** (d'après (Susnjak, 2022) [traduction]) ;
- son appellation résulte d'une contraction entre «chat» pour conversation et «GPT» [*Generative Pre-trained Transformer*], qui est un «modèle de transduction de langage prédictif» (transduction : transformation d'une donnée représentée sous une forme en une donnée représentée sous une autre forme) ;
- son algorithme basé sur un **«transformer»** (Vaswani et al., 2017) correspond à un modèle d'architecture de réseau de neurones qui utilise le **mécanisme d'auto-attention permettant d'établir des relations entre les mots et de déterminer le sens d'un mot en fonction du contexte**, en se focalisant sur les mots les plus pertinents, permettant ainsi de s'approcher de façon crédible du langage naturel ;
- il fonctionne selon un **modèle de prédiction de mots**, à partir des instructions/ consignes ou «prompts» de l'utilisateur en langage naturel, combinant dans sa phase d'entraînement apprentissage par renforcement et apprentissage supervisé à partir des rétroactions humaines ;
- il est donc important de rappeler que, comme toute technologie IA, les textes produits sont générés et renforcés à partir d'**actions et de travail humains** (des personnes qui conçoivent l'algorithme jusqu'à celles qui identifient, vérifient des données, ou interagissent avec le programme).

Le socle technologique de ChatGPT est le résultat d'**«une combinaison de deux modèles** : un modèle **linguistique** (dont les références s'arrêtent en 2021), GPT, modèle géant de prédiction de texte et modèle encyclopédique, et un modèle **conversationnel**

(qui s'affine en fonction du retour des utilisateurs), InstructGPT, version «redressée» de GPT-3 à partir de l'annotation de textes générés.»

C'est aussi le résultat d'«une série de **strates issues de la recherche en linguistique et en informatique** :

- la **statistique sémantique**, avec le concept de «fenêtre contextuelle» (le sens procède de la position relative des mots les uns avec les autres) ;
- la **vectorisation de mots** (*word embedding* ou « plongement lexical ») permettant de caractériser un mot au sein d'espaces contextuels en fonction des mots voisins ;
- les **réseaux de neurones** pour reconnaître les relations syntaxiques et sémantiques précises entre les mots et gérer les transferts linguistiques (les réponses aux instructions des utilisateurs peuvent ainsi s'appuyer sur un espace sémantique multilingue) ;
- les **modèles transformers** (2017) fondés sur le mécanisme d'attention et modélisant les formules syntaxiques ;
- le développement des **grands modèles de langage**, à partir de 2018 avec BERT (Google) puis GPT-2 et 3 (OpenAI).»

D'après (Langlais, 2023)

Le 14 mars 2023 OpenAI annonce la montée en version de son modèle avec le lancement de **GTP-4** (service payant à ce jour ou disponible dans la nouvelle version du moteur de recherche Bing) : « **modèle multimodal** à grande échelle qui peut accepter des entrées d'image et de texte et produire des sorties de texte. Bien que moins capable que les humains dans de nombreux scénarios du monde réel, GPT-4 présente des performances de niveau humain sur divers critères professionnels et académiques, y compris la réussite d'un examen du barreau simulé avec un score autour des 10% des meilleurs candidats. » (OpenAI, 2023)

6.2.3 Les autres grands modèles de langage

Parmi les autres grands modèles de langage, on peut citer notamment, en open source :

- **Bert** (Bidirectional Encoder Representations from Transformers), développé par Google en 2018 (Devlin et al., 2019) ;
- **Bloom**, initié par la start-up franco-américaine Hugging Face (projet BigScience, 2021) entraîné sur la machine « Jean Zay » «un des plus puissants supercalculateurs d'Europe» (CNRS, 2022) ;
- **LLaMA** de Meta AI (Touvron et al., 2023).

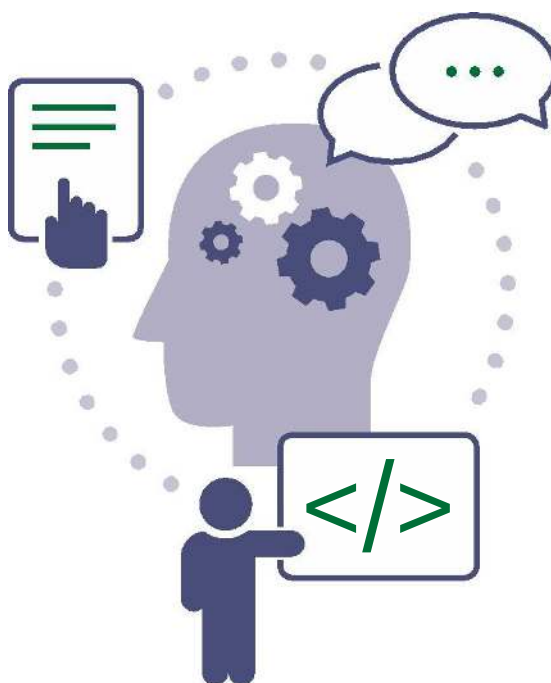
Pour plus de détails consulter aussi (Pôle d'expertise de la régulation numérique, 2023)

6.2.4 Les réactions et interrogations

Depuis l'ouverture de l'application ChatGPT les réactions et **les débats ont été très nombreux** dans les médias et dans de nombreux milieux professionnels, parmi lesquels la recherche et l'éducation.

« Ces IA génératives multilingues, multitâches et interactives posent de **nombreuses questions** : les données choisies pour les entraîner, la répartition des langues dans un système multilingue, les paramètres d'optimisation des systèmes, la propriété des contenus générés, etc. De plus, le pouvoir génératif des IA est souvent amplifié par des **filtres qui permettent de censurer certains sujets** et des modules de déduction logique dans le but de vérifier la véracité des énoncés. Une poignée d'humains (ingénieurs, transcripteurs, évaluateurs) ont créé ce type de système utilisé par des millions de personnes. Ces systèmes d'intelligence artificielle utilisés massivement posent donc des **défis éthiques majeurs** parmi lesquels la transformation de la notion de production d'information, le rapport à la vérité et les risques massifs liés à la désinformation et la manipulation.»
(Deville, 2023)

Les réactions ont été **de la fascination**, face au caractère crédible et vraisemblable des « conversations » produites et à la pertinence - au moins apparente - de certaines réponses, **à l'inquiétude** en raison des mésusages possibles (possibilité de triche pour réaliser des travaux scolaires, de fraude aux examens, doute sur la paternité des écrits produits dans tel ou tel contexte personnel ou professionnel) et des incertitudes sur les technologies, les données mobilisées et la fiabilité des réponses produites.



6.2.5 Quelles potentialités pédagogiques ?

Ce type d'application offre plusieurs potentialités pédagogiques très activement explorées et partagée actuellement par la communauté des enseignants :

- un **soutien à la créativité** en proposant, à partir des requêtes et questions posées, une première base de rédaction ou de code ;
- un **soutien à l'apprentissage du français et des langues étrangères** par la fonction de suivi conversationnel ;
- une **assistance** pour amorcer ou optimiser une tâche rédactionnelle ;
- un **nouveau défi pour l'éducation aux médias et à l'information (EMI)** et un cadre pour la **formation à l'esprit critique**, qui nécessite de la part des enseignants et de leurs élèves de vérifier systématiquement le contenu et la pertinence des réponses fournies.

Ces potentialités ont déjà fait l'objet de plusieurs publications, parmi lesquelles (Kasneci et al., 2023) qui pointe les possibilités suivantes : apprentissage personnalisé (analyse de travaux et suggestion de matériels pédagogiques adaptés), planification des cours (assistance à la création de plans et d'activités inclusives), apprentissage des langues (résumés et traductions). [Traduction]

Par ailleurs, plusieurs **guides pédagogiques** proposent des pistes d'utilisation pour l'enseignant. Ainsi le guide produit par (Herft, 2023), enseignant australien, et traduit en français, propose une série d'exemples d'instructions à saisir pour en faire un assistant pédagogique et didactique. Ex. : «Créez un plan de cours pour une classe... qui couvre le concept de ... et comprend une variété d'activités et d'évaluations.»

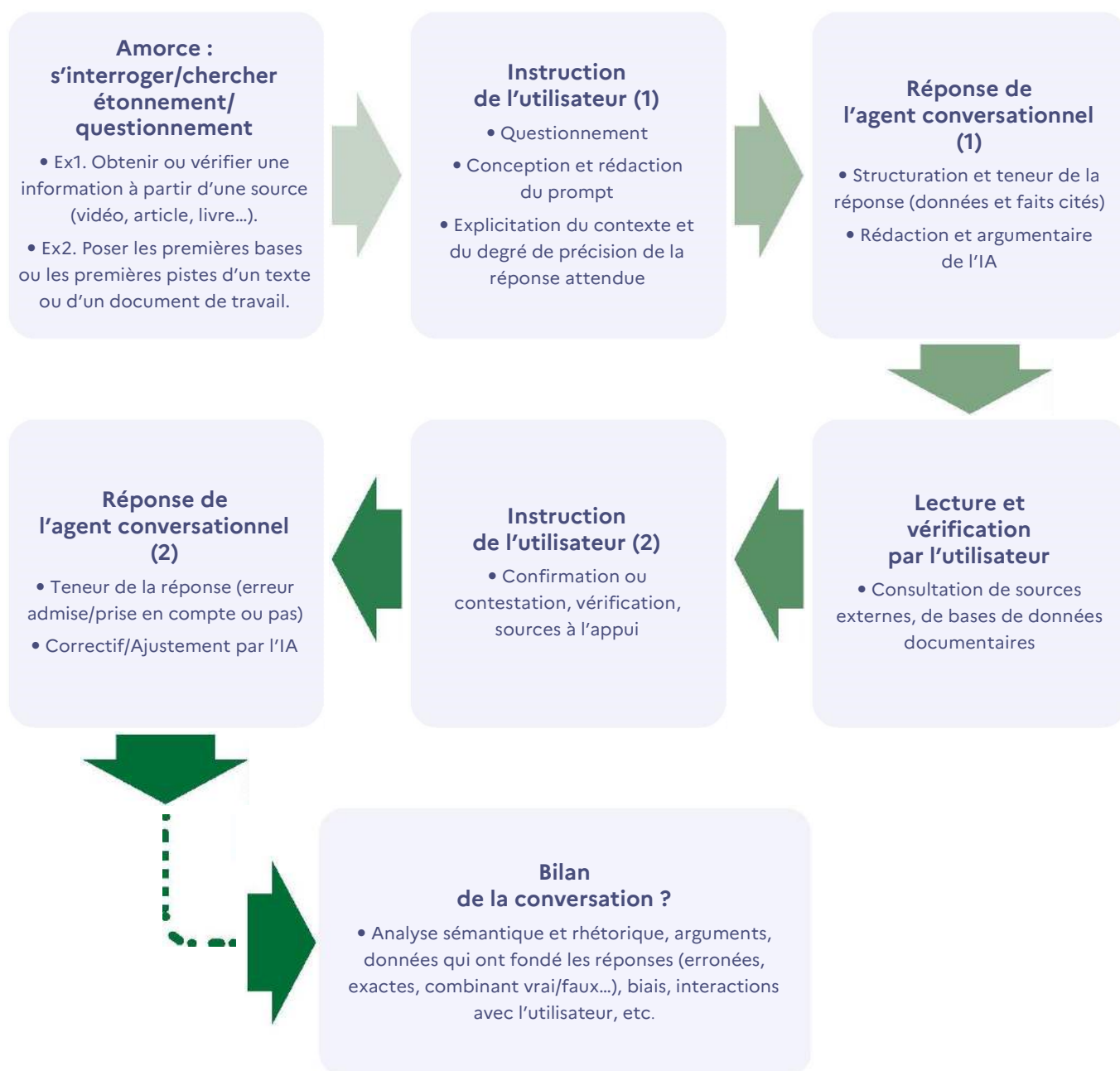
Les perspectives des grands modèles de langage pour les personnes handicapées

«Afin d'autonomiser les apprenants handicapés, [ils] peuvent être utilisés en combinaison avec des solutions de conversion de la parole en texte ou du texte en parole pour aider les personnes souffrant de déficience visuelle. En combinaison avec les possibilités de tutorat en groupe et à distance mentionnées précédemment, les modèles de langage peuvent être utilisés **pour développer des stratégies d'apprentissage inclusives** avec un soutien adéquat dans des tâches telles que l'écriture adaptative, la traduction et la mise en évidence de contenus importants dans différents formats. Cependant, il est important de noter que l'utilisation de grands modèles de langage **doit être accompagnée de l'aide de professionnels** tels que des orthophonistes, des éducateurs et d'autres spécialistes qui peuvent adapter la technologie aux besoins spécifiques des handicaps de l'apprenant.» (Kasneci et al., 2023) [Traduction]

6.2.6 Quels limites et points de vigilance (dimensions éthiques, pédagogiques et scientifiques) ?

Plusieurs tests ont été réalisés pour évaluer les performances et limites de ChatGPT (GPT 3.5) (Allouche, 2023). Le schéma ci-dessous propose une modélisation de parcours utilisateur critique, intégrant des consignes qui peuvent être simples ou complexes, allant d'une simple demande d'information à une problématisation.

Modélisation de parcours utilisateur critique avec un agent conversationnel



À l'issue de ces premiers tests, plusieurs limites et points de vigilance peuvent être relevés (sachant que dans le cadre du **RGPD** l'application n'est pas utilisable dans un cadre scolaire, seules sont possibles des utilisations à titre individuel, sous le régime du contrat privé) :

- pour l'instant les réponses fournies ne sont **pas basées sur des données en temps réel** via une connexion au web (sous réserve de vérification, il s'agit des données issues du web, antérieures à 2022) ;
- elles contiennent des **biais** en raison de la nature des données traitées, même si des filtres sont appliqués pour limiter les biais les plus flagrants ;
- les **erreurs factuelles** et possibles «hallucinations» nécessitent quasi systématiquement une vérification des contenus par l'utilisateur, ce qui peut constituer à la fois une limite et un atout dans un cadre pédagogique pour développer l'esprit critique et établir une distance avec les contenus générés ;
- le risque de produire et de diffuser des infox ou des informations tronquées, erronées ou non vérifiables ;
- une telle technologie, en raison même de ses performances et de la masse des données traitées (sans transparence sur la nature de celles-ci), soulève inévitablement les questions du référencement (peut-on se référer aux réponses produites alors qu'elles ne sont pas systématiquement sourcées ni reproductibles ?), de la paternité des contenus (à qui appartiennent-ils, peut-on attribuer la paternité d'une œuvre à une IA ?) (Stokel-Walker, 2023) ou du **droit d'auteur** (les réponses générées à partir de ces données massives respectent-elles le droit d'auteur lorsque celui-ci s'applique ? par exemple sur une œuvre scientifique ou littéraire qui n'est pas sous licence libre) ;
- **l'absence de sources** - ou le caractère possiblement erroné de celles qui sont citées - constitue une véritable difficulté pour une utilisation dans un cadre académique (éducation, recherche, documentation), sauf si l'utilisateur en fait la demande expresse, et sous réserve que celui-ci vérifie une à une les références citées.

En réponse à ce problème d'autres applications concurrentes, comme Perplexity (<https://www.perplexity.ai/>), proposent de donner des sources (mais sans non plus la garantie sur la pertinence ou la complétude de celles-ci).

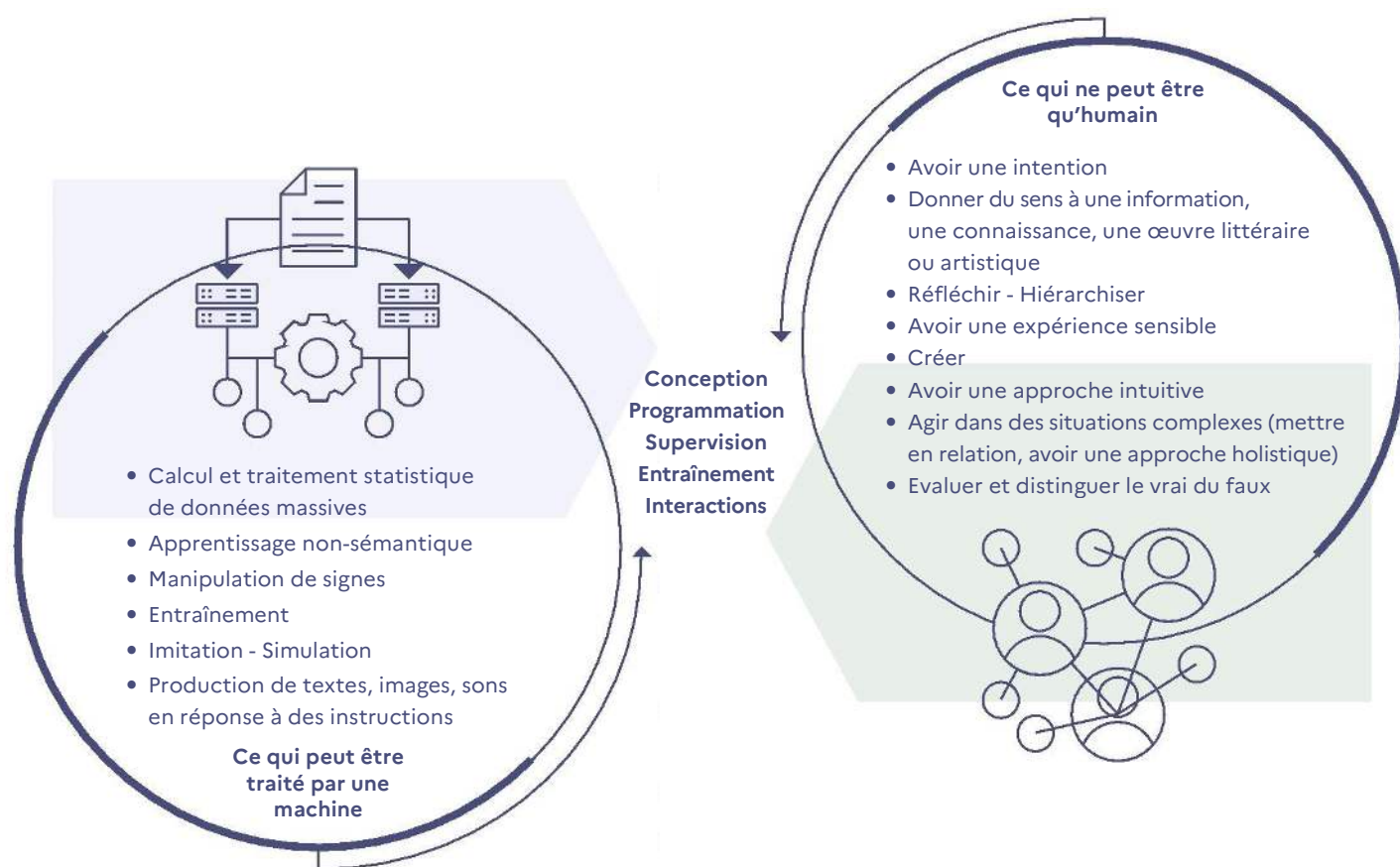
On peut noter aussi l'**impact environnemental** de ce type de modèle, non spécifique à ChatGPT :

«On estime que l'entraînement du modèle GPT-3 d'OpenAI a nécessité autant d'énergie que le voyage aller-retour d'une voiture sur la lune, générant ainsi l'équivalent de 85 000 kg d'émissions de CO2 (Quach, 2020). Il est désormais largement admis que la consommation d'énergie constitue un défi majeur pour l'IA axée sur les données (Strubell et al., 2019).» (Holmes & Tuomi, 2022) traduction par (Bocquet, 2023)

En résumé, les principales pistes à explorer relèvent surtout de l'assistance (voir tableau p. 25), de l'inspiration, du complément ou de l'approfondissement. En effet, un tel saut technologique invite surtout les chercheurs et les pédagogues à, davantage encore, innover, collaborer, s'informer et se former, **mettre à distance et faire preuve d'esprit critique** pour accompagner les élèves et les étudiants dans leurs apprentissages et leur formation.

Comme le suggère le schéma ci-dessous, parmi les défis à relever figure la **distinction à établir avec les élèves entre ce qui relève exclusivement de l'humain** : donner du sens à une information, une connaissance, une œuvre littéraire ou artistique, **et ce qui peut effectivement être confié à une IA** : le traitement et le calcul de données massives, au-delà des capacités humaines, à des fins de prédiction et d'optimisation, mais sans intention ni compréhension de ce qui est produit.

Ce qui peut relever ou pas d'une IA



Références

Allouche E. (2023, janvier 5). Sens et finalités du numérique en éducation – Hors-série : Tests et simulations d'« entretien » avec ChatGPT (Open AI) [Billet]. *Éducation, numérique et recherche*.
<https://edunumrech.hypotheses.org/7635>

Association for the Advancement of Artificial Intelligence & Computer Science Teachers Association. (2020). *AI4K12*. AI4K12.
<https://ai4k12.org/>

Association Française pour l'Intelligence Artificielle (AFIA). (2022, janvier 6). Journée Enseignement et Formation en IA : « IA pour l'enseignement ». *AfIA*.
<https://afia.asso.fr/efia-2022/>

Boa Sorte P., Farias M. A. de F., Santos A. E. dos Santos, J. do C. A. & Dias, J. S. dos S. R. (2021). Artificial intelligence in academic writing : What is in store with the gpt-3 algorithm? *Revista EntreLínguas*, 7(00).
<https://periodicos.fclar.unesp.br/entrelinguas/article/download/15352/11410/51225>

Bocquet F. (2023, janvier 23). État de l'art et de la pratique de l'intelligence artificielle dans l'éducation (Holmes & Tuomi, 2022) [Traduction] [Billet]. *Éducation, numérique et recherche*.
<https://edunumrech.hypotheses.org/8350>

Boisson J. & DNE-TN2. (2023, janvier 12). Intelligence artificielle, traitement automatique des langues et agents conversationnels [Billet]. *Éducation, numérique et recherche*.
<https://edunumrech.hypotheses.org/8235>

Cardon D., Cointet J.-P. & Mazieres A. (2018). La revanche des neurones : L'invention des machines inductives et la controverse de l'intelligence artificielle. *Réseaux*, 5(211).
<https://doi.org/10.3917/res.211.0173>

Cisel M., & Baron G.-L. (2019). Vers des intelligences artificielles pour l'enseignement de la démarche d'investigation. *Educational Journal of the University of Patras UNESCO Chair*, 0, Article 0.
<https://doi.org/10.26220/june.2990>

CNRS. (2022, juillet 12). *Livraison du plus grand modèle de langue multilingue « open science » jamais entraîné*. CNRS.
<https://www.cnrs.fr/fr/livraison-du-plus-grand-modele-de-langue-multilingue-open-science-jamais-entraine>

Collin S. & Marceau E. (2021). L'intelligence artificielle en éducation : Enjeux de justice. *Formation et profession*, 29(2), 1.
<https://doi.org/10.18162/fp.2021.a230>

Cour des comptes. (2023). *La stratégie nationale de recherche en intelligence artificielle*. Cour des comptes.
<https://www.ccomptes.fr/fr/publications/la-strategie-nationale-de-recherche-en-intelligence-artificielle>

de la Higuera C. & Bocquet F. (2020, juillet 9). L'éducation, la formation des enseignants et l'apprentissage de l'intelligence artificielle : Un aperçu des questions clés [Billet]. *Éducation, numérique et recherche*.

<https://edunumrech.hypotheses.org/1973>

Devlillers L. (2023, février 20). *ChatGPT et « intelligences » artificielles : Comment déceler le vrai du faux*. The Conversation.

<http://theconversation.com/chatgpt-et-intelligences-artificielles-comment-deceler-le-vrai-du-faux-200181>

Devlin J., Chang M.-W., Lee K. & Toutanova K. (2019). *BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv.

<https://doi.org/10.48550/arXiv.1810.04805>

Directorate-General for Education, Youth, Sport and Culture (European Commission). (2022). *Lignes directrices pour les enseignants et les éducateurs en matière de lutte contre la désinformation et de promotion de l'habileté numérique par l'éducation et la formation*. Publications Office of the European Union.

<https://op.europa.eu/en/publication-detail/-/publication/a224c235-4843-11ed-92ed-01aa75ed71a1/language-fr>

DNE-TN2. (2021, juin 30). *Le deep learning, un tournant épistémologique pour les SHS ?* (Huma-Num LAB, 2020) [Billet]. *Éducation, numérique et recherche*.

<https://edunumrech.hypotheses.org/3203>

DNE-TN2. (2022, janvier 20). *Les enjeux éducatifs à l'ère de l'Intelligence Artificielle : GTnum #Scol_IA – Maison de l'Intelligence artificielle* [Billet]. *Éducation, numérique et recherche*.

<https://edunumrech.hypotheses.org/3898>

DNE-TN2. (2023, mars 9). *Intelligence artificielle et éducation ouverte : Retour sur la journée du GTnum #IA_EO (26/01/23)* [Billet]. *Éducation, numérique et recherche*.

<https://edunumrech.hypotheses.org/8605>

Ganascia J.-G. (2022). *Intelligence artificielle et épistémologie. Allers-retours indispensables. 1*.

<https://hal.science/hal-03760357>

Gouvernement - Direction générale des Entreprises. (2023). *Éduquer à l'Intelligence artificielle*. Stratégie nationale pour l'intelligence artificielle.

<https://www.entreprises.gouv.fr/fr/numerique/enjeux/developper-ecosysteme-de-talents>

Gozalo-Brizuela R. & Garrido-Merchan E. C. (2023). *ChatGPT is not all you need. A State of the Art Review of large Generative AI models*.

<https://doi.org/10.48550/ARXIV.2301.04655>

Grand lexique français de l'Intelligence artificielle. (2021). DataFranca Wiki.

<https://datafranca.org/wiki/Accueil>

Herft A. (2023). *Guide de l'enseignant. L'usage de ChatGPT « ce qui marche le mieux »* (A. Gagné, Trad.).

<https://labua.univ-angers.fr/wp-content/uploads/2023/01/ChatGPT-Guide-de-lenseignant-FR.pdf>

Holmes W., Persson J., Chounta I.-A., Wasson B. & Dimitrova V. (2022). *Artificial intelligence and education. A critical view through the lens of human rights, democracy and the rule of law*. Council of Europe.

<https://rm.coe.int/artificial-intelligence-and-education-a-critical-view-through-the-lens/1680a886bd>

Holmes W. & Tuomi I. (2022). State of the art and practice in AI in education. *European Journal of Education*, (n/a).

<https://doi.org/10.1111/ejed.12533>

Inria (2020). *L'Intelligence Artificielle... avec intelligence !* (MOOC). FUN-MOOC.

<https://www.fun-mooc.fr/courses/course-v1:inria+41021+session01/about>

Inria Flowers (Réalisateur). (2023a, mars 8). *Comment fonctionne ChatGPT? Un tour d'horizon en moins de 5mn*.

<https://www.youtube.com/watch?v=K8gOvC8gvB4>

Inria Flowers (Réalisateur). (2023b, mars 13). *Le prompting, ou l'art de se faire comprendre par ChatGPT : Explication en 5mn*.

https://www.youtube.com/watch?v=8lQ9i_QoA3A

Kasneci E., Seßler K., Küchemann S., Bannert M., Dementieva D., Fischer F., Gasser U., Groh G., Günemann S., Hüllermeier E., Krusche S., Kutyniok G., Michaeli T., Nerdel C., Pfeiffer J., Poquet O., Sailer M., Schmidt A., Seidel T., ... Kasneci G. (2023). *ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education*.

<https://doi.org/10.35542/osf.io/5er8f>

Khosravi H., Shum S. B., Chen G., Conati C., Gasevic D., Kay J., Knight S., Martinez-Maldonado R., Sadiq S. & Tsai, Y.-S. (2022). Explainable Artificial Intelligence in education. *Computers and Education: Artificial Intelligence*, 100074.

<https://doi.org/10.1016/j.caeai.2022.100074>

Langlais P.-C. (2023, février 7). *ChatGPT : Comment ça marche ?* [Billet]. *Sciences communes*.

<https://scoms.hypotheses.org/1059>

Le Blanc B. (2014). Alan Turing : Les machines à calculer et l'intelligence. *Hermès, La Revue*, 68(1), 123-126.

<https://doi.org/10.3917/herm.068.0123>

Le Cun Y., Brizard C. & Brizard C. (2019). *Quand la machine apprend : La révolution des neurones artificiels et de l'apprentissage profond*. Odile Jacob.

Lepage Alexandre (Réalisateur). (2022, mars 3). *L'intelligence artificielle en éducation*.

<https://www.youtube.com/watch?v=ZAu0xllcls>

OCDE. (2023, mars 28). *Is Education Losing the Race with Technology? : AI's Progress in Maths and Reading*. OCDE.

<https://www.oecd.org/education/is-education-losing-the-race-with-technology-73105f99-en.htm>

OpenAI. (2023). *GPT-4 Technical Report*. OpenAI.

<https://cdn.openai.com/papers/gpt-4.pdf>

Ouyang F., & Jiao P. (2021). Artificial intelligence in education : The three paradigms. *Computers and Education: Artificial Intelligence*, 2, 100020.
<https://doi.org/10.1016/j.caeai.2021.100020>

Pôle d'expertise de la régulation numérique. (2023). ChatGPT ou la percée des modèles d'IA conversationnels. Gouvernement français, 6.
https://www.peren.gouv.fr/rapports/2023-04-06_Eclairage%20sur_CHATGPT_FR.pdf

Popenici S. A. D. & Kerr S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(1), 22.
<https://doi.org/10.1186/s41039-017-0062-8>

Romero M., Aloui H., Heiser L., Galindo L. & Lepage A. (2021). *Un bref parcours sur les ressources, pratiques et acteurs en IA et éducation*.
<https://hal.archives-ouvertes.fr/hal-03190014>

Romero M., Heiser L., Lepage A., Gagnebien A., Bonjour A., Lagarrigue A., Palaude A., Boulord C., Gagneur C.-A., Mercier C., Caucheteux C., Guidoni-Stoltz D., Tressols F., Alexandre F., Céci J.-F., Metral J.-F., Camponovo J., Henry J., Fouché L., ... Borgne Y.-A. L. (2023). *Enseigner et apprendre à l'ère de l'intelligence artificielle (GTnum #Scol_IA): Vol. Livre blanc*.
<https://hal.science/hal-04013223>

Stokel-Walker C. (2023). ChatGPT listed as author on research papers: Many scientists disapprove. *Nature*.
<https://doi.org/10.1038/d41586-023-00107-z>

Susnjak T. (2022). *ChatGPT: The End of Online Exam Integrity?* (arXiv:2212.09292). arXiv.
<https://doi.org/10.48550/arXiv.2212.09292>

Touvron H., Lavril T., Izacard G., Martinet X., Lachaux M.-A., Lacroix T., Rozière B., Goyal N., Hambro E., Azhar F., Rodriguez A., Joulin A., Grave E. & Lample G. (2023). *LLaMA: Open and Efficient Foundation Language Models* (arXiv:2302.13971). arXiv.
<https://doi.org/10.48550/arXiv.2302.13971>

UNESCO. (2019). *Consensus de Beijing sur l'intelligence artificielle et l'éducation*. UNESCO Bibliothèque Numérique.
<https://unesdoc.unesco.org/ark:/48223/pf0000368303>

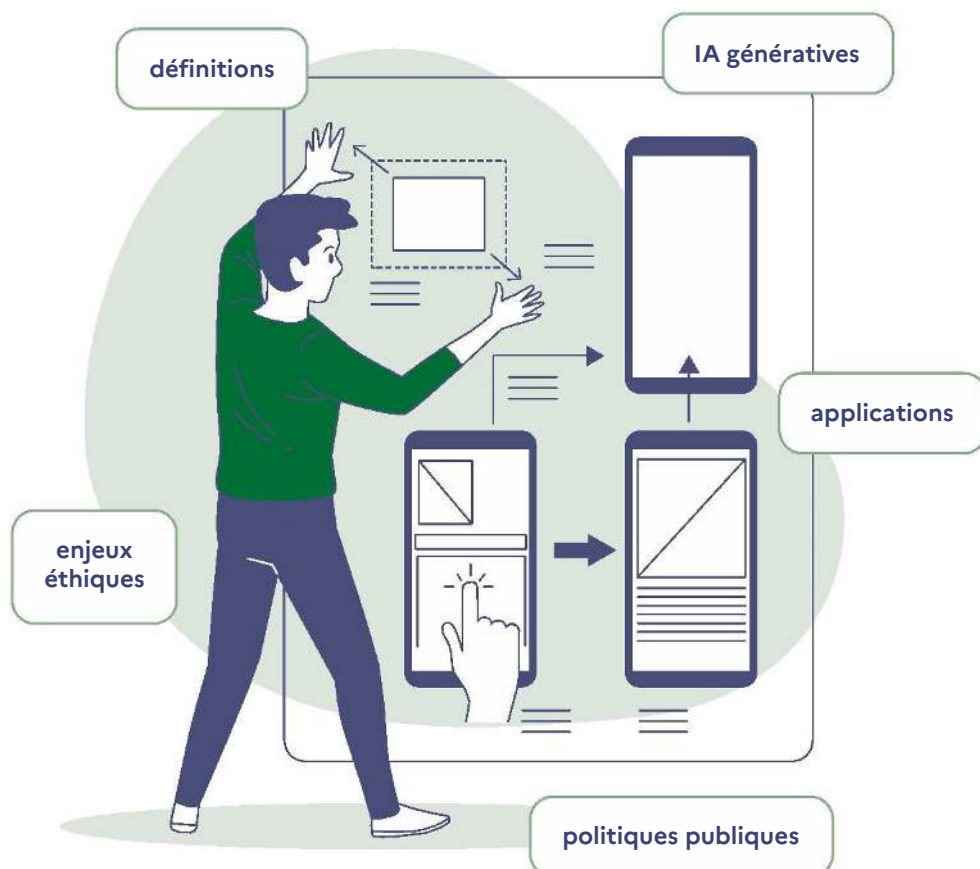
UNESCO. (2021). *IA et éducation : Guide pour les décideurs politiques*. UNESCO Bibliothèque Numérique.
<https://unesdoc.unesco.org/ark:/48223/pf0000380006>

UNICEF. (2021, novembre). *Orientations stratégiques sur l'IA destinée aux enfants 2.0*. UNICEF.
<https://www.unicef.org/globalinsight/fr/rapports/orientations-strat%C3%A9giques-sur-lia-destin%C3%A9e-aux-enfants>

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser L. & Polosukhin I. (2017). *Attention Is All You Need*.
<https://doi.org/10.48550/ARXIV.1706.03762>

Référence de l'article : DNE-TN2 (2023). Intelligence artificielle et éducation : Apports de la recherche et enjeux pour les politiques publiques [Billet]. Éducation, numérique et recherche.

<https://edunumrech.hypotheses.org/8726>



Carnet Hypothèses
Éducation, numérique
et recherche



<https://edunumrech.hypotheses.org/>

Contact



Direction du numérique pour l'éducation
(Ministère de l'Éducation nationale et de la Jeunesse)
Bureau du soutien à l'innovation numérique
et à la recherche appliquée (DNE TN2)

@Edu_Num

 Licence
Ouverte 2.0