

Nom du comité : Organisation des Nations unies pour l'éducation, la science et la culture (UNESCO)

Problématique : L'Établissement de lignes directrices éthiques pour l'utilisation de logiciels d'analyse du comportement de l'IA dans le secteur de l'éducation

Nom des présidents : Diego Saugnac, Gemma Elena Joyce

L'Établissement de lignes directrices éthiques pour l'utilisation de logiciels d'analyse du comportement de l'IA dans le secteur de l'éducation

1. Introduction

L'utilisation de l'IA dans l'éducation est le sujet de recherches depuis une trentaine d'années et porte actuellement ses fruits avec la mise en place de systèmes de reconnaissance de comportement basés sur l'IA. Aujourd'hui, la mise en œuvre des outils IA au sein du système d'éducation augmente de façon spectaculaire avec l'essor de la numérisation. Il est maintenant évident que le secteur est en train de se moderniser, s'éloignant des méthodes éducatives conventionnelles, ce qui a ouvert de nouvelles opportunités pour l'introduction de logiciels IA dans ce secteur. Cependant, cette modernisation n'est viable que si des directives légales sont imposées à la technologie, afin d'éviter un dilemme éthique.

a. Mots clés

i. Intelligence artificielle (IA) = l'intelligence démontrée par un programme informatique qui imite l'intelligence humaine

ii. Apprentissage Adaptatif Intelligent (AIA) = une méthode d'enseignement qui utilise des algorithmes informatiques pour interagir avec l'élève et offrir des ressources et des activités d'apprentissage personnalisées pour répondre aux besoins uniques de chaque élève

iii. EMFACS (Emotion Facial Action Coding System) = un algorithme de reconnaissance faciale de IA conçu pour reconnaître les émotions humaines à partir des mouvements faciaux

iv. Système de Tutorat Intelligents (ITS) = un système informatique basé sur l'IA qui fournit un soutien ou un retour d'information immédiat et personnalisé aux étudiants sans l'intervention de tuteurs humains

v. biais discriminatoires de l'IA = un biais de production qui peut survenir en raison de l'exposition du système d'IA à des données discriminatoires

vi. modélisation cognitive = un domaine de l'informatique qui traite de la simulation de la résolution de problèmes humains et du traitement mental dans un modèle informatique

vii. modèle de l'apprenant = un modèle éducatif qui se concentre sur l'apprentissage autonome, donnant à l'étudiant une plus de control de son éducation

viii. modèle pédagogique = un modèle d'éducation ou le tuteur est principal contributeur à l'apprentissage de l'élève

ix. vision d'ordinateur = le domaine d'études entourant la façon dont les ordinateurs voient et comprennent les images et les vidéos numériques

x. système d'apprentissage personnalisé = un système éducatif dans lequel les technologies IA personnalisent le programme d'études et les ressources en fonction des besoins des élèves

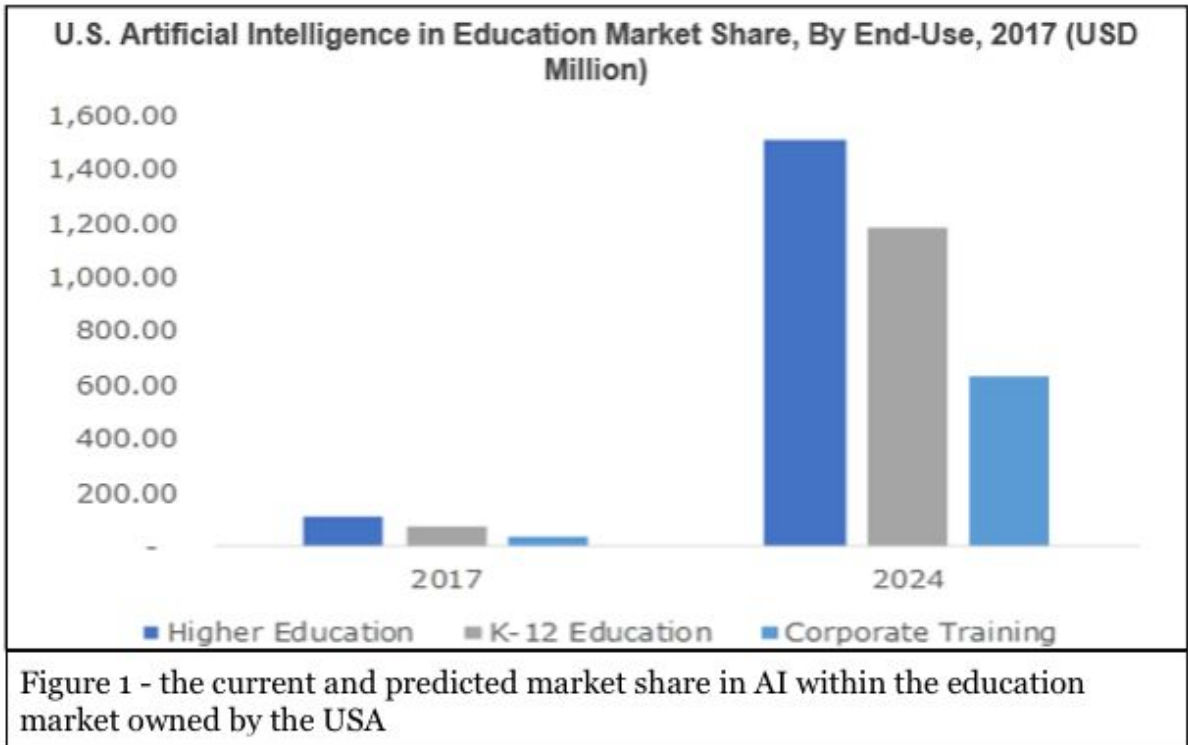
2. Vue d'ensemble de la question

b. Histoire et origines

Comme le concept d'intelligence artificielle (IA) est relativement nouveau, il existe peu de directives éthiques établies qui restreignent le développement et la mise en œuvre des technologies d'analyse comportementale de l'IA. Cette limitation centrale s'est traduite par l'absence d'une législation appropriée

pour le régler. À mesure que les algorithmes et processus clés sont optimisés, tels que le traitement du langage, le raisonnement, la planification, la modélisation cognitive, la synthèse du contenu des leçons, l'enseignement dispensé au moyen d'appareils mobiles ou l'enseignement efficace au moyen de systèmes de tutorat intelligents (ITS), les technologies basées sur l'IA risquent d'affecter considérablement le secteur éducatif si leur application à grande échelle est étendue. La présence économique des systèmes basés sur l'IA dans le secteur de l'éducation est segmentée en fonction de l'utilisateur final et du modèle éducatif, qui comprend le modèle de l'apprenant, le modèle pédagogique et le modèle du domaine. Dans ce contexte, il est divisée en secteurs de l'enseignement supérieur, primaire et secondaire. Au fur et à mesure, comme leur influence s'accroît, des fonds plus importants seront investis pour promouvoir le développement de logiciels d'analyse comportementale basés sur l'IA dans divers domaines, notamment dans le secteur de l'éducation, où ils promettent de transformer le système standardisé.

Ces dernières années, on a assisté à une augmentation de l'utilisation des systèmes d'IA dans le secteur de l'éducation. Avec l'introduction de l'apprentissage adaptatif intelligent (AIA), il y a eu des développements technologiques rapides et des niveaux de collaboration sans précédent entre les fournisseurs et les établissements, ce qui a permis la modernisation la plus rapide de l'éducation à ce jour. Actuellement, les principaux facteurs limitant la croissance des systèmes de reconnaissance comportementale basés sur l'IA dans l'éducation sont les contraintes budgétaires, les méthodes d'éducation traditionnelles, la grande fiabilité du travail humain, les implications éthiques et le manque de confiance dans ces systèmes. Toutefois, ces facteurs sont souvent ignorés, car les institutions et l'industrie considèrent la perspective de l'IA comme une occasion d'accroître leur présence sur le marché. Comme l'illustre le graphique ci-dessous, l'importance de l'IA sur le marché de l'éducation est appelée à augmenter considérablement.



c. Impact des logiciels d'analyse comportementale de l'IA sur l'économie, la société et le gouvernement

Ce type de technologie, lorsqu'elle est appliquée dans le secteur de l'éducation, visait à l'origine à permettre l'adaptation du programme scolaire et des méthodes d'enseignement, afin d'adapter l'expérience éducative de l'élève et de permettre un apprentissage plus efficace. Ici, les systèmes d'analyse comportementale de l'IA peuvent être utilisés pour interpréter les expressions faciales comme des signaux pour différentes émotions, comme une prédiction du comportement, ou comme une indication que l'élève comprend une tâche. Lorsqu'ils sont pris en compte, ces facteurs peuvent illustrer les besoins d'un élève, ce qui permet d'adapter le programme scolaire à son style d'apprentissage, de minimiser les files d'attente pour les comportements perturbateurs et de réduire les taux d'abandon scolaire dans les écoles et collèges. De plus, le logiciel sera en mesure de fournir instantanément un feedback ciblé et personnalisé aux étudiants, ce qui s'est avéré être un élément clé d'un apprentissage réussi. Le "Emotion Facial Action Coding System"

(EMFACS) est l'une des principales ressources utilisées par les développeurs de ce logiciel, et les institutions utilisent cet algorithme dans les systèmes de reconnaissance du comportement pour évaluer les candidats potentiels pour une application scolaire basée sur la reconnaissance faciale.

En raison de leur capacité à accroître l'efficacité de l'éducation, les experts prédisent que ces systèmes vont avoir un impact important sur l'économie s'ils sont mis en œuvre à plus grande échelle, car moins de fonds devront être investis dans les ressources et méthodes d'apprentissage, qui peuvent se révéler inefficaces pour enseigner le curriculum. Cependant, la technologie basée sur l'IA elle-même peut entraîner des coûts élevés, avec le fardeau supplémentaire de financer les ressources nécessaires pour adapter le programme à l'apprenant individuel.

Outre son impact sur l'économie, cette technologie devrait avoir des effets sociaux importants à petite et à grande échelle. Premièrement, les systèmes de reconnaissance comportementale basés sur l'intelligence artificielle peuvent faciliter l'intégration des enfants scolarisés à la maison dans un programme et une atmosphère scolaire normale en permettant une simulation de classe virtuelle et des programmes personnalisés adaptés aux compétences de l'élève.

Un autre effet potentiel de ces systèmes est la transformation des écoles en centres d'apprentissage spécialisés décentralisés, étant donné qu'un plus grand nombre d'élèves suivent un programme personnalisé, ce qui réduit la pression sur le gouvernement pour financer les écoles. Bien que cette transition puisse désavantager les étudiants dans certains domaines, elle permet à l'État d'investir l'argent des impôts dans d'autres domaines, favorisant ainsi le développement. Un désavantage social supplémentaire qui pourrait découler de l'implémentation de ces technologies c'est la possibilité que les élèves ne bénéficient pas de compétences précieuses acquises auprès des enseignants, comme la résolution de problèmes et la pensée critique, ce qui déduit de l'expérience scolaire d'un élève. Actuellement, les systèmes d'IA ne peuvent pas reproduire l'individualité des enseignants dans leur comportement envers les élèves, ce qui signifie que les élèves sont moins susceptibles de refléter des comportements individuels comme la résilience ou des réactions émotionnelles appropriées. En outre, les programmes d'études personnalisés, qui sont le résultat de systèmes d'analyse de comportement basés sur l'IA, peuvent accroître les inégalités entre les enfants ayant des

capacités d'apprentissage différentes, un effet qui est réduit par le système standardisé actuel. En outre, la décentralisation de l'éducation devrait réduire la socialisation entre les élèves, ce qui suscite des préoccupations en matière de santé mentale et de développement de compétences sociales telles que la coopération. Avec moins d'interaction entre les apprenants suivant des programmes personnalisés, les enfants sont moins susceptibles d'être influencés par les normes sociales et culturelles, ce qui peut creuser davantage le fossé social entre ces élèves et ceux qui suivent des systèmes éducatifs standardisés.

De plus, les systèmes d'analyse comportementale de l'IA gardent le risque d'erreur lorsqu'ils basent les interprétations comportementales sur les signaux faciaux et, tout comme les systèmes humains, ils sont susceptibles d'ignorer les avertissements comportementaux ou de mal interpréter les signaux. Néanmoins, les experts soutiennent que cette technologie est beaucoup plus fiable que son équivalent actuel, car, dans la plupart des cas, les conclusions auxquelles parvient le logiciel d'IA sont d'abord vérifiées par des humains, et ce sont eux qui prennent la décision finale concernant une intervention. Non seulement cela limite la marge d'erreur pour produire des normes sans précédent, mais cela ouvre aussi des possibilités d'emploi pour les organismes de réglementation humains dans le processus de vérification des données.

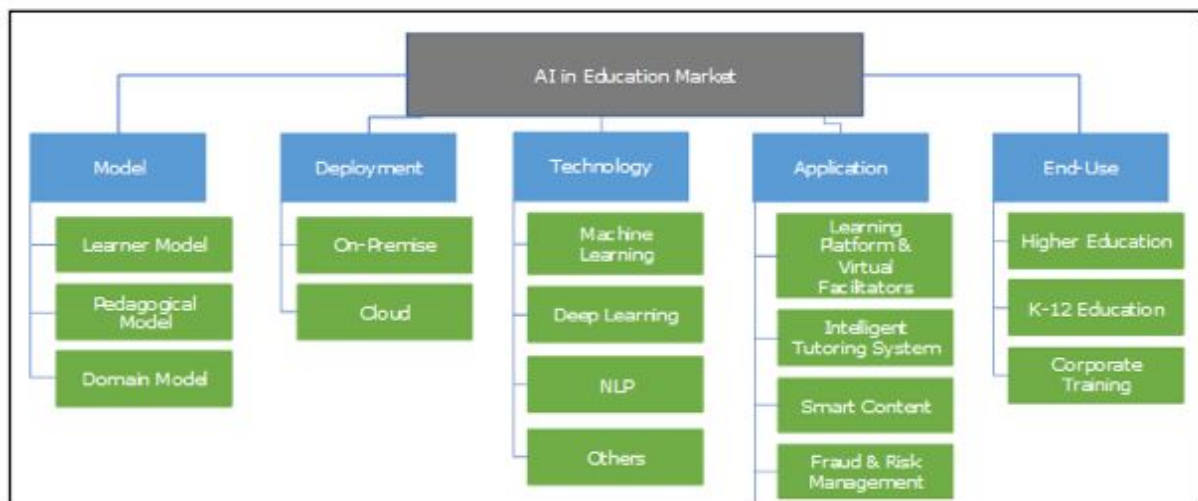


Figure 2 - the categories into which AI within the education sector can be broken down

Pourtant, les logiciels d'analyse comportementale ne sont pas la seule application des systèmes d'IA dans le secteur de l'éducation. L'une des

principales utilisations de cette technologie est d'aider les tuteurs en accomplissant des tâches qui exigent beaucoup de temps et de travail, comme le classement et le traitement des données sur les élèves, ce qui permet aux enseignants de consacrer plus de temps au soutien individuel des élèves.

d. Effets des systèmes d'analyse comportementale de l'IA sur la population générale

Outre leur utilisation dans le secteur de l'éducation, des systèmes d'analyse comportementale basés sur l'IA sont en cours d'introduction dans divers domaines pour répondre à différentes applications, ce qui leur permet d'avoir un impact plus profond sur la population générale. Actuellement, ils commencent à être testés dans des systèmes de vidéosurveillance publique, où le logiciel analyse en permanence les données afin de détecter les comportements agressifs et de déclencher une alarme. Un exemple de la mise en œuvre de cette technologie est celui de la gare du Châtelet, qui fait partie du vaste réseau du métro parisien.

Une autre utilisation viable des logiciels d'IA consiste à identifier les comportements criminels, et ces systèmes sont actuellement introduits dans divers pays afin de prévoir et d'éviter ainsi les actes criminels ou terroristes. Cependant, s'appuyant sur l'apprentissage automatique pour établir une relation entre les caractéristiques du visage et le comportement, ces systèmes peuvent être peu fiables et biaisés, car les biais discriminatoires existants dans la base de données peuvent influencer le résultat.

Récemment, ces technologies ont également été utilisées pour prédire le comportement d'achat humain afin de permettre la conception de publicités plus efficaces. Jusqu'à présent, peu de modèles économiques ont réussi à modéliser avec succès le comportement humain, mais maintenant, avec l'introduction de systèmes de reconnaissance du comportement basés sur l'IA, cela devient possible.

Les systèmes d'analyse du comportement basés sur l'IA, ainsi que les autres applications de l'IA dans le secteur de l'éducation, permettent d'adapter et de décentraliser l'éducation pour mieux adapter les horaires des enfants à leurs

besoins de sommeil, améliorant ainsi leur concentration, leur mémoire et leur santé générale. Si elles sont mises en œuvre à plus grande échelle, elles sont susceptibles de transformer radicalement les modèles sociaux de la population générale, forçant les parents à assumer la responsabilité de l'éducation de leurs enfants avec l'efficacité croissante des systèmes décentralisés de scolarisation et d'enseignement à domicile fonctionnant avec des logiciels basés sur l'IA. Bien que les experts soutiennent la thèse que la participation parentale est bénéfique pour l'enfant, elle empêche les parents de travailler, diminuant ainsi le pourcentage de la population qui est économiquement active, et réduisant le revenu familial. Ces questions ne sont pas les seuls problèmes concernant l'utilisation de l'IA dans le secteur de l'éducation ; aux États-Unis, l'introduction de ces technologies pourrait menacer la sécurité d'emploi de 3,1 millions d'enseignants d'éducation secondaire public et de 0,4 million d'enseignants privés, ainsi que celle de 3,4 emplois de support.

e. Conséquences de l'absence de directives éthiques sur la question

Bien que ces systèmes soient prometteurs et susceptibles d'avoir un impact social positif, ils soulèvent également plusieurs problèmes. Premièrement, les algorithmes qui forment la base des systèmes de reconnaissance faciale se reposent sur l'apprentissage de machine pour filtrer les ensembles de données. Bien que cette méthode se soit avérée efficace pour fournir des informations directement applicables et utiles, elle peut aussi amplifier le rôle des biais discriminatoires dans les ensembles de données. Il s'agit là d'un problème, qui appelle donc à légiférer pour garantir le développement responsable des systèmes d'analyse comportementale de l'IA. Dans le même ordre d'idées, les petits défauts et la faible qualité de résolution des données d'image transmises au logiciel d'IA peuvent les amener à mal interpréter les données, ce qui peut entraîner des interventions inutiles. En outre, en l'absence de directives éthiques concernant la mise en œuvre de l'IA dans le secteur de l'éducation, des questions peuvent se poser concernant l'influence de l'IA sur les tuteurs humains, car les conclusions auxquelles parvient le logiciel peuvent entrer en conflit avec les convictions de l'enseignant, et donc soulever la question de priorité entre le jugement humain et celui de l'IA.

Une autre conséquence de l'absence de lignes directrices éthiques pour ces systèmes c'est la crainte d'une surveillance excessive et de l'atteinte à la vie

privée. Ces systèmes fonctionnent en continu et effectuent une analyse approfondie des images, ce qui peut révéler des détails personnels en violation de la volonté d'une personne. En fait, les préoccupations soulevées à ce sujet ont inspiré Michal Kosinski et Yilun Wang, de l'Université Stanford aux États-Unis, à mettre au point un système conçu pour exposer les dangers de d'un système basé sur la reconnaissance faciale. L'expérience visait à attirer l'attention du public aux risques pour la vie privée associés à la mise en œuvre massive de ce type de technologie. Le système a tenté d'identifier la sexualité de certains individus en se basant uniquement sur les traits du visage d'une personne et sur ses mouvements. Le logiciel a connu du succès avec une marge d'erreur de seulement 20 %, ce qui soulève de sérieuses questions éthiques. Les deux étudiants ont affirmé que ces systèmes violeraient le droit à la vie privée.

f. Répercussions positives potentielles de la résolution du problème concernant l'économie, la société et le gouvernement

Qu'il s'agisse d'implants conçus pour améliorer la reconnaissance faciale ou de prothèses sophistiquées qui assurent la mobilité des patients handicapés, les technologies basées sur l'IA sont répandues dans de nombreux domaines et promettent des changements sociaux, économiques et politiques révolutionnaires. Si, avec la collaboration des branches de l'ONU, des États membres, des ONG et des organisations financées par les gouvernements, des directives éthiques pour l'utilisation des systèmes d'analyse comportementale de l'IA dans le secteur de l'éducation sont élaborées et programmées dans les systèmes d'IA, tout le potentiel de la technologie pourrait être exploité afin d'exercer un changement social positif. Non seulement ces systèmes pourraient transformer les modèles d'éducation standardisés actuels avec des techniques modernes, mais ils pourraient aussi modifier complètement la structure du système, grâce à la mise en œuvre de programmes d'apprentissage personnalisés. Grâce aux systèmes de tutorat intelligents basés sur l'analyse du comportement de l'intelligence artificielle, l'étudiant peut bénéficier d'un soutien immédiat et d'un retour d'information personnalisé pour une efficacité maximale afin d'optimiser l'apprentissage et améliorer le niveau général de formation.

3. Études de cas:

g. Les lignes directrices éthiques de l'UE concernant l'IA

Bien qu'elles ne visent pas spécifiquement les logiciels d'analyse comportementale de l'IA, les directives éthiques établies par l'Union européenne (UE) en avril 2019 fournissent une base pour la législation à l'avenir. Selon ces lignes directrices, pour être digne de confiance, l'IA doit être légale, elle doit respecter toutes les réglementations applicables, et elle doit respecter les valeurs éthiques, d'un point de vue technique et social. Au total, 7 exigences ont été proposées, dans le but de fournir un aperçu éthique applicable à des domaines différents. Premièrement, l'UE déclare que les systèmes d'IA doivent donner du pouvoir aux êtres humains en respectant les lignes directrices établies par les Nations unies en matière de droits de l'homme, et des mécanismes de surveillance efficaces doivent être mis en place pour contrôler ces systèmes. En outre, ces technologies doivent être résilientes et la sécurité doit être assurée par un mécanisme de soutien, afin d'éviter des dommages involontaires aux humains. Un autre point soulevé a été l'importance de protéger les données et d'assurer des mécanismes adéquats de gouvernance des données afin de protéger la vie privée et de réglementer l'accès aux données. Dans le même ordre d'idées, le développeur du système d'IA doit faire preuve de transparence en ce qui concerne ses données, les capacités et les limites de ses logiciels, et son modèle d'affaires, qui peut être facilité par l'utilisation de mécanismes de traçabilité. En outre, il faut éviter les préjugés discriminatoires en encourageant les concepteurs d'IA à favoriser la diversité dans leurs systèmes, en permettant à la population générale d'y avoir accès et de détenir des parts dans la technologie. En outre, selon l'UE, les systèmes d'IA doivent profiter aux humains des générations actuelles et futures et prendre en compte l'environnement vivant et non vivant. Enfin, des mécanismes doivent être mis en place pour assurer la responsabilisation à l'égard des technologies d'IA et de leurs résultats, afin d'éviter une gestion inadéquate des systèmes et, partant, un dilemme éthique.

h. Une enquête de ProPublica révèle des biais discriminatoires dans les outils d'évaluation de la récidive

En 2016, ProPublica a identifié des biais discriminatoires dans les données produites par un système de prévision du comportement d'IA utilisé pour prévoir la récidive en prison. En analysant les conclusions auxquelles est parvenu le système, ils ont constaté qu'il était biaisé à l'encontre des accusés noirs, ce qui, depuis sa proclamation, a soulevé diverses questions éthiques. Des systèmes tels que cette technologie basée sur l'intelligence artificielle sont couramment utilisés dans les salles d'audience aux États-Unis et dans d'autres pays du monde. En tant qu'outil d'aide à la décision sur les personnes qui peuvent être libérées à tous les stades du système judiciaire, les conclusions de ces systèmes de prédiction du comportement sont tirées en analysant les caractéristiques du visage du condamné à l'aide d'un logiciel de reconnaissance faciale et, à partir des connaissances stockées dans sa base de données, le système d'IA utilise un apprentissage automatique pour déterminer le résultat le plus probable. Au sein du système judiciaire, ce logiciel n'est qu'un des facteurs qui influencent la décision d'un juge concernant un condamné, et il est accompagné d'une évaluation des besoins de réadaptation de l'accusé, du jugement du juge lui-même, ainsi que d'autres mesures selon le pays où se tient le procès. Ce n'est toutefois pas le cas des systèmes d'analyse du comportement basés sur l'IA utilisés pour d'autres applications et, pour l'instant, rares sont ceux qui disposent de mécanismes de sauvegarde pour s'assurer que les conclusions auxquelles parvient le système sont exactes. Par conséquent, cette étude de cas peut être utilisée pour illustrer l'importance d'assurer des systèmes de vérification pour les systèmes d'IA dans le secteur de l'éducation afin d'éviter les biais discriminatoires basés sur leur prévalence dans les ensembles de données.

i. Les chercheurs d'IBM forment les systèmes d'IA à suivre les directives éthiques

L'un des principaux défis auxquels nous sommes actuellement confrontés en ce qui concerne les technologies d'intelligence artificielle est celui des entreprises qui maintiennent des systèmes éthiques tout en continuant à attirer l'attention de leurs clients dans des publicités ciblées basées sur l'intelligence artificielle. Pour résoudre ce dilemme, les chercheurs d'IBM ont collaboré avec le MIT Media Lab pour mettre au point une technique de recommandation de l'IA qui optimise les résultats en fonction des préférences

personnelles de l'utilisateur tout en garantissant leur conformité avec les directives éthiques. Récemment, sous la direction de Francesca Rossi, la fonctionnalité du logiciel a été démontrée dans un système de recommandation de films permettant aux parents de définir des contraintes de contenu pour leurs enfants. Les tentatives antérieures d'intégration de ces lignes directrices dans les algorithmes d'IA reposaient sur l'établissement de règles statiques pour le système, qui s'est avéré fonctionnel dans divers contextes, mais qui avait aussi ses limites car les ensembles de données analysées par le système étaient trop vastes, rendant difficile la restriction du système. Chez IBM, ils ont adopté une nouvelle approche, en utilisant l'apprentissage machine pour définir des directives éthiques basées sur des exemples. En programmant le système d'IA en deux étapes, l'une consistant à exposer le système à des exemples définissant les lignes directrices auxquelles le moteur de recommandation doit se conformer, et la seconde consistant à permettre à l'IA d'examiner les exemples par apprentissage machine et de les utiliser comme base pour élaborer ses propres lignes directrices éthiques. En examinant cette technique, il n'est pas difficile de voir le potentiel qu'elle recèle pour établir des lignes directrices éthiques pour l'utilisation des systèmes d'analyse comportementale de l'IA dans différents domaines, dont le secteur de l'éducation.

4. Solutions possibles :

1. Création du Comité des Nations Unies pour l'éthique de l'intelligence artificielle (UNCEAI)

La mise en place d'un nouvel organe des Nations Unies chargé d'établir des directives éthiques pour des secteurs spécifiques et des applications de l'IA est une solution possible à ce problème. Bien que l'UE soit actuellement l'entité chef de file dans l'établissement de ces lignes directrices, elles doivent être appliquées dans les systèmes d'IA en général, ce qui peut devenir problématique car de nombreuses questions spécifiques peuvent se poser selon l'objectif pour lequel le système d'IA est conçu et le secteur où il est utilisé. Avec la participation de tous les États membres de l'ONU, des organisations financées par les gouvernements et des ONG qui peuvent

prouver leur implication actuelle dans le développement de l'IA ou leur intention de poursuivre ce travail à court terme, cette nouvelle branche peut avoir un aperçu des règles fondamentales qui doivent être établies afin de garantir que l'IA respecte les codes moraux. Des réunions régulières pour débattre des résolutions proposées par les membres du Comité, avec des résolutions adoptées après qu'un consensus majoritaire a été atteint, assureront l'efficacité du processus de vote et l'établissement des lignes directrices à court terme.

Comme il s'agit d'une direction générale désignée pour résoudre le problème, il sera efficace d'établir les lignes directrices dans un court laps de temps. En outre, nous pouvons également supposer que, en raison de la diversité de la représentation, de nombreux points de vue différents se rassembleront pour produire un ensemble de règlements viables à l'échelle mondiale pour chaque question particulière. Toutefois, pour mettre en place ce comité, un financement sera nécessaire de la part des Etats membres, des organisations donatrices et des particuliers, ce qui pourrait entraver la création de la branche. De plus, bien que la diversité de la représentation puisse être un atout, elle peut aussi désavantager le processus décisionnel, car les malentendus et les arguments pourraient ralentir la dynamique des discussions.

2. Programmer les systèmes d'IA sur la base de directives éthiques par le biais de l'apprentissage automatique autonome.

Des techniques prometteuses permettant de programmer des systèmes d'IA avec des directives éthiques récemment mises au point par IBM Research ont le potentiel d'accélérer le rythme auquel les technologies basées sur l'IA sont approuvées et commercialisées. Actuellement, l'un des principaux obstacles à la mise en œuvre de systèmes d'analyse comportementale de l'IA dans le secteur de l'éducation est le développement nécessaire pour modifier leurs algorithmes de manière à ce qu'ils fonctionnent sur la base de directives éthiques. Grâce à ce nouveau système, le logiciel est formé en étant exposé à des exemples pertinents de situations dans lesquelles une directive éthique est applicable, ce qui permet à l'IA d'établir de façon autonome ces directives par apprentissage automatique.

Contrairement aux techniques précédentes, qui consistent à programmer l'IA à l'aide de réglages précis, cette approche permet aux systèmes de s'adapter à différentes situations et de réagir en conséquence en se fondant sur les exemples directs, ce qui englobe un éventail plus large de résultats possibles. En outre, cette méthode s'est également avérée plus rapide, car auparavant, les développeurs devaient modifier l'algorithme afin d'exposer la machine à la directive sous différents angles, afin de minimiser les risques que l'IA trouve des failles et désobéisse aux instructions initiales. D'autre part, pour l'instant, la technique est limitée à des situations spécifiques et, à l'avenir, les développeurs espèrent permettre que le même système puisse être appliqué dans différents secteurs.

3. Application des systèmes de vérification pour justifier les conclusions des systèmes d'IA

Des préjugés discriminatoires aux erreurs d'interprétation - L'IA peut être sujette à des erreurs si elle est déclenchée par des tendances dans leur base de données ou des erreurs de données techniques. Par conséquent, il est impératif d'introduire un système de vérification permettant aux humains de juger des résultats atteints par l'IA et de les neutraliser s'ils sont jugés inappropriés. Ce sentiment est partagé par Audrey Azoulay, Directrice générale de l'UNESCO : " Le moment est venu de définir les principes éthiques qui doivent servir de fondement et de cadre à cette perturbation, de faire en sorte que l'IA serve les choix collectifs, fondés sur des valeurs humanistes ". C'est particulièrement vrai dans le secteur de l'éducation où le conflit entre le jugement de l'enseignant et celui du système d'analyse comportementale de l'IA est un scénario plausible.

En permettant aux humains de vérifier les résultats produits par le système d'IA, la probabilité d'erreur ou de biais discriminatoire est minimisée, ce qui permet au logiciel de fonctionner plus efficacement. De plus, le pouvoir que détient l'enseignant sur le système lui permet d'outrepasser sa décision, évitant ainsi les conséquences potentielles d'une erreur fonctionnelle au sein même du système d'IA. Cependant, ce processus de vérification serait coûteux et prendrait beaucoup de temps, selon le milieu dans lequel il est appliqué et l'environnement humain dans lequel le logiciel d'analyse du comportement

d'IA fonctionne. De plus, l'être humain peut abuser de son pouvoir sur la machine en annulant sa conclusion contre les intérêts du développeur d'IA ou contre les directives éthiques sur la base desquelles elle a été programmée.



Figure 3 - the 2019 UNESCO conference focused on designing a strategy for the future of AI

Principaux acteurs internationaux :

i) IBM

IBM participe à la conception de techniques de programmation de logiciels d'IA afin de lui permettre de suivre des directives éthiques basées sur des exemples donnés, permettant ainsi à l'IA de créer de façon autonome les règles à partir desquelles elle fonctionnera.

ii) Microsoft Corporation

Microsoft Corporation est impliquée dans le développement d'outils basés sur l'intelligence artificielle pour le système éducatif, tels que Cortana.

iii) L'UE

L'Union européenne participe à l'établissement de lignes directrices éthiques sur lesquelles les futurs systèmes d'IA pourront s'appuyer.

iv) Les États-Unis

Les États-Unis d'Amérique utilisent des systèmes de prédiction du comportement basés sur l'IA qui dépendent de la reconnaissance faciale dans les salles d'audience pour prédire la probabilité de récidive des personnes condamnées.

v) Arabie Saoudite

L'Arabie saoudite a récemment investi 51,1 milliards d'USD pour développer le système éducatif actuel et 6,4 milliards d'USD supplémentaires pour permettre la modernisation du secteur, ce qui peut être facilité par l'introduction de systèmes d'analyse comportementale de l'IA et d'autres technologies basées sur l'IA.

6. Lignes directrices pour la recherche :

- **Quelles sont les applications des outils d'analyse comportementale de l'IA dans le secteur de l'éducation ?**
- **Quels pays, organisations ou individus investissent dans les outils d'analyse comportementale de l'IA dans le secteur de l'éducation ?**
- **Quelles sont les lignes directrices éthiques déjà en place pour les outils d'analyse comportementale de l'IA et comment peut-on les appliquer au secteur de l'éducation ?**

- **Comment les directives éthiques concernant l'utilisation des outils d'analyse comportementale de l'IA peuvent-elles être appliquées dans le secteur de l'éducation ?**
- **Quelles sont les autres applications des technologies basées sur l'IA dans le secteur de l'éducation ?**

7. Bibliographie :

j. Contenu écrit

<https://en.unesco.org/courier/2018-3/ethical-risks-ai>

<https://www.alliedmarketresearch.com/artificial-intelligence-in-the-education-sector-market>

<https://venturebeat.com/2017/07/23/14-ways-ai-will-impact-the-education-sector/>

<https://www.goldsteinresearch.com/report/artificial-intelligence-in-education-sector-market-outlook-2024-global-opportunity-and-demand-analysis-market-forecast-2016-2024>

<https://www.forbes.com/sites/barbarakurshan/2016/03/10/the-future-of-artificial-intelligence-in-education/#3cb7ebd82e4d>

<https://www.raconteur.net/technology/top-5-sectors-using-artificial-intelligence>

<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

<https://becominghuman.ai/predicting-buying-behavior-using-machine-learning-a-case-study-on-sales-prospecting-part-i-3bf455486e5d>

<https://venturebeat.com/2018/07/16/ibm-researchers-train-ai-to-follow-code-of-ethics/>

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

k. Figures, dessins, et images

Figure 1

<https://www.gminsights.com/industry-analysis/artificial-intelligence-ai-in-education-market>

Figure 2

<https://www.gminsights.com/industry-analysis/artificial-intelligence-ai-in-education-market>

Figure 3

<http://itedgenews.ng/2019/03/07/unesco-conference-artificial-intelligence-urges-human-centred-governance-ai/>