

L'intelligence artificielle et les outils éducatifs numériques pour la prévention du cyberharcèlement

Projet No : 2023-1-ES01-KA220-SCH-000160237

1. Schéma d'évaluation de l'IA dans l'éducation, présentation des tâches et méthodologie

1.1. Introduction et objectif général

Dans le paysage numérique contemporain, caractérisé par une intégration technologique omniprésente, **le déploiement consciencieux de l'intelligence artificielle (IA) dans le secteur éducatif apparaît comme une préoccupation centrale** en termes de compréhension des opportunités, mais aussi avec une attention requise sur les risques associés.

Le projet AI@Mediator, visant à explorer l'opportunité de lutter contre le cyberharcèlement par la mise en œuvre d'un chatbot basé sur l'IA, **se trouve au cœur de cette question complexe**. En tant que membres du Consortium AI@Mediators, nous sommes appelés non seulement à contribuer au débat, mais aussi à fournir des outils et des méthodologies utiles qui peuvent soutenir ceux qui ont commencé plus tard cette réflexion et ce chemin sur la mise en œuvre de l'IA dans le domaine de l'éducation, en particulier auprès de la population vulnérable des mineurs.

Dans le cadre de cette initiative ambitieuse, le développement d'un schéma d'évaluation méticuleux pour l'IA dans l'éducation revêt une signification particulière. **Ce document représente notre perspective, notre moyen de guider** les lecteurs à comprendre si les outils ou pratiques d'IA qu'ils adoptent, ou envisagent d'adopter, ont d'un côté les exigences minimales pour être mis en œuvre avec succès, de l'autre, s'ils respectent les normes éthiques les plus élevées en termes de confidentialité et de valeurs européennes.

1.2. Objectifs spécifiques

Le schéma d'évaluation vise à fournir une référence centrale, pour faciliter l'évaluation éclairée et l'intégration éthique des technologies IA dans les contextes éducatifs. En définissant des paramètres clairs pour une utilisation éthique de l'IA, le schéma alimente l'objectif global du projet de cultiver un environnement en ligne sécurisé et de favoriser la résilience numérique parmi des groupes cibles diversifiés. Cela implique deux sous-objectifs :

- **Suggérer une stratégie d'évaluation pour discerner la dimension théorique des technologies IA dans les domaines éducatifs**, avec un accent prononcé sur la confidentialité des données, les considérations éthiques et l'alignement avec les valeurs et politiques de l'Union européenne. Cette approche est stratégiquement positionnée pour fournir un cadre de déploiement responsable de l'IA, s'inscrivant ainsi dans les objectifs globaux du projet AI@Mediator.

- **Établir un ensemble de critères d'évaluation pour mesurer l'impact des outils et méthodologies IA déployés, selon les trois critères de capacité, d'utilité et d'adoption.** Cela vise à contribuer à comprendre si l'approche déployée est la plus appropriée en termes de coût d'opportunité, d'exigences, de besoins du groupe cible, de cadre et d'environnement social/culturel.

2. Évaluation de la dimension éthique

Cette analyse complète explore les risques et opportunités associés à trois dimensions : la confidentialité des données, l'éthique et les recommandations de l'UE. Cette activité vise à souligner leur rôle central dans la mise en œuvre réussie des initiatives d'IA, en particulier dans le domaine de la prévention du cyberharcèlement. En examinant l'interaction entre la confidentialité des données, l'éthique et les valeurs européennes, les établissements d'enseignement peuvent identifier efficacement les risques, déployer des mesures d'atténuation et favoriser une culture d'utilisation responsable de l'IA.

Il est important de comprendre que toute évaluation éthique peut être affectée par des biais personnels ou culturels, ce qui peut entraîner des conclusions partiellement inadéquates. Ainsi, toute évaluation éthique doit être bien détaillée, encadrée et commentée pour être significative, et toute analyse doit être mise à jour et actualisée périodiquement en fonction de l'évolution de la situation et des besoins.

2.1. Confidentialité des données

Risques majeurs : L'un des principaux risques associés à la confidentialité des données et à l'IA est le potentiel d'accès non autorisé ou de partage non intentionnel ou de fuite d'informations sensibles, entraînant des violations de la confidentialité et même une exposition personnelle des mineurs. Des mesures de protection des données inadéquates et un manque de sagesse générale en matière de données pourraient entraîner des situations alarmantes, compromettant les droits à la vie privée des élèves et des éducateurs et affectant même leur vie quotidienne. De plus, une mauvaise gestion des données pourrait violer des exigences réglementaires telles que le Règlement Général sur la Protection des Données (RGPD) ou d'autres lois nationales connexes, exposant les établissements d'enseignement à des répercussions juridiques et financières.

Mesures d'atténuation :

- **Mettre en œuvre des mesures robustes de protection des données :** Les établissements d'enseignement doivent établir des protocoles stricts pour la collecte, le stockage et le traitement des données afin de protéger la vie privée des élèves. Cela comprend le chiffrement, les contrôles d'accès et des audits de sécurité réguliers pour prévenir l'accès non autorisé et les violations.
- **Obtenir le consentement éclairé :** Avant de collecter toute donnée d'élève, les établissements doivent obtenir le consentement explicite des élèves ou de leurs tuteurs, décrivant l'objectif et la portée de l'utilisation des données.
- **Minimiser la rétention des données :** Pour atténuer le risque d'exposition des données, les établissements doivent adopter une politique de rétention minimale des données, ne conservant les données que le temps nécessaire pour atteindre les objectifs éducatifs.

Questions d'investigation suggérées :

- Comment les données des élèves sont-elles collectées, stockées et traitées dans l'outil ou la pratique éducative basée sur l'IA ?
- Des mesures de sécurité adéquates sont-elles en place pour protéger les données des élèves contre tout accès non autorisé ou toute violation ?
- Le consentement explicite est-il obtenu des élèves ou de leurs tuteurs avant de collecter et d'utiliser leurs données ?
- Quelles procédures sont en place pour la rétention et la suppression des données afin de minimiser le risque d'exposition ?
- Comment l'outil ou la pratique d'IA assure-t-il l'anonymisation des données et la préservation de la vie privée tout au long de son cycle de vie ?

2.2. Éthique

Risques majeurs : Sur le plan éthique, l'utilisation de l'IA dans l'éducation soulève des préoccupations importantes concernant l'équité, la transparence et la responsabilité. Les algorithmes d'IA peuvent présenter des biais ou des résultats discriminatoires, perpétuant les inégalités et renforçant les dynamiques de pouvoir existantes (sujets très sensibles si l'on se réfère aux dynamiques de harcèlement). De plus, l'opacité des processus de prise de décision de l'IA peut éroder la confiance et réduire l'agentivité humaine, notamment dans des domaines sensibles tels que l'évaluation et le soutien aux élèves.

Mesures d'atténuation :

- **Évaluer les biais :** Avant de déployer des systèmes d'IA, les établissements d'enseignement doivent effectuer des évaluations approfondies pour identifier et atténuer les biais dans les algorithmes et les sources de données.
- **Améliorer la transparence :** Les établissements doivent prioriser la transparence dans les processus de prise de décision de l'IA, en fournissant des explications claires sur le fonctionnement des algorithmes.
- **Établir des comités d'examen éthique :** Les établissements d'enseignement devraient établir des comités d'examen interdisciplinaires chargés d'évaluer les implications éthiques des initiatives d'IA.

Questions d'investigation suggérées :

- Comment l'outil ou la pratique éducative basée sur l'IA aborde-t-il les biais potentiels dans ses algorithmes et ses sources de données ?
- Y a-t-il une transparence concernant les processus de prise de décision de l'outil IA, y compris les facteurs influençant ses sorties ?
- Des mécanismes sont-ils en place pour assurer l'équité, la responsabilité et la supervision humaine dans la prise de décision pilotée par l'IA ?
- L'outil d'IA a-t-il fait l'objet d'un examen ou d'une évaluation éthique pour évaluer sa conformité avec les normes et directives éthiques ?

2.3. Recommandations de l'UE

Risques majeurs : Du point de vue des recommandations de l'UE, l'utilisation de l'IA dans l'éducation doit s'aligner sur les cadres réglementaires et les directives politiques établis par l'Union européenne. Les recommandations de l'UE soulignent

l'importance de respecter les droits et valeurs fondamentaux, notamment la protection des données, la vie privée et la non-discrimination, ainsi que des aspects plus spécifiques et techniques liés à la réglementation de l'IA attendue (AI-Act). Les établissements d'enseignement sont censés se conformer à des règlements tels que le RGPD et adhérer aux principes éthiques décrits dans des documents tels que les Lignes directrices sur l'éthique pour une IA digne de confiance.

Mesures d'atténuation :

- **Adhérer à la conformité RGPD :** Les établissements d'enseignement doivent assurer la conformité avec le Règlement Général sur la Protection des Données (RGPD) et d'autres lois pertinentes sur la protection des données.
- **Suivre les Lignes directrices européennes sur l'éthique de l'IA :** Les établissements doivent aligner leurs pratiques d'IA avec les Lignes directrices européennes sur l'éthique de l'IA, qui décrivent les principes éthiques et les recommandations pour le développement et le déploiement de l'IA.
- **S'engager dans une formation et une sensibilisation continues :** Les parties prenantes de l'éducation devraient investir dans des programmes de formation et des initiatives de sensibilisation pour s'assurer que le personnel et les élèves sont informés de leurs droits et responsabilités concernant l'utilisation de l'IA.

Questions d'investigation suggérées :

- L'outil ou la pratique éducative basée sur l'IA est-il conforme aux réglementations pertinentes en matière de protection des données telles que le RGPD ?
- Y a-t-il un alignement avec les Lignes directrices européennes sur l'éthique de l'IA dans le développement et le déploiement de l'outil ou de la pratique d'IA ?
- L'établissement s'est-il engagé dans des initiatives de formation et de sensibilisation continues pour assurer la conformité avec les recommandations de l'UE ?

3. Évaluation de la dimension d'impact

Dans le domaine en expansion des solutions basées sur l'IA, l'intégration de l'intelligence artificielle a émergé comme un outil et une stratégie prometteurs pour améliorer le bien-être et la sécurité des élèves, notamment pour lutter contre le problème omniprésent du cyberharcèlement. Alors que les établissements d'enseignement cherchent à exploiter les outils d'IA pour créer des environnements en ligne plus sûrs et promouvoir des expériences d'apprentissage positives, **il devient impératif d'évaluer l'efficacité, l'impact potentiel et les implications éthiques de ces technologies.** À cette fin, le schéma d'évaluation délimite un cadre structuré comprenant **trois composantes clés : la capacité, l'utilité et l'adoption.**

3.1. Capacité

La composante capacité du schéma d'évaluation se concentre sur **l'évaluation de l'efficacité technique des systèmes d'IA déployés dans les environnements éducatifs.** Cela implique d'évaluer la compétence de l'IA à détecter et à traiter efficacement les cas de cyberharcèlement. Les facteurs clés examinés comprennent

l'exactitude, la fiabilité et les performances globales des algorithmes d'IA dans l'identification et l'atténuation des incidents de cyberharcèlement. De plus, **la composante capacité examine la compatibilité des systèmes d'IA avec l'infrastructure et les plateformes éducatives existantes**. Les considérations de compatibilité englobent des aspects tels que l'interopérabilité des données, la scalabilité du système et la facilité de mise en œuvre.

3.2. Utilité

La composante utilité du schéma d'évaluation **explore l'utilité pratique des applications d'IA dans les contextes éducatifs. Cette évaluation se concentre sur l'évaluation des avantages tangibles et des résultats découlant du déploiement des outils d'IA**, notamment dans le domaine de la prévention du cyberharcèlement et du bien-être des élèves. En plus de la prévention du cyberharcèlement, **la composante utilité examine également les impacts éducatifs plus larges facilités par les technologies d'IA**. Cela comprend l'évaluation de leur efficacité à offrir des expériences d'apprentissage personnalisées, à améliorer l'engagement des élèves et à améliorer les performances académiques.

3.3. Adoption

La composante adoption du schéma d'évaluation se concentre sur **l'évaluation de la mesure dans laquelle les technologies d'IA sont adoptées et intégrées dans les pratiques éducatives**. Cette évaluation examine le niveau d'acceptation et d'utilisation des outils d'IA par les éducateurs, les élèves et d'autres parties prenantes au sein de la communauté scolaire. De plus, **la composante adoption mesure également l'alignement avec les valeurs de l'UE, les normes éthiques et les cadres réglementaires régissant la confidentialité et la protection des données**. Cela implique d'évaluer si les applications d'IA adhèrent aux principes d'équité, de transparence et de responsabilité dans leur mise en œuvre et leur utilisation.

4. Activités d'évaluation

DIMENSION THÉORIQUE

Critères d'évaluation de la conformité en matière de confidentialité des données

Activité	Description	Résultat attendu	KPIs
Évaluation des exigences minimales en matière de confidentialité des données	Cette activité implique la réalisation d'une évaluation complète de l'outil ou de la pratique éducative basée sur l'IA pour assurer la conformité avec les exigences minimales de confidentialité des données. Elle comprend l'évaluation des pratiques de collecte, de stockage	Un rapport complet décrivant l'état actuel de la confidentialité des données dans l'outil ou la pratique d'IA, mettant en évidence les domaines de conformité et les lacunes ou risques potentiels.	Pourcentage d'exigences de confidentialité des données satisfaites. Nombre de risques ou vulnérabilités identifiés. Statut de mise en œuvre des mesures recommandées.

	et de traitement des données, ainsi que des mesures de sécurité mises en place pour protéger les données des élèves.		
--	--	--	--

Critères d'évaluation de la conformité éthique

Activité	Description	Résultat attendu	KPIs
Cartographie et compréhension des éléments potentiels de risque éthique	Cette activité implique la cartographie et l'analyse des risques éthiques potentiels associés à l'utilisation de l'IA dans les environnements éducatifs. Elle comprend l'identification des biais potentiels, de la discrimination, du manque de transparence ou d'autres préoccupations éthiques inhérentes à l'outil ou à la pratique éducative basée sur l'IA.	Un rapport complet d'évaluation des risques éthiques mettant en évidence les dilemmes éthiques potentiels, les biais et les risques associés à l'outil ou à la pratique d'IA, ainsi que des recommandations de stratégies d'atténuation.	Nombre de risques ou préoccupations éthiques identifiés. Évaluation de la gravité des risques éthiques identifiés. Statut de mise en œuvre des stratégies d'atténuation recommandées.

Critères d'évaluation des recommandations de l'UE

Activité	Description	Résultat attendu	KPIs
Benchmark entre les solutions mises en œuvre et les dernières recommandations politiques de l'UE	Cette activité consiste à comparer les caractéristiques et fonctionnalités des solutions IA mises en œuvre dans les établissements éducatifs avec les dernières recommandations politiques fournies par l'Union européenne (UE). Elle comprend l'évaluation de l'alignement avec les réglementations, directives et principes éthiques de l'UE.	Un rapport d'analyse comparative mettant en évidence le degré d'alignement entre les solutions IA mises en œuvre et les dernières recommandations politiques de l'UE, identifiant les domaines de conformité et les domaines d'amélioration potentiels.	Pourcentage d'alignement avec les recommandations politiques de l'UE. Identification des domaines où les solutions mises en œuvre s'écartent des recommandations de l'UE. Statut de mise en œuvre des recommandations.

DIMENSION D'IMPACT

Critères d'évaluation de la capacité

Activité	Description	Résultat attendu	KPIs
Évaluation technique des systèmes d'IA	Réaliser une évaluation technique complète des systèmes d'IA déployés dans les environnements éducatifs. Cela comprend	Rapport d'observation décrivant les capacités et limitations techniques des systèmes d'IA, incluant des	Taux de précision des algorithmes d'IA dans la détection des incidents de cyberharcèlement.

	l'évaluation des algorithmes, des taux de précision et des métriques de performance des outils d'IA dans la détection et le traitement des incidents de cyberharcèlement.	recommandations d'amélioration.	Métriques de performance telles que les taux de faux positifs et faux négatifs. Compatibilité avec l'infrastructure éducative existante.
--	---	---------------------------------	--

Critères d'évaluation de l'utilité

Activité	Description	Résultat attendu	KPIs
Évaluation de l'impact des applications d'IA	Évaluer l'utilité pratique et l'impact des applications d'IA dans les contextes éducatifs. Cela implique d'évaluer leur efficacité à promouvoir le bien-être des élèves, à réduire les cas de cyberharcèlement et à améliorer les résultats éducatifs.	Rapport d'évaluation mettant en évidence les avantages tangibles et les résultats découlant du déploiement des outils d'IA, avec des recommandations pour optimiser leur utilité.	Réduction des cas de cyberharcèlement suite à la mise en œuvre des applications d'IA. Amélioration du bien-être et de la sécurité des élèves. Amélioration des performances académiques et de l'engagement.

Critères d'évaluation de l'adoption

Activité	Description	Résultat attendu	KPIs
Enquête sur la perception des parties prenantes	Administrer des enquêtes aux éducateurs, élèves et autres parties prenantes pour évaluer leurs perceptions et leur acceptation des technologies d'IA dans les environnements éducatifs. Cela comprend la mesure des attitudes envers l'adoption de l'IA, l'identification des obstacles à la mise en œuvre et la sollicitation de retours sur l'utilisabilité.	Rapport d'analyse des enquêtes résumant les perceptions des parties prenantes et les recommandations pour promouvoir l'adoption des technologies d'IA.	Niveau d'acceptation et de volonté d'adopter les technologies d'IA parmi les éducateurs et les élèves. Nombre d'obstacles clés identifiés à l'adoption et à la mise en œuvre. Retours sur l'utilisabilité et l'expérience utilisateur des applications d'IA.

5. Conclusion et avertissement

En conclusion, le Schéma d'évaluation de l'IA dans l'éducation, présenté dans ce document, constitue un cadre fondateur essentiel pour évaluer les aspects éthiques, de confidentialité et de conformité des outils et pratiques éducatifs basés sur l'IA.

Bien que le Schéma d'évaluation de l'IA dans l'éducation s'efforce de fournir un cadre complet, il est essentiel de reconnaître la complexité et la diversité inhérentes au paysage éducatif. En reconnaissant qu'aucun schéma unique ne peut pleinement prendre en compte chaque nuance, les établissements sont encouragés à adopter une approche flexible. Ils doivent adapter et intégrer les activités d'évaluation

suggérées pour s'aligner sur leurs contextes uniques, cultures institutionnelles et besoins spécifiques de leurs groupes cibles diversifiés.

Les considérations éthiques sont au cœur de tout schéma d'évaluation, étant donné la nature intrinsèquement qualitative des dilemmes éthiques. Dans le domaine de l'IA dans l'éducation, l'examen éthique exige un examen méticuleux, tenant compte des biais potentiels, des nuances culturelles et des sensibilités évolutives de la société.

Il est crucial de reconnaître que ce document représente une étape initiale dans le parcours en cours du projet AI@Mediators. Alors que les paysages technologiques et réglementaires continuent d'évoluer rapidement, les parties prenantes sont invitées à rester informées et à suivre l'évolution du projet. En restant engagées et réactives aux tendances et développements émergents, les parties prenantes peuvent s'assurer que leurs approches de l'IA dans l'éducation restent pertinentes, efficaces et éthiquement solides.

6. Références

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20(1), 38.

JISC (2022) AI in tertiary Education. A summary of the current state of play.

Martínez-Comesaña, M., Rigueira-Díaz, X., Larrañaga-Janeiro, A., Martínez-Torres, J., Ocaranza-Prado, I., & Kreibel, D. (2023). Impacto de la inteligencia artificial en los métodos de evaluación en la educación primaria y secundaria: revisión sistemática de la literatura. *Revista de Psicodidáctica*, 28(2), 93-103.

UNESCO (2023). ChatGPT and artificial intelligence in higher education.

UNESCO (2023). Guidance for generative AI in education and research.

Van Laar, E., Van Deursen, A. J., Van Dijk, J. A., & De Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in human behavior*, 72, 577-588.

Hinduja, S., & Patchin, J. W. (2015). *Cyberbullying prevention and response: Expert perspectives*. Routledge.

Smith, P. K., & Steffgen, G. (Eds.). (2013). *Cyberbullying through the new media: Findings from an international network*. Psychology Press.

Cénat, J. M., Hébert, M., Blais, M., Lavoie, F., Guerrier, M., & Derivois, D. (2014). Cyberbullying, psychological distress and self-esteem among youth in Quebec schools. *Journal of affective disorders*, 169, 7-9.

Juvonen, J., & Gross, E. F. (Eds.). (2008). Extending the school grounds? Bullying experiences in cyberspace. *Journal of School Health*, 78(9), 496-505.

Patchin, J. W., & Hinduja, S. (2011). *Cyberbullying prevention and response: Expert perspectives*. Routledge.

Belsey, B. (2004). *Cyberbullying: An emerging threat to the always on generation*.

Spears, B., Slee, P., Owens, L., & Johnson, B. (2009). Behind the scenes and screens: Insights into the human dimension of covert and cyberbullying. *Social Inclusion*, 35-47.

Financé par l'Union européenne. Les opinions et points de vue exprimés ne représentent que ceux de l'auteur(s) et ne reflètent pas nécessairement ceux de l'Union européenne ou de l'Agence nationale SEPIE espagnole. Ni l'Union européenne, ni l'Agence nationale SEPIE ne peuvent être tenus responsables.

Projet No : 2023-1-ES01-KA220-SCH-000160237