

AI@MEDIATORS

GUIDE DES PRATIQUES ÉTHIQUES POUR L'UTILISATION DE L'IA DANS L'ÉDUCATION EN EUROPE



CONCEPTION GRAPHIQUE PARA
IDEC S.A.

WWW.AIMEDIATORSPROJECT.EU

[AI.MEDIATORS](https://www.facebook.com/AI.MEDIATORS)



[AI@MEDIATORS](https://www.linkedin.com/company/AI-MEDIATORS)



[AI.MEDIATORS](https://www.instagram.com/AI.MEDIATORS)





À PROPOS DE NOTRE PROJET

L'intelligence artificielle (IA) transforme de nombreux aspects de nos vies, et son impact sur l'éducation est particulièrement significatif, notamment pour les adolescents, qui sont vulnérables à ses risques et à ses avantages. L'IA offre des bénéfices tels que l'accès universel à l'information, les systèmes de tutorat intelligent, les facilitateurs virtuels et l'automatisation des tâches administratives. Ces avantages peuvent améliorer l'éducation, mais il est crucial d'évaluer et de surveiller attentivement les changements que l'IA introduit afin d'éviter des conséquences négatives sur les comportements et la pensée de la population.

Ce guide fournit aux enseignants, apprenants et éducateurs les outils nécessaires pour comprendre les principaux problèmes liés à l'IA. Il souligne l'importance de développer et d'utiliser l'IA de manière éthique et responsable, en accord avec les valeurs démocratiques et les droits de l'homme, ainsi que de prendre des mesures pour maximiser ses avantages tout en minimisant ses risques.

Au cours des deux dernières décennies, les interactions sociales entre les jeunes ont été influencées par l'utilisation d'Internet, notamment via les réseaux sociaux, en raison de la profonde transformation des télécommunications au 21^e siècle. L'accès à Internet et aux plateformes telles que Facebook, Instagram et TikTok a créé de nouvelles formes de socialisation, qui peuvent reproduire des comportements violents traditionnels, tels que le harcèlement et le sexisme, ou engendrer de nouveaux comportements nuisibles, comme l'addiction à Internet. Ces dynamiques sont médiées par de nouvelles formes de communication et la structure des réseaux sociaux.

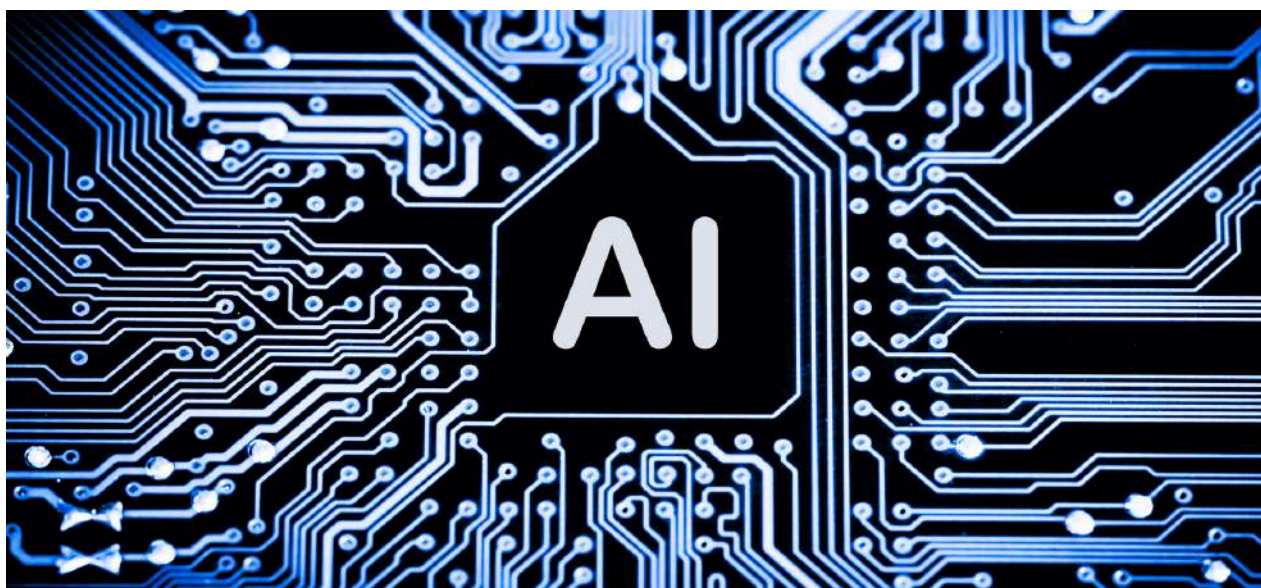
Le cyberharcèlement sur les réseaux sociaux est défini comme un comportement délibéré et hostile, où un individu ou un groupe utilise les réseaux sociaux pour communiquer à plusieurs reprises un contenu agressif dans l'intention de causer du tort ou un malaise à une personne cible. Ce type de harcèlement est souvent associé à une utilisation problématique d'Internet. Les victimes de cyberharcèlement peuvent présenter des symptômes tels que la dépression, l'insomnie, l'anxiété générale et sociale, une faible estime de soi, un sentiment de solitude, une insatisfaction vis-à-vis de la vie, l'absentéisme scolaire, le manque de concentration, des sentiments de honte et de culpabilité, des idées suicidaires et des comportements antisociaux, entre autres.

Les auteurs du cyberharcèlement ont également des niveaux significativement plus élevés d'insatisfaction vis-à-vis de la vie, une faible estime de soi, de la dépression, de la paranoïa, de l'anxiété, du psychotisme, de la sensibilité interpersonnelle, de l'hostilité, des idées suicidaires, de la consommation de drogues et des activités criminelles, comparés à ceux qui ne sont pas impliqués dans le cyberharcèlement. Les comportements de cyberharcèlement peuvent se manifester de différentes manières. Directement, ils prennent la forme d'insultes, de menaces, de chantage, de harcèlement sexuel ou de lynchage public lorsqu'il implique un grand nombre de personnes. Indirectement, le cyberharcèlement peut inclure la création de rumeurs, la divulgation de contenus sensibles de tiers sans leur consentement ou la privation de la socialisation d'une personne (marginalisation), entre autres formes.

Dans ce contexte, trois figures principales sont identifiées dans le cyberharcèlement : la victime, qui est la personne visée par le comportement violent ; l'agresseur, qui est celui qui adopte le comportement violent ; et les observateurs, qui sont ceux qui témoignent du cyberharcèlement sans y participer directement, agissant comme spectateurs. Ces derniers peuvent être subdivisés en participants (aidant l'agresseur ou la victime) ou non-participants. Il est crucial de noter que ces rôles ne sont pas statiques, car la plupart des personnes impliquées dans des situations de cyberharcèlement peuvent jouer les trois rôles à des moments différents. Quel que soit le rôle assumé, il est possible d'aborder le cyberharcèlement sous différents angles afin de le prévenir et de l'atténuer.

Dans le contexte du cyberharcèlement, l'IA peut être un outil puissant pour lutter contre ce problème. Les algorithmes d'IA peuvent analyser de grands volumes de données en temps réel pour identifier des schémas de comportement abusif, permettant des interventions précoces. De plus, les systèmes d'IA peuvent offrir un soutien aux victimes via des chatbots qui fournissent des conseils et des ressources, ainsi que détecter des signes de détresse émotionnelle dans les messages et les publications. Ces capacités peuvent contribuer à créer un environnement scolaire plus sûr et plus protecteur, contribuant à la lutte contre le cyberharcèlement et à l'amélioration du bien-être des élèves.

Ci-dessous se trouve un ensemble de principes avec des questions directrices qui peuvent vous aider à identifier les limites et les points forts si vous avez besoin d'utiliser l'intelligence artificielle dans votre école. Rappelez-vous que ces questions ne sont pas définitives, mais une aide à la bonne pratique éthique que vous devriez adapter aux circonstances de votre école.



PRÉVENTION DES PRÉJUDICES

Les élèves dépendant d'un usage d'intelligence artificielle pour lutter contre le cyberharcèlement ne doivent pas causer de préjudices (psychologiques, physiques, sociaux, économiques, etc.) Aux élèves, aux parents, aux enseignants ou à toute autre personne utilisant le système. C'est pourquoi les risques possibles de son utilisation doivent être éliminés ou minimisés.

• Les enseignants ont-ils discuté des préjudices à court terme que l'utilisation de l'IA peut avoir ? Et à long terme ?

- Son utilisation interfère-t-elle avec une activité importante pour le développement de l'élève ?
- Peut-elle mettre en danger la santé physique ou mentale de l'individu et ses relations interpersonnelles ?
- Savons-nous dans quelle mesure l'application peut faire des erreurs/hallucinations ?
- L'utilisation du système d'IA conduit-elle les adolescents à se sentir démotivés dans leur apprentissage / dans leur utilisation des outils numériques / à être moins impliqués dans les relations sociales ?

Exemple pratique : pour lutter contre le cyberharcèlement, on peut recourir à des programmes qui utilisent des groupes de discussion où des insultes sont échangées entre élèves. Les enseignants doivent en tenir compte à l'avance afin d'évaluer comment y faire face. Il convient également d'observer, à court et long terme, comment ces dispositifs influencent le comportement des individus, notamment car certaines applications mobiles, avec ou sans IA, peuvent générer de la dépendance et ne sont donc pas recommandées pour lutter contre le cyberharcèlement.

PRINCIPE DE BIEN-ÊTRE

L'application d'un outil d'IA ne devrait pas seulement signifier l'absence de préjudice, mais aussi l'amélioration de la condition humaine, qu'il s'agisse de développer l'aspect physique, mental, social et/ou émotionnel. La lutte contre le cyberharcèlement elle-même justifie ce principe, mais nous devons être conscients de la réalisation des objectifs proposés :

- L'application de l'IA vise-t-elle un aspect du bien-être des élèves ? Y parvient-elle ?
- L'application améliore-t-elle les relations interpersonnelles de l'élève ?
- Dans le cas d'une évaluation coûts-bénéfices, l'utilisation de l'application en vaut-elle la peine pour le bien-être de l'apprenant, et le coût représente-t-il un préjudice significatif pour l'intégrité de l'individu ?
- En termes de temps et d'effort d'utilisation du programme d'IA, est-il aussi efficace que possible dans ses bénéfices pour l'individu et pour la société ?
- Le système d'IA favorise-t-il une communication claire et non violente entre les élèves et les adultes (enseignants, personnel, parents) ?

Exemple pratique : l'école pourrait mettre en place un chatbot alimenté par l'IA qui offre un soutien émotionnel et des stratégies d'adaptation aux élèves. En collaboration avec les psychologues de l'Éducation nationale.



RESPECT DE L'AUTONOMIE HUMAINE

Il s'agit de maintenir la capacité des individus à agir inconditionnellement sur leurs propres actions, sans pression extérieure, afin d'être responsables de leur comportement.

- L'application a-t-elle des éléments addictifs ?
- Le consentement éclairé a-t-il été recueilli auprès des personnes concernées ?
- L'utilisation de l'application d'IA est-elle obligatoire pour l'apprenant, et existe-t-il une alternative ?
- Les élèves sentent-ils qu'ils peuvent interagir librement avec l'application ou se sentent-ils observés ?
- Les apprenants ont-ils une personne de référence lorsqu'ils ont des questions sur l'utilisation de l'application d'IA ?

Exemple pratique : avant de mettre en place un système d'IA pour combattre le cyberharcèlement, l'école devrait obtenir le consentement éclairé des élèves et des parents, en expliquant son objectif et son utilisation. L'IA devrait être conçue pour être non intrusive, en s'assurant que les élèves peuvent se désinscrire s'ils se sentent mal à l'aise, respectant ainsi leur autonomie.

NON-DISCRIMINATION

C'est l'obligation morale de s'assurer que les systèmes d'IA pour lutter contre le cyberharcèlement ne discriminent pas les mineurs en créant une situation de désavantage sur la base (entre autres) de : race, nationalité, genre, orientation sexuelle, religion ou convictions politiques. Une attention particulière doit également être accordée aux besoins spécifiques que certains élèves peuvent avoir pour participer.

- Peut-il être utilisé par tous les apprenants avec quasiment aucune barrière technologique ?
- Peut-il être adapté aux élèves handicapés ou ayant des besoins particuliers ?
- Les avantages de l'IA sont-ils équitablement répartis entre tous les élèves qui l'utilisent ?
- L'IA a-t-elle été entraînée de manière non biaisée afin de ne pas discriminer un groupe ?
- Existe-t-il des mécanismes pour détecter les biais et discriminations possibles dans l'utilisation ou les résultats de l'IA ?

Exemple pratique : l'application doit être traduite dans le plus grand nombre de langues possible pour garantir l'inclusion et l'accès de tous les élèves à l'application, sans distinction.



RESPONSABILITÉ PROACTIVE

TRANSPARENCE

Une plus grande transparence contribue à des sociétés plus pacifiques, justes, démocratiques et inclusives en permettant le contrôle public et en réduisant la corruption et la discrimination. Elle repose sur le principe que les utilisateurs doivent être informés clairement et de manière compréhensible de l'utilisation de l'IA et de tout ce qu'elle implique.

- L'application d'IA montre-t-elle facilement qu'elle utilise l'IA et comment elle fonctionne ?
- Les élèves, les parents et les enseignants savent-ils quelles données sont collectées et à quel fin ?
- Toutes les personnes concernées connaissent-elles les risques et les limites de l'utilisation de l'IA ?
- Quels mécanismes l'école met-elle en place pour vérifier que le système d'IA fonctionne correctement ?
- Les enseignants et/ou les représentants des élèves savent-ils pourquoi l'IA donne ces résultats ?

Exemple pratique : les élèves sont informés, avant d'utiliser l'IA, de son fonctionnement, de ses avantages, de ses risques et de ses limites. Cette démonstration d'honnêteté est déjà une façon d'éduquer les élèves aux valeurs et de renforcer la confiance. Plus la confiance dans un tel programme d'IA est grande, plus il est utilisé et plus il est efficace dans la lutte contre le cyberharcèlement.

Il s'agit de l'engagement éthique requis des acteurs sociaux impliqués dans le développement, la distribution et la mise en œuvre de ces applications. Ces agents sociaux doivent anticiper afin de réduire ou d'éliminer tout risque que ces technologies peuvent poser pour les adolescents, car une application d'IA seule ne peut pas comprendre les biais ou les différences de pouvoir qu'elle peut engendrer et peut, par inadvertance, favoriser des situations de cyberharcèlement.

- Y a-t-il une équipe technique chargée du stockage, de la sécurité et du traitement des données des élèves ?
- Un formulaire de consentement éclairé a-t-il été élaboré pour les parents et les élèves, expliquant les avantages, les risques et les limites de l'IA ?
- Si un logiciel externe est utilisé, la fiabilité du logiciel est-elle connue ?
- Y a-t-il des réunions régulières pour évaluer si l'application répond aux exigences minimales ou si des problèmes ont été identifiés ?
- Y a-t-il une personne légalement responsable en cas d'incident ?

Exemple pratique : les écoles organisent des réunions mensuelles pour examiner l'impact de l'outil d'IA sur le comportement des élèves. Ils discuteraient des problèmes qui se posent et apporteraient les ajustements nécessaires pour s'assurer que l'outil est utilisé de manière éthique et efficace. Cette approche proactive contribue à maintenir un environnement d'apprentissage sûr et favorable.



PARTICIPATION ACTIVE

Les utilisateurs des applications d'IA ne doivent pas être considérés comme de simples agents passifs, mais impliqués (ou représentés) dans la conception, le développement et l'utilisation, afin d'éviter les problèmes liés aux nouveaux scénarios posés par les systèmes d'IA.

- L'application a-t-elle été conçue ou sélectionnée par les élèves en fonction de leurs besoins ?
- Les élèves ont-ils été informés ou formés à tirer le meilleur parti de l'IA ?
- Les enseignants savent-ils comment fonctionne l'application et quel est leur rôle ?
- Les parents savent-ils comment fonctionne l'application ? Ont-ils un rôle ?
- L'application a-t-elle été testée et recommandée par une institution représentant les intérêts des élèves ?

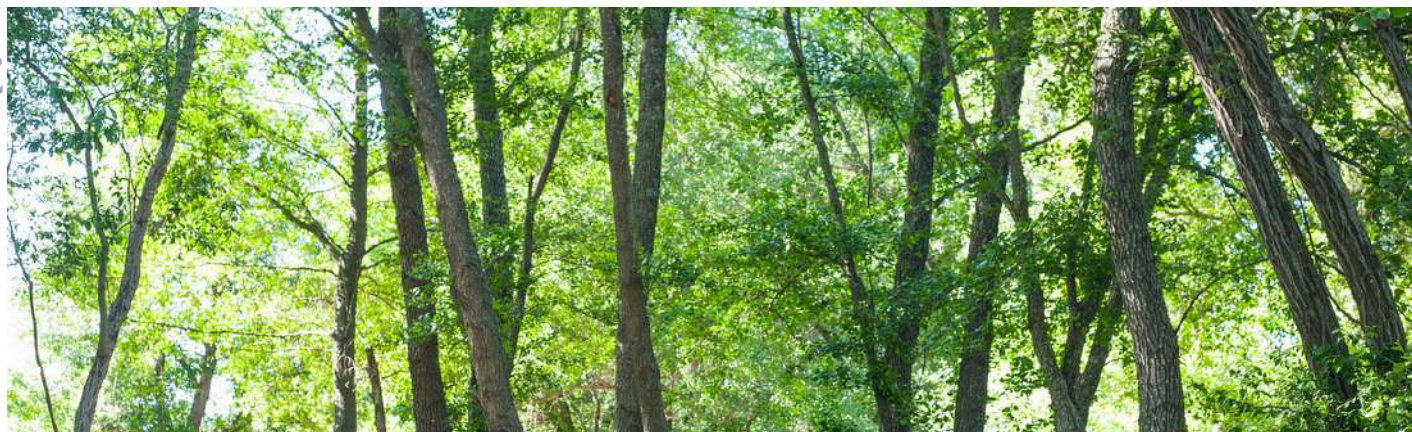
Exemple pratique : il devrait y avoir une communication entre enseignants et élèves pour savoir comment le cyberharcèlement se produit, ses conséquences et pour savoir comment choisir la meilleure application d'IA pour lutter contre le cyberharcèlement.

EXPLICABILITÉ

C'est la capacité à expliquer clairement ce que fait un système d'IA, avec quels types de données, pour quels usages, et avec quels risques possibles. Sans cette compréhension minimale, le consentement de l'utilisateur (élève, parent, enseignant) ne peut pas être véritablement « éclairé », car il ne sait pas ce qu'il accepte. Il ne s'agit pas de donner tous les détails techniques, mais de présenter les points clés de manière compréhensible.

- Quelles données des élèves (texte, voix, image, résultats scolaires, comportement) sont collectées par l'outil, et pourquoi le faire précisément ?
- Comment puis-je expliquer ces usages de données à des élèves de différents âges, avec des mots qu'ils comprennent ?
- Les parents/tuteurs ont-ils reçu une explication claire, et ont-ils la possibilité de refuser ou de limiter l'usage de l'outil ?
- Quels risques principaux identifiés (vie privée, dépendance, stigmatisation via des profils automatisés) dois-je évoquer avec les élèves et les parents ?
- Comment puis-je aider les élèves à repérer eux-mêmes une utilisation problématique de l'IA (par exemple, une recommandation qui leur semble injuste ou étrange) ?

Exemple pratique : avant d'activer un outil IA, vous expliquerez clairement aux élèves : ce que fait l'outil ; le rôle de l'élève ; la manière dont sont exploitées les données ; les risques (erreurs possibles, propositions inadaptées, risque de dépendance à l'outil).



VIE PRIVÉE

C'est la protection des données personnelles des individus, en particulier des enfants et des adolescents, contre la collecte, l'utilisation et le traitement non autorisés. Les outils d'IA doivent non seulement justifier que les données qu'ils collectent sont légitimes, et à des fins spécifiques consenties par l'enfant et le tuteur légal. Ils doivent également garantir qu'ils ne collectent pas de données autres que celles nécessaires à la lutte contre le cyberharcèlement.

- L'IA respecte-t-elle les réglementations en matière de protection des données et d'autres ordonnances similaires ?
- Des politiques claires sont-elles en place pour le stockage, l'accès et le partage des données ?
- Y a-t-il des professionnels techniques au sein de l'institution qui peuvent assumer la responsabilité de prévenir les violations de la sécurité ?
- Les parents, les enseignants et les élèves savent-ils quelles données sont stockées et quelles non ?
- Est-il possible de ne pas collecter les données personnelles des élèves ou de les anonymiser ?

Exemple pratique : les données sont stockées sur un serveur sécurisé par la direction de l'école, qui est maintenu à jour contre d'éventuelles cyberattaques. La divulgation d'informations confidentielles est à éviter, car elle peut conduire à de la discrimination et du harcèlement sur les réseaux sociaux, encourageant ainsi le cyberharcèlement.

PROTECTION DE L'ENVIRONNEMENT

L'utilisation de l'intelligence artificielle implique une dépense énergétique considérable, de sorte que les développeurs et les distributeurs doivent être responsables d'utiliser l'IA de la manière la plus durable et la plus respectueuse de l'environnement possible, en minimisant son impact environnemental, ainsi qu'en maximisant son rapport coût-efficacité.

- L'institution sait-elle quel est le coût énergétique de l'introduction de ce système d'IA et si celui-ci peut être compensé par des économies ailleurs ?
- Y a-t-il un moyen de prolonger la durée de vie ou d'améliorer l'efficacité de l'équipement utilisant l'IA ?
- Les conséquences de l'utilisation de l'IA contribuent-elles à améliorer la société ?
- La société fournissant l'outil démontre-t-elle que l'outil et son utilisation sont respectueux de l'environnement ?

Exemple pratique : vérifier que l'équipement utilisant des systèmes d'IA utilise des outils d'économie d'énergie. Par ailleurs, promouvoir des valeurs plus empathiques, bienveillantes et collectives améliore non seulement la lutte contre le cyberharcèlement mais aussi la sensibilisation à l'environnement.



Co-funded by
the European Union

MERCI

Financé par l'union européenne. Les opinions et points de vue exprimés sur ce site web n'engagent que les auteurs et ne reflètent pas nécessairement ceux de l'union européenne ou du service espagnol pour l'internationalisation de l'éducation (SEPIE). Ni l'union européenne ni l'agence nationale SEPIE ne peuvent être tenues responsables.

Projet no : 2023-1-es01-ka220-sch-000160237



PARTENAIRES



Universidad de Alicante



AI@MEDIATORS

