

UNE INTELLIGENCE ARTIFICIELLE ÉTHIQUE :

Les actions de l'UNESCO

L'Organisation des Nations Unies pour l'éducation, la science et la culture a mis en place un cadre normatif universel pour l'utilisation de l'Intelligence artificielle d'une manière éthique. Un partenariat avec la Commission Européenne a été conclu dans ce cadre et un budget de 4 millions d'euros est alloué à l'application de ces dites recommandations. Le respect de la condition éthique est au centre de nombreuses initiatives menées par l'UNESCO.

Compte tenu de l'utilité qu'elle présente dans différents secteurs y compris la culture et l'éducation, l'intérêt pour l'intelligence artificielle ne cesse de croître. Après une première édition ayant eu lieu en décembre 2022 en République Tchèque, se tient le Deuxième Forum mondial sur l'éthique de l'Intelligence artificielle en Slovénie en février 2024. Ceci s'intègre dans le cadre des efforts menés, depuis plusieurs années, par l'Organisation des Nations Unies pour l'Education, la Science et la Culture (UNESCO) en faveur du développement de cette technologie dans tous les pays membres.

L'éthique comme socle commun

La condition éthique est un élément autour duquel plusieurs efforts sont menés dans l'objectif d'organiser l'usage de cette technologie polyvalente. L'essor que connaît celle-ci et l'aspect diversifié et souvent expérimental peuvent instaurer une utilisation anarchique dépassant les normes éthiques. D'où l'intérêt de la mise en place d'un cadre réglementaire délimitant l'exploitation de l'AI et la faisant répondre à un cadre éthique.

Selon ces recommandations, profiter des avantages majeurs et appliquer cette utilisation à plusieurs secteurs d'activité se fera ainsi au moyen de garde-fous moraux se basant sur des valeurs comme la protection des droits de l'Homme et de la dignité, la transparence et l'équité.

L'initiative de l'UNESCO permet de mettre en place un instrument normatif mondial à la disposition de ses 193 Etats membres. A travers les recommandations qui le composent, les décideurs politiques disposent d'un support pratique pour transformer en actions, les principes fondamentaux en matière de gouvernance des données, environnement et écosystèmes, éducation et recherche, santé et bien-être social.

Ce cadre normatif repose sur quatre valeurs : respecter les droits de l'Hommes et la dignité humaine, vivre dans des sociétés pacifiques justes et indépendantes, assurer la diversité et l'inclusion et veiller à l'existence d'un environnement et des écosystèmes qui prospèrent.

Cette conception se veut dynamique afin de présenter des politiques évolutives au rythme des avancées technologiques, des mutations qu'elles imposent et des conséquences qui en résultent.

Les dix recommandations

Cette approche s'articule autour de dix principes fondamentaux :

- Les principes de proportionnalité et d'innocuité : utilisation pour des buts légitimes en tenant compte des risques liés à l'utilisation.
- La sûreté et la sécurité : Les acteurs de l'IA doivent garantir une utilisation sans préjudices liés à la sûreté ni vulnérabilité liée aux attaques et aux failles dans la sécurité.
- Le droit au respect de la vie privée et protection des données : Des cadres de protection des données et des mécanismes de gouvernance appropriés doivent être mis en place dans les systèmes d'IA.
- La gouvernance et la collaboration multipartites et adaptatives : L'utilisation des données doit se faire dans le respect du droit international et de la souveraineté nationale. Il est, dans ce cadre, recommandé, de faire en sorte que la gouvernance de l'IA s'opère d'une manière inclusive au moyen de la participation des différentes parties prenantes.
- La responsabilité et la redevabilité : Il est préconisé que les systèmes d'IA soient vérifiables et traçables par le biais de mécanismes de surveillance et d'évaluation d'impact.
- La transparence et l'explicabilité : Ces deux éléments doivent être adaptés au contexte pour trouver un équilibre approprié dans le cadre de l'utilisation.
- La surveillance et les décisions humaines : Des personnes physiques ou des entités juridiques doivent porter les responsabilités physiques et juridiques lors de tous les stades du cycle de vie des systèmes d'IA.
- La durabilité : Les technologies de l'IA devront être évaluées continuellement pour répondre aux objectifs de développement durable (ODD) des Nations Unies.

- La sensibilisation et l'éducation : Il est important de garantir au public la compréhension du fonctionnement de cette technologie au moyen d'un engagement civique favorisant l'acquisition des compétences numériques, la formation à l'éthique de l'IA et l'éducation aux médias.
- - L'équité et la non-discrimination : Promouvoir la justice sociale, garantir l'équité, lutter contre les discriminations et veiller à l'inclusivité tels doivent être les engagements des acteurs de l'IA afin que les bénéfices liés à cette technologie soient accessibles à tous.

Un process au profit de tous

L'UNESCO définit en faveur des Etats membres onze domaines d'actions stratégiques dans le cadre desquels la prise en compte des valeurs précitées est essentielle : Economie et emploi, santé et bien-être social, évaluation de l'impact éthique, gouvernance, politique des données, développement et coopération internationale, environnement et écosystème, genre, culture, éducation et recherche, communication et information.

Afin de faire en sorte que les potentialités que présente l'AI en matière d'éducation soient à la portée du plus grand nombre d'utilisateurs, l'UNESCO s'est engagée à aider les Etats membres dans ce sens. Ce support s'opère dans le cadre de l'agenda Education 2030 qui repose sur 17 objectifs de développement durable.

Deux méthodes pratiques ont été mises en place pour garantir la mise en œuvre effective de la recommandation :

- Méthode d'évaluation de l'état de préparation : C'est un process dont les résultats aideront à placer les états membres sur une échelle de préparation à une mise en œuvre éthique et responsable de l'IA. Les résultats auxquels aboutira ce process permettront à l'UNESCO d'adapter les mesures de renforcement mises à la disposition des différents pays.
- L'évaluation de l'impact éthique : Ce processus s'adresse aux équipes de projet d'IA pour qu'en collaboration avec les pays concernés, ils puissent évaluer les impacts des systèmes IA et envisager les préventions utiles.

Parmi les recommandations de l'UNESCO figure l'impératif d'assurer « l'accès de tous à une éducation de qualité, sur un pied d'égalité, et de promouvoir les possibilités d'apprentissage tout au long de la vie ». Cela permet d'asseoir une utilisation équitable de l'IA en faisant de ce progrès technologique un moyen pour annihiler les inégalités en matière d'accès à la connaissance. Inclusion et équité sont les priorités de cette vision se définissant comme humaniste et visant une pratique innovante en matière d'enseignement et d'apprentissage.

Les lignes directrices de cette approche ont été détaillées en marge du Consensus de Beijing (document final de la Conférence internationale sur l'intelligence artificielle), dans le cadre d'une publication intitulée « AI et éducation ; Guide pour les décideurs politiques ».

Une série d'initiatives

Par ailleurs, l'UNESCO a mis en place une plateforme collaborative (Women4Ethical AI) qui se définit comme un réseau de femmes pour une IA éthique. Le but de cette démarche est d'aider les gouvernements et les entreprises à atteindre l'égalité des genres dans la conception et dans le déploiement de cette technologie. Dix-sept expertes en IA mettront en place un référentiel de bonnes pratiques. Cette plateforme se basera sur le développement d'algorithmes non discriminatoires et œuvrera à l'accessibilité de la technologie à des filles, des femmes et des groupes dits sous-représentés. Les dix-sept femmes en charge de cette plateforme viennent de domaines professionnels différents : enseignement supérieur, société civile, secteur privé, organismes de régulation.

En outre, l'UNESCO a lancé une initiative de collaboration entre ses services spécialisés et des entreprises opérant dans le domaine de l'IA en Amérique latine. Il s'agit du Conseil ibéro-américain des entreprises pour l'éthique de l'IA, un cercle constitué comme un lieu d'échange d'expériences et d'exploration de pistes de collaboration. Au centre de ces objectifs : les pratiques éthiques au sein de cette industrie. Ce rassemblement actuellement présidé par Microsoft et Telefonica mettra en œuvre des moyens techniques afin de concevoir et diffuser un outil d'évaluation de l'impact éthique de l'IA. Il entend aussi participer à la mise en place de réglementations régionales dites intelligentes et ce pour favoriser l'implémentation d'un environnement compétitif en matière de technologie mais aussi en matière de responsabilité et d'éthique.

L'UNESCO et la Commission européenne ont signé un accord pour accélérer la mise en œuvre des recommandations en faveur de l'utilisation éthique de l'IA. Ce cadre normatif sera ainsi appliqué dans les pays membres. Trente ont déjà commencé à légiférer sur la base de ces recommandations pour que l'intelligence artificielle respecte les libertés et bénéficie à tous.

Afin d'accompagner les pays à faibles revenus dans le cadre de leurs législations nationales en faveur d'une utilisation éthique de l'IA, un budget de 4 millions d'euros a été alloué. De nombreuses initiatives bénéficieront de financements dans ce cadre. Parmi elles, le projet « Experts en éthique de l'IA sans frontières ». La vocation de ce regroupement est de fournir un appui et des conseils à la demande et de façon adaptée aux politiques publiques pour que les institutions des Etats membres puissent appliquer l'ensemble des recommandations.

Pour que cette pratique puisse être généralisée et pour qu'elle puisse être évolutive, l'UNESCO projette d'organiser un forum mondial pour réunir de manière annuelle les acteurs du domaine de l'IA.