

Charte Intelligence Artificielle IRF ZAAO

I - Préambule

Les outils d'intelligence artificielle (IA) mis à disposition du grand public depuis 2023, dans un mouvement initié par OpenAI avec ChatGPT et depuis repris par de nombreux acteurs, sont de fait entrés dans les usages de nombre d'apprenants et personnels, dans le cadre privé mais aussi dans le cadre professionnel.

La présente charte expose le positionnement de l'IRF ZAAO (section II) et présente un ensemble de consignes générales (section III) puis de consignes spécifiques à l'usage dans le cadre pédagogique (section IV), et dans le cadre des activités administratives et techniques (section V), dans le cadre juridique national, européen et international actuellement en vigueur, à savoir le Règlement européen sur l'intelligence artificielle¹.

II - Positionnement de l'IRF ZAAO

Les IA sont des outils offrant des performances remarquables dans divers usages. Elles doivent à ce titre être vues comme permettant d'assister efficacement les membres de l'IRF ZAAO dans leurs activités et d'obtenir potentiellement des gains de temps et d'efficacité. Elles offrent également un potentiel de développement de l'innovation. Ce constat peut d'ailleurs être élargi au-delà des seules IA génératives.

Toutefois, les capacités de ces outils pourraient être sûrement interprétées et déboucher sur des usages, non pas d'assistance mais de substitution aux démarches intellectuelles attendues des personnels et des stagiaires dans le cadre de leurs missions ou de leurs formations.

Si l'assistance est souhaitable, la substitution ne peut être envisagée ni admise. Il ne s'agit pas là d'une prise de position morale mais bien d'un positionnement pratique et pragmatique dicté par les principes de l'intégrité scientifique. En effet, bien que leurs performances augmentent régulièrement, ces outils restent sujets à des erreurs ou sources de plagiat et doivent être l'objet d'une vigilance constante et soutenue.

Par ailleurs, nombre d'informations, de données, ne sauraient être transmises à ces outils d'IAs hébergés hors de l'IRF et pour lesquelles la confidentialité ne peut être systématiquement garantie. Enfin, **sauf indication contraire**, les travaux réalisés par les stagiaires en vue de l'évaluation des connaissances et compétences acquises ne sauraient être produits par de tels outils, ce qui constituerait un cas de triche avéré et passible de sanctions.

Le monde de la formation est par essence porteur des trois principes que sont la transparence quant à l'utilisation de l'intelligence artificielle, le respect de la confidentialité dans tous les champs et l'éthique d'une manière générale pour l'enseignement.

Les technologies d'IA génératives, ainsi que leurs usages évoluent rapidement. Cette charte a vocation à évoluer en conséquence au cours des années à venir.

¹ Règlement (UE 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissement des règles harmonisées concernant l'intelligence artificielle et modifiant les règlements (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/1139 et (UE) 2019/2144 et les directives 2014/90/UE, (UE) 2016/797 et (UE) 2020/1828 (règlement sur l'intelligence artificielle) Texte présentant de l'intérêt pour l'EEE.

III - Consignes générales



Tout document issu d'une IA générative doit faire l'objet d'une **relecture critique et attentive**, d'une part à cause des risques de plagiat attribuables à la personne qui s'approprierait le contenu généré, d'autre part à cause des risques d'hallucinations (voir annexes). Les propositions, les références, les faits énoncés doivent être vérifiés.

Ne jamais transmettre de **données personnelles²** ni de **documents professionnels**, a fortiori de documents confidentiels. Des données relevant du domaine public peuvent être transmises. Pour tout autre type de document, même sous licence ouverte (par exemple Créative Commons), la transmission est déconseillée (une licence ouverte ne signifie pas nécessairement que tous les droits soient cédés). En cas de doute sur le caractère confidentiel ou public d'une information, il est possible de s'adresser au fournisseur de l'information.

Dans le cas de la **transmission à l'IA de données ou documents rédigés par l'utilisateur**, il est conseillé à ce dernier d'utiliser toutes les mesures à sa disposition pour qu'ils ne soient pas réutilisés par l'IA dans un contexte autre que celui du dialogue avec l'utilisateur susnommé.

En d'autres termes, il est conseillé de faire en sorte que le document ne soit pas intégré au système d'apprentissage global de l'IA (perte de propriété intellectuelle, voire partage de données personnelles). Pour cela on peut par exemple, dans ChatGPT, désactiver l'option « historique de chat et formation » dans les paramètres (rubrique "contrôle des données"). Ce mécanisme est connu sous le nom de "opt-out". Chat GPT ne conservera plus les conversations, et par conséquent ne les utilisera plus pour s'entraîner. Si l'on souhaite garder l'historique de conversation sans qu'elles ne soient utilisées par Chat GPT, cela est possible mais nécessite un peu plus de manipulations : A ce jour le lien pour lancer cette démarche est : <https://privacy.openai.com/policies?modal=take-control> (puis "do not train on my content").

Fondamentalement, ne pas transmettre des informations à une IA générative reste le meilleur moyen de protéger ces informations.

Pour tout texte généré par une IA générative, le préciser par un mécanisme de citation classiques. Cette obligation peut être levée dans certains types d'usage : simple traduction, amélioration du niveau de langue sans apport des changements sur le fond.

Concernant la **génération d'images ou de vidéos**, il convient de citer l'IA ayant permis la génération. Vérifier que le média ne contient pas de personnes réelles reconnaissables.

A ce jour, l'IRF ZAAO n'offre pas d'accès institutionnel aux IA. La création d'un compte pour l'utilisation d'un outil relève donc d'une décision privée.

Sur le plan éthique, il faut être conscient que les résultats sont susceptibles de suivre certains préjugés historiques sur le sujet ; la neutralité souvent prêtée aux méthodes d'intelligence artificielle est extrêmement relative. ***Il faut donc faire preuve d'esprit critique.*** Il est recommandé, lorsqu'une IA a été utilisée pour générer de nouvelles idées ou un nouveau texte, de conserver un exemplaire de cette production dans sa forme brute, pour des questions de traçabilité. Cette recommandation ne s'applique pas lorsque l'IA est simplement utilisée pour améliorer un texte personnel.

Les utilisateurs reconnaissent être informés des risques de biais inhérents aux données d'entraînement de l'IA. Les utilisateurs ont la responsabilité de surveiller, identifier, et corriger tout contenu potentiellement biaisé, discriminatoire ou offensant produit par l'IA.

² Une donnée personnelle est toute information se rapportant à une personne physique identifiée ou identifiable (issue du site de la CNIL <https://www.cnil.fr/fr/definition/donnee-personnelle>).

IV - Consignes spécifique à l'usage dans le cadre pédagogique

Le recours à des outils mobilisant l'Intelligence Artificielle (IA) est par principe interdit sauf si son utilisation est explicitement autorisée dans le sujet proposé aux stagiaires. Le recours non autorisé à l'intelligence artificielle constitue une fraude. Si l'utilisation de l'IA est autorisée, les éléments de la copie qui en sont issus devront être signalés comme n'importe quel emprunt ou citation d'une source externe. ».

En dehors du contrôle de connaissances, **il revient à l'enseignant de définir les consignes** d'usage de l'IA durant les travaux dirigés ou pratiques ou dans tout autre type d'exercice.

Un outil d'IA ne saurait se substituer au stagiaire pour mettre en œuvre les compétences mobilisées dans un exercice donné.

Par exemple, il n'est pas judicieux de faire réaliser une traduction à une IA si l'objectif est justement de mettre en œuvre un vocabulaire acquis récemment, ou pour réaliser des exercices d'initiation à la programmation.

En revanche, il peut être pertinent d'utiliser une IA pour générer un texte que le stagiaire devra ensuite traduire par lui-même, ou critiquer ; ou encore pour bâtir une trame d'application informatique dans laquelle le stagiaire viendra compléter lui-même les éléments correspondant à l'enseignement en cours.

Les outils d'IA ne peuvent pas se substituer aux enseignants pour le suivi des formations ou contrôle de acquis ou autres devoirs.

V - Consignes pour les personnels administratifs et techniques

Les consignes pour les personnels administratifs et techniques correspondent essentiellement aux consignes générales.

Ne pas transmettre de données personnelles ou de documents professionnels autres que publics (pas de documents internes, pas de documents confidentiels reçus d'un tiers, ...).

En conséquence, la production de résumés de documents internes ou confidentiels est proscrite, de même que la traduction de documents internes ou confidentiels.

Le résumé, la traduction ou la synthèse de documents publics est possible, mais cet usage suppose une relecture critique : s'assurer que le propos est conforme à l'original.

L'usage comme moteur de recherche est à manipuler avec la plus grande prudence, à cause des risques d'hallucinations (voir annexes). Il se peut également que l'IA générative ait été entraînée avec des textes provenant de l'étranger et fournisse une réponse correcte... mais pas pour la France (correspondant par exemple à la législation Suisse ou Belge).

Une IA générative peut être utilisée pour produire un plan d'action, une trame de projet, etc. Comme mentionné dans les consignes générales, les documents générés doivent faire l'objet d'une **relecture critique et attentive**.

Références

Sources : Charte intelligence artificielle de l'Université d'Orléans – Avec l'accord de l'Université d'Orléans

CNIL. *Intelligence artificielle : Le plan d'action de la CNIL*. 16 mai 2023.

<https://www.cnil.fr/fr/intelligence-artificielle-le-plan-daction-de-la-cnil>

Déclaration de Montréal pour un développement responsable de l'Intelligence Artificielle. Une initiative de l'Université de Montréal. 2018 : <https://declarationmontreal-iaresponsable.com/la-declaration/>

Comité national pilote d'éthique du numérique (CNPEN). *Avis n°7 Systèmes d'intelligence artificielle générative : enjeux d'éthique*. 30 juin 2023.

https://www.ccne-ethique.fr/sites/default/files/2023-09/CNPEN_avis7_06_09_2023_web-rs2.pdf

ALLEA. *The European Code of Conduct for Research Integrity*. Révisé Edition 2023. Juin 2023. [DOI 10.26356/ECOC](https://doi.org/10.26356/ECOC)

Office Français de l'Intégrité Scientifique. *L'intégrité scientifique. Les textes de référence*.

<https://www.ofis-france.fr/les-textes-de-reference/>

Office Français de l'Intégrité Scientifique. *Systèmes d'intelligence artificielle générative : quelques points de vigilance*. L'OFIS fait le point. Février 2024 :

https://www.ofis-france.fr/wp-content/uploads/2024/02/LOfisfaitlepointFevrier2024_AI.pdf

La Higuiera, Colin de, Iyer, Jotsna. *IA pour les enseignants : un manuel ouvert*, 2eme édition. AI4T.

Janvier 2024. <https://pressbooks.pub/iapourlesenseignants/>

Université de Bretagne Sud. *Cartographie : IA génératives dans l'enseignement*. 2024.

<https://www.mindomo.com/fr/mindmap/ia-generatives-dans-lenseignement-afe76157fc7d4d3e99d300cbe75596a1>

Éléments techniques de compréhension

Les techniques actuelles d'IA génératives reposent sur des approches statistiques : à partir d'un grand nombre d'exemples (textes, images, ...) fournis pour son apprentissage, une IA générative mémorise des éléments et associations fréquentes. Ensuite, elle réagira aux requêtes (appelées *prompts*) en générant une suite d'éléments dont elle juge l'association pertinente par rapport à la requête. Dans le cas d'une IA générant du texte, elle produira au fur et à mesure une suite de mots (pour être précis, il s'agit de « tokens », qui peuvent être des morceaux de mots, des éléments de ponctuations, ...) en les enchaînant d'une manière probable, vis-à-vis du prompt, du début de réponse déjà produit, et de l'historique de conversation.

En conséquence, une IA générative n'est pas un « super-dictionnaire » ni un moteur de recherche qui fournirait un raccourci vers une source d'information. Elle n'est pas non plus, jusqu'à preuve du contraire, une entité intelligente qui comprendrait la question posée de manière conceptuelle et retournerait une réponse argumentée. Elle n'est qu'un système qui renvoie une suite de mots répondant de manière fortement probable à la propre suite de mots de l'utilisateur.

Il est donc possible que la réponse sonne juste mais ne corresponde pas à des faits avérés, surtout s'il est demandé à l'IA de s'aventurer hors des concepts généraux.

Enfin, et même si cela est tentant, il ne faut pas prêter à une IA générative des intentions ni des sentiments. Elle ne vous aime pas, elle ne vous déteste pas. Elle ne vous dit pas la vérité et elle ne vous ment pas. Elle ne fait que remplir au mieux le rôle qu'on lui a assigné : vous répondre de manière vraisemblable en fonction des exemples à partir desquels elle a appris.

Concept d'hallucination

Une IA générative formule toujours une réponse aux questions qui lui sont posées, quitte à ce que cette réponse soit totalement erronée. Ce type de dérive relève de la catégorie plus large des hallucinations.

On utilise le terme d'hallucination pour décrire un résultat aberrant fourni par une IA générative. Dans la génération de texte, cela peut prendre la forme d'une production ne correspondant à aucune réalité, par exemple une dissertation expliquant pourquoi les œufs de vache sont plus gros que les œufs de poule, ou plus subtilement l'évocation de textes de référence ayant en fait été inventés par l'IA. Dans le domaine de la génération d'image, cela peut se manifester par des éléments irréels, comme un personnage à trois bras ou une perspective impossible.

Si cette aberration peut être bénigne, voire souhaitée dans une production de fiction, cela est en revanche problématique si l'on s'attend à une réponse reflétant la réalité.

Comment citer un usage de l'Intelligence Artificielle

Si l'utilisation d'IA génératives est autorisée pour la réalisation d'un travail, il convient de documenter l'emploi des outils d'intelligence artificielle au cours de l'élaboration du travail demandé, de référencer les IA utilisées et de citer les contenus générés par IA et intégrés dans le travail soumis à évaluation.