Limites, risques et bonnes pratiques

1. Introduction

Lorsque l'on choisit d'utiliser des IAG, il faut être conscient que, comme tout outil, **elles comportent des limites et des risques**, et donc qu'en regard il existe une série de **bonnes pratiques essentielles** à suivre pour garantir une utilisation optimale.

Dans ce module, nous allons donc parcourir ces différents aspects, tout en gardant en tête que ces listes ne sont pas forcément exhaustives et qu'elles pourront évoluer avec le temps. Dans un premier temps, nous allons explorer les différentes **limites des IAG**, puis nous identifierons **les risques qui peuvent découler de leur utilisation**, et enfin nous recenserons l**es bonnes pratiques à mettre en place** pour essayer de pallier à ces différents inconvénients.

2. Limites

Il existe de nombreuses limites aux outils d'IAG, certaines pourraient être atténuées avec les évolutions rapides des outils, mais d'autres renvoient aux **contraintes inhérentes à ces systèmes**, liées à leur conception et à leur mode de fonctionnement. Bien qu'elles soient capables de générer des textes fluides et de synthétiser des informations, ces intelligences artificielles ne possèdent ni compréhension réelle, ni jugement critique, ni accès continu aux connaissances actualisées, … ni intelligence. Elles peuvent générer des erreurs, simplifier ou déformer des situations complexes, reproduire des biais, ou fournir des informations obsolètes. Ces limites rappellent que les IAG **ne peuvent pas se substituer au raisonnement humain** et que leur utilisation nécessite vigilance, vérification des contenus et cadre réfléchi d'intégration dans les pratiques pédagogiques et scientifiques.

2.1. Hallucinations

Qu'est-ce que c'est ?

Les hallucinations désignent la production, par une IAG, **d'informations incorrectes ou inventées**, même lorsque le modèle s'exprime avec assurance. Elles apparaissent parce qu'un modèle génératif prédit des suites de mots ou des images plausibles plutôt qu'il ne vérifie la réalité des faits. Les illustrations les plus parlantes de ces hallucinations se voient dans la génération des images : personnages avec trois mains, écritures illisibles, etc.

Cette notion renvoie à un fonctionnement fondamental qu'il faut absolument toujours avoir en tête lorsque l'on utilise un outil d'IAG : l'IAG génère du texte cohérent sur le plan linguistique, **sans garantie de correspondance avec le monde réel.**

En quoi c'est une limite ?

Les hallucinations constituent une limite importante car elles **compromettent la fiabilité des contenus produits**, notamment lorsqu'ils sont réutilisés ensuite tels quels. L'utilisateur peut avoir le sentiment que la réponse est

exacte grâce à sa cohérence interne, et parce que les IAG utilisent souvent des tons péremptoires et assurés, alors qu'elle repose sur des informations imaginaires.

Exemple

Lors de la rédaction d'un support de cours, une IAG peut fournir une bibliographie contenant des articles inexistants, dotés de titres, de revues et de DOI inventés. Si l'utilisateur ne vérifie pas systématiquement ces références, elles peuvent se retrouver intégrées au support distribué aux étudiants.

2.2. Biais

Qu'est-ce que c'est ?

Comme vu dans la partie précédente, les biais correspondent aux **déformations produites par une IAG lorsqu'elle génère du contenu** influencé par les données sur lesquelles elle a été entraînée, le modèle sur lequel elle est basée ou par son interaction avec l'utilisateur. Les biais peuvent prendre la forme de stéréotypes, de sur-représentations ou d'angles d'approche qui privilégient certains points de vue au détriment d'autres. Les biais sont des **tendances structurelles** inscrites dans le fonctionnement même du modèle, qui reflète notamment les données dont il est issu.

En quoi c'est une limite ?

Les biais sont une limite car ils peuvent **orienter ou restreindre la manière dont un sujet est présenté**, ce qui affecte la neutralité, la diversité des perspectives ou l'équilibre argumentatif. Lorsque l'utilisateur attend une réponse générée, il peut recevoir une proposition qui donne l'illusion d'être objective alors qu'elle privilégie implicitement un cadre particulier. Cela peut entraîner des explications partielles, ou renforcer involontairement des représentations simplificatrices.

Exemple

Une IAG interrogée sur les causes d'un phénomène social peut privilégier une explication issue d'un courant de recherche dominant dans les publications anglophones, alors que d'autres traditions théoriques, présentes dans d'autres aires linguistiques, proposent des interprétations tout aussi robustes mais moins visibles dans les données d'entraînement (comme les travaux sur la traite esclavagiste, dont la majorité des sources provient d'universitaires occidentaux et donc qui biaise les résultats à cette seule vision).

2.3. Corpus arrêté à une date donnée

Qu'est-ce que c'est ?

Les sources consultées par les IAG sont « **figées dans le temps** », c'est-à-dire qu'elles s'appuient sur un corpus arrêté à une date donnée, même lorsque son interface donne l'impression d'un accès direct et continu à l'actualité ou aux dernières publications scientifiques (par exemple, la phrase "Rechercher sur le web" s'affiche lorsque l'IAG ChatGPT est en train de générer une réponse). Cela signifie que les résultats générés **dépendent de l'état des données au moment de l'entraînement**, sans intégration des évolutions récentes. Le modèle peut donc mobiliser des informations dépassées ou ne pas reconnaître l'existence de résultats plus récents qui modifient la manière d'aborder un sujet.

En quoi c'est une limite ?

Cela peut induire des **décalages entre ce que produit l'IAG et l'état réel d'un champ de recherche**. Dans des disciplines où les connaissances évoluent rapidement, l'outil peut proposer une vision stabilisée de débats qui ne le sont plus, ou présenter comme consensuelles des positions aujourd'hui révisées. L'utilisateur risque alors d'intégrer dans un cours ou une analyse des éléments qui ne **correspondent plus aux pratiques ou aux résultats** les plus à jour, ce qui peut affaiblir la qualité du travail produit.

Exemple

Lors de la préparation d'un cours de droit, une IAG peut décrire comme en vigueur une réforme abandonnée ou profondément modifiée récemment, ou alors il peut être dangereux d'utiliser les IAG pour générer des analyses de situations géopolitiques actuelles qui ne prendraient pas en compte les évolutions récentes (conflits en cours, accords commerciaux, etc.).

2.4. Absence de compréhension

Qu'est-ce que c'est ?

L'absence de compréhension signifie qu'une IAG ne saisit **ni l'intention, ni la logique profonde d'un concept**, ni les implications d'une situation. Elle ne manipule pas des idées mais des formes linguistiques : elle repère des **corrélations** entre mots sans accéder aux mécanismes ou aux significations que mobilise un raisonnement humain. Autrement dit, lorsqu'elle produit un texte qui semble pertinent, elle ne s'appuie pas sur une compréhension du phénomène mais sur **la probabilité que telle formulation suive telle autre.** Cette absence de compréhension réelle limite sa capacité à interpréter correctement des questions ambiguës, à détecter des incohérences ou à raisonner au-delà des modèles qu'elle a observés. Par ailleurs, plus la demande formulée est complexe, plus **la sortie générée va dépendre du prompt en lui-même** : une légère variation dans la question peut entraîner des écarts importants dans la réponse, non parce que le sujet change réellement, mais parce que le modèle interprète différemment les intentions implicites contenues dans les mots ou la structure de la

requête. Cette dépendance montre que l'IAG ne travaille pas à partir **d'un cadre conceptuel robuste mais d'un ajustement linguistique** qui s'adapte à la surface du texte fourni.

En quoi c'est une limite ?

C'est une limite car elle entraîne une fragilité : l'IAG peut donner des réponses **qui paraissent solides mais qui ne reposent sur aucun fondement conceptuel.** Elle peut par exemple reformuler un argument sans discerner qu'il est contradictoire, ou synthétiser deux théories incompatibles comme si elles étaient complémentaires. L'utilisateur peut alors prendre pour une analyse ce qui n'est qu'une **juxtaposition de segments plausibles**, sans véritable articulation intellectuelle. Cela affecte particulièrement les activités nécessitant une cohérence d'ensemble, un raisonnement structuré ou une interprétation fine des enjeux soulevés. Pour un travail d'analyse ou de préparation pédagogique, cette incompréhension et la dépendance à la formulation du prompt peut entraîner des difficultés à stabiliser une notion : un même concept formulé en termes différents peut donner lieu à des positionnements divergents, rendant l'exploitation de l'outil plus incertaine et obligeant l'utilisateur à multiplier les reformulations pour atteindre un résultat fiable.

Exemple

En travaillant sur un support de cours consacré aux méthodes d'analyse qualitative, une IAG peut présenter l'entretien semi-directif comme « permettant de mesurer objectivement les préférences d'un groupe », alors que cette méthode repose précisément sur l'exploration subjective et située des discours. Comme le modèle ne comprend pas les finalités méthodologiques, il assemble des éléments qui semblent aller ensemble mais qui contredisent les principes mêmes de la méthode. Si l'enseignant ne repère pas cette incohérence, il peut transmettre une interprétation erronée aux étudiants.

3. Risques

Comme son nom l'indique, le risque dans l'utilisation des IAG définit un danger, plus ou moins prévisible, qui est probable ou possible. Il s'agit donc ici de **répertorier les effets nocifs potentiels**, à court ou moyen terme, d'un usage systématique ou régulier des IAG dans sa pratique numérique quotidienne.

Le but ici n'est pas d'être alarmiste ni prophétique, mais de mettre en lumière des points d'attention à garder à l'esprit lorsque l'on fait une utilisation récurrente des outils d'intelligence artificielle générative.

3.1. Dépendance à l'outil

Il y a quelques mois, en juin 2025, une pré-étude portée notamment par des chercheurs du MIT a été largement diffusée et commentée. En substance, cette étude (réalisée sur un panel réduit de 54 participants) montre que

l'utilisation de ChatGPT dans l'écriture d'un essai réduit l'engagement cognitif (ainsi, par ailleurs, que le sentiment de "posséder" son travail) et, pire, que l'utilisation répétée d'un outil d'IAG amène à une "dette cognitive" qui ne permet pas de revenir à un niveau d'engagement cognitif "normal", même lorsque l'on n'utilise plus cet outil. Évidemment, ces conclusions ont été largement reprises et simplifiées (voire franchement contrefaites) et présentées comme preuve que les IAG rendraient bêtes, paresseux et amnésiques.

Sans aller jusque dans la caricature, cette pré-étude vient néanmoins pointer l'un des risques les plus « évidents » lorsque l'on parle des IAG : si on utilise les intelligences artificielles génératives pour tout, on ne saura presque plus rien faire. Les mêmes craintes peuvent être appliquées dans d'autres domaines : la calculatrice présente sur les smartphones ferait baisser le niveau en calcul mental, le correcteur orthographique intégré dans tout environnement numérique ferait baisser le niveau en grammaire et orthographe, etc.

La dépendance à l'outil renvoie au risque que l'utilisateur s'appuie **excessivement** sur l'IAG pour produire analyses, contenus ou synthèses, au point de **réduire son propre engagement cognitif et critique**. Cette dépendance peut s'installer progressivement, car l'outil offre rapidité et confort, donnant l'illusion de fiabilité et de compétence immédiate. Elle devient problématique dès lors que l'utilisateur **cesse de vérifier les informations, de confronter différentes sources ou de mobiliser ses propres savoirs et méthodes d'analyse**, transférant à la machine une partie de la responsabilité intellectuelle qu'il devrait conserver.

Ce risque est particulièrement sensible dans un contexte académique, car il peut affaiblir la qualité des productions et la capacité à enseigner, analyser ou évaluer de manière autonome. Il peut également favoriser une homogénéisation des contenus, puisque l'IAG tend à produire des **réponses standardisées** basées sur ses corpus, réduisant la diversité des points de vue et la créativité conceptuelle.

Exemple : Un étudiant qui utilise systématiquement l'IAG pour rédiger des synthèses bibliographiques peut progressivement se désengager de la lecture critique des articles originaux, acceptant les résumés proposés comme représentatifs sans vérifier leur exactitude ni leurs nuances.

3.2. Fuite de données personnelles ou confidentielles

La fuite de données personnelles désigne le risque que des informations sensibles, transmises à l'IAG, soient **stockées, analysées ou potentiellement accessibles par des tiers**, en raison du fonctionnement des serveurs ou des protocoles utilisés par le fournisseur de l'outil. Ce risque n'est pas lié à une intention malveillante mais à la manière dont les données sont traitées et conservées : même des informations partagées dans un cadre pédagogique ou de recherche peuvent être enregistrées, utilisées pour améliorer le modèle, ou, dans certains contextes, **exposées à des vulnérabilités**. L'utilisateur peut ainsi divulguer sans le savoir des éléments confidentiels ou protégés, ce qui peut avoir des conséquences juridiques, éthiques ou professionnelles.

Il s'agit d'un risque majeur car il concerne à la fois **la protection des individus et la responsabilité de l'enseignant-chercheur.** Les données des étudiants, les résultats d'expérimentations ou les informations internes à un projet peuvent être traitées par l'IAG de manière qui échappe au contrôle de l'utilisateur. La confiance dans l'outil ne doit donc pas remplacer les précautions nécessaires en matière de confidentialité et de conformité réglementaire.

Exemple : Lors de la correction de travaux, un enseignant pourrait copier-coller des extraits de devoirs dans l'IAG pour générer des commentaires ou des synthèses. Si ces extraits contiennent des données personnelles identifiables (nom, numéro étudiant, contenu sensible), elles peuvent être conservées par le service de l'outil et exposées à des tiers ou utilisées dans un futur entraînement du modèle, créant une violation involontaire de la confidentialité des étudiants.

3.3. Plagiat

Le plagiat correspond à la possibilité que le contenu généré par une IAG **reproduise, volontairement ou non, des passages existants provenant de textes protégés par le droit d'auteur**, ou qu'il soit repris intégralement sans attribution. Les IAG n'ont pas la capacité de distinguer ce qui est original de ce qui est protégé, ce qui peut aboutir à des formulations proches ou identiques à des sources existantes. Ce phénomène expose l'utilisateur à des problèmes de respect de la propriété intellectuelle, en particulier s'il publie ou diffuse le contenu généré comme s'il était entièrement de sa création.

Le risque est accentué dans le contexte universitaire, où la **production originale et la citation des sources** sont des principes essentiels. Même lorsqu'il n'y a pas d'intention de tricher, l'enseignant ou l'étudiant peut se retrouver en infraction vis-à-vis des règles de l'université ou des lois sur le droit d'auteur, simplement en se reposant sur l'IAG pour générer des contenus scientifiques ou pédagogiques.

Exemple : Dans la préparation d'un cours, un enseignant utilise l'IAG pour rédiger un résumé de la littérature sur une problématique sociologique. Si certaines phrases sont très proches de publications existantes et sont distribuées aux étudiants sans mention des sources, cela constitue un plagiat involontaire.

3.4. Désinformation

La désinformation renvoie au risque de voir circuler **des informations fausses ou trompeuses qui, en raison de leur présentation cohérente et assurée, sont perçues comme crédibles par les lecteurs**. Ce phénomène s'explique par la capacité des IAG à produire des textes linguistiquement plausibles sans disposer d'un réel accès à la vérité des faits. Ils peuvent ainsi mêler éléments exacts et inventions, ou encore présenter des interprétations comme des constats établis, donnant l'illusion d'un savoir solide alors que le contenu est en

partie, voire totalement, erroné. La désinformation constitue dès lors un enjeu majeur, car elle peut se diffuser rapidement et **peser sur la compréhension des faits, les choix individuels ou la construction des connaissances**.

Par exemple, des erreurs peuvent se retrouver dans des supports de cours, des synthèses de lecture ou des analyses scientifiques, parfois sans être immédiatement détectées. Même des lecteurs avertis peuvent être trompés par un raisonnement apparemment cohérent et convaincant, ce qui rappelle l'importance d'un regard critique et d'une vérification systématique des contenus générés par IAG.

Exemple : Dans la préparation d'un cours sur les transformations économiques du XXe siècle, une IAG peut produire des chiffres de croissance ou des données sur l'emploi qui semblent plausibles mais qui sont inventés ou obsolètes. Si l'enseignant intègre ces informations dans ses supports sans vérification, les étudiants reçoivent des données incorrectes, ce qui compromet la qualité de l'enseignement et la rigueur scientifique du cours.

3.5. Humanisation et attachement émotionnel

Il existe enfin un autre risque, qui peut paraître un peu moins lié au contexte universitaire « pur », mais qui peut affecter les aspects de santé mentale et de bien être des utilisateurs, en particulier des étudiants : c'est **l'attachement émotionnel aux outils d'IAG**, en relation avec l' « effet ELIZA ». Il s'agit de la tendance à **attribuer à l'IAG des qualités humaines**, comme l'intention, la compréhension ou l'empathie, alors que le modèle ne fait que générer des séquences de mots basées sur des probabilités statistiques.

Ce biais cognitif est un risque car il peut conduire à considérer les suggestions de l'outil comme **des conseils éclairés et bienveillants** plutôt que comme **des productions automatisées**. La confiance excessive induite par cette humanisation peut réduire la vigilance critique et amener l'utilisateur à déléguer des décisions intellectuelles ou émotionnelles à une entité dépourvue de discernement. Il existe de nombreux témoignages d'utilisateurs ayant déclaré un attachement fort, une relation « patient/psychologue », voire une idylle amoureuse avec une IAG.

Ce risque est particulièrement problématique dans des contextes où le jugement humain, la nuance ou la validation par les pairs sont essentiels. L'illusion d'interlocuteur intelligent peut **masquer la nature statistique de l'outil**, encourageant des comportements de passivité, de délégation de responsabilités ou de prise de décision irrationnelle seulement suggérés par une IAG. De plus, l'utilisation d'une IAG comme un confident, un ami ou un médecin peut amener à **des dérives graves**, comme l'ont montré plusieurs cas d'utilisateurs confortés dans leurs idées suicidaires par des IAG au cours de l'année 2025.

Rappel : Si vous ou l'un de vos proches souffrez de détresse émotionnelle ou êtes en proie à des pensées suicidaires, vous pouvez appeler le 3114 (accessible gratuitement 24/24 et 7 jours sur 7).

4. Bonnes pratiques

Maintenant que nous avons vu les limites et les risques liés aux IAG, nous allons énumérer quelques bonnes pratiques qui permettent de **pallier aux défauts des IAG** et de garantir une utilisation qui s'appuie à la fois sur une bonne connaissance de ce que les IAG peuvent faire et ce qu'elles ne peuvent pas faire, et sur une utilisation transparente et responsable des outils.

4.1. Utilisation à bon escient

Il est essentiel de mobiliser les IAG comme des **outils d'assistance complémentaires**, et non comme une source unique ou un substitut au jugement et à la création humaine. Utilisée de manière ciblée, sur des tâches où elle apporte un réel bénéfice (rapidité, aide à la synthèse, génération de pistes, etc.) elle peut permettre de soutenir le travail intellectuel tout en laissant à l'utilisateur **la responsabilité finale de l'analyse, de la validation et de l'interprétation**.
Cette posture suppose une connaissance claire des limites et des risques propres aux IAG (que vous pouvez retrouver dans les premiers chapitres de ce module), afin de calibrer le niveau de confiance accordé aux contenus générés. Cette approche critique et informée permet ainsi de tirer parti des capacités de l'IAG sans confondre **assistance et substitution**, et d'en faire un outil maîtrisé, responsable et fiable.

Ainsi, on peut délimiter des champs "interdits": déléguer une responsabilité aux IAG (comme corriger des copies, noter les étudiants, faire un choix pédagogique, etc.) constitue un usage inapproprié de ces outils. **La responsabilité doit toujours rester du côté humain**, car elle suppose conscience, intention et capacité à rendre compte de ses choix.

Exemple : Dans une étape de création d'un cours, on peut utiliser une IAG pour se donner des idées sur l'organisation du plan général et l'ordre des séances. Il est indispensable ensuite de retravailler ce plan pour l'adapter, vérifier que tous les objectifs pédagogiques sont balayés, que la progression est cohérente et adaptée au niveau des étudiants, etc.

4.2. Vérification systématique

Cela peut paraître être du bon sens mais il est indispensable, lorsque l'on utilise une IAG (mais, au final, comme pour n'importe quelle source utilisée) de **vérifier systématiquement les résultats générés** avant de les utiliser ou de les diffuser. En pratiquant cette vérification, on ne se contente pas de juger de la fluidité ou de la cohérence du texte, mais on confronte chaque affirmation à des sources fiables, à des données actualisées ou à sa propre expertise. Cela permet de transformer l'IAG en un **outil d'assistance fiable** plutôt qu'en un **générateur autonome de savoir**, et de maintenir la qualité scientifique et pédagogique des productions.

Exemple : Lors de la préparation d'un cours, on obtient d'une IAG une liste de références bibliographiques. En vérifiant chaque article, on peut s'apercevoir que certains DOI sont incorrects et que certaines publications citées sont inventées, ou que les références ne sont pas correctement reproduites. Cette étape de vérification permet de corriger et de compléter la liste avant de la diffuser aux étudiants, garantissant l'exactitude et la fiabilité des ressources.

4.3. Transparence dans l'usage

La transparence dans l'usage consiste à **informer clairement les interlocuteurs ou le public lorsqu'un contenu a été produit ou assisté par une IAG**. Cette habitude permet de situer le statut de l'information, d'éviter les confusions entre production humaine et génération automatique, et de maintenir **la confiance dans les échanges**. La transparence favorise également la responsabilité : l'utilisateur reconnaît que l'outil a participé à la création du contenu, ce qui encourage une relecture critique et une vérification des informations. Sans cette transparence, les productions peuvent être perçues comme entièrement humaines et fiables et peuvent induire en erreur à la fois sur le niveau d'expertise réel du concepteur et sur la fiabilité des contenus.

De plus, en tant qu'enseignant, lorsque vous signifiez que vous avez utilisé les IAG, vous participez à éviter le tabou autour de l'utilisation des IAG et vous encouragez donc les étudiant à être eux-même transparents dans leur utilisation.

Exemple : L'Université de Montréal a créé ces étiquettes qui peuvent servir à signaler facilement les utilisations d'IAG dans un document :

4.4. Prompt engineering

Le prompt engineering (ou ingénierie de prompt) part du principe que **de bons prompts donnent de bons résultats**. En substance, on considère qu'il existe des éléments essentiels à fournir aux IAG dans la rédaction de son prompt pour qu'elle génère un résultat au plus proche de la demande. Un prompt bien conçu permet de limiter les risques d'hallucinations, de réponses partielles ou hors sujet, et facilite la génération d'informations adaptées au besoin formulé. Cela renforce également le contrôle de l'utilisateur sur la production, en réduisant l'incertitude inhérente à la sensibilité des IAG à la formulation des questions.

Exemple : Vous verrez dans les prochains modules des conseils pour la conception de prompts en fonction du but recherché. En règle générale, ces prompts se basent sur la méthode "ACTIF" : Identité (rôle que doit tenir l'IAG), Action (nature de la demande), Contexte (éléments permettant de préciser le public visé, la période, l'environnement professionnel, etc.), Tonalité (le style d'écriture qui doit être utilisé) et Format (la mise en forme ou la structure des éléments générés)

5. Conclusion

Dans ce module, nous avons donc vu que les IAG portaient intrinsèquement des limites fortes dont il faut absolument tenir compte quand on les utilise : outre les biais et les hallucinations, les IAG **ne comprennent pas ce qu'on leur demande**, elles ne font que produire des suite probables de caractères qui simulent un langage plausible.

Cette notion fondamentale doit être comprise et intégrée, ce qui permet de se prémunir d'une partie des risques inhérents à l'utilisation d'un ou plusieurs outils d'IAG. Par ailleurs, en adoptant une attitude critique systématique vis-à-vis des contenus générés par IA, et en l'utilisant de façon éclairée et précise, on peut alors **gagner un outil utile**, performant dans de nombreux domaines et qui permet un certain gain de temps dans de nombreux aspects de l'enseignement.

Dans le module suivant, nous allons voir comment on peut utiliser les IAG dans la création et l'animation d'un cours, et comment on peut affiner sa façon de prompter pour obtenir les meilleurs résultats possibles.