



INRAE



Guide du bon usage des assistants à base d'intelligences artificielles génératives

Novembre 2025



Guide du bon usage des assistants à base d'intelligences artificielles génératives

Ce document s'adresse à tous les personnels INRAE voulant utiliser un assistant basé sur une intelligence artificielle générative (IAG) dans le contexte professionnel. Il souligne les risques et impacts que chacun doit connaître et évaluer avant de les utiliser et propose également des recommandations générales d'utilisation. Des cas d'usage ont été identifiés et aideront le lecteur à mieux cerner les opportunités et risques qu'offrent ces technologies. Chaque cas d'usage pouvant être lu séparément, certaines recommandations sont volontairement répétées.

Les recommandations proposées renvoient et ne se substituent pas aux réglementations existantes (règlement général sur la protection des données à caractère personnel, droit d'auteur, droit *sui generis* sur les bases de données, etc.) et ne traitent pas de la mise à disposition ou de la mise en service d'IAG faisant l'objet de la réglementation européenne n°2024/1699 sur l'IA (RIA ou AI Act) pour laquelle des obligations spécifiques s'appliquent.

Rédacteurs :

Micael Aliouat, Direction des Affaires juridiques
Colette Cadiou, Marjolaine Hamelin, Hadi Quesneville, Tristan Salord, Alban Thomas, Direction pour la Science ouverte
Armelle Carnet, Direction Responsabilité sociétale et environnementale
Christelle Chapuis, Direction d'Appui au pilotage
Jocelyn De Goër De Herve, UMR EPIA
Remy Decoupes, UMR TETIS
Gaétane Desgeorge, Jean-Michel Vansteene, Direction des Systèmes d'information
Nathalie Gandon, déléguée à la protection des données
Claire Lurin, déléguée à la déontologie, à l'intégrité scientifique et à l'éthique
Thomas Lallart, responsable de la Sécurité des systèmes d'information

Sommaire

Introduction	5
Recommandations générales	6
Focus Responsabilité sociétale et environnementale et IAG	8
Focus relatif à l'usage de données à caractère personnel	9
Synthèse des solutions recommandées suivant la sensibilité des données	10
Cas d'usage à INRAE	12
1/ Recherche d'informations	12
1.1/ Recherche d'informations (cas général)	13
1.2/ Recherche de références bibliographiques	14
2/ Rédaction de texte, compte-rendu, rapport, contrat, e-mail, etc	15
2.1/ Rédaction	15
2.2/ Traduction	17
3/ Écriture d'articles scientifiques	18
3.1/ Écriture d'articles	18
3.2/ Génération de figures scientifiques	20
4/ Évaluation d'article, de projet	21
5/ Génération d'images et de contenus audio ou vidéo	21
6/ Brainstorming et génération d'idées	22
7/ Développement de code informatique	23
8/ Traitement de données	25
Ressources bibliographiques	26
Sur les usages génériques	26
Sur les usages en science	26
Sur les usages dans l'enseignement	27
Sur les usages par les informaticiens	27
Sur le plan juridique	27
Ressources libres pour se former	28
Pour aller encore plus loin	29
Données d'entraînement : nécessité d'obtenir les autorisations des titulaires	29
Propriété et protection des résultats générés et des prompts	29
Choisir un assistant IAG en toute transparence et en toute sécurité	29



Introduction

L'intelligence artificielle (IA), née dans les années 1950, s'intéresse à reproduire des comportements « intelligents » à l'aide de l'ordinateur. Depuis un peu plus d'une décennie, nous entendons surtout parler d'intelligence artificielle au travers de l'apprentissage automatique (machine learning), et plus précisément des réseaux de neurones profonds (deep learning). Ces technologies connaissent des succès importants dans de nombreuses applications, mais l'IA recouvre un champ de méthodes bien plus vaste.

Ces toutes dernières années, ce qu'on appelle intelligence artificielle générative (IAG) suscite particulièrement l'intérêt mais aussi des craintes. En effet, des assistants virtuels, tels que ChatGPT qui est parmi les plus connus, ont d'importantes capacités conversationnelles et peuvent donner l'illusion d'une intelligence humaine. Si les précédentes IA se bornaient à une tâche bien définie, certaines IAG sont présentées aujourd'hui comme des assistants personnels généralistes.

Ce document propose au lecteur des recommandations générales et spécifiques au contexte INRAE pour l'usage de solutions d'IAG, en fonction des cas d'usage les plus courants et tenant compte de la sensibilité des données traitées.

D'une façon générale, il ne faut jamais utiliser des données dont le niveau de sensibilité n'est pas « public » avec les offres gratuites en ligne. Cela reste vrai aussi pour la majorité des offres payantes.

À noter que les différents produits (et les modèles associés) évoqués dans ce document peuvent évoluer très rapidement. Malgré le soin apporté aux messages délivrés dans ce document, il est important de garder à l'esprit que certaines recommandations pourraient devenir obsolètes dans un proche avenir. Des mises à jour seront donc proposées.

Recommandations générales

De façon générale, l'utilisation d'une IAG soulève plusieurs points d'attention dont l'impact dépend des utilisations.

Notez bien que **l'utilisateur est seul responsable devant la loi des contenus qu'il produit à l'aide de l'IAG et du respect de la réglementation relative à ses usages**. Cette responsabilité inclut l'attribution correcte des idées et de la paternité, en veillant à l'exactitude des faits, en se basant sur des sources authentiques. L'utilisateur doit donc systématiquement **vérifier l'exactitude des résultats produits**. Selon les solutions choisies (interne et/ou externe), l'utilisateur doit également veiller à ce que les prompts¹ ne contiennent pas **d'informations sensibles comme des données confidentielles ou des données à caractère personnel** (DCP)².

Ainsi pour chaque usage, l'utilisateur est invité à se questionner sur les critères suivants :

- **Qualité et pertinence des résultats** : les résultats générés dépendent de la performance de l'IAG et de ses données d'entrée (quantité, qualité, fraîcheur, etc.). Des résultats biaisés, incomplets ou même erronés peuvent être obtenus.
- **Confidentialité et sécurité** : la majorité des outils pose des questions de confidentialité et de sécurité car potentiellement soumis à des lois étrangères de transfert et d'utilisation des données. C'est le cas des outils hébergés sur des serveurs hors Union européenne mais aussi des outils hébergés dans l'Union européenne dès lors qu'ils sont édités ou simplement accessibles à une société non européenne (ou sous contrôle d'une société-mère non européenne). Il est donc important d'identifier **le niveau de sensibilité des données** que l'on souhaite soumettre avant d'effectuer une requête, celle-ci pouvant constituer une violation d'un secret protégé par la loi ou d'une obligation contractuelle de confidentialité. Par ailleurs, **la divulgation de données est aussi susceptible de compromettre le dépôt d'un brevet ou l'existence d'un savoir-faire secret**.
- **Transparence** : les modèles d'IAG sont souvent perçus comme des « boîtes noires », ce qui rend difficile la compréhension de la manière dont les résultats sont générés et le respect des règles de rigueur et d'honnêteté propres à l'intégrité scientifique. **Les systèmes d'IAG permettant de citer les sources ayant produit la réponse doivent être privilégiés**.
- **Droits d'auteur et de propriété intellectuelle** : l'utilisation de contenus protégés pour entraîner les modèles ou générer des résultats peut poser des problèmes juridiques, particulièrement importants dans le cas d'IAG ne citant pas (correctement) ses sources. **Une mauvaise utilisation de l'IAG peut conduire à du plagiat** (volontaire ou involontaire). Il est donc nécessaire, autant que cela est possible, de vérifier que les résultats générés n'incorporent pas des œuvres protégées par le droit d'auteur et ses droits voisins (textes, images, sons, code logiciel...), ou encore des données protégées par le droit des producteurs de bases de données (données issues d'une extraction substantielle d'une base de données) qui nécessiterait qu'INRAE obtienne une autorisation de leurs titulaires pour pouvoir les utiliser.

1. Prompt : instruction, rédigée en langage naturel par un humain dans une interface homme machine, transmise à une IAG pour interagir avec elle, afin d'obtenir une réponse.

2. <https://donnees-personnelles.intranet.inrae.fr/aspects-reglementaires-et-grands-principes/definitions-et-abreviations>



- **Impact environnemental :** l'entraînement et l'utilisation de grands modèles d'IAG nécessitent des ressources importantes, ce qui a un impact environnemental direct et indirect non négligeable : consommation d'énergie et d'eau, émission de gaz à effet de serre, etc.

Pour certains cas d'usage, des outils sans IA, ni IAG, peuvent donner des résultats aussi bons que des outils mobilisant de l'IA. Pour des tâches peu complexes, il est possible d'envisager des modèles très spécialisés et de privilégier des modèles plus légers et moins énergivores comme Mistral Small, DistilBERT ou TinyBERT au lieu de versions plus lourdes comme ChatGPT.

En pratique, Il est nécessaire de :

- **Garantir la confidentialité** des données le nécessitant, en prenant garde à la diffusion non maîtrisée de données (respect du droit de propriété intellectuelle, secret des affaires, protection des données y compris à caractère personnel, respect de la vie privée, etc.).
- Appliquer la même **rigueur**, avec ou sans IA :
 - Vérifier et valider l'exactitude des réponses générées via l'usage d'IA. Cela passe par la fiabilité et la transparence de la méthode utilisée ainsi que par l'identification de biais potentiels. L'humain est toujours responsable de sa production, même avec utilisation des outils d'IA. Il convient de vérifier les sources d'information ou d'idées issues d'outils d'IA. Demander à l'IAG utilisée qu'elle cite les sources des contenus utilisés pour générer la réponse et, si une licence existe, la licence applicable à ces contenus.
 - Veiller à ne pas diffuser des données non maîtrisées du fait d'hallucinations³ des outils d'IA ou en raison de sources non identifiées.
 - Pouvoir documenter, lors de l'usage d'une IA, les outils IA, algorithmes, méthodes et prompts utilisés.
 - Être attentif aux marquages apposés par les outils utilisant de l'IAG⁴. S'il y a lieu, des logos existent pour préciser le degré d'utilisation de l'IA :

Exemples de labels francophones⁵



Exemples de labels anglophones⁶



- **Utiliser les outils d'IA avec parcimonie** compte tenu des impacts environnementaux. L'usage d'une IAG n'est pas toujours pertinent, il est nécessaire d'évaluer son apport et de privilégier son usage aux cas apportant le plus de valeur ajoutée. De plus, certaines IAG permettent de différer la réponse attendue plutôt que d'avoir une réponse immédiate, ce qui permet d'économiser de la puissance de calcul.
- **Se former** aux principes et limites des outils d'IA pour en faire un usage raisonné et responsable. Progressivement, INRAE proposera des ressources en ligne et via la formation tout au long de la vie (FTLV).
- En cas de doute et lors d'usages impliquant des données sensibles, **lire les conditions générales d'utilisation (CGU)** avant d'utiliser un outil d'IAG, d'autant qu'elles sont régulièrement modifiées.

Ces recommandations sont valables pour tous les usages, décrits ou non dans ce document.

³. Hallucination (définition Wikipédia) : Dans le domaine de l'intelligence artificielle, une « hallucination » est une réponse fautive ou trompeuse qui est présentée comme un fait certain. Par exemple, un « agent conversationnel/chatbot qui génère un chiffre d'affaires » pour une entreprise sans avoir de données à ce sujet.

⁴. Les fournisseurs d'IAG ont désormais l'obligation d'apposer un marquage sur les contenus générés par IA sous un format lisible par machine (art. 50.2 du RIA).

⁵. Source : bibliothèques de l'université de Montréal

⁶. <https://ai-label.org/>

Focus Responsabilité sociétale et environnementale et IAG

Il est très difficile d'évaluer l'impact écologique (émission de gaz à effet de serre, consommation de ressources en eau, minerais et terres rares, électricité, etc.) des outils utilisant l'IAG et de les comparer précisément.

- Pour évaluer cet impact, le site [Compare:IA](#) permet une première approche et propose des ordres de grandeurs.
- Sasha Luccioni, chercheuse spécialisée dans l'impact environnemental de l'IA et membre de [l'OECD.AI](#), estime qu'une IAG utilise 30 fois plus d'énergie qu'un moteur de recherche classique et que créer une image en haute définition avec une IA consomme autant d'énergie que la recharge complète d'un téléphone portable.
- Selon l'Agence internationale de l'énergie (AIE), les datacenters poussés par l'IA et les cryptomonnaies ont consommé près de 460 TWh d'électricité en 2022, soit 2 % de la production mondiale. Cette consommation pourrait doubler d'ici 2026.

Ainsi, l'utilisation des IAG, bien qu'elle offre de nombreuses opportunités, présente des enjeux importants en termes d'impacts environnementaux et sociaux. Certaines études scientifiques commencent à démontrer un impact sur les capacités cognitives, telle que celle d'Inria et datacraft qui souligne que « l'utilisation excessive de ces outils peut entraîner une perte progressive de compétences ou favoriser une paresse cognitive ».⁷

Pour un usage respectueux de ces dimensions, il est crucial :

- De **faire preuve de discernement** et de **questionner le besoin réel**/la pertinence avant d'utiliser une IAG compte tenu des impacts, risques, etc.
- D'adopter une approche responsable, soucieuse de réduire l'empreinte écologique/environnementale des technologies utilisées.
- De veiller à ce que le contenu produit ne relaie pas, malgré soi, des normes culturelles, sociales et éthiques potentiellement discriminantes ; de garantir l'équité, l'inclusion et la justice sociale.
- De se former. INRAE, via la FTLV, proposera prochainement des webinaires et des formations, notamment sur les usages respectueux de l'IA.

Enfin, selon les besoins, des outils n'utilisant pas l'IAG peuvent donner des résultats aussi bons et être plus vertueux.

⁷. Source Inria : L'intelligence artificielle au travail, accompagner & sécuriser les initiatives des collaborateurs, juillet 2025.
<https://datacraft.paris/ia-generative-initiatives-collaborateurs/>



Focus relatif à l'usage de données à caractère personnel

INRAE se doit de protéger les données personnelles de ses agents et de ses partenaires/tiers. Il est **strictement interdit d'insérer dans un prompt des données personnelles** de type matricule, adresse personnelle, numéro de sécurité sociale, curriculum vitae (CV), lettre de motivation, dossier de candidature, entretien professionnel, etc. **Ces usages à hauts risques ne sont, à ce stade, pas autorisés à INRAE.**

Pour plus d'informations et/ou des conseils

cil-dpo@inrae.fr

Synthèse des solutions recommandées suivant la sensibilité des données

Le tableau ci-après résume, suivant la sensibilité des données (basée sur l'échelle de sensibilité commune CNRS, INRAE et Inria ⁸), les solutions possibles pour répondre aux principaux cas d'usage identifiés au sein d'INRAE.

Les solutions citées sont classées par ordre alphabétique.

Sensibilité	Public	Diffusion limitée	Confidentiel ⁹	Diffusion restreinte
Cas d'usage				
Recherche d'information	Le Chat, ChatGPT, NotebookLM, Perplexity (Recherche), Wolfram Alpha	Offre institutionnelle INRAE à venir Offre LLMAaaS d'OUTSCALE (partenariat Mistral AI et Dassault Systèmes)	Offre institutionnelle INRAE à venir Offre LLMAaaS d'OUTSCALE (partenariat Mistral AI et Dassault Systèmes)	Accompagnement requis par le Fonctionnaire Sécurité Défense et le Responsable de la sécurité des SI INRAE
Recherche bibliographique	Consensus, Elicit, Perplexity (Recherche approfondie), SciSpace AI, Scite, Scopus AI, Web of Science Research Assistant	Offre institutionnelle INRAE à venir	Offre institutionnelle INRAE à venir	
Rédaction de documents	Le Chat, ChatGPT, Claude, Copilot, Gemini, Perplexity (Recherche et Laboratoires), QuillBot	Offre institutionnelle INRAE à venir Offre LLMAaaS d'OUTSCALE (partenariat Mistral AI et Dassault Systèmes)	Offre institutionnelle INRAE à venir Offre LLMAaaS d'OUTSCALE (partenariat Mistral AI et Dassault Systèmes)	
Traduction	ChatGPT, DeepL, Google Translate, QuillBot	DeepL Pro ou supérieur	DeepL Pro ou supérieur	
Écriture d'articles scientifiques	ChatGPT, Elicit, Paperpal, Perplexity (Recherche et Laboratoires), QuillBot, Rubriq	Offre institutionnelle INRAE à venir	Offre institutionnelle INRAE à venir	
Génération de figures scientifiques	Adobe Firefly, BioRender, DALL-E 3, Flux 1.1, Gemini, Midjourney, Wolfram Alpha	Offre institutionnelle INRAE à venir	Offre institutionnelle INRAE à venir	
Développement de code informatique	ChatGPT, Claude Code, Cursor (avec privacy mode), GitHub Copilot (licence business), Mistral, Perplexity	Offre institutionnelle INRAE à venir Possibilité d'installer des modèles sur infrastructures INRAE (Devstral ou Qwen 2.5 Coder) Offre LLMAaaS d'OUTSCALE (partenariat Mistral AI et Dassault Systèmes) Cursor (avec privacy mode), GitHub Copilot (licence business)	Offre institutionnelle INRAE à venir Possibilité d'installer des modèles sur des infrastructures INRAE (Devstral ou Qwen 2.5 Coder) Offre LLMAaaS d'OUTSCALE (partenariat Mistral AI et Dassault Systèmes)	

⁸. <https://cybersecurite.intranet.inrae.fr/service-et-document/sensibilite-de-l-information-les-bons-reflexes/sensibilite-de-l-information>

⁹. Ne concerne pas les données réglementées par la protection du patrimoine scientifique et technique, les données de santé, les données à caractère personnel « sensibles » au sens de la CNIL pour lesquelles un accompagnement spécifique est nécessaire.



Quelques axes complémentaires de réflexion à prendre en compte :

- Pour des cas d'usage impliquant la recherche d'information avec génération de texte, les solutions montrant les étapes de résolution et citant les sources de façon claires doivent être privilégiées (exemple : Perplexity). Il est par ailleurs recommandé de vérifier les sources car certaines peuvent être très éloignées voire « anecdotiques » par rapport à l'information recherchée.
- Globalement, l'usage du Copilot intégré à Windows n'est pas recommandé. En effet, son intégration avancée à l'écosystème Microsoft fait peser des risques quant à la sécurité et à la confidentialité des données. Enfin, les retours d'expérience collectés dans la sphère publique laissent à penser que cette solution serait moins performante en comparaison d'autres recommandées dans ce guide.
- Sur la souveraineté de l'offre Mistral : les modèles proposés en SaaS sont hébergés chez des hébergeurs non souverains. Ainsi, les différentes offres de Mistral AI ne sont à utiliser que pour des données d'une sensibilité « public » uniquement. D'autres possibilités sont à privilégier telles que l'offre [LLMaaS d'OUTSCALE](#) qui propose les modèles de Mistral AI sur des hébergements souverains ou encore l'hébergement des modèles open source de Mistral AI en interne INRAE.

Cas d'usage à INRAE

En complément des recommandations générales qui s'appliquent à tous les cas d'usage, le lecteur est invité à se questionner selon les cas d'usages décrits ci-après.

Pour chaque cas d'usage, des recommandations spécifiques sont précisées, ce qui peut volontairement donner à voir certaines répétitions de conseils entre plusieurs cas d'usage. De ce fait, le lecteur peut lire chaque cas d'usage séparément.

1/ Recherche d'informations

Les moteurs de recherche conversationnels marquent une évolution majeure dans le domaine de la recherche d'informations en ligne. Ils se distinguent des moteurs traditionnels en combinant l'intelligence artificielle avec les capacités de recherche classiques. Les assistants IAG sont capables de rechercher et synthétiser des informations, à partir de requêtes formulées en langage naturel, sans que l'utilisateur ait besoin de maîtriser le vocabulaire du domaine. S'ils ne se substituent pas à un moteur de recherche, ils peuvent le compléter avantageusement. En effet, ces outils innovants peuvent générer des réponses synthétiques et personnalisées, plutôt que de simples listes de liens. Leur interface conversationnelle permet aux utilisateurs d'engager un « dialogue », avec la possibilité de poser des questions de suivi pour approfondir leurs recherches.

Ces moteurs se révèlent particulièrement utiles pour l'analyse et la synthèse de grandes quantités d'informations, la création de contenus et l'organisation thématique des recherches. Ils peuvent également traiter des documents et des images fournies par l'utilisateur, enrichissant ainsi les possibilités de recherche et d'analyse.

Un assistant IAG peut également servir à la recherche d'information au sein d'un corpus de documents (exemple : publications scientifiques d'une unité de recherche). Pour garantir que les informations soient bien extraites du corpus et que les résultats ne soient pas générés à partir des « connaissances » de l'IAG, une solution du type RAG ¹⁰ doit dans ce cas être privilégiée.

Principes à respecter pour toute recherche d'informations :

- il est indispensable de croiser les sources et vérifier les informations obtenues : une IAG doit toujours être utilisée avec précaution et les résultats analysés de façon critique et rigoureuse ;
- il est important de choisir l'assistant IAG le plus adapté à l'utilisation souhaitée. En particulier, la transparence sur les sources et la confidentialité des recherches sont des critères déterminants. Ainsi, il faut être vigilant car certains outils stockent les requêtes effectuées et peuvent même les partager.

¹⁰. Le RAG (Retrieval Augmented Generation) ou génération augmentée de récupération est une technique d'intelligence artificielle qui combine un modèle de langage avec un système de recherche d'informations pour fournir des réponses plus précises et actualisées en s'appuyant sur des documents et bases de données spécifiques. Le RAG permet à un agent conversationnel/chatbot de « consulter » des documents internes (manuels, procédures, FAQ, bases de données) en temps réel avant de générer une réponse, plutôt que de se contenter uniquement de ses connaissances générales figées lors de son entraînement initial.



1.1/ Recherche d'informations (cas général)

Société	Produit	Modèle(s) utilisé(s) (LLM)	Description	Politique tarifaire	Sensibilité des données
Mistral AI Société française, hébergement américain	Le Chat	Plusieurs modèles Mistral de différentes tailles suivant le niveau de tarification	Le Chat offre un assistant IA personnalisable (recherche web, génération d'image, interpréteur de code, canvas)	Free, Pro, Team ou Entreprise	Public
Open AI Société américaine	ChatGPT	Plusieurs modèles GPT accessibles suivant le niveau de tarification	Suivant la tarification, recherche approfondie, recherche en temps réel sur le web.	Free, Plus, Pro, Team, Entreprise	Public
Google Société américaine	NotebookLM	Modèles de Gemini	Assistant de recherche IA personnalisé et prise de notes. Génération de résumés (texte et audio), explications et réponses suivant le contenu téléversé par les utilisateurs	Free, Pro	Public
Perplexity AI, Inc. Société américaine	Perplexity	Plusieurs modèles possibles comme Claude, GPT, Gemini, Grok, etc.	Moteur de recherche alimenté par l'IA qui combine la puissance de plusieurs modèles avec des recherches en temps réel pour fournir des réponses précises et sourcées Recherches autonomes approfondies, création de rapports, tableaux ou d'applications web Personnalisation pour projets spécifiques (« Spaces »)	Gratuit, Pro, Entreprise Pro	Public
Wolfram Alpha Société britannique	Wolfram Alpha	Non précisé ¹¹	Analyse et calcul précis de données scientifiques Attention, limité aux bases de connaissances intégrées	Basic, Pro, Pro premium	Public

Conseils :

- Pour des prompts contenant des données ou documents publics, Perplexity est recommandé pour des recherches avec citations académiques. Wolfram|Alpha est pertinent pour des analyses quantitatives spécifiques.
- La recherche de références bibliographiques ne devrait pas être confiée à un modèle (LLM) grand public comme ChatGPT mais à des outils spécialisés pour cet usage. L'accès aux articles scientifiques est essentiel pour les chercheurs. Plusieurs IAG permettent de faciliter cette tâche en proposant des moteurs de recherche avancés, des résumés d'articles et des analyses de citations. Cependant, l'accès aux articles sous paywall et la confidentialité des requêtes restent des enjeux majeurs. Le tableau ci-dessous présente le détail des différentes solutions disponibles.

¹¹. Non précisé = lorsque les modèles utilisés ne sont pas communiqués par l'éditeur, il est difficile de vérifier le droit d'auteur sur les données d'entraînement ou encore la transparence du modèle.

1.2/ Recherche de références bibliographiques

Société	Produit	Modèle(s) utilisé(s) (LLM)	Description	Politique tarifaire	Sensibilité des données
Consensus Entreprise américaine	Consensus	LLM de ChatGPT	Synthèse des consensus scientifiques	Free, Premium, Teams, Enterprise	Public
Elicit Research, PBC Entreprise américaine	Elicit	Non précisé ¹²	Recherche avancée d'articles, synthèse automatique d'études. Moins efficace pour articles très récents	Basic (free), Plus, Pro, Team, Enterprise	Public
Perplexity AI, Inc. Entreprise américaine	Perplexity	Plusieurs modèles possibles comme Claude, GPT, Gemini, Grok, etc.	Moteur de recherche alimenté par l'IA qui combine la puissance de plusieurs modèles avec des recherches en temps réel pour fournir des réponses précises et sourcées Recherches autonomes approfondies, création de rapports, tableaux ou d'applications web Personnalisation pour projets spécifiques («Spaces»)	Free, Pro, Enterprise Pro	Public
PubGenius Inc. Entreprise américaine	SciSpace AI	Non précisé	Recherche avancée et extraction d'informations clés	Basic (free), Premium, Advanced, Teams	Public
Scite, LLC Entreprise américaine, filiale du groupe Research Solutions	Scite	lo3-mini	Analyse des citations et impact des articles Moins performant pour la recherche documentaire globale	Personal (payant), Organization	Public
Elsevier Entreprise néerlandaise, filiale du groupe britannique RELX	Scopus AI	LLM de ChatGPT	Recherche approfondie dans une base de données scientifiques Uniquement en anglais	Payant	Public
Clarivate Entreprise britannique	Web of Science Research Assistant	Non précisé	Analyse avancée des citations et métriques d'impact	Payant	Public

Conseils pour des prompts contenant des données ou documents publics uniquement (par exemple, les articles en open access):

- pour une recherche rapide et accessible, Perplexity et Elicit sont des options intéressantes ;
- pour une analyse approfondie des citations et des métriques d'impact, Scite et SciSpace AI sont à privilégier ;
- pour une recherche bibliographique exhaustive et académique, Scopus AI et Web of Science Research Assistant sont les plus performants, mais nécessitent un abonnement ;
- Notez que des IAG généralistes peuvent être utilisées séparément à chacune des étapes: par exemple pour

¹². Non précisé = lorsque les modèles utilisés ne sont pas communiqués par l'éditeur, il est difficile de vérifier le droit d'auteur sur les données d'entraînement ou encore la transparence du modèle.



formuler une équation de recherche que l'on injectera ensuite dans une base de données bibliographiques ou pour générer une synthèse à partir de PDF d'articles que l'on aura trouvés préalablement dans une base de données bibliographiques et que l'utilisateur lui fournira.

La Direction pour la Science ouverte organise des formations et teste ces outils. Des recommandations sont à venir avec des formations intégrant l'IA dans les pratiques de recherches bibliographiques.

2/ Rédaction de texte, compte-rendu, rapport, contrat, e-mail, etc.

2.1/ Rédaction

L'assistant IAG peut être utile à toutes les étapes de la rédaction d'un texte et aider à se conformer à un format attendu : proposition de plan, reformulation de parties pour en homogénéiser le contenu, corrections orthographiques et/ou syntaxiques, etc. L'IA peut aussi être très utile pour synthétiser plusieurs documents, pour résumer ou traduire un texte, pour en ajuster le ton, et s'adapter à différents types de lecteurs. Une relecture attentive est incontournable pour des corrections éventuelles et une reformulation.

Nous avons régulièrement besoin de rédiger des textes variés, tels que des e-mails, des rapports, des résumés de projets, des demandes de financement ou encore des présentations. L'IA peut être un atout précieux pour structurer les idées, améliorer la clarté du texte et accélérer la rédaction.

Principes à respecter dans le cadre de l'utilisation d'une IAG pour la rédaction de documents :

- Il est indispensable de se questionner sur la sensibilité des documents soumis à l'IA (prompt) et sur la confidentialité des requêtes. Seuls des documents dans le domaine public peuvent être soumis à des outils dont la confidentialité n'est pas garantie.
- Il est important de savoir quelles informations, fournies ou disponibles sur internet, sont utilisées par l'IA pour générer le texte. Il convient donc de privilégier les IA qui citent leurs sources.
- Il est nécessaire de relire de façon rigoureuse le document produit pour corriger les erreurs potentiellement générées par l'outil.

La rédaction de comptes-rendus par une **IA (intégrée à un système de visio-conférence ou non)** doit être systématiquement signalée aux participants au préalable. Par ailleurs, sauf à utiliser un système souverain (à l'inverse de Zoom ou Teams), seules des discussions d'ordre public peuvent donner la possibilité d'utiliser un système d'IA pour rédiger un compte-rendu, ce dernier devant ensuite être impérativement relu et corrigé par les participants. De plus, il est strictement interdit d'utiliser une solution d'IA permettant de déduire les émotions ou de mesurer le taux d'engagement d'une personne pendant une réunion virtuelle.

Société	Produit	Modèle(s) utilisé(s) (LLM)	Description	Politique tarifaire	Sensibilité des données
Mistral AI Entreprise française	Le Chat	Plusieurs modèles dont Mistral large ou encore Mistral Small	Assistant IA (multimodal) personnalisable Création d'agents IA personnalisables Déployable n'importe où – auto-hébergé, dans un cloud public ou privé, ou en tant que service hébergé dans le cloud Mistral	Free, Pro, Team, Enterprise	Public
Open AI Société américaine	ChatGPT	Plusieurs modèles GPT accessibles suivant le niveau de tarification	Rédaction de documents, reformulation	Free, Plus, Pro, Team, Enterprise	Public
Anthropic Entreprise américaine	Claude	Sonnet 4, Opus 4	Analyse de longs documents, réponses nuancées, assistance à l'écriture	Free, Pro, Max, Enterprise	Public
Microsoft Entreprise américaine	Copilot	Principalement les modèles de ChatGPT	Génération rapide dans Office Biais stylistiques possibles	Free, Pro	Public
Google Entreprise américaine	Gemini	Gemini	Capacités multimodales natives (texte, image...) Moins performant en rédaction scientifique	Standard, Enterprise	Public
Perplexity AI, Inc. Entreprise américaine	Perplexity	Plusieurs modèles possibles comme Claude, GPT, Gemini, Grok, etc.	Moteur de recherche alimenté par l'IA qui combine la puissance de plusieurs modèles avec des recherches en temps réel pour fournir des réponses précises et sourcées Recherches autonomes approfondies, création de rapports, tableaux ou d'applications web Personnalisation pour projets spécifiques («Spaces»)	Gratuit, Pro, Enterprise Pro	Public
QuillBot Entreprise américaine, filiale de Learneo, Inc	QuillBot	Non précisé ¹³	Outil efficace pour reformuler et améliorer le style. Moins adapté pour rédiger de longs documents	Gratuit, Premium	Public

Conseils : pour les données publiques, ChatGPT et Claude 4 sont les meilleurs choix pour la rédaction de documents complets, tandis que QuillBot est un bon complément pour améliorer le style et la reformulation. Pour des données plus sensibles, des modèles open source de Mistral AI peuvent être une solution s'ils sont déployés sur un datacenter INRAE ou dans des clouds de confiance (SecNumCloud).

Plusieurs assistants IAG avancés sont spécifiquement conçus pour transformer des documents textuels en présentations (ex : Slidespilot, Gamma, Prezo).

¹³. Non précisé = lorsque les modèles utilisés ne sont pas communiqués par l'éditeur, il est difficile de vérifier le droit d'auteur sur les données d'entraînement ou encore la transparence du modèle.



2.2/ Traduction

Tous les modèles d'IAG ne sont pas explicitement conçus pour des tâches de traduction. En effet, la traduction de textes scientifiques nécessite une grande précision terminologique et contextuelle. Certaines IAG sont particulièrement performantes pour assurer une traduction fidèle et fluide, tandis que d'autres offrent des fonctionnalités supplémentaires, comme la reformulation et l'adaptation stylistique.

Société	Produit	Modèle(s) utilisé(s) (LLM)	Description	Politique tarifaire	Sensibilité des données
Open AI Entreprise américaine	ChatGPT	Plusieurs modèles GPT accessibles suivant le niveau de tarification	Traduction naturelle et fluide, capable de reformuler et expliquer les termes Moins précis sur certains termes techniques	Free, Plus, Pro, Team, Enterprise	Public
DeepL SE Entreprise allemande	DeepL	Non précisé ¹⁴	Traduction précise avec un bon respect du contexte. Moins performant pour des textes très spécialisés	Gratuit, Starter, Advanced, Ultimate, Enterprise	Public
DeepL SE Entreprise allemande	DeepL Pro	Non précisé	Traduction précise avec un bon respect du contexte. Moins performant pour des textes très spécialisés	Advanced, Ultimate, Enterprise	Diffusion limitée et Confidentiel
Google Entreprise américaine	Google Translate	Non précisé	Très rapide mais manque de précision pour les textes spécialisés	Non précisé	Public
QuillBot Entreprise américaine, filiale de Learneo, Inc	QuillBot	Non précisé	Outil de reformulation avancée, utile pour améliorer le style et la clarté. Moins efficace pour les longues traductions complètes	Gratuit, Premium	Public

Conseils: Pour des prompts contenant des données publiques, DeepL est la meilleure option pour des traductions précises. ChatGPT peut être utile pour reformuler et clarifier certains passages. QuillBot est recommandé pour améliorer le style, tandis que Google Translate peut dépanner pour des traductions rapides mais doit être utilisé avec précaution pour les textes scientifiques spécialisés.

Les textes retravaillés avec l'aide d'un assistant IAG doivent faire l'objet d'une relecture attentive et ne doivent pas contenir de données sensibles, personnelles ou confidentielles.

¹⁴. Non précisé = lorsque les modèles utilisés ne sont pas communiqués par l'éditeur, il est difficile de vérifier le droit d'auteur sur les données d'entraînement ou encore la transparence du modèle.

3/ Écriture d'articles scientifiques

La génération de textes et d'images pour des publications dans des revues scientifiques soulève des problèmes de droit d'auteur, notamment en terme d'autorat, de réutilisation et de plagiat, dont bon nombre restent non résolus. **Les chercheurs doivent appliquer de façon stricte la transparence et la rigueur requises par l'intégrité scientifique mais aussi se reporter aux règles en vigueur de la revue scientifique ou de l'éditeur qui diffuse la revue dans laquelle ils souhaitent publier leur article. Certaines revues interdisent l'utilisation de l'IAG ; d'autres l'autorisent en obligeant à déclarer son utilisation dans les remerciements.**

La confidentialité des données dans la requête saisie dans une IAG est essentielle car l'outil peut manipuler des résultats non publiés. Dans tous les cas, une attention doit être portée aux sources et à la vérification des informations.

Avant d'utiliser un outil d'IAG pour préparer un article scientifique, les auteurs doivent consulter les instructions de la revue cible pour vérifier les usages autorisés. De manière générale, les règles en vigueur chez la plupart des éditeurs de revues à l'heure de la rédaction de ce guide sont les suivantes :

- Toute utilisation de l'IAG pour un usage au-delà de la simple reformulation, édition ou traduction d'un texte doit faire l'objet d'une déclaration par l'auteur et de l'autorisation de la revue cible.
- Les assistants IAG ne peuvent en aucun cas être déclarés comme auteurs ou co-auteurs de publications scientifiques.
- L'utilisation des assistants IAG pour la modification ou la manipulation de données et résultats de recherche constituent une falsification, c'est-à-dire un manquement grave à l'intégrité scientifique.

3.1/ Écriture d'articles

Les textes scientifiques utilisent un vocabulaire particulier, et leur écriture nécessite clarté, précision et rigueur. Il convient d'utiliser des IAG qui ont été entraînées sur ce type de documents pour un meilleur résultat. Certaines IAG peuvent aider à structurer un texte, améliorer la qualité rédactionnelle et vérifier la cohérence des arguments. Cependant, **leur usage doit être accompagné d'une relecture critique, notamment pour éviter les erreurs factuelles et les biais d'interprétation.**



Société	Produit	Modèle(s) utilisé(s) (LLM)	Description	Politique tarifaire	Sensibilité des données
Open AI Société américaine	ChatGPT	Plusieurs modèles GPT accessibles suivant le niveau de tarification	Rédaction de documents	Free, Plus, Pro, Team, Enterprise	Public
Elicit Research, PBC Entreprise américaine	Elicit	Non précisé ¹⁵	Synthèse d'articles, structuration d'arguments scientifiques. Moins performant pour la rédaction fluide	Basic (free), Plus, Pro, Team, Enterprise Public	Public
Cactus communication Singapour	Paperpal	LLM propre	Rédaction d'articles scientifiques, amélioration de la rédaction, synthèse de PDF et extraction d'informations, recherche de références bibliographiques, vérification de plagiat	Free, Prime	Public
Perplexity AI, Inc. Entreprise américaine	Perplexity	Plusieurs modèles possibles comme Claude, GPT, Gemini, Grok, etc.	Moteur de recherche alimenté par l'IA qui combine la puissance de plusieurs modèles avec des recherches en temps réel pour fournir des réponses précises et sourcées Recherches autonomes approfondies, création de rapports, tableaux ou d'applications web Personnalisation pour projets spécifiques («Spaces»)	Gratuit, Pro, Entreprise Pro	Public
QuillBot Entreprise américaine, filiale de Learneo, Inc	QuillBot	Non précisé	Reformulation avancée, amélioration de la clarté et de la lisibilité Moins adapté aux textes longs et complexes	Gratuit, Premium	Public
American Journal Experts Entreprise américaine	Rubriq	Non précisé	Amélioration du style rédactionnel	Standard, Premium	Public

Conseils : Perplexity (notamment la fonctionnalité «Laboratoires») est recommandé pour la rédaction de textes scientifiques bien structurés, tandis que QuillBot peut être utile pour améliorer la lisibilité. Elicit est un bon choix pour organiser des arguments basés sur des sources scientifiques. Paperpal propose davantage de fonctionnalités (correction grammaticale, détection de plagiat, intégration dans Word...) mais sa version gratuite est limitée.

¹⁵. Non précisé = lorsque les modèles utilisés ne sont pas communiqués par l'éditeur, il est difficile de vérifier le droit d'auteur sur les données d'entraînement ou encore la transparence du modèle.

3.2/ Génération de figures scientifiques

La production de figures scientifiques est essentielle pour illustrer des concepts, des résultats expérimentaux et des modèles théoriques dans les articles et présentations. Plusieurs outils basés sur l'IAG permettent de faciliter cette tâche en générant des schémas précis, des illustrations et des infographies adaptées aux besoins des chercheurs.

Société	Produit	Modèle(s) utilisé(s) (LLM)	Description	Politique tarifaire	Sensibilité des données
Adobe Entreprise américaine	Adobe Firefly	Non précisé ¹⁶	Génération d'illustrations et retouches avancées avec des fonctionnalités adaptées aux scientifiques. Payant pour un usage intensif	Gratuit, Standard, Pro, Premium	Public
Science Suite Inc. Entreprise canadienne	BioRender	Non précisé	Outil spécialisé pour les figures scientifiques avec des bibliothèques de modèles adaptées à la biologie Payant pour certaines fonctionnalités avancées	Free, individual, Lab, Institution	Public
Open AI Entreprise américaine	DALL-E	DALL-E 3	Génération d'illustrations scientifiques par IA sur la base d'une description textuelle. Manque de précision pour les schémas complexes et techniques	Plus, Pro	Public
Stable Diffusion (Black Forest Labs) Entreprise allemande	Flux 1.1 Pro AI Image Generator	Flux 1.1	Génération rapide d'illustrations avec un grand réalisme	Gratuit, standard, professionnel	Public
Google Entreprise américaine	Gemini	Gemini Nano, Gemini Pro et Gemini Ultra	Capacités multimodales natives (texte, image ...) Moins performant en rédaction scientifique	Standard, Enterprise	Public
Midjourney, Inc. Entreprise américaine	Midjourney	Midjourney	Génération de visuels, illustrations, dans tous les styles artistiques	Basic, Standard, Pro, Mega	Public
Wolfram Alpha Entreprise britannique	Wolfram Alpha	Non précisé	Export LaTeX/PDF, calculs symboliques. Courbe d'apprentissage raide	Pro, Pro premium	Public

Conseils : Dans le cas de données publiques, pour des figures précises et adaptées à la biologie, BioRender est l'outil recommandé. Pour des illustrations générées par IAG, DALL-E peut être une alternative intéressante, bien que limitée pour des schémas complexes. Adobe Firefly est pertinent pour des retouches et améliorations graphiques avancées.

¹⁶. Non précisé = lorsque les modèles utilisés ne sont pas communiqués par l'éditeur, il est difficile de vérifier le droit d'auteur sur les données d'entraînement ou encore la transparence du modèle.



4/ Évaluation d'article, de projet

Cet usage comporte de nombreux enjeux déontologiques et éthiques. Dans le cadre d'une évaluation, l'IAG doit être utilisée avec la plus grande prudence car, sans jugement contextuel et humain, le risque d'une évaluation consensuelle sans valeur ajoutée est grand. **Si, malgré ces réserves, une IAG est utilisée dans le cadre d'une évaluation, il est indispensable de combiner ses résultats avec une expertise humaine pour garantir une évaluation précise, adaptée et rigoureuse. De plus, la plus grande transparence sur l'utilisation de l'IAG est essentielle pour maintenir la confiance dans le processus d'évaluation.**

L'usage d'un assistant IAG pour évaluer un document contenant des informations non publiées (projet, article, données, etc.) n'est possible que si on peut garantir que l'usage de l'IAG ne causera pas de rupture de confidentialité. Ainsi **l'usage d'assistants IAG ne présentant pas les garanties de confidentialité nécessaires est à proscrire pour l'évaluation de projets de toute nature.**

L'usage des assistants IAG pour la rédaction de rapports d'évaluation par les pairs d'articles scientifiques n'est généralement pas autorisé par les maisons d'édition, principalement pour des raisons de confidentialité. Dans le cas où l'article à évaluer est déjà en accès libre, par exemple sous forme de preprint, l'usage d'un assistant IAG pour la rédaction du rapport d'évaluation ne rompt pas la confidentialité. Cependant, les évaluateurs doivent, quel que soit le cas, **toujours se conformer aux instructions de la revue pour les reviewers.**

Il est par ailleurs interdit d'utiliser un quelconque système d'IAG (souverain ou non) dans le cadre de l'évaluation individuelle d'un agent tant que cette possibilité n'est pas formellement autorisée par INRAE.

Principes à respecter :

- Il est indispensable de combiner évaluation humaine et évaluation par l'outil pour produire une réflexion rigoureuse et experte.
- L'IAG choisie doit être adaptée à la sensibilité de la donnée.
- Il est indispensable d'être transparent sur l'utilisation de l'IAG, vis-à-vis du demandeur de l'évaluation et vis-à-vis des scientifiques porteurs des projets ou des articles.

Conseils : Aucune solution n'est volontairement proposée pour cet usage car il n'est pas recommandé.

En tout état de cause, que ce soit pour l'évaluation d'un projet ou d'un article scientifique, l'usage de l'IAG ne doit pas se substituer à l'analyse critique réalisée par les experts. Cet usage doit donc se restreindre à celui de la rédaction/amélioration de texte (cas d'usage 2).

5/ Génération d'images et de contenus audio ou vidéo

En cas d'utilisation d'images générées par IAG dans un document diffusé en dehors de l'Institut, il est recommandé de consulter les conditions générales d'utilisation de l'assistant IAG utilisé pour connaître la politique d'utilisation des contenus générés.

Pour garantir un usage responsable et respectueux des normes légales et sociales, voici des recommandations clés à suivre.

- Il faut **informer clairement sur l'origine des contenus générés**. Si les images, les sons ou les vidéos sont basés sur des œuvres existantes, les utilisateurs doivent être informés de leur provenance, et les auteurs originaux doivent être crédités.
- Il est recommandé **d'ajouter des métadonnées ou des balises explicites** qui indiquent que le contenu (image, audio ou vidéo) a été créé ou modifié par une IAG. D'une manière générale, il faut être en mesure de fournir des informations permettant de retracer la source des contenus générés, comme l'algorithme ou le modèle utilisé. Cela garantit la traçabilité en cas de litige.
- Lorsque des images, des voix ou des vidéos de personnes sont générées ou utilisées, il est important de **respecter leur vie privée et leur droit à l'image**. Si les contenus générés sont basés sur des personnes réelles (par exemple, synthèse vocale d'une personne publique, images inspirées d'un individu), il est impératif d'obtenir leur consentement ou de les rendre non identifiables pour garantir le respect de la vie privée.
- Les contenus générés doivent être respectueux de l'éthique, de la déontologie, et de l'intégrité scientifique.

La génération d'image et de vidéo correspond à des usages très impactants d'un point de vue environnemental. Il convient donc de circonscrire cet usage des IAG aux seules images pouvant aider à la compréhension de vos propos, d'éviter la génération d'images uniquement illustratives et de proscrire tout usage posant des questions d'intégrité scientifique ou de respect des personnes.

Conseils: Aucune solution n'est volontairement recommandée pour cet usage du fait des problématiques induites de droit d'auteur et de droit à l'image. De plus, les solutions du marché sont réputées très énergivores.

6/ Brainstorming et génération d'idées

L'IAG peut servir pour le brainstorming en recherche. Ces systèmes peuvent — lorsqu'ils sont utilisés de manière appropriée — aider à générer une variété d'idées, de perspectives et de solutions potentielles, particulièrement utiles lors des premières étapes de la planification de la recherche. Un chercheur pourrait saisir sa question de recherche dans l'assistant IAG et recevoir des suggestions sur les approches expérimentales, les méthodologies potentielles ou les questions de recherche alternatives. Par exemple : *«Analyser les recherches récentes sur la consolidation de la mémoire et l'influence des émotions sur l'apprentissage et la mémorisation. Sur la base de cette analyse, générer de nouvelles hypothèses pour des études potentielles portant sur les mécanismes neuro-biologiques»*.

Néanmoins, cette utilisation soulève des questions éthiques majeures auxquelles il est important de réfléchir avant toute utilisation. Elle peut favoriser une uniformisation de la pensée en reproduisant des schémas existants, limitant ainsi la créativité et l'originalité des recherches. Par ailleurs, les biais et hallucinations intrinsèques à l'IAG risquent de perpétuer des stéréotypes, des données ou hypothèses partielles voire fausses, rendant indispensable une utilisation critique et précautionneuse de ces outils. Ainsi, les chercheurs peuvent éventuellement les utiliser comme **point de départ de leur réflexion mais certainement pas comme des solutions ultimes**.

Conseils: Aucune solution n'est recommandée pour cet usage car beaucoup peuvent être utilisées et il est difficile d'évaluer leur pertinence. En tout état de cause, la génération d'idées avec l'assistance d'une IAG obéit comme les autres cas d'usages à l'obligation d'analyse critique et de réappropriation des résultats générés.



7/ Développement de code informatique

L'IA peut réduire le temps de développement de code informatique (scripts, développement logiciel ...), en générant tout ou partie de codes (commandes, fonctions voire même une application, notamment web) ou de configurations. Il est également possible de faire reprendre tout ou partie de ses codes pour résoudre une erreur ou en vue d'une amélioration de la qualité du développement.

Principes à respecter pour toute génération de code informatique par une IA :

- **Il faut pouvoir lire le code et le comprendre.** Tout comme pour la production de contenus textuels, il est possible d'obtenir un code bien écrit mais erroné. De plus, il est important de garantir la sécurité du code informatique produit pour ne pas introduire des failles pouvant être exploitées lors d'une cyberattaque.
- L'écriture de code informatique répond à des standards qui visent à assurer les conditions de reproductibilité et de lisibilité. Le code produit par une IA ne correspond pas forcément à ces standards. Il est de la responsabilité de l'utilisateur de **s'assurer que le code produit et diffusé respecte les standards** métiers et les bonnes pratiques de développement logiciel.
- Il faut veiller à ce que **le code généré ne puisse pas être soumis à un droit d'auteur existant**. Il est recommandé pour cela d'utiliser des IA entraînées sur des codes libres de droits.
- Il faut également considérer **les questions de confidentialité et de licence de réutilisation** avant de soumettre un code à une IA.
- Une vigilance est nécessaire lors de l'installation d'extensions d'assistance de code basée sur l'IA dans un environnement de développement (IDE). Bien que le code ne soit pas nécessairement confidentiel, l'IDE peut contenir des secrets applicatifs (jetons d'authentification, clés API, mots de passe, clés de chiffrement ou informations sensibles dans les fichiers de configuration, etc.), des informations internes telles que les chemins d'accès à des composants (bases de données, serveurs, etc.) et d'autres variables d'environnement sensibles. Ces éléments représentent un risque de perte de confidentialité en cas de captation par un tiers ou cyberattaquant.
- Il est nécessaire de rester vigilant à l'utilisation de jeux de données/de tests, qui parfois peuvent être des données réelles : ces données ne devraient pas être traitées par l'IA.
- Enfin, il est dangereux de solliciter la génération de code par une IA de modules sensibles et en particulier :
 - des modules de gestion de droit d'administration et de droit d'accès,
 - des modules cryptographiques,
 - des modules de traitement de données sensibles.

Ces recommandations s'appliquent non seulement aux programmes informatiques, mais plus largement à tout langage informatique (HTML, YAML, SQL, etc.).

Société	Produit	Modèle(s) utilisé(s) (LLM)	Description	Politique tarifaire	Sensibilité des données
Open AI Société américaine	ChatGPT	Plusieurs modèles GPT accessibles suivant le niveau de tarification	Utilisation pour une aide au débogage	Free, Plus, Pro, Team, Enterprise	Public
Anthropic Entreprise américaine	Claude Code	Claude Opus, Claude Sonnet, et Claude Haiku	Agent d'assistance au code, directement intégré dans les environnements de développement	Pro, Max, Team, Enterprise	Si hébergement interne : confidentiel sensible
Anysphere Entreprise américaine	Cursor	Non précisé ¹⁷	Environnement de développement intégrant de l'IAG et proposant notamment une interface de chat contextuel	Free, Pro, Ultra	Public et diffusion limitée uniquement avec le mode privacy activé
Mistral AI Entreprise française	Devstral	Devstral	Modèle open source pour les tâches de développement logiciel	Gratuit (hébergement On Premise ou SecNumCloud à prévoir)	Si hébergement interne : confidentiel sensible
Partenariat GitHub et OpenAI	GitHub Copilot	Version modifiée de GPT 3	Agent d'assistance au code, directement intégré dans les environnements de développement	Free, Team, Enterprise	Public
Mistral AI Entreprise française	Mistral code	Codestral, Codestral Embed, Devstral, Mistral Medium	Modèle d'IA de codage aux capacités agentiques et spécialement conçu pour résoudre des problèmes logiciels réels	Free, Pro, Team, Enterprise	Si hébergement interne : confidentiel sensible Si Mistral code chez Mistral AI : public
Partenariat Mistral AI et Dassault Systèmes	LLMaaS par Outscale	Codestral	Génération de code, détection de bugs, complétion de code, explication de code, optimisation de code	Payant	Jusqu'au confidentiel critique
Perplexity AI Entreprise américaine	Perplexity	Plusieurs modèles possibles comme Claude, GPT, Gemini, Grok, etc.	Utilisation pour une aide au débogage	Gratuit, Pro, Entreprise Pro	Public
Alibaba Entreprise chinoise	Qwen	Qwen 2.5 Coder	Version spécialisée dans la génération de code, à intégrer dans les environnements de développement	Gratuit (hébergement On Premise ou SecNumCloud à prévoir)	Si hébergement interne : confidentiel sensible

Conseils : La génération de code ne doit pas faire l'impasse sur les bonnes pratiques du génie logiciel particulièrement la séparation entre le code, l'environnement d'exécution et les données. Il est conseillé de lire le guide d'usage des IAG d'assistance au codage de l'ANSSI et du BSI et d'appliquer les recommandations de

¹⁷. Non précisé = lorsque les modèles utilisés ne sont pas communiqués par l'éditeur, il est difficile de vérifier le droit d'auteur sur les données d'entraînement ou encore la transparence du modèle.



sécurité proposées : <https://cyber.gouv.fr/actualites/lanssi-et-le-bsi-publient-leurs-recommandations-de-securite-concernant-les-assistants-de>

Pour l'installation d'une extension d'assistance de code basée sur l'IA dans un IDE, il est important d'isoler les fichiers sensibles (variables d'environnement, configuration contenant les secrets applicatifs, etc.) du projet indexé par l'extension. Aussi, il est recommandé de privilégier un mode « prompt manuel » (par rapport à une assistance contextuelle automatique) en veillant également à ne pas saisir ce type d'informations dans la requête. Le risque peut également être réduit en utilisant des secrets fictifs ou éphémères en phase de développement et en désactivant la télémétrie lorsque cela est possible.

8/ Traitement de données

L'IA peut être utilisée pour l'étiquetage, la catégorisation de textes et d'images (assignation d'une catégorie à un texte/une image à partir d'une liste établie). Cette application peut permettre d'automatiser des processus qui sont traditionnellement longs et coûteux en matière de ressources humaines. Cependant, la fiabilité de l'IA dans ces tâches nécessite une considération minutieuse et une validation au cas par cas.

La principale préoccupation concernant le jugement basé sur l'IA dans l'étiquetage est sa dépendance à la qualité et au biais des données d'entraînement. Les systèmes d'IA pourraient reproduire les biais inhérents présents dans leurs ensembles de données d'entraînement, conduisant à un étiquetage biaisé ou inexact. **Les utilisateurs doivent valider la performance de l'IA — en comparant la sortie avec des benchmarks étiquetés par des humains pour garantir la précision et l'impartialité.**

L'IA peut servir d'outil supplémentaire dans le processus d'assurance qualité des données, en aidant à l'identification des erreurs, des incohérences ou des biais dans les ensembles de données. Sa capacité à traiter rapidement d'énormes volumes de données lui permet de repérer des problèmes potentiels qui pourraient être ignorés lors de curations manuelles. L'IA doit être, pour cet usage, un élément d'une stratégie d'examen de données plus large. Il est essentiel de corroborer les anomalies détectées par l'IA avec des vérifications manuelles et des évaluations d'experts.

Conseils : Aucune solution n'est recommandée pour cet usage car beaucoup peuvent être utilisées et il est difficile d'évaluer leur pertinence. En tout état de cause, la classification ou étiquetage de données avec l'assistance d'une IA obéit comme les autres usages à l'obligation d'analyse critique et de réappropriation des résultats générés.

Ressources bibliographiques

Sur les usages génériques

- Voici quelques sites et plateformes utiles pour tester des IAG Open Source et garantir une transparence d'usage :
 - Hugging Face <https://huggingface.co/models> : plateforme hébergeant de nombreuses IAG open source, accompagnées de documentation sur les données d'entraînement et les droits d'utilisation. Ce site propose également des données sur les impacts environnementaux des IAG : <https://huggingface.co/spaces/genai-impact/ecologits-calculator>
 - EleutherAI : organisation à l'origine d'IAG open source comme GPT-Neo, offrant une transparence maximale.
 - Compar:IA, le comparateur d'IA conversationnelles : <https://www.comparia.beta.gouv.fr/>
- Sur une approche d'éthique de l'utilisation des systèmes d'IAG, le comité consultatif national d'éthique a publié l'avis du CNPEN en juin 2023 : <https://www.ccne-ethique.fr/publications/avis-ndeg7-du-cnpn-systemes-dintelligence-artificielle-generative-enjeux-dethique?taxo=56>
- Sur la prévention des biais, ce blog propose en anglais des explications et recommandations intéressantes <https://www.promptfoo.dev/blog/prevent-bias-in-generative-ai/>
- France Num, le portail de la transformation numérique des entreprises propose un **guide du débutant pour créer des prompts efficaces** : <https://www.francenum.gouv.fr/guides-et-conseils/pilotage-de-lentreprise/gestion-traitement-et-analyse-des-donnees/comment-creer>
- La direction du numérique de l'Etat (DINUM) publie régulièrement des **actualités sur l'action du numérique de l'État en faveur de l'IA** : <https://www.numerique.gouv.fr/intelligence-artificielle/>
- Un panorama des solutions IA « sur étagère » pour les administrations publiques est proposé par ALLiance, l'incubateur IA de l'État : <https://alliance.numerique.gouv.fr/cartographie/panorama-des-solutions-ia-sur-etagere-pour-les-administrations-publiques/>
- Une bibliothèque de prompts (promptothèque) du ministère en charge de l'écologie : <https://promptotheque-dsfr-bebd50.gitlab-pages.din.developpement-durable.gouv.fr/>

Sur les usages en science

- La Commission européenne a publié des **lignes directrices pour un usage responsable en science** afin de préserver l'intégrité de la recherche : <https://european-research-area.ec.europa.eu/news/living-guidelines-responsible-use-generative-ai-research-published>
- La Commission européenne a publié le rapport **Artificial intelligence in science – Promises or perils for creativity ?** explorant l'impact de l'IA sur la créativité scientifique. https://research-and-innovation.ec.europa.eu/knowledge-publications-tools-and-data/publications/all-publications/artificial-intelligence-science-promises-or-perils-creativity_en



- L'Association européenne des éditeurs scientifiques (EASE) a publié des recommandations pour l'usage de l'IA générative dans la communication scientifique : <https://ease.org.uk/2024/09/recommendations-on-the-use-of-ai-in-scholarly-communication/>
- Le Cirad a publié en octobre 2024 des points de vigilance face aux problèmes connus des IAG et des bonnes pratiques dans le domaine de la recherche : <https://agritrop.cirad.fr/609884/>
- L'Office français de l'intégrité scientifique a organisé un colloque en avril 2025 sur le thème « Comment l'IA générative transforme les pratiques de recherche : nouveaux enjeux d'intégrité scientifique ? ». Les présentations et les vidéos sont disponibles : <https://www.ofis-france.fr/colloque/colloque-2025/>

Sur les usages dans l'enseignement

- La CNIL a publié des conseils sur l'usage de l'IA dans l'enseignement : <https://www.cnil.fr/fr/deux-faq-utilisation-des-systemes-dia-scolaire>
- Le ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche a publié en juin 2025 le cadre d'usage de l'intelligence artificielle en éducation : <https://www.education.gouv.fr/publication-du-cadre-d-usage-de-l-intelligence-artificielle-en-education-450652>

Sur les usages par les informaticiens

- Pour les administrateurs de SI, l'ANSSI a publié un guide pour la mise en œuvre des systèmes d'IA générative (sécurisation d'architecture, bonnes pratiques à mettre en œuvre depuis la conception jusqu'à l'entraînement de modèles, etc.) : <https://cyber.gouv.fr/publications/recommandations-de-securite-pour-un-systeme-dia-generative>

Sur le plan juridique

- Le site Vie publique, au cœur du débat public a publié une FAQ sur le **cadre juridique européen** de l'IA : <https://www.vie-publique.fr/questions-reponses/292157-intelligence-artificielle-le-cadre-juridique-europeen-en-7-questions>
- La loi n°2024-449 du 21 mai 2024 visant à sécuriser et à réguler l'espace numérique et notamment l'article 31 sur le recours à un service d'informatique en nuage fourni par un prestataire privé : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000049563368>
- La Doctrine Cloud de l'État, diffusée par la circulaire du Premier Ministre actualisée en mai 2023 <https://www.legifrance.gouv.fr/circulaire/id/45446>
- Le règlement européen n°2024/1689 (« AI Act ») établissant des règles harmonisées pour les développeurs et déployeurs d'IA : <https://eur-lex.europa.eu/eli/reg/2024/1689/oj?locale=fr>
- La CNIL a publié les fiches pratiques IA, recommandations spécifiques sur l'application du RGPD : <https://www.cnil.fr/fr/les-fiches-pratiques-ia>

Ressources libres pour se former

- Espace de ressources IA Sup, dispositif de documentation mutualisé à destination des communautés de l'enseignement supérieur : <https://universitenumérique.fr/ia-sup-ressource-intelligence-artificielle/>
- Institut Montaigne - Objectif IA : initiez-vous à l'intelligence artificielle (Public = tous agents publics ; durée = 6 h) : <https://www.campus.numerique.gouv.fr/catalogue/objectif-ia-initiez-vous-%C3%A0-lintelligence-artificielle/>
- Inria – l'intelligence artificielle... avec intelligence (durée = 10 h) : <https://www.fun-mooc.fr/fr/cours/lintelligence-artificielle-avec-intelligence/>
- Inria – Découvrir les IA génératives (1 h) sur la plateforme interministérielle de formation Mentor : <https://mentor.gouv.fr/course/view.php?id=8890§ion=1>

La liste des formations accessibles en ligne sera complétée dans les prochaines versions de ce document.



Pour aller encore plus loin

Données d'entraînement : nécessité d'obtenir les autorisations des titulaires

Par principe, toute utilisation de données couvertes par des droits de propriété intellectuelle (droit d'auteur, base de données notamment) ne peut avoir lieu sans l'autorisation écrite des titulaires des droits de propriété intellectuelle. Cela inclut l'utilisation à des fins d'entraînement d'une IAG. Des exceptions ou des nuances peuvent permettre une utilisation mais nécessitent de vérifier au cas par cas si les conditions sont remplies. Pour plus d'informations et/ou des conseils, n'hésitez pas à contacter le référent données de votre unité (RDO) et/ou de votre département (RDS).

Propriété et protection des résultats générés et des prompts

À ce jour, ni le législateur ni la jurisprudence ne se sont prononcés sur la propriété des résultats générés par l'IAG.

Au regard du droit d'auteur, toute œuvre de l'esprit (images, sons, textes, ...) est protégeable si elle est originale, c'est-à-dire si elle dispose d'une physionomie d'ensemble propre qui découle de choix créatifs d'un auteur personne physique. Ainsi, les œuvres découlant du hasard ou de contraintes physiques (exemple : logique contraignante d'un langage informatique, choix contraints par la fonction technique du produit) sont exclues car ne résultant pas de choix libres d'un auteur s'inscrivant dans une démarche créative.

Sans préjuger des futures évolutions du droit en la matière, en pratique, une manière de se prémunir contre cette objection et afin de faire valoir son droit d'auteur, il convient d'être en mesure d'expliquer en quoi chacun des aspects du résultat généré est un choix délibéré d'un auteur humain et non le fait d'une contrainte ou le choix aléatoire d'une IAG. Par ailleurs, il est nécessaire de toujours retravailler le résultat généré par l'IAG et ne jamais le réutiliser tel quel.

Choisir un assistant IAG en toute transparence et en toute sécurité

Il est primordial de choisir des modèles dont les « poids » calculés lors de l'entraînement sont publics, et idéalement dont les données d'entraînement sont documentées, que l'on qualifie d'Open Source. Ils offrent ainsi une plus grande visibilité sur les données utilisées et les processus d'entraînement. Il est à noter que ce sont souvent des modèles plus petits et donc plus économes en ressources, existant en plusieurs variantes en fonction des domaines d'application.

Les CGU des prestataires sont un aspect essentiel à surveiller, car elles définissent les limites d'utilisation des IAG, notamment les droits sur les données, la confidentialité, et les usages commerciaux. Elles peuvent être modifiées régulièrement, imposant aux utilisateurs de rester vigilants. Il est important de supposer qu'un prestataire peut potentiellement s'appuyer sur les CGU d'un autre. Par exemple, des entreprises comme OpenAI ou Microsoft utilisent des services d'infrastructures cloud tiers comme Amazon Web Services (AWS) ou Microsoft Azure. Cela signifie que les CGU de ces services tiers influencent également les droits et la sécurité des données.

Une compréhension approfondie des CGU des prestataires est donc essentielle pour garantir que les données sont traitées conformément aux attentes légales.

Le choix entre l'utilisation d'un modèle « local » (sur des stations de travail ou des serveurs INRAE) ou via une infrastructure distante (cloud) dépend de la sensibilité des données manipulées. Pour les données à caractère personnel et/ou sensibles, il est nécessaire d'utiliser des modèles sur des infrastructures de confiance respectant les normes de l'ANSSI pour ce type de données. Elles peuvent être hébergées chez des prestataires commerciaux si elles ont le label SecNumCloud.

Si les données ne sont pas sensibles (contenus génériques, analyse textuelle non confidentielle), les infrastructures cloud (Google Cloud AI, Amazon AWS, Microsoft Azure) peuvent être utilisées pour exécuter des modèles à grande échelle. Il est cependant recommandé pour des raisons de sécurité, mais surtout de souveraineté et de critères éthiques ou économiques, d'utiliser des clouds européens. Cela permet de bénéficier de la puissance de calcul nécessaire pour traiter des modèles de grande taille sans avoir à gérer l'infrastructure matérielle. Toutefois, il est crucial de vérifier que le prestataire respecte les normes de sécurité (chiffrement des données, conformité aux standards ISO/IEC 27001, SOC 2, etc.) pour garantir une gestion sécurisée des données. Pour plus d'informations et/ou des conseils : rssi@inrae.fr.





147, rue de l'université
75338 Paris Cedex 7
Tél.: +33 1 (0)1 42 75 90 00

Rejoignez-nous sur:



inrae.fr

Institut national de recherche pour
l'agriculture, l'alimentation et l'environnement



**RÉPUBLIQUE
FRANÇAISE**

*Liberté
Égalité
Fraternité*

INRAE