# Recommandations pour l'usage des IA génératives comme assistant personnel au sein d'INRAE

Hadi Quesneville, DipSO, INRAE

« L'IA générative à l'Université Paris-Saclay : un outil du quotidien pour les chercheur·euses ! »
24 mars 2025

# Des avantages et des opportunités

- **Une productivité améliorée**
  - Augmentation de la productivité,
  - Production de nouvelles idées,
  - Grande précision par la prise en compte d'une quantité d'information plus importante.

- **Des nouvelles modalités d'interaction**
  - Outil d'aide à l'apprentissage,
  - Nouvelle interactivité avec les systèmes informatiques et automatiques,
  - Peuvent dialoguer entre eux.

# Des limites et des risques

- **Des risques juridiques.**
  - Droit d'auteur,
  - RGPD,
  - Confidentialité

- **Une question de souveraineté.**
  - Reproductibilité des résultats,
  - Maintien de services et d'activité.

- **Des erreurs**
  - Des risques d'erreurs factuelles, appelées hallucinations,
  - Des risques déontologiques, intégrité scientifique,
  - Des risques d'information obsolète (date de l'apprentissage du modèle).

# Des limites et des risques (suite)

- **Des questions d'éthiques**
  - Impact environnemental de ces pratiques,
  - Impacts sociétaux et éthique,
  - Risques de reproduction de normes sociales discriminantes.

- **Des problèmes de données**
  - L'IA générative a besoin de données en quantités considérables pour être **fiable**,
  - Le degré d'ouverture du modèle pour être **crédible**,
  - **Augmentation massive** des productions (articles, données, …).

# Recommandations générales

# Développer une expertise

- **Continuer la veille sur les LLM,** et leur usage en tant qu'assistant personnel**.**
  - Il est difficile de prédire aujourd'hui quel(s) modèle(s)/service(s) seront les plus pertinents pour assister des chercheurs.
- **Développer une expertise sur l'évaluation des LLM**.
  - Suivre dans la durée l'évolution des performances et des biais des systèmes d'IA afin d'anticiper de nouveaux risques.
- **Maîtriser la spécialisation des modèles.**
  - Il est très coûteux d'améliorer la qualité d'un LLM de fondation. Mais ils doivent être spécialisé pour des usages particuliers (technique de RAG et « fine tuning ») et cette maîtrise est clef pour le bon usage de ces IA.

# Une gouvernance de l'IA

- **Maîtriser le flux de données.**
  - L'utilisation des IA implique l'utilisation de données qui n'échappe pas au **cadre régi par la gouvernance des données**.
- **Sensibiliser, former et accompagner les utilisateurs à l'usage des IA génératives.**
  - La qualité des réponses de l'IA générative est fortement influencée par l'entrée utilisateur ou le prompt. Il est nécessaire de **proposer des formations** pour aider les utilisateurs à maîtriser cet outil.
- **Structurer une gouvernance de l'IA.**
  - Une **gouvernance de l'IA pour limiter les risques**. Le lien fort existant avec les données suggère d'établir une gouvernance commune.
  - **Suivre les usages de l'IA au sein de l'Institut**, mais aussi les **initiatives nationales et internationales** qui se mettent en place pour l'accompagnement à leur utilisation, leur contrôle et leur développement.
  - **Participer à l'organisation au sein de l'ESR** du partage des connaissances et des expériences sur ces nouveaux outils.
  - **Donner des recommandations pour leur bon usage** au sein de l'Institut.

# Quelques principes généraux en terme de gouvernance de l'IA

# Responsabilité individuelle

- L'utilisateur est le seul responsable
  - de l'exactitude des contenus générés à l'aide de l'IAG
  - du respect de la réglementation.
- Il doit se tenir informé des évolution des règles en vigueur
  - Conditions Générales d'Utilisation (CGU), protection des données personnelles,….
- Anticiper les conséquences sociales, éthiques et environnementales de leurs usages.
  - Contenus générés ne soient pas contraires à la déontologie et l'intégrité scientifique (diffusion de fausses informations, pratiques frauduleuses, etc…).
  - Résultats produits ne soient pas socialement dommageables (reproduction de normes potentiellement discriminantes, etc.).

# Respect du droit d'auteur

- Nécessité de vérifier que les résultats générés n'incorporent pas de données protégées
  - vérifiez les droits de réutilisation des sources listées par l'IAG dans le résultat généré que vous utilisez.

- Propriété et protection des résultats générés et des prompts
  - Garder une trace des dialogues menés avec l'assistant d'IAG
  - Etre 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 de l'IAG.
  - Toujours retravailler le résultat généré par l'IAG et ne jamais le réutiliser tel quel.
- Nécessité d'obtenir les autorisations des titulaires pour les données d'entrainement

# Confidentialité

## Risque de rupture de confidentialité

- La loi interdit de divulguer
  - des données personnelles et/ou sensibles ( cf. échelle de sensibilité commune CNRS, INRAE et INRIA)
  - données dont INRAE n'a pas les droits du titulaire
- L'entrée d'informations confidentielles dans le prompt d'un assistant IAG tiers constitue en effet une forme de divulgation de ces informations.
  - Expose les individus et l'institution à des risques potentiels de violation de la vie privée et de la sécurité.
- Ne pas utiliser un service comme par exemple ChatGPT, Copilote, Gemini, …
  - pour préparer un compte-rendu de réunion (attention à AI companion de Zoom),
  - pour synthétiser un article obtenu via un abonnement INRAE,
  - pour rédiger un projet de recherche comportant des aspects confidentiels,
  - pour traiter/analyser des données de recherche sensibles.

# Le choix d'un assistant IAG

- Préférer des modèles « *open sources* »
  - les données d'entraînement sont documentées et les poids du modèle sont publics.
- Pour les données à caractère personnel et/ou sensibles,
  - IA sur des infrastructures de confiances 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.
  - 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.

# Responsabilité sociétale et environnementale (RSE)

- Privilégier les serveurs locaux
  - Réduire la bande passante réseau nécessaire pour envoyer et recevoir des données vers des serveurs distants, ce qui a un impact environnemental direct.
- Privilégier des modèles plus légers et moins énergivores
  - Nombre de paramètres faible
- Identifier et corriger les biais présents dans les données ou dans les résultats générés

# Quelques recommandations selon les cas d’usages

# Recherche d'information sur internet

- Utiliser des services proposées par des tiers.
  - Des services généralistes : ChatGPT, Perplexity, …
  - Des services spécialisés : bibliographie, réglementations, …
- Ne pas injecter de données sensibles dans les prompts
  - Ne pas transmettre de documents sensibles pour interroger ces systèmes,
  - Faire attention à la rédaction des prompts.
- Services
  - Veilles scientifiques, économique, …
  - Acquisition de connaissances.

# Rédaction de documents généraux, guides, contrats

- Utile à la rédaction d'un document
  - Proposition de plan, reformulation de parties, s'adapter à différents types de lecteurs, ajuster le ton,
  - Synthétiser plusieurs documents, pour résumer ou traduire un texte,
  - Les outils spécialisés pour la traduction doivent être privilégiés.
- Plusieurs assistants IAG sont spécifiquement conçus pour les manuscrits scientifiques.
- Une relecture attentive pour des corrections éventuelles restent incontournables.

# Rédaction d'articles scientifiques

- Déclaration par l'auteur et autorisation de la revue pour l'utilisation de l'IAG
  - Usage au-delà de la simple reformulation édition ou traduction d'un texte.
- Ne peuvent pas être déclarés comme auteurs ou co-auteurs de publications scientifiques.
- Manquement grave à l'intégrité scientifique.
  - La falsification, la modification ou la manipulation de données et résultats de recherche

# Evaluation d'article, de projet, etc...

- Document contenant des informations non publiées (projet, article, données, etc...)
  - Garantir que l'usage de l'IAG  ne causera pas de rupture de confidentialités.
  - Même sur une infrastructure de confiance, l'usage de l'IAG ne doit pas se substituer à l'analyse critique réalisée par les experts.
- Rédaction de rapports d'évaluation par les pairs d'articles scientifiques
  - Interdit par les maisons d'édition pour des raisons de confidentialité.
  - Sauf s'il est déjà en accès libre (ex: preprint)

# Quelques propositions d’actions

# Développer des services reposant sur des infrastructures de confiance

## Une IA INRAE souveraine

- Services INRAE (ou ESR) pour travailler sur des documents sensibles
  - Respect du RGPD,
  - Confidentialité.
- Services
  - Synthèse de corpus documentaires,
  - Recherche d'information dans des documents,
  - Aide rédactionnelle : plan d'un document, amélioration du style, de la grammaire, de l'orthographe, traductions....

# Développement de code informatique

- Utiliser des services dédiés qui respectent le droit d'auteur et les licences.
  - Ex: service payant proposé par Gitlab Duo

- Service proposé par la forge INRAE
  - Génération de code,
  - Correction de bug.

# Accompagnement

- Sensibiliser aux questions réglementaires
  - Présentations dans de nombreux cercles (Réunions de DU, Référents données, …
  - Des documents de recommandations.

- Former les agents de l'institut à l'usage des prompts
  - Formations interne (en construction).
  - Formations de partenaires externes.

# Remerciements

**Groupe de Travail IA générative**

Micael Aliouat

Colette Cadiou

Remy Decoupes

Jocelyn De Goer De Herve

Nathalie Gandon

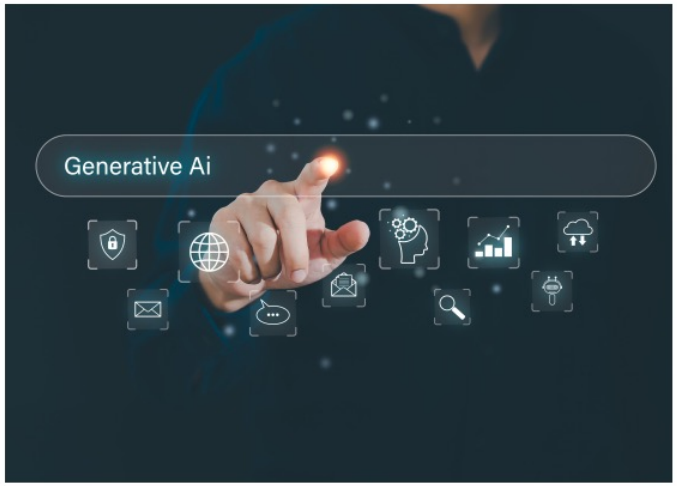
Marjolaine Hamelin

Hadi Quesneville

Tristan Salord

Alban Thomas

**https://science-ouverte.inrae.fr/fr**
DOI : 10.17180/ztym-j930




Accompagnement

**Recommandations INRAE pour l'usage de l'IA générative comme assistant personnel**

20 septembre 2024

INRAE publie ses premières lignes directrices en vue d'un cadrage de l'utilisation des intelligences artificielles génératives comme assistant personnel.