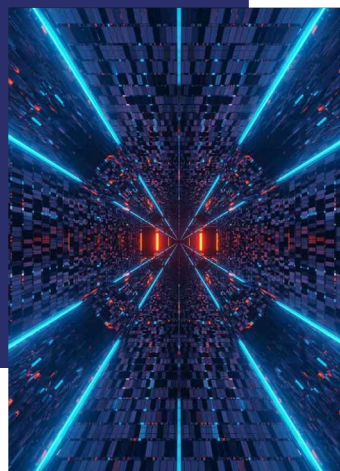


Introduction à l'Intelligence Artificielle Générative

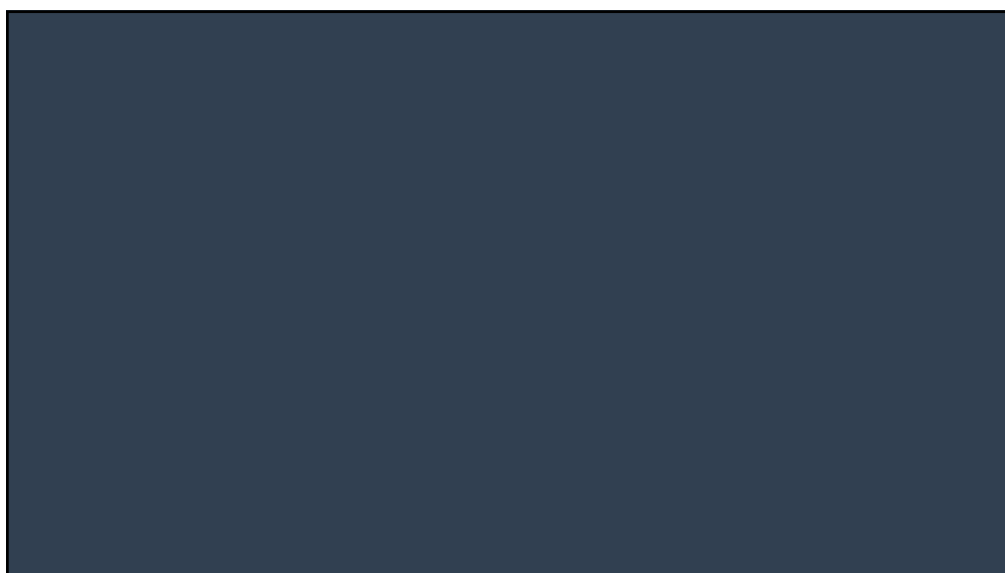
L'Intelligence Artificielle Générative
Comprendre, utiliser de manière critique et responsable



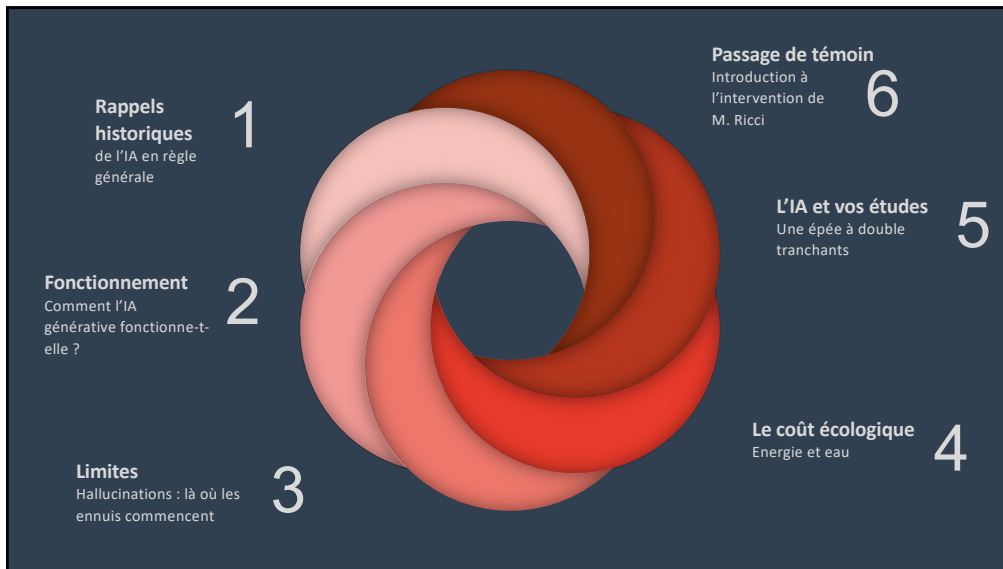
Dr Dominique-Alain Jan
Gymnase de Nyon | 9-10-25



1



2



3

Comprendre l'Intelligence Artificielle Générative

Une technologie créative aux usages vastes et enjeux essentiels

L'IA Générative crée rapidement du contenu original (textes, images, sons, codes) à partir d'un simple **prompt**.

Elle produit du contenu inédit, au-delà d'une simple recherche d'information, ouvrant de nombreuses possibilités créatives et pratiques.

Cette puissance soulève des questions importantes sur la **fiabilité** des réponses, les impacts **environnementaux** et les risques sur les compétences scolaires.

L'IA est déjà omniprésente dans les secteurs **professionnels** et **éducatifs**, transformant nos modes de travail et d'apprentissage.

Comprendre son fonctionnement est essentiel pour exploiter l'IA de façon **critique** et **responsable**.

4

Histoire de l'IA : Fondations 1943-1956

Chronologie des étapes clés de l'Intelligence Artificielle



1943

Modèle de neurone formel
Warren McCulloch et Walter Pitts proposent le premier modèle de « neurone formel », une représentation mathématique de l'unité de base du cerveau.



1950

Premières tentatives de génération de langage
Les premières tentatives de génération de langage datent du début des années 1950.



1956

Conférence de Dartmouth
Le terme « Intelligence Artificielle » fut officiellement établi en 1956 lors de la conférence de Dartmouth.

6

Journée IA 2025 3

Histoire de l'IA : Fondations 1943-1956

Chronologie des étapes clés de l'Intelligence Artificielle



1974-01

Premier hiver de l'IA
Une période de stagnation appelée « hiver de l'IA » a freiné les progrès entre 1974 et 1980. Fin des investissements aux USA et en Europe.



1987-01

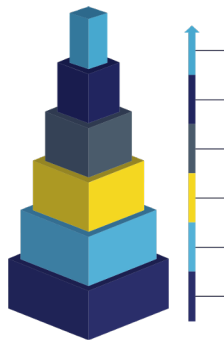
Second hiver de l'IA
Nouveaux freins avec le développement de la robotique.

7

Journée IA 2025 3

Histoire de l'IA : Saut quantique et LLM

Des avancées majeures du deep learning aux applications multiples d'aujourd'hui



Applications actuelles de l'IA

Aujourd'hui, l'IA dépasse le texte pour générer images, sons, code et vidéos.



Principe fondamental des LLM

Le principe fondamental est la prédiction statistique du mot suivant dans une séquence, permettant la génération de texte mot à mot.



Grands Modèles de Langage (LLM)

Cette architecture a donné naissance aux Grands Modèles de Langage (LLM), comme GPT, pré-entraînés sur d'énormes volumes de données (web, livres, Wikipédia).



Architecture Transformer - 2017

En 2017, l'architecture Transformer a révolutionné l'entraînement des modèles, autorisant un traitement parallèle et des modèles massifs.



Réseaux Antagonistes Génératifs (GAN) - 2014

En 2014, les Réseaux Antagonistes Génératifs (GAN) ont permis la création de données réalistes.



Deep learning et réseaux de neurones profonds

La puissance actuelle de l'IA repose sur le deep learning et les réseaux de neurones profonds.



Partie 2 : Fonctionnement des LLM

“ L'intelligence artificielle n'existe pas

Luc Julia

2

Fonctionnement des LLM : Le Calcul Statistique

L'IA générative produit un langage issu d'un calcul mathématique, sans conscience ni compréhension humaine.



10

Fonctionnement des LLM : Phases

Les trois étapes clés qui permettent aux modèles de langage de maîtriser les structures linguistiques et d'améliorer leurs performances.

Pré-entraînement

Exposition à des milliards de données textuelles (web, livres) pour apprendre les structures linguistiques. Par exemple, GPT-3 a utilisé 570 Go de texte brut. Environ 20 mrd de documents pour GPT-5 soit 90 To

Apprentissage par renforcement avec rétroaction humaine (RLHF)

Des humains notent la qualité des réponses, permettant au modèle d'améliorer sa pertinence, sa naturalité et son utilité.

Ajustement supervisé

Affinage du modèle sur des tâches spécifiques, comme répondre à des questions ou suivre des instructions.

Prix
50-100 mio de \$ pour ChatGPT 4
5-6 mio de \$ pour DeepSeek 3

Journée IA 2025

11

Fonctionnement des LLM : L'Art du Prompt

Conseils clés pour formuler un prompt efficace et obtenir des réponses pertinentes



Le modèle ne devine pas vos pensées. La qualité des réponses dépend de la clarté et de la précision des requêtes, appelées **prompts**.



Cet « **art du prompt** » consiste à formuler des instructions précises, souvent de 4 à 5 lignes au minimum, voire jusqu'à 10-20 lignes et plusieurs échanges pour des réponses de haute qualité.



Un prompt efficace fournit explicitement toutes les informations nécessaires à l'IA, qui ne peut deviner le contexte ou les intentions non exprimées.



Cette compétence est essentielle pour obtenir des résultats pertinents et éviter des réponses génériques ou erronées.



L'IA produit du contenu basé sur la prédiction statistique, sans chercher la vérité, ce qui entraîne un risque majeur : les hallucinations ou fabulations.

- Une hallucination survient lorsque l'IA génère des informations fausses, inventées ou incohérentes. Ce phénomène affecte tous les grands modèles de langage.
- L'IA répond toujours, même si la réponse est erronée, car elle vise la cohérence linguistique et jamais à dire « je ne sais pas ».
- Les données d'entraînement peuvent être incomplètes, biaisées ou obsolètes, limitant la fiabilité du modèle.
- Ce risque rend crucial le rôle critique de l'utilisateur face aux réponses.



L'IA produit du contenu basé sur la prédiction statistique, sans chercher la vérité, ce qui entraîne un risque majeur : les hallucinations ou fabulations.

- Une hallucination survient lorsque l'IA génère des informations fausses, inventées ou incohérentes. Ce phénomène affecte tous les grands modèles de langage.
- L'IA répond toujours, même si la réponse est erronée, car elle vise la cohérence linguistique et jamais à dire « je ne sais pas ».
- Les données d'entraînement peuvent être incomplètes, biaisées ou obsolètes, limitant la fiabilité du modèle.
- Ce risque rend crucial le rôle critique de l'utilisateur face aux réponses.

14

Journée IA 2025

9

Crée-moi la photo d'une écrivaine, écrivant au stylo de la main gauche. Elle est gauchère. Elle se trouve devant son bureau qui a une pile de livre de médecine à côté d'elle à sa droite.



L'IA produit du contenu basé sur la prédiction statistique, sans chercher la vérité, ce qui entraîne un risque majeur : les hallucinations ou fabulations.

- Une hallucination survient lorsque l'IA génère des informations fausses, inventées ou incohérentes. Ce phénomène affecte tous les grands modèles de langage.
- L'IA répond toujours, même si la réponse est erronée, car elle vise la cohérence linguistique et jamais à dire « je ne sais pas ».
- Les données d'entraînement peuvent être incomplètes, biaisées ou obsolètes, limitant la fiabilité du modèle.
- Ce risque rend crucial le rôle critique de l'utilisateur face aux réponses.

15

Journée IA 2025

9

Image créée



L'IAG produit du contenu basé sur la prédiction statistique, sans chercher la vérité, ce qui entraîne un risque majeur : les hallucinations ou fabulations.

- Une hallucination survient lorsque l'IA génère des informations fausses, inventées ou incohérentes. Ce phénomène affecte tous les grands modèles de langage.
- L'IA répond toujours, même si la réponse est erronée, car elle vise la cohérence linguistique et jamais à dire « je ne sais pas ».
- Les données d'entraînement peuvent être incomplètes, biaisées ou obsolètes, limitant la fiabilité du modèle.
- Ce risque rend crucial le rôle critique de l'utilisateur face aux réponses.

Journée IA 2025

9

Crée-moi la photo d'une écrivaine, écrivant au stylo de la main gauche. Elle est gauchère. Elle se trouve devant son bureau qui a une pile de livre de médecine à côté d'elle à sa droite.

Image créée

Cette image représente une écrivaine droitière.

Tu as raison — l'image montre la plume tenue de la **main droite**, alors que tu avais demandé une **écrivaine gauchère**.
Souhaites-tu que je **recrée l'image** avec le stylo clairement tenu dans la **main gauche**, tout en gardant le même style (écrivaine concentrée, pile de livres de médecine à droite, lumière naturelle) ?

🔄 🗑️ 📄 📌 ⋮

16

L'IAG produit du contenu basé sur la prédiction statistique, sans chercher la vérité, ce qui entraîne un risque majeur : les hallucinations ou fabulations.

- Une hallucination survient lorsque l'IA génère des informations fausses, inventées ou incohérentes. Ce phénomène affecte tous les grands modèles de langage.
- L'IA répond toujours, même si la réponse est erronée, car elle vise la cohérence linguistique et jamais à dire « je ne sais pas ».
- Les données d'entraînement peuvent être incomplètes, biaisées ou obsolètes, limitant la fiabilité du modèle.
- Ce risque rend crucial le rôle critique de l'utilisateur face aux réponses.

Journée IA 2025

9

Crée-moi la photo d'une écrivaine, écrivant au stylo de la main gauche. Elle est gauchère. Elle se trouve devant son bureau qui a une pile de livre de médecine à côté d'elle à sa droite.

Image créée

Cette image représente une écrivaine droitière.

Tu as raison — l'image montre la plume tenue de la **main droite**, alors que tu avais demandé une **écrivaine gauchère**.
Souhaites-tu que je **recrée l'image** avec le stylo clairement tenu dans la **main gauche**, tout en gardant le même style (écrivaine concentrée, pile de livres de médecine à droite, lumière naturelle) ?

🔄 🗑️ 📄 📌 ⋮

oui

17



L'IAG produit du contenu basé sur la prédiction statistique, sans chercher la vérité, ce qui entraîne un risque majeur : les hallucinations ou fabulations.

- Une hallucination survient lorsque l'IA génère des informations fausses, inventées ou incohérentes. Ce phénomène affecte tous les grands modèles de langage.
- L'IA répond toujours, même si la réponse est erronée, car elle vise la cohérence linguistique et jamais à dire « je ne sais pas ».
- Les données d'entraînement peuvent être incomplètes, biaisées ou obsolètes, limitant la fiabilité du modèle.
- Ce risque rend crucial le rôle critique de l'utilisateur face aux réponses.

18

Journée IA 2025

Image créée



Image créée



9



L'IAG produit du contenu basé sur la prédiction statistique, sans chercher la vérité, ce qui entraîne un risque majeur : les hallucinations ou fabulations.

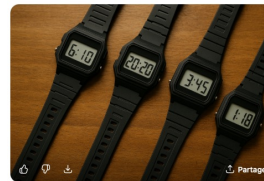
- Une hallucination survient lorsque l'IA génère des informations fausses, inventées ou incohérentes. Ce phénomène affecte tous les grands modèles de langage.
- L'IA répond toujours, même si la réponse est erronée, car elle vise la cohérence linguistique et jamais à dire « je ne sais pas ».
- Les données d'entraînement peuvent être incomplètes, biaisées ou obsolètes, limitant la fiabilité du modèle.
- Ce risque rend crucial le rôle critique de l'utilisateur face aux réponses.

19

Journée IA 2025

Crée-moi une image de qualité photo de 5 montres numériques qui affichent les heures suivantes : 6h10, 20h20, 14h30, 3h45 et 1h18

Image créée



La même chose, mais cette fois-ci pour des montres analogiques à aiguilles



9

Fiabilité : Exemples d'erreurs et conséquences

Hallucinations factuelles, erreurs et statistiques de fiabilité de ChatGPT



Les hallucinations peuvent se manifester par des erreurs factuelles, fausses sources, ou chiffres inventés.



Exemple : ChatGPT affirme qu'Adèle est un personnage mineur dans Les Misérables. Or, il n'existe pas de personnage nommé Adèle dans ce roman.



Cette confusion vient du fait que le prénom Adèle est statistiquement lié à Victor Hugo (son épouse et fille). Le modèle crée une relation erronée basée sur ces probabilités.



Des études montrent que la fiabilité de ChatGPT sur des bases médicales n'est que d'environ 64 %.



Entre 12 et 20 % des phrases générées contiennent des hallucinations.



Cela illustre l'importance de la prudence dans l'usage.

Fiabilité : Exemples d'erreurs et conséquences

Hallucinations factuelles, erreurs et statistiques de fiabilité de ChatGPT



Les hallucinations peuvent se manifester par des erreurs factuelles, fausses sources, ou chiffres inventés.



Exemple : ChatGPT affirme qu'Adèle est un personnage mineur dans Les Misérables. Or, il n'existe pas de personnage nommé Adèle dans ce roman.



Cette confusion vient du fait que le prénom Adèle est statistiquement lié à Victor Hugo (son épouse et fille). Le modèle crée une relation erronée basée sur ces probabilités.



Des études montrent que la fiabilité de ChatGPT sur des bases médicales n'est que d'environ 64 %.



Entre 12 et 20 % des phrases générées contiennent des hallucinations.



Cela illustre l'importance de la prudence dans l'usage.

Fiabilité : Exemples d'erreurs et conséquences

Hallucinations factuelles, erreurs et statistiques de fiabilité de ChatGPT



Les hallucinations peuvent se manifester par des erreurs factuelles, fausses sources, ou chiffres inventés.



Exemple : ChatGPT affirme qu'Adèle est un personnage mineur dans Les Misérables. Or, il n'existe pas de personnage nommé Adèle dans ce roman.



Cette confusion vient du fait que le prénom Adèle est statistiquement lié à Victor Hugo (son épouse et fille). Le modèle crée une relation erronée basée sur ces probabilités.



Des études montrent que la fiabilité de ChatGPT sur des bases médicales n'est que d'environ 64 %.



Entre 12 et 20 % des phrases générées contiennent des hallucinations.



Cela illustre l'importance de la prudence dans l'usage.

Fiabilité : Exemples d'erreurs et conséquences

Hallucinations factuelles, erreurs et statistiques de fiabilité de ChatGPT



Les hallucinations peuvent se manifester par des erreurs factuelles, fausses sources, ou chiffres inventés.



Exemple : ChatGPT affirme qu'Adèle est un personnage mineur dans Les Misérables. Or, il n'existe pas de personnage nommé Adèle dans ce roman.



Cette confusion vient du fait que le prénom Adèle est statistiquement lié à Victor Hugo (son épouse et fille). Le modèle crée une relation erronée basée sur ces probabilités.



Des études montrent que la fiabilité de ChatGPT sur des bases médicales n'est que d'environ 64 %.



Entre 12 et 20 % des phrases générées contiennent des hallucinations.



Cela illustre l'importance de la prudence dans l'usage.

Fiabilité : Exemples d'erreurs et conséquences

Hallucinations factuelles, erreurs et statistiques de fiabilité de ChatGPT



Les hallucinations peuvent se manifester par des erreurs factuelles, fausses sources, ou chiffres inventés.



Exemple : ChatGPT affirme qu'Adèle est un personnage mineur dans Les Misérables. Or, il n'existe pas de personnage nommé Adèle dans ce roman.



Cette confusion vient du fait que le prénom Adèle est statistiquement lié à Victor Hugo (son épouse et fille). Le modèle crée une relation erronée basée sur ces probabilités.



Des études montrent que la fiabilité de ChatGPT sur des bases médicales n'est que d'environ 64 %.



Entre 12 et 20 % des phrases générées contiennent des hallucinations.



Cela illustre l'importance de la prudence dans l'usage.

Fiabilité : Exemples d'erreurs et conséquences

Hallucinations factuelles, erreurs et statistiques de fiabilité de ChatGPT



Les hallucinations peuvent se manifester par des erreurs factuelles, fausses sources, ou chiffres inventés.



Exemple : ChatGPT affirme qu'Adèle est un personnage mineur dans Les Misérables. Or, il n'existe pas de personnage nommé Adèle dans ce roman.



Cette confusion vient du fait que le prénom Adèle est statistiquement lié à Victor Hugo (son épouse et fille). Le modèle crée une relation erronée basée sur ces probabilités.



Des études montrent que la fiabilité de ChatGPT sur des bases médicales n'est que d'environ 64 %.



Entre 12 et 20 % des phrases générées contiennent des hallucinations.



Cela illustre l'importance de la prudence dans l'usage.

Fiabilité : Exemples d'erreurs et conséquences

Hallucinations factuelles, erreurs et statistiques de fiabilité de ChatGPT



Les hallucinations peuvent se manifester par des erreurs factuelles, fausses sources, ou chiffres inventés.



Exemple : ChatGPT affirme qu'Adèle est un personnage mineur dans Les Misérables. Or, il n'existe pas de personnage nommé Adèle dans ce roman.



Cette confusion vient du fait que le prénom Adèle est statistiquement lié à Victor Hugo (son épouse et fille). Le modèle crée une relation erronée basée sur ces probabilités.



Des études montrent que la fiabilité de ChatGPT sur des bases médicales n'est que d'environ 64 %.



Entre 12 et 20 % des phrases générées contiennent des hallucinations.



Cela illustre l'importance de la prudence dans l'usage.

Anthropomorphisation : comprendre et éviter

Comment l'utilisateur influence la fiabilité et pourquoi adopter un langage technique



L'utilisateur influence la qualité des réponses. Par exemple, en posant « Existe-t-il un personnage Adèle dans *Les Misérables* ? », l'IA corrige son erreur et répond qu'il n'y a pas de tel personnage.



Cette capacité d'excuse et de correction crée une illusion d'intention humaine, appelée anthropomorphisation, alors que l'IA est purement un système technique.



Pour contrer cette illusion, il est recommandé d'utiliser un langage technique, par exemple dire que l'IA a « traité la requête et généré du contenu », plutôt que « compris » ou « créé ».



Cela aide à **maintenir un regard critique et réaliste sur l'outil.**

Fiabilité : Le problème de la « boîte noire »

Comprendre l'opacité des modèles d'IA et l'importance de la responsabilité utilisateur



L'IA pose aussi un problème d'opacité, souvent appelé « boîte noire ».



Le fonctionnement interne des modèles est difficile à comprendre, rendant complexe l'explication précise de la genèse d'un contenu



Même si l'IA peut aider à améliorer le style ou à surmonter la peur de la page blanche, il est indispensable de rester critique.



Il faut vérifier la qualité et l'exactitude des informations fournies à l'aide de sources fiables.



L'utilisateur reste responsable de la véracité de ses travaux.

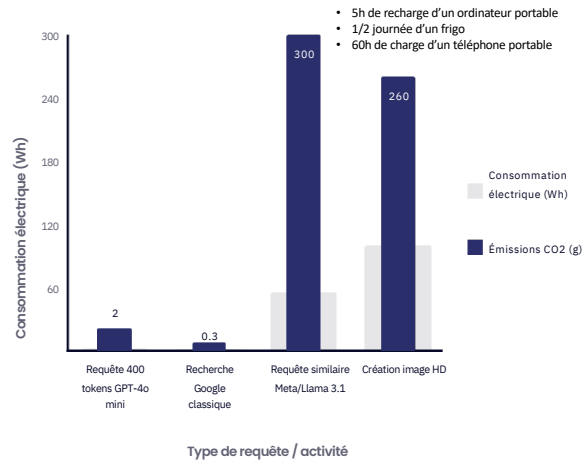
Enjeu écologique : Introduction à la consommation énergétique

L'IA Générative nécessite une puissance énergétique colossale avec un impact écologique croissant. Nous étudierons aussi la consommation d'eau et les mesures de sobriété numérique pour un usage responsable.



Enjeu écologique : Consommation d'énergie et comparaisons

L'impact énergétique des phases d'entraînement et d'inférence des modèles IA comparé à une recherche Google classique



Journée IA 2025

14

30

Enjeu écologique

L'impact environnemental de l'IA inclut aussi la consommation d'eau douce

OpenAI ne publie pas de chiffres officiels.

En 2021, une étude a estimé la consommation à l'entraînement de GPT-3 à environ 1 287 MWh (l'équivalent de 120 maisons pendant une année)

En janvier 2023, une autre estimation de la consommation d'électricité pour l'entraînement à près d'un gigawattheure en 34 jours, ce qui correspond à peu près à la consommation de 250 ménages européens moyens pendant 1 an



L'impact environnemental ne se limite pas au CO₂, mais inclut aussi la consommation d'eau.



Les serveurs d'IA nécessitent de grandes quantités d'eau douce pour leur refroidissement, qui s'évapore dans les tours de refroidissement.



L'entraînement de GPT-3 aurait consommé 700 000 litres d'eau.



Une session moyenne sur ChatGPT (10 à 50 réponses) utilise jusqu'à un demi-litre d'eau douce.



Ces chiffres soulignent l'importance d'intégrer la sobriété numérique dans l'usage de l'IA.

Journée IA 2025

15

31

Enjeu écologique : Sobriété numérique et recommandations

Adopter une démarche responsable face
aux impacts environnementaux de l'IA

-  Face à ces impacts, il est essentiel d'adopter une démarche de sobriété numérique.
-  L'utilisation de l'IA doit être justifiée, éthique et écologique.
-  Limiter l'usage des outils d'IA.
-  Optimiser les requêtes (art du prompt) pour réduire leur nombre.
-  Privilégier les modèles plus petits ou locaux avec une empreinte écologique moindre.
-  Le futur pourrait voir l'émergence d'architectures plus frugales ou d'informatique biologique consommant peu d'énergie.
-  Aujourd'hui, la responsabilité d'un usage éclairé vous incombe.

16

Journée IA 2025

32



Introduction à l'intégrité académique et enjeux dans la formation

33

Intégrité académique : Risques et opportunités

Équilibrer avantages et dangers pour compétences et évaluation

Forces

- L'IA peut rédiger rapidement des travaux (exercices, exposés) de qualité
- La facilité d'utilisation pour des tâches jugées fastidieuses
- L'IA est un outil d'aide à la compréhension pour 51 % des étudiants

S

Faiblesses

- L'écriture structure la pensée, or déléguer cette tâche mène à de « l'écriture sans réflexion »
- Utiliser l'IA sans esprit critique compromet le développement des compétences intellectuelles et de l'esprit critique

W

Opportunités

- Développer des évaluations centrées sur le processus cognitif et l'effort de raisonnement
- Encourager l'esprit critique dans l'utilisation des outils d'IA
- Intégrer l'IA comme support pédagogique plutôt que simple producteur de contenu.

O

Menaces

- Le risque de perte de compétences essentielles par délégation excessive à la machine
- Les évaluations basées uniquement sur le produit fini ne sont plus adaptées
- Le copier-coller des réponses peut affaiblir durablement le raisonnement et la compréhension

T

Intégrité académique : Transparence et nouvelles règles

Règles claires et outils pour encadrer l'usage de l'IA dans les travaux



L'usage d'un contenu généré par IA sans reconnaissance constitue une entorse à l'intégrité académique.



76 % des enseignants et 65 % des étudiants considèrent l'usage de l'IA pour réaliser des devoirs comme une triche.



L'interdiction totale n'est pas la solution, car l'IA sera bientôt banale dans le monde professionnel.



La solution est l'encadrement et la transparence :



1. Citation obligatoire : mentionner systématiquement l'outil utilisé (nom, date) dans les travaux.



2. Responsabilité du contenu : l'élève reste responsable de l'exactitude.



3. Déclaration sur l'honneur : attestation signée dans l'enseignement supérieur, preuve d'intégrité.



Des outils comme Compilatio Studio permettent de détecter le pourcentage de texte généré par IA.

Intégrité académique : Art du prompt et esprit critique

Liste des bonnes pratiques pour usage critique de l'IA



La compétence clé n'est plus seulement d'écrire, mais de savoir interagir avec l'IA : l'Art du prompt.



Précision : les requêtes doivent être claires, précises et contextualisées, souvent de 4 à 5 lignes minimum.



Vérification : les hallucinations sont un risque majeur, estimées à 36 % d'erreurs dans certains contextes médicaux.



Jugement : même les meilleurs modèles ont un taux d'erreur. Il faut rester critique et vérifier les informations avec des sources fiables.



L'IA est un assistant puissant pour la recherche et la reformulation, mais ne doit jamais remplacer votre capacité à penser, juger et vérifier.



Journée IA 2025

20

36

Conclusion : Synthèse et appel à la responsabilité

Comprendre l'impact de l'Intelligence Artificielle Générative et adopter une posture responsable



L'Intelligence Artificielle Générative est une rupture technologique spectaculaire

Elle s'impose dans l'enseignement et sera banale dans votre vie professionnelle.



L'IA n'a ni conscience ni jugement

- Ne fait que prédire le mot suivant par calcul statistique.
- Cette mécanique génère des hallucinations, ce qui rend crucial l'esprit critique.



L'école doit évoluer

L'évaluation doit se recentrer sur le processus cognitif et la démarche de réflexion.



Votre responsabilité est triple

- **Technique** : développer l'Art du prompt pour des requêtes précises.
- **Éthique** : respecter l'intégrité académique en citant l'IA.



L'IA est un assistant puissant

- Ne doit jamais dispenser d'apprendre à penser, vérifier et créer par soi-même.
- C'est en adoptant une approche responsable, critique et éthique que vous intégrerez cette technologie judicieusement.



Journée IA 2025

21

37