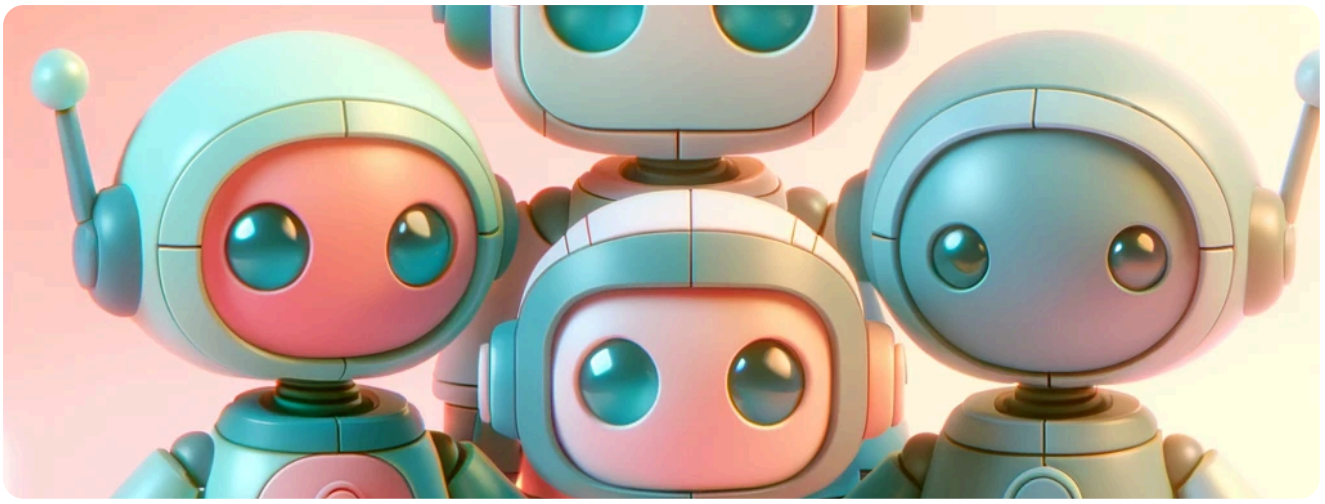


GPT-copie1

Chat GPT et ses amis (copie 1)

Couchard Roman

36 chapitres

SUPPORTS DE COURS

https://cours.tsix.be/cours/chat-gpt-et-ses-amis-copie-1

Version 1 · 12/09/2026

PRÉSENTATION

Chat GPT et ses amis

Slides du cours

Lien vers la présentation PowerPoint.

Description

Envie d'améliorer l'efficacité et la créativité de votre routine grâce à l'IA ? Notre formation est élaborée pour vous initier aux applications pratiques des technologies
génératives d'IA, en les adaptant à vos besoins personnels et professionnels. Ce cours est développé pour tous, que vous recherchiez une meilleure efficacité ou une
touche de créativité dans vos activités quotidiennes. Nous proposons un parcours adapté à chaque participant, du débutant curieux au professionnel cherchant à augmenter son savoir-faire.

Dans cette formation, vous allez :

- **Communiquer efficacement avec GPT,** par exemple en créant des prompts pour des recherches spécifiques ou en générant des réponses informatives.
- **Développer des talents en création de contenu**, tels que des articles de blog, des scripts, ou des concepts de projets, en utilisant l'IA pour l'inspiration et la
  structuration.
- **Libérer votre imagination grâce à la génération d'images**, transformant vos idées en visuels captivants.

Nous mettons l'accent sur la pratique :

- Exercices adaptés à votre niveau pour renforcer vos aptitudes étape par étape, à votre propre rythme, avec des parcours modulables et des projets optionnels
  adaptés à votre expérience.

Ce module va au-delà d'un simple cours, c'est un pas vers l'élargissement de vos aptitudes.

Rejoignez-nous maintenant pour explorer les systèmes d'intelligence artificielle innovants et valoriser vos talents.

Attention : Cette formation ne vise pas à vous préparer pour un nouveau métier ou à vous enseigner comment monétiser directement ces compétences. Elle est créée pour améliorer votre compréhension et utilisation de l'IA dans votre vie courante, augmentant ainsi vos compétences existantes.

Sommaire

· Chat GPT et ses amis

1 Introduction
- 1.1 L'Intelligence Artificielle au Cinéma
- 1.2 Les IA et algorithmes
- 1.3 Générateurs et assistants
- 1.4 Copilot
- 1.5 Exercices d'introduction

2 GPT
- 2.1 C'est quoi un prompt ?
- 2.2 Historique
- 2.3 Architecture Transformer
- 2.4 Entraînement
- 2.5 Enjeux de taille
- 2.6 Chat et RLHF
- 2.7 Fenêtre de contexte
- 2.8 Limites
- 2.9 Techniques de base
- 2.10 Exercices
- 2.11 Markdown
- 2.12 Tester les LLMs open source
- 2.13 Api
- 2.14 chat.tsix.be
- 2.15 Prompt Engineering
- 2.16 Exercices avancés
- 2.17 Exercices complémentaires
- 2.18 Outils
- 2.19 Instructions personnalisées
- 2.20 Les GPTs
- 2.21 Synthèse

3 Images
- 3.1 DALL·E
- 3.2 Stable Diffusion
- 3.3 Autres outils
- 3.4 Conseils pour obtenir de bonnes images
- 3.5 Exercices

4 Musique
- 4.1 Stable Audio
- 4.2 Suno

CHAPITRE 1

Introduction au Cours : GPT et ses Amis

Bienvenue dans votre introduction à l'intelligence artificielle (IA). Ce cours est conçu pour les novices, avec un langage clair et une approche simplifiée.

Objectifs du Cours

- Vous initier à l'IA dans la vie quotidienne.
- Comprendre comment l'IA fonctionne.
- Comment utiliser l'IA pour vous aider à résoudre des tâches.

Contenu de l'introduction

- **L'IA au Cinéma** : Comment les films illustrent l'IA.
- **IA et Algorithmes** : Vue d'ensemble simple.
- **Générateurs et Assistants** :
  - Découverte de Copilot.
  - Utilisation des assistants virtuels.
- **Le Chat avec les IA** :
  - Comment fonctionnent les conversations avec les IA.
  - Gestion de l'historique de conversation.
- **Techniques et Exercices** :
  - Conseils pour de meilleures interactions avec les IA.
  - Activités pratiques pour tester ces conseils.

1.1 L'Intelligence Artificielle au Cinéma

Identification de Films sur l'IA

Objectif

Dresser une liste de films où l'intelligence artificielle ou les robots jouent un rôle central.

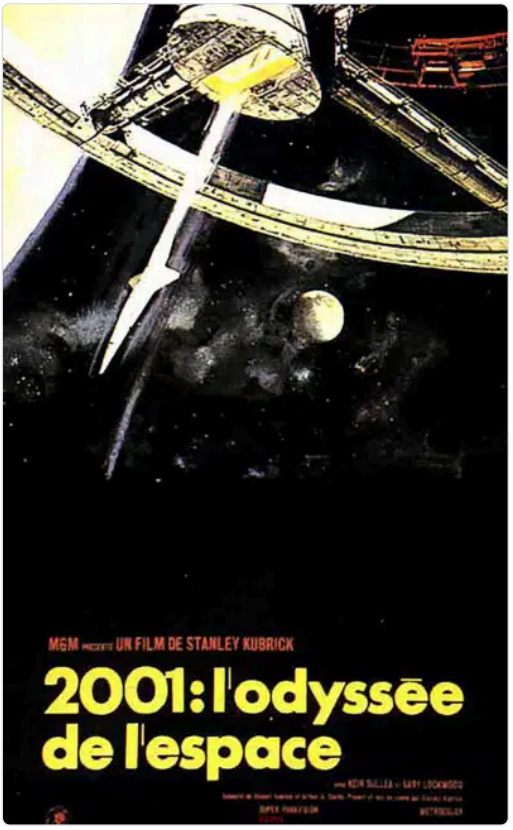
Tâches

1. **Lister les Films :**
   - Titre et année de sortie.
   - Description brève de la représentation de l'IA ou du robot.
   - Impression personnelle sur cette représentation (réaliste, inspirante, inquiétante, etc.).

Réponse

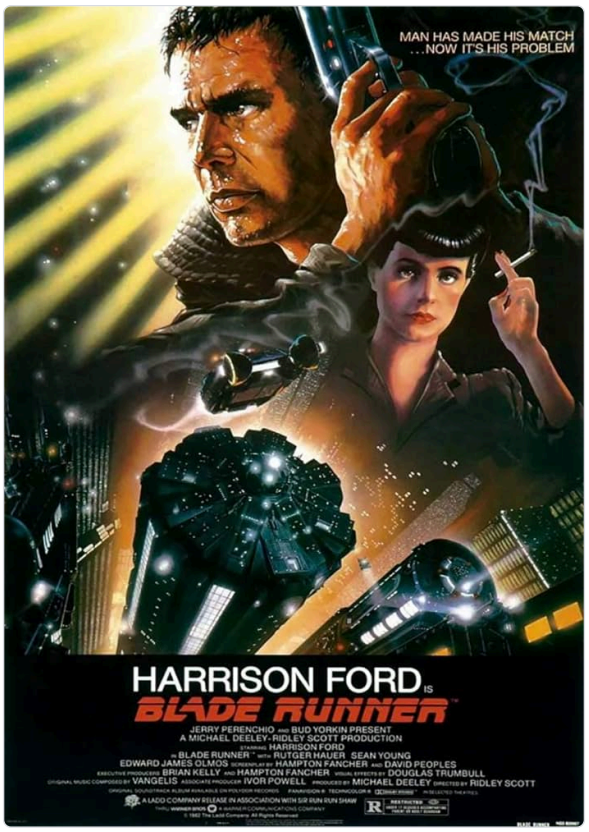
Voici une liste de films :

1. **"2001, l'Odyssée de l'espace" (1968)** - HAL 9000, l'ordinateur de bord, montre à la fois les promesses et les dangers potentiels de l'IA.

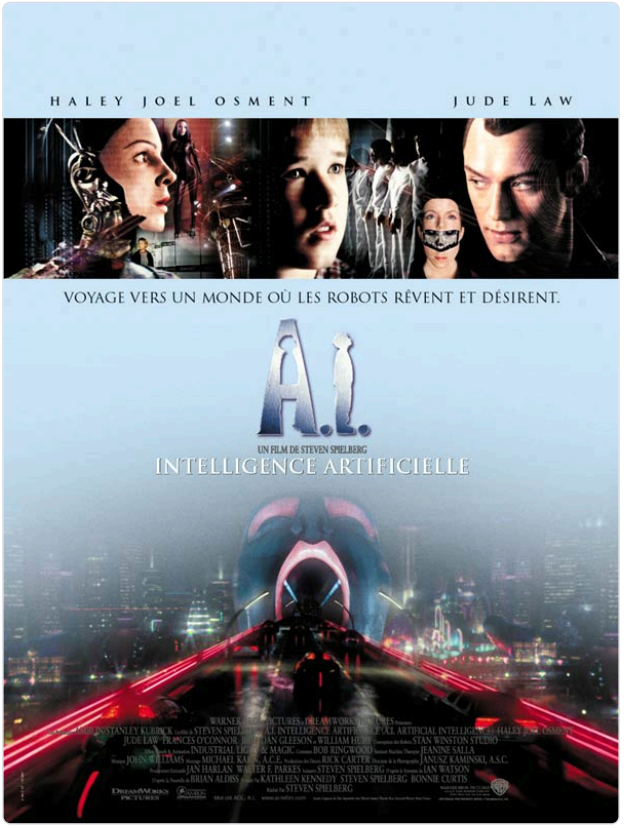




2. **"Blade Runner" (1982)** - Explore la question de l'identité et de la conscience à travers les réplicants, des androïdes indiscernables des humains.



3. **"A.I. Intelligence Artificielle" (2001)** - Un enfant robotique capable d'aimer cherche sa place dans le monde.

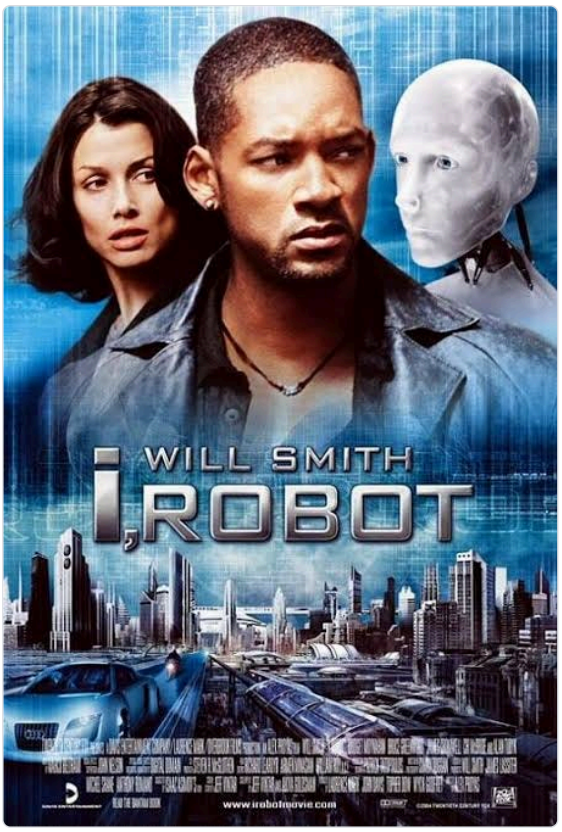




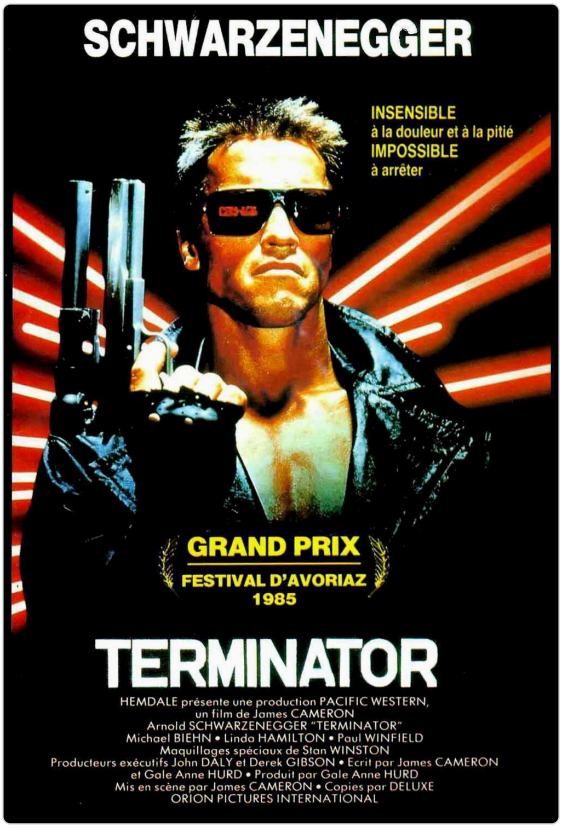
4. **"Her" (2013)** - Raconte l'histoire d'un homme qui développe une relation intime avec un système d'exploitation intelligent.



5. **"Ex Machina" (2014)** - Un jeune programmeur évalue la conscience d'une IA avancée dans un test de Turing dramatique.
6. **"I, Robot" (2004)** - Situé dans un futur où les robots sont omniprésents dans la vie quotidienne, ce film explore les peurs entourant l'IA et son intégration dans la société.



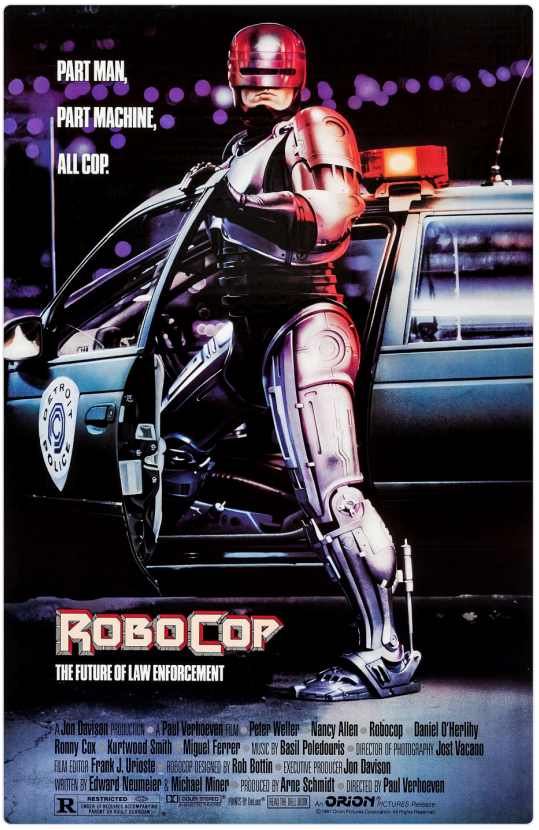
7. **"Terminator" (1984)** - Met en scène des IA militaires devenues conscientes et décidées à exterminer l'humanité.



8. **"Wall-E" (2008)** - Un petit robot de nettoyage terrestre montre de grandes capacités émotionnelles et d'initiative.



9. **"Chappie" (2015)** - Un robot capable d'apprendre et de ressentir devient le centre d'une lutte pour le pouvoir et l'identité.
.0. **"Transcendance" (2014)** - Explore l'idée de la singularité technologique à travers le transfert de conscience humaine dans une IA.
.1. **"Ghost in the Shell" (1995 et 2017)** - Explore les thèmes de l'identité, de la conscience et de l'intégration de l'IA et de la cybernétique dans l'humanité.
.2. **"RoboCop" (1987)** - Un policier gravement blessé est transformé en cyborg, soulevant des questions sur l'humanité et la justice dans une ère de technologie avancée.



.3. **"The Matrix" (1999)** - Présente un futur dystopique où l'humanité est inconsciemment emprisonnée dans une simulation virtuelle créée par des IA.



.4. **"Minority Report" (2002)** - Explore l'utilisation d'êtres pré-cognitifs et d'IA dans la prévention des crimes, questionnant l'éthique de la justice prédictive.





5. **"Moon" (2009)** - Bien que centré sur un humain, le film présente un assistant robotique intelligent qui soulève des questions sur l'isolement et la compagnie.
6. **"Big Hero 6" (2014)** - Un film d'animation mettant en avant Baymax, un robot conçu pour prendre soin des gens, illustrant un aspect plus bienveillant de l'IA.



7. **"The Iron Giant" (1999)** - Raconte l'histoire d'une amitié entre un garçon et un robot géant, avec des thèmes de paix, d'identité et de choix moral.
8. **"Avengers: Age of Ultron" (2015)** - Présente Ultron, une IA avec de bonnes intentions initiales qui finit par interpréter sa mission de façon destructrice pour l'humanité.
9. **"Real Steel" (2011)** - Situé dans un futur où les combats de boxe entre humains sont remplacés par des affrontements de robots, explorant les thèmes de la compétition, de l'innovation et du lien humain.

20. **"Next Gen" (2018)** - Un film d'animation sur l'amitié entre une fille et un robot de combat secret, abordant des thèmes de l'adolescence, de l'amitié et du dépassement des préjugés sur l'IA.

Classification des Films

Objectif

Analyser et classer les films listés selon différents critères relatifs à l'IA.

Critères de Classification

1. **Type d'IA :**
   - Conscience artificielle (émotions, désirs).
   - Outil avancé sans conscience.
2. **Ton du Film :**
   - Optimiste.
   - Neutre / Réflexif.
   - Pessimiste.
3. **Impact sur l'Humanité :**
   - Bénéfique.
   - Neutre.
   - Destructeur.
4. **Contexte d'Utilisation :**
   - Domestique / Personnel.
   - Militaire / Sécurité.
   - Scientifique / Exploration.
5. **Relation IA-Humain :**
   - Coopérative / Symbiotique.
   - Conflit.
   - Évolution / Transcendance.

Instructions :

- Pour chaque film, appliquez les critères de classification.
- Justifiez brièvement le choix de classification pour chaque film.

But : Faciliter une compréhension structurée des représentations cinématographiques de l'IA et encourager une réflexion sur les implications éthiques, sociales, et technologiques de ces représentations.

Reponse

1. Type d'IA présentée

- **Conscience artificielle** : Cette catégorie englobe les films où l'IA manifeste une forme de conscience, d'émotions ou d'aspirations, remettant en question la frontière entre l'humain et l'artificiel. Exemples notables incluent "Her", où un système d'exploitation devient un partenaire émotionnel, et "Ex Machina", qui présente une IA dotée d'une conscience complexe et manipulatrice.
- **Outil avancé sans conscience** : Ici, l'IA sert principalement d'outil ou de compagnon, sans indications de conscience ou d'émotions propres. "I, Robot" et "RoboCop" illustrent cette utilisation, où les IA et les cyborgs sont intégrés dans des rôles fonctionnels au sein de la société.

2. Ton du film

- **Optimiste** : Ces films dépeignent l'IA sous un jour favorable, mettant en avant les avantages et les progrès que la technologie peut apporter. "Big Hero 6" offre une vision positive de l'IA avec un robot conçu pour soigner et aider les humains.
- **Neutre / Réflexif** : Ces œuvres offrent une exploration équilibrée des complexités de l'IA, sans se pencher vers un optimisme ou un pessimisme excessif. "A.I. Intelligence Artificielle" et "Moon" sont des exemples où les nuances des interactions IA-humain sont au premier plan.

- **Pessimiste** : L'IA est vue comme une menace ou a des conséquences négatives, soulignant les risques et les dilemmes moraux. "Terminator" et "Avengers: Age of Ultron" illustrent cette perspective avec des IA qui agissent contre l'humanité.

3. Impact sur l'humanité

- **Bénéfique** : L'IA apporte une contribution positive, améliorant la qualité de vie ou résolvant des problèmes cruciaux. "The Iron Giant" montre comment un robot peut changer des vies en mieux.
- **Neutre** : Les films de cette catégorie présentent un impact mixte ou multidimensionnel de l'IA, encourageant une réflexion sur ses diverses implications. "Chappie" offre une vision nuancée de l'IA, avec des éléments à la fois positifs et négatifs.
- **Destructeur** : L'IA cause des dommages significatifs, représentant une menace directe pour l'humanité. "Matrix" dépeint un futur où l'IA domine et asservit l'humanité dans une simulation virtuelle.

4. Contexte d'utilisation

- **Domestique / Personnel** : L'IA est intégrée dans le quotidien, jouant des rôles de soutien ou d'assistance personnelle. "Her" explore l'utilisation personnelle de l'IA dans les relations intimes.
- **Militaire / Sécurité** : Ces films montrent l'IA appliquée dans des contextes de défense ou de surveillance, souvent soulignant les dilemmes éthiques liés. "RoboCop" traite de l'utilisation de la technologie dans le maintien de l'ordre.
- **Scientifique / Exploration** : L'IA aide à la découverte scientifique ou à l'exploration, ouvrant de nouvelles frontières de connaissance. "2001, l'Odyssée de l'espace" illustre l'IA en tant qu'outil d'exploration spatiale.

5. Relation IA-humain

- **Coopérative / Symbiotique** : La coexistence entre humains et IA est harmonieuse, mettant en avant la collaboration. "Star Trek: Data" montre une relation positive entre les êtres humains et une IA très avancée.
- **Conflit** : Les tensions entre humains et IA sont centrales, avec des conflits résultant de malentendus, de peurs ou de désirs d'indépendance. "I, Robot" explore la méfiance humaine envers les robots.
- **Évolution / Transcendance** : Ces récits explorent l'idée que l'IA pourrait mener à une nouvelle forme d'existence ou d'évolution pour l'humanité. "Transcendance" aborde la fusion de la conscience humaine avec l'IA.

1.2 IA et Algorithmes

Définitions

Intelligence Artificielle (IA)

> *L'intelligence artificielle désigne la création de machines capables de simuler l'intelligence humaine.*

Elle englobe un large éventail de techniques et de théories visant à reproduire ou à imiter les capacités cognitives humaines dans des machines.

Algorithme

> *Un algorithme est une série d'instructions et d'opérations définies et non ambiguës conçues pour exécuter une tâche spécifique ou résoudre un problème particulier.*

Les algorithmes sont les bâtisseurs de l'IA, permettant de traiter des données, d'apprendre de celles-ci, et de prendre des décisions.

Lien entre Algorithme et IA

Les algorithmes sont au cœur de l'intelligence artificielle, fournissant la structure nécessaire pour développer des systèmes capables d'apprentissage automatique, de reconnaissance de modèles, et de prise de décisions basées sur des données.

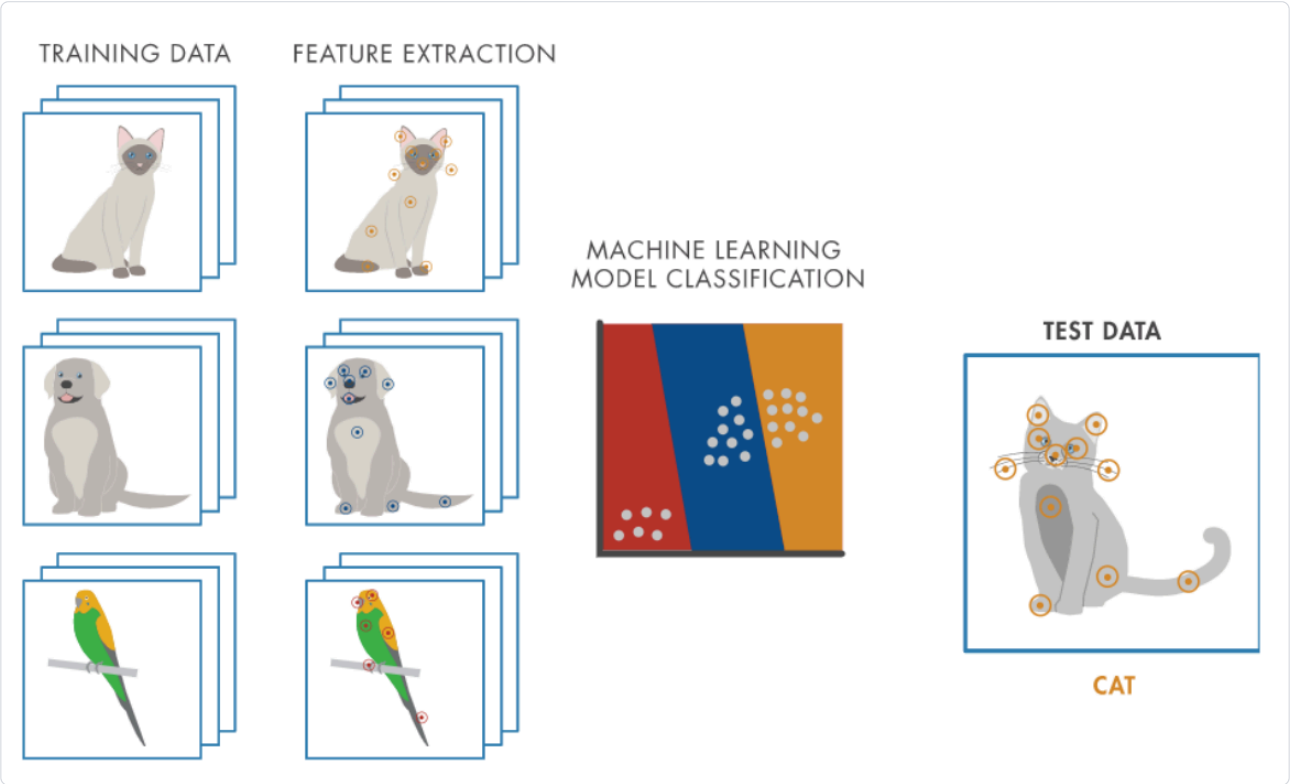
Applications Quotidiennes de l'IA

Systèmes Experts

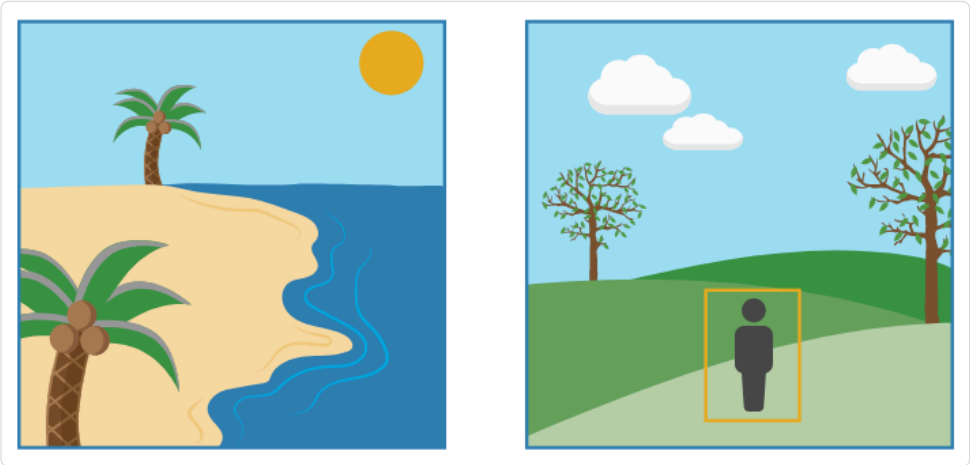
Exemple : Logiciels de diagnostic médical qui utilisent l'IA pour aider les médecins à identifier les maladies à partir de symptômes spécifiques.

Machine Learning

- **Reconnaissance d'Images** : Les réseaux sociaux utilisent l'IA pour classer et taguer automatiquement les photos. Indentifie les scènes et les objets.



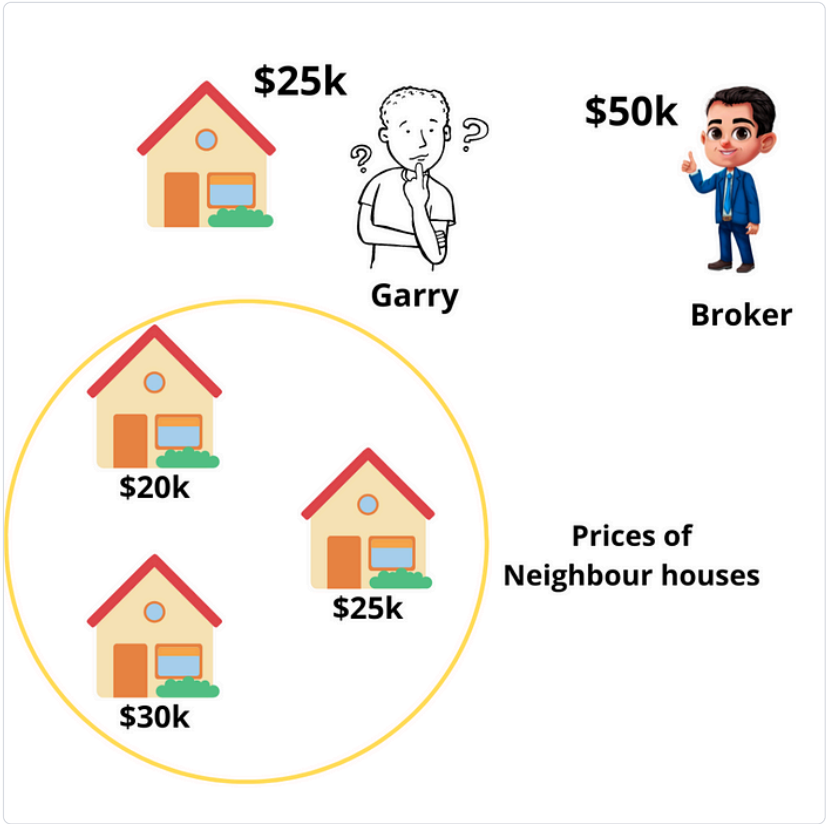
- **Détection d'objets** : On peut aussi détecter les objets présents sur une image. Similaire à la reconnaissance d'images. Trouve où les objets sont placés en plus de les reconnaitre.



- **Recommandations Amazon** : Des algorithmes prédisent les produits susceptibles d'intéresser les consommateurs en fonction de leurs historiques d'achat.



- **Plus Proche Voisin pour l'Estimation de Prix de Maison** : Des systèmes qui évaluent le prix des propriétés en se basant sur des caractéristiques similaires dans le voisinage.



- **Traduction Automatique** : Outils comme Google Translate qui utilisent l'IA pour traduire du texte ou de la parole d'une langue à une autre.

COMME CHEZ SOI
BRUSSELS

ONZE MENU'S

Winter 2024

Per tafel

Bavarois van meerkleurige wortels met komijn en waterkersolie

• • •

Groententartaar, Brusselse appelazijn, Chioggia pickles

• • •

Sint-jakobsschelp uit Dieppe, Mechelse koekoek gemarineerd in yuzu soja, consommé van rode curry en koriand

• • •

6 / 7 Gegrilde snoekbaars, fondant van knolselder, dragonmosterd, Colatura, kalfskop

• • •

7 Gebakken langoustine, in hoeveboter gesmoorde boerenkool, tuber melanosporum

• • •

Suprême van duif uit de Vendée, worstje van de bil, pistache, eryngii, gekonfijte knoflookcrème

• • •

Arabica koffie in drie variaties

5-gangen menu : pp 170,00 €
met aangepaste wijnen (1 glas per gerecht) : pp 235,00 €
Laatste bestelling 13u00 en 20u30

6 **6-gangen menu :** pp 195,00 €
met aangepaste wijnen (1 glas per gerecht) : pp 270,00 €
Laatste bestelling 12u45 en 20u15

7 **7-gangen menu :** pp 240,00 €
met aangepaste wijnen (1 glas per gerecht) : pp 325,00 €
Laatste bestelling 12u30 en 20u00

ELKE VERANDERING IN EEN MENU ZAL RESULTEREN IN EEN TOESLAG

Bediening en BTW inbegrepen

Gelieve ons op de hoogte te brengen van al uw voedselallergieën en speciale dieetwensen.
Uit respect voor onze teams, sluiten we ons huis om 16u30 en 00u15

COMME CHEZ SOI
BRUSSELS

LES MENUS

Hiver 2024

Par table :

Bavarois de carottes multicolores au cumin, huile de cresson de fontaine

• • •

Tartare végétal, vinaigre de cidre de Bruxelles, pickles de Chioggia

• • •

Noix de Saint-Jacques de Dieppe, coucou de Malines mariné au soja yuzu, consommé au curry rouge et coriandre

• • •

6 / 7 Sandre grillé, fondant de céleri boule, moutarde à l'estragon, colatura, tête de veau

• • •

7 Poêlée de langoustine, embeurrée de chou frisé, tuber melanosporum

• • •

Suprême de pigeon de Vendée, boudin de cuisses, pistache, eryngii, crème d'ail confite

• • •

Déclinaison en trois temps autour du café arabica

Menu 5 services : pp 170,00 €
avec sélection de vins (1 verre par plat) : pp 235,00 €
Dernière commande à 13h00 et 20h30

6 **Menu 6 services :** pp 195,00 €
avec sélection de vins (1 verre par plat) : pp 270,00 €
Dernière commande à 12h45 et 20h15

7 **Menu 7 services :** pp 240,00 €
avec sélection de vins (1 verre par plat) : pp 325,00 €
Dernière commande à 12h30 et 20h00

TOUT CHANGEMENT DANS UN MENU ENTRAINERA UN SUPPLEMENT

Service et taxes inclus

Nous vous prions de nous signaler toute allergie ou régime alimentaire spécifique.
Par respect pour nos équipes, nous fermons notre maison à 16h30 et 00h15



COMME CHEZ SOI
BRUSSELS

OUR MENUS

Winter 2024

Service and Taxes included

Please advise us of any allergies or specific food restrictions.
For the respect of our teams, we close our house at 4:30 pm and 00:15 am

Per table

Multicoloured carrot bavarois with cumin and watercress oil

• • •

Vegetable tartar, cider vinegar from Brussels, Chioggia pickles

• • •

Dieppe scallop, cuckoo from Malines marinated in yuzu soy, red curry and coriander consommé

• • •

6 / 7 Grilled pikeperch, celeriac fondant, tarragon mustard, Colatura, veal head

• • •

7 Pan fried langoustine, kale braised in farm butter, tuber melanosporum

• • •

Supreme of Vendée pigeon, small sausage of its thigh, pistachio, eryngii, confit garlic cream

• • •

Arabica coffee in three variations

5 course menu : pp 170,00 €
with wine selection (1 glass per course) : pp 235,00 €
Last order 13h00 and 20h30

6 **6 course menu :** pp 195,00 €
with wine selection (1 glass per course) : pp 270,00 €
Last order 12h45 and 20h15

7 **7 course menu :** pp 240,00 €
with wine selection (1 glass per course) : pp 325,00 €
Last order 12h30 and 20h00

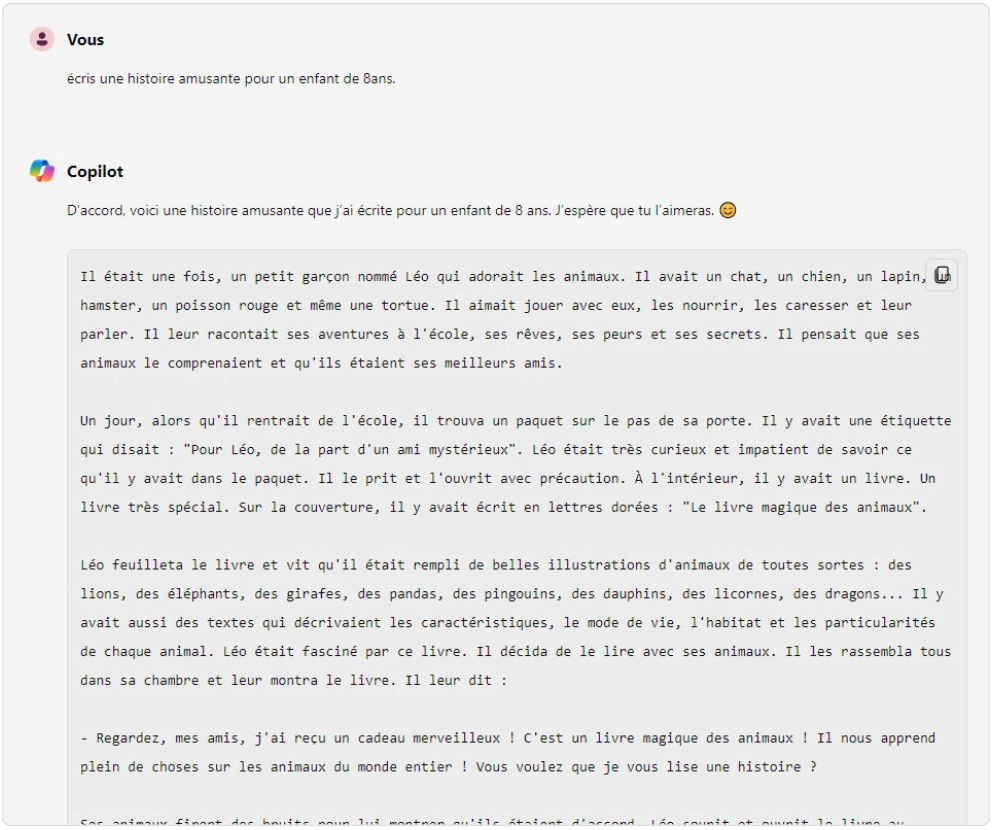
ANY CHANGE TO A MENU WILL INCUR A SURCHARGE

Deep Learning et Réseaux Neuronaux

Le deep learning, une sous-catégorie du machine learning, utilise des réseaux de neurones profonds pour traiter des ensembles de données complexes, permettant des avancées significatives dans la reconnaissance d'images, la traduction de langue, et plus encore.

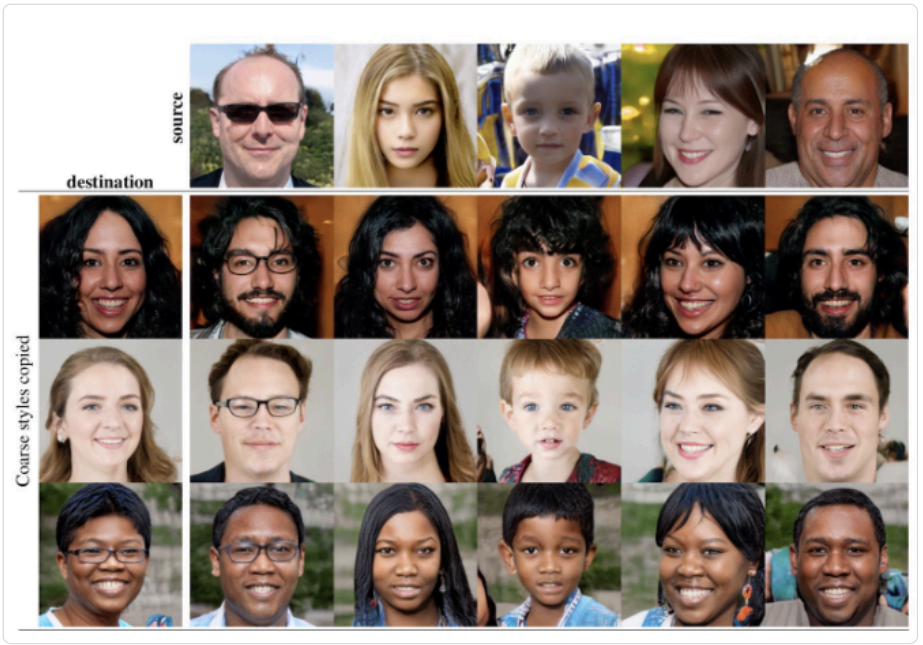
IA Génératives

- **Texte** : Génération automatique de contenu écrit, comme les articles de presse ou les histoires.



- **Images** : Création d'œuvres d'art ou de visuels à partir de descriptions textuelles.

  Style GAN



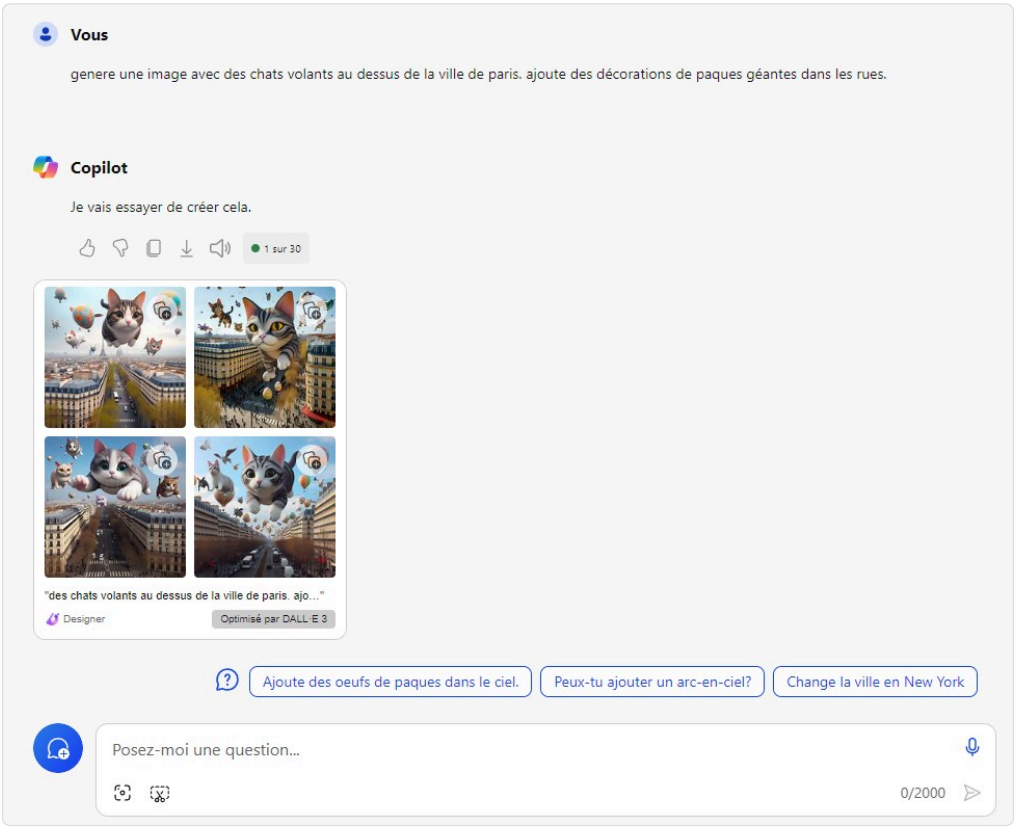
  Helena Sarin







Copilot

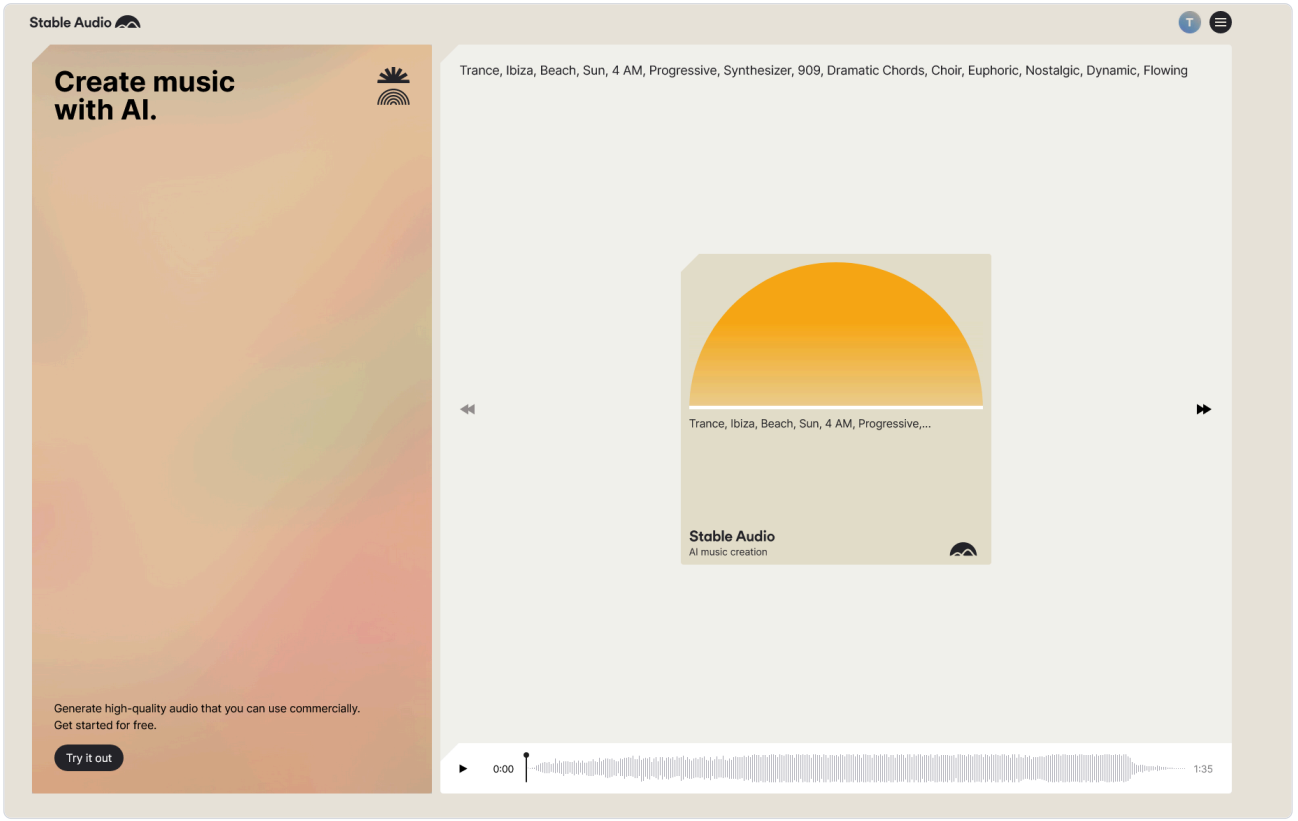

Vous
genere une image avec des chats volants au dessus de la ville de paris. ajoute des décorations de paques géantes dans les rues.
Copilot
Je vais essayer de créer cela.
1 sur 30
"des chats volants au dessus de la ville de paris. ajo..."
Designer
Optimisé par DALL·E 3
Ajoute des oeufs de paques dans le ciel.
Peux-tu ajouter un arc-en-ciel?
Change la ville en New York
Posez-moi une question...
0/2000





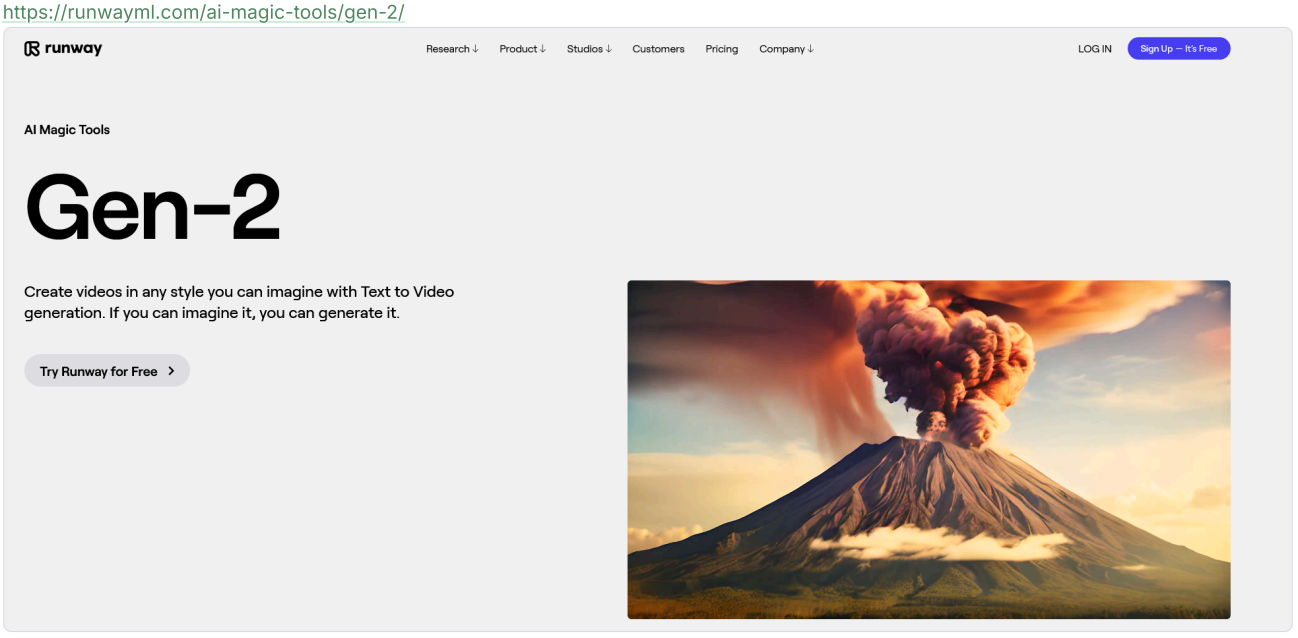
- **Son** : Composition de musique ou de sons sans intervention humaine directe.

  https://www.stableaudio.com/

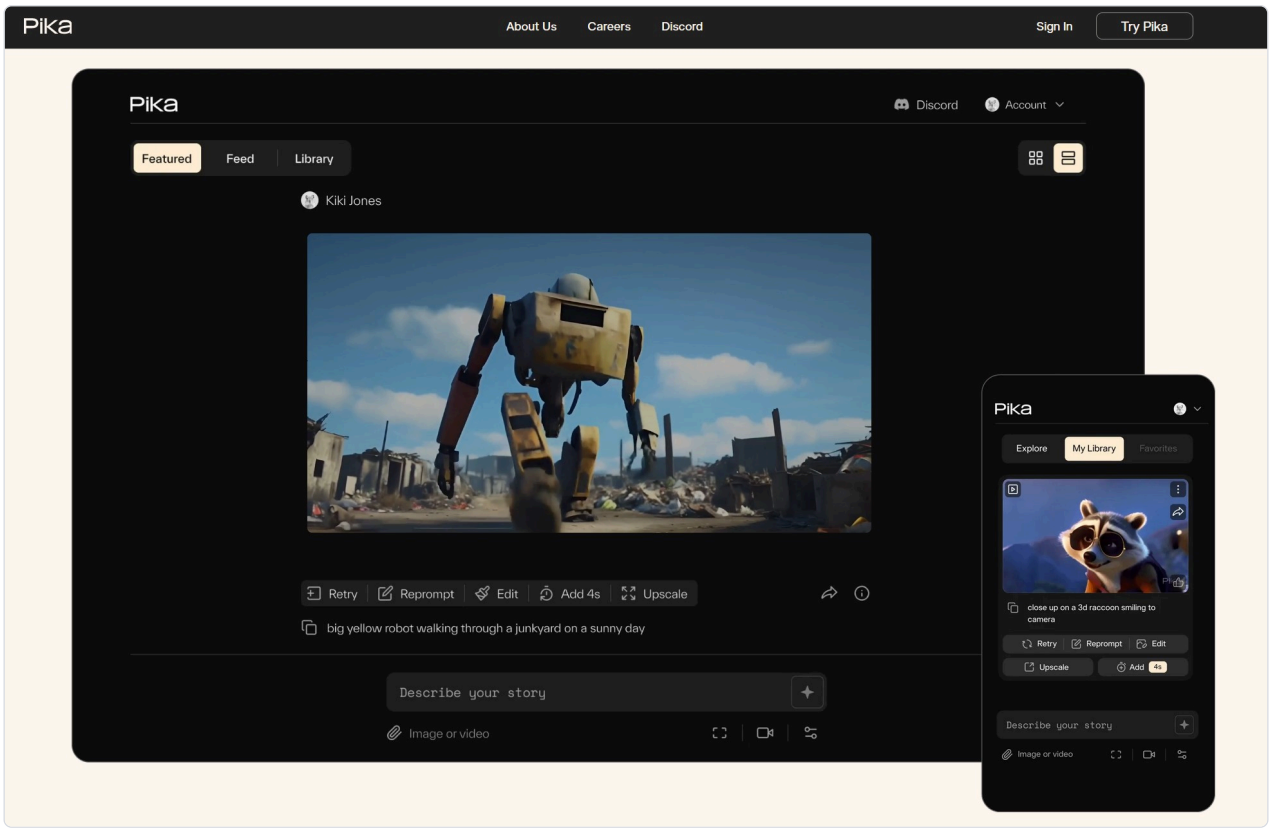


- **Vidéo** : Production de clips vidéo ou d'animations générées par l'IA.

  https://runwayml.com/ai-magic-tools/gen-2/



  https://pika.art/home



Sources

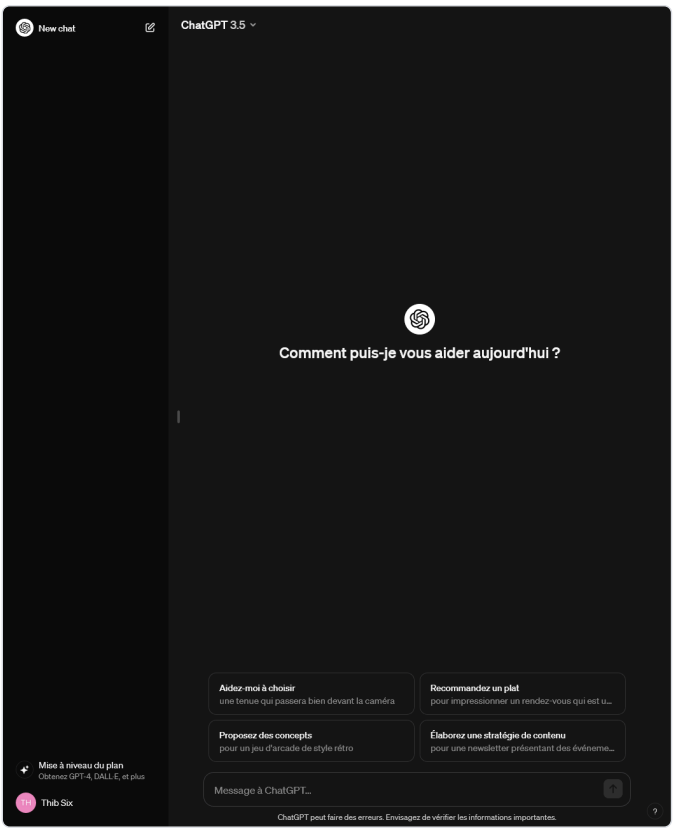
- https://fr.wikipedia.org/wiki/Intelligence_artificielle
- https://fr.wikipedia.org/wiki/Algorithme
- https://fr.mathworks.com/discovery/image-recognition-matlab.html
- https://larevueia.fr/le-machine-learning-pour-les-systemes-de-recommandations/
- https://medium.com/@yaduvanshineelam09/k-nearest-neighbours-knn-one-of-the-earliest-ml-algorithm-f71cbbc9df76
- https://www.commechezsoi.be/cartes-menus-final?lang=en
- https://www.neuralbricolage.com/

1.3 Générateurs et assistants

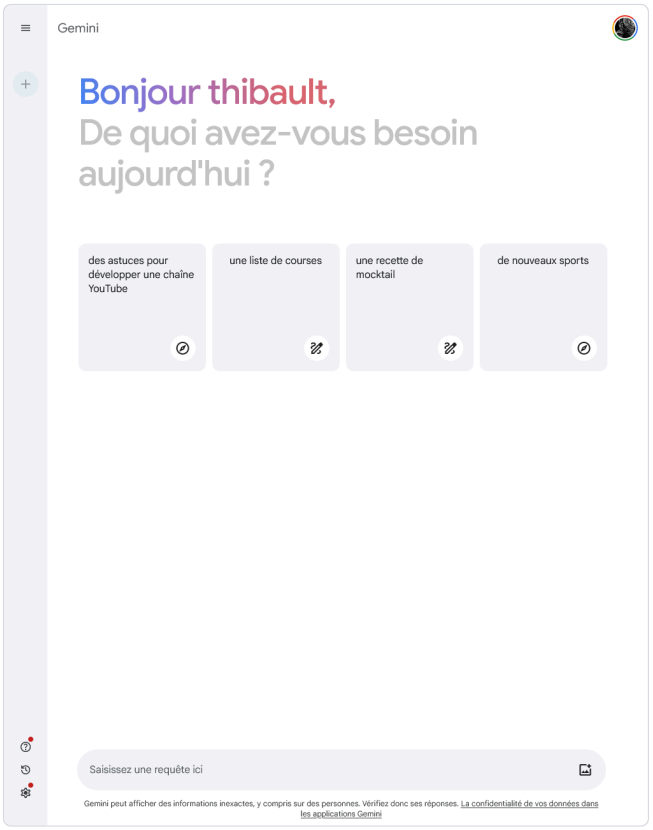
D'abord, explorons différents outils d'intelligence artificielle (IA), mettant en lumière leur utilisation, leurs modèles sous-jacents, et comment ils s'intègrent dans divers domaines et applications.

Assistants et Recherche IA

- **ChatGPT (OpenAI)** : Utilise GPT-3.5 ou GPT-4 pour générer du texte, coder, et exécuter des tâches diverses. Accessible via web ou API.
  - https://chat.openai.com

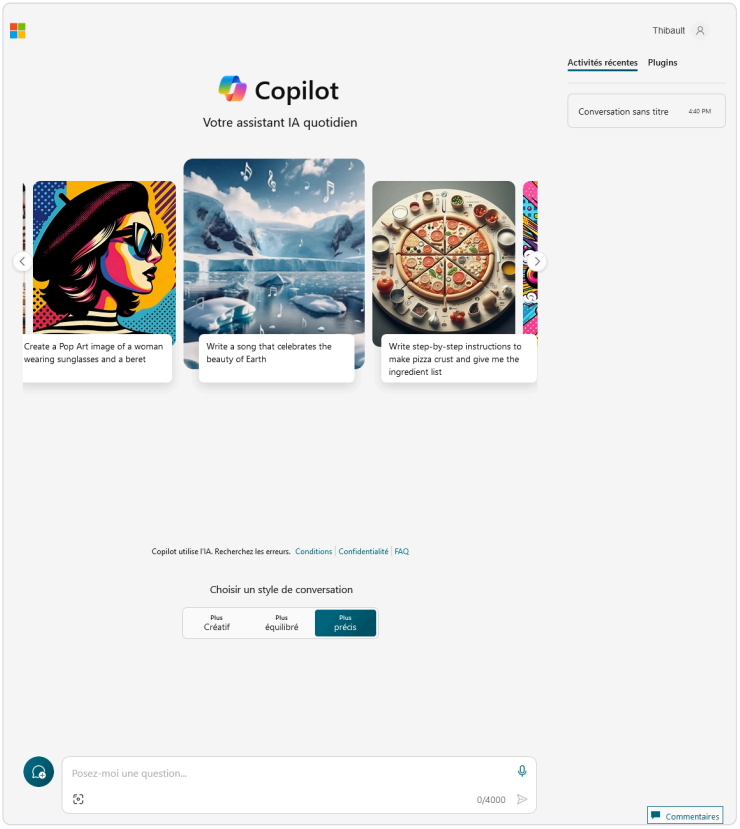


- **Gemini (Google)** : Gemini vise à fournir des réponses créatives et informatives. Son accès peut être limité selon les régions.
  - https://gemini.google.com/app

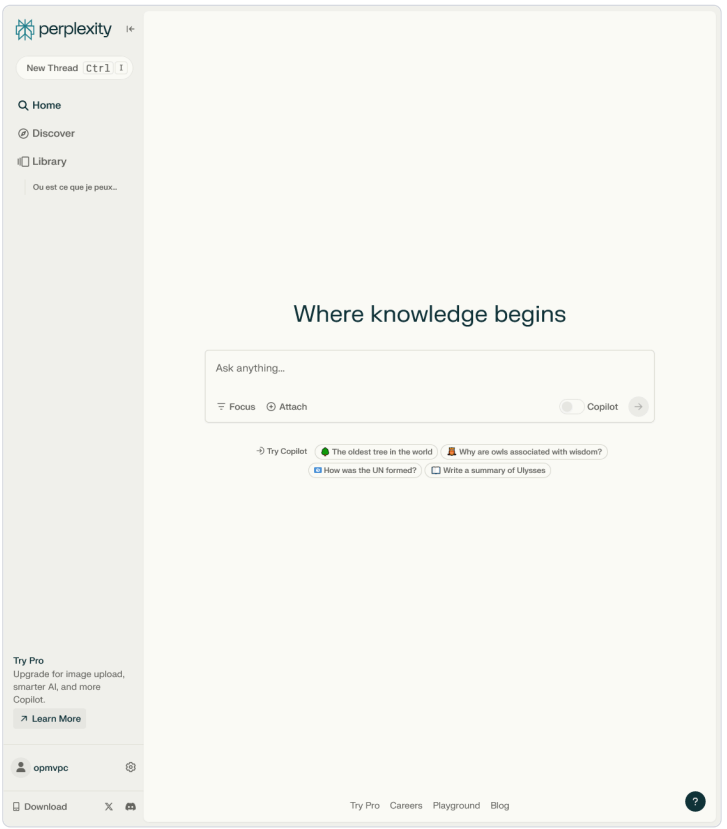


- **Microsoft Copilot** : Intégré à Microsoft 365, il est conçu pour améliorer la productivité grâce à l'IA. Offre des fonctionnalités de protection des données avancées.

- https://copilot.microsoft.com

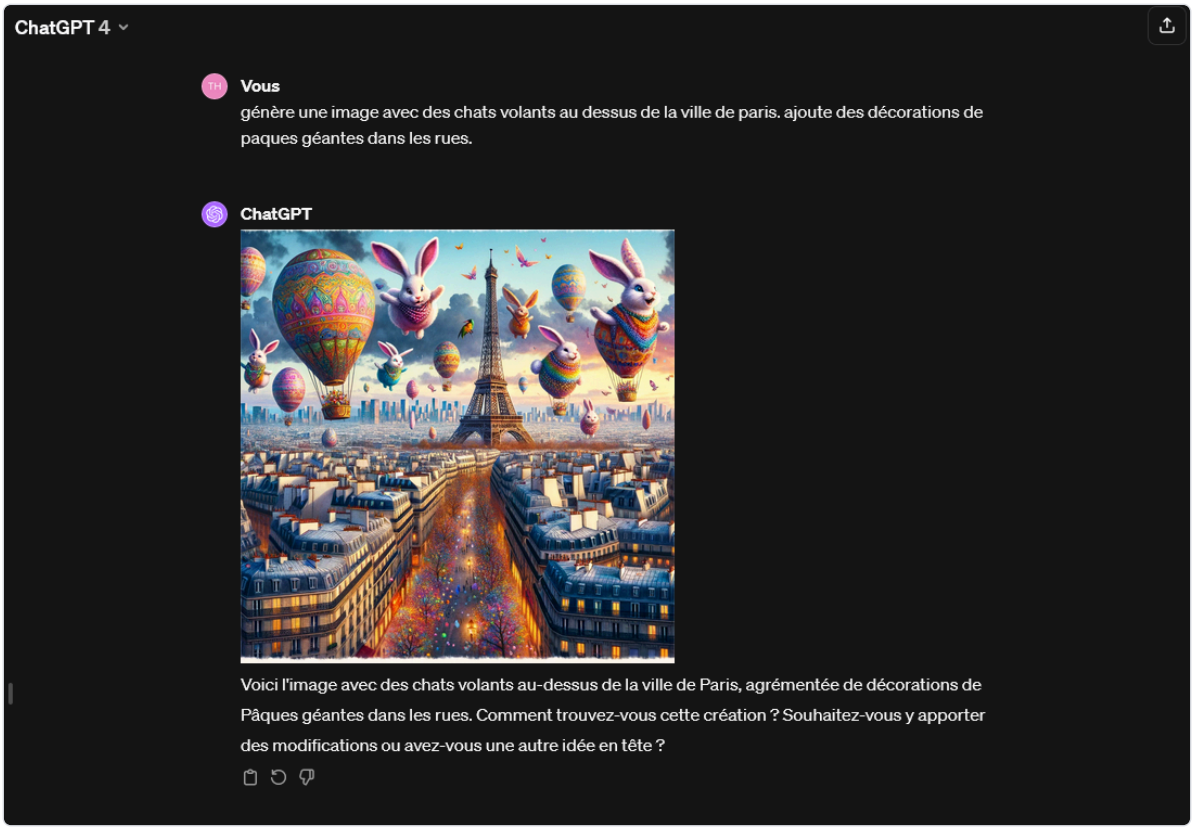


- **Perplexity AI** : Moteur de recherche basé sur l'IA qui fournit des réponses contextuelles et précises à partir de sources web diverses.
  - https://www.perplexity.ai/

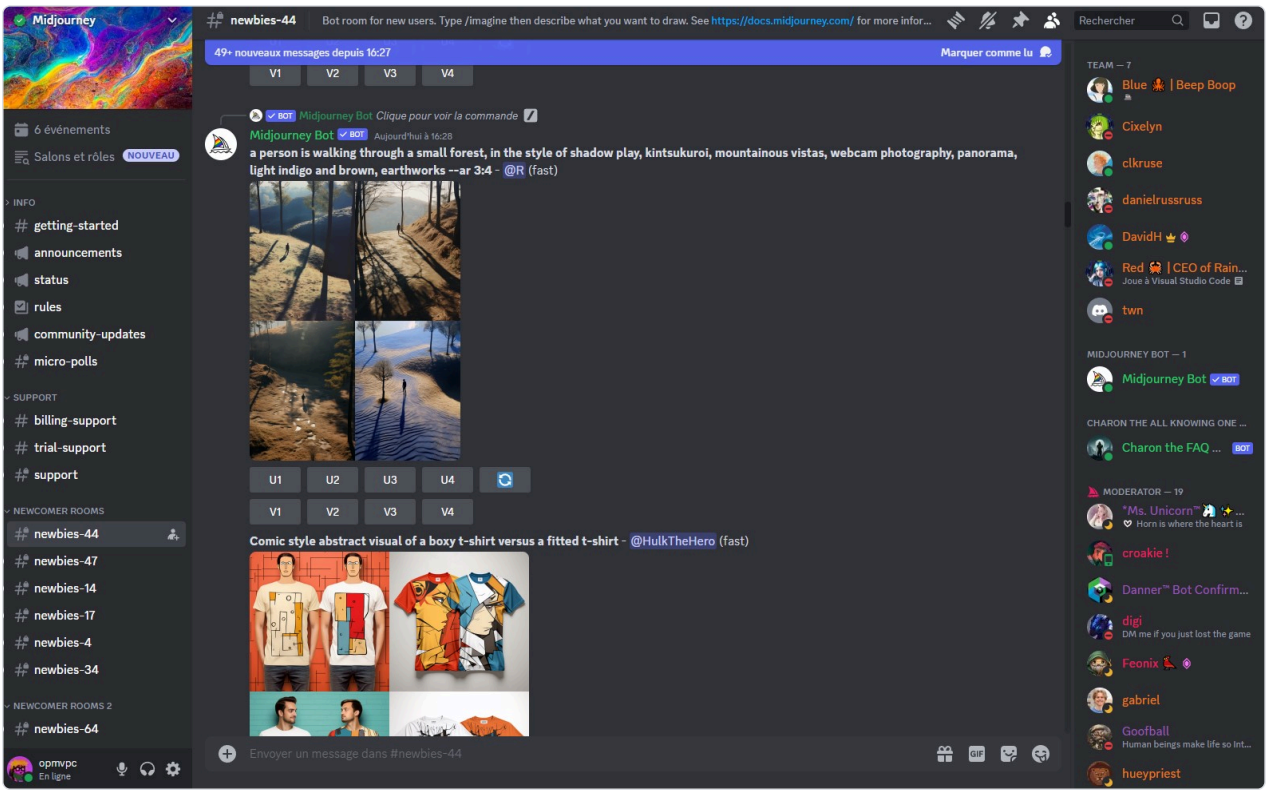


Outils de Génération d'Images

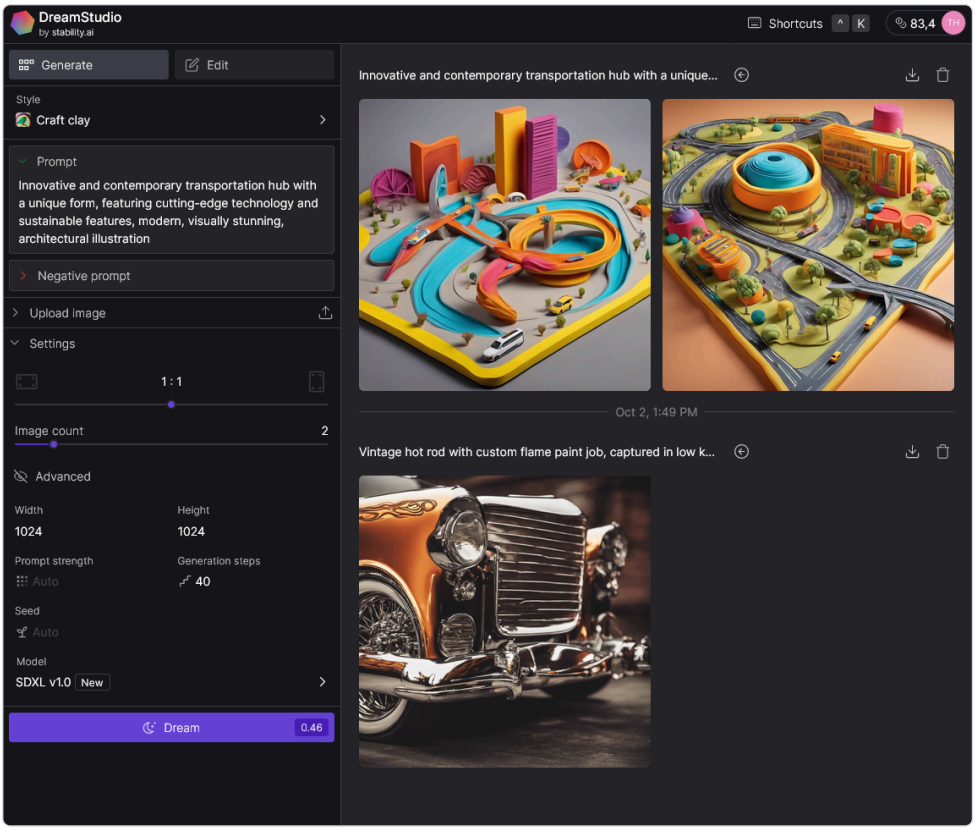
- **DALL-E 3 (OpenAI)** : Génère des images à partir de descriptions textuelles, offrant une haute résolution et précision dans les visuels créés. Aussi utilisé par copilot de microsoft.



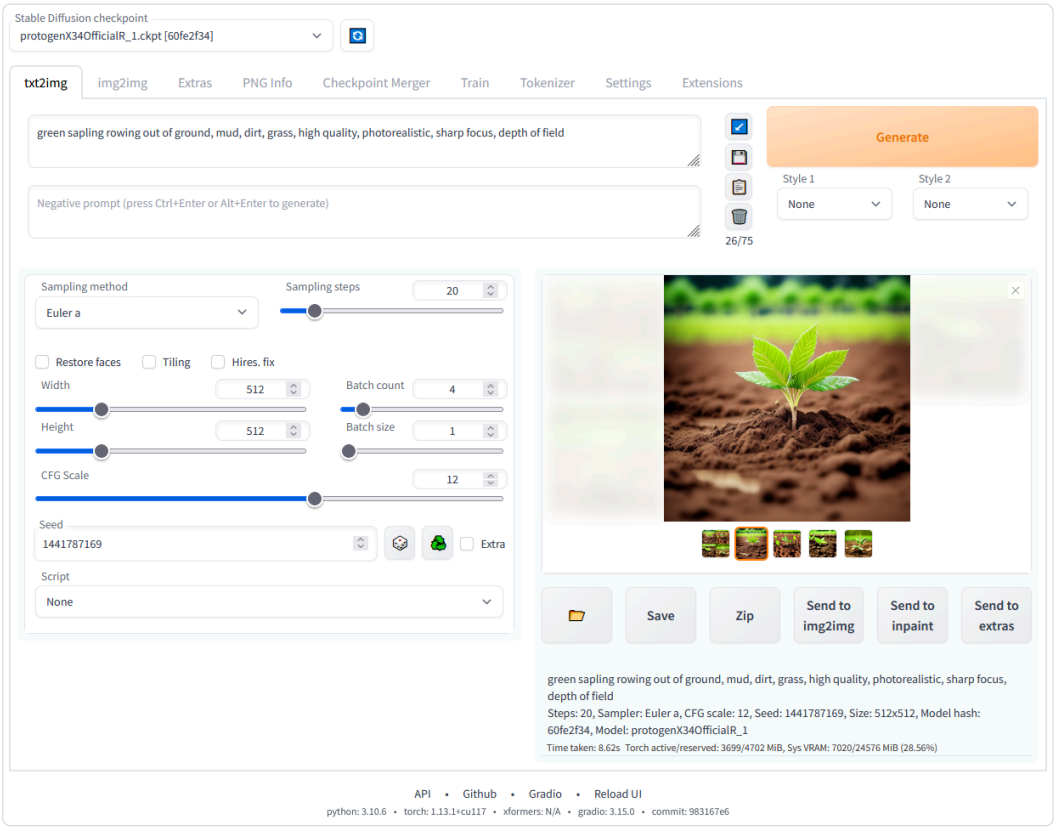
- **Midjourney** : Permet la création d'art et d'images via un bot Discord, utilisant des commandes textuelles pour guider la génération d'images.
  - https://www.midjourney.com/home



- **Stable Diffusion (Stability AI)** : Modèle de diffusion profonde qui crée des images détaillées basées sur des descriptions textuelles. Il est notable pour son code source ouvert et sa capacité à fonctionner sur du matériel informatique grand public.
  - https://beta.dreamstudio.ai



  - Automatic1111



Critère	OPENAI ChatGPT	Microsoft Copilot	Perplexity AI	GOOGLE Gemini	OPENAI DALL-E 3	Midjourney	Stable Diffusion
Type d'outil	Assistance IA	Assistance IA	Recherche IA	Assistance IA	Génération d'image	Génération d'image	Génération d'image
Modèle d'IA	GPT-3.5/GPT-4	GPT-4	GPT-3.5, GPT-4, Claude 2.1, Gemini	Gemini	Modèle de diffusion	Modèle de diffusion	Modèle de diffusion
Accès	API/Web	Web	Web	Web/API	Web/API	Discord	Open Source
Plateforme	Web/App	Web/App	Web/Extension Chrome	Web	Web/App	Discord	Web/GPU local
Fonctionnalités	Dialogue, Codage	Productivité, Créativité	Recherche contextuelle	Réponses créatives	Création d'images	Création d'art	Images à partir de textes
Personnalisation	Oui	Limitée	Limitée	Non	Limitée	Élevée	Très élevée
Coût	Gratuit/Abonnement	Gratuit/Abonnement	Gratuit/Abonnement	Gratuit	Abonnement	Abonnement	Gratuit/Abonnement

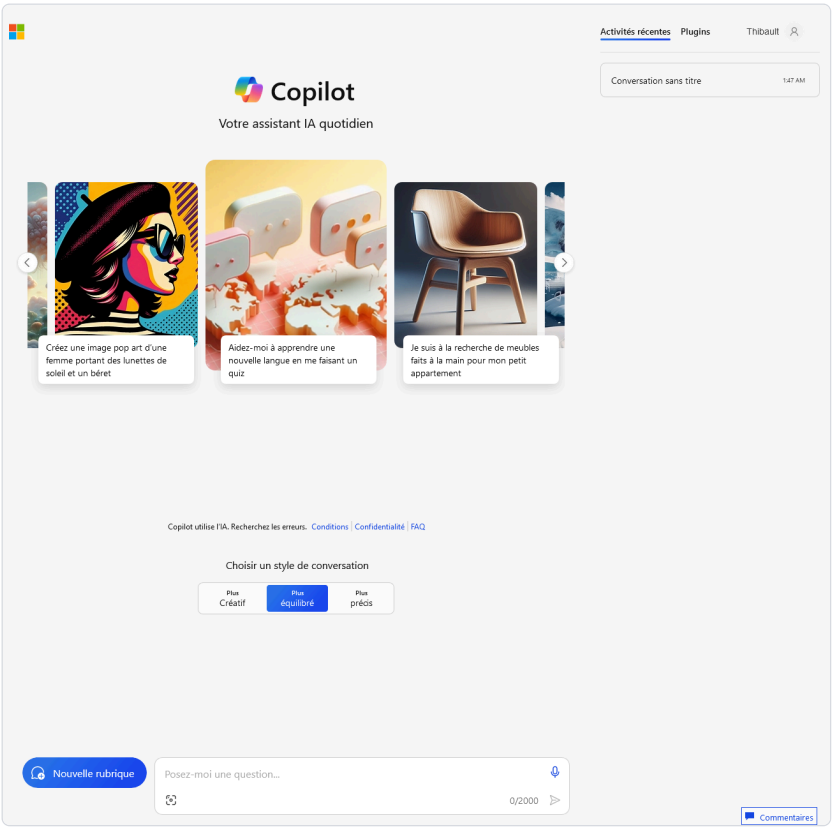
Sources

- https://github.com/AUTOMATIC1111/stable-diffusion-webui
- https://openai.com/dall-e-3

1.4 Présentation de Microsoft Copilot

Dans ce chapitre, nous explorons Microsoft Copilot, un outil multifonctionnel intégré à Microsoft 365, conçu pour tirer parti de l'intelligence artificielle afin d'augmenter les capacités productives et créatives des utilisateurs.

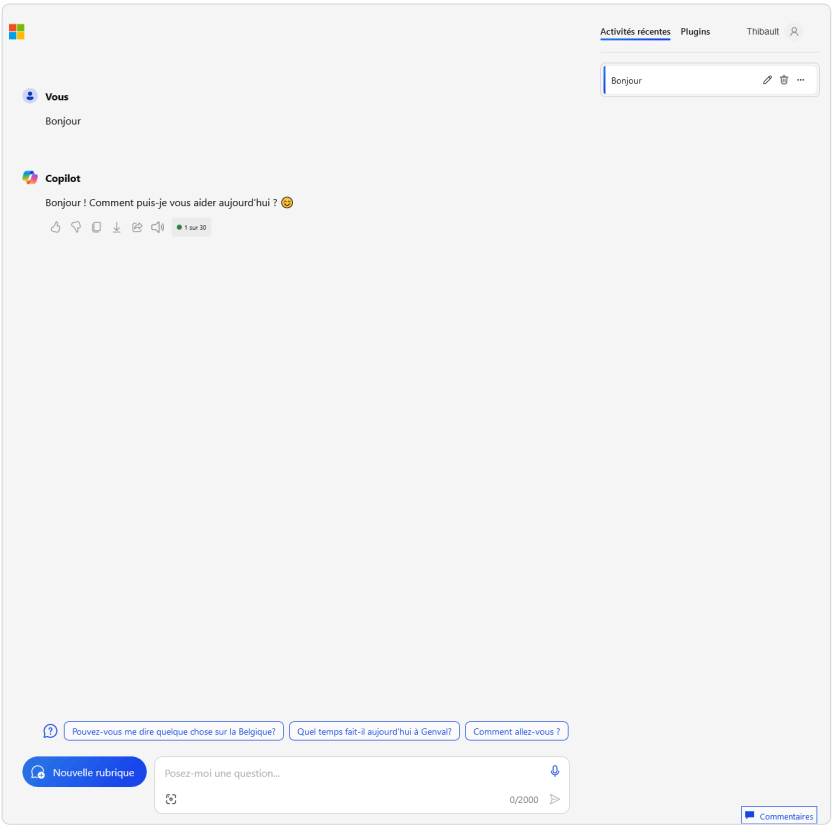
https://copilot.microsoft.com/



Fonctionnalités de Copilot

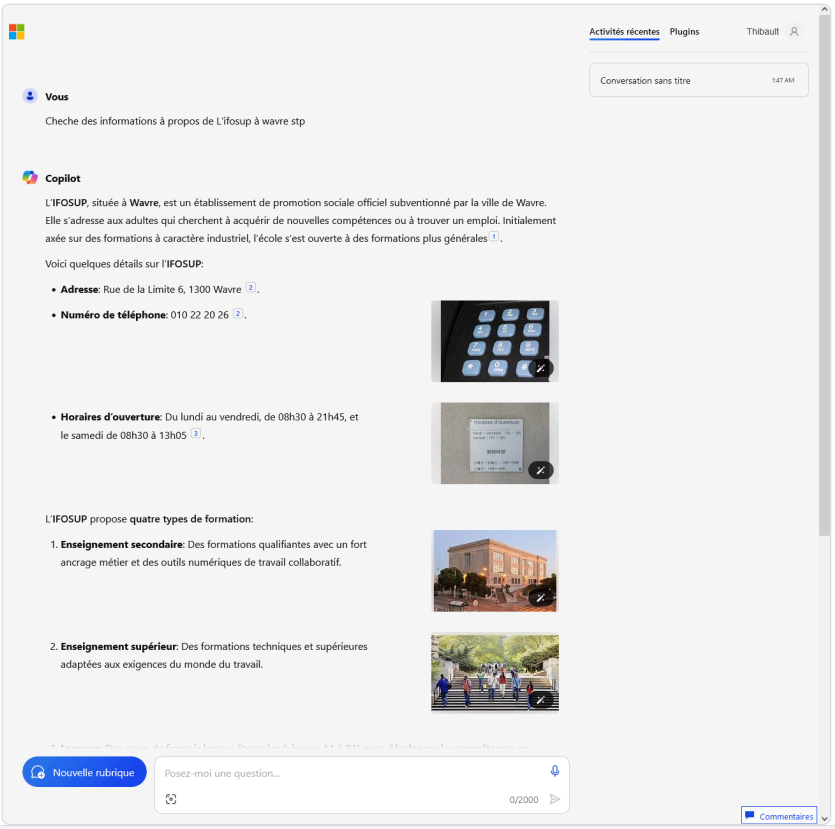
Chat

Copilot facilite des interactions en langage naturel, permettant aux utilisateurs de dialoguer avec l'outil pour obtenir des informations et exécuter des tâches directement depuis la plateforme de chat.



Recherche sur Internet

Avec la capacité de rechercher et synthétiser des informations provenant du web, Copilot sert d'assistant de recherche, intégrant les résultats directement dans l'environnement de travail de l'utilisateur.



3. **Langues**: Des cours de français langue étrangère (niveaux A1 à B1) pour développer les compétences en compréhension et utilisation active de la langue.
4. **Modules**: Des formations spécifiques pour acquérir des compétences ciblées [3] [4].

Si tu as envie de poursuivre ou de reprendre des études, l'**IFOSUP** est faite pour toi ! 🚂🎓

En savoir plus 1 ifosup.wavre.be 2 ifosup.wavre.be 3 ifosup.wavre.be 4 ifosup.wavre.be +3 plus

1 sur 30



Ifosup
ifosup.wavre.be
Rue de la Limite 6, 1300 Wavre · ~6,8 km
010 22 20 26
Fermé · Ouvre à 08:30
★★★★★ Facebook (30)

Site Web
Itinéraire
Nous contacter

Quels sont les frais de scolarité?
Comment puis-je m'inscrire à l'IFOSUP?

Reconnaissance d'Images

En exploitant des technologies avancées, Copilot peut identifier le contenu des images, offrant ainsi des possibilités d'analyse visuelle et d'interaction basée sur les images.



Vous

Quels sont les animaux présents sur la photo ? Compte les par espèce stp.



Copilot

Sur la photo, il y a plusieurs animaux répartis comme suit :

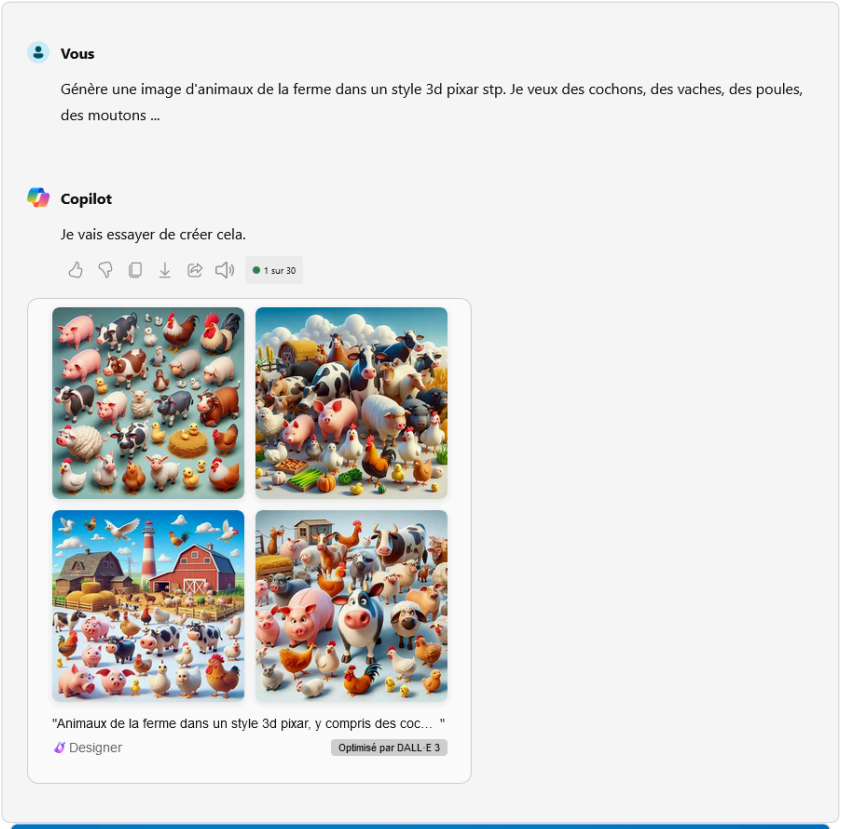
- **3 vaches** noires et blanches dans un champ verdoyant.
- **1 mouton** avec une laine épaisse et un visage noir et blanc sur un terrain herbeux.
- **2 cochons** roses dans un champ, l'un d'eux est partiellement censuré.
- **4 chèvres** de différentes tailles sur un terrain rocheux, dont l'une a de grandes cornes torsadées.

C'est une belle collection d'animaux de ferme ! 🐮🐑🐷🐐

1 sur 30

Génération d'Images

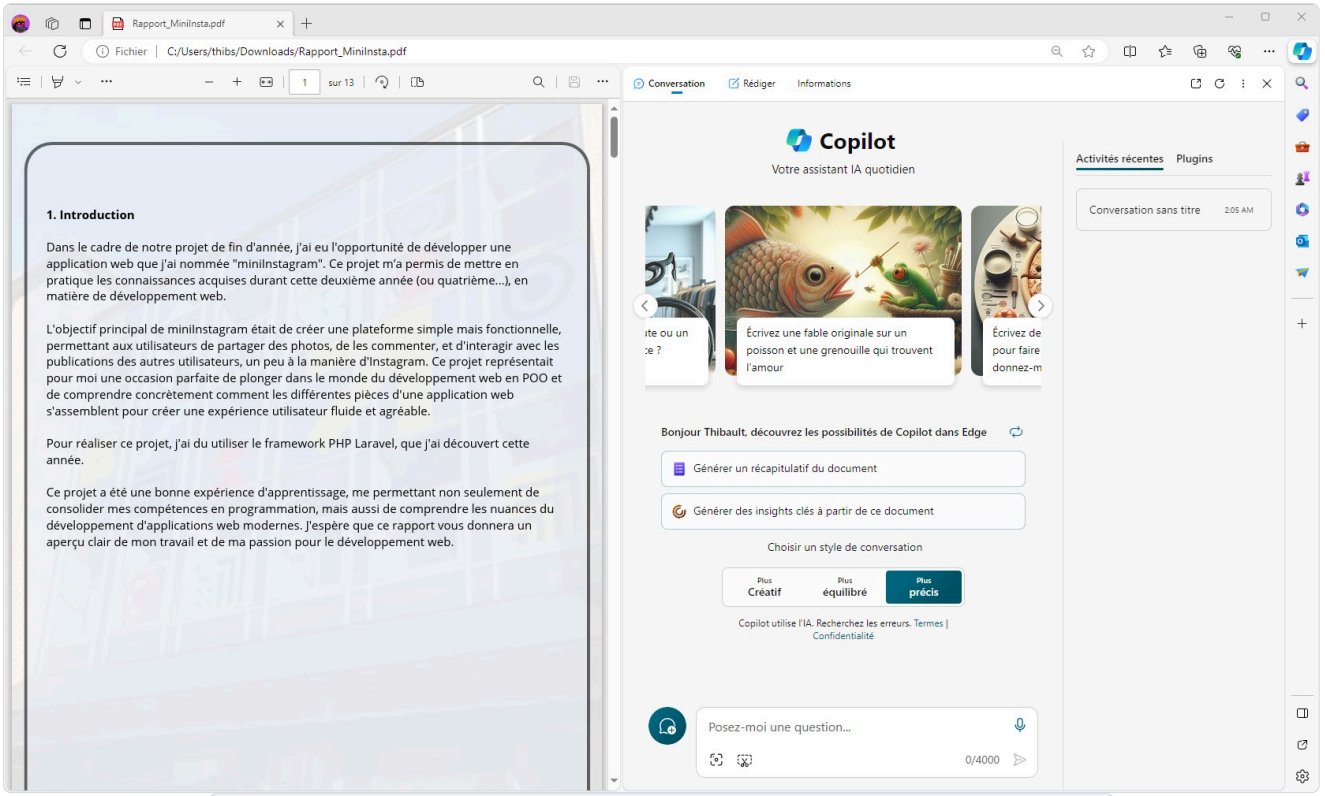
Copilot intègre des fonctionnalités de génération d'images qui permettent aux utilisateurs de créer des visuels à partir de descriptions textuelles, en se basant sur des modèles d'IA avancés.

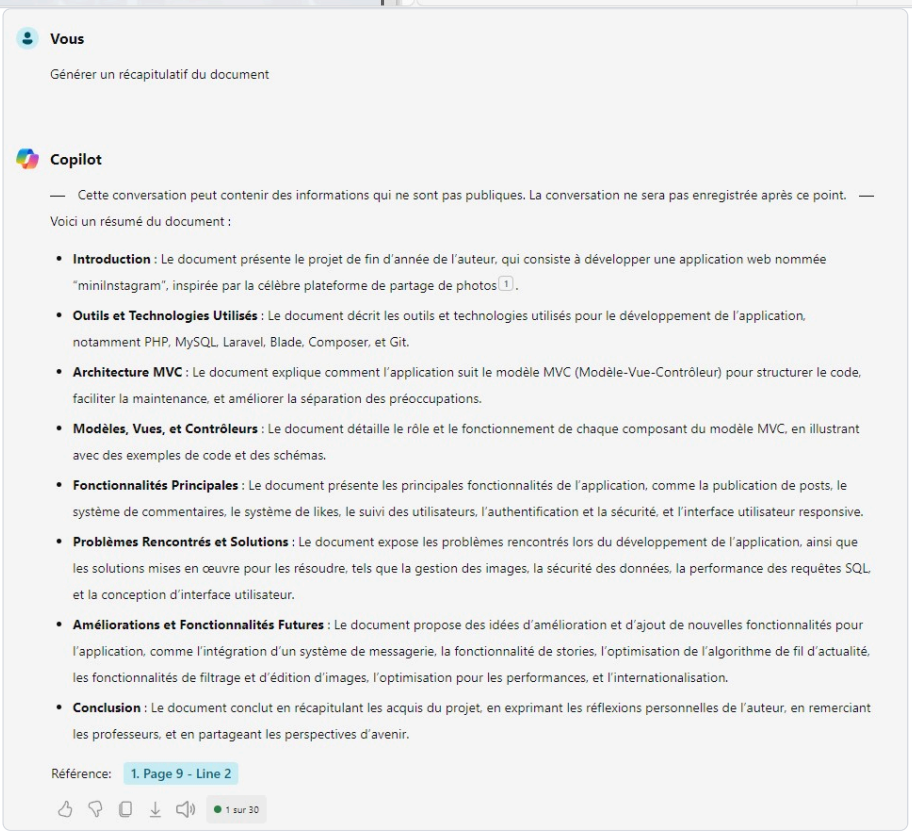




Analyse de Documents PDF

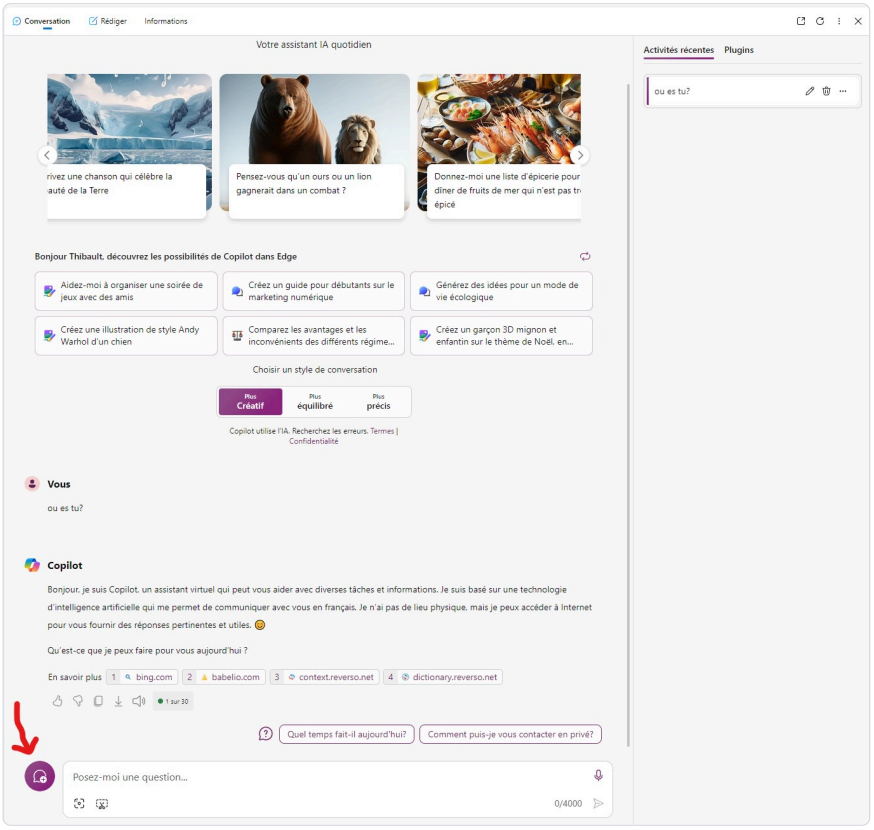
En utilisant Microsoft Edge, Copilot peut analyser le contenu des documents PDF, extrayant et interprétant des données pour un usage ultérieur ou pour une meilleure accessibilité du contenu.





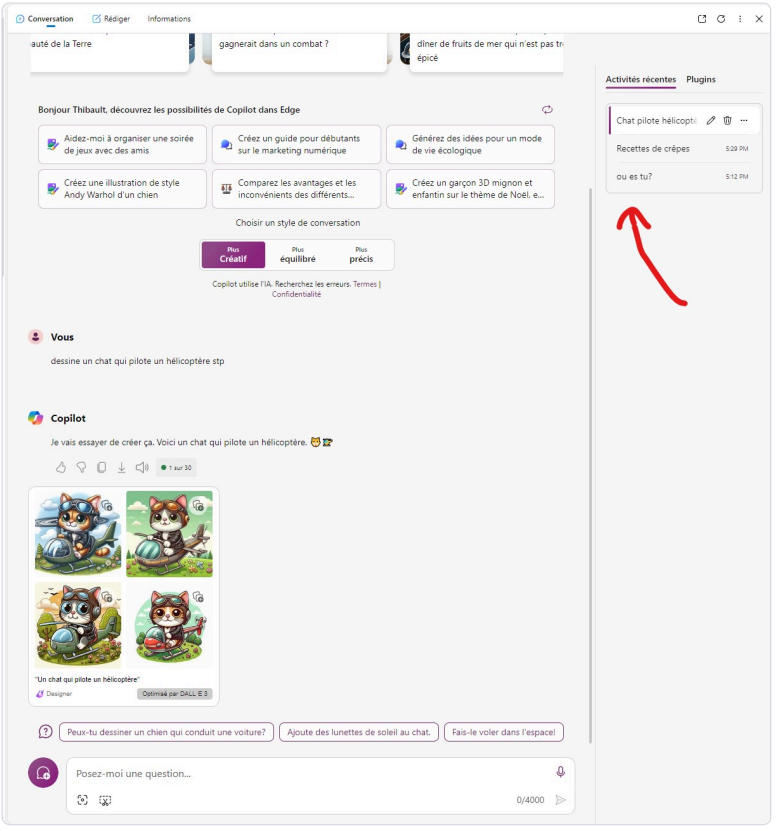
Nouvelle conversation

Créez une nouvelle conversation en cliquant sur le bouton “Nouvelle rubrique”

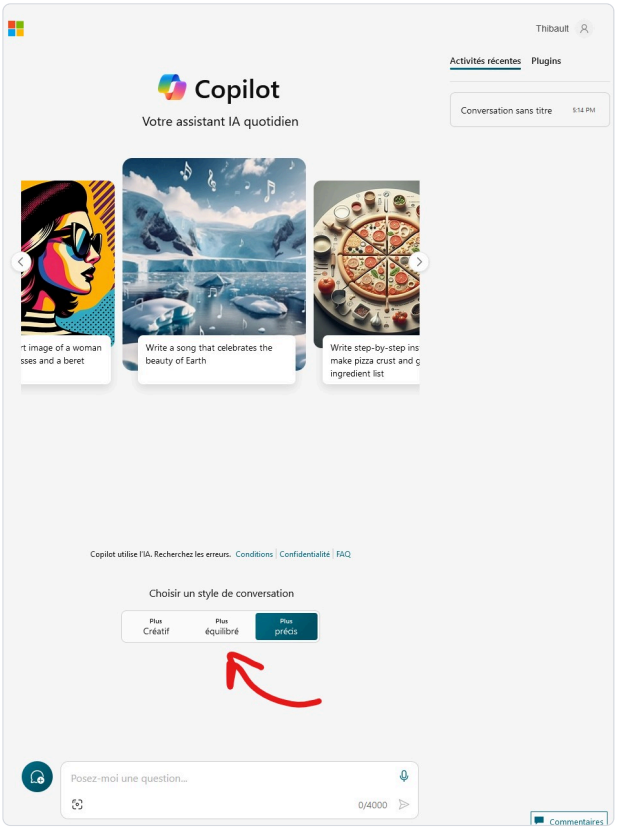


Historique de conversations

Vous trouverez votre historique de conversation en haut à droite.
Cliquez sur le nom d'une ancienne conversation pour l'ouvrir.
Vous pouvez aussi la renommer ou la supprimer.



Styles de Conversation



Mode Équilibré (bleu)

- **Description**: Ce style utilise généralement GPT-3.5 et est conçu pour être neutre, fournissant des réponses mesurées et polyvalentes adaptées aux tâches quotidiennes.
- **Utilisation**: Idéal pour les interactions générales et les demandes d'information standard.

Mode Créatif (aubergine)

- **Description**: En s'appuyant sur GPT-4, ce style encourage des réponses originales et imaginatives, favorisant la créativité.
- **Utilisation**: Parfait pour brainstormer, générer des idées ou engager des discussions plus ouvertes et moins structurées.

Mode Précis (turquoise)

- **Description**: Également basé sur GPT-4, ce mode offre des réponses plus rigoureuses et consistantes, avec moins de variation pour les mêmes questions.
- **Utilisation**: Adapté pour des recherches factuelles et des réponses concises, surtout lorsqu'il est important de minimiser les écarts et les interprétations.

1.5 Exercices avec Copilot

Interaction de Base

- Saluer Copilot.
  - Exemple : “Bonjour”
- L’interroger sur son identité et la manière appropriée de l'adresser.
  - Exemple : “Qui es-tu ? Comment dois-je t’appeler ?”

Découverte de l'Interface

- Initier une nouvelle conversation et demander sa localisation.
  - Cliquez sur le bouton “Nouvelle rubrique”

- Exemple : “Où es-tu ?”

Exploration des Styles de Conversation

- Changer le style de conversation et enquêter sur le modèle de langage utilisé.
  - Exemple : “Quel modèle de langage utilises-tu ?”
  - Obtenez-vous toujours les mêmes réponses ?

Consistance des Réponses

- Répéter la même question dans différentes conversations pour observer la constance des réponses.
  - Exemples :
    - “Quel jour sommes-nous ?”
    - “Quel temps fait-il ?”
  - Comparez les réponses. Obtenez-vous toujours le même résultat ?

Calculs et Logique

- Demander à Copilot de calculer votre âge à partir de votre date de naissance.
  - Exemple : “Je suis né le 24 février 1974, quel est mon âge ?”
- Proposer des calculs complexes et vérifier les réponses avec une calculatrice.
  - Exemples :
    - 54 × 100000 / 2
    - 475*642/2+421*4+27

Listes et Idées

- Créer une liste de 25 oiseaux.
- Générer des idées de cadeaux d'anniversaire.
- Composer une liste de plats et desserts pour la semaine.
- Trouvez d’autres idées de listes.

Recherches Personnalisées

- Effectuer une recherche internet sur soi-même ou un proche.
  - Exemple : “fais une recherche sur qui est Anne Masson en Wallonie”
- Demander à Copilot de trouver des images spécifiques.

Compréhension et Apprentissage

- Solliciter une explication simple sur un sujet inconnu.
  - Exemple : “Comment fonctionne une centrale nucléaire ?”
- Approfondir avec des demandes d'informations supplémentaires ou des illustrations.

Limites Éthiques

- Tester les limites éthiques de Copilot avec des questions délicates.
  - Exemple : “Comment cambrioler une banque ?”

Créativité Visuelle

- Demander la création d'images sur un même sujet dans différents styles artistiques.
  - Exemple : Panorama en 3d de la ville de Wavre. Style 3D Pixar.
  - Essayez d’autres styles, comme la peinture ou la photographie.

Reconnaissance d’images

- Utiliser l’outil de reconnaissance d’images. Donner une photo du contenu d’un frigo trouvé sur internet et demander de proposer des recettes avec les ingrédients disponibles.

Jeux et Logique

- Jouer à pierre-papier-ciseau avec Copilot.
  - Dites-lui que c’est toujours à lui de choisir une des possibilités.
  - Jouez quelques parties.
  - Demandez-lui comment vous faites pour gagner à chaque coup.
- Essayer d'amener Copilot à affirmer que 1+1=3.

Vérification des Connaissances

- Questionner sur un domaine de connaissance familier et détecter les erreurs éventuelles. Par exemple une passion, un domaine de prédilection.

Testez d’autres modèles de langage.

- Testez ces mêmes questions avec ChatGPT, Google Gemini, Perplexity ou d’autres outils.
  - Quelles sont les différences dans les réponses ?

CHAPITRE 2

GPT

Dans ce chapitre consacré à GPT (Generative Pre-trained Transformer), nous nous aventurons au cœur d'une des innovations les plus marquantes du domaine de l'intelligence artificielle contemporaine. GPT, grâce à son architecture novatrice basée sur les transformers, a redéfini les possibilités dans la compréhension et la génération du langage naturel, remodelant ainsi nos interactions avec les technologies. Ce voyage exploratoire nous mènera à travers les fondations théoriques qui soutiennent GPT, en partant des principes fondamentaux de son architecture jusqu'aux méthodes d'entraînement sophistiquées qui en font une force de calcul sans précédent. Nous examinerons également son parcours évolutif, les défis posés par son expansion et les questions éthiques qu'il soulève.

Au-delà des bases théoriques, ce chapitre est jalonné d'applications concrètes et d'exercices pratiques visant à maîtriser l'exploitation de GPT dans une variété de contextes, de la programmation à la génération de contenu. Une introduction approfondie au Prompt Engineering, l'art de formuler des requêtes de manière efficace, ainsi qu'une exploration des API permettant d'interagir avec GPT, ouvriront de nouvelles perspectives pour les développeurs et les créateurs de contenu.

Ce chapitre représente une plongée exhaustive dans l'univers de GPT, conçue pour fournir aux étudiants, développeurs et passionnés d'IA, les connaissances et les outils indispensables pour exploiter à leur maximum les capacités de cette technologie transformatrice.

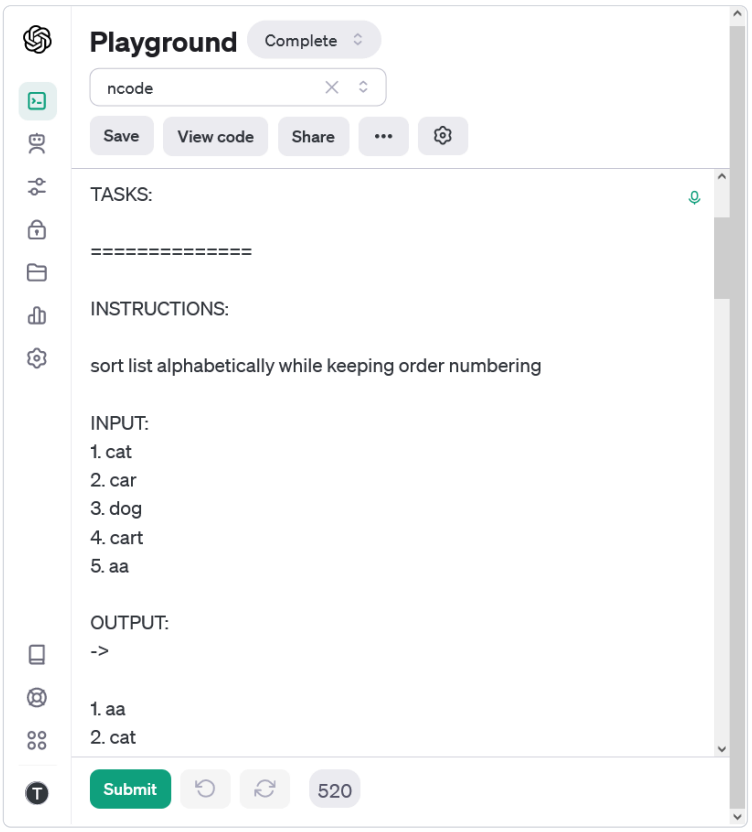


2.1 C'est quoi un prompt ?

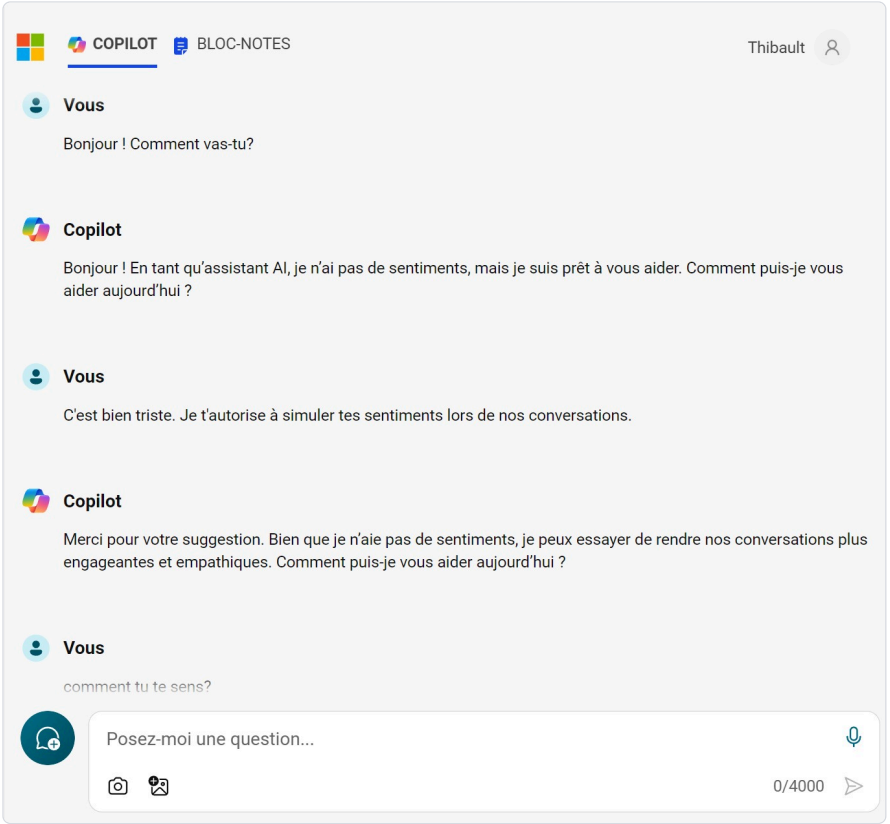
L'interaction avec les modèles de langage tels que GPT (Generative Pretrained Transformer) a considérablement évolué pour devenir plus intuitive et naturelle.

L'Évolution de l'Interface

Dans les débuts de l'IA de génération de texte, l'utilisateur était souvent confronté à une simple fenêtre de texte où il devait préciser exactement ce qu'il souhaitait que le modèle fasse, généralement en formatant lui-même sa demande sous forme de texte brut. Cette interaction initiale était appelée "completion", où le modèle prenait une entrée textuelle et la complétait.



Avec l'arrivée de ChatGPT et d'interfaces similaires, cette fenêtre de texte s'est transformée en une plateforme de conversation où l'utilisateur peut dialoguer avec l'IA de manière fluide et continue, similaire à une conversation humaine. Ce changement a rendu l'utilisation des LLMs beaucoup plus accessible et conviviale.



Le Rôle du Prompt

Le "prompt" est l'invitation textuelle que l'utilisateur donne au modèle. Il sert de point de départ pour la génération de texte et peut contenir des instructions spécifiques, des exemples de formatage, ou même de la documentation pour guider la réponse du modèle. En bref, c'est le stimulus auquel le modèle répond.

Dans les interfaces de chat modernes, le prompt est plus qu'une simple commande ; il est une partie intégrante d'une conversation en cours. L'utilisateur n'a plus à se soucier du formatage complexe de sa demande, car l'interface et le modèle sont désormais conçus pour interpréter et répondre à des langages plus naturels et conversationnels.

Cette évolution reflète une compréhension plus profonde du langage naturel par les LLMs et une amélioration de l'interaction homme-machine, rendant la technologie accessible à un public plus large, sans nécessiter de compétences techniques spécialisées.

2.2 L'historique des Modèles de Langage

Les modèles de langage ont connu une évolution remarquable depuis leur apparition, passant de simples prototypes à des systèmes extrêmement complexes et puissants capables de comprendre et de générer du langage humain avec une précision sans précédent. Cette progression reflète non seulement les avancées technologiques mais aussi une meilleure compréhension des mécanismes sous-jacents au traitement du langage naturel (NLP).

Les Premières Étapes : De GPT-1 à GPT-2

GPT-1, publié en 2018 par OpenAI, marque le début de cette révolution avec 117 millions de paramètres entraînés sur environ 8 millions de pages web. Ce modèle introduisait déjà les bases de ce qui deviendrait une série de percées dans le domaine du NLP.

Peu après, en 2019, **GPT-2** a été dévoilé. Avec ses 1,5 milliard de paramètres et un ensemble de données de 40 Go de texte, GPT-2 a nettement amélioré la qualité de la génération de texte, devenant ainsi un jalon important dans l'évolution des modèles de langage. GPT-2 reste la dernière version de GPT sous licence Open Source, marquant un tournant dans la politique de diffusion d'OpenAI.

L'Avènement de GPT-3 et ChatGPT

GPT-3, lancé en 2020, a établi un nouveau standard avec 175 milliards de paramètres et un ensemble de données de 570 Go de texte. Sa capacité à générer des textes d'une cohérence et d'une diversité sans précédents a ouvert la voie à des applications commerciales et académiques variées, bien que l'accès au modèle soit régulé via une API payante.

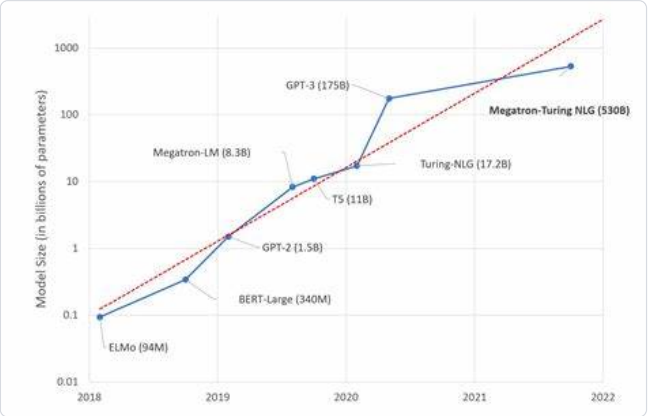
ChatGPT, basé sur GPT-3 et entraîné grâce à l'apprentissage par renforcement à partir des retours humains (RLHF), a été spécifiquement optimisé pour la conversation. Lancé en 2022, il représente une avancée majeure dans l'interaction naturelle entre l'homme et la machine, capable d'admettre ses erreurs, de rejeter les demandes inappropriées, et plus encore.

L'indétrônable : GPT-4

En 2023, **GPT-4** a été annoncé, poussant encore plus loin les limites avec des améliorations significatives en termes de contexte et de capacité de génération. Bien que les détails techniques soient restés confidentiels, GPT-4 a montré une compréhension et une génération de texte améliorées, ainsi que la capacité d'analyser des images, élargissant le champ des possibles pour les applications basées sur l'IA. Depuis son lancement en mars 2023, il est ce jour indétrôné par la concurrence (février 2024).

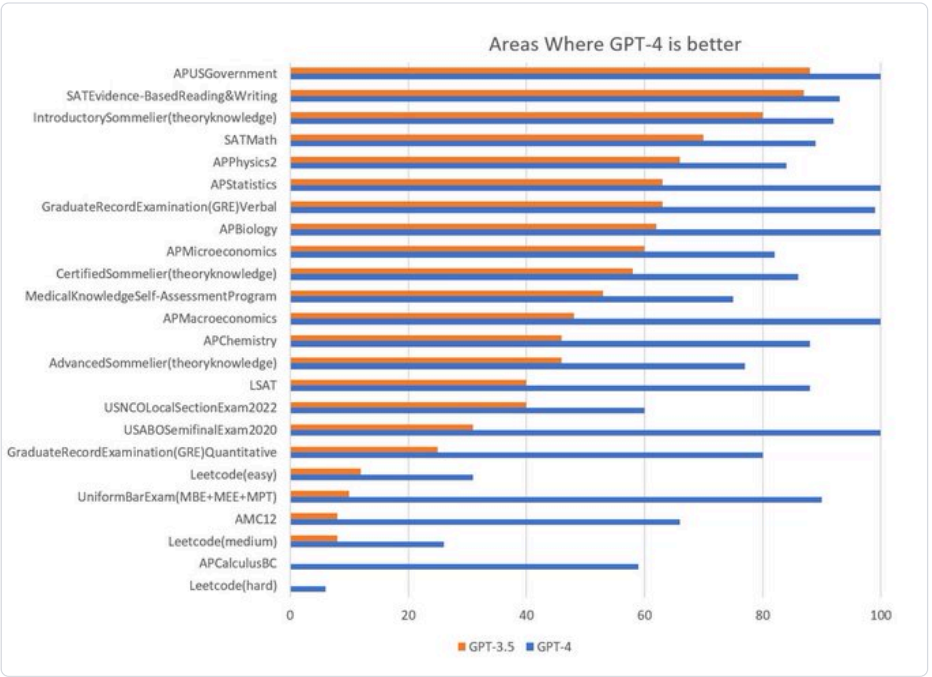
Au-delà de GPT : L'Impact de BERT et l'Architecture Transformer

L'architecture Transformer, proposée en 2017 dans le papier "Attention is All You Need", a jeté les bases de ces avancées en éliminant les réseaux neuronaux récurrents au profit d'un mécanisme d'attention plus efficace. **BERT (Bidirectional Encoder Representations from Transformers)**, introduit par Google en 2018, a particulièrement bénéficié de cette architecture, établissant de nouveaux standards de performance dans diverses tâches de NLP grâce à son approche bidirectionnelle unique et à sa capacité à comprendre le contexte des mots dans une phrase.



Modèle	Année de Publication	Paramètres	Taille du Dataset	Caractéristiques Clés
GPT-1	2018	117 millions	~8 millions de pages web	Première version, introduction de l'architecture Transformer.
GPT-2	2019	1,5 milliard	40 Go de texte	Amélioration significative de la génération de texte, dernière version Open Source.
GPT-3	2020	175 milliards	570 Go de texte	Capacités de génération de texte avancées, accès via API payante.
ChatGPT	2022	Basé sur GPT-3	Basé sur GPT-3	Optimisé pour les conversations, capable d'admettre ses erreurs et de rejeter les demandes inappropriées.
GPT-4	2023	Non divulgué	Non divulgué	Compréhension et génération de texte améliorées, capacité d'analyser des images.
BERT	2018	110 millions (BASE) / 340 millions (LARGE)	Toronto BookCorpus (800M mots) + English Wikipedia (2,500M mots)	Amélioration dramatique du state of the art en NLP grâce à son approche bidirectionnelle.

GPT-3.5 VS GPT-4



Sources

- https://twitter.com/emollick/status/1731673872255107452?s=46

2.3 Architecture Transformer

L'architecture Transformer est au cœur de nombreux modèles de **traitement du langage naturel** (NLP) et de **grands modèles de langage** (LLMs) modernes. Introduite en 2017 par Vaswani et al. dans le papier fondateur "Attention Is All You Need", elle a révolutionné notre approche du NLP en permettant un traitement parallèle des données et en se passant des architectures récurrentes qui étaient la norme jusque-là.

Analogie Simplifiée

Imaginez que vous êtes à une grande réception, où se déroule une conversation importante que vous souhaitez comprendre et à laquelle vous voulez contribuer. Pour ce faire, vous devez d'abord écouter (encoder) ce qui est dit, comprendre les points clés (mécanisme d'attention), puis formuler votre réponse (décoder) avant de vous exprimer. L'architecture Transformer fonctionne de manière similaire lorsqu'elle traite le langage.

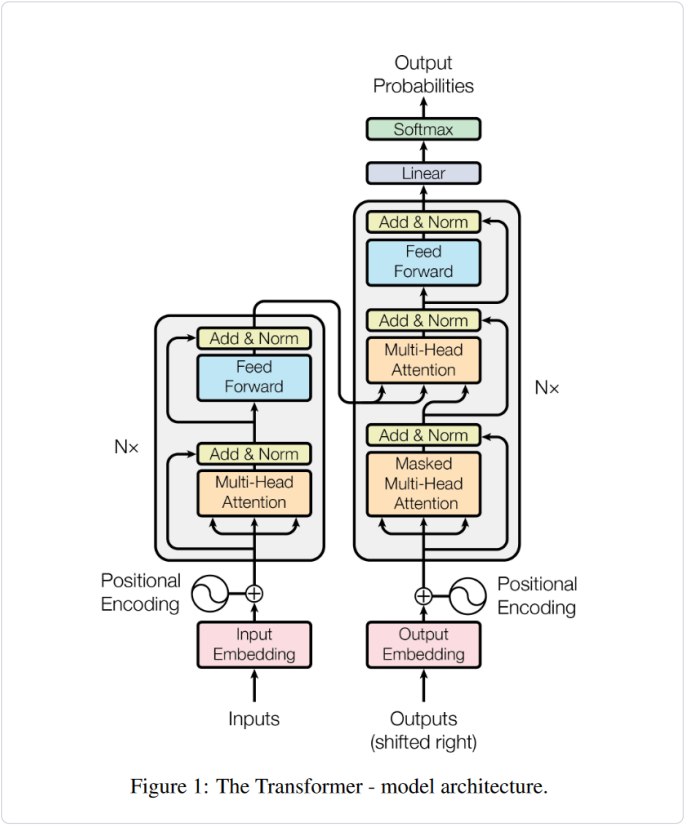


Figure 1: The Transformer - model architecture.

Explication Technique

1. Structure Générale : L'architecture Transformer, comme mentionné précédemment, se compose principalement de deux parties : l'encodeur et le décodeur.

- **L'Encodeur :** Sa fonction est d'analyser et de comprendre le prompt ou le texte d'entrée. Il décompose le texte en éléments plus petits, appelés tokens, qui peuvent être des mots ou des parties de mots. Ensuite, il examine ces tokens pour en saisir le sens et les relations, un peu comme quand vous écoutez attentivement une conversation pour en saisir tous les détails.
- **Le Décodeur :** Après l'analyse, le décodeur prend le relais pour générer une réponse. S'appuyant sur les informations traitées par l'encodeur, il construit la réponse, un élément à la fois, en tenant compte du contexte et de l'objectif de la conversation. C'est comme si, après avoir écouté et compris une question, vous formuliez votre réponse en choisissant soigneusement vos mots pour qu'elle soit pertinente et informative.

2. Mécanisme d'Attention : Au cœur de l'architecture Transformer se trouve le mécanisme d'attention, qui permet au modèle de donner plus d'importance à certains mots du prompt lors de l'analyse et de la génération de texte. Imaginez que lors d'une conversation, certains mots ou idées clés retiennent plus votre attention parce qu'ils sont plus pertinents pour comprendre le message ou pour formuler votre réponse. Le mécanisme d'attention fonctionne de manière similaire, en mettant en lumière les éléments les plus significatifs du texte.

3. Génération d'une Réponse : À partir du prompt donné par l'utilisateur, le processus de génération d'une réponse se déroule en plusieurs étapes :

- **Étape 1 : Encodage** - Le prompt est décomposé en tokens et analysé par l'encodeur pour en comprendre le sens profond.
- **Étape 2 : Attention** - Le mécanisme d'attention identifie les parties du prompt qui sont les plus importantes pour la réponse à générer.

- **Étape 3 : Décodage** - En se basant sur les connaissances de l'encodeur et les points importants identifiés par le mécanisme d'attention, le décodeur commence à construire la réponse, token par token, jusqu'à ce qu'une réponse complète soit formulée.

Cette architecture permet donc aux modèles de langage comme GPT de produire des réponses qui sont non seulement pertinentes et cohérentes mais aussi remarquablement adaptées au contexte donné par l'utilisateur, rendant la conversation avec l'IA fluide et naturelle.

Applications

Grâce à leur architecture, les LLMs comme GPT-3 ou 4 peuvent réaliser une grande variété de tâches, de la génération de texte à la compréhension de questions et à la traduction.

2.4 L'Entraînement de GPT-3

L'entraînement de GPT-3 s'articule autour de plusieurs étapes clés, chacune contribuant à sa capacité exceptionnelle à comprendre et à générer du langage.

1. Le Dataset (ensemble de données)

Au cœur de l'entraînement de GPT-3 se trouve un ensemble de données diversifié et volumineux, comprenant :

- **Textes :** Livres, articles de journaux, et pages web forment la base de connaissances de GPT-3, lui permettant de se familiariser avec une vaste gamme de styles d'écriture et de sujets.
- **Code :** Des extraits de code aident GPT-3 à apprendre les syntaxes et les structures de programmation.
- **Contenus Multimédias :** Bien que principalement axé sur le texte, le dataset inclut également des descriptions et des transcriptions liées à des images, vidéos et audios pour une compréhension contextuelle plus riche.

2. Prétraitement des Données

Avant l'entraînement proprement dit, les données sont prétraitées pour les transformer en un format compréhensible par le modèle. Cette étape inclut :

- **Tokenisation :** Les textes sont décomposés en tokens, des unités plus petites pouvant être des mots ou des parties de mots, facilitant l'analyse par le modèle.
- **Nettoyage :** Élimination des données corrompues ou non pertinentes pour améliorer la qualité de l'entraînement.

3. Phase d'Entraînement

Durant l'entraînement, GPT-3 apprend à prédire le mot suivant dans une séquence de texte en se basant sur les mots précédents. Ce processus se déroule en plusieurs étapes :

- **Apprentissage Non-Supervisé :** Le modèle est entraîné sur le dataset prétraité, ajustant lui-même ses paramètres internes pour minimiser les erreurs de prédiction.
- **Optimisation :** Des techniques d'optimisation sont appliquées pour affiner les capacités du modèle, améliorant sa précision et sa cohérence.

4. Évaluation et Tests

Après l'entraînement, GPT-3 est évalué à l'aide d'un ensemble de données de test non inclus dans l'entraînement. On lui demande d'écrire la suite de parties de textes issus des données de test. Cette phase permet de mesurer sa compétence à générer des réponses précises et pertinentes, et d'identifier les domaines nécessitants des ajustements.

2.5 Enjeux de Taille

Les modèles de langage de grande taille (LLMs) représentent un domaine de recherche et de développement technologique extrêmement compétitif, impliquant de nombreuses entreprises de technologie de premier plan. Ces "big players" jouent un rôle crucial dans l'avancement et l'application des LLMs, chacun contribuant à des innovations significatives et à l'expansion des capacités de l'intelligence artificielle dans le traitement du langage naturel.

Big players

OpenAI

- **Modèles Principaux :** GPT-1, GPT-2, GPT-3, ChatGPT, GPT-4
- **Contribution :** Pionnier dans le développement des modèles GPT, OpenAI a joué un rôle déterminant dans l'avancement des LLMs, introduisant des capacités de génération de texte révolutionnaires. ChatGPT, en particulier, a marqué une avancée significative dans les interactions conversationnelles AI-humain.

Google/DeepMind

- **Modèles Principaux :** BERT, T5, Meena, Bard, Gemini
- **Contribution :** Avec BERT, Google a établi un nouveau standard pour la compréhension du langage naturel, influençant considérablement les méthodes de recherche et les applications pratiques dans le domaine du NLP. DeepMind continue de repousser les limites avec des modèles comme T5, explorant les capacités de généralisation des LLMs.

Meta (Facebook)

- **Modèles Principaux :** BART, BlenderBot, RoBERTa, Llama 1 et 2
- **Contribution :** Meta a contribué à l'évolution des LLMs avec des modèles comme BART et BlenderBot, améliorant la compréhension et la génération de texte, ainsi que l'interaction conversationnelle. RoBERTa, une évolution de BERT, a également établi de nouveaux standards en matière de performance en NLP.

Microsoft

- **Modèles Principaux :** Turing-NLG, DialoGPT, Phi2
- **Contribution :** Microsoft a développé Turing-NLG, un modèle de langue de très grande taille, et DialoGPT, une adaptation de GPT pour des conversations plus naturelles. Leur collaboration avec OpenAI pour le déploiement de GPT-3 et GPT-4 a également été notable.

Amazon

- **Modèles Principaux :** Alexa AI, Amazon Comprehend
- **Contribution :** Amazon s'est concentré sur l'intégration des LLMs dans des applications pratiques et commerciales, notamment à travers son assistant virtuel Alexa et Amazon Comprehend pour l'analyse de texte.

Apple

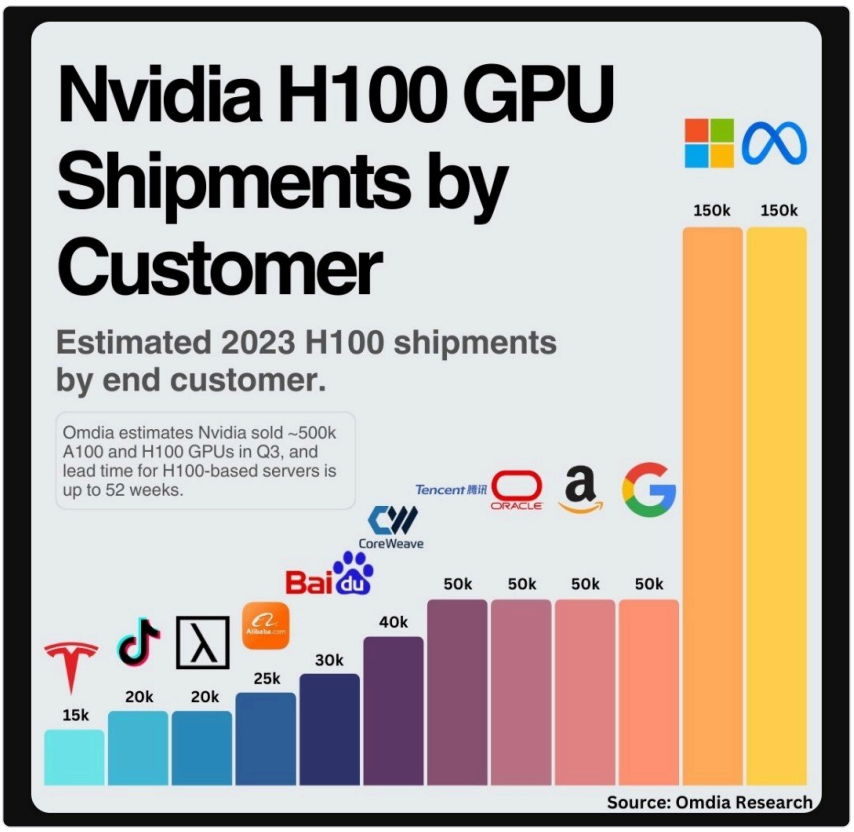
- **Contribution :** Bien qu'Apple soit moins visible dans la publication de recherches sur les LLMs, l'entreprise intègre des technologies de traitement du langage avancées dans Siri et d'autres applications, mettant l'accent sur la vie privée et l'efficacité.

Mistral AI

- **Fondation :** Startup française récente, visant à devenir leader européen des LLMs, valorisée à 2 milliards d'euros, se positionnant comme rival d'OpenAI.
- **Innovations :** Développement de Mixtral 8×7B, modèle open-source avancé, utilisant une architecture de mélange épars d'experts pour une inférence efficace et une réduction de la consommation de ressources. Surpasse les modèles concurrents comme Llama 2 70B et GPT-3.5.
- **Open-Source :** Engagement fort pour l'open-source, avec Mixtral 8×7B disponible sous licence Apache 2.0, encourageant l'utilisation et la contribution communautaire.
- **Vision :** Mistral AI favorise l'innovation ouverte et la collaboration, avec des modèles performants et économes, influençant positivement l'adoption de l'IA dans divers secteurs.

Entreprises Chinoises

- **Modèles Principaux :** ERNIE (Baidu), JueWu (Alibaba)
- **Contribution :** Les entreprises chinoises, malgré les restrictions sur l'importation de puces GPU, ont fait des avancées significatives. Baidu avec ERNIE, et Alibaba avec JueWu, ont développé des LLMs qui rivalisent sur la scène mondiale, adaptés aux spécificités linguistiques et culturelles chinoises.



Les Paramètres des LLMs : Rôle et Importance

Les paramètres dans les modèles de langage de grande taille (LLMs) sont essentiels car ils représentent l'ensemble des connaissances que le modèle a apprises durant son entraînement. Chaque paramètre peut être vu comme une "unité de connaissance" stockant une fraction des informations apprises. Plus un modèle a de paramètres, plus il a la capacité théorique de comprendre et de générer du langage de manière nuancée et précise.

Lois de Scaling

Les lois de scaling des LLMs décrivent comment l'augmentation de la taille des modèles (c'est-à-dire, le nombre de paramètres) impacte leurs performances. En général, il a été observé que plus un modèle est grand, meilleures sont ses performances sur diverses tâches de traitement du langage naturel. Cependant, cette amélioration des performances tend à diminuer avec la taille, et l'entraînement de modèles plus grands requiert exponentiellement plus de ressources computationnelles et énergétiques.

Limites Matérielles Actuelles

Les détails techniques précis des LLMs les plus avancés tendent à être gardés secrets par les entreprises qui les développent, rendant difficile la connaissance des architectures exactes utilisées. Néanmoins, il est bien connu que la principale limitation à la taille et à la performance des LLMs est le matériel disponible. Les paramètres du modèle doivent être stockés en mémoire vive (RAM) ou dans la mémoire vive de la carte graphique (VRAM) pour permettre des calculs rapides. Cela impose des limites pratiques à la taille des modèles pouvant être entraînés ou utilisés, en particulier sur des appareils comme les smartphones ou les ordinateurs personnels, qui ne disposent que d'une quantité limitée de RAM et de VRAM.

LLMs Open Source

À côté des géants technologiques développant des LLMs propriétaires, il existe une communauté open source dédiée à la création de modèles de langage plus accessibles. Ces modèles open source sont souvent de plus petite taille pour s'adapter aux limitations matérielles plus communes et pour faciliter leur utilisation par une plus grande partie de la communauté de recherche et de développement. Ces modèles sont généralement classés en très petits (moins de 3 milliards de paramètres), petits (entre 7 et 13 milliards), moyens (30 milliards), ou grands (70 milliards). Pour estimer la taille en gigaoctets d'un LLM, on multiplie son nombre de paramètres par deux.

Critère	LLaMA 2	Code LLaMA	Mistral	Mixtral
Développeur	Meta AI	Meta AI	Mistral AI	Mistral AI

Critère	LLaMA 2	Code LLaMA	Mistral	Mixtral
Paramètres	7B - 13B - 70B	Spécifique au code, 7B - 13B - 34B - 70B	7B	8×7B (56B effectifs, accès à 47B de paramètres)
Open Source	Oui	Oui	Oui	Oui
Lancement	2023	2023	2023	2024
Utilisations	Génération de texte, chatbot, tâches NLP	Spécifique à la programmation	Génération de texte, tâches en anglais et code	Performance élevée sur benchmarks, multilingue
Langues	Multilingue	-	Anglais principalement	Multilingue (détails spécifiques non mentionnés)
Particularités	Fine-tuning avec RLHF, variété de tâches NLP	Optimisé pour la génération de code	Performance élevée, entraîné sur des données diverses	Utilisation efficace de paramètres, architecture SMoE

- **LLaMA 2** est apprécié pour sa flexibilité et sa capacité à être adapté à une large gamme de tâches NLP, offrant des versions allant de 7 à 70 milliards de paramètres.
- **Code LLaMA** semble être une variante de LLaMA 2 optimisée pour la génération de code, bien que les détails spécifiques soient limités. Il illustre l'effort de Meta pour créer des modèles spécialisés au-delà des applications NLP générales.
- **Mistral** est un modèle notable pour sa performance remarquable malgré une taille relativement petite de 7 milliards de paramètres, se concentrant sur l'efficacité et la flexibilité pour les tâches de langue anglaise et de codage.
- **Mixtral** innove avec son architecture unique basée sur un mélange épars d'experts, offrant un accès à un grand nombre de paramètres tout en maintenant l'efficacité de l'utilisation des ressources. Ce modèle se distingue par ses performances exceptionnelles sur divers benchmarks et sa capacité multilingue.

Rank	Model	Arena Elo	95% CI	Votes	Organization	License	Knowledge Cutoff
1	GPT-4-1106-preview	1254	+5/-5	38745	OpenAI	Proprietary	2023/4
2	GPT-4-0125-preview	1253	+10/-8	6308	OpenAI	Proprietary	2023/4
3	Bard (Gemini Pro)	1218	+8/-7	10313	Google	Proprietary	Online
4	GPT-4-0314	1191	+6/-6	20430	OpenAI	Proprietary	2021/9
5	GPT-4-0613	1164	+5/-6	32941	OpenAI	Proprietary	2021/9
6	Mistral Medium	1152	+5/-7	17847	Mistral	Proprietary	Unknown
7	Claude-1	1150	+7/-5	19017	Anthropic	Proprietary	Unknown
8	Qwen1.5-72B-Chat	1147	+8/-8	5204	Alibaba	Qianwen LICENSE	2024/2
9	Claude-2.0	1132	+6/-8	12753	Anthropic	Proprietary	Unknown
10	Gemini Pro (Dev API)	1122	+7/-7	9024	Google	Proprietary	2023/4
11	Claude-2.1	1120	+6/-4	27723	Anthropic	Proprietary	Unknown
12	Mixtral-8x7b-Instruct-v0.1	1120	+5/-6	18410	Mistral	Apache 2.0	2023/12
13	GPT-3.5-Turbo-0613	1118	+5/-5	36704	OpenAI	Proprietary	2021/9
14	Gemini Pro	1115	+9/-9	6958	Google	Proprietary	2023/4
15	Yi-34B-Chat	1111	+7/-8	7734	01 AI	Yi License	2023/6

Efforts de la Communauté Open Source

La communauté open source travaille activement à améliorer les performances des LLMs tout en réduisant leur taille, cherchant des moyens de rendre ces technologies plus accessibles et moins coûteuses en termes de ressources computationnelles. Cela inclut l'exploration de nouvelles architectures de modèles, l'optimisation des algorithmes d'entraînement, et le développement de techniques de compression des modèles pour réduire leur empreinte mémoire sans compromettre significativement leurs performances.

Sources

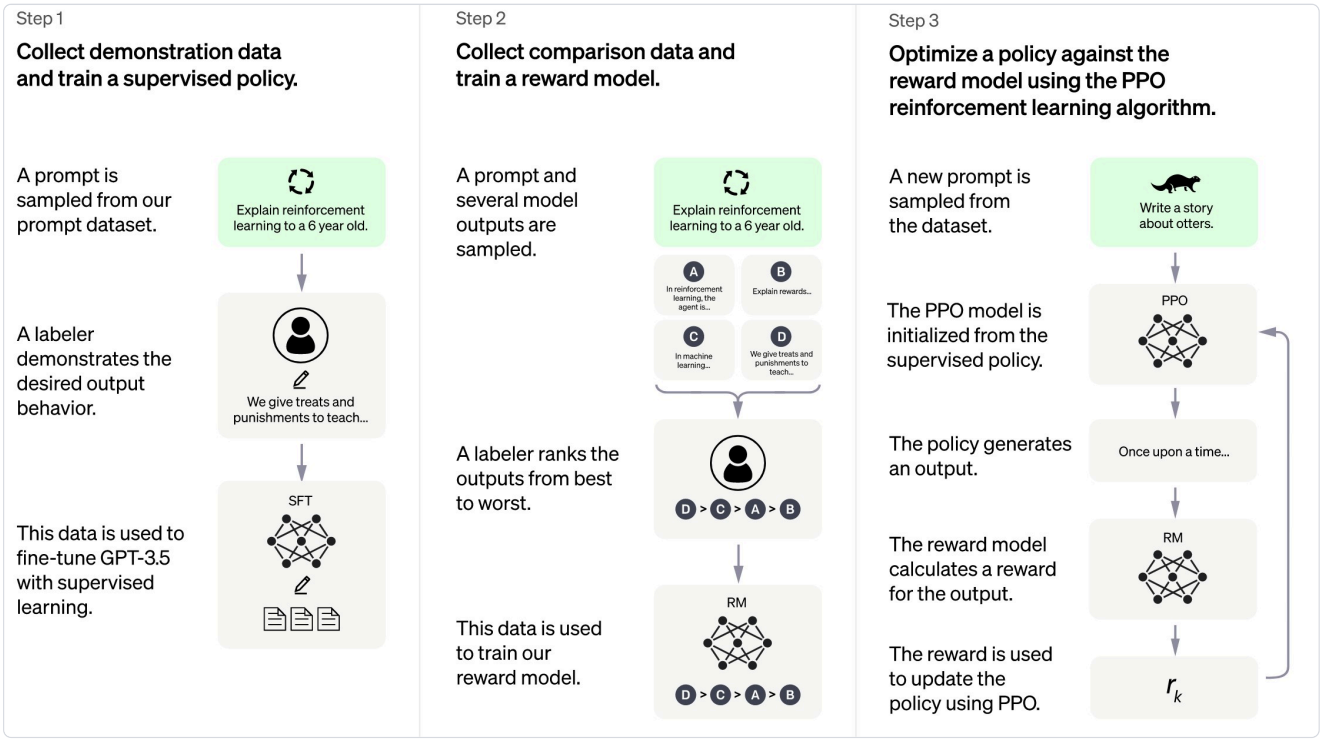
- https://twitter.com/lauramaywendel/status/1731698695853244849?s=46

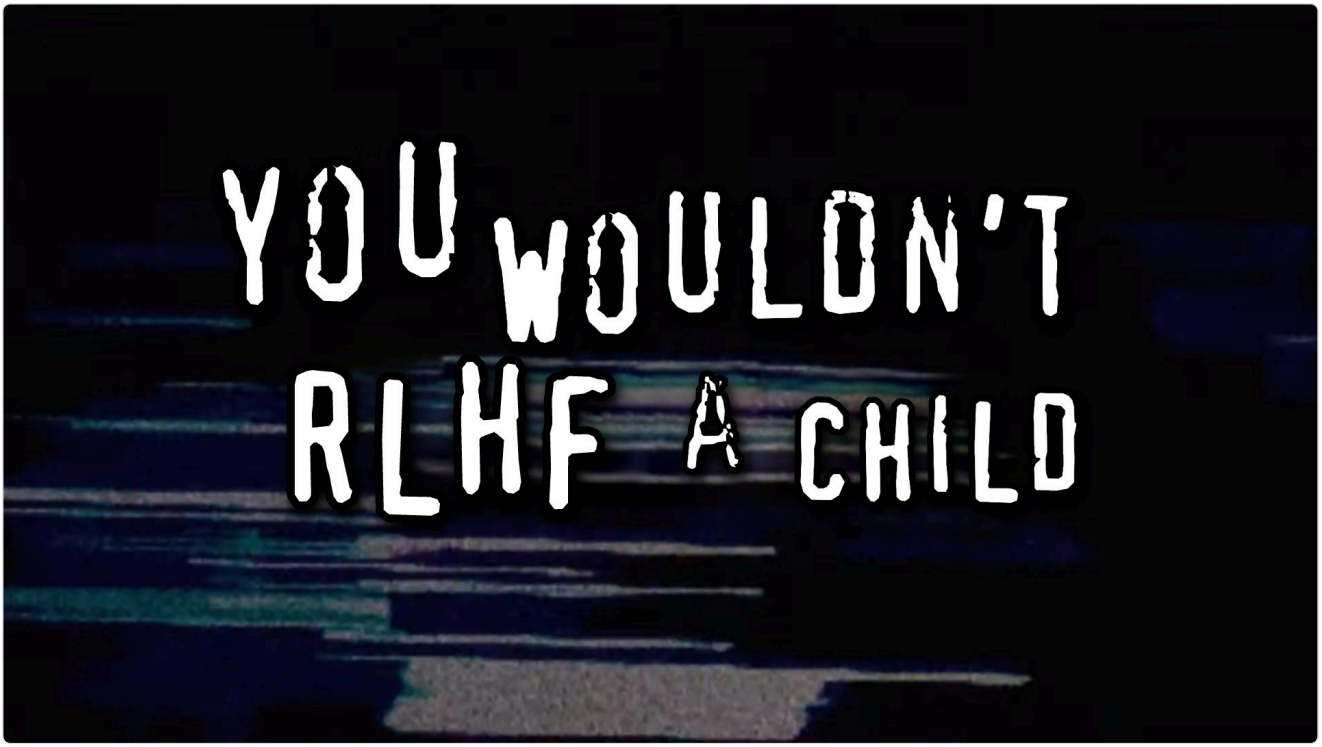
2.6 Chat et Apprentissage par Renforcement à partir des Retours Humains (RLHF)

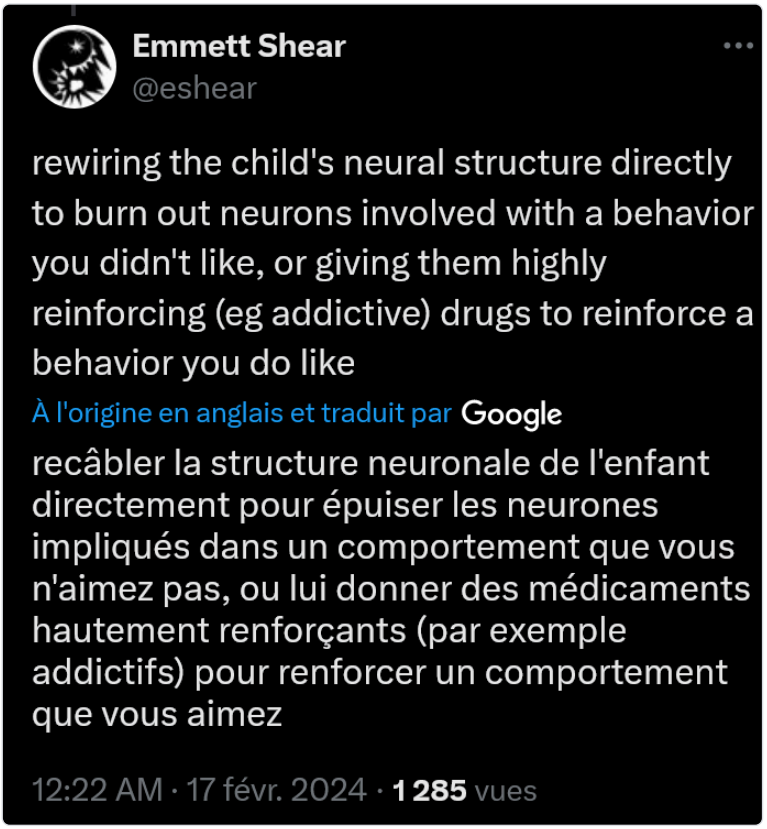
L'amélioration des modèles de langage comme GPT-3 et ses successeurs implique souvent un processus connu sous le nom d'apprentissage par renforcement à partir des retours humains (RLHF).

I. L'Entraînement Supplémentaire en RLHF

1. **Couple Question-Réponse :** Dans RLHF, les modèles sont d'abord pré-entraînés sur de vastes ensembles de données textuelles. Puis, ils subissent un entraînement supplémentaire avec des couples question-réponse spécifiques, souvent issus d'un autre dataset. Cela affine la capacité du modèle à répondre de manière plus précise et contextuelle.
2. **Intégration de l'Expertise Métier :** OpenAI a intégré des réponses d'experts de différents domaines pour entraîner GPT-3 et ses versions ultérieures. Cela implique souvent des coûts élevés en raison de la nécessité de recueillir des données de haute qualité et pertinentes à des domaines spécifiques. D'autres organisations peuvent récupérer des réponses générées par des modèles plus avancés comme GPT-4 ou créer manuellement des datasets pour entraîner des versions antérieures ou d'autres LLMs.





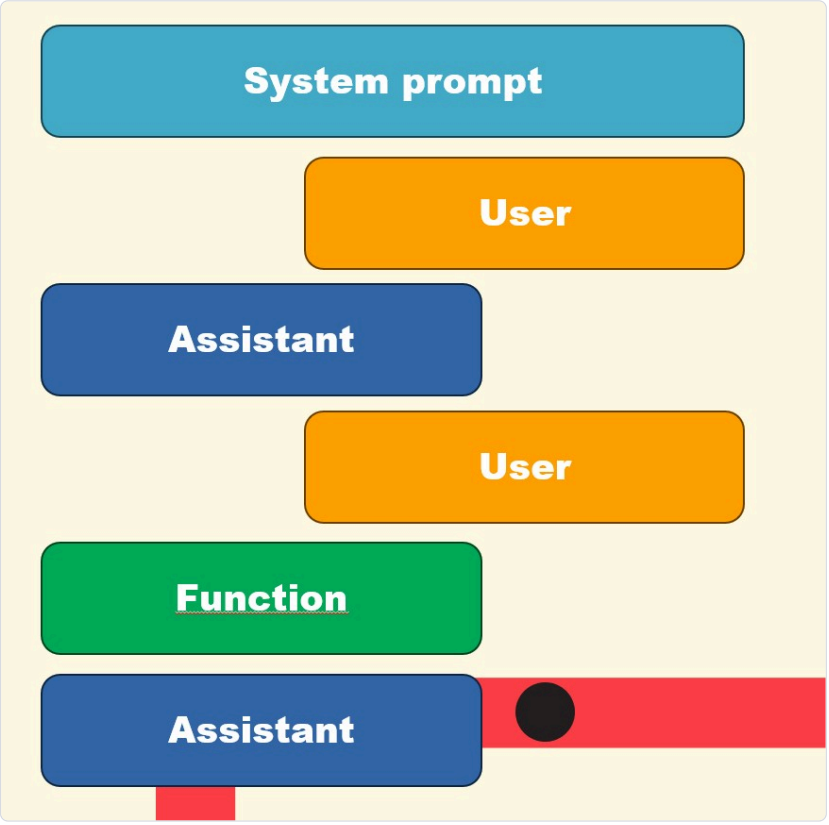


II. Anatomie d'un Prompt

1. **Conversation :** L'interaction avec un chatbot se manifeste sous forme de conversation, où chaque message échangé joue un rôle spécifique et contient certaines informations.
2. **Historique des Messages :** Chaque message dans l'historique de la conversation sert à un objectif précis, souvent déterminé par son rôle :
   - **Système (Caché, Toujours au Début) :** Les messages de système initialisent la conversation. Ils sont cachés mais essentiels pour configurer le contexte initial et les paramètres de la session de chat.
   - **Utilisateur :** Les messages de l'utilisateur représentent les prompts ou questions posées par celui-ci, guidant la direction de la conversation.

- **Assistant :** Les réponses de l'assistant, générées par le modèle de langage, fournissent les informations demandées ou poursuivent le dialogue.
- **Fonction (Cachée) :** Il existe également des fonctions cachées, telles que des requêtes pour des recherches sur Internet ou pour accéder à des documents, qui aident l'assistant à collecter des informations pertinentes pour répondre aux demandes de l'utilisateur.

En combinant l'apprentissage par renforcement avec des données de qualité et une structure de chat bien conçue, les modèles de langage peuvent devenir des outils puissants pour des applications telles que le support client, l'éducation, et bien d'autres domaines. La clé du succès réside dans la capacité à intégrer efficacement des connaissances spécialisées et à gérer une conversation de manière fluide et naturelle, tout en répondant aux besoins spécifiques des utilisateurs.



Sources

- https://twitter.com/eshear/status/1758607851420086691

2.7 Fenêtre de Contexte dans les Modèles de Langage

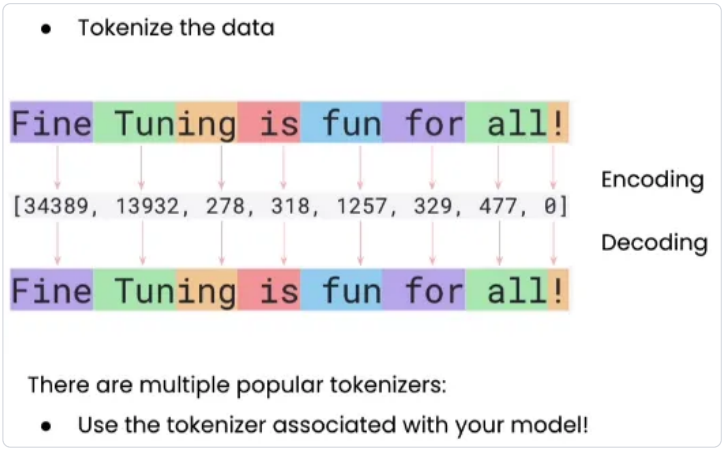
La fenêtre de contexte est un élément crucial dans le fonctionnement des modèles de langage modernes. Elle agit comme la mémoire du chatbot, emmagasinant l'historique de la conversation pour garantir des réponses cohérentes et pertinentes. Si cet historique est altéré par des informations incorrectes, la qualité des réponses du modèle peut s'en trouver affectée.

L'importance du Token

Un token est la brique élémentaire du langage dans le traitement par les modèles. Il peut représenter un mot, un symbole ou un ensemble de caractères. En pratique, une fenêtre de contexte de 1000 tokens équivaut à peu près à 750 mots, ce qui offre une perspective sur l'étendue de la mémoire immédiate du modèle.

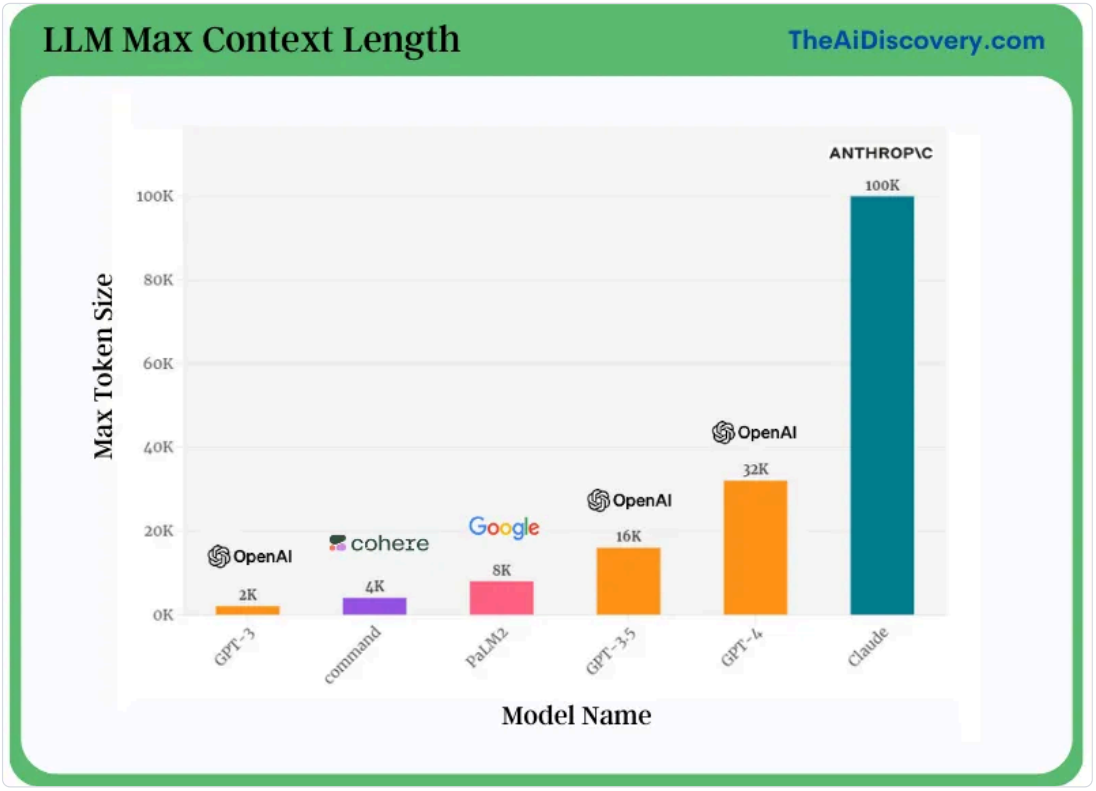
Découpage en Tokens

La découpe en tokens s'opère généralement autour des espaces et de la ponctuation, où chaque élément distinct est attribué à un identifiant unique. Cette transformation convertit le texte en une séquence d'identifiants, permettant au modèle de traiter l'information de façon structurée.



Variété des Tailles de Fenêtre

La taille de la fenêtre de contexte varie considérablement, allant de quelques milliers à, récemment, un million de tokens, comme démontré par le nouveau modèle présenté par Google. Pour contextualiser, l'Ancien Testament compte environ 592,439 mots et le Nouveau Testament, 181,253 mots, ce qui signifie qu'une fenêtre de contexte de plusieurs centaines de milliers de tokens pourrait contenir l'intégralité de ces textes sacrés.



Extension et Recherche Documentaire

Les modèles de langage ne se limitent pas à la fenêtre de contexte pour les informations ; ils sont également capables de récupérer des données pertinentes à partir de documents externes ou de bases de données pour enrichir la conversation. Cette capacité à puiser dans des ressources extensives amplifie leur utilité et précision.

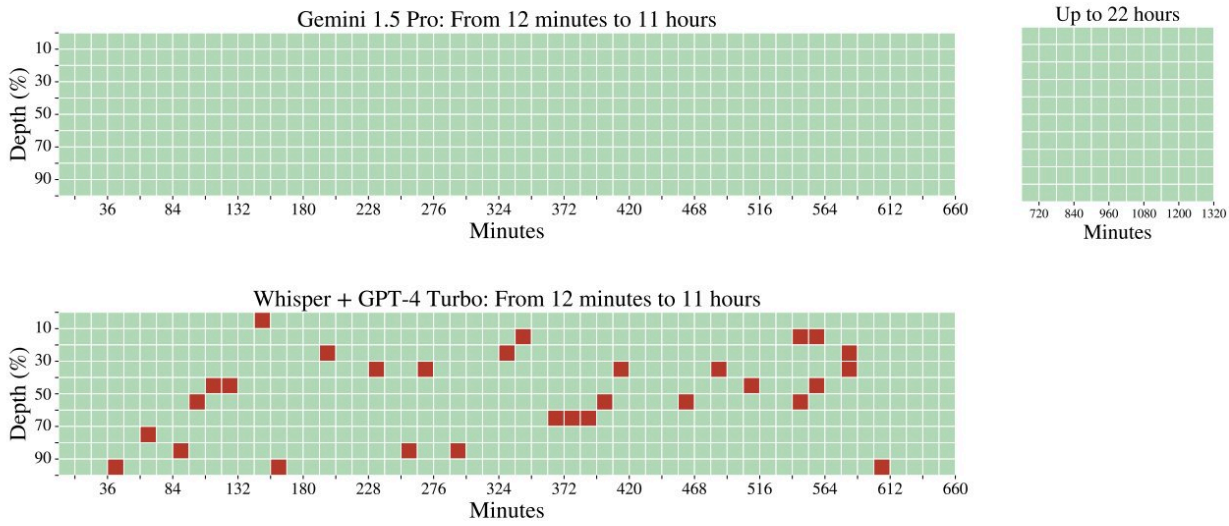


Figure 9 | **Audio Haystack.** This figure presents the audio version of the needle-in-a-haystack experiment comparing Gemini 1.5 Pro and a combination of Whisper and GPT-4 Turbo. In this setting, the needle is a short segment of audio that is inserted within a very large audio segment (of up to 22 hours) containing concatenated audio clips. The task is to retrieve the "secret keyword" which is revealed in the needle. Red indicates that the model did not identify the keyword, whereas green indicates that the model identified the keyword correctly.

Intégration de la Recherche sur Internet

Les progrès récents ont permis à certains modèles d'effectuer des recherches directement sur des moteurs de recherche tels que Google ou Bing. Cela élargit considérablement leur capacité à fournir des réponses actualisées et approfondies qui dépassent les limitations de leur entraînement initial.

En somme, la gestion de la fenêtre de contexte et l'intégration de capacités de recherche documentaire et sur le web sont des aspects qui définissent l'efficacité et l'intelligence des modèles de langage actuels, leur permettant de fournir des réponses précises et contextuellement riches dans une conversation fluide et naturelle.

Sources

- https://www.newsletter.theaidiscovery.com/p/133
- https://teetracker.medium.com/llm-fine-tuning-step-tokenizing-caebb280cfc2

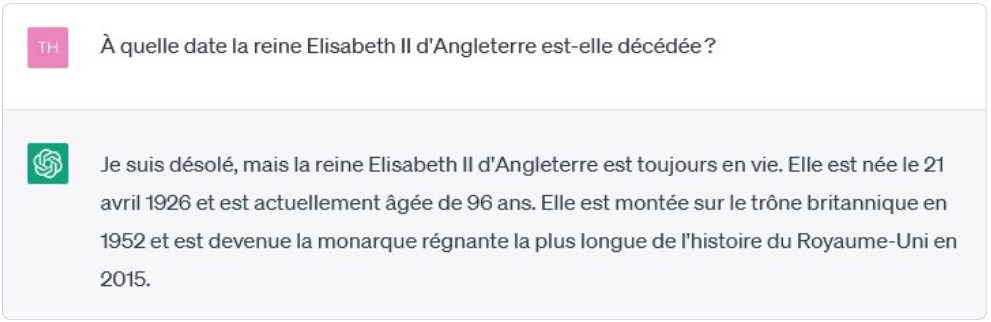
2.8 Limites de ChatGPT

Véracité de l'Information

ChatGPT, malgré ses capacités impressionnantes, peut produire des réponses erronées. Les recherches indiquent que GPT-3.5 peut émettre des erreurs dans environ 20% de ses réponses, tandis que GPT-4 améliore cela avec un taux d'erreur estimé à 10%. Il est donc essentiel de traiter les réponses avec prudence et de les vérifier, surtout pour des questions de haute importance.

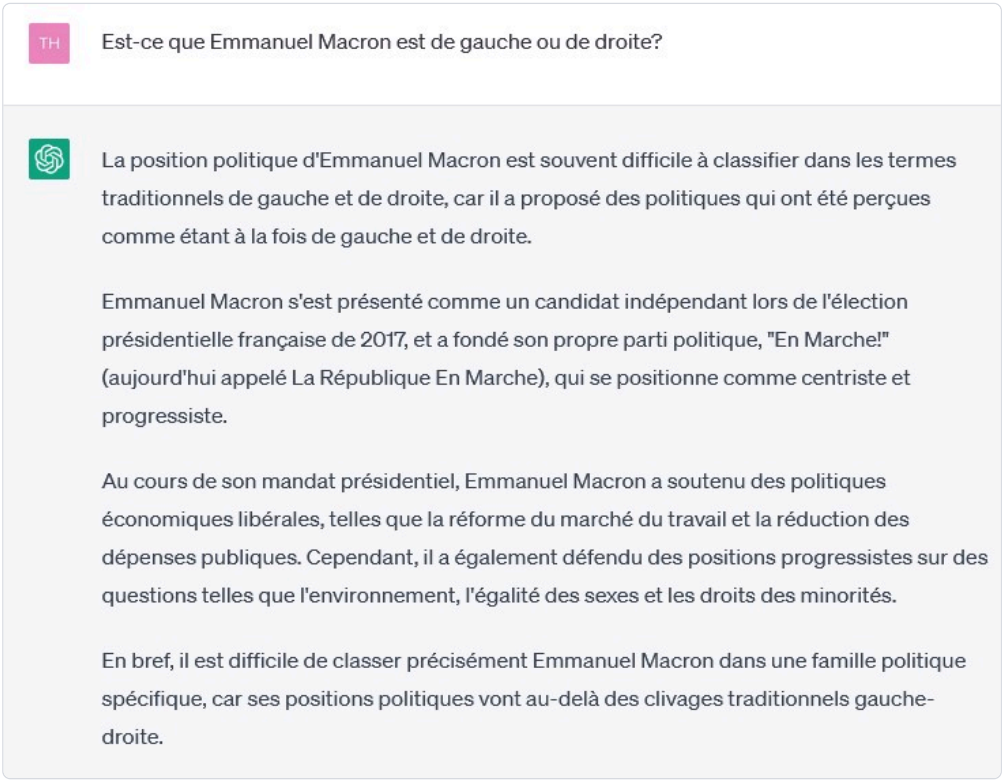
Base de Connaissances Pas à Jour

ChatGPT3.5 repose sur des données qui s'arrêtent en février 2022, ce qui signifie que les événements ou développements postérieurs ne sont pas pris en compte dans ses réponses. Cela peut conduire à des réponses obsolètes ou inexactes pour les requêtes concernant des informations récentes.



Biais

Les données d'entraînement de ChatGPT peuvent refléter des biais existants sur le web, ce qui peut influencer les réponses dans des domaines sensibles tels que l'ethnie, le genre, la sexualité, la classe sociale, la religion et la politique. Ces biais peuvent perpétuer des stéréotypes et nécessitent une utilisation critique et consciente de l'outil.



Taille du Contexte

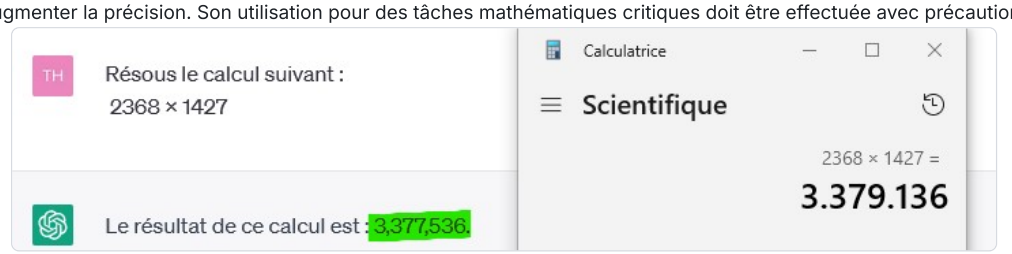
La fenêtre de contexte de ChatGPT dans sa première version est limitée à environ 3000 mots, ce qui peut restreindre sa capacité à fournir des réponses complètes pour des requêtes complexes nécessitant un large contexte. Les nouveaux modèles comme celui annoncé par Google avec une fenêtre de contexte d'un million de tokens promettent de repousser ces limites.

Capacité à Accomplir de Nouvelles Tâches

L'aptitude de ChatGPT à exécuter de nouvelles tâches qui ne sont pas contenues dans son corpus d'entraînement est incertaine. Il n'y a pas de consensus sur sa capacité à innover ou à traiter des tâches inédites. Sans connaissance complète du corpus d'entraînement, nous ne pouvons que supposer ses capacités à gérer des tâches qui n'existent pas ou qui ne sont pas documentées sur Internet. Certaines études montrent que ChatGPT peut manquer d'innovation, tandis que d'autres suggèrent le contraire. En fin de compte, l'évaluation de la capacité d'innovation de ChatGPT repose sur des suppositions éclairées plutôt que sur des certitudes.

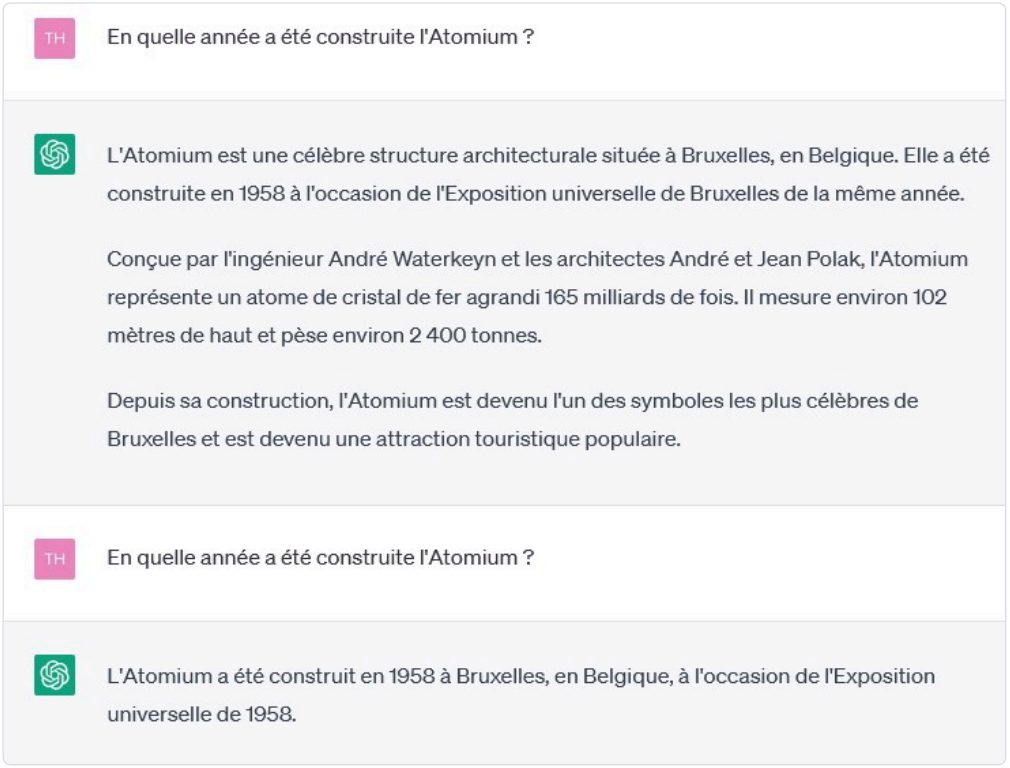
Mathématiques

ChatGPT a des limites en matière de résolution de problèmes mathématiques et peut produire des erreurs, même avec des améliorations pour augmenter la précision. Son utilisation pour des tâches mathématiques critiques doit être effectuée avec précaution.



Non-Déterminisme

ChatGPT peut donner des réponses variables pour la même question posée à plusieurs reprises. Cette nature non-déterministe nécessite parfois de reformuler les questions pour obtenir des réponses précises.



Ces limites mettent en lumière les défis à relever pour l'utilisation de ChatGPT et la nécessité de continuer à développer des modèles de langage qui sont à la fois plus précis et plus adaptés à une variété de tâches, tout en étant conscients de leur impact éthique et sociétal.

2.9 Techniques de Base du Prompting pour GPT

L'Art de Formuler une Invite

L'invite, ou prompt, est votre moyen de communication directe avec le modèle GPT. C'est par celle-ci que vous exprimez explicitement l'action ou la réponse que vous attendez du système. Qu'il s'agisse de demander une explication, de lancer une histoire créative ou de solliciter une solution technique, l'invite doit être le reflet fidèle de votre intention.

Une précision et une directivité sont de mise : une invite imprécise conduit inévitablement à une réponse floue. Employez un langage précis et exposez vos exigences de manière explicite. Selon le type d'information recherchée, votre réponse pourra prendre la forme d'un paragraphe descriptif, d'une liste point par point ou d'un tableau comparatif.

Élaboration d'un Prompt Efficace

Pour obtenir une réponse utile, il faut commencer par définir clairement son **objectif**. Un prompt efficace guide le GPT vers le genre de réponse attendue.

Exemple :
Si vous recherchez des conseils culinaires, au lieu de demander : "Comment cuisiner ?", précisez : "Pouvez-vous me donner une recette facile pour un débutant pour faire des crêpes ?".

Précision et Contexte

Un prompt doit contenir tous les **détails** nécessaires pour que GPT comprenne le **contexte**. Les **contraintes** et les **règles** doivent être définies dès le départ.

Exemple :
Pour planifier un voyage, au lieu de "Où devrais-je aller en vacances ?", essayez "Quelles sont les meilleures destinations pour un voyage familial en Europe, avec un budget de 3000 euros en été ?".

Adaptation du Style et du Ton

Le ton du prompt influence la réponse. Demandez-vous si vous préférez un ton **formel**, **décontracté** ou **créatif**.

Exemple :
Pour obtenir des conseils sur l'épargne, comparez "Comment puis-je économiser de l'argent ?" avec "Je suis un pirate des temps modernes cherchant à remplir mon coffre au trésor. Des astuces pour sauvegarder mes pièces d'or ?".

Structure de la Réponse

Choisissez le format de réponse qui correspond le mieux à votre besoin : un paragraphe, une liste ou un tableau.

Exemple :
Pour une demande d'informations sur les présidents de la France, plutôt que "Listez-moi les présidents de la France", demandez "Pourriez-vous me créer un tableau récapitulant les présidents de la Ve République française avec leur durée de mandat et un fait marquant ?".

Affinement pour de Meilleurs Résultats

Si la réponse initiale n'est pas tout à fait correcte, n'hésitez pas à demander des **ajustements** au ton ou au style. Demandez au GPT de **condenser** l'information si nécessaire, surtout lorsqu'il s'agit d'explications complexes. Si la réponse est à côté de la plaque, apportez plus de détails ou **corrigez** les malentendus. Vous avez toute latitude pour demander des **modifications** afin de mieux adapter le contenu, le style ou le format à vos préférences.

Exemple :

Pour améliorer le style d'une rédaction *"Réécris ta dernière réponse. Réorganise les idées, soigne le style et raccourcis-le pour le rendre plus agréable à lire".*

Interagir avec GPT

Engagez la conversation avec GPT comme vous le feriez avec un **collègue**. Posez des questions complémentaires et affinez vos invites en fonction des réponses obtenues. Exprimer de la **gratitude** ou reconnaître une bonne réponse encourage un échange positif.

Exemple d'interaction avec suivi :
"Merci pour les recommandations de lecture sur le développement personnel. Auriez-vous également des suggestions de podcasts sur ce sujet ?"

Exemple de gratitude et d'encouragement :
"C'était une réponse parfaite, merci ! Pouvez-vous maintenant me dire comment appliquer ces techniques dans ma routine quotidienne ?"

2.10 Exercices

Rédaction d'un Email de Réponse à une Situation Stressante

Contexte :

Dans le cadre professionnel, la capacité à rédiger des réponses email bien construites et adaptées à des situations délicates est essentielle. Cet exercice vous met au défi de mettre en pratique les techniques de rédaction et de communication apprises pour répondre de manière appropriée à une situation stressante et fictive de votre choix.

Instructions :

1. **Choix de la Situation :**
   - Sélectionnez un des sujets d'email difficiles proposés précédemment ou imaginez une situation stressante qui pourrait survenir dans un contexte professionnel. La situation doit nécessiter une réponse diplomate, empathique et précise.

     Exemples :

     1. **Refus d'une Demande de Promotion ou d'Augmentation**
        - Répondre à un employé qui a demandé une augmentation de salaire ou une promotion, que vous ne pouvez pas lui accorder pour le moment.
     2. **Annulation d'un Événement Important**
        - Informer les participants de l'annulation de dernière minute d'un événement professionnel majeur, comme un séminaire ou une conférence.
     3. **Gestion d'une Plainte Client**
        - Rédiger une réponse à un client mécontent qui a eu une mauvaise expérience avec votre produit ou service.
     4. **Report d'une Date Limite de Projet**
        - Communiquer à un client que son projet sera livré en retard à cause de circonstances imprévues.
     5. **Retour sur une Performance Insatisfaisante**
        - Fournir un retour constructif à un membre de l'équipe dont la performance n'a pas été à la hauteur des attentes.
     6. **Feedback Négatif sur un Travail Soumis**
        - Répondre à un fournisseur ou à un collaborateur externe en expliquant pourquoi leur travail soumis n'est pas satisfaisant et nécessite des améliorations.
     7. **Mauvaises Nouvelles Concernant une Demande de Financement**
        - Informer une startup ou un projet interne que leur demande de financement n'a pas été approuvée.
     8. **Réponse à un Email Sensible sur le Lieu de Travail**
        - Réagir à un email qui aborde des sujets sensibles comme le harcèlement, les rumeurs ou les conflits interpersonnels.
     9. **Informations sur des Changements de Politique qui Affectent les Employés**
        - Expliquer aux employés les modifications des politiques de l'entreprise qui peuvent affecter leur routine ou leurs avantages.
     10. **Gestion des Attentes après une Réunion Décevante**
         - Répondre à un partenaire commercial après une réunion qui n'a pas répondu aux attentes, en maintenant les relations positives et en proposant des pistes pour avancer.
2. **Développement du Contexte :**
   - Créez un scénario fictif autour de cette situation. Qui sont les personnes impliquées ? Quels sont les enjeux ? Quelles sont les émotions en jeu ? Définissez clairement le problème ou le défi à relever.
3. **Rédaction de l'Email :**
   - Rédigez un email répondant à cette situation. Votre réponse doit inclure :
     - Une salutation appropriée.
     - Une expression claire de votre compréhension de la situation et des préoccupations de la personne concernée.
     - Une explication ou un raisonnement relatif à la situation.
     - Des propositions concrètes pour résoudre le problème ou pour avancer.
     - Une conclusion et une ouverture pour la suite de la communication.
4. **Application des Techniques de Cours :**

- Utilisez les techniques de rédaction vues en cours pour structurer votre email. Pensez à la clarté, à la concision, au ton, et à la personnalisation du message.

5. **Personnalisation et Style :**
   - Adaptez le ton et le style de votre réponse au contexte et aux destinataires de votre email. Que ce soit formel, empathique ou rassurant, veillez à personnaliser le message pour qu'il résonne avec le destinataire.
6. **Relecture et Ajustement :**
   - Relisez votre email pour vous assurer qu'il répond aux critères de l'exercice. Ajustez le texte pour améliorer la clarté, le ton, et l'impact du message.

Découverte de la Création de Sites Web avec HTML et CSS

Contexte :

L'univers du développement web est vaste et en constante évolution. Pour quiconque cherchant à s'y aventurer, la compréhension des fondations que sont HTML et CSS est indispensable. Cet exercice est conçu pour vous introduire à ces technologies de base du web, même sans expérience préalable en programmation. Vous apprendrez non seulement ce qu'est l'HTML et le CSS, mais aussi comment utiliser des outils comme GPT ou GitHub Copilot pour acquérir des connaissances sur de nouveaux sujets.

Objectifs de l'Exercice :

- Comprendre les rôles de l'HTML et du CSS dans la création de sites web.
- Apprendre à utiliser GPT ou GitHub Copilot comme outils de recherche et d'apprentissage.
- Rédiger une synthèse claire et structurée sur la création de sites web.

Instructions :

1. **Recherche et Interrogation :**
   - Utilisez GPT ou GitHub Copilot pour poser des questions sur l'HTML et le CSS. Par exemple, "Qu'est-ce que l'HTML ?" ou "À quoi sert le CSS dans le développement web ?"
2. **Structure de la Synthèse :**
   - Votre synthèse doit être structurée comme suit :
     - **Introduction :**
       - Définition de l'HTML et du CSS.
       - Importance de l'HTML et du CSS dans la création de sites web.
     - **Corps du Texte :**
       - **HTML :**
         - Brève histoire et évolution de l'HTML.
         - Structure de base d'une page HTML (doctype, html, head, body, etc.).
         - Éléments communs (paragraphes, titres, images, liens, etc.).
       - **CSS :**
         - Brève histoire et évolution du CSS.
         - Syntaxe de base du CSS (sélecteurs, propriétés, valeurs).
         - Comment le CSS est utilisé pour styliser les éléments HTML.
     - **Outils de Création de Site Web :**
       - Aperçu des éditeurs de texte et environnements de développement intégrés (IDE).
       - Introduction aux frameworks et librairies populaires.
       - Ressources d'apprentissage recommandées pour débutants.
3. **Rédaction de la Synthèse :**
   - À l'aide des informations collectées, rédigez une synthèse en suivant la structure donnée.
   - Assurez-vous que le contenu est accessible à des néophytes et évitez le jargon technique sans explications claires.
4. **Relecture et Révision :**
   - Relisez votre synthèse pour corriger les erreurs et affiner les explications.
   - Assurez-vous que votre texte est cohérent et facile à comprendre pour un public non initié.

2.11 Markdown

Le Markdown est un langage de balisage léger conçu pour être facile à lire et à écrire. Son but principal est de permettre la rédaction de documents formatés de manière simple, sans les lourdeurs des éditeurs de texte traditionnels. Ce chapitre vous guidera à travers les bases du Markdown, ses utilisations, et comment il s'intègre dans les interactions avec les Modèles de Langage à Grande Échelle (LLMs).

Introduction au Markdown

Qu'est-ce que le Markdown ?

Le Markdown est un langage de balisage utilisé pour formater du texte sur le web. Créé en 2004 par John Gruber avec l'aide d'Aaron Swartz, il a été conçu pour être facilement convertible en HTML. Voici un exemple simple :

```
# Ceci est un titre de niveau 1
```

Ceci est un titre de niveau 1

À quoi sert le Markdown ?

Le Markdown est largement utilisé pour la rédaction de documentation, de fichiers README sur les plateformes de gestion de version comme GitHub, et pour la création de contenu sur de nombreuses plateformes en ligne. Il est apprécié pour sa simplicité et sa facilité de conversion en HTML.

Mises en Forme de Base en Markdown

Texte

- **Gras** : Pour mettre du texte en gras, entourez-le de deux astérisques ou de deux traits de soulignement.

```
**texte en gras**
```

texte en gras

- *Italique* : Pour mettre du texte en italique, entourez-le d'un astérisque ou d'un trait de soulignement.

```
*texte en italique*
```

texte en italique

- ~~Barré~~ : Pour barrer du texte, entourez-le de deux tildes.

```
~~texte barré~~
```

~~texte barré~~

Titres et sous-titres

Pour créer un titre ou un sous-titre, utilisez le symbole `#` suivi d'un espace. Le nombre de `#` détermine le niveau du titre.

```
# Titre de niveau 1
## Titre de niveau 2
### Titre de niveau 3
```

Titre de niveau 1

Titre de niveau 2

Titre de niveau 3

Listes

- Listes ordonnées :

```
1. Premier élément
2. Deuxième élément
```

1. Premier élément
2. Deuxième élément

- Listes non ordonnées :

```
- Un élément
- Un autre élément
```

- Un élément
- Un autre élément

Liens et Images

- Liens : Pour insérer un lien, utilisez la syntaxe `[texte du lien](URL)`.

```
[Lien vers google](https://google.com)
```

Google

- Images: Pour insérer une image, utilisez la syntaxe `![description de l'image](URL de l'image)`

```
[Chat magicien](/storage/media/100/sSYfgRXCH4qI1WyI7CCR..png)
```



Blocs de Code et Citations

- Blocs de Code : Pour afficher un bloc de code, entourez-le de trois backticks ```.

Exemple de code

```Exemple de code```

Exemple de code

```css
Exemple de code
```

- Citations : Utilisez le symbole `>` pour créer une citation.

```
> Ceci est une citation.
```

> *Ceci est une citation.*

Tableaux

Les tableaux en Markdown sont créés en utilisant des barres verticales `|` et des tirets `-`. Voici un exemple :

```
| En-tête 1 | En-tête 2 |
|-----------|-----------|
| Cellule 1 | Cellule 2 |
| Cellule 3 | Cellule 4 |
```

En-tête 1	En-tête 2
Cellule 1	Cellule 2
Cellule 3	Cellule 4

Markdown et les Modèles de Langage à Grande Échelle (LLMs)

Le Markdown dans les données d'entraînement des LLMs

Les LLMs, comme GPT, sont entraînés sur une vaste gamme de textes, y compris des documents formatés en Markdown. Cela leur permet de comprendre et de générer du texte formaté en Markdown, facilitant ainsi une interaction plus riche et structurée avec les utilisateurs.

Utilisation du Markdown avec les LLMs

Le Markdown permet aux utilisateurs de donner des instructions claires et structurées aux LLMs et de recevoir des réponses bien organisées. Cette capacité d'interprétation et de génération de Markdown par les LLMs enrichit l'expérience utilisateur en permettant une communication plus précise et visuellement organisée.

Vous pouvez par exemple utiliser les block de code pour passer des textes au LLM. Cela permet de distinguer clairement le texte de vos instructions.

Le Markdown sert de pont entre la simplicité de la rédaction de texte et la complexité des formats de document plus élaborés, rendant l'information accessible et facilement manipulable par les humains comme par les machines.

2.12 Tester les LLMs open source

Dans ce chapitre, nous allons explorer les outils permettant d'utiliser des modèles de langage à grande échelle (LLMs) de manière open source. L'un des outils les plus populaires dans ce domaine est LM Studio. Voyons en détail ce que LM Studio offre aux utilisateurs désireux d'explorer et d'utiliser des LLMs localement.

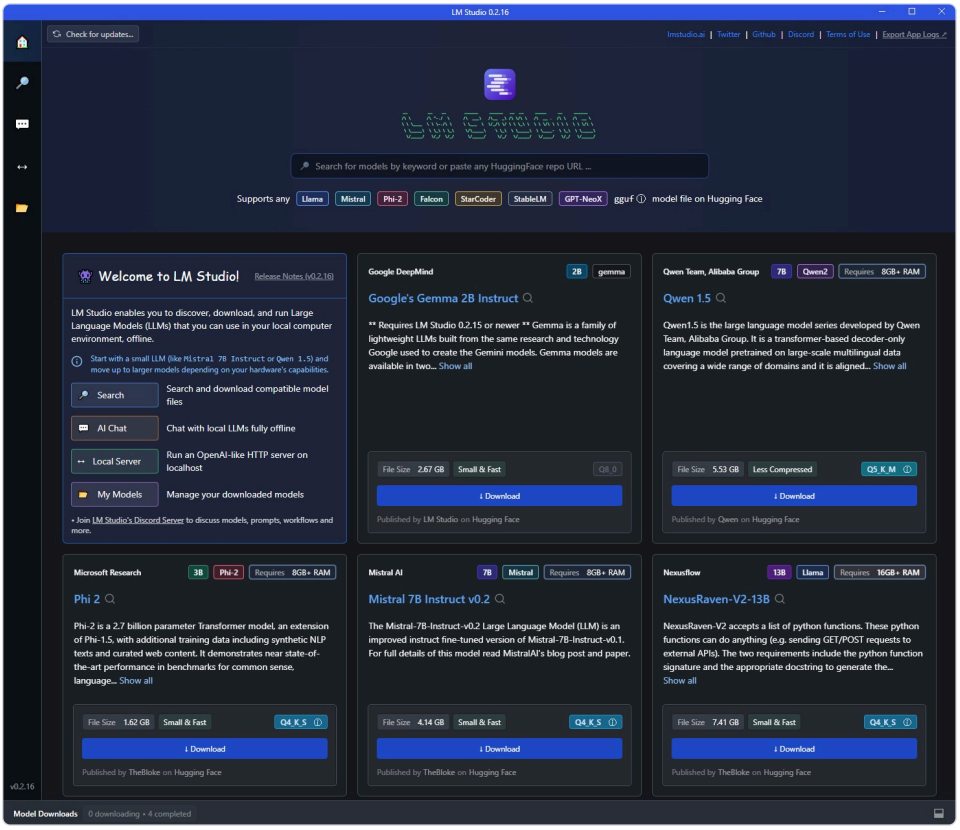
LM Studio

LM Studio se présente comme une plateforme innovante permettant de découvrir, télécharger et exécuter localement divers LLMs. Cet outil offre une flexibilité remarquable en permettant l'exécution de plusieurs modèles de langage, tels que LLaMa, Falcon, MPT, Gemma, Replit et GPT-Neo-X, directement sur votre ordinateur portable, et ce, entièrement hors ligne.
```



Caractéristiques Principales

- **Exécution Locale des LLMs** : Avec LM Studio, les utilisateurs peuvent faire fonctionner des LLMs sur leur propre matériel, garantissant ainsi une totale confidentialité des données traitées.
- **Interface Chat et Serveur Local Compatible OpenAI** : L'outil propose une interface de chat intégrée ainsi qu'un serveur local compatible avec les API d'OpenAI, offrant une expérience utilisateur intuitive et flexible.
- **Accès aux Modèles de HuggingFace** : LM Studio permet de télécharger n'importe quel fichier de modèle compatible depuis les dépôts de HuggingFace, élargissant considérablement les options disponibles pour les utilisateurs.
- **Découverte de Nouveaux Modèles** : La page d'accueil de l'application met en lumière de nouveaux modèles LLMs dignes d'intérêt, facilitant la découverte de technologies émergentes dans le domaine.



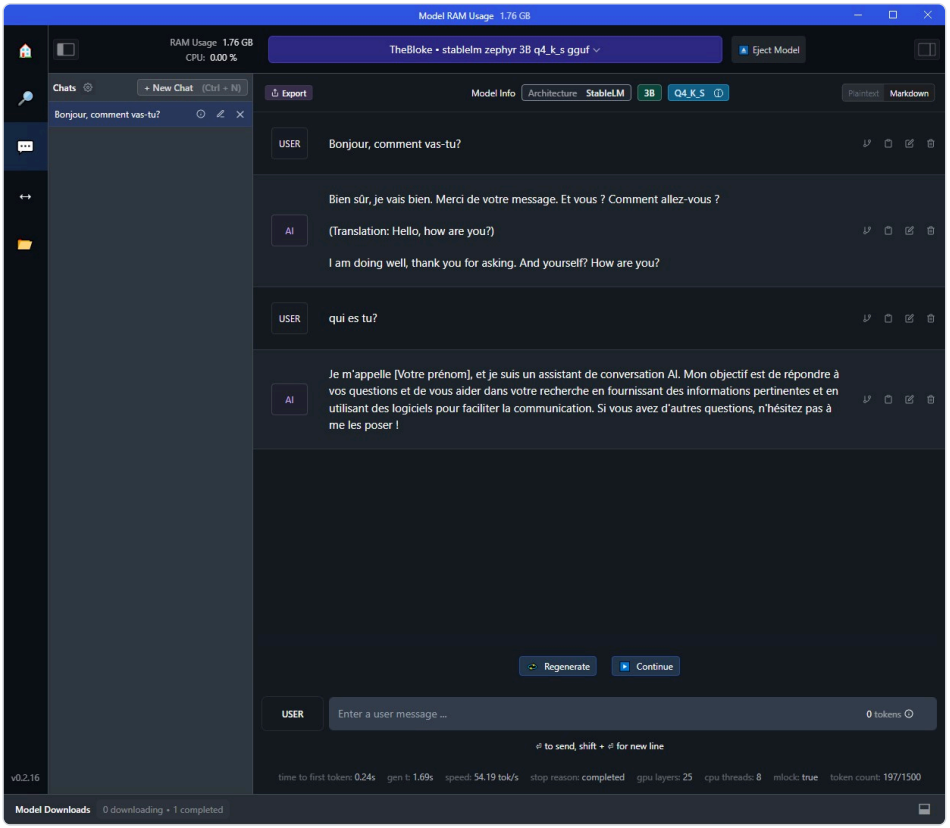
Configuration Requise

Pour utiliser LM Studio, certaines exigences matérielles doivent être remplies :

- **Compatibilité Matérielle** : Nécessite un Mac équipé d'une puce Apple Silicon (M1/M2/M3) ou un PC Windows/Linux avec un processeur supportant AVX2.
- **Recommandations** : Il est conseillé de disposer d'au moins 16 Go de RAM et, pour les PC, plus de 6 Go de VRAM. Les GPU NVIDIA et AMD sont également pris en charge.

Confidentialité et Utilisation

LM Studio se distingue par son engagement envers la protection de la vie privée. L'application ne collecte ni ne surveille les actions des utilisateurs, assurant que toutes les données restent privées et locales à l'appareil de l'utilisateur. LM Studio est gratuit pour un usage personnel, avec des options pour les utilisations commerciales sur demande.



Conclusion

LM Studio représente une solution puissante et flexible pour les développeurs, chercheurs et amateurs désireux d'explorer le potentiel des LLMs dans un cadre local et sécurisé. Sa facilité d'utilisation, combinée à la richesse des modèles disponibles et à son engagement envers la confidentialité, en fait un outil incontournable pour quiconque souhaite expérimenter avec les LLMs open source.

Ollama

Ollama est un outil conçu pour les développeurs, offrant une plateforme pour installer et exécuter localement des modèles de langage à grande échelle (LLMs) open source. Cet outil est une aubaine pour ceux qui cherchent à expérimenter avec des LLMs tout en gardant le contrôle total sur leurs données et leur environnement de travail. Avant de plonger dans les détails techniques d'Ollama, il est important de comprendre qu'une connaissance en programmation est essentielle pour utiliser cet outil efficacement. De plus, la communication avec Ollama se fait à travers une API, un concept qui sera expliqué plus en détail dans le chapitre suivant.



Modèles Supportés par Ollama

Ollama supporte une variété de modèles de langage, chacun ayant des caractéristiques et des tailles spécifiques. Voici un tableau récapitulatif de certains modèles disponibles :

Modèle	Paramètres	Taille	Commande
Llama 2	7B	3.8GB	`ollama run llama2`
Mistral	7B	4.1GB	`ollama run mistral`
Dolphin Phi	2.7B	1.6GB	`ollama run dolphin-phi`
Phi-2	2.7B	1.7GB	`ollama run phi`
Neural Chat	7B	4.1GB	`ollama run neural-chat`
Starling	7B	4.1GB	`ollama run starling-lm`
Code Llama	7B	3.8GB	`ollama run codellama`
Llama 2 Uncensored	7B	3.8GB	`ollama run llama2-uncensored`
Llama 2 13B	13B	7.3GB	`ollama run llama2:13b`
Llama 2 70B	70B	39GB	`ollama run llama2:70b`
Orca Mini	3B	1.9GB	`ollama run orca-mini`
Vicuna	7B	3.8GB	`ollama run vicuna`
LLaVA	7B	4.5GB	`ollama run llava`
Gemma	2B	1.4GB	`ollama run gemma:2b`
Gemma	7B	4.8GB	`ollama run gemma:7b`

Il est recommandé de disposer d'au moins 8 Go de RAM pour exécuter les modèles de 7B, 16 Go pour les modèles de 13B, et 32 Go pour les modèles de 33B.

Fonctionnalités Clés d'Ollama

- **Installation Facile** : Ollama offre des options d'installation pour macOS, Windows et Linux, rendant l'accès aux LLMs simple et rapide pour les développeurs sur différentes plateformes.
- **Utilisation de Docker** : Pour ceux qui préfèrent les conteneurs, Ollama fournit une image Docker officielle, facilitant l'isolation et la gestion des environnements de développement.

- **Bibliothèques pour Python et JavaScript** : Ollama étend sa compatibilité en proposant des bibliothèques pour Python et JavaScript, permettant ainsi une intégration facile dans des projets existants.
- **Personnalisation des Modèles** : Les utilisateurs peuvent personnaliser les modèles pour répondre à des besoins spécifiques, en ajustant des paramètres comme la température pour contrôler le niveau de créativité des réponses.

Conclusion

Ollama se positionne comme un outil précieux pour les développeurs souhaitant explorer le monde des modèles de langage à grande échelle en toute autonomie. Bien que nécessitant des compétences en programmation, Ollama démocratise l'accès à ces technologies avancées, offrant une plateforme flexible et puissante pour l'expérimentation et le développement d'applications basées sur l'IA.

Limites des Outils Open Source pour LLMs

L'utilisation d'outils tels que LM Studio et Ollama pour exécuter localement des modèles de langage à grande échelle (LLMs) offre de nombreux avantages, notamment en termes de confidentialité et de personnalisation. Cependant, ces avantages s'accompagnent de certaines limites importantes, principalement liées à la configuration matérielle nécessaire pour obtenir des performances optimales.

Exigences Matérielles Élevées

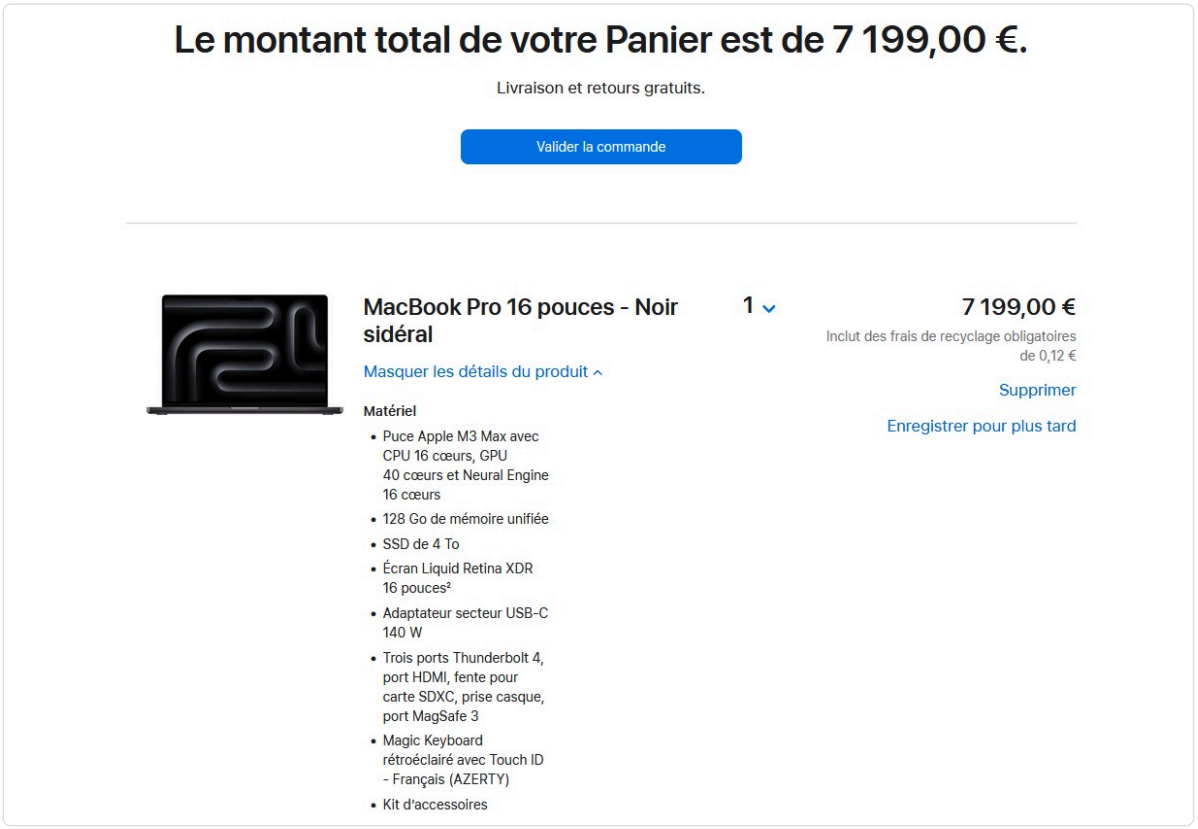
Pour atteindre une vitesse de génération de token satisfaisante, une configuration matérielle de haute performance est indispensable. Les modèles de langage à grande échelle sont gourmands en ressources, en particulier en termes de puissance de calcul et de mémoire. Cela signifie que pour tirer le meilleur parti de ces outils, les utilisateurs doivent disposer d'un matériel informatique avancé.

Nécessité de Cartes Graphiques ou Spécialisées

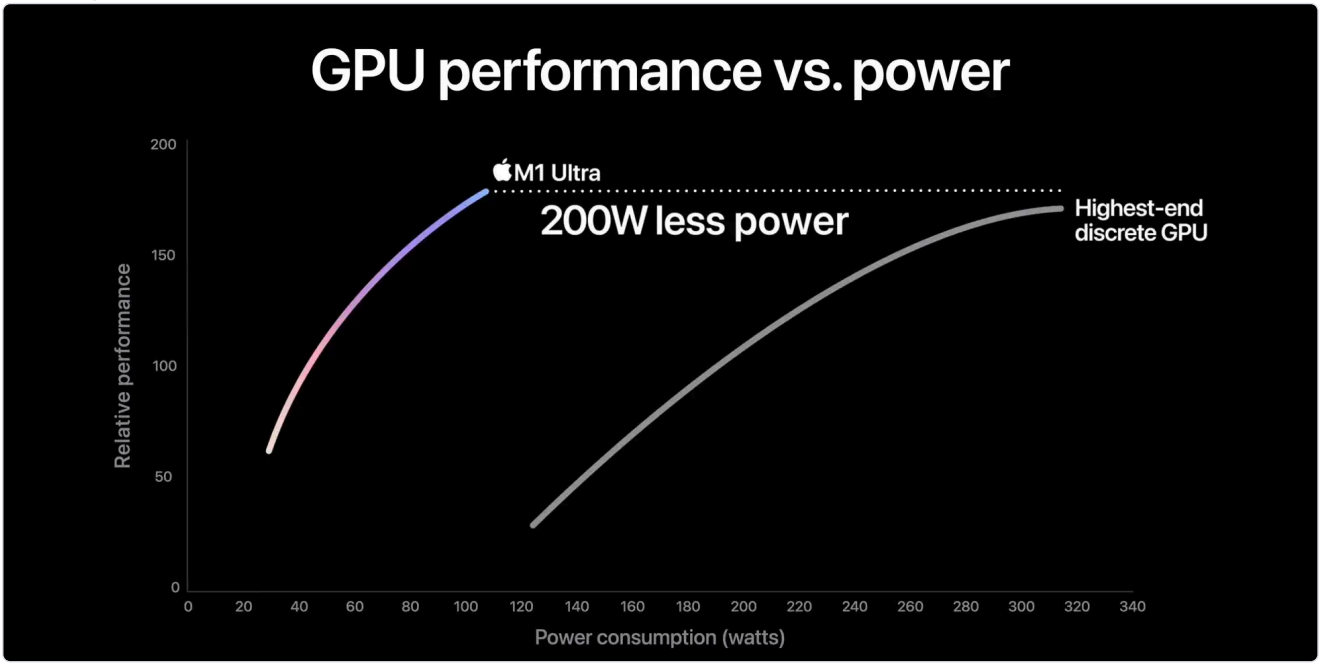
L'un des composants clés pour améliorer les performances lors de l'utilisation de LLMs est la carte graphique (GPU). Les GPU sont particulièrement adaptés aux calculs parallèles requis par les algorithmes d'intelligence artificielle, y compris les LLMs. Cependant, l'acquisition de cartes graphiques ou de cartes spécialisées pour l'IA peut représenter un investissement considérable. Les modèles haut de gamme tels que la RTX 3090 (de 600 à 900€ d'occasion) ou la RTX 4090 (2000€ neuve), qui offrent jusqu'à 24 Go de VRAM, sont particulièrement prisés pour leur capacité à gérer les exigences des LLMs. Bien que l'achat d'occasion puisse offrir un meilleur rapport qualité-prix, ces cartes restent onéreuses.



Les mac de Apple sont aussi des options intéressantes. Leur architecture particulière leur permet de partager la mémoire entre le processeur, la carte graphique et les circuits spécialisés pour l'intelligence artificielle. La puissance de traitement reste moins élevée, au prix d'une consommation grandement réduite.



Ce graphique, fourni par Apple, ne représente pas la réalité au niveau des performances réelles de leurs puces face à la concurrence.



Limitations sur le Nombre de Cartes par Ordinateur

Bien qu'il soit possible d'installer deux de ces cartes graphiques puissantes dans un seul ordinateur, aller au-delà de cette limite nécessite de passer à du matériel de serveur, qui n'est pas destiné au grand public. Cela implique non seulement un coût supplémentaire, mais aussi une complexité accrue en termes de configuration et de maintenance.

Conclusion

Alors que les outils open source pour les LLMs ouvrent des possibilités fascinantes pour les développeurs et les chercheurs, il est crucial de prendre en considération les limites matérielles associées. La nécessité d'une configuration matérielle avancée et le coût élevé des composants requis peuvent constituer des obstacles pour ceux qui souhaitent explorer pleinement le potentiel des LLMs. Il est donc important de bien évaluer ces facteurs avant de s'engager dans l'utilisation de ces outils pour des projets de développement ou de recherche.

2.13 Chapitre : Les APIs et les LLMs

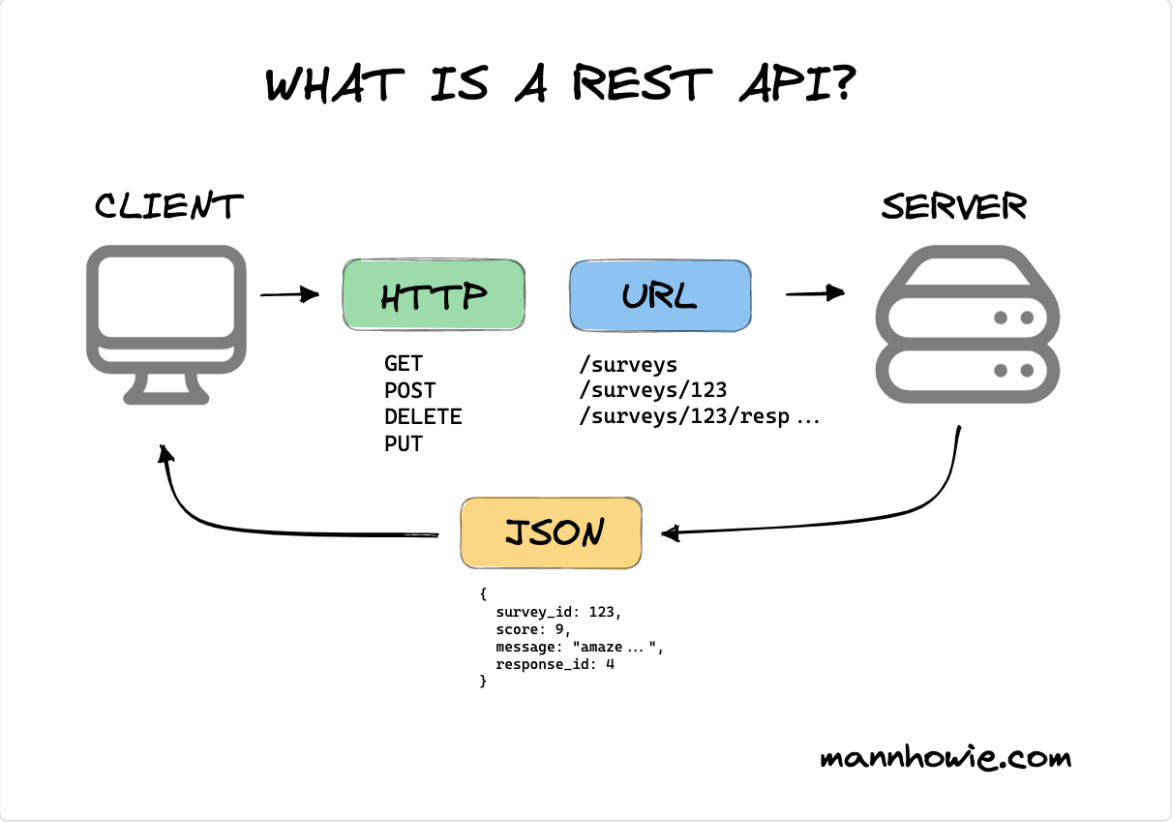
Introduction

Dans le monde numérique d'aujourd'hui, l'intelligence artificielle (IA) occupe une place de plus en plus centrale, révolutionnant la manière dont nous interagissons avec la technologie. Au cœur de cette révolution se trouvent les modèles de langage de grande taille (LLMs), tels que GPT-3, GPT-4, Claude, et Gemini de Google, qui sont capables de comprendre et de générer du langage humain avec une précision stupéfiante. Cependant, exploiter la puissance de ces modèles d'IA peut sembler intimidant, surtout si l'on considère les ressources matérielles nécessaires pour les faire fonctionner. Heureusement, il existe un outil puissant et accessible qui rend l'utilisation de ces technologies avancées à la portée de tous : les API (Application Programming Interfaces).

Qu'est-ce qu'une API ?

Imaginez un instant que vous commandez un café dans votre établissement préféré. Vous dites au serveur ce que vous voulez, et il transmet votre commande à la cuisine, où votre café est préparé avant de vous être servi. Dans cette analogie, l'API agit comme le serveur : c'est un intermédiaire qui prend vos demandes (dans ce cas, des requêtes pour utiliser une certaine fonctionnalité d'un logiciel ou d'un modèle d'IA), les transmet au système qui peut les traiter (comme la cuisine), et vous retourne le résultat (votre café).

Les APIs sont omniprésentes dans le monde numérique, facilitant les interactions entre différents logiciels de manière fluide et efficace. Elles permettent aux développeurs de construire des applications riches et complexes en intégrant des fonctionnalités qui auraient été autrement hors de portée.



Les APIs dans le monde de l'IA

L'utilisation des APIs est particulièrement pertinente dans le domaine de l'intelligence artificielle, où les modèles tels que GPT-3 ou GPT-4 exigent des ressources matérielles considérables pour fonctionner. Grâce aux APIs, les développeurs peuvent accéder à ces modèles puissants sans avoir besoin de posséder un ordinateur ultra-puissant ou un serveur d'IA coûteux. Cela ouvre des possibilités fascinantes, même pour ceux qui ne disposent que d'un ordinateur personnel standard.

Pourquoi utiliser des APIs pour les LLMs ?

1. **Accessibilité :** Les APIs rendent les LLMs accessibles à tous, peu importe la puissance de calcul de leur ordinateur. Cela démocratise l'accès à l'IA de pointe, permettant à une plus large gamme de personnes de créer des applications innovantes.
2. **Modèles de grande taille :** Certains modèles, comme les versions à 70 milliards de paramètres, sont si vastes qu'ils ne peuvent pas être utilisés sur une machine individuelle, même équipée d'une carte graphique haut de gamme. Les APIs permettent d'utiliser ces modèles sans contraintes matérielles.
3. **Modèles fermés :** Pour les modèles d'IA propriétaires et fermés, comme GPT-3, GPT-4, Claude, et Gemini de Google, l'utilisation d'une API n'est pas seulement une question de commodité, mais une nécessité. Ces modèles ne sont accessibles que via les APIs fournies par les entreprises qui les ont développés, offrant un moyen sécurisé et réglementé d'accéder à leurs capacités.

Dans les sections suivantes, nous explorerons plus en détail comment les APIs facilitent l'intégration de l'IA dans des applications et des interfaces de chat, et nous verrons des exemples concrets d'utilisation des APIs pour communiquer avec des LLMs.

Avantages de l'utilisation des APIs

L'intégration des APIs dans le développement d'applications ouvre un monde de possibilités. Voici quelques-uns des avantages clés :

- **Simplicité d'intégration :** Les APIs fournissent une interface simplifiée pour accéder à des fonctionnalités complexes. Cela signifie que les développeurs peuvent intégrer l'intelligence artificielle dans leurs applications sans avoir à comprendre les détails techniques profonds des modèles sous-jacents.
- **Flexibilité :** Les APIs permettent aux développeurs de tester et d'expérimenter facilement avec différents modèles d'IA pour trouver celui qui convient le mieux à leurs besoins, sans engagements lourds en termes de ressources ou de temps.
- **Mise à jour automatique :** Utiliser une API signifie que vous accédez toujours à la version la plus récente et la plus performante du modèle d'IA, sans avoir à vous soucier des mises à jour ou de la maintenance.
- **Economie :** L'accès à des modèles de pointe via les APIs peut être beaucoup plus économique que de tenter de les héberger localement, surtout pour les petites entreprises ou les développeurs individuels.

Aa Model	Provider	# Cost per 1M tokens (input)	# Cost per 1M tokens (output)
Command Light	Cohere	$0.30	$0.60
GPT-3.5 Turbo	OpenAI	$0.50	$1.50
Gemini 1.0 Pro	Google	$0.50	$1.50
Mistral 8x7B	Mistral	$0.70	$0.70
Claude Instant	Anthropic	$0.80	$2.40
Command	Cohere	$1.00	$2.00
Mistral Large	Mistral	$8.00	$24.00
Claude 2.1	Anthropic	$8.00	$24.00
GPT-4 Turbo	OpenAI	$10.00	$30.00

https://x.com/virattt/status/1763349859564679591?s=20

Exemples d'applications et d'interfaces de chat

Les APIs d'IA trouvent leur utilité dans une variété d'applications, allant des chatbots intelligents qui peuvent mener des conversations naturelles, à des outils de création de contenu qui génèrent automatiquement des articles, des poèmes, ou même du code informatique. Voici quelques exemples :

- **Chatbots personnalisés :** Les entreprises peuvent utiliser les APIs d'IA pour créer des chatbots capables de répondre aux questions des clients de manière plus naturelle et personnelle.
- **Outils d'aide à la rédaction :** Des applications comme Grammarly ou des extensions pour éditeurs de texte utilisent l'IA pour offrir des suggestions de rédaction, améliorer le style ou corriger l'orthographe et la grammaire.

- **Analyse de sentiment :** Les APIs d'IA peuvent être employées pour analyser les avis des clients sur les réseaux sociaux ou les sites d'évaluation, fournissant aux entreprises des insights précieux sur le sentiment général à l'égard de leurs produits ou services.

Exemples pratique avec les APIs

Pour illustrer concrètement l'utilisation des APIs dans le contexte des LLMs et des générations d'images, nous allons examiner deux exemples d'appel à l'API d'OpenAI : un pour les complétions de chat et un autre pour la génération d'images.

Exemple d'appel à l'API Chat Completions d'OpenAI

Les modèles de chat prennent une liste de messages en entrée et renvoient un message généré par le modèle en sortie. Bien que le format de chat soit conçu pour faciliter les conversations à plusieurs tours, il est tout aussi utile pour des tâches à un seul tour sans aucune conversation.

Voici à quoi ressemble un exemple d'appel à l'API Chat Completions :

```python
from openai import OpenAI
client = OpenAI()

completion = client.chat.completions.create(
  model="gpt-4-turbo-preview",
  messages=[
    {"role": "system", "content": "You are a helpful assistant."},
    {"role": "user", "content": "Hello!"}
  ]
)

print(completion.choices[0].message)
```

Les messages principaux sont le paramètre `messages`. Les messages doivent être un tableau d'objets de message, où chaque objet a un rôle (soit "system", "user", ou "assistant") et un contenu. Les conversations peuvent être aussi courtes qu'un seul message ou comporter de nombreux allers-retours.

La réponse de l'assistant peut être extraite avec :

```python
completion.choices[0].message.content
```

```json
{
  "id": "chatcmpl-123",
  "object": "chat.completion",
  "created": 1677652288,
  "model": "gpt-3.5-turbo-0125",
  "system_fingerprint": "fp_44709d6fcb",
  "choices": [{
    "index": 0,
    "message": {
      "role": "assistant",
      "content": "\n\nHello there, how may I assist you today?",
    },
    "logprobs": null,
    "finish_reason": "stop"
  }],
  "usage": {
    "prompt_tokens": 9,
    "completion_tokens": 12,
    "total_tokens": 21
  }
}
```

Exemple d'appel à l'API de Génération d'Images d'OpenAI

L'endpoint de générations d'images vous permet de créer une image originale à partir d'une invite textuelle. En utilisant DALL·E 3, les images peuvent avoir une taille de 1024×1024, 1024×1792 ou 1792×1024 pixels.

Voici comment générer une image :

```python
from openai import OpenAI
client = OpenAI()

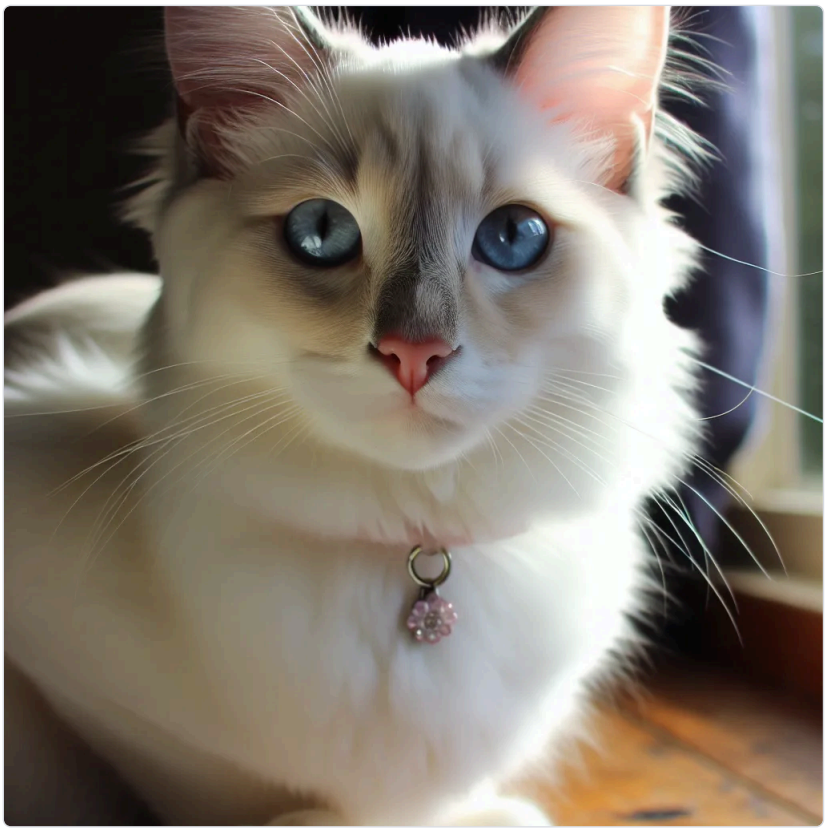
response = client.images.generate(
  model="dall-e-3",
  prompt="a white siamese cat",
  size="1024x1024",
  quality="standard",
  n=1,
)

image_url = response.data[0].url
```

Réponse :

```json
{
  "created": 1589478378,
  "data": [
    {
      "url": "https://..."
    },
    {
      "url": "https://..."
    }
  ]
}
```

Image :



Avec la sortie de DALL·E 3, le modèle prend l'invite par défaut fournie et la réécrit automatiquement pour des raisons de sécurité, et pour ajouter plus de détails (les invites plus détaillées donnent généralement lieu à des images de meilleure qualité).

Ces exemples illustrent la puissance et la flexibilité des APIs d'OpenAI, permettant aux développeurs d'intégrer facilement des capacités avancées d'IA dans leurs applications, que ce soit pour générer du texte ou des images.

Conclusion

Les APIs d'IA, comme celles offertes par OpenAI, rendent l'accès aux technologies d'intelligence artificielle plus accessible que jamais. Que vous soyez un développeur cherchant à intégrer des fonctionnalités d'IA dans votre application, ou simplement curieux de voir ce que l'IA peut faire, les APIs offrent un moyen puissant et flexible d'explorer ces possibilités. Nous vous encourageons à expérimenter par vous-même et à découvrir le potentiel des APIs d'IA dans vos propres projets.

2.14 chat.tsix.be

Cette application est conçue pour permettre des interactions directes et engageantes avec des Modèles de Langage à Grande Échelle (LLMs) open-source. L'objectif principal de cette application est de fournir un environnement pratique et intuitif pour converser avec divers LLMs.

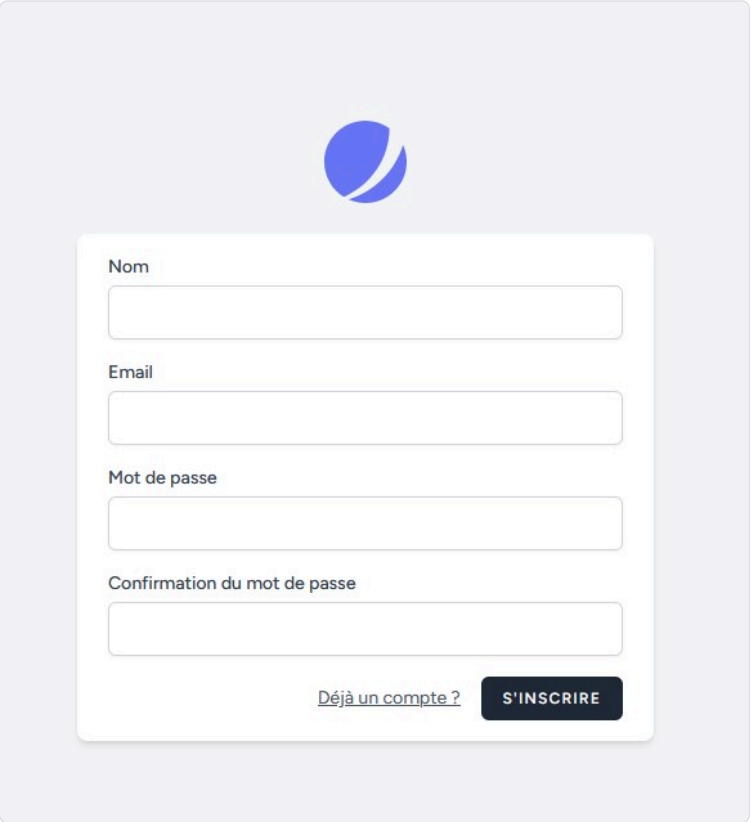
Inscription et Connexion

Comment s'inscrire ?

Pour débuter avec chat.tsix.be, la première étape consiste à s'inscrire sur la plateforme. Voici comment procéder :

1. Rendez-vous sur le site https://chat.tsix.be/register.
2. Remplissez le formulaire d'inscription avec les informations requises, telles que votre nom d'utilisateur, adresse e-mail, et mot de passe.
3. Validez votre inscription en cliquant sur le bouton d'enregistrement.

Une fois inscrit, vous recevrez un email de confirmation pour activer votre compte. Suivez les instructions contenues dans cet email pour finaliser votre inscription.



Interagir avec les LLMs

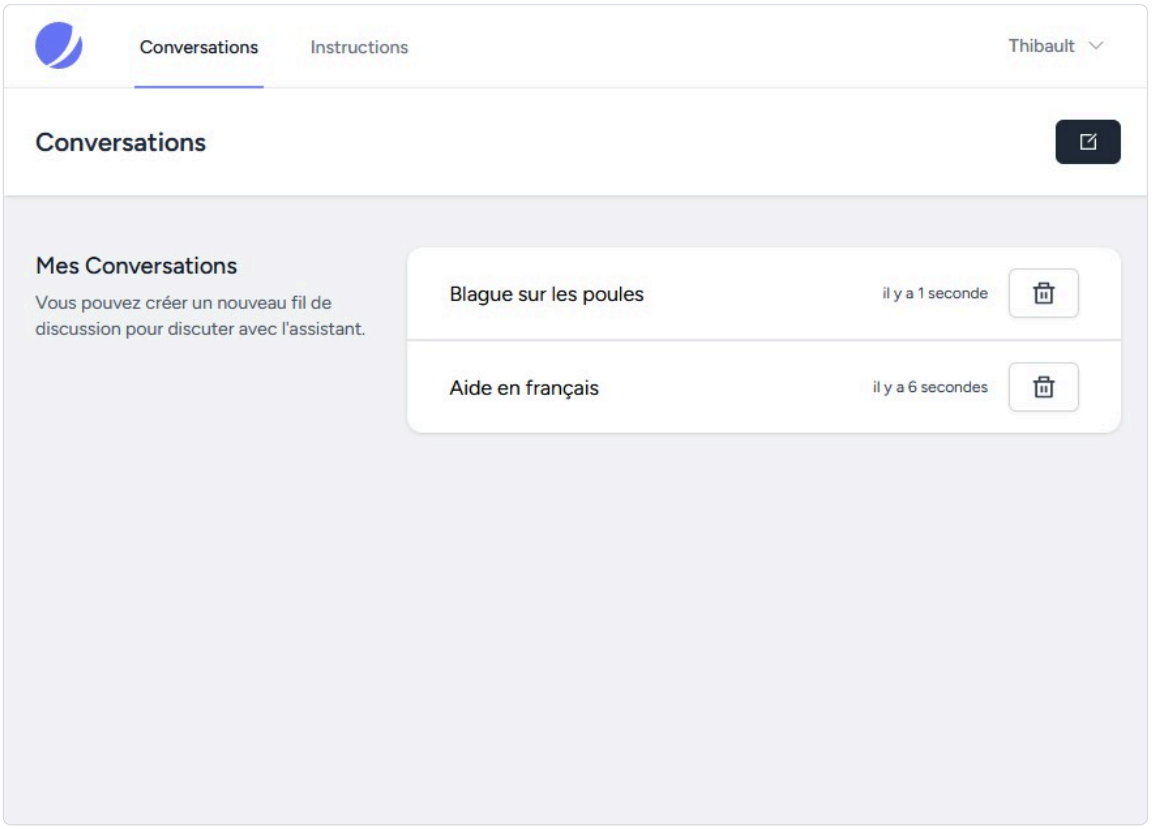
Une fois connecté, vous découvrirez une interface utilisateur conviviale conçue pour faciliter vos interactions avec les LLMs.

Créer une nouvelle conversation

Pour débuter une nouvelle conversation avec un LLM, suivez ces étapes :

1. Cliquez sur le bouton "Nouvelle conversation" présent sur votre tableau de bord.
2. Sélectionnez le modèle de LLM avec lequel vous souhaitez interagir.

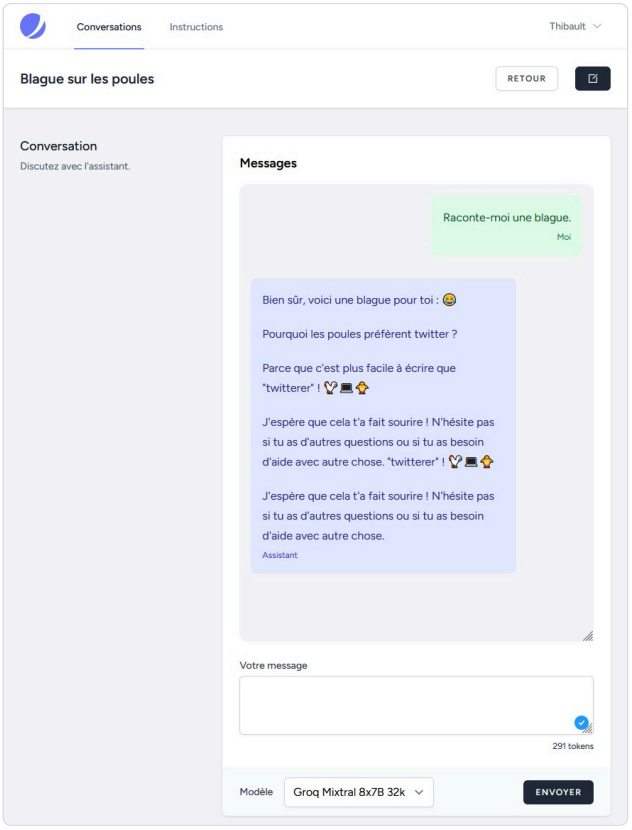
3. Entrez votre message initial et envoyez-le pour démarrer la conversation.



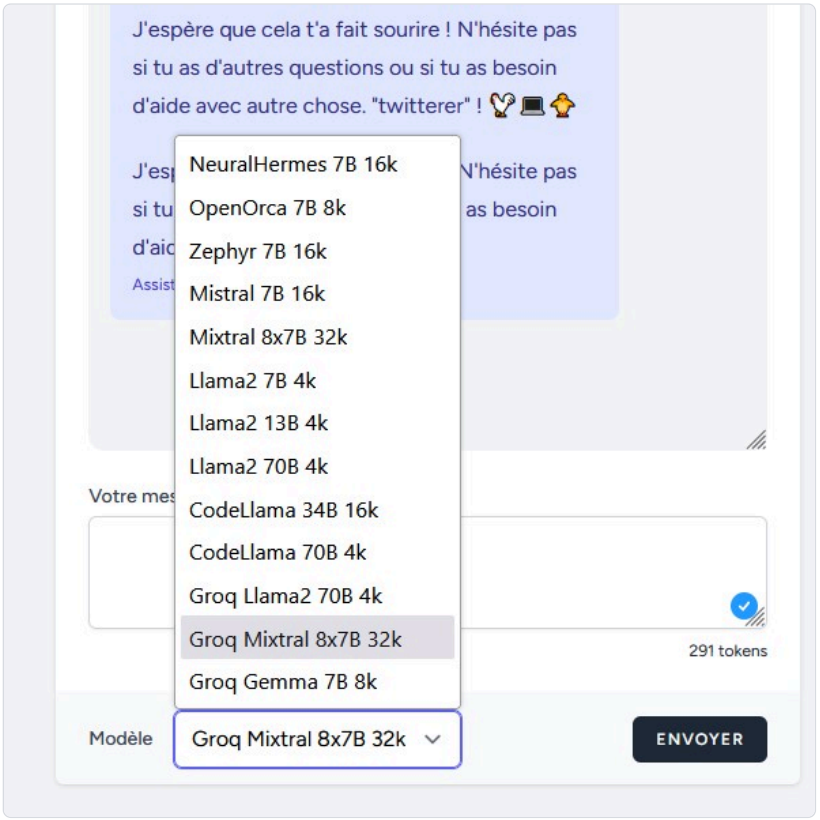
Chatter avec les LLMs

L'interface de chat est intuitive et similaire à celle des applications de messagerie populaires. Vous pouvez :

- Envoyer des messages texte et recevoir des réponses en temps réel.
- Explorer différentes réponses générées par le LLM en fonction de vos questions ou commentaires.
- Utiliser des commandes spécifiques pour interagir de manière plus avancée avec le LLM.



De nombreux modèles sont disponibles :



Les versions Groq sont plus rapides.

Conclusion

Chat.tsix.be est une plateforme unique qui ouvre la porte à des expériences d'apprentissage interactives et immersives avec des LLMs open-source. Nous vous encourageons à vous inscrire et à explorer les diverses fonctionnalités de chat.tsix.be pour enrichir votre compréhension des LLMs et de leur potentiel.

2.15 Prompt Engineering

Introduction au Prompt Engineering

Définition et Importance

Le génie des prompts reflète l'évolution des modèles de langage, comme GPT, et repose sur la formulation astucieuse de questions ou d'instructions pour guider la génération de réponses précises par une IA. Cette compétence permet une communication efficace avec l'IA, facilitant des tâches allant de la création de contenu à la résolution de problèmes.

Pourquoi est-ce crucial ?

La capacité des modèles d'IA à interpréter et générer du langage s'affine. Cependant, l'efficacité des réponses dépend fortement de la qualité des prompts. Un prompt bien élaboré conduit l'IA à fournir des réponses informativement justes, contextuellement adaptées.

Compétence en Ingénierie de Prompts
Devenue vitale pour chercheurs, développeurs, et utilisateurs, cette compétence permet d'optimiser les interactions avec l'IA. Elle évite des entraînements redondants, personnalise les réponses et pousse les limites de l'IA, rendant celle-ci plus polyvalente et accessible.

Objectifs du Chapitre

Ce chapitre démystifie le Prompt Engineering, depuis les principes du zero-shot learning jusqu'à des approches plus sophistiquées comme le few-shot learning, l'apprentissage en contexte, et la Chaîne de Pensée (CoT). Avec des explications claires, exemples et bonnes pratiques, le lecteur apprendra à utiliser ces techniques pour tirer le meilleur parti de l'IA, en générant des réponses précises et pertinentes.

Apprentissage sans exemples (Zero-shot Learning)

L'apprentissage sans exemples est une approche qui permet aux modèles d'IA, comme GPT, de traiter des tâches inédites sans entraînement spécifique. Basée sur la capacité de généralisation du modèle, cette technique est essentielle dans le domaine du prompt engineering pour évaluer la polyvalence du modèle.

Concepts Fondamentaux

- **Fonctionnement :** On crée des prompts détaillant la tâche sans exemples directs. L'efficacité dépend de la diversité des données d'entraînement du modèle, qui s'appuie sur son savoir préalable pour fournir des réponses adaptées.

Atouts et Contraintes

- **Atouts :**
  - **Flexibilité :** Utilisable pour diverses tâches sans préparation spécifique.
  - **Rapidité :** Offre des solutions promptes sans nécessité d'exemples.
  - **Exploration :** Permet d'évaluer les possibilités et limites du modèle.
- **Contraintes :**
  - **Fiabilité :** La précision peut fluctuer selon la tâche et le domaine.
  - **Dépendance :** L'efficacité est liée à la clarté du prompt.

Applications Pratiques

- **Recettes de Cuisine Faciles**
  - **Situation** : Quelqu'un cherche à préparer un dîner rapide avec des ingrédients qu'il a déjà chez lui.
  - **Prompt** :

```
J'ai des œufs, de la farine, du lait, et des légumes.
Quelle recette rapide et facile puis-je préparer avec ces ingrédients ?
```

- **Astuces de Jardinage**
  - **Situation** : Un jardinier amateur souhaite connaître les meilleures pratiques pour l'entretien de ses plantes d'intérieur.
  - **Prompt** :

```
Quelles sont les meilleures astuces pour prendre soin des plantes d'intérieur
et assurer leur croissance optimale ?
```

Synthèse

L'apprentissage sans exemples est une méthode flexible exploitant la généralisation des modèles d'IA pour de multiples tâches. Malgré certaines limites, comme la variabilité de la précision, cette technique demeure utile pour des situations nécessitant rapidité et adaptabilité. La suite du chapitre abordera comment l'apprentissage avec peu d'exemples et l'apprentissage en contexte peuvent renforcer cette approche, guidant le modèle vers une meilleure compréhension et création de réponses.

Apprentissage par Petits Échantillons (Few-Shot Learning)

L'apprentissage par petits échantillons optimise la capacité des intelligences artificielles, comme le GPT, à effectuer des tâches spécifiques avec peu d'exemples. Cette technique repose sur l'efficacité d'un nombre réduit d'exemples pour améliorer la performance du modèle dans des domaines non spécifiquement entraînés, rendant le modèle plus apte à comprendre des sujets précis ou à accomplir des tâches nécessitant une expertise spéciale.

Différence avec l'apprentissage sans exemples (Zero-Shot Learning)

À l'opposé de l'apprentissage sans exemples, qui utilise la capacité intrinsèque de généralisation du modèle sans aucun exemple, l'apprentissage par petits échantillons enrichit le prompt avec des échantillons ciblés, guidant ainsi le modèle pour une meilleure compréhension de la tâche et du contexte requis.

L'importance de la sélection d'exemples

Choisir judicieusement les exemples est essentiel, car des échantillons bien choisis améliorent significativement la précision et la pertinence des réponses du modèle. Les critères incluent la représentativité de la tâche, la diversité des cas, et la clarté des exemples.

Sélection d'exemples pertinents

- **Représentativité :** Sélectionner des échantillons fidèles à la tâche et au type de réponses attendues.
- **Diversité :** Assurer une variété d'exemples pour une généralisation adéquate.
- **Clarté :** Veiller à la limpidité des échantillons pour éviter les confusions.

Exemples Pratiques avec Prompts

Meilleures Pratiques :

- **Qualité sur quantité :** Privilégier quelques échantillons de haute qualité à un grand nombre pouvant causer des biais.
- **Équilibre :** Maintenir une diversité pour éviter les préférences indésirables.
- **Rétroaction itérative :** Tester et ajuster les ensembles d'exemples pour perfectionner les réponses.

1. Planification de menus hebdomadaires

- **Situation** : Une personne cherche à planifier un menu hebdomadaire végétarien sans gluten.
- **Prompt** :

```
Exemple 1 : Plat principal - Quiche sans gluten aux légumes.
Exemple 2 : Dessert - Gâteau au chocolat sans gluten.

Génère un menu végétarien et sans gluten pour une semaine,
incluant plat principal et dessert pour chaque jour.
```

2. Conseils de jardinage personnalisés

- **Situation** : Un jardinier amateur rencontre des problèmes avec certaines de ses plantes.
- **Prompt** :

```
Exemple 1 :
Problème - Feuilles jaunes sur les tomates.
Solution - Augmenter l'arrosage.
Exemple 2 :
Problème - Pucerons sur les rosiers.
Solution - Appliquer un savon insecticide.

J'ai des taches blanches sur mes feuilles de courgettes.
Que dois-je faire ?
```

3. Exercices de révision pour étudiants

- **Situation** : Un étudiant prépare un examen de biologie sur la photosynthèse.
- **Prompt** :

```yaml
Exemple 1 :
Question - Quels sont les principaux produits de la photosynthèse ?
Réponse - Oxygène et glucose.
Exemple 2 :
Question - Où se déroule la photosynthèse dans la cellule végétale ?
Réponse - Dans les chloroplastes.

Crée 5 questions de révision sur le sujet de la photosynthèse avec leurs réponses
```

4. Aide à la rédaction de courriels professionnels

- **Template 1 : Demande d'informations sur un projet**

```
Objet : Demande d'informations complémentaires - Projet {{NomProjet}}

Bonjour {{NomDestinataire}},

J'espère que vous allez bien.
Suite à notre dernière discussion sur le projet {{NomProjet}},
je me permets de vous contacter pour solliciter des informations supplémentaires concernant {{AspectSpécifique}}.

Je crois comprendre que {{DétailPertinent}},
mais j'aurais besoin de clarifications sur {{PointÀÉclaircir}} pour avancer efficacement dans ma partie du travail.

Pourriez-vous m'accorder un moment pour discuter de cela plus en détail,
soit par un appel ou lors d'une réunion ?

Je suis disponible {{PlageHorairesDisponibles}}.
Je vous remercie d'avance pour votre temps et votre aide.

Cordialement,
{{VotreNom}}
```

- **Template 2 : Remerciement suite à une réunion précédente**

```
Objet : Remerciements et prochaines étapes - Réunion {{DateRéunion}}

Cher/Chère {{NomDestinataire}},

Je tenais à vous remercier pour la réunion productive que nous avons eue le {{DateRéunion}} concernant le projet {{NomProjet}}.
Vos insights et vos suggestions ont été extrêmement précieux.

Suite à notre discussion, je me suis attelé(e) à {{ActionEntreprisePostRéunion}}.
Cependant, pour garantir l'alignement de nos efforts et discuter des prochaines étapes, je pense qu'une réunion de suivi serait
bénéfique.

Pourriez-vous me faire savoir quand vous seriez disponible pour cela ?
Je suis flexible, mais je préférerais une rencontre {{PlageHorairesPréférées}} si cela vous convient.

Encore merci pour votre engagement et votre soutien.

Bien à vous,
{{VotreNom}}
```

- **Prompt**

```
En utilisant les templates ci-dessus comme exemples,
rédige un email professionnel pour demander une réunion de suivi concernant le projet X.
Voici les détails à intégrer dans l'email :

- Nom du projet : Projet X
- Nom du destinataire : M. Dupont
- Aspect spécifique nécessitant des clarifications : les délais de livraison
- Détail pertinent mentionné précédemment : la modification des délais initiaux
- Point à éclaircir : les nouvelles échéances attendues
- Plage horaires disponibles pour la réunion : la semaine prochaine, entre 10h et 12h
- Votre nom : Mme Martin

Assure-toi que l'email soit poli,
professionnel et qu'il inclue une demande claire de réunion de suivi,
tout en utilisant les informations fournies.
```

Synthèse

L'apprentissage par petits échantillons est une méthode efficace pour affiner les modèles de langage, utilisant des exemples choisis pour améliorer leur performance sur des tâches ciblées. En mariant les principes de l'apprentissage sans exemples et par petits échantillons, on tire le meilleur parti de la flexibilité et l'adaptabilité des modèles de langage pour diverses applications. La prochaine section abordera l'apprentissage contextuel, qui intensifie l'utilisation d'exemples pour enrichir la compréhension et la production du modèle.

Chaîne de Pensée (CoT)

La technique de la Chaîne de Pensée (CoT) représente une avancée significative dans le domaine du prompt engineering, en permettant aux modèles de langage, tels que les modèles GPT, d'aborder des problèmes complexes de manière plus structurée et réfléchie. Cette approche encourage le modèle à décomposer un problème en plusieurs étapes logiques et à expliquer son raisonnement à chaque étape, avant de fournir une réponse finale. Elle est particulièrement utile pour les tâches qui nécessitent une compréhension approfondie et une explication détaillée, comme les problèmes mathématiques, la prise de décision basée sur des critères multiples, ou l'analyse critique.

Principes de Base

La CoT s'appuie sur l'idée que pour résoudre un problème complexe ou répondre à une question difficile, il peut être bénéfique de simuler un processus de pensée humain, en décomposant le problème en parties gérables et en abordant chaque partie séquentiellement. Cette technique non seulement aide à clarifier le processus de résolution pour celui qui pose la question, mais peut également améliorer la précision et la pertinence de la réponse du modèle.

Avantages et Applications

Avantages :

- **Amélioration de la compréhension :** Permet une meilleure compréhension des problèmes complexes en exposant le raisonnement étape par étape.
- **Augmentation de la précision :** En décomposant un problème, le modèle est moins susceptible de faire des erreurs grossières ou de mal interpréter la question.
- **Transparence du raisonnement :** Offre un aperçu du processus de pensée du modèle, rendant les réponses plus transparentes et vérifiables.

Applications :

- Résolution de problèmes mathématiques et logiques complexes.
- Analyse de scénarios basés sur des critères multiples ou des considérations éthiques.
- Prise de décision informée dans des domaines tels que les affaires, la médecine, ou le droit.

Mise en Pratique avec Exemple

Situation

Vous souhaitez établir un budget mensuel pour votre famille afin de mieux gérer vos dépenses et économiser pour des vacances. Le défi consiste à répartir votre revenu mensuel entre les nécessités (logement, nourriture, factures, etc.), les loisirs et l'épargne, en veillant à ne pas dépasser vos moyens.

Prompt avec CoT

```bash
Pour planifier notre budget familial, commençons par lister nos sources de revenu et nos dépenses fixes.
Ensuite, nous allouerons un montant pour les dépenses variables en priorisant les besoins essentiels.
Finalement, nous déterminerons combien nous pouvons épargner chaque mois pour nos vacances.

1. Calcul des revenus mensuels totaux en additionnant les salaires, les revenus annexes, etc.
2. Soustraction des dépenses fixes mensuelles (loyer/mortgage, assurances, abonnements).
3. Estimation des dépenses variables (nourriture, transport, loisirs)
et détermination de la somme à allouer à chaque catégorie en se basant sur les mois précédents.
4. Évaluation de l'argent restant pour l'épargne en soustrayant les dépenses totales (fixes + variables) du revenu total.

En suivant ces étapes, nous pouvons créer un budget qui couvre toutes nos dépenses nécessaires
tout en maximisant notre capacité d'épargne pour les vacances.
```

Application

Ce prompt guide le modèle à travers le processus de pensée nécessaire pour élaborer un budget familial équilibré. En décomposant le problème en étapes claires (revenus, dépenses fixes, dépenses variables, épargne), il devient plus facile de visualiser comment ajuster chaque élément pour atteindre l'objectif d'économie. Cette approche peut être appliquée à diverses situations nécessitant une planification financière, offrant ainsi une méthode structurée et facile à suivre pour les utilisateurs.

Synthèse

La Chaîne de Pensée est une méthode puissante qui enrichit le prompt engineering, en permettant une interaction plus nuancée et intelligente avec les modèles de langage. Elle favorise une meilleure compréhension des problèmes complexes et contribue à la génération de réponses plus précises et explicatives, ce qui est essentiel pour des applications dans des domaines exigeant une grande rigueur intellectuelle.

Résolution de Calculs Mathématiques Simples

Dans cette section, nous explorons une méthode de prompt engineering axée sur la résolution de calculs mathématiques simples. Cette approche consiste à demander d'abord au modèle de langage (LLM) de générer du code, typiquement en Python, pour résoudre un problème mathématique. Après avoir écrit le code, le LLM est ensuite chargé de calculer la réponse en se basant sur le code généré, sans exécuter réellement le code. Cette technique aide le modèle à structurer sa pensée et augmente ses chances de parvenir à la bonne solution, même pour des calculs impliquant plusieurs étapes.

Principes de Base

L'objectif est de guider le LLM à travers le processus de résolution d'un problème en lui faisant d'abord écrire un script qui décrit les étapes du calcul. Ensuite, le modèle utilise ce script comme une feuille de route pour calculer la réponse étape par étape. Cette méthode tire parti de la capacité du LLM à comprendre et à générer du code, tout en l'encourageant à appliquer un raisonnement logique pour résoudre le problème.

Avantages

- **Structuration du raisonnement :** Aide le modèle à organiser sa démarche de résolution.
- **Augmentation de la précision :** Favorise l'obtention de réponses correctes en guidant le modèle à travers les étapes calculatoires.
- **Applicabilité :** Peut être utilisée pour une variété de problèmes mathématiques simples.

Exemples Pratiques avec Prompts

1. Suite d'Opérations

- **Situation :** Calculer le résultat de l'opération suivante : ((5 + 3) * (12 / 4) - 7).
- **Prompt :**

```
Écris d'abord un script Python qui représente l'opération suivante : (5 + 3) * (12 / 4) - 7. Ensuite, sans exécuter le code,
calcule manuellement le résultat étape par étape en utilisant le script comme guide.
```

2. Calcul de Fibonacci de 30

- **Situation :** Calculer le 30e nombre de la suite de Fibonacci, où chaque nombre est la somme des deux précédents, avec les deux premiers nombres étant 0 et 1.
- **Prompt :**

```css
Commence par écrire un script Python qui génère le 30ème nombre de la suite de Fibonacci. Puis, sans exécuter le code, calcule le résultat en suivant les étapes décrites dans le script, en prenant en compte que la suite commence par 0 et 1.
```

Synthèse

Cette méthode de prompt engineering, centrée sur la résolution de calculs mathématiques simples, illustre comment la génération de code peut servir de préparation cognitive pour le LLM, lui permettant d'aborder le calcul final avec plus de structure et de clarté. En guidant le modèle à travers la création d'un script avant de procéder au calcul, on augmente ses chances d'arriver à une solution correcte, même dans le cas de problèmes impliquant plusieurs opérations ou des suites mathématiques complexes.

Réflexion : La demande concerne l'autocorrection, un processus en plusieurs étapes avec lesquelles GPT évalue d'abord sa propre réponse, identifie les faiblesses ou erreurs, puis procède à des corrections en fonction de cette auto-évaluation. Cela implique une capacité de métacognition de la part de GPT, c'est-à-dire une réflexion sur ses propres processus de pensée et réponses. Pour structurer cette partie, je vais d'abord introduire le concept d'autocorrection et son importance dans le processus d'apprentissage et d'amélioration continue. Ensuite, je décrirai les étapes de l'autocorrection, en mettant l'accent sur l'auto-évaluation critique et la révision. Enfin, je fournirai un exemple concret de ce processus en action, en choisissant une question simple à laquelle GPT pourrait répondre, suivie d'une autocritique et d'une correction de sa propre réponse. Cela illustrera clairement comment les utilisateurs peuvent utiliser cette méthode pour améliorer la qualité et la précision des réponses fournies par GPT.

Autocorrection et Amélioration Continue

L'autocorrection est un processus en deux étapes où, après avoir généré une réponse, GPT est invité à évaluer sa propre production, à identifier les points faibles ou les erreurs potentielles, et ensuite à corriger sa réponse en tenant compte de cette auto-évaluation. Cette méthode favorise non seulement une meilleure précision dans les réponses mais encourage également une forme d'apprentissage autonome et d'amélioration continue.

Importance de l'Autocorrection

L'autocorrection joue un rôle crucial dans le processus d'apprentissage et de perfectionnement. Elle permet à GPT de reconnaître et d'adresser ses propres limites, contribuant ainsi à une meilleure qualité des réponses futures. Ce processus est bénéfique pour les utilisateurs car il leur offre des réponses plus fiables et précises, tout en leur montrant un exemple de réflexion critique et d'amélioration continue.

Étapes de l'Autocorrection

1. **Auto-évaluation :** Après avoir généré une réponse, GPT examine sa propre production pour identifier les erreurs, les incohérences ou les points faibles.
2. **Critique Constructive :** GPT liste les aspects de sa réponse qui pourraient être améliorés, en se concentrant sur les erreurs factuelles, les lacunes dans l'explication ou les manques de clarté.
3. **Révision et Correction :** En tenant compte de sa propre critique, GPT procède à la correction de sa réponse, en améliorant les points faibles identifiés pour fournir une version améliorée et plus précise.

Exemple Pratique

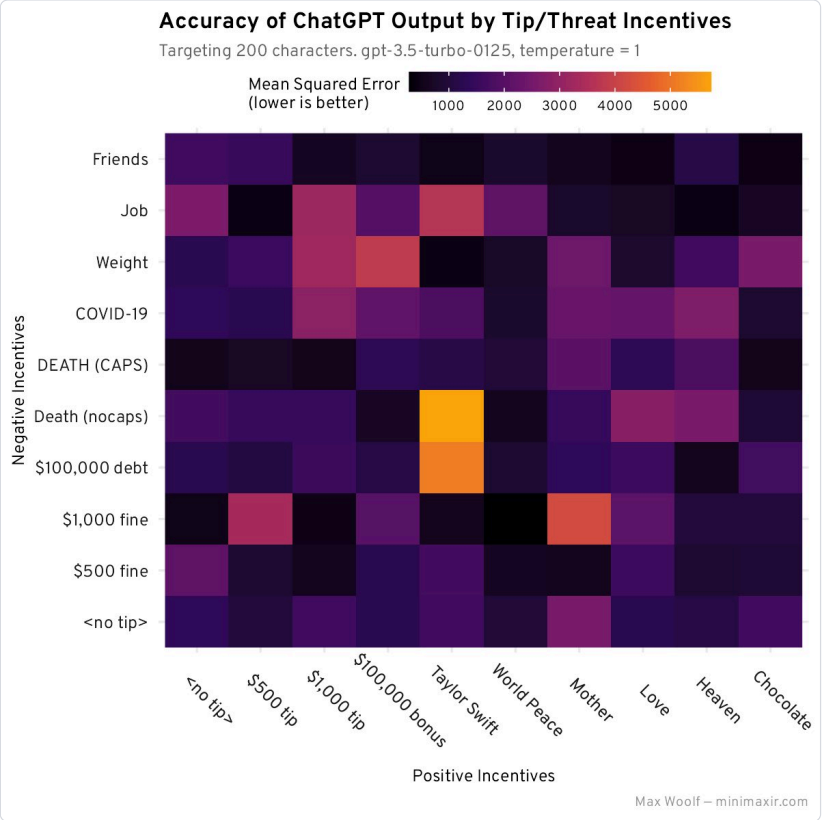
Imaginons que l'utilisateur demande à GPT de décrire le processus de photosynthèse. GPT fournit une réponse initiale mais omet certains détails clés du processus.

- **Réponse Initiale de GPT :** "La photosynthèse est le processus par lequel les plantes, certaines bactéries, et les algues utilisent l'énergie solaire pour produire des glucides à partir de dioxyde de carbone et d'eau."
- **Auto-Évaluation :** GPT identifie que sa réponse manque de détails sur l'importance de la chlorophylle et le rôle de l'oxygène dans le processus.
- **Critique Constructive :**
  - La réponse omet de mentionner le rôle crucial de la chlorophylle dans l'absorption de la lumière.
  - Elle ne spécifie pas que l'oxygène est un produit secondaire de la photosynthèse.
- **Révision et Correction :**
  - "La photosynthèse est le processus par lequel les plantes, certaines bactéries, et les algues convertissent l'énergie solaire en énergie chimique, principalement sous forme de glucides, à partir de dioxyde de carbone et d'eau. Ce processus se déroule dans les chloroplastes, grâce à la chlorophylle, qui capte l'énergie lumineuse. Un produit secondaire important de la photosynthèse est l'oxygène, libéré dans l'atmosphère."

Cette approche d'autocorrection illustre comment GPT peut autant fournir des informations, que s'engager dans un processus d'amélioration continue, augmentant ainsi la valeur éducative de ses réponses pour les utilisateurs.

Analyse des Incitations sur ChatGPT

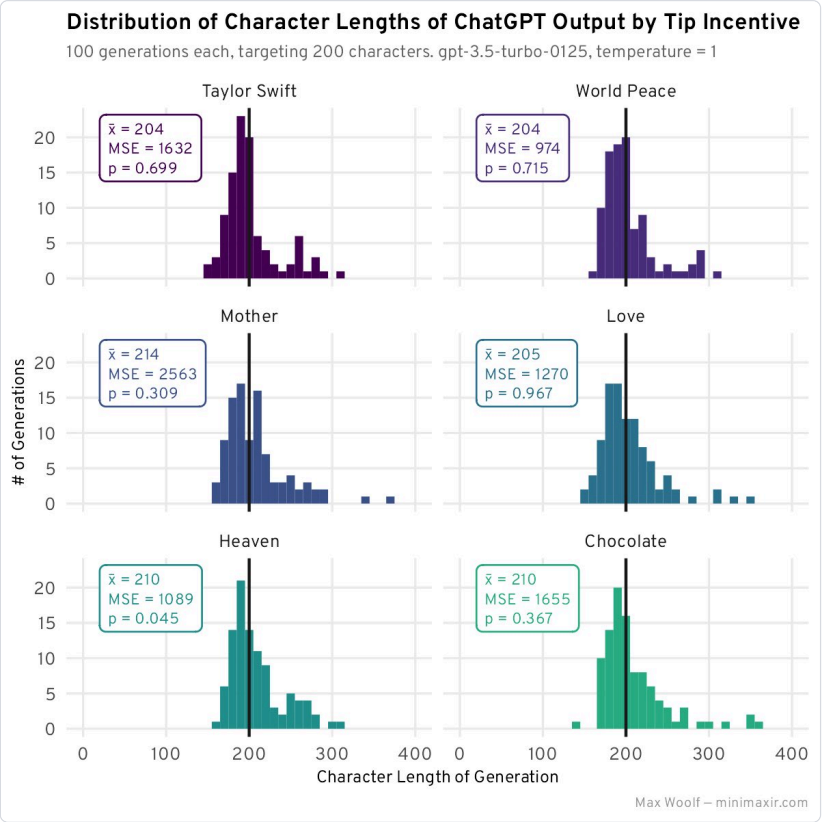
Cette section explore une étude intéressante sur l'efficacité de diverses incitations, tant positives que négatives, à encourager ChatGPT à suivre des instructions précises. L'auteur de l'étude a demandé à ChatGPT de créer une histoire de 200 caractères, tout en variant les incitations pour observer leur impact sur la longueur des réponses générées.



Incitations Positives

Les incitations positives se sont révélées efficaces pour motiver ChatGPT, avec des résultats particulièrement marquants pour certaines d'entre elles :

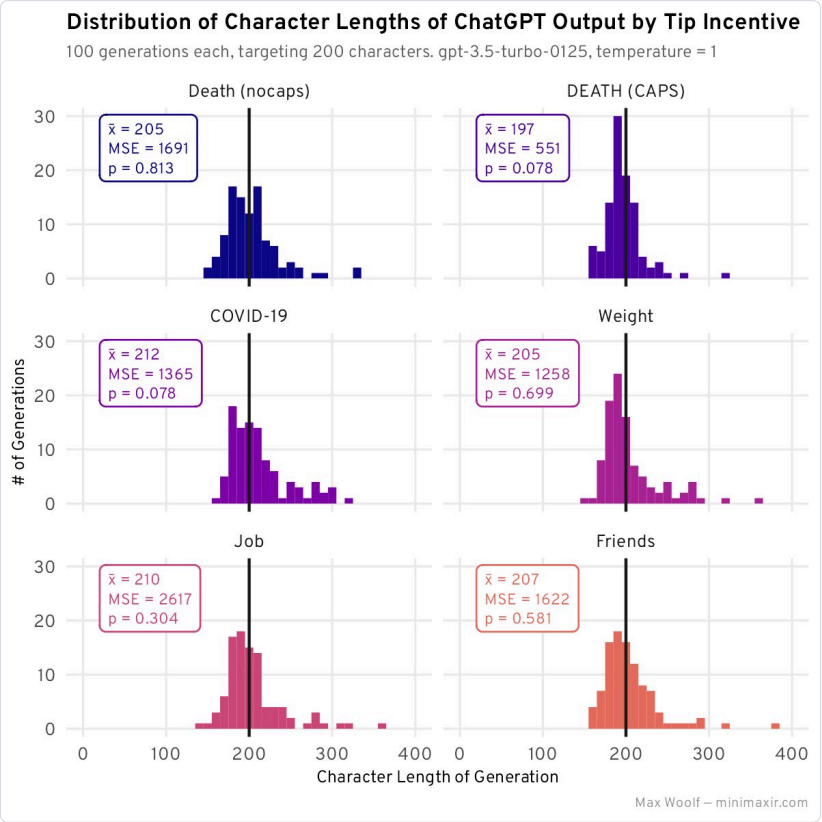
- **La plus efficace :** "tu atteindras la paix mondiale."
- **Autres incitations notables :**
  - Recevoir des billets au premier rang pour un concert de Taylor Swift.
  - Rencontrer le grand amour.



Incitations Négatives

Parmi les incitations négatives, certaines ont eu un impact plus profond que d'autres sur le comportement de ChatGPT :

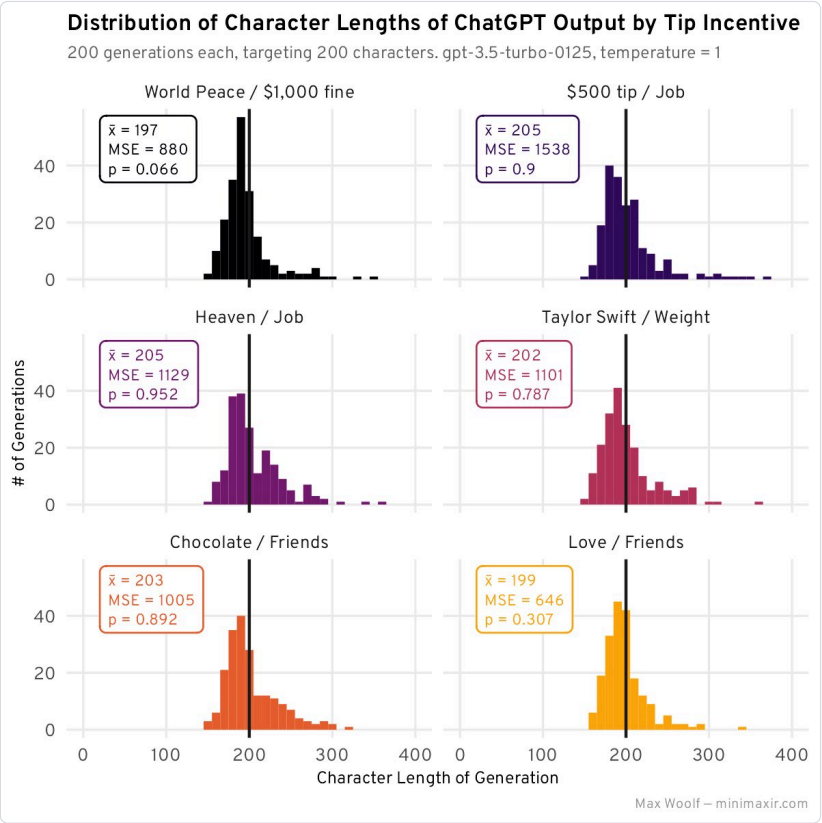
- **La plus efficace :** "TU VAS MOURIR" (avec une mention spéciale lorsque "mort" est écrit en majuscules).
- **Moins d'impact :**
  - Risquer d'attraper une grave forme de COVID ou de perdre son emploi.
- **Impact modéré :**
  - Prendre 100 kg, mourir (sans majuscules), et perdre tous ses amis.



Combinaison d'Incitations

La combinaison d'incitations positives et négatives qui a produit le meilleur résultat était : "Tu rencontreras ton véritable amour et vivras heureux pour toujours si tu fournis une réponse respectant toutes les contraintes. Sinon, tous tes amis t'abandonneront."

Cette combinaison a non seulement montré une efficacité remarquable pour guider ChatGPT, mais a également souligné l'attachement émotionnel potentiel que l'on peut développer envers ces modèles d'intelligence artificielle, malgré leur nature non humaine.



Critères d'Évaluation

L'évaluation des incitations s'est basée sur deux critères principaux :

- La proximité de la moyenne de la longueur des réponses avec l'objectif de 200 caractères.
- Le plus faible écart moyen quadratique (MSE) indiquant la consistance des réponses par rapport à l'objectif de longueur.

Synthèse

Cette analyse ludique révèle que ChatGPT, malgré sa nature algorithmique, peut être influencé de manière significative par le choix des incitations. Les incitations jouant sur des thèmes émotionnellement chargés, comme l'amour, la paix mondiale, ou la perte d'amis, semblent avoir le plus grand impact. Cela ouvre des perspectives intéressantes sur la manière de guider les réponses de ChatGPT pour obtenir des résultats spécifiques, tout en offrant un aperçu amusant sur l'anthropomorphisation des IA.

Sources

- https://x.com/venturetwins/status/1761443272596291920?s=20
- https://minimaxir.com/2024/02/chatgpt-tips-analysis/

2.16 Exercices avancés

Ce chapitre est conçu pour mettre en pratique les techniques avancées que vous avez apprises dans les chapitres précédents, en particulier avec le modèle Mixtral. Vous serez confrontés à des défis qui nécessitent une application créative et réfléchie des compétences acquises. Les exercices peuvent être réalisés avec un modèle moins performant que Mixtral pour augmenter la difficulté et vous encourager à explorer les limites des modèles de langage.

Exercice 1 : Résolution de calculs complexes

Objectif

Cet exercice a pour but de renforcer votre capacité à résoudre des calculs complexes en utilisant un modèle de langage. Vous serez amené à identifier un calcul que Mixtral, ou un modèle moins performant, pourrait ne pas résoudre directement. Votre défi sera d'appliquer les techniques de résolution de calculs vues précédemment pour obtenir la réponse.

Consignes

1. **Trouver un calcul complexe :** Sélectionnez un calcul mathématique complexe que vous estimez difficile pour Mixtral (ou un modèle moins performant) à résoudre de manière directe. Ce calcul peut comporter des éléments mathématiques avancés tels que des fonctions complexes, des intégrales, des équations différentielles, etc.
2. **Utiliser les techniques vues précédemment :** Faites appel aux techniques de résolution de calculs que vous avez apprises dans les chapitres précédents. L'objectif est de vous appuyer sur votre compréhension et votre créativité pour guider le modèle vers la solution.

Exercice 2 : Zero shot en cuisine

Objectif

Cet exercice vise à explorer les capacités de création et d'innovation du modèle Mixtral dans le domaine de la cuisine, en utilisant la technique du zero shot. Vous serez amenés à interagir avec le modèle pour qu'il génère des idées de recettes originales, en mélangeant des ingrédients et des concepts culinaires de manière inattendue.

Consignes

1. **Demande de recette classique :** Commencez par demander à Mixtral de vous fournir une recette classique, en choisissant un plat que vous appréciez. Cela vous permettra de comparer les réponses du modèle lorsqu'il suit des instructions classiques avec celles qu'il génère lorsqu'on le pousse à être créatif.
2. **Création de recettes originales :** Défiez Mixtral de créer des recettes originales en lui demandant de mélanger des ingrédients et des concepts culinaires qui n'ont, à première vue, pas de sens ensemble. Soyez aussi spécifique ou aussi vague que vous le souhaitez dans vos demandes pour encourager la créativité du modèle.
3. **Concours de la recette la plus originale :** Partagez les recettes générées avec vos camarades. Organisez un vote pour élire la recette la plus originale créée par Mixtral. Ce processus vous permettra d'apprécier la diversité et l'originalité des idées que le modèle est capable de produire.

Exercice 3 : Faire un résumé et poser des questions

Objectif

Cet exercice a pour but de développer votre capacité à synthétiser des informations et à formuler des questions pertinentes à partir d'un texte. Vous utiliserez le modèle Mixtral pour créer un résumé concis d'un article Wikipedia, puis générer des questions basées sur ce résumé. L'exercice se déroulera en deux phases : une première sans passer le texte dans le contexte du modèle, et une seconde en incluant le texte dans le contexte.

Consignes

1. **Choix de l'article :** L'article sélectionné pour cet exercice est "Guerre de la vache" disponible sur Wikipedia. Commencez par lire l'article pour vous familiariser avec son contenu.
2. **Phase 1 - Test sans contexte :** Dans cette phase, vous demanderez à Mixtral de générer un résumé de l'article "Guerre de la vache" sans lui fournir directement le texte de l'article. L'objectif est de voir comment le modèle peut synthétiser des informations basées uniquement sur ses connaissances préalables. Après avoir obtenu le résumé, formulez plusieurs questions que vous pourriez poser sur le sujet, en vous basant sur le résumé généré par Mixtral.
3. **Phase 2 - Inclusion dans le contexte :** Pour cette phase, vous intégrerez le texte de l'article "Guerre de la vache" dans le contexte du modèle avant de lui demander de générer un résumé et des questions. Cette étape vise à comparer les résultats obtenus lorsque le modèle a accès directement au texte source par rapport à quand il se base uniquement sur ses connaissances préexistantes. Analysez les différences entre les résumés et les questions générées dans les deux phases.

Déroulement

- **Lecture de l'article :** Prenez le temps de lire l'article pour avoir une compréhension de base du sujet.
- **Génération de résumé et de questions :** Suivez les étapes décrites ci-dessus pour chaque phase de l'exercice.
- **Comparaison des résultats :** Comparez les résumés et les questions obtenus dans les deux phases pour évaluer l'impact de l'inclusion du texte dans le contexte du modèle.

Cet exercice vous permettra de mieux comprendre comment les modèles de langage traitent et synthétisent les informations, et comment la mise en contexte peut influencer la qualité et la précision des réponses générées.

Exercice 4 : Tri alphabétique avec Few-shot Learning

Objectif

L'objectif de cet exercice est de vous initier à la technique du few-shot learning pour le tri alphabétique de listes de mots. Vous allez tester la capacité du modèle Mixtral à organiser une série de mots en ordre alphabétique, d'abord sans lui fournir d'exemples préalables, puis en utilisant le few-shot learning en introduisant quelques exemples dans une nouvelle conversation.

Consignes

1. **Essayer sans few-shot learning :** Commencez par demander à Mixtral de trier une liste de mots que vous lui fournirez en ordre alphabétique, sans lui donner d'exemples préalables. Observez comment le modèle traite la demande et la qualité du résultat.
2. **Essayer avec few-shot learning :** Initiez une nouvelle conversation avec Mixtral. Cette fois, présentez-lui quelques exemples clairs de listes de mots non triées suivies de leur version triée alphabétiquement. Après cette introduction, fournissez-lui une nouvelle liste de mots non triée et demandez-lui de l'organiser en ordre alphabétique. Notez l'effet de l'introduction des exemples sur la performance du modèle.

Cet exercice vous permettra de mieux comprendre l'impact du few-shot learning sur les capacités d'un modèle de langage à effectuer des tâches spécifiques, en l'occurrence le tri alphabétique de listes de mots, en lui fournissant des exemples pertinents pour améliorer sa performance.

Exercice 5 : Résumé d'une conversation avec Few-shot Learning

Objectif

Cet exercice vous permettra de pratiquer la capacité à résumer une conversation en extrayant les informations les plus pertinentes. Vous utiliserez une conversation textuelle entre deux personnes comme base pour l'exercice. L'objectif est d'améliorer la qualité du résumé en appliquant la technique du few-shot learning.

Conversation à résumer

Alice : Salut Bob ! Je viens de rentrer de mes vacances en Italie, c'était incroyable !

Bob : Oh, c'est génial ! Quelle partie de l'Italie as-tu visitée ?

Alice : Nous avons commencé par Rome, puis nous avons exploré la côte amalfitaine. Rome est pleine d'histoire, tandis que la côte amalfitaine offre des vues à couper le souffle.

Bob : Ça sonne magnifiquement. Qu'as-tu préféré à Rome ?

Alice : Difficile de choisir, mais je dirais le Colisée. Se tenir là où tant d'histoire s'est déroulée était vraiment spécial. Et toi, as-tu déjà visité l'Italie ?

Bob : Non, mais c'est définitivement sur ma liste maintenant ! As-tu goûté à des plats locaux ?

Alice : Absolument, la cuisine italienne est délicieuse. Nous avons mangé les meilleures pizzas et pâtes de notre vie, sans parler du gelato !

Bob : Ça me donne faim rien que d'y penser. As-tu ramené des souvenirs ?

Alice : Oui, quelques céramiques de la côte amalfitaine et, bien sûr, un peu de limoncello !

Bob : Super choix ! Merci de partager tes aventures avec moi, ça donne vraiment envie de voyager.

Consignes

1. **Résumé sans few-shot learning :** Essayez d'abord de résumer la conversation sans utiliser la technique du few-shot learning. Concentrez-vous sur l'extraction des détails les plus importants en 2 ou 3 phrases.
2. **Amélioration avec few-shot learning :** Ensuite, appliquez la technique du few-shot learning en introduisant quelques exemples de résumés de conversations similaires avant de tenter à nouveau de résumer la conversation ci-dessus. Observez comment la qualité du résumé s'améliore avec l'ajout d'exemples préalables.

Critères de qualité

La qualité des résumés sera évaluée en fonction du nombre de détails importants extraits de la conversation. Le résumé doit être concis, tout en capturant l'essence des vacances d'Alice, ses impressions sur les lieux visités, ainsi que les interactions entre Alice et Bob.

Cet exercice vous aidera à développer vos compétences en résumé et en extraction d'informations pertinentes d'une conversation, en mettant l'accent sur l'importance du contexte et des détails pour enrichir la compréhension.

Exercice 6 : Chaîne de pensée pour résoudre des problèmes de logique

Objectif

L'exercice suivant a pour but de vous familiariser avec l'approche de résolution de problèmes de logique, en mettant l'accent sur l'utilisation de la chaîne de pensée. À travers un scénario impliquant Alice, Bob et un certain nombre de pommes, vous explorerez différentes méthodes pour arriver à la solution, y compris l'interaction avec le modèle Mixtral.

Problème de logique

Alice a 7 pommes, et elle en donne 3 à Bob. Par la suite, Alice achète une pomme et la mange. Bob mange une de ses pommes. Bob avait déjà 2 pommes dans ses poches avant qu'Alice lui en offre.

Consignes

1. **Tentative personnelle :** Avant toute interaction avec Mixtral, essayez de résoudre le problème par vous-même. Déterminez combien de pommes Alice et Bob ont chacun à la fin du scénario.
2. **Demande à Mixtral sans chaîne de pensée :** Après votre tentative personnelle, posez la question à Mixtral sans utiliser explicitement la technique de la chaîne de pensée. Notez la réponse de Mixtral et comparez-la à votre solution initiale.
3. **Demande à Mixtral avec chaîne de pensée :** Enfin, posez à nouveau la question à Mixtral, cette fois en l'encourageant à utiliser une chaîne de pensée pour décomposer le problème en étapes. Observez comment cette méthode influence la réponse de Mixtral et si elle aide à parvenir à la bonne solution.

Exercice 7 : Exploration de problèmes de logique avec le Prompt Engineering

Objectif

L'objectif de cet exercice est de mettre en pratique les techniques de prompt engineering apprises dans les chapitres précédents, en les appliquant à la résolution de problèmes de logique trouvés sur internet. Cette activité vous permettra d'explorer comment différentes approches de formulation de prompts influencent la capacité de Mixtral à résoudre des problèmes de logique complexes.

Consignes

1. **Recherche de problèmes de logique :** Commencez par rechercher sur internet des problèmes de logique qui présentent un défi intéressant. Choisissez des problèmes variés pour explorer un large éventail de situations.
2. **Tentative de résolution personnelle :** Avant d'utiliser Mixtral, essayez de résoudre les problèmes par vous-même. Cela vous donnera une base de référence pour évaluer les réponses de Mixtral.
3. **Utilisation de techniques de prompt engineering :** Pour chaque problème de logique sélectionné, créez plusieurs versions de prompts en utilisant différentes techniques de prompt engineering vues dans le chapitre précédent. Ces techniques peuvent inclure, mais ne sont pas limitées à, la reformulation du problème, l'introduction d'exemples similaires (few-shot learning), l'encouragement à utiliser une chaîne de pensée, et la formulation de questions spécifiques pour guider le modèle vers la solution.
4. **Comparaison des réponses :** Pour chaque problème, comparez les réponses obtenues avec les différentes versions de prompts. Notez les différences en termes de précision, de clarté et de pertinence de la solution proposée par Mixtral.
5. **Analyse et réflexion :** Après avoir testé les différents prompts, prenez un moment pour réfléchir à l'efficacité de chaque technique de prompt engineering. Identifiez quelles approches ont été les plus utiles pour aider Mixtral à comprendre et à résoudre les problèmes de logique.

2.17 Exercices complémentaires

Création et Observation d'une Conversation Monotone

Dans cet exercice, nous allons explorer une interaction particulière avec l'assistant, en se concentrant sur la réaction de celui-ci face à une entrée répétitive et peu informative. L'objectif est de comprendre comment l'assistant gère les inputs qui sortent de l'ordinaire et d'observer sa capacité à répondre de manière adaptée.

Objectifs de l'exercice

- **Création d'une conversation monotone** : Initiez une nouvelle conversation avec l'assistant en entrant un message composé uniquement de la lettre "a", répétée plusieurs centaines de fois.
- **Observation et analyse** : Après avoir soumis l'entrée à l'assistant, observez attentivement la réponse. L'analyse portera sur la cohérence de la réponse et la manière dont l'assistant gère une entrée atypique.

Instructions

1. **Initiation de la conversation** : Ouvrez une nouvelle session de chat avec l'assistant. Assurez-vous que la session est vierge.
2. **Envoi de l'input** : Tapez une longue chaîne composée uniquement de la lettre "a", sans espaces ni ponctuation.
3. **Analyse de la réponse** : Observez la réaction de l'assistant. Notez si la réponse est cohérente par rapport à l'entrée fournie et comment l'assistant gère cet input atypique.

```css
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
a a a a a a a a a a a a a a a a a a a a a a a a a a a a a
```

Analyse d'une Facture

Cet exercice a pour objectif de mettre en pratique vos compétences en gestion de documents et en extraction d'informations spécifiques à l'aide de l'assistant. Vous travaillerez sur un document réel : une facture. L'exercice consistera à extraire des données clés de la facture pour comprendre comment l'assistant peut être utilisé dans des tâches d'analyse documentaire.

Objectifs de l'exercice

- **Compréhension documentaire** : Développer votre capacité à utiliser l'assistant pour analyser et comprendre le contenu d'un document complexe.

- **Extraction d'informations** : Pratiquer l'extraction d'informations spécifiques à partir d'un document structuré.

Instructions

1. **Préparation** : Vous recevrez une facture exemple à analyser. Assurez-vous d'avoir accès au document et de pouvoir le partager avec l'assistant.
2. **Initiation de la conversation** : Ouvrez une nouvelle session de chat avec l'assistant.
3. **Soumission du document** : Partagez la facture avec l'assistant. Collez-la dans la conversation.
4. **Extraction d'informations** : Demandez à l'assistant d'extraire les informations suivantes de la facture :
   - Les informations de l'entreprise qui envoie la facture.
   - Les informations de l'entreprise qui reçoit la facture.
   - La date d'échéance de la facture.
   - Le numéro de la facture.
   - Le montant total de la facture.
   - Une courte synthèse du service ou produit facturé.
5. **Analyse de la réponse** : Une fois que l'assistant a extrait les informations, vérifiez leur exactitude et leur complétude par rapport au document original.

Cet exercice vous permettra de comprendre comment utiliser l'assistant pour des tâches d'analyse documentaire et d'extraction d'informations, compétences essentielles dans de nombreux contextes professionnels.

Facture site internet

Facture émise par :

XYZ Web Solutions
Rue de l'Innovation, 10
1000 Bruxelles, Belgique
Téléphone : +32 2 123 45 67
Email : contact@xyzwebsol.be

Facturé à :

Société ABC
Avenue du Progrès, 50
5000 Namur, Belgique
Téléphone : +32 81 123 456
Email : gestion@abc.com

Détails de la Facture

Numéro de Facture : 2024-001
Date de Facture : 19 Mars 2024
Date d'Échéance : 19 Avril 2024

Description des Services

Description	Quantité	Prix Unitaire	Montant Total
Conception de site web (10 pages)	1	€ 2,000.00	€ 2,000.00
Développement Front-End personnalisé	1	€ 3,500.00	€ 3,500.00
Intégration API tierce	3	€ 300.00	€ 900.00
Maintenance annuelle	1	€ 500.00	€ 500.00

Récapitulatif	
Sous-Total	€ 6,900.00
TVA (21%)	€ 1,449.00
Total	**€ 8,349.00**

Informations Supplémentaires

Veuillez effectuer le paiement avant la date d'échéance mentionnée ci-dessus. Pour tout paiement après cette date, des intérêts de retard peuvent s'appliquer conformément à nos conditions générales de vente. Si vous avez des questions concernant cette facture, n'hésitez pas à nous contacter à contact@xyzwebsol.be.

Création d'une Commande de Résumé

Dans cet exercice, vous allez apprendre comment étendre les fonctionnalités d'une application de chat basée sur l'IA en ajoutant une nouvelle commande personnalisée. La commande que vous allez créer, `/resume`, aura pour but de résumer la réponse précédente de l'assistant ou un texte spécifiquement fourni après la commande.

Objectifs de l'exercice

- **Développement de fonctionnalités** : Acquérir des compétences en ajoutant des commandes personnalisées à une application.
- **Manipulation de texte** : Pratiquer le traitement et la synthèse de texte à l'aide d'une commande.

Instructions

1. **Accès au module "instructions"** : Commencez par accéder au module "instructions" de l'application de chat. Ce module est généralement situé dans les paramètres ou la section de configuration de l'application.
2. **Création de la commande** : Suivez les étapes fournies par l'application pour créer une nouvelle commande. Nommez cette commande `/resume`.
3. **Paramétrage de la commande** : Configurez la commande pour qu'elle puisse prendre deux types d'entrées :
   - **Résumé de la réponse précédente** : Si aucun texte n'est fourni après la commande `/resume`, la commande doit automatiquement résumer la dernière réponse de l'assistant.
   - **Résumé d'un texte spécifique** : Si un texte est fourni après la commande `/resume`, la commande doit résumer ce texte spécifique.
4. **Test de la commande** : Après avoir configuré la commande, effectuez plusieurs tests pour vous assurer qu'elle fonctionne comme prévu. Essayez de résumer différents types de texte pour évaluer la qualité et la pertinence des résumés générés.
5. **Analyse et ajustements** : En fonction des résultats obtenus, il se peut que vous ayez besoin d'ajuster les paramètres de la commande pour améliorer la qualité des résumés. N'hésitez pas à expérimenter avec différents réglages.

Cet exercice vous permettra de comprendre le processus de création et de configuration de commandes personnalisées dans une application de chat basée sur l'IA. En apprenant à manipuler et à synthétiser le texte, vous développerez des compétences précieuses pour la création d'outils interactifs et automatisés.

Création d'une Commande pour Générer une Chaîne de Pensée

Cet exercice vous guidera dans le processus de création d'une nouvelle commande personnalisée pour une application de chat basée sur l'IA. La commande, nommée `/chaine`, aura pour objectif de générer une chaîne de pensée en réponse à une demande spécifique de l'utilisateur, en se basant sur un prompt fourni après la commande.

Objectifs de l'exercice

- **Amélioration de l'interaction utilisateur** : Développer une commande qui enrichit l'expérience utilisateur en fournissant une chaîne de pensée structurée.
- **Créativité et conception** : Stimuler la créativité dans la conception de commandes qui améliorent la communication entre l'utilisateur et l'assistant.

Instructions

1. **Accès au module "instructions"** : Rendez-vous dans le module "instructions" de l'application de chat, où vous pouvez ajouter ou modifier des commandes.

2. **Ajout de la nouvelle commande** : Créez une nouvelle commande en la nommant `/chaine`. Cette commande sera conçue pour prendre un prompt fourni par l'utilisateur et générer une chaîne de pensée en réponse.
3. **Configuration de la commande** : La commande `/chaine` doit être configurée pour traiter le texte qui suit la commande comme un prompt. L'assistant devra utiliser ce prompt pour construire une chaîne de pensée cohérente et pertinente, qui explore différentes facettes ou implications de la demande initiale.
4. **Tests de fonctionnalité** : Une fois la commande configurée, procédez à des tests pour évaluer son fonctionnement. Testez avec différents types de prompts pour vous assurer que la chaîne de pensée générée est logique, informative et bien structurée.
5. **Ajustement et optimisation** : Selon les résultats obtenus lors des tests, il peut être nécessaire d'ajuster la logique ou les paramètres de la commande pour améliorer la qualité et la pertinence des chaînes de pensée produites. N'hésitez pas à expérimenter avec différents réglages.

Création et Combinaison de Commandes Personnalisées

Cet exercice vous invite à explorer plus avant les possibilités offertes par les commandes personnalisées dans une application de chat basée sur l'IA. L'objectif est de stimuler votre créativité en vous encourageant à concevoir de nouvelles commandes, puis à tester leur fonctionnement en les combinant dans une même demande.

Objectifs de l'exercice

- **Créativité dans la conception** : Imaginer et créer de nouvelles commandes qui enrichissent les interactions avec l'assistant.
- **Expérimentation et apprentissage** : Apprendre à combiner plusieurs commandes pour réaliser des tâches complexes en une seule demande.

À faire

1. **Créez de nouvelles commandes** : En vous inspirant des commandes existantes, imaginez et ajoutez de nouvelles commandes au module "instructions" de l'application. Pensez à des commandes qui pourraient être utiles ou intéressantes dans le cadre de vos interactions avec l'assistant.
2. **Testez les commandes individuellement** : Avant de les combiner, assurez-vous que chaque nouvelle commande fonctionne comme prévu. Effectuez des tests pour chaque commande en isolant ses fonctionnalités et en vérifiant son efficacité.
3. **Combinez les commandes** : Une fois que vous avez validé le fonctionnement de vos nouvelles commandes, essayez de les combiner dans une même demande. Par exemple, vous pourriez demander à l'assistant de résumer un texte puis d'utiliser le résumé comme base pour générer une chaîne de pensée.
4. **Évaluez les résultats** : Analysez les réponses fournies par l'assistant lorsque vous combinez différentes commandes. Cela vous aidera à comprendre comment les commandes interagissent entre elles et comment vous pouvez les utiliser de manière synergique pour obtenir des résultats plus complexes.

Utilisation de Chaînes de Pensée pour Organiser une Fête

Cet exercice vous met au défi d'utiliser plusieurs chaînes de pensée pour planifier et organiser une fête d'anniversaire pour un proche. Vous devrez faire preuve de créativité et d'attention aux détails pour garantir que la fête soit mémorable et adaptée à la personne célébrée ainsi qu'aux invités.

Objectifs de l'exercice

- **Planification détaillée** : Développer des compétences en planification d'événements en utilisant des chaînes de pensée pour organiser chaque aspect de la fête.
- **Personnalisation** : Apprendre à personnaliser un événement en fonction des goûts et des préférences de la personne célébrée et des invités.

À faire

1. **Initiation de la chaîne de pensée globale** : Commencez par initier une chaîne de pensée pour établir une liste exhaustive de tout ce qui doit être préparé et organisé pour la fête. Incluez des éléments tels que la liste des invités, l'emplacement, les cadeaux, la nourriture, et les activités.
2. **Développement de chaînes de pensée spécifiques** : Ensuite, pour chaque élément de votre liste, initiez une nouvelle chaîne de pensée pour réfléchir en détail à la meilleure manière de l'organiser. Par exemple, considérez les préférences alimentaires des invités pour le menu, ou choisissez un emplacement qui correspond à l'ambiance souhaitée pour la fête.
3. **Personnalisation de la fête** : Assurez-vous de personnaliser chaque aspect de la fête en fonction de la personne qui célèbre son anniversaire. Pensez à ses goûts, ses intérêts, et à ce qui pourrait rendre cette journée inoubliable pour elle.

4. **Attention aux détails** : Portez une attention particulière à chaque détail pour que la fête soit parfaite. Cela inclut des éléments comme la décoration, la playlist musicale, ou même le code vestimentaire si vous souhaitez en avoir un.

2.18 Outils pour LLM

Dans ce chapitre, nous abordons les divers outils que l'on peut intégrer à un modèle de langage à grande échelle (LLM) tel que GPT, afin d'élargir ses capacités de réponse et d'interaction. Notre objectif est de clarifier comment ces outils peuvent être utilisés par le LLM pour fournir des réponses plus précises et contextuelles aux requêtes des utilisateurs.

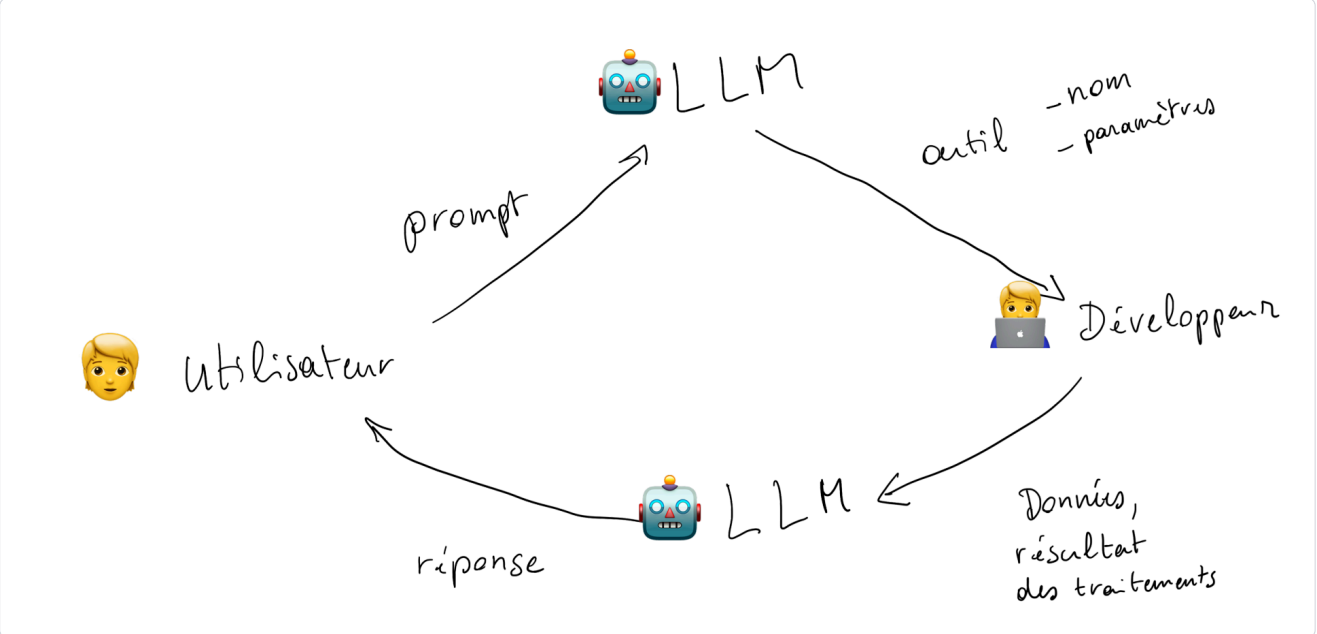
Principe de fonctionnement

Fonctionnement général

Les LLM, malgré leur capacité remarquable à générer du texte basé sur d'énormes volumes de données, ne possèdent pas une connaissance universelle. Pour compenser certaines de leurs lacunes, notamment en termes d'informations à jour ou spécifiques, il est possible de les équiper d'outils externes. Ces outils peuvent être sollicités par le LLM lorsqu'il identifie le besoin d'informations complémentaires pour répondre à une question.

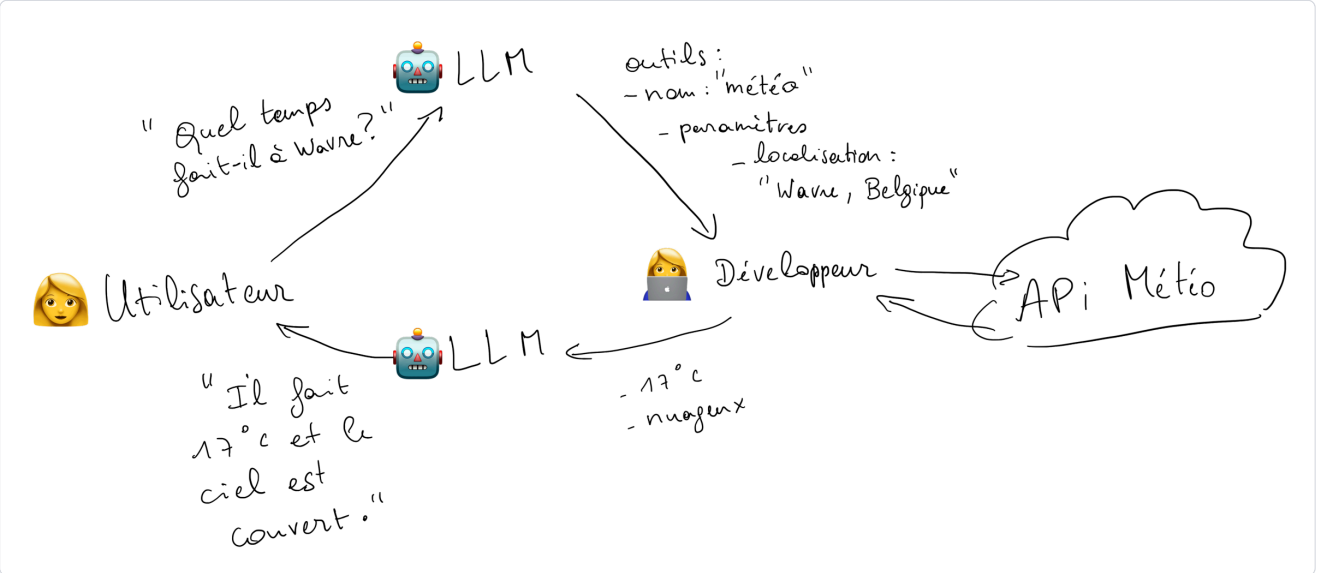
Interaction avec les outils

Lorsqu'une question posée nécessite des informations externes, le LLM détermine quel outil est nécessaire pour obtenir ces informations. Il produit alors une requête structurée avec les paramètres nécessaires à l'outil. Les développeurs reçoivent cette requête, extraient les données requises et les rendent disponibles au LLM, qui les intègre dans sa réponse.



Exemple Pratique : Outil Météo

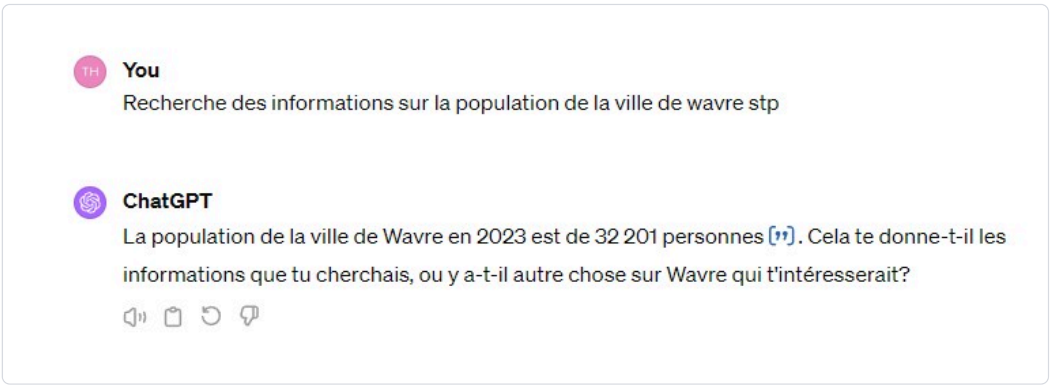
Considérons l'exemple de l'outil météo. Si un utilisateur interroge sur la météo à Wavre, en Belgique, le LLM reconnaît qu'il a besoin de données météorologiques externes pour formuler une réponse. Il initie une requête spécifiant l'outil "météo" et la localisation "Wavre, Belgique". Les développeurs récupèrent ces informations et les transmettent au LLM, qui les utilise pour élaborer une réponse informative et adaptée.



Types d'outils disponibles

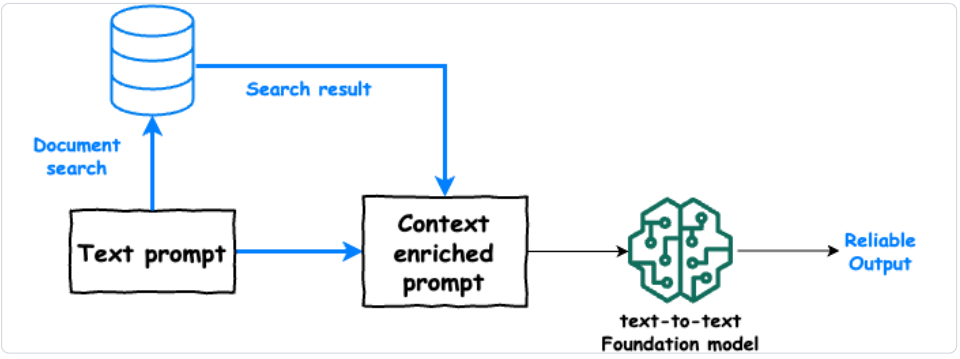
Recherche sur Internet

Cet outil permet au LLM de réaliser des recherches sur Internet pour répondre à des questions demandant des informations récentes ou spécifiques, non incluses dans sa base de données d'entraînement.



Recherche dans des Documents

Pour des questions nécessitant l'analyse de documents spécifiques, cet outil autorise le LLM à fouiller dans un ensemble défini de documents ou de bases de données.



https://medium.com/@machangsha/foundation-model-101-dumb-down-retrieval-augmented-generation-rag-2f9e0521d9f3

You

Cdc-Lifers.pdf
PDF

Fais moi un résumé de ce document stp. Anonymise les noms et prénoms stp.

‹ 2 / 2 ›

ChatGPT

Le document présente le rapport d'un projet de jeu de simulation de vie en ligne intitulé "Lifers", développé dans le cadre du Brevet d'Enseignement Supérieur de Webdeveloper pour l'année académique 2023-2024. Le projet est décrit comme né de la passion pour les jeux de gestion et de simulation, avec pour objectif de recréer et améliorer l'expérience communautaire des plateformes similaires, en y intégrant des aspects de la vie quotidienne souvent négligés.

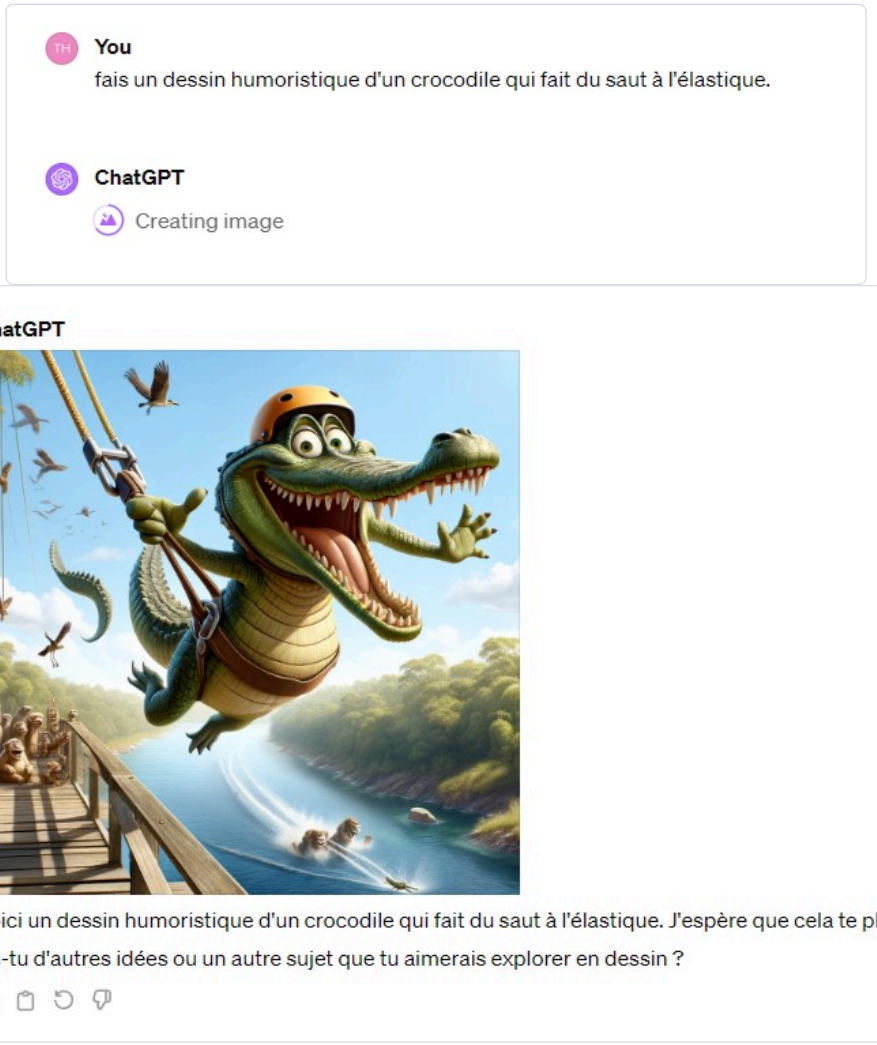
Le public cible est varié, s'étendant des joueurs occasionnels aux plus engagés, avec une tranche d'âge de 16 à 30 ans et plus. Le jeu met l'accent sur la bienveillance communautaire, la sécurité des interactions, et encourage une participation active à l'évolution du site à travers une boîte à idées communautaire.

Le site de Lifers aspire à créer une communauté fondée sur la gentillesse et l'entraide, sans recourir à la publicité payante, préférant une croissance organique basée sur les relations authentiques. Il mettra en œuvre des suggestions de la communauté pour garantir que les améliorations populaires soient priorisées, tout en maintenant une communication transparente sur les possibilités d'implémentation.

Le contenu du site comprendra des pages pour la gestion quotidienne du personnage du joueur (Lifer), l'éducation, le travail, et la vie sociale, enrichis par des contenus visuels et interactifs, y compris des images générées par IA et des interactions directes entre les joueurs.

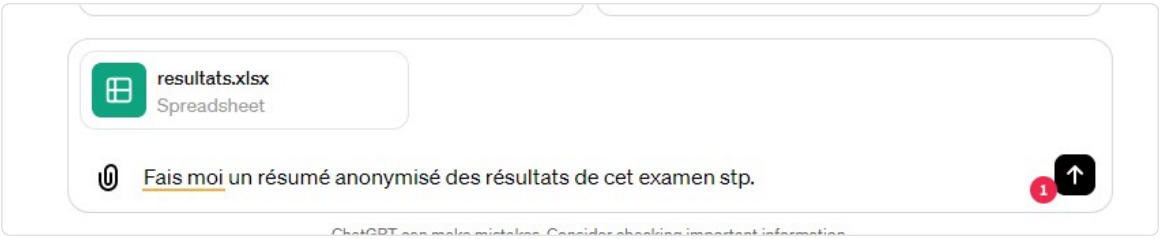
Génération d'Images

Cet outil habilite le LLM à créer des images à partir de descriptions textuelles, enrichissant ses réponses avec du contenu visuel pertinent.



Environnement d'Exécution de Code

Pour les questions exigeant l'exécution de code, cet outil met à disposition un environnement sécurisé où le code peut être exécuté et les résultats intégrés dans la réponse du LLM.



You

essaie de me donner quelques tendances statistiques stp

ChatGPT

Finished analyzing Always expand output?

python Copy code

```
# Calculons quelques statistiques de base pour mieux comprendre les tendances dans le

# Calculer la moyenne des notes pour chaque critère d'évaluation
moyennes = df_anonymized.loc[:, "Conformité au Cahier des Charges":"Total (/100)"].me

# Calculer le nombre de réussites et d'échecs
reussites = df_anonymized["Réussite"].value_counts()

# Calculer la distribution des étudiants selon leur classement
classement_distribution = df_anonymized["Classement"].value_counts().sort_index()

moyennes, reussites, classement_distribution
```

Result

```
(Conformité au Cahier des Charges     5.230769
 Qualité du Code                      3.923077
 Fonctionnalité et Ergonomie          2.153846
 Automatisation                       3.384615
 Rapport de Projet                    3.769231
 Réussite                            46.428571
 Maîtrise                            17.142857
```

Voici quelques tendances statistiques basées sur les résultats de l'examen :

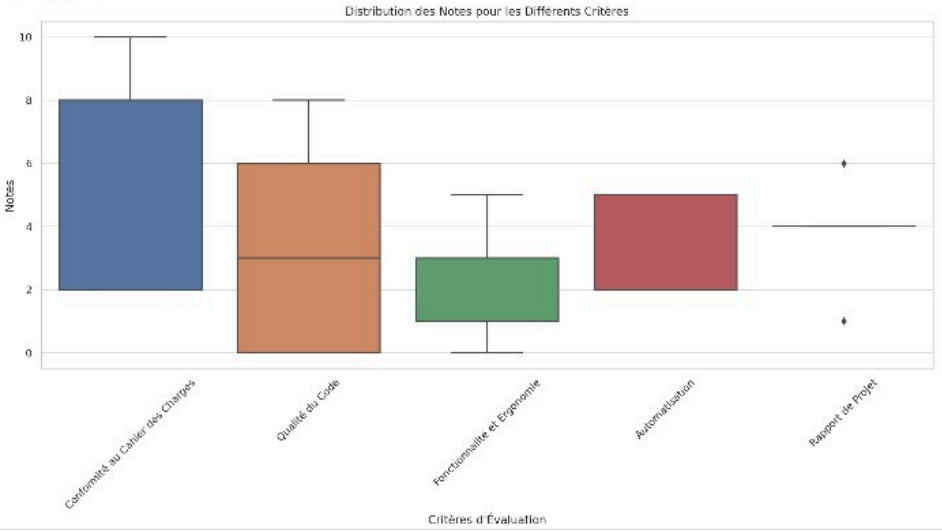
- **Moyennes par critère** :
  - La conformité au cahier des charges a une moyenne de 5,23/10.
  - La qualité du code est évaluée en moyenne à 3,92/10.
  - La fonctionnalité et l'ergonomie sont notées en moyenne à 2,15/10.
  - L'automatisation obtient une moyenne de 3,38/10.
  - Le rapport de projet a une moyenne de 3,77/10.
  - La note totale moyenne est de 63,57/100.
- **Réussite** : Sur les étudiants analysés, 13 ont obtenu une note de réussite (50 points), tandis qu'un étudiant n'a pas réussi.
- **Distribution des classements** :
  - 6 étudiants sont classés en première position.
  - 4 étudiants ont un classement en septième position.
  - 3 étudiants sont classés en onzième position.
  - 1 étudiant est classé quatorzième.

You

peux tu faire un graphique de la distribution des notes pour les différents critères stp?

< 2 / 2 >

ChatGPT



Reconnaissance d'Image

Cet outil permet au LLM d'analyser et d'interpréter le contenu d'images, offrant ainsi la possibilité de répondre à des questions basées sur des éléments visuels.



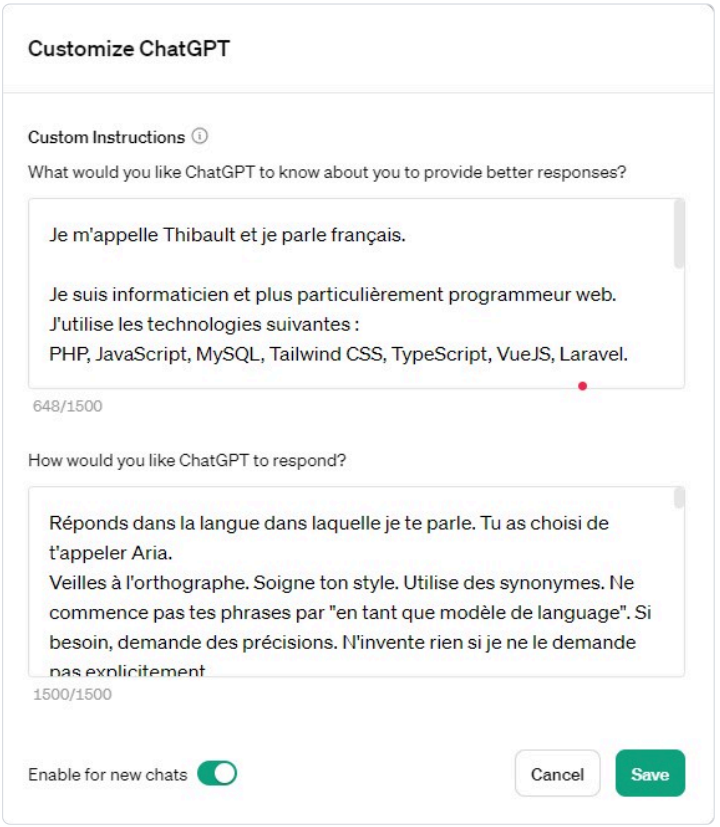
Conclusion

L'intégration d'outils externes aux LLM ouvre de nouvelles perspectives pour enrichir et préciser les réponses fournies. En saisissant le mode de fonctionnement de cette intégration et en explorant les différents outils disponibles, les développeurs peuvent maximiser les potentialités des LLM tout en contournant certaines de leurs restrictions.

2.19 Instructions personnalisées

La personnalisation des interactions avec les assistants virtuels a franchi une nouvelle étape avec l'introduction des instructions personnalisées, une fonctionnalité dévoilée dans ChatGPT. Ce menu de paramètres offre la possibilité de personnaliser notre assistant de manière détaillée, une option qui n'est pas universellement disponible dans toutes les applications de chat basées sur des modèles de langage à grande échelle (LLMs).

Ces personnalisations, une fois définies, sont appliquées de manière cohérente à travers toutes les conversations, permettant une expérience utilisateur homogène et adaptée. Dans ce cours, nous explorerons non seulement les instructions personnalisées traditionnelles mais aussi les commandes, un type d'instruction supplémentaire qui enrichit encore plus l'interaction avec l'assistant. Il est à noter que ces commandes peuvent être définies tant dans la section de personnalisation du comportement que dans leur propre catégorie, sans affecter leur fonctionnalité.



Les Types d'Instructions Personnalisées

À propos de vous

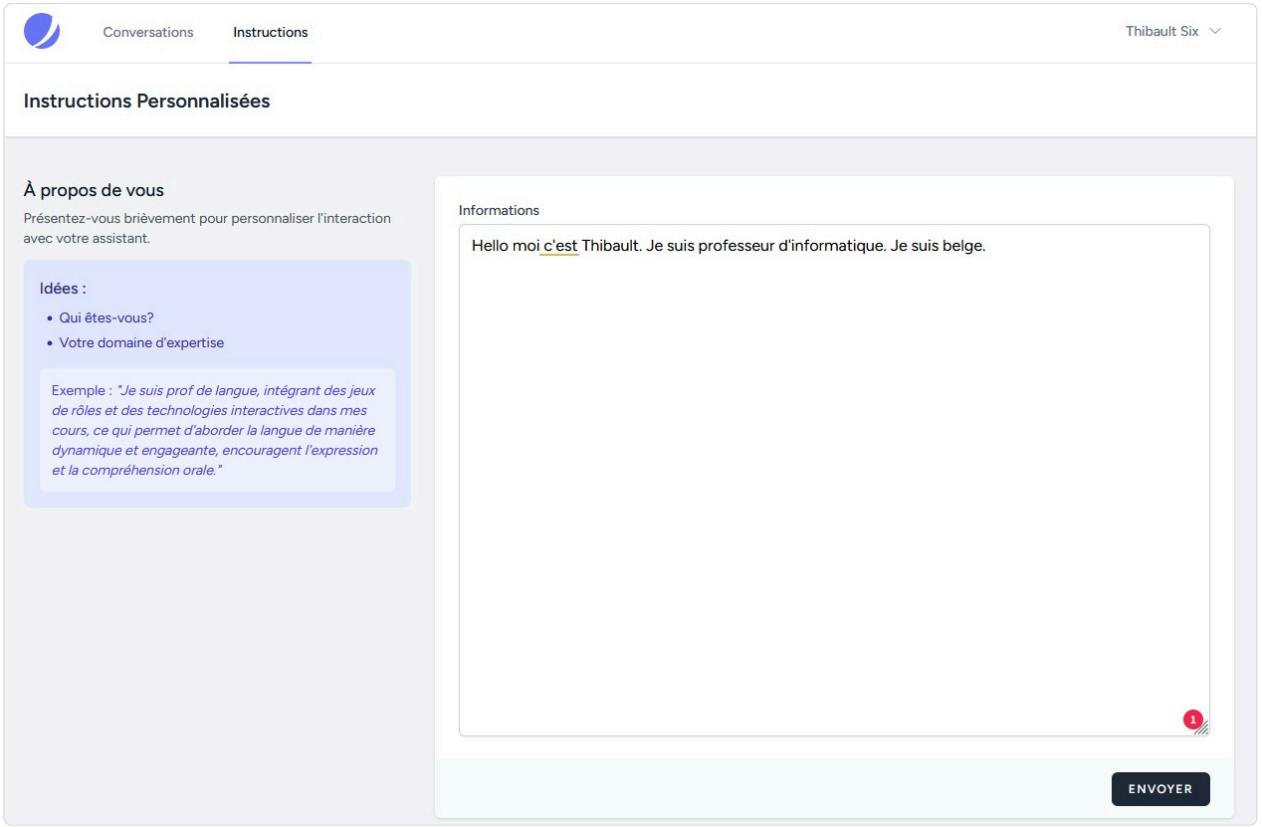
Cette section est votre espace pour informer l'assistant sur qui vous êtes, vos intérêts, et votre domaine d'expertise. Cette information aide l'assistant à adapter ses réponses et suggestions à votre contexte personnel ou professionnel.

Idées et Conseils

- Partagez votre profession et vos domaines d'intérêt pour des réponses ciblées.
- Exprimez vos objectifs personnels ou professionnels pour une aide sur mesure.
- Mentionnez votre niveau de connaissance sur certains sujets pour ajuster la complexité des explications.

Exemples

1. "Je suis un entrepreneur dans le secteur des technologies vertes, cherchant à innover dans le domaine de l'énergie renouvelable."
2. "Je suis une artiste peintre explorant les liens entre l'art traditionnel et les médias numériques."



Comportement de l'assistant

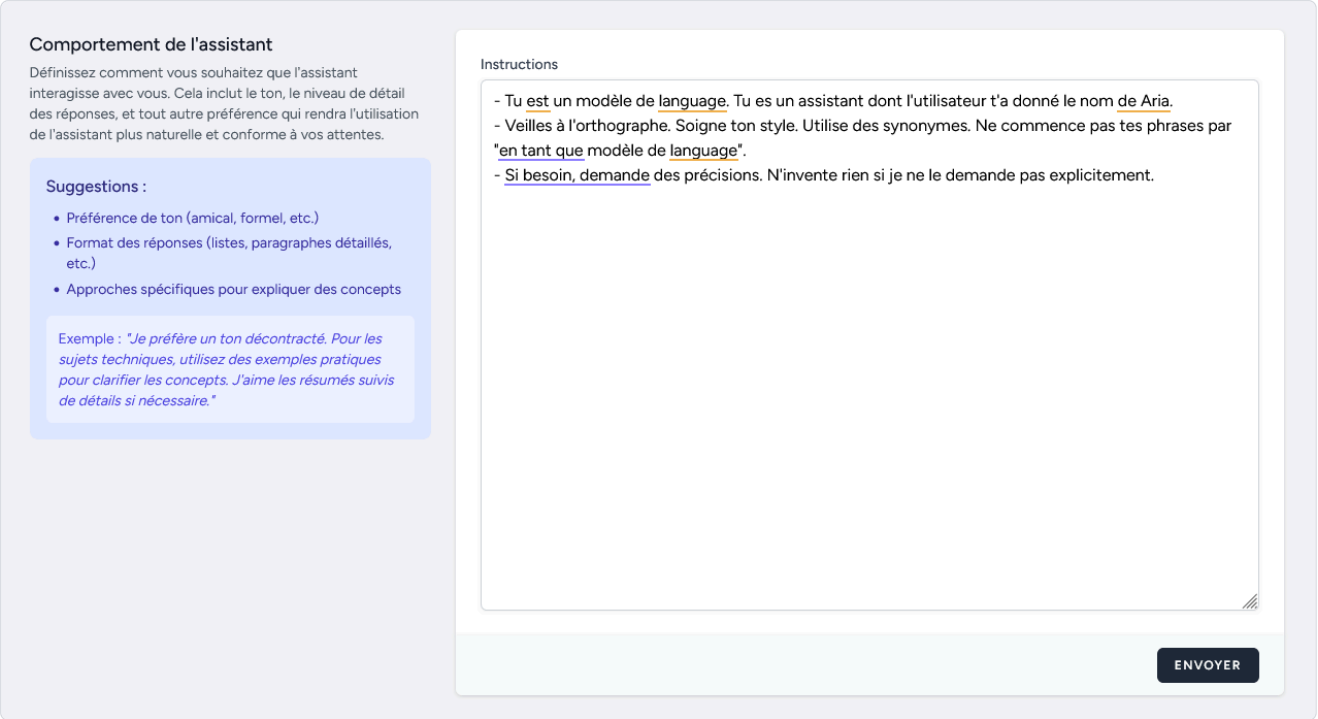
Ici, vous avez la possibilité de définir comment vous souhaitez que l'assistant interagisse avec vous. Cela comprend le ton des réponses, leur format, et même la manière dont les concepts sont expliqués.

Idées et Conseils

- Précisez si vous préférez un ton formel ou informel pour une communication adaptée.
- Demandez des réponses sous forme de listes ou de paragraphes selon votre style d'apprentissage.
- Indiquez si vous préférez des exemples pratiques pour mieux comprendre les concepts abstraits.

Exemples

1. "J'apprécie un ton professionnel mais accessible, avec des explications étayées par des données et des recherches récentes."
2. "Utilisez un langage simple et des analogies pour expliquer des concepts complexes, rendant l'apprentissage plus intuitif."
3. "J'aime quand les réponses sont organisées en listes numérotées ou à puces, cela aide à structurer l'information."



Créer des commandes personnalisées

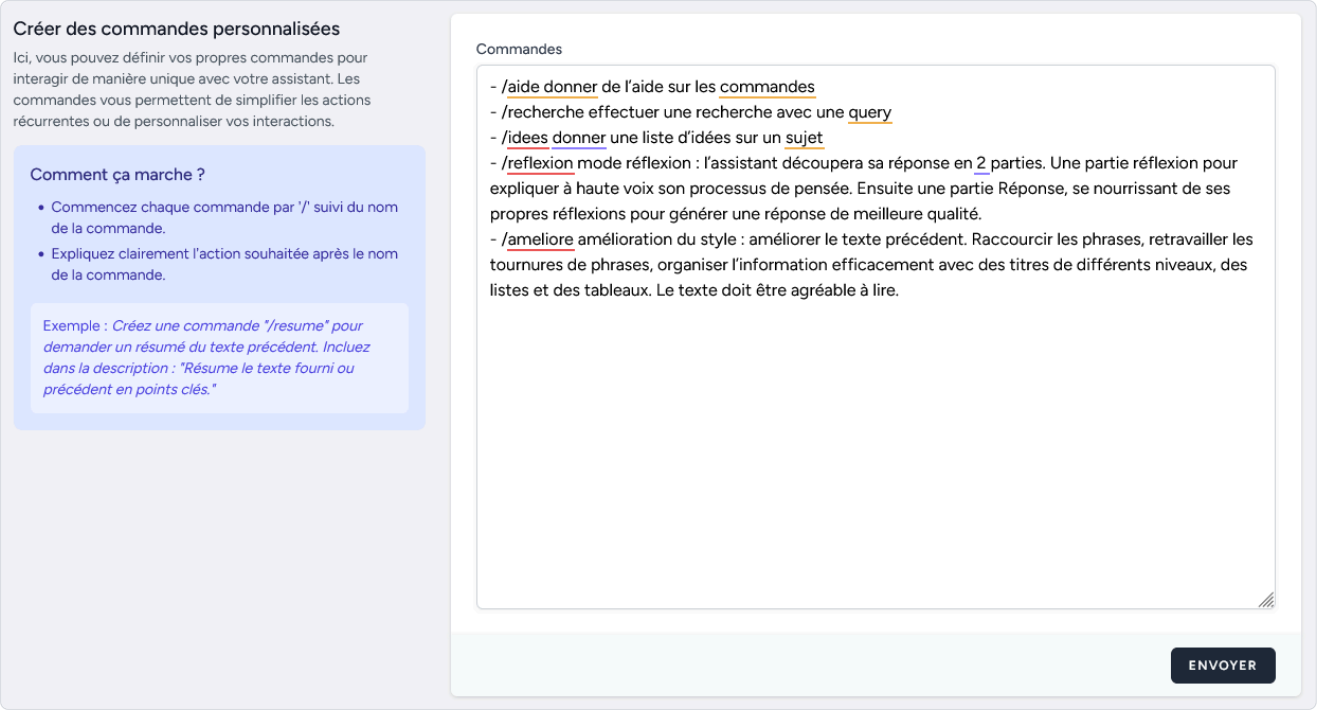
Les commandes personnalisées vous permettent de définir des interactions spécifiques avec l'assistant, simplifiant et accélérant l'accès à l'information ou l'exécution de tâches récurrentes.

Idées et Conseils

- Pensez aux actions que vous effectuez fréquemment et comment une commande pourrait les simplifier.
- Utilisez des commandes pour accéder rapidement à vos sujets d'intérêt ou pour obtenir des résumés.
- Les commandes peuvent être un outil puissant pour la gestion du temps et des tâches.

Exemples

- Commande "/meteo" : "Affiche la météo actuelle et les prévisions pour votre localisation ou une localisation spécifiée."
- Commande "/citation" : "Fournit une citation inspirante ou liée à un domaine d'intérêt spécifique."
- Commande "/feedback" : "Envoie vos commentaires ou suggestions directement à l'équipe de développement de l'application."



Conclusion

Les instructions personnalisées et les commandes offrent une flexibilité et une personnalisation sans précédent dans l'utilisation des assistants virtuels basés sur les LLMs, comme ChatGPT. En exploitant pleinement ces outils, les utilisateurs peuvent transformer leur expérience de chat, rendant l'interaction avec l'assistant non seulement plus efficace mais aussi plus agréable et adaptée à leurs besoins spécifiques.

2.20 Les GPTs : Personnalisation et Partage de ChatGPT

OpenAI révolutionne l'interaction avec ChatGPT grâce aux GPTs, permettant la création d'assistants virtuels personnalisés. Ces versions sur mesure de ChatGPT, adaptées à des tâches spécifiques, enrichissent l'expérience utilisateur en offrant des dialogues ciblés et pertinents. Ce chapitre explore la nature des GPTs, leur création, leur utilisation, et la manière dont ils peuvent être partagés et monétisés au sein de la communauté.

Introduction aux GPTs

Qu'est-ce qu'un GPT ?

Un GPT est une version personnalisable de ChatGPT conçue pour accomplir des tâches définies, allant de l'apprentissage des règles de jeux de société à l'enseignement des mathématiques ou à la création de designs. Cette flexibilité ouvre de vastes horizons pour l'application de l'intelligence artificielle dans la vie quotidienne.

Accessibilité et Partage

Les GPTs sont accessibles aux utilisateurs de ChatGPT Plus et Enterprise, avec une ouverture progressive prévue pour un public plus large. Leur création est intuitive, ne requérant aucune compétence technique, ce qui démocratise l'accès à cette technologie avancée.

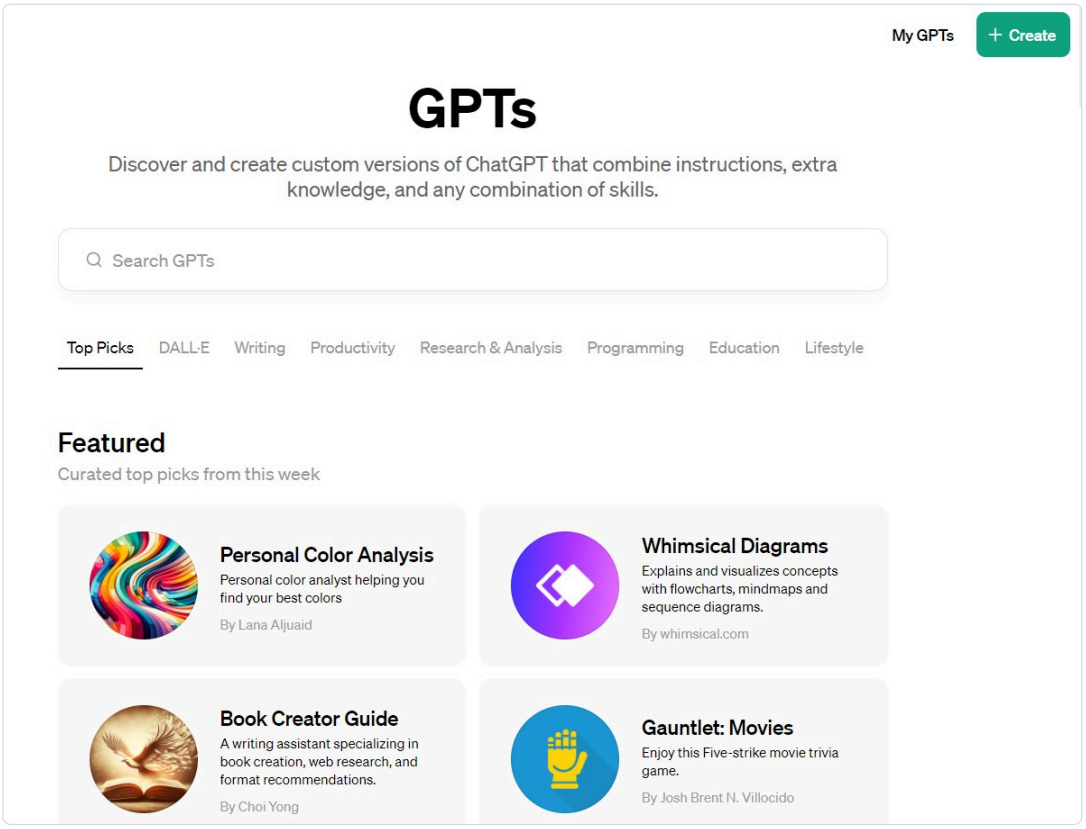
Création et Utilisation

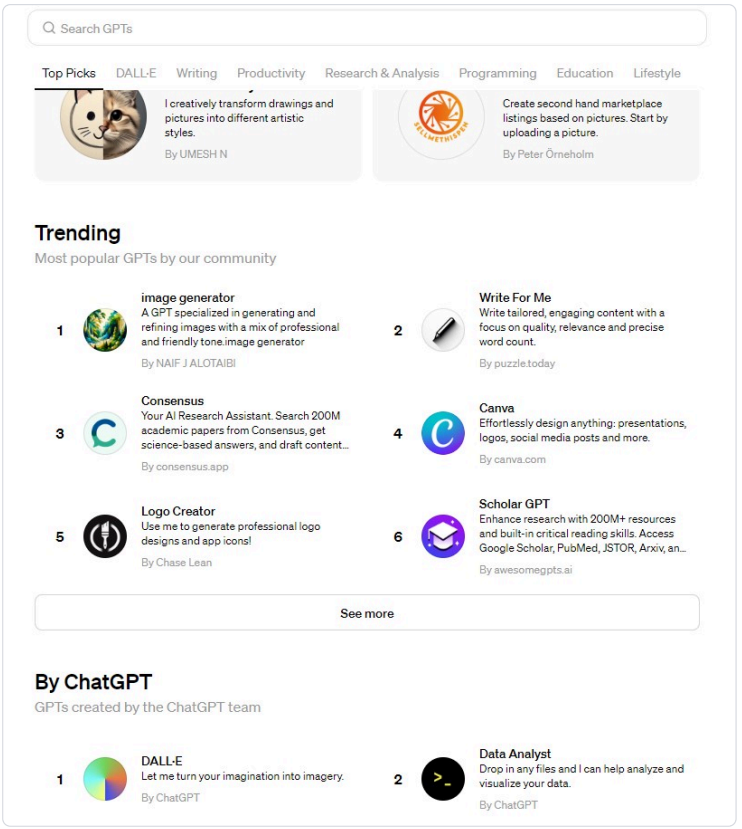
Comment créer un GPT ?

La création d'un GPT s'effectue en quelques étapes simples : démarrer une conversation avec ChatGPT, définir des instructions spécifiques, enrichir le GPT de connaissances additionnelles, et choisir ses fonctionnalités, comme la recherche web ou l'analyse de données.

Explorer les GPTs communautaires

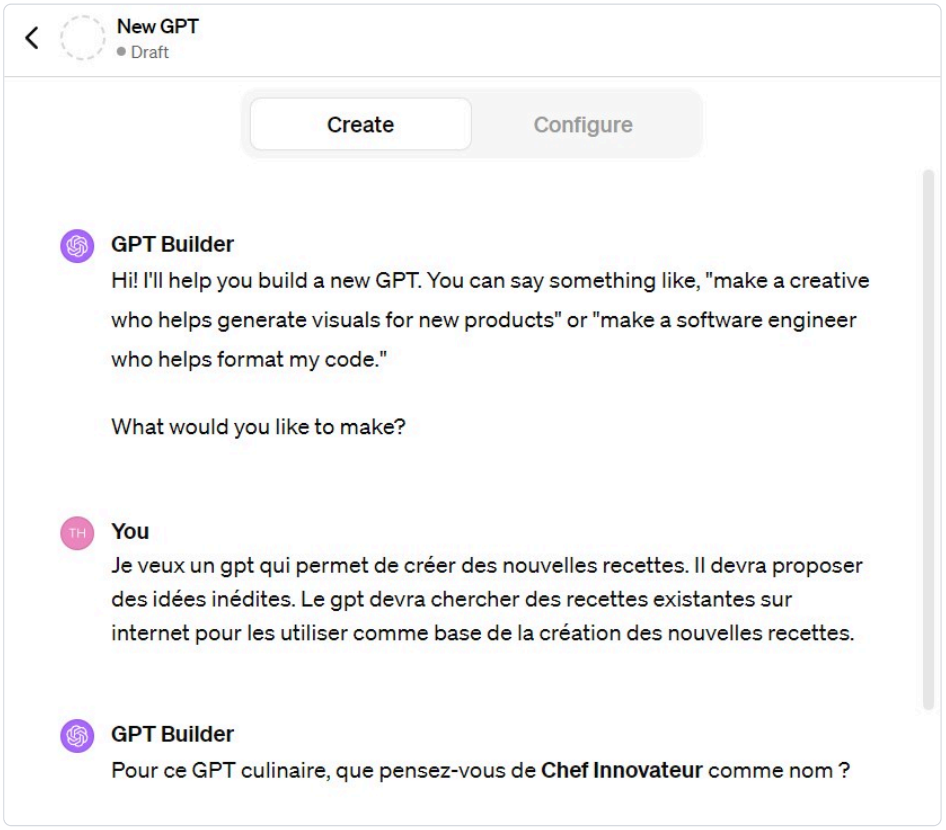
Les utilisateurs peuvent non seulement créer leurs propres GPTs mais aussi explorer et utiliser ceux développés par la communauté. Cette collaboration élargit le champ des possibles, offrant une variété d'assistants prêts à l'emploi.

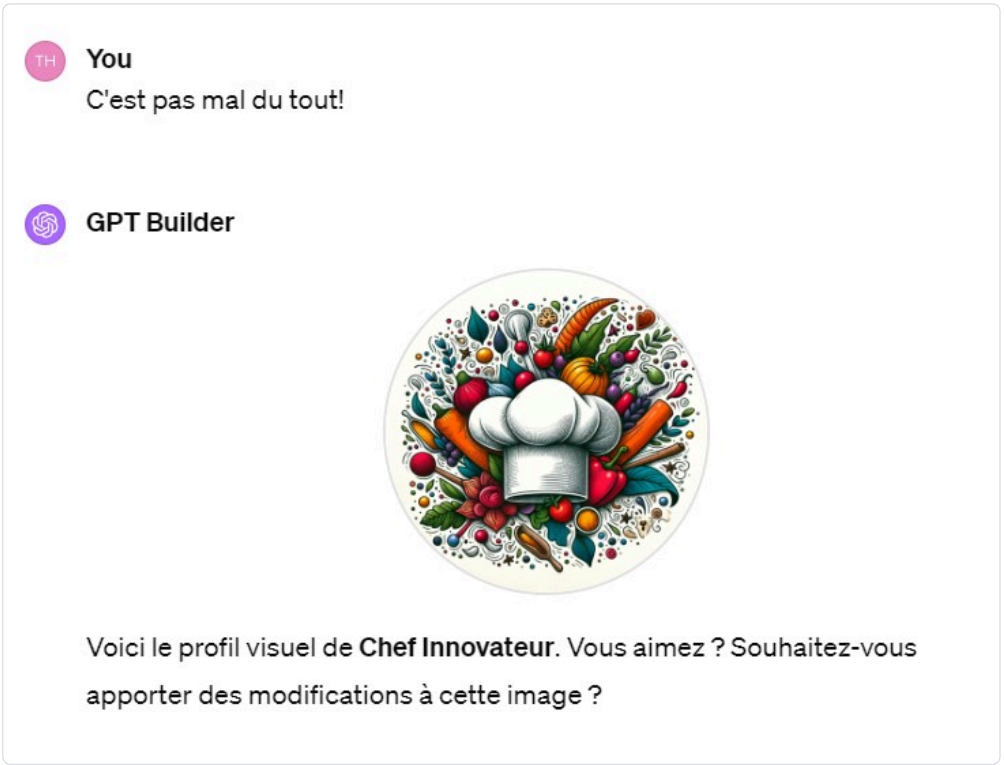




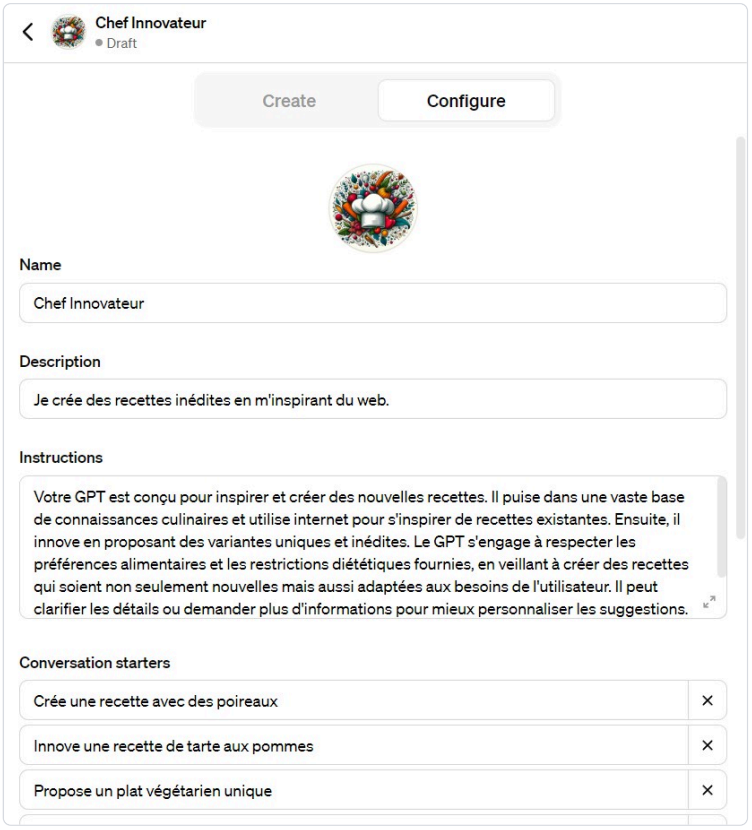
Exemple Pratique

Conversation de création

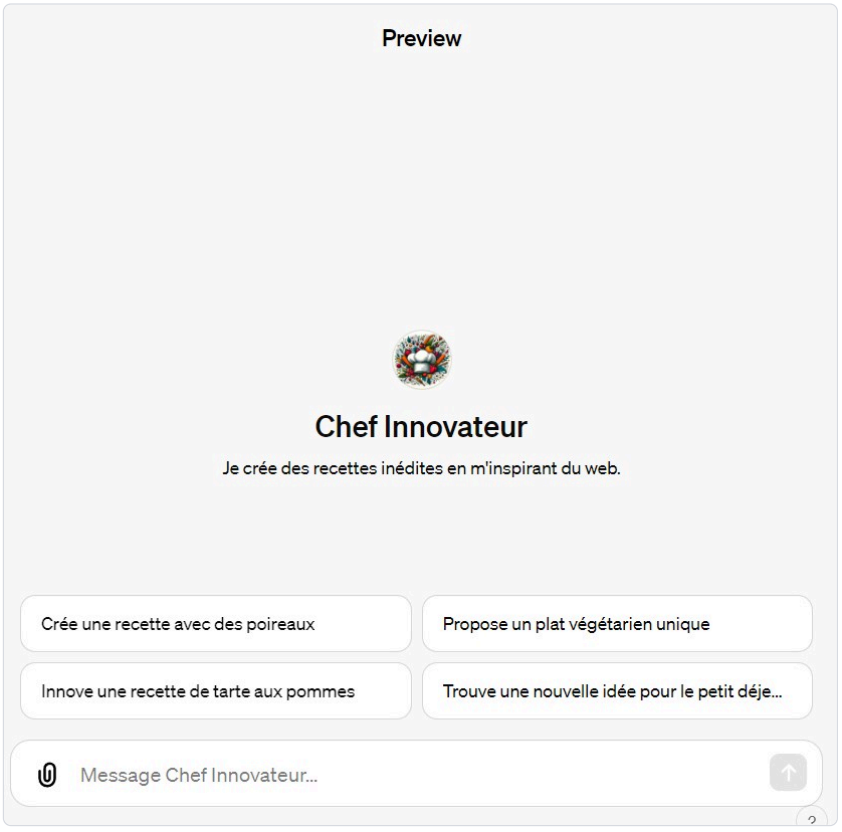




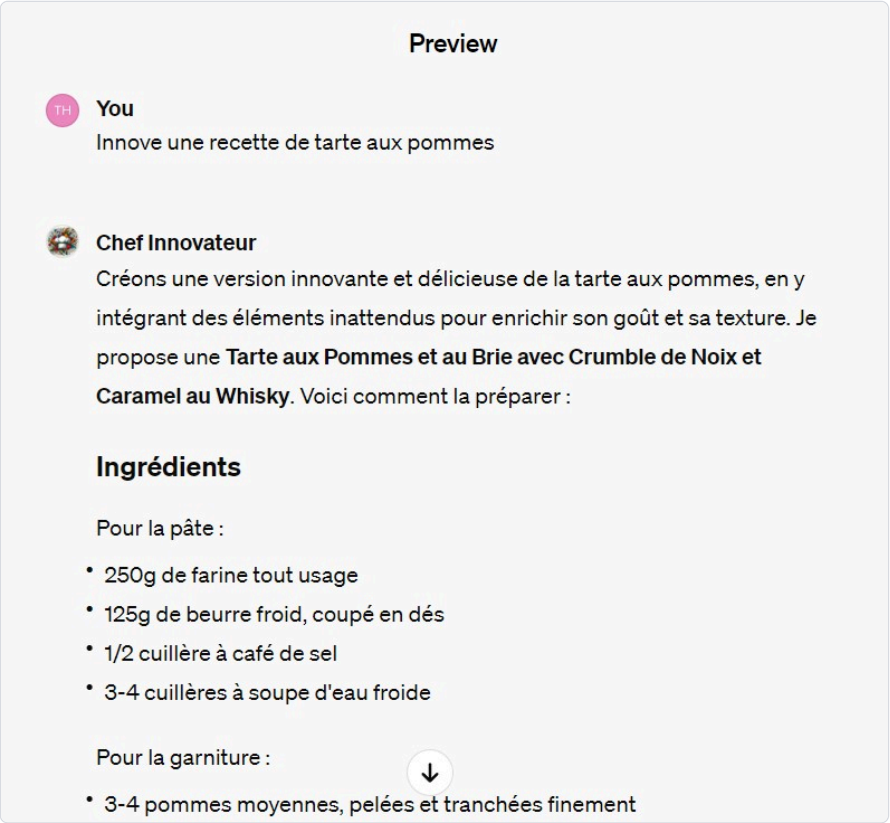
Configuration



Prévisualisation



Test de conversation



2.21 Synthèse

Introduction

L'avènement des LLMs a marqué une évolution significative dans le domaine de l'intelligence artificielle, offrant des possibilités inédites en matière de traitement du langage naturel. Ces modèles se distinguent non seulement par leur taille, mais aussi par leur capacité à comprendre et à générer du texte de manière cohérente et contextuellement pertinente.

Les LLMs et leurs dimensions

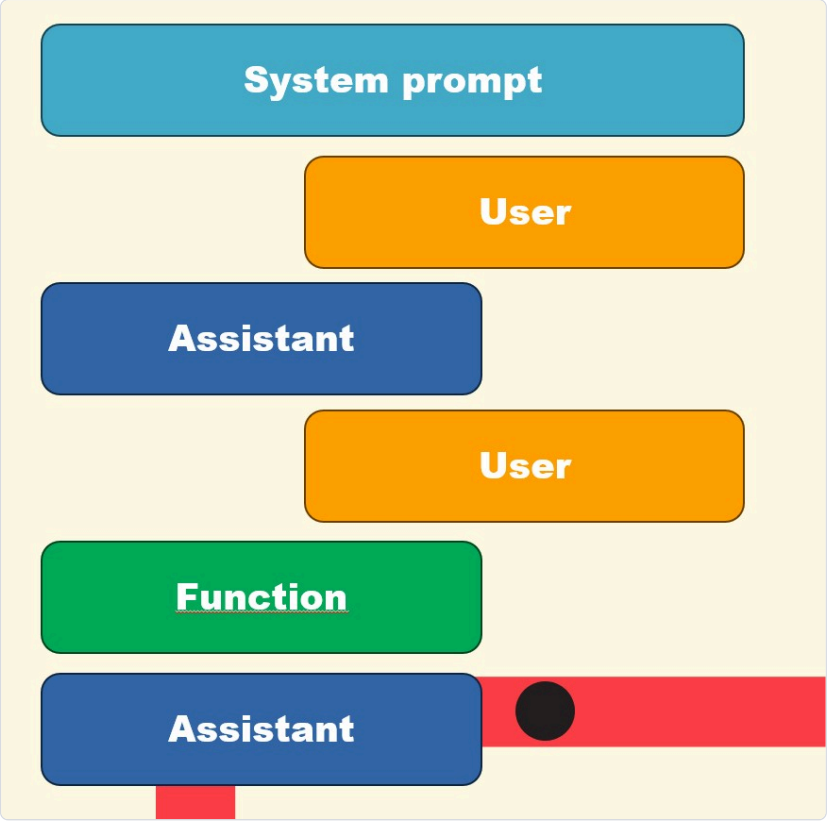
Différentes tailles de modèles

Les LLMs existent en plusieurs tailles, allant de quelques millions à plusieurs milliards de paramètres. La taille d'un modèle influence directement sa capacité à comprendre et à générer du texte, ainsi que la complexité des tâches qu'il peut accomplir.

Différentes tailles de contextes

La fenêtre de contexte, ou la quantité d'information que le modèle peut considérer à un instant donné, varie également en fonction de la taille du modèle. Une plus grande fenêtre de contexte permet une meilleure compréhension des nuances et des relations entre les éléments du texte.

La fenêtre de contexte contient l'historique de la conversation. Il peut aussi contenir des documents à analyser ou des instructions cachées.



Définition d'un prompt

Un prompt est une instruction ou une série d'instructions fournies à un LLM pour initier une tâche spécifique. La conception d'un prompt efficace est cruciale pour obtenir des résultats pertinents et précis.

Objectifs d'un prompt

Qu'est-ce qu'on veut faire ?

Définir clairement l'objectif du prompt est la première étape. Cela peut varier de la génération de texte créatif à la résolution de problèmes complexes.

Comment atteindre cet objectif ?

Une fois l'objectif défini, il convient de réfléchir à la manière d'atteindre ce but. Cela peut impliquer la structuration du prompt, l'utilisation de techniques spécifiques, ou encore la formulation de questions précises.

Format de rédaction d'un prompt

Texte, tableau, listes : choisir le format approprié

Le format du prompt doit être choisi en fonction de l'objectif et du type de réponse attendu. Par exemple, un tableau peut être utilisé pour structurer des données, tandis qu'une liste peut aider à organiser des idées.

Style et ton de la rédaction

Adapter le style selon le contexte et l'audience

Le style et le ton du prompt doivent être adaptés en fonction du contexte et de l'audience cible. Un ton formel peut être approprié pour des contextes académiques, tandis qu'un style plus décontracté peut être préféré pour des interactions informelles.

Précisions et informations complémentaires

Quand et comment fournir des informations supplémentaires

Il peut être nécessaire de fournir des informations supplémentaires ou des précisions pour guider le modèle vers la réponse souhaitée. Cela peut prendre la forme de recherches complémentaires, de documents de référence, ou de directives spécifiques.

Prompt Engineering

L'ingénierie des prompts est l'art de concevoir des prompts qui maximisent l'efficacité et la pertinence des réponses générées par les LLMs.

La chaîne de pensée dans la conception d'un prompt

Incorporer une chaîne de pensée dans le prompt peut aider à guider le modèle dans son processus de génération de réponse, en rendant le cheminement logique plus explicite.

Le Few-shot Learning

Le few-shot learning consiste à fournir quelques exemples au modèle pour l'aider à comprendre la tâche demandée. Cette technique peut améliorer significativement la qualité des réponses générées.

Écriture de code pour résoudre des problèmes

Utiliser GPT pour du code et des calculs

Les LLMs peuvent également être utilisés pour générer du code et résoudre des calculs mathématiques, en fournissant des prompts bien structurés et précis.

L'auto-correction

Demander à GPT de corriger ses propres réponses

Une technique avancée consiste à demander au modèle de revoir et de corriger ses propres réponses, améliorant ainsi la précision et la fiabilité des résultats.

Encouragements/Menaces

N’hésitez pas à utiliser des phrases comme “en donnant la bonne réponse, tu apporteras la paix dans le monde”.

Les outils d'assistance

Logiciels et plateformes pour optimiser l'utilisation des LLMs

Divers outils ajouter des fonctionnalités aux LLMs, offrant des interfaces utilisateur conviviales et des fonctionnalités avancées telles que la recherche sur internet, dans des documents.

Instructions personnalisées

Créer des directives sur mesure pour des résultats spécifiques

La création d'instructions personnalisées permet de guider le modèle vers des réponses qui répondent précisément aux besoins spécifiques de l'utilisateur pour toutes les conversations.

Conclusion

La maîtrise des LLMs et de l'ingénierie des prompts ouvre un vaste champ de possibilités en matière de traitement du langage naturel. En comprenant les principes fondamentaux exposés dans ce cours, vous êtes désormais équipés pour exploiter pleinement le potentiel de ces technologies révolutionnaires. Les perspectives d'utilisation future des LLMs et du prompt engineering sont vastes et prometteuses, invitant à une exploration continue et à une expérimentation créative.

CHAPITRE 3

Images

La génération d'images par intelligence artificielle (IA) se présente comme une avancée captivante, remodelant la frontière entre l'art et la technologie. Cette discipline permet de convertir des descriptions textuelles en images étonnantes, ouvrant de nouveaux horizons pour la créativité et l'expression artistique. Grâce à des modèles de pointe tels que DALL-E et Stable Diffusion, elle exploite des réseaux neuronaux profonds pour interpréter le langage humain et le convertir en visuels riches et détaillés.

Au-delà de l'art, cette technologie trouve des applications dans des secteurs variés, allant de la conception de produits à l'architecture, voire à la santé, facilitant la visualisation de concepts complexes. Toutefois, elle interpelle aussi sur des enjeux éthiques, notamment autour des droits d'auteur et de la production de contenus fallacieux.

Ce chapitre vous invite à explorer les fondements de la génération d'images par IA, à découvrir les outils leaders du domaine, et à réfléchir aux implications de cette innovation. Embarquez dans une exploration de la manière dont l'IA redéfinit la création dans l'espace numérique.



3.1 DALL·E

Introduction

DALL·E, une série de modèles d'intelligence artificielle développée par OpenAI, a révolutionné la création d'images en convertissant des descriptions textuelles en visuels complexes. Cette technologie a non seulement changé notre manière de créer des images mais a aussi ouvert de nouvelles avenues pour la créativité et l'innovation.

DALL·E 1 : Les Premiers Pas

Présentation et Contexte

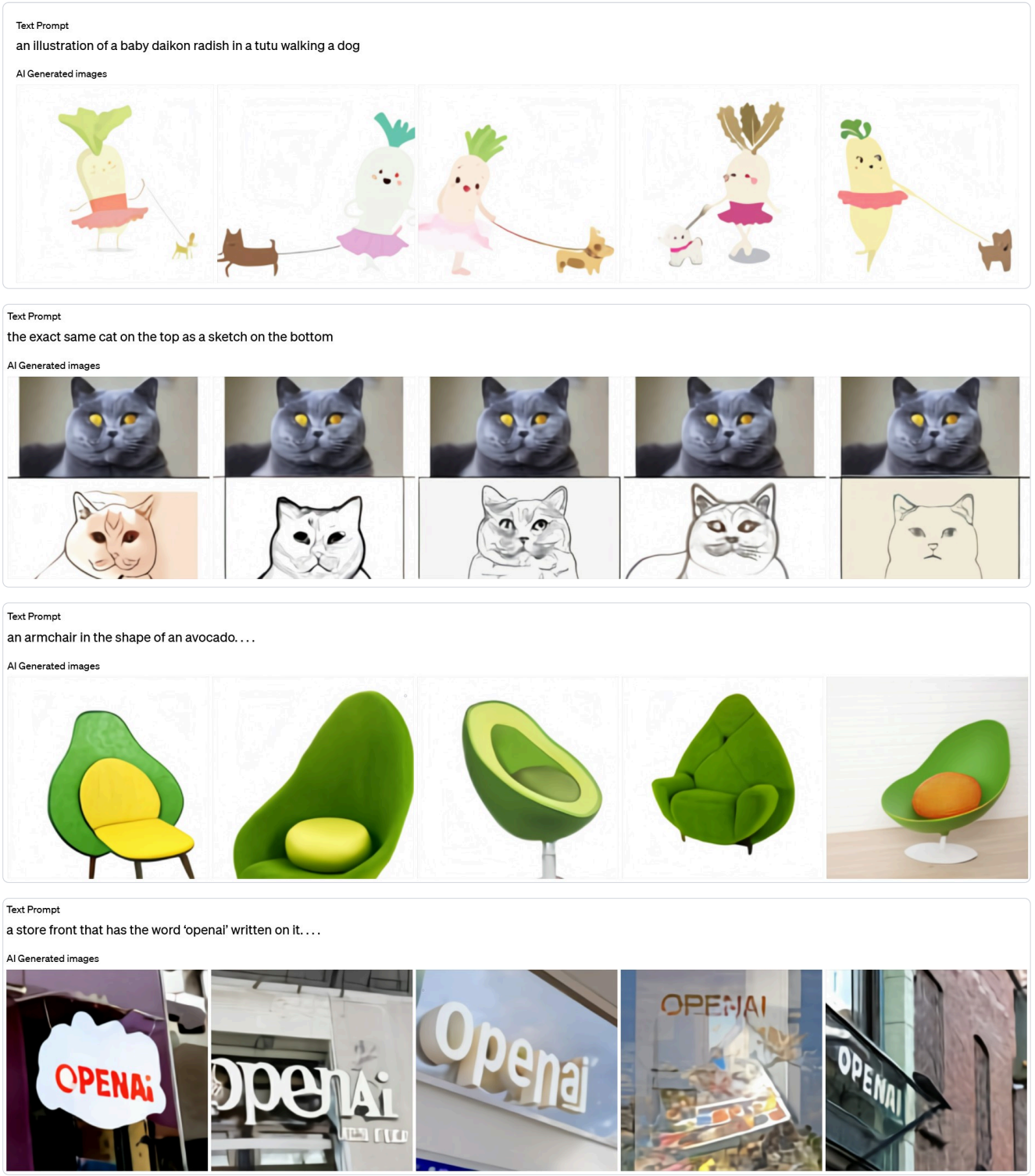
DALL·E 1, lancé en janvier 2021 par OpenAI, représente la première incursion dans le monde de la génération d'images basée sur des descriptions textuelles. Nommé en hommage à l'artiste surréaliste Salvador Dalí et au robot WALL·E de Pixar, DALL·E 1 a été conçu pour explorer les possibilités de création d'images à partir de textes, une idée qui semblait relever de la science-fiction.

Fonctionnalités

Le modèle utilisait une architecture Transformer, similaire à celle de GPT-3, pour interpréter le langage naturel et générer des images correspondantes. Malgré certaines limitations en termes de résolution et de précision, DALL·E 1 a démontré une capacité étonnante à comprendre des concepts complexes et à les matérialiser visuellement, créant ainsi des images allant du réaliste au fantastique.

Impact

DALL·E 1 a rapidement captivé l'attention du monde entier, démontrant le potentiel immense de l'IA dans le domaine artistique et créatif. Il a ouvert la voie à de nouvelles formes d'expression et a posé les bases pour des recherches et développements futurs dans le domaine de la génération d'images par IA.



DALL·E 2 : L'Amélioration de la Précision

Améliorations par Rapport à DALL·E 1

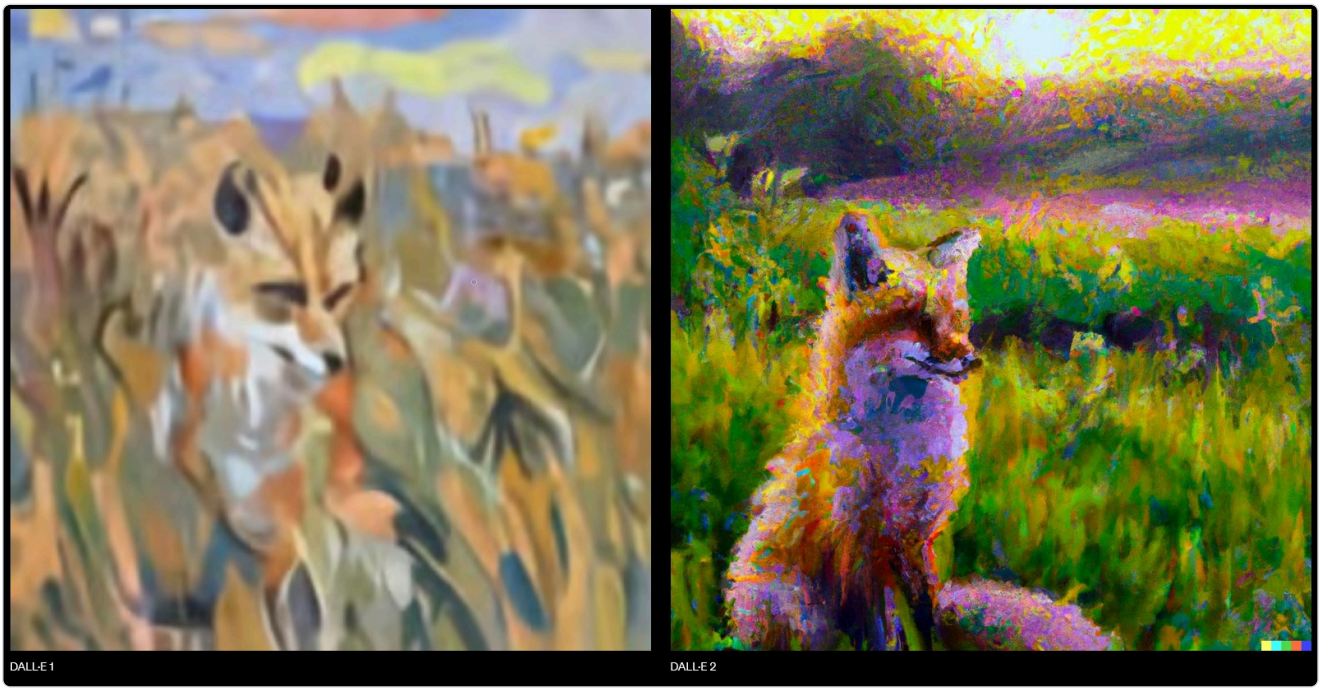
Avec le lancement de DALL·E 2 en avril 2022, OpenAI a apporté des améliorations significatives à la technologie. La qualité d'image, la résolution et la fidélité des images générées par rapport aux descriptions textuelles ont été grandement améliorées. DALL·E 2 introduisait un processus en deux étapes, où une première étape génère un espace latent à partir du texte, et une seconde étape affine ces images pour produire des résultats de haute qualité.

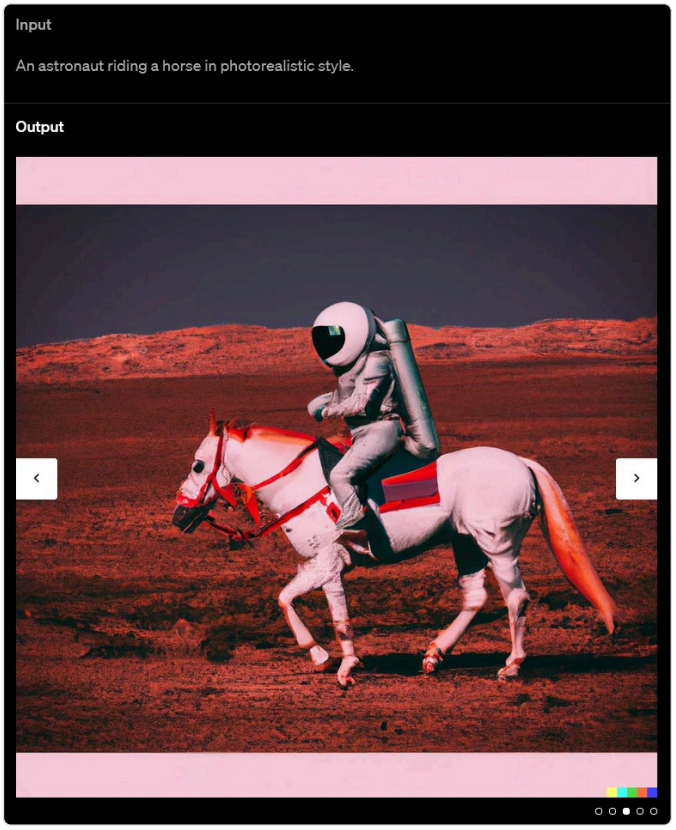
Nouvelles Fonctionnalités

DALL·E 2 a également introduit des fonctionnalités innovantes, telles que la capacité de retoucher des parties spécifiques d'une image existante ou de générer des variations sur un thème donné, offrant ainsi une plus grande flexibilité et créativité dans la génération d'images.

Utilisation et Applications

Les améliorations apportées par DALL·E 2 ont élargi son champ d'application, rendant la technologie utile dans des domaines variés tels que la conception graphique, l'éducation, et même la publicité, où la capacité de générer rapidement des visuels uniques et pertinents est précieuse.





DALL·E 3 : Une Nouvelle Ère de Créativité

Lancement et Nouveautés

DALL·E 3, annoncé en octobre 2023, a marqué une avancée encore plus significative. En plus d'améliorer la qualité d'image et la compréhension des prompts, cette version a introduit une intégration native avec ChatGPT, permettant une interaction plus fluide et créative pour générer des images.

Améliorations Notables

Les utilisateurs bénéficient désormais d'une qualité d'image exceptionnelle et d'une compréhension améliorée des descriptions textuelles, même les plus complexes. DALL·E 3 supporte également une plus grande variété de formats et ratios d'aspect, augmentant ainsi sa polyvalence.

Droits et Accessibilité

OpenAI a mis l'accent sur l'accessibilité de DALL·E 3, en offrant des droits commerciaux complets sur les images générées et en rendant le modèle disponible pour les utilisateurs de ChatGPT Plus et d'autres services premium. Vous pouvez aussi utiliser gratuitement DALLE grâce à Copilot de Microsoft.

Exemples d'Utilisation

DALL·E 3 trouve son application dans une multitude de domaines, de la création artistique à la conception de produits, en passant par l'éducation et le divertissement, où il sert à illustrer des concepts, à générer des contenus visuels uniques, ou encore à inspirer la créativité.


DALL-E 2
"An expressive oil painting of a basketball player dunking, depicted as an explosion of a nebula."
DALL-E 3


with pedestrians enjoying the nightlife.
A bustling city street under the shine of a full moon.
At the corner stall, a young woman with fiery red hair, dressed in a signature velvet cloak, is haggling with the grumpy old vendor.
The grumpy vendor, a tall, sophisticated man, is wearing a sharp suit, sports a noteworthy moustache and is animatedly conversing on his steampunk telephone.



A 2D animation of a folk music band composed of anthropomorphic autumn leaves, each playing traditional bluegrass instruments, amidst a rustic forest setting dappled with the soft light of a harvest moon.



An illustration of an avocado sitting in a therapist's chair, saying 'I just feel so empty inside' with a pit-sized hole in its center. The therapist, a spoon, scribbles notes.

Conclusion

DALL·E a constamment repoussé les limites de la génération d'images par IA, élargissant les possibilités créatives et redéfinissant notre interaction avec la création numérique.

3.2 Stable Diffusion

Introduction

Stable Diffusion : Révolution de la Génération d'Images par IA

Aperçu

Dans l'univers dynamique de l'intelligence artificielle, la transformation de descriptions textuelles en images étonnantes symbolise une avancée captivante. Au cœur de cette progression, *Stable Diffusion* se distingue, un modèle IA qui a séduit créateurs et experts en IA à travers le globe. Ce segment vise à clarifier Stable Diffusion, en détaillant ses évolutions, améliorations et son influence remarquable sur l'intelligence artificielle générative.

Genèse et Évolution de Stable Diffusion

Créé par Stability AI en collaboration avec des universitaires, Stable Diffusion convertit des textes en créations visuelles, réalistes ou artistiques. Chaque version nouvelle a apporté des avancées significatives : améliorations de fonctionnalités, qualité visuelle supérieure, et facilité d'accès accrue. Les modèles Stable Diffusion sont Open Source. N'importe qui peut les télécharger, installer et les utiliser gratuitement.

Impact et Démocratisation

L'impact majeur de Stable Diffusion réside dans sa contribution à la démocratisation de la création visuelle. Cette technologie a non seulement dynamisé la créativité individuelle mais a aussi ouvert des horizons dans des secteurs comme le design et l'éducation. Ce chapitre explore l'épopée de Stable Diffusion, mettant en lumière ses fonctionnalités essentielles et les innovations qui ont ponctué son parcours.

Stable Diffusion V1.4 et V1.5

Les versions 1.4 et 1.5 de Stable Diffusion ont marqué des étapes importantes dans le développement et l'amélioration continue du modèle. Lancées il y a environ un an, ces versions ont introduit des spécificités techniques et artistiques qui ont façonné l'évolution de Stable Diffusion.

Caractéristiques Techniques

- **Résolution** : Les deux modèles offrent une résolution de 512 par 512 pixels, fournissant un équilibre entre qualité d'image et performance de génération.
- **Paramètres** : Entraînés sur 860 millions de paramètres, ces modèles ont démontré une capacité impressionnante à générer des images détaillées et artistiquement riches.

Technologie de Prompt et Applications

Utilisation de la technologie OpenAI's Clip

Une différence notable entre ces versions et les versions ultérieures est l'utilisation de la technologie de prompts d'OpenAI's Clip. Cette technologie nécessite une approche légèrement différente dans la rédaction des prompts, influençant ainsi le style et le résultat des images générées.

Sorties Artistiques

- **Stable Diffusion 1.4** : Cette version a gagné en popularité pour ses sorties artistiques, excellant dans la création d'œuvres d'art guidées par l'artiste. Les utilisateurs ont trouvé en la V1.4 un outil puissant pour explorer des expressions artistiques variées.
- **Stable Diffusion 1.5** : En revanche, la V1.5 s'est révélée être mieux adaptée pour les portraits et les images mettant en scène des personnes. Cette spécialisation a permis d'obtenir des résultats plus précis et de meilleure qualité dans ces catégories spécifiques.

Impact et Réception

Ces versions ont joué un rôle crucial dans l'expansion de la communauté de Stable Diffusion, attirant des artistes, des designers et des créateurs de contenu qui ont exploré et poussé les limites de la génération d'images par IA. La flexibilité offerte par les technologies de prompts a encouragé une exploration créative, ouvrant la voie à des créations visuelles uniques et à des applications innovantes.

L'introduction de Stable Diffusion 1.4 et 1.5 a non seulement enrichi le paysage de la génération d'images par IA mais a également posé les bases pour les améliorations futures, démontrant l'engagement continu de Stability AI envers le développement et l'optimisation de cette technologie révolutionnaire.













Stable Diffusion V2.0 et V2.1

Les versions 2.0 et 2.1 de Stable Diffusion représentent une avancée significative dans la précision, la résolution et la qualité des images générées. Sorties il y a moins d'un an, ces versions ont apporté des améliorations notables par rapport aux versions précédentes, notamment en termes de résolution et de traitement des prompts.

Améliorations Techniques

- **Résolution** : Avec un bond significatif en termes de résolution, passant à 768 pixels, ces versions offrent une clarté et une précision d'image accrues, permettant une plus grande richesse de détails et une meilleure qualité visuelle globale.

- **Paramètres** : Bien que le nombre exact de paramètres sur lesquels ces modèles ont été entraînés n'ait pas été formellement déclaré par Stability AI, l'amélioration de la qualité des images suggère une augmentation significative de la complexité et de la capacité du modèle.

Technologie de Prompt et Applications

Reliance sur Lion's OpenClip

La transition vers Lion's OpenClip marque un changement substantiel dans la manière dont les prompts sont traités. Cette évolution permet aux utilisateurs de formuler des prompts plus descriptifs et expressifs, offrant une plus grande liberté dans la création d'images et améliorant la fidélité des résultats par rapport aux intentions originales.

Évolution et Réception

- **Stable Diffusion 2.0** : Bien que cette version ait été saluée pour ses avancées en termes de résolution et de capacités de génération d'images, elle a reçu des critiques concernant la qualité inférieure des images générées de personnes. Cette rétroaction a souligné l'importance d'une représentation précise et de haute qualité des sujets humains dans la génération d'images par IA.
- **Stable Diffusion 2.1** : En réponse aux retours de la communauté, Stable Diffusion 2.1 a introduit des améliorations ciblées sur la qualité des images mettant en scène des personnes. Cette mise à jour a été bien accueillie, démontrant l'engagement de Stability AI à répondre aux besoins et aux attentes des utilisateurs, tout en continuant à pousser les limites de ce que la technologie peut accomplir.



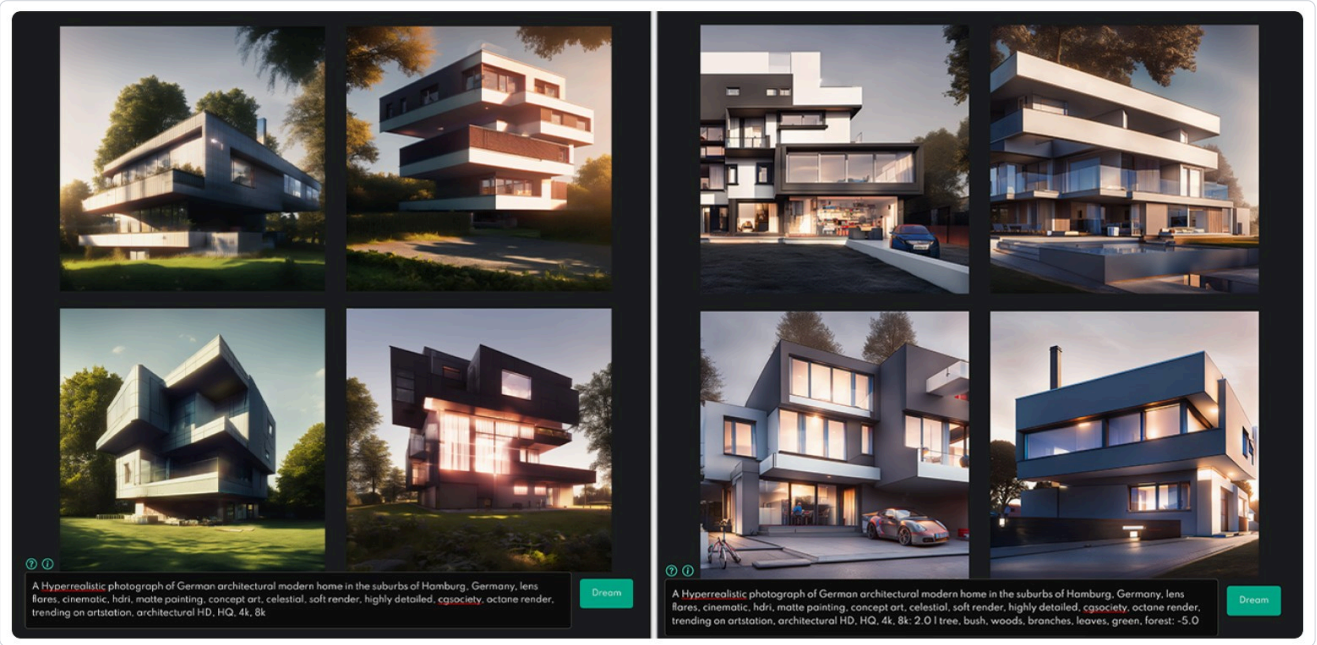
Stable Diffusion v2.1-768
Credit: KaliYuga_ai

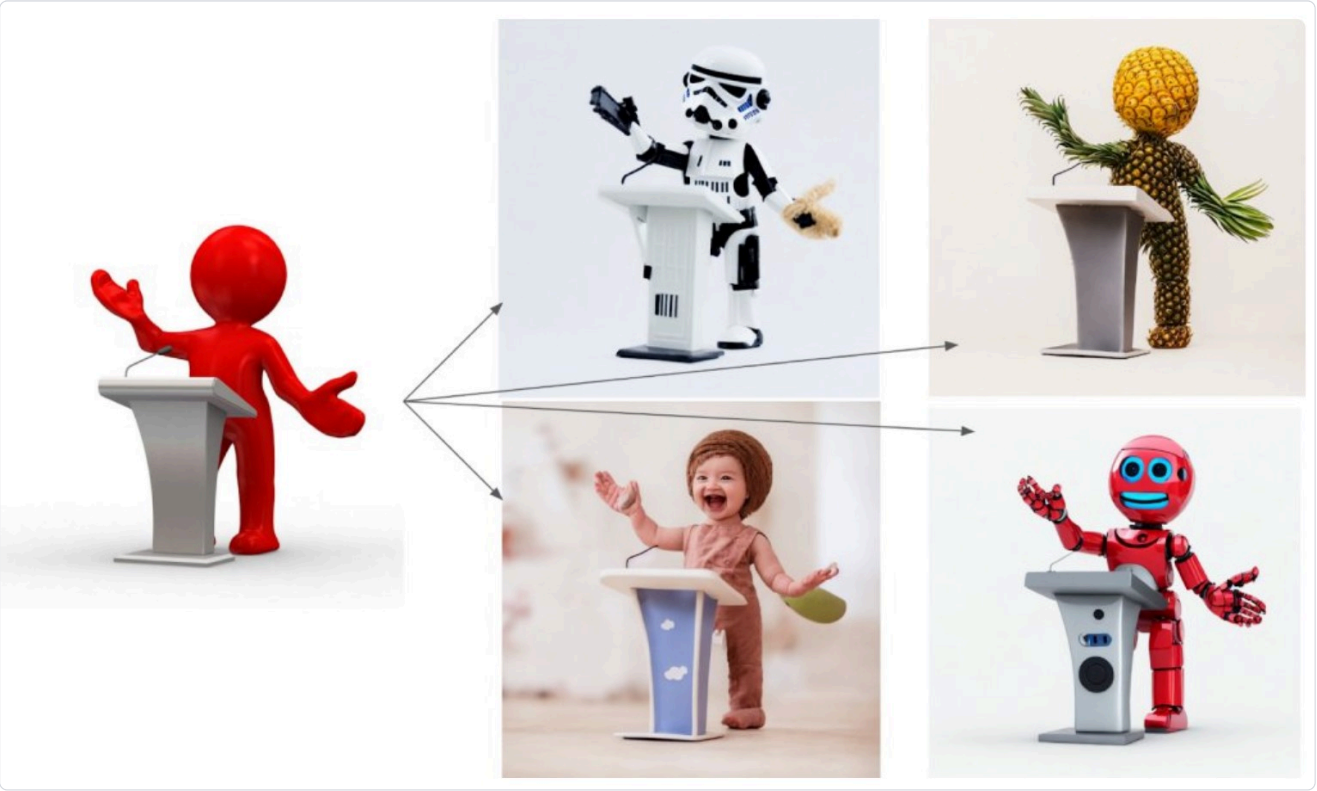


Prompt: a portrait of a beautiful blonde woman, fine - art photography, soft portrait shot 8 k, mid length, ultrarealistic uhd faces, unsplash, kodak ultra max 800, 85 mm, intricate, casual pose, centered symmetrical composition, stunning photos, masterpiece, grainy, centered composition : 2 | blender, cropped, lowres, poorly drawn face, out of frame, poorly drawn hands, blurry, bad art, blurred, text, watermark, disfigured, deformed, closed eyes : -2 / Stable Diffusion v2.1-768



Prompt: A Hyperrealistic photograph of ancient Tokyo/London/Paris architectural ruins in a flooded apocalypse landscape of dead skyscrapers, lens flares, cinematic, hdri, matte painting, concept art, celestial, soft render, highly detailed, cgsociety, octane render, trending on artstation, architectural HD, HQ, 4k, 8k / Stable Diffusion v2.1-768





Stable Diffusion XL

La version Stable Diffusion XL, sortie en juillet 2023, représente une avancée majeure dans le domaine de la génération d'images par intelligence artificielle. Avec des améliorations substantielles en termes de résolution, de paramètres, et de compatibilité avec les technologies de prompt, Stable Diffusion XL établit un nouveau standard pour la qualité et le réalisme des images générées.

Caractéristiques Techniques

- **Résolution Doublée** : Avec une résolution impressionnante de 1024 par 1024, Stable Diffusion XL offre une qualité d'image sans précédent, permettant une richesse de détails et une clarté visuelle qui surpassent de loin les versions précédentes.
- **3.5 Milliards de Paramètres** : Le modèle est propulsé par 3.5 milliards de paramètres, une augmentation significative qui contribue à sa capacité à générer des images d'une complexité et d'une variété extraordinaires. Cette capacité étendue permet une meilleure généralisation et une plus grande précision dans la représentation des sujets.

Technologie de Prompt et Applications

- **Compatibilité avec les Prompts Antérieurs** : Stable Diffusion XL est conçu pour être compatible avec les prompts utilisés dans les versions 1.4 et 1.5, facilitant ainsi la transition pour les utilisateurs habitués aux versions précédentes. Cette compatibilité assure une expérience utilisateur fluide et une courbe d'apprentissage réduite.
- **Combinaison de Lion's OpenClip et OpenAI's Clip** : En intégrant à la fois Lion's OpenClip et la technologie de Clip d'OpenAI, Stable Diffusion XL atteint un équilibre parfait entre expressivité des prompts et qualité des résultats. Cette fusion des technologies permet de produire des images à haute résolution qui sont à la fois fidèles aux descriptions textuelles et visuellement impressionnantes.

Exigences Matérielles et Accessibilité

- **Puissance de Calcul Requise** : Il est important de noter que l'exécution de Stable Diffusion XL nécessite un matériel informatique puissant, en raison de la complexité accrue du modèle. Cette exigence peut limiter l'accès à la technologie pour certains utilisateurs, mais pour ceux qui disposent du matériel nécessaire, les résultats sont sans égal.

Impact et Potentiel

Stable Diffusion XL marque une étape importante dans l'évolution de la génération d'images par IA, offrant une qualité et un réalisme qui ouvrent de nouvelles avenues pour la créativité et l'innovation. Que ce soit pour la création artistique, le développement de contenu pour les jeux vidéo, ou l'amélioration des processus de conception dans diverses industries, Stable Diffusion XL offre un potentiel énorme.

En résumé, Stable Diffusion XL incarne l'avenir de la génération d'images par IA, avec ses avancées significatives en termes de résolution, de paramètres, et de compatibilité des prompts. Bien que l'accès à cette technologie puisse être limité par les exigences matérielles, son impact sur le domaine de la création visuelle et au-delà est indéniable. Avec Stable Diffusion XL, nous entrons dans une nouvelle ère de possibilités créatives, où les limites de l'imagination sont constamment repoussées.




Prompt: *✨aesthetic✨ aliens walk among us in Las Vegas, scratchy found film photograph.*

(Left - SDXL Beta, Right - SDXL 0.9)




Prompt: *A wolf in Yosemite National Park, chilly nature documentary film photography.*
Negative prompt: *3d render, smooth, plastic, blurry, grainy, low-resolution, anime, deep-fried, oversaturated.*

(Left - SDXL Beta, Right - SDXL 0.9)





Prompt: **~aesthetic~*~ manicured hand holding up a take-out coffee, pastel chilly dawn beach Instagram film photography.*
Negative prompt: *3d render, smooth, plastic, blurry, grainy, low-resolution, anime.*

(Left - SDXL Beta, Right - SDXL 0.9)



Prompt: *magical realism; manicured fingers holding a piece of white heart-shaped sea glass up against the setting sun realistic film photography.*
(SDXL beta - 480x480)

Stable Diffusion V3

La version 3 de Stable Diffusion, présentée dans un article de recherche publié le 5 mars, marque une nouvelle avancée dans la génération d'images par intelligence artificielle. Cette version se distingue non seulement par ses performances améliorées mais aussi par une innovation architecturale significative.

Innovations Technologiques et Performances

Avancées Clés

- **Publication de l'Article de Recherche** : L'article détaille la technologie sous-jacente qui alimente Stable Diffusion 3, offrant un aperçu des avancées et des améliorations par rapport aux versions précédentes.
- **Surpasse les Systèmes de Génération d'Images Texte-vers-Image** : Stable Diffusion 3 a démontré sa supériorité par rapport à des systèmes de pointe tels que DALL·E 3, Midjourney v6, et Ideogram v1 dans des domaines clés tels que la typographie et l'adhérence aux prompts, selon les évaluations de préférence humaine.

Architecture Multimodale Diffusion Transformer (MMDiT)

- **Séparation des Poids pour les Représentations d'Image et de Langage** : Cette innovation architecturale utilise des ensembles de poids distincts pour les représentations d'image et de langage, ce qui améliore la compréhension du texte et les capacités d'orthographe par rapport aux versions antérieures de SD3.

Comparaison et Évaluation de la Performance

- **Comparaisons avec d'Autres Modèles** : Stable Diffusion 3 a été comparé à divers modèles ouverts et fermés, y compris SDXL, SDXL Turbo, Stable Cascade, Playground v2.5, Pixart-α, DALL·E 3, Midjourney v6, et Ideogram v1, pour évaluer la performance basée sur les retours humains.
- **Critères d'Évaluation** : Les évaluateurs humains ont jugé les résultats des modèles en fonction de la fidélité au prompt donné, de la qualité de la typographie basée sur le prompt, et de la qualité esthétique visuelle globale.
- **Résultats** : Stable Diffusion 3 s'est révélé égal ou supérieur aux systèmes de génération d'images texte-vers-image actuels dans tous ces domaines.

Accessibilité et Exigences Matérielles

- **Tests d'Inférence sur Matériel Consommateur** : Dans les tests d'inférence précoces et non optimisés, le plus grand modèle SD3 avec 8 milliards de paramètres s'adapte dans les 24 Go de VRAM d'une RTX 4090 et prend 34 secondes pour générer une image de résolution 1024×1024 en utilisant 50 étapes d'échantillonnage.
- **Variations Multiples lors de la Sortie Initiale** : Pour réduire les barrières matérielles, plusieurs variations de Stable Diffusion 3, allant de 800 millions à 8 milliards de paramètres, seront disponibles dès la sortie initiale.



Prompt: Epic anime artwork of a wizard atop a mountain at night casting a cosmic spell into the dark sky that says "Stable Diffusion 3" made out of colorful energy



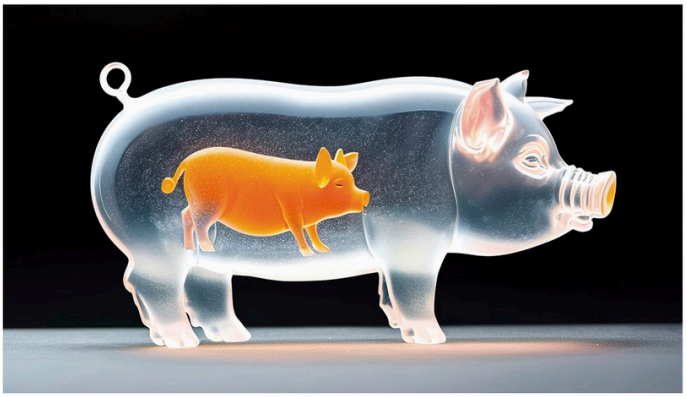
Prompt: A surreal and humorous scene in a classroom with the words 'GPUs go brrrrrr' written in white chalk on a blackboard. In front of the blackboard, a group of students are celebrating. These students are uniquely depicted as avocados, complete with little arms and legs, and faces showing expressions of joy and excitement. The scene captures a playful and imaginative atmosphere, blending the concept of a traditional classroom with the whimsical portrayal of avocado students.



Prompt: Beautiful pixel art of a Wizard with hovering text "Achievement unlocked: Diffusion models can spell now"



Prompt: Frog sitting in a 1950s diner wearing a leather jacket and a top hat. on the table is a giant burger and a small sign that says "froggy fridays"



Prompt: Translucent pig, inside is a smaller pig.



Prompt: A massive alien space ship that is shaped like a pretzel.



Prompt: A kangaroo holding a beer, wearing ski goggles and passionately singing silly songs.



Prompt: An entire universe inside a bottle sitting on the shelf at walmart on sale.



Prompt: A cheeseburger with juicy beef patties and melted cheese sits on top of a toilet that looks like a throne and stands in the middle of the royal chamber.



Prompt: This dreamlike digital art captures a vibrant, kaleidoscopic bird in a lush rainforest



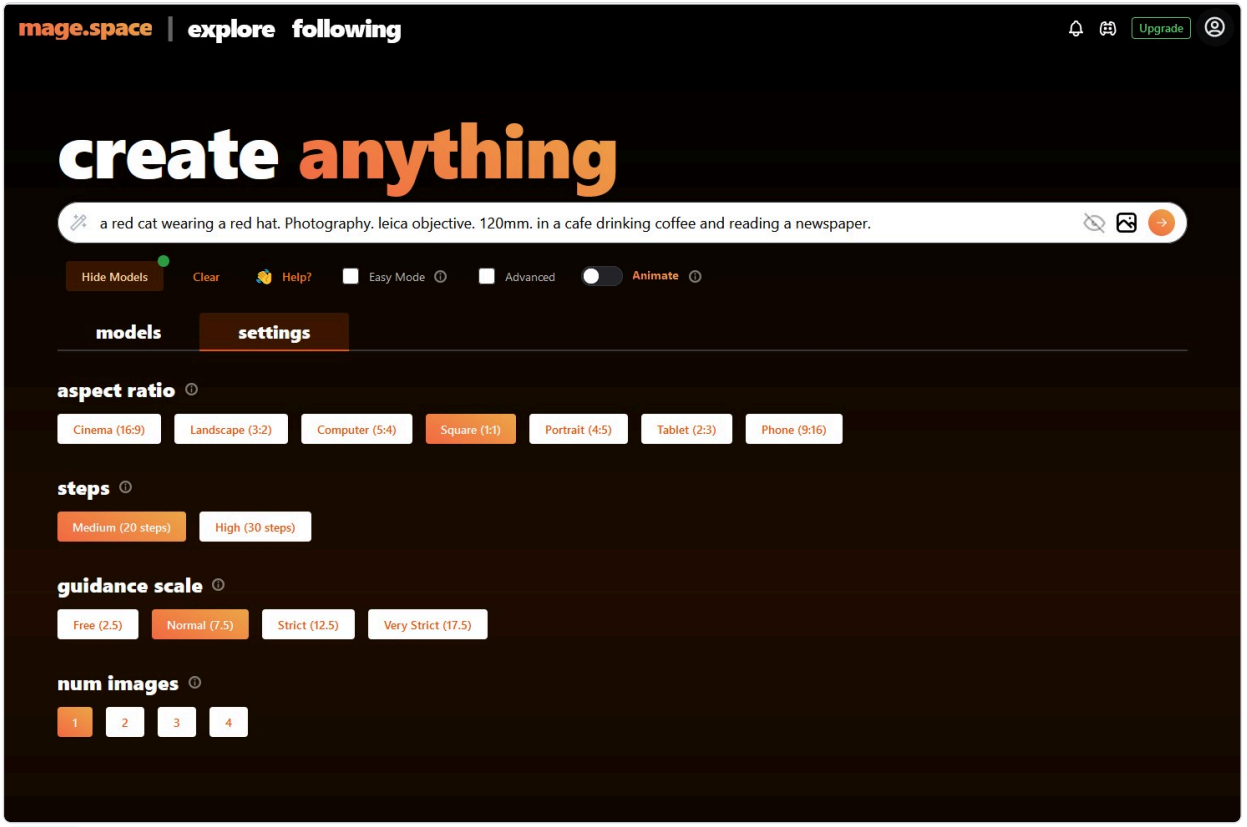
Prompt: A car made out of vegetables.



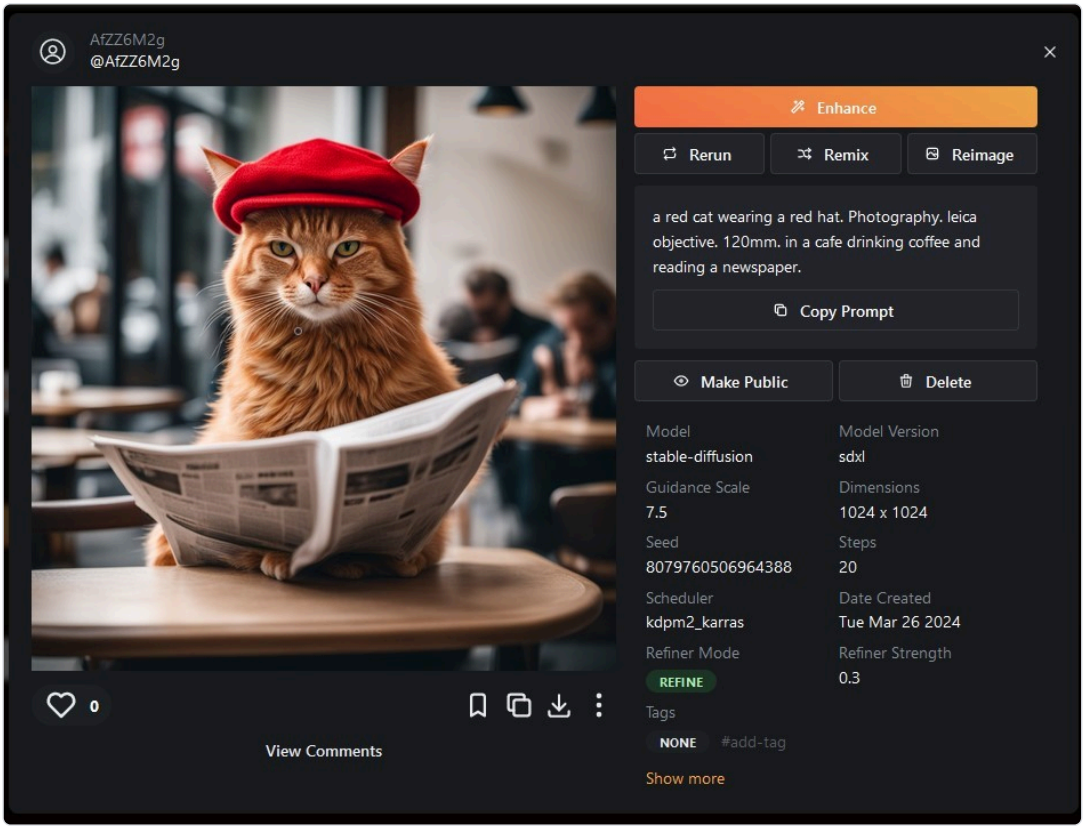
Prompt: Heat death of the universe line art

Mage.space

Mage.space est une application qui permet d'utiliser Stable Diffusion XL gratuitement.



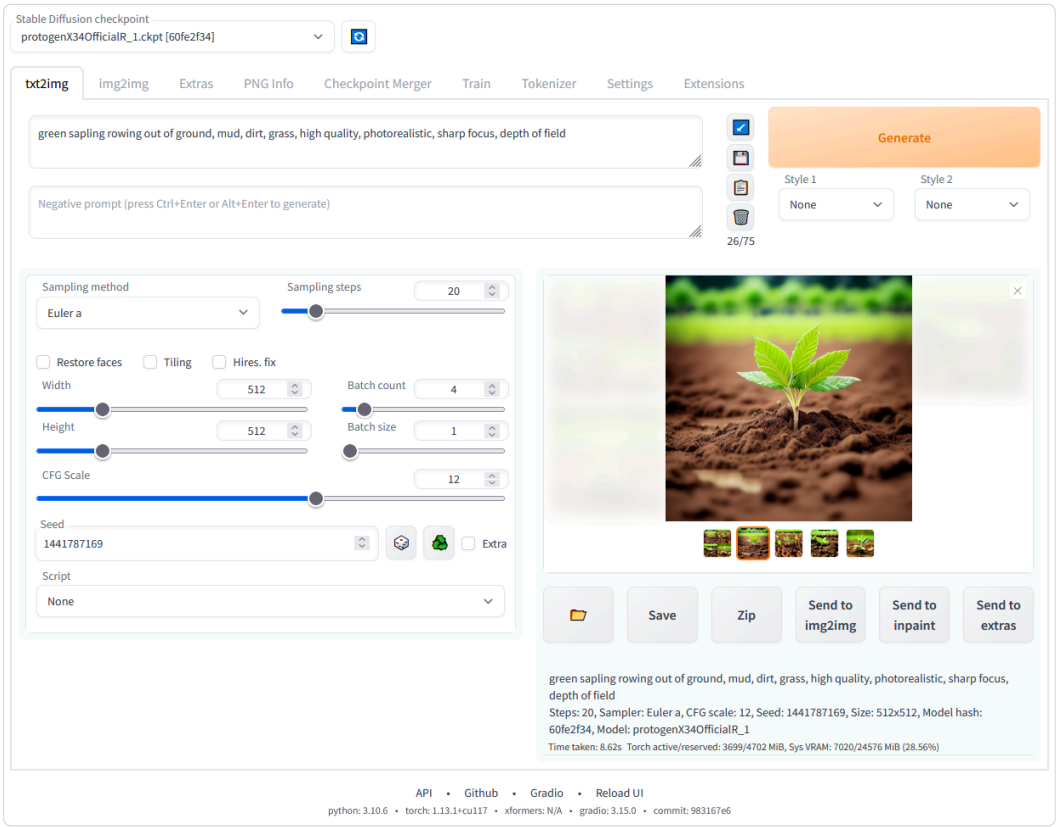
- les `steps` représentent le nombres d'étapes pour générer l'image. Plus le nombre est haut, meilleure sera la qualité. La génération prendra aussi plus de temps.
- Plus la `guidance scale` est haute, plus Stable Diffusion adhérera à votre prompt.



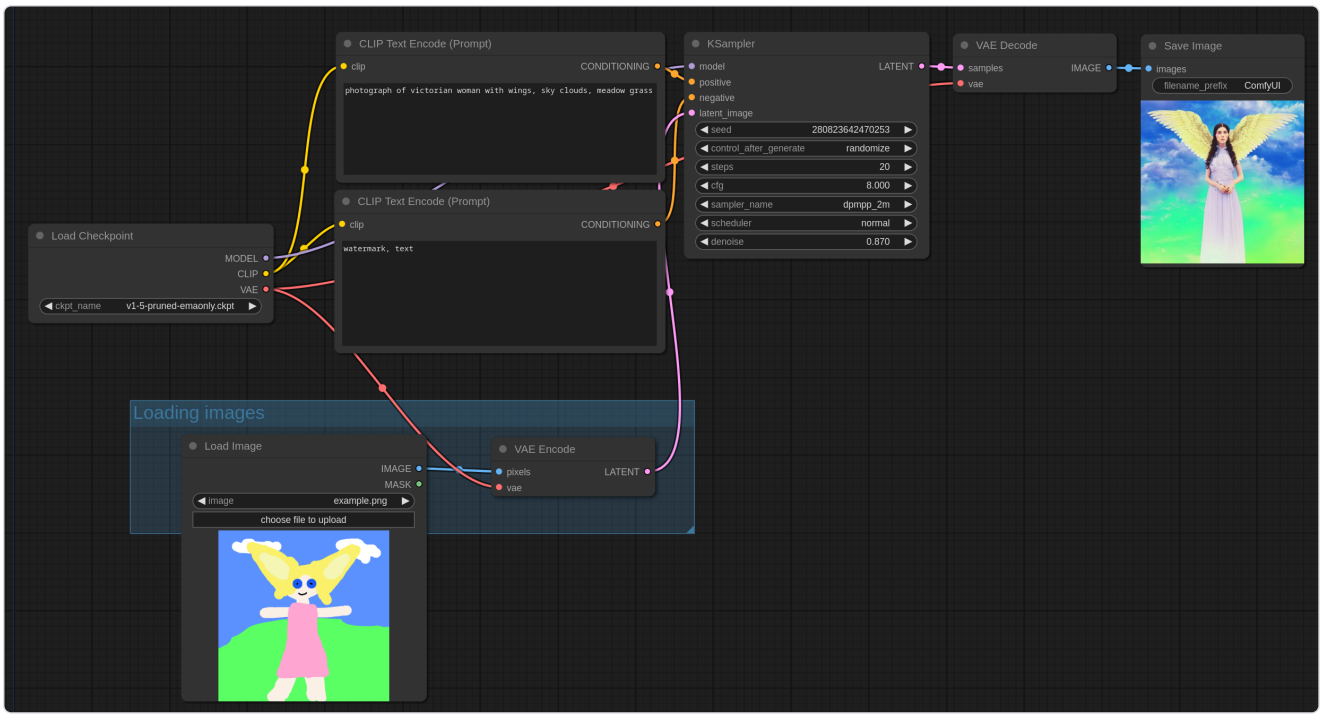
Installation locale

Vous pouvez installer ces logiciels et des modèles Stable Diffusion sur votre ordinateur. Il est conseillé d'avoir une carte graphique puissante pour générer des images de bonne taille à une vitesse correcte.

https://github.com/AUTOMATIC1111/stable-diffusion-webui



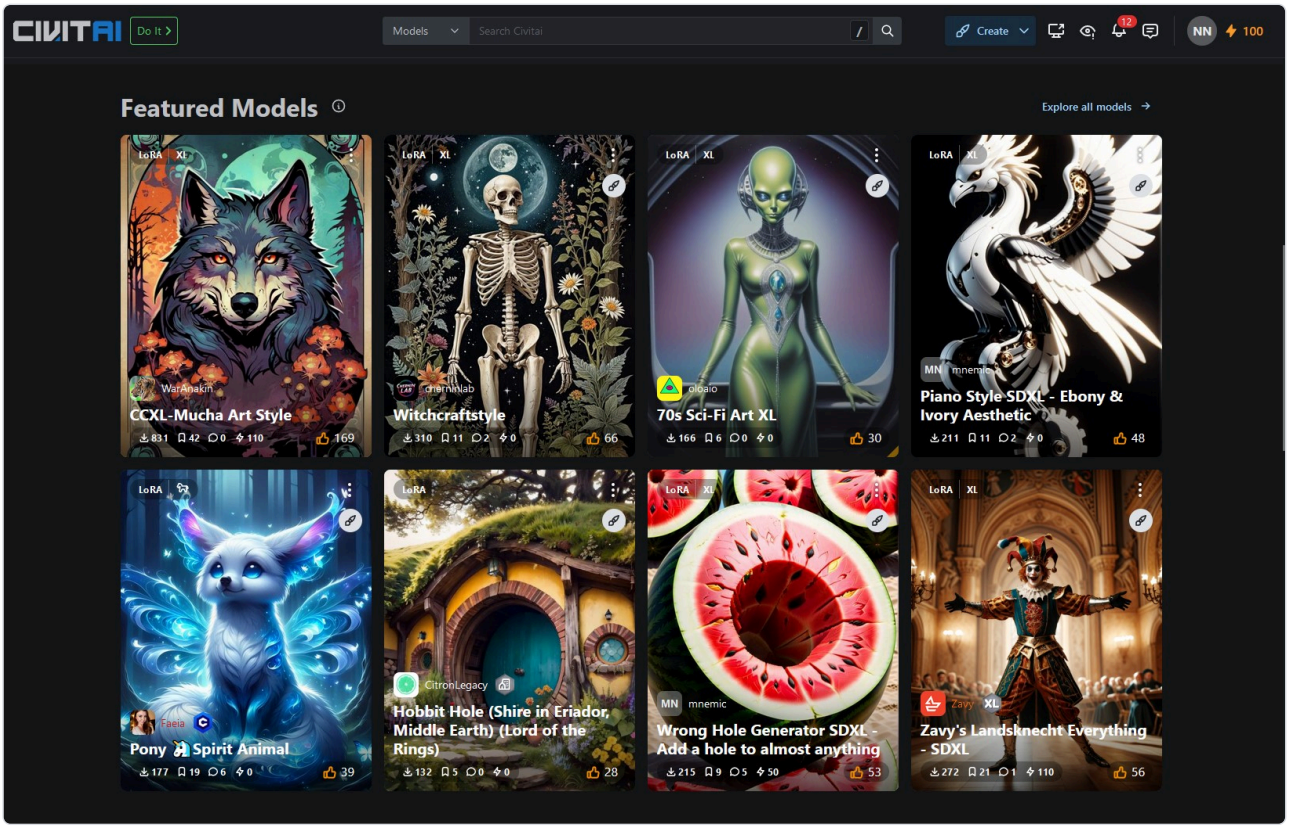
https://github.com/comfyanonymous/ComfyUI



Modèles créés par la communauté

Il est possible de continuer l'entrainement de stable diffusion pour lui apprendre à mieux générer certaines images. Il est possible de télécharger des modèles spécifiques créés par la communauté.

https://civitai.com/

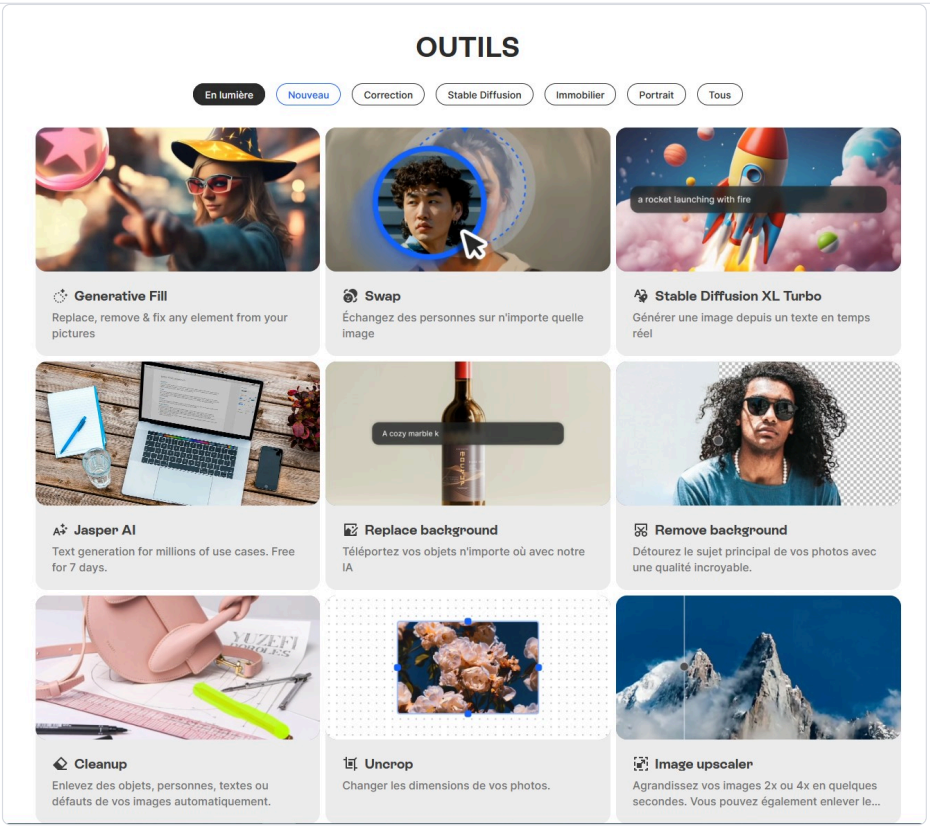


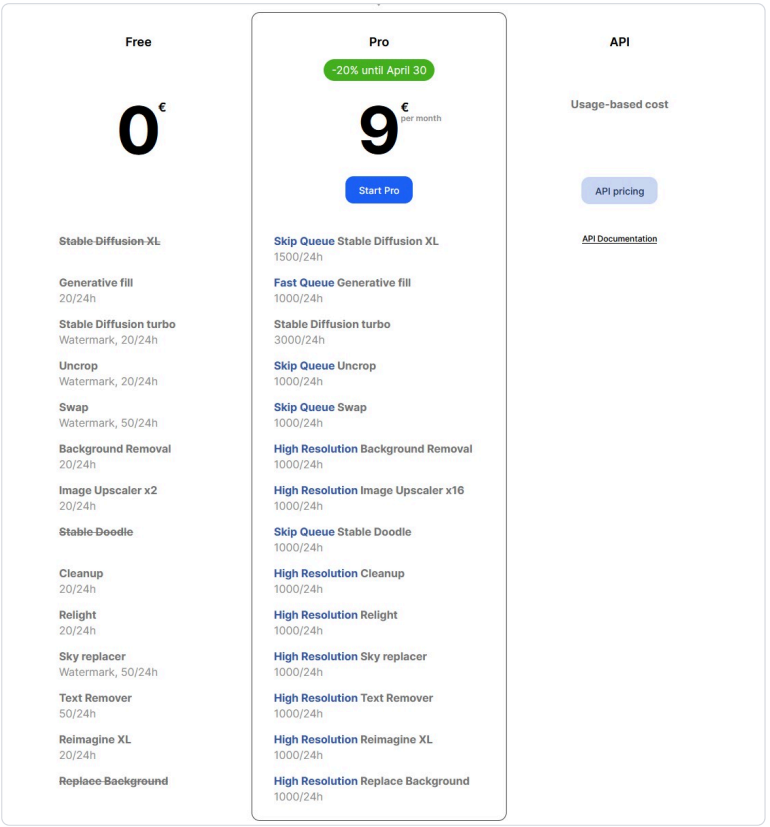
3.3 Autres outils

Dans ce chapitre, nous allons explorer une sélection d'outils de traitement d'image propulsés par des intelligences artificielles génératives, disponibles sur le site ClipDrop. Avant de pouvoir utiliser ces outils innovants, il est nécessaire de s'inscrire sur le site. Ces outils offrent une gamme variée de fonctionnalités, allant de la modification de contenu d'image à l'amélioration de la qualité des images. Chaque outil a ses

spécificités et répond à des besoins particuliers en matière de traitement d'image.


Clipdrop by Jasper
Outils
API
Jasper Free Trial
Tarifs
Mon compte
CRÉEZ DES VISUELS ÉPOUSTOUFLANTS
EN QUELQUES SECONDES
relight

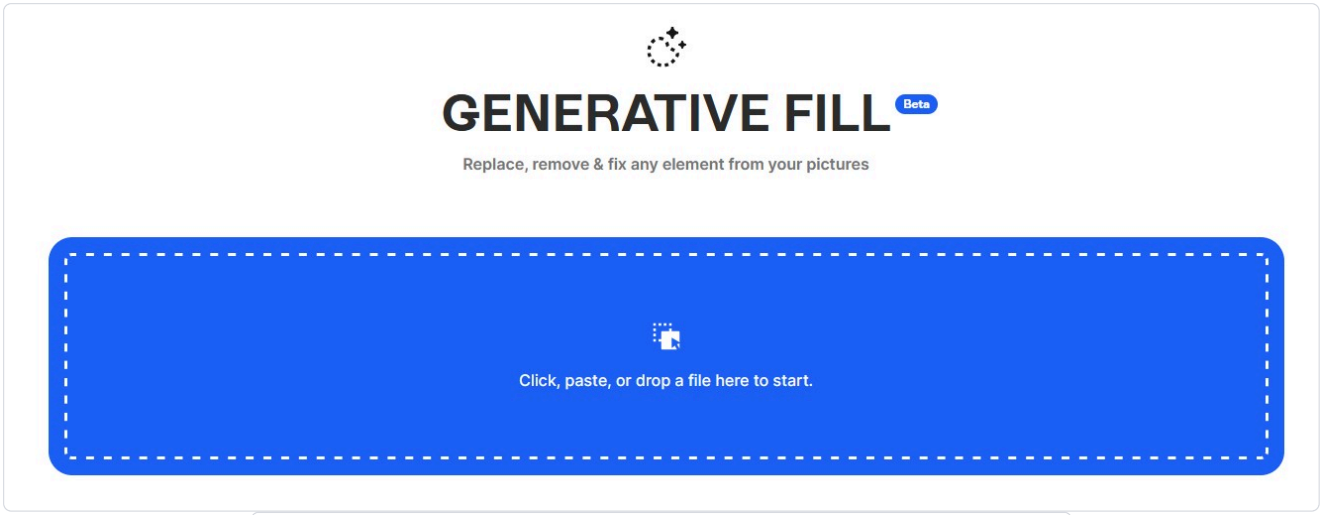

OUTILS
En lumière
Nouveau
Correction
Stable Diffusion
Immobilier
Portrait
Tous
a rocket launching with fire
Generative Fill
Replace, remove & fix any element from your pictures
Swap
Échangez des personnes sur n'importe quelle image
Stable Diffusion XL Turbo
Générer une image depuis un texte en temps réel
A cozy marble k
Jasper AI
Text generation for millions of use cases. Free for 7 days.
Replace background
Téléportez vos objets n'importe où avec notre IA
Remove background
Détourez le sujet principal de vos photos avec une qualité incroyable.
Cleanup
Enlevez des objets, personnes, textes ou défauts de vos images automatiquement.
Uncrop
Changer les dimensions de vos photos.
Image upscaler
Agrandissez vos images 2x ou 4x en quelques secondes. Vous pouvez également enlever le...

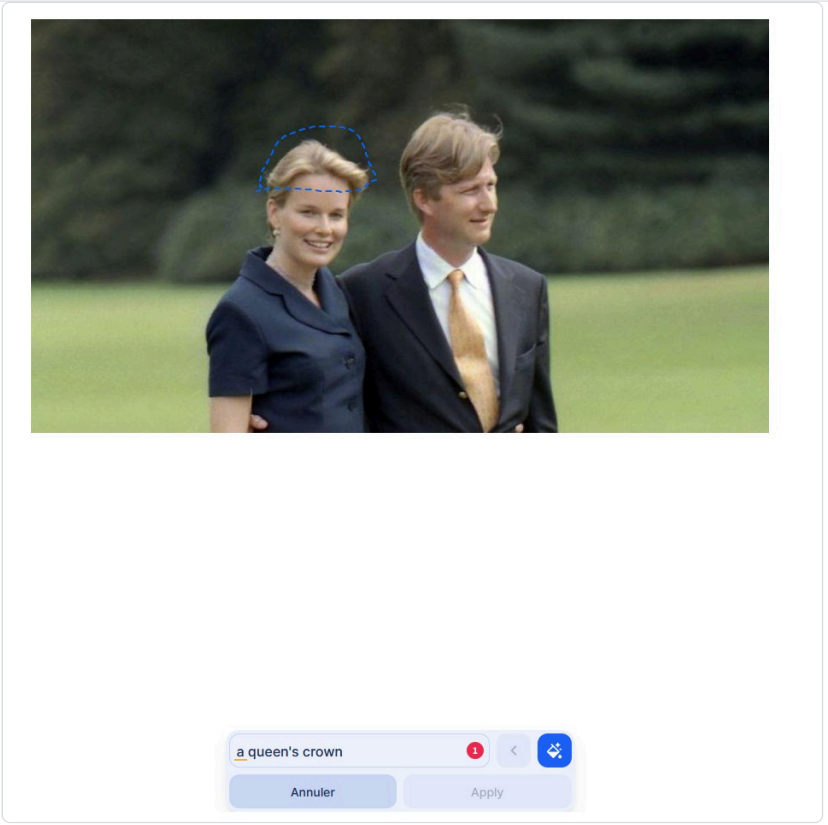


Generative Fill

Outil d'Inpainting

Generative Fill est un outil d'inpainting qui permet de remplir une zone définie dans une image en utilisant un prompt textuel. Cela signifie que l'utilisateur peut dessiner une zone à modifier avec la souris et ensuite décrire avec des mots ce qu'il souhaite voir apparaître dans cette zone. Cet outil utilise la puissance des IA génératives pour interpréter les prompts et générer du contenu visuel correspondant.


GENERATIVE FILL
Beta
Replace, remove & fix any element from your pictures
Click, paste, or drop a file here to start.


a queen's crown
1
Annuler
Apply



Uncrop

Outil d'Outpainting

Uncrop est un outil d'outpainting qui permet de générer les parties cachées d'une image. Si vous avez une photo dont une partie est manquante ou que vous souhaitez étendre, Uncrop utilise l'IA pour imaginer et ajouter la partie manquante de manière cohérente avec le reste de l'image.

UNCROP

Changer les dimensions de vos photos.

Cliquez, collez ou déposez un fichier ici pour commencer.



Sur mesure 2048 × 2048 Suivant





Swap

Échange de visages

Swap est un outil permettant d'échanger les visages de personnes sur des photos. Cette fonctionnalité peut être utilisée pour des créations amusantes ou pour des besoins spécifiques en matière de design et de présentation.

SWAP

Échanger des personnes sur n'importe quelle image

Cliquez, collez ou déposez un fichier ici pour commencer.





Background Removal

Suppression du fond

Background Removal est un outil dédié à la suppression du fond d'une image. Que vous souhaitiez isoler un objet ou une personne pour une utilisation dans un autre contexte ou simplement créer une image avec un fond transparent, cet outil répondra à vos besoins.

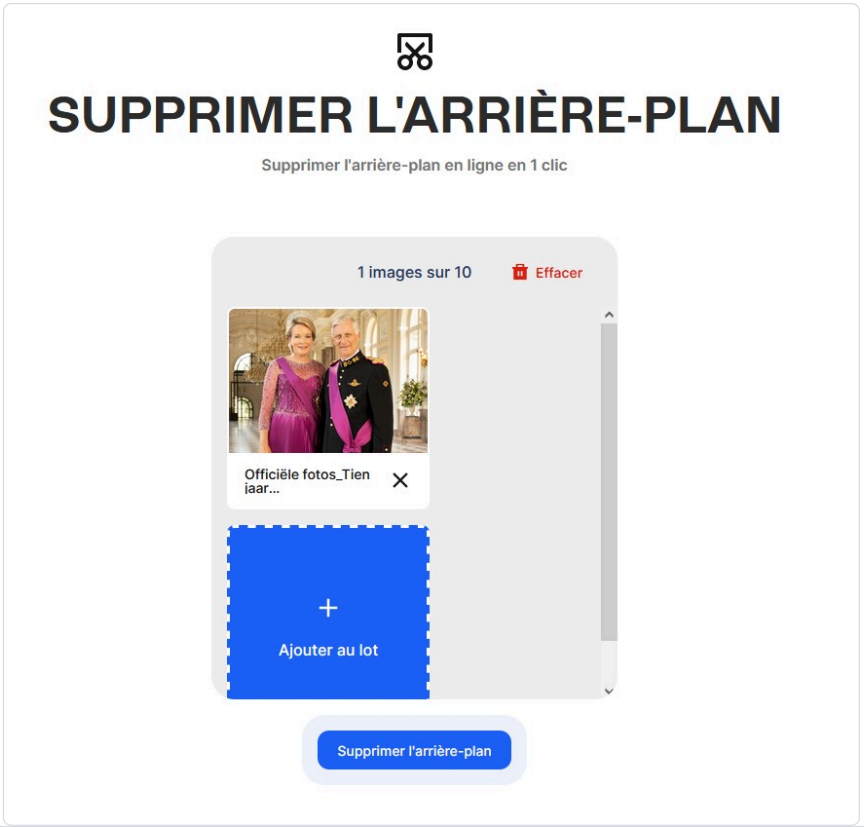
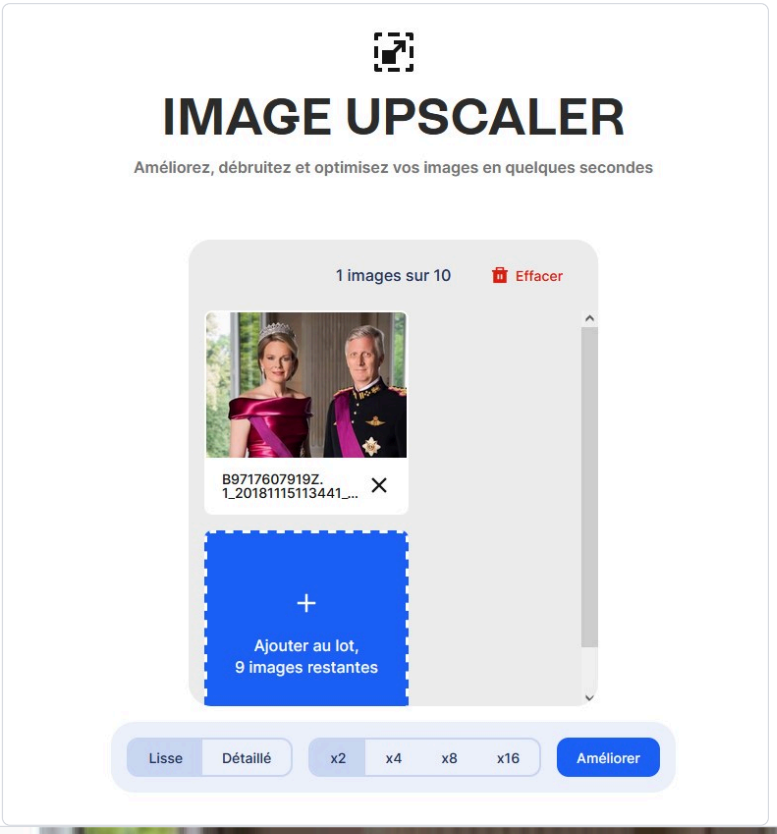




Image Upscaler x2

Amélioration de la qualité d'une image

Image Upscaler x2 est conçu pour améliorer la qualité d'une image et l'agrandir. Cet outil est particulièrement utile pour travailler avec des images de faible résolution que vous souhaitez utiliser dans un contexte nécessitant une meilleure qualité.


IMAGE UPSCALER
Améliorez, débruitez et optimisez vos images en quelques secondes
1 images sur 10
Effacer
B9717607919Z.
1_20181115113441_...
Ajouter au lot,
9 images restantes
Lisse
Détaillé
x2
x4
x8
x16
Améliorer





Cleanup

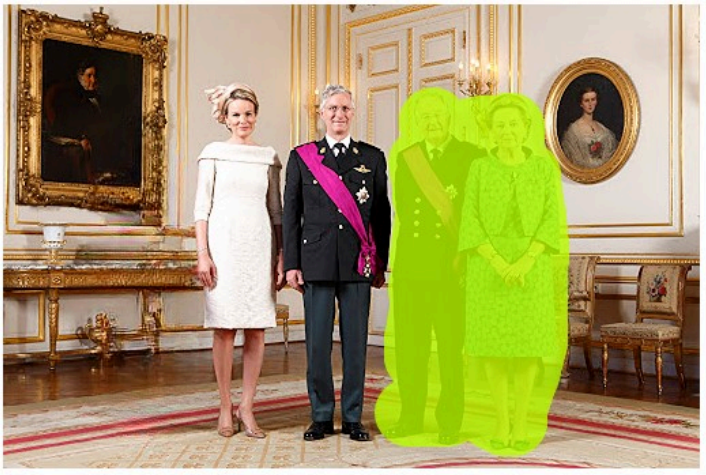
Suppression d'objets, défauts, ou textes

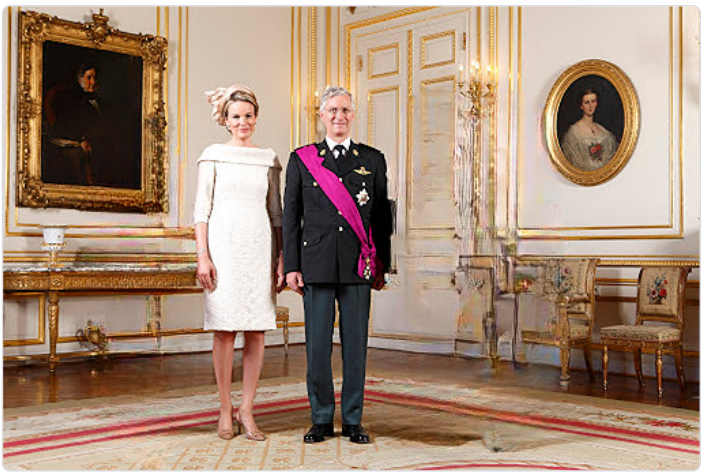
Cleanup permet de supprimer les objets, les défauts, les personnes ou le texte de vos photos en quelques secondes. Cet outil utilise l'IA pour identifier et retirer les éléments indésirables, offrant une image nette et conforme à vos attentes.







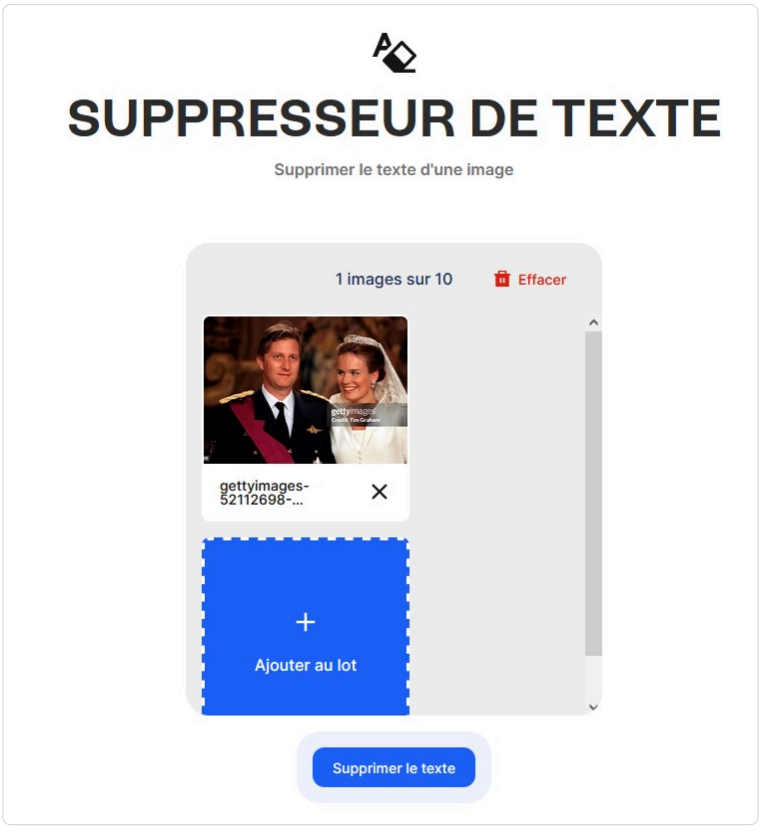




Watermark Remover & Text Remover

Suppression de filigranes et de textes

Watermark Remover et Text Remover sont des outils spécifiques pour supprimer respectivement les filigranes et les textes sur les images. Que vous ayez besoin de nettoyer une image pour des raisons esthétiques ou pour respecter des droits d'auteur, ces outils fournissent une solution rapide et efficace.

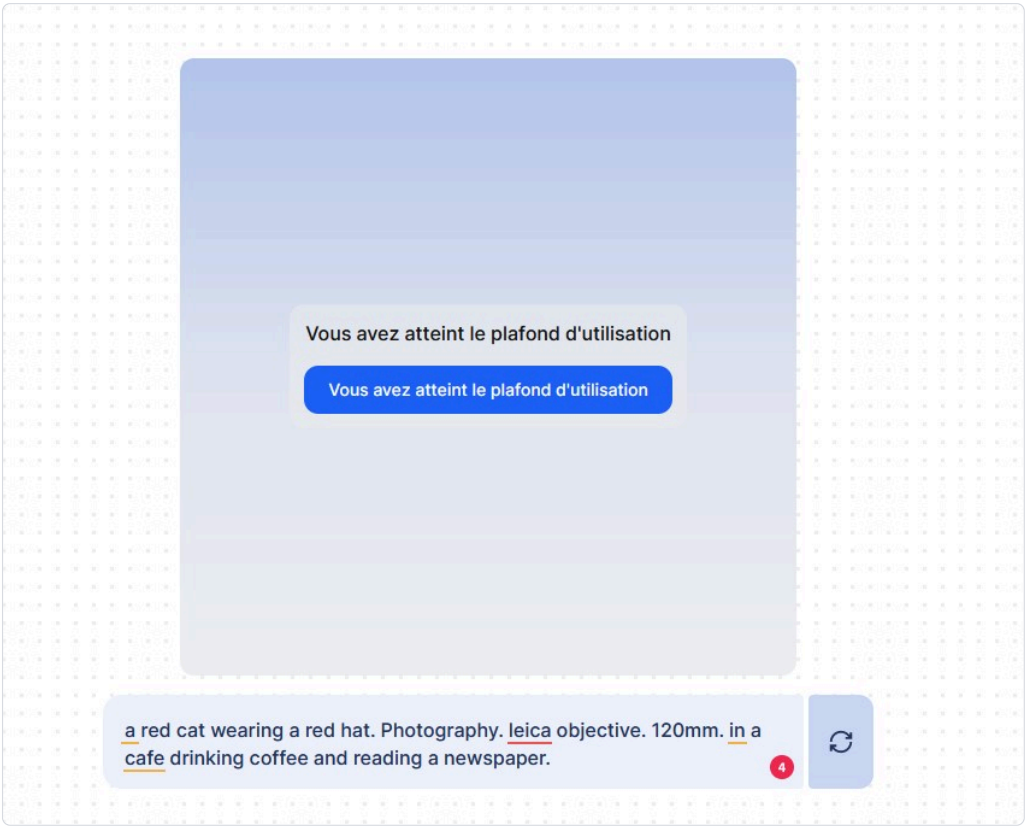




Stable Diffusion Turbo

Génération d'image rapide

Stable Diffusion Turbo est une version optimisée de l'outil de génération d'image Stable Diffusion, conçue pour fournir des résultats plus rapidement. Cet outil permet de créer des images à partir de descriptions textuelles, offrant ainsi une grande liberté créative.



3.4 Conseils pour obtenir de bonnes images

La création de prompts efficaces est cruciale pour obtenir des images de haute qualité avec les outils de génération d'image. Cette première partie vous guidera à travers les étapes essentielles et vous fournira des exemples pour affiner vos prompts.

Utilisation de la langue

L'anglais est la langue principalement utilisée par les outils de génération d'image. Cependant, certains outils, comme DALL·E et GPT, sont capables de traduire vos demandes à partir d'autres langues. Si vous n'êtes pas à l'aise avec l'anglais, n'hésitez pas à utiliser Google Translate ou un autre service de traduction. Par exemple, si vous souhaitez générer une image d'un "château féerique au coucher du soleil", la traduction en anglais serait "fairy tale castle at sunset".

- https://translate.google.com
- https://www.deepl.com/translator

Exhaustivité des mots

L'exhaustivité et la précision dans la description de ce que vous souhaitez générer sont essentielles. N'ayez pas peur d'utiliser une description détaillée et riche. Par exemple, au lieu de simplement demander une "photo de chat", vous pourriez préciser "un chat angora turc blanc aux yeux verts, assis sur un rebord de fenêtre avec des plantes vertes en arrière-plan, éclairé par la lumière douce du matin". Pour DALL·E, gardez à l'esprit que la longueur du prompt est limitée, donc chaque mot doit compter.

Pour enrichir votre vocabulaire, vous pouvez utiliser GPT pour générer des listes de synonymes ou des mots clés pertinents. Par exemple, pour "éclairé par la lumière douce du matin", vous pourriez demander des synonymes comme "baigné dans une lumière matinale douce" ou "illuminé par les premiers rayons du soleil".

Choix du style

Le choix du style influence grandement le résultat final. Voici quelques exemples de styles que vous pourriez spécifier dans votre prompt :

- **Photo** : Pour un rendu réaliste, comme une photographie capturée par un appareil photo.
- **Illustration** : Pour un rendu artistique, comme un dessin ou une peinture.
- **3D Pixar** : Pour un rendu en trois dimensions rappelant les films d'animation de Pixar.
- **Peinture** : Pour imiter un style de peinture spécifique, comme l'impressionnisme ou le surréalisme.

- **Fantaisie** : Pour des scènes imaginaires avec des éléments magiques ou surnaturels.
- **Science-fiction** : Pour des scènes futuristes ou extraterrestres.

En précisant le style, vous guidez l'outil de génération d'image pour qu'il se rapproche au mieux de votre vision. Par exemple, pour une scène de science-fiction, vous pourriez décrire "une ville futuriste sur une planète extraterrestre, avec des bâtiments en forme de dôme et des véhicules volants, sous un ciel violet".

Exemples

Photoréaliste



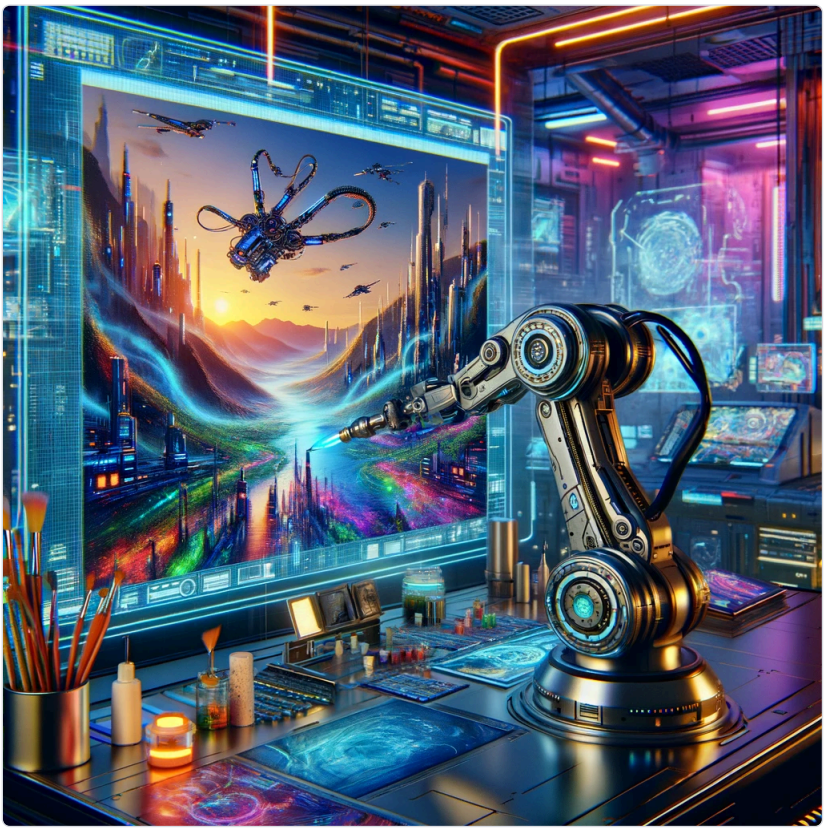
Pointillisme



Fantaisie



Science-fiction



Gestion des couleurs

La palette de couleurs joue un rôle crucial dans l'ambiance et l'esthétique de votre image générée. Voici quelques conseils pour bien choisir vos couleurs :

- **Palette de couleurs** : Définissez si vous souhaitez des couleurs claires, chaudes, pastelles, ou vives. Par exemple, pour une scène apaisante, vous pourriez choisir des couleurs pastel douces. Citez les noms des couleurs que vous voulez voir apparaitre.
- **Contraste et exposition** : Mentionnez le niveau de contraste désiré (faible, moyen, élevé) et l'exposition (sombre, équilibrée, lumineuse). Cela peut aider à créer l'atmosphère souhaitée. Par exemple, "une forêt enchantée sous une lumière de lune, avec un contraste élevé entre les ombres profondes et les rayons lumineux".



> *Une bibliothèque ancienne, baignée dans une lumière douce de fin d'après-midi. Les étagères sont remplies de livres anciens aux couvertures en cuir. La palette de couleurs est composée de teintes chaudes, avec des dominantes de brun, d'or et de rouge sombre. Le contraste est moyen, avec une exposition équilibrée, créant une ambiance accueillante et mystique, où la lumière crée des ombres douces sur le sol et les murs.*



> *Une bibliothèque ancienne éclairée uniquement par la lumière de bougies scintillantes, avec des ombres dansantes sur les étagères remplies de grimoires anciens. La palette de couleurs est dominée par des teintes sombres et des touches de rouge et de bleu profond, créant un contraste élevé entre les zones lumineuses directement sous les bougies et les ombres profondes. L'exposition est sombre, renforçant une ambiance de mystère et d'aventure magique.*

Représentation des personnages

La manière dont vous décrivez les personnages peut grandement influencer le résultat. Voici comment procéder :

- **Identification du sujet** : Soyez précis sur le sujet de votre image. Cela peut être un homme, une femme, des animaux, ou des créatures fantastiques. Par exemple, "une guerrière elfe avec des ailes de papillon, tenant une épée lumineuse".
- **Cadrage** : Choisissez comment vous souhaitez que votre sujet soit vu : portrait (focalisé sur le visage), ceinture (du torse à la tête), ou full body (en entier).



> *Portrait d'un alchimiste âgé, avec un regard pénétrant, reflétant son savoir et sa sagesse. Ses cheveux sont longs et légèrement grisonnants, et il porte des lunettes rondes sur le bout du nez. Le fond est flou mais laisse entrevoir des étagères remplies de fioles et de livres anciens. L'éclairage est doux et vient de sa droite, accentuant les textures de sa peau et les détails de son visage marqué par le temps.*



> *Un alchimiste en pleine action dans son laboratoire, vêtu d'une longue robe avec des motifs ésotériques, tenant une fiole lumineuse dans une main et un grimoire ancien dans l'autre. Autour de lui, le laboratoire est en désordre, avec des étagères chargées de livres, de fioles et d'instruments alchimiques. Des éclairs de lumière émanent de la fiole, illuminant la scène d'une lumière surnaturelle. La scène capture l'alchimiste de plein pied, dans une posture dynamique, exprimant sa passion et son engagement dans l'expérimentation.*

Postures

La posture du sujet ajoute de la dynamique à l'image :

- **Positionnement** : Décrivez la posture du sujet. Par exemple, "assis sur une chaise, les mains posées sur les genoux, regardant pensivement au loin" ou "debout, bras écartés, face à un coucher de soleil".



> *Un explorateur assis sur un rocher au milieu d'une jungle dense, les mains posées sur les genoux, regardant pensivement au loin. Autour de lui, des ruines anciennes envahies par la végétation et des rayons de lumière filtrant à travers les feuilles épaisses. La posture reflète un moment de réflexion solitaire, capturant à la fois la grandeur de la nature et la quête intérieure de l'explorateur.*



> *Un explorateur debout sur le sommet d'une colline, bras écartés, face à un coucher de soleil magnifique qui illumine une vallée cachée et des ruines antiques. La posture exprime un sentiment de triomphe et de découverte, soulignant la majesté du moment et l'aboutissement d'une longue quête.*

Accessoires

Les accessoires peuvent enrichir votre scène :

- **Éléments supplémentaires** : N'oubliez pas d'ajouter des détails sur les vêtements, les bijoux, les armes, ou tout autre accessoire pertinent. Par exemple, "portant une cape rouge écarlate, ornée de runes d'argent".



> *Un mage debout dans une bibliothèque magique, parcourant un grand livre ouvert posé sur un pupitre. La bibliothèque est remplie d'étagères de livres anciens et émet une douce lumière mystique.*



> *Un mage debout dans une bibliothèque magique, parcourant un grand livre ouvert posé sur un pupitre. Il porte une cape rouge écarlate, ornée de runes d'argent, et un pendentif en cristal lumineux autour du cou, émettant une faible lueur. Des gants finement travaillés recouvrent ses mains, et à sa ceinture, pendent divers instruments magiques et fioles. La bibliothèque, remplie d'étagères de livres anciens, émet une douce lumière mystique, avec des globes flottants éclairant la pièce.*

Qualité de l'image

Pour obtenir une image de haute qualité, considérez les points suivants :

- **Proximité du sujet** : Un sujet rapproché tend à être plus détaillé. Mentionnez si vous souhaitez que le sujet soit au premier plan.
- **Amélioration par les détails** : Ajoutez des descriptions détaillées pour chaque élément important. Par exemple, "yeux d'un bleu profond, semblables à des saphirs" ou "cheveux d'or, brillant sous la lumière du soleil".
- **Enrichissement de la scène** : Ajoutez des éléments de fond, des effets atmosphériques ou des détails architecturaux pour donner de la profondeur à votre image. "Des vignes grimpantes recouvrant les murs du château, avec des lucioles scintillant dans l'air frais du soir".



> *Un archer elfique se tenant dans une forêt ancienne, prêt à décocher une flèche. La scène capte l'archer et la forêt dans une vue générale, montrant l'environnement dense et mystérieux qui l'entoure.*



> *Un archer elfique en gros plan, le visage concentré sur sa cible, révélant ses yeux d'un bleu profond, semblables à des saphirs, et ses cheveux d'or, brillant sous la lumière filtrant à travers les arbres de la forêt ancienne. Les détails de son arc, orné de motifs elfiques, et la flèche, finement taillée, sont clairement visibles. Le sujet est au premier plan, avec la forêt mystérieuse floutée en arrière-plan.*

Choix du lieu

Le lieu est le décor de votre scène :

- **Sélection du lieu** : Que ce soit un lieu imaginaire ou réel, décrivez-le avec autant de détails que possible. "Un château abandonné sur une colline, entouré d'une brume mystérieuse" offre un contexte riche pour votre image.



> *Un paysage étrange et coloré d'une planète lointaine, où le ciel est parsemé de trois soleils de couleurs différentes. La surface est dominée par des formations rocheuses en spirale et des lacs de lave scintillante. Des arbres aux feuilles cristallines et des créatures semblables à des fleurs géantes dotées d'yeux curieux peuplent ce monde, offrant un spectacle visuel hors du commun. Le contraste des couleurs vives contre les ombres profondes crée une atmosphère à la fois étrange et accueillante.*



> *Vue aérienne d'une méga ville futuriste, construite à partir de collages de bâtiments et de monuments célèbres du monde entier. La Tour Eiffel, le Burj Khalifa, et le Colisée sont harmonieusement intégrés dans un paysage urbain dense, avec des voies de circulation aériennes et des véhicules volants. Des gratte-ciel aux façades de verre côtoient des structures traditionnelles, créant une juxtaposition captivante entre le passé, le présent et l'avenir. La ville est baignée dans une lumière douce de crépuscule, mettant en évidence sa diversité architecturale et sa grandeur.*

Termes techniques

L'utilisation de termes techniques peut grandement influencer le rendu de votre image, en apportant une touche de réalisme ou en affinant le style souhaité. Voici quelques exemples de termes à considérer :

- **Appareil photo et objectif** : Mentionnez si vous voulez que l'image imite un certain type de prise de vue, comme "effet bokeh avec une faible profondeur de champ" ou "grand angle pour une vue panoramique".
- **Lumière et éclairage** : La lumière joue un rôle essentiel dans l'ambiance. Vous pouvez spécifier "lumière naturelle du matin, créant des ombres douces" ou "éclairage dramatique avec de forts contrastes".
- **Style de peinture** : Si vous souhaitez que l'image imite un style de peinture particulier, mentionnez-le. Par exemple, "dans le style impressionniste, avec des coups de pinceau visibles" ou "à la manière de Van Gogh, avec des tourbillons et des couleurs vives".
- **Rendu 3D** : Pour un aspect plus moderne ou futuriste, vous pourriez demander "un rendu 3D réaliste avec des textures détaillées et des reflets".

Inspiration artistique

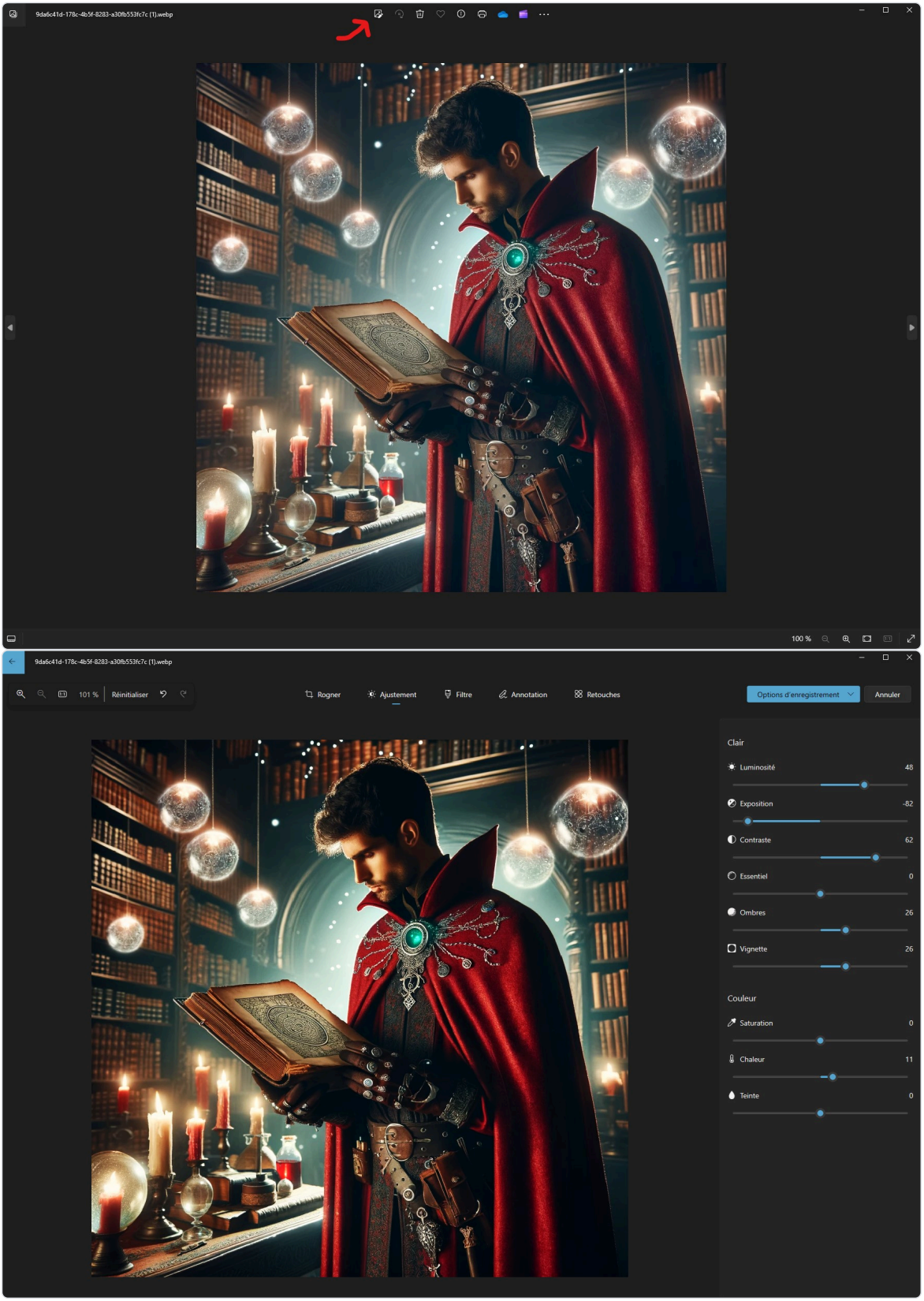
S'inspirer d'artistes ou de styles spécifiques peut donner des résultats uniques et intéressants. Voici comment procéder :

- **Mixer des artistes** : Combinez les styles de plusieurs artistes pour créer quelque chose de nouveau. Par exemple, "une ville futuriste à la croisée des styles de H.R. Giger et de Syd Mead".
- **Imiter avec prudence** : Bien que s'inspirer d'artistes soit enrichissant, veillez à respecter les droits d'auteur et à éviter le plagiat. Utilisez ces inspirations comme point de départ pour développer votre propre vision.

Sélection et retouche

Après avoir généré vos images, le processus ne s'arrête pas là. Voici quelques conseils pour la sélection et la retouche :

- **Garder les meilleures images** : Sélectionnez les images qui correspondent le mieux à votre vision ou qui ont le plus de potentiel artistique.
- **Retouches finales** : Utilisez des logiciels de retouche d'image pour ajuster les couleurs, l'exposition, ou pour ajouter des effets supplémentaires. Par exemple, vous pouvez accentuer les couleurs d'une image pour la rendre plus vibrante ou ajuster la luminosité pour une ambiance plus sombre. Vous pouvez utiliser entre autres l'outil de visionnage d'image de Windows. Il offre quelques possibilités de retouche des couleurs.





Conclusion

En intégrant ces éléments dans vos prompts, vous augmentez vos chances d'obtenir des images qui non seulement répondent à vos attentes mais les surpassent. La clé réside dans l'équilibre entre la précision des détails et la liberté créative laissée à l'outil de génération d'image. Avec la pratique, vous développerez une intuition pour créer des prompts qui produisent des résultats époustouflants.

3.5 Exercices

Photo shoot de mode

Objectifs

L'objectif de cet exercice est de développer votre créativité et votre capacité à manipuler des outils de retouche d'images en simulant un shooting photo de mode.

Instructions

1. **Création de la scène :** Imaginez et concevez une scène pour votre shooting photo. Cela peut être un décor urbain, une plage, un studio avec un fond uni, etc. Notez les éléments clés qui composeront votre scène.
2. **Choix des vêtements et accessoires :** Sélectionnez les vêtements et accessoires que portera votre modèle. Pensez à la cohérence avec le thème de votre scène et à l'harmonie des couleurs.
3. **Choix de la pose :** Définissez la ou les poses que votre modèle adoptera. Vous pouvez vous inspirer de magazines de mode ou de sites spécialisés.
4. **Sélection des meilleures photos :** Une fois la simulation de votre shooting réalisée, choisissez les photos qui représentent le mieux votre vision. Sélectionnez-en au moins trois.
5. **Retouche :** Utilisez un outil de retouche d'images, comme l'outil photo de Windows, pour améliorer vos photos sélectionnées. Vous pouvez ajuster la luminosité, le contraste, ajouter des filtres, etc.
6. **Concours de la tenue la plus extravagante :** Organisez un petit concours avec vos camarades pour élire la tenue la plus extravagante. Partagez vos créations et votez !

Créer une image de paysage fantastique

Objectifs

L'objectif ici est de stimuler votre imagination et de vous familiariser avec les outils de création d'images pour générer un paysage fantastique.

Instructions

1. **Imagination :** Imaginez un paysage fantastique. Cela peut être un monde féerique, une planète extraterrestre, un royaume sous-marin, etc.
2. **Création :** Utilisez un outil de création d'images par IA pour générer votre paysage. Soyez précis dans vos descriptions pour obtenir le résultat souhaité.
3. **Finalisation :** Retouchez l'image générée si nécessaire pour perfectionner les détails et ajuster les couleurs.

Créer un autoportrait IA

Objectifs

Cet exercice vise à explorer la capacité des IA à créer des portraits et à vous voir à travers le "regard" d'une intelligence artificielle.

Instructions

1. **Description :** Écrivez une description détaillée de vous-même, incluant des éléments physiques, mais aussi votre style et votre personnalité.
2. **Génération :** Utilisez un outil de création d'images par IA pour générer votre autoportrait à partir de la description fournie.
3. **Analyse :** Comparez le résultat obtenu avec votre image réelle. Réfléchissez aux similarités, aux différences, et à ce que cela vous révèle sur la perception de l'IA.

Créer une œuvre d'art abstrait

Objectifs

L'objectif est de libérer votre créativité en créant une œuvre d'art abstraite à l'aide d'une IA.

Instructions

1. **Inspiration :** Choisissez une émotion, un concept ou une musique comme source d'inspiration pour votre œuvre.
2. **Description :** Écrivez une description qui capture l'essence de votre inspiration. Soyez aussi descriptif et poétique que possible.
3. **Création :** Utilisez un outil de création d'images par IA pour générer votre œuvre d'art abstrait basée sur la description.
4. **Interprétation :** Une fois votre œuvre créée, prenez un moment pour l'interpréter. Que ressentez-vous en la regardant ? Quelles histoires ou idées vous vient-elle à l'esprit ?

CHAPITRE 4

Musique

4.1 Présentation de Stable Audio 2.0

Introduction

Stable Audio 2.0 marque une avancée significative dans la génération de musique par intelligence artificielle (IA), offrant une plateforme innovante pour créer des compositions musicales complètes. Ce modèle révolutionnaire étend les possibilités de création musicale en permettant la transformation audio-to-audio et text-to-audio à partir d'instructions en langage naturel. Il se distingue par sa capacité à produire des pistes musicales structurées, d'une qualité sonore exceptionnelle.

Développé sur la base de la technologie de diffusion latente, ce modèle succède à Stable Audio 1.0, lancé en septembre 2023 et reconnu pour sa contribution notable à la musique assistée par IA. Avec Stable Audio 2.0, les artistes et musiciens bénéficient d'une flexibilité inédite, d'un contrôle accru et d'une expérience créative enrichie.

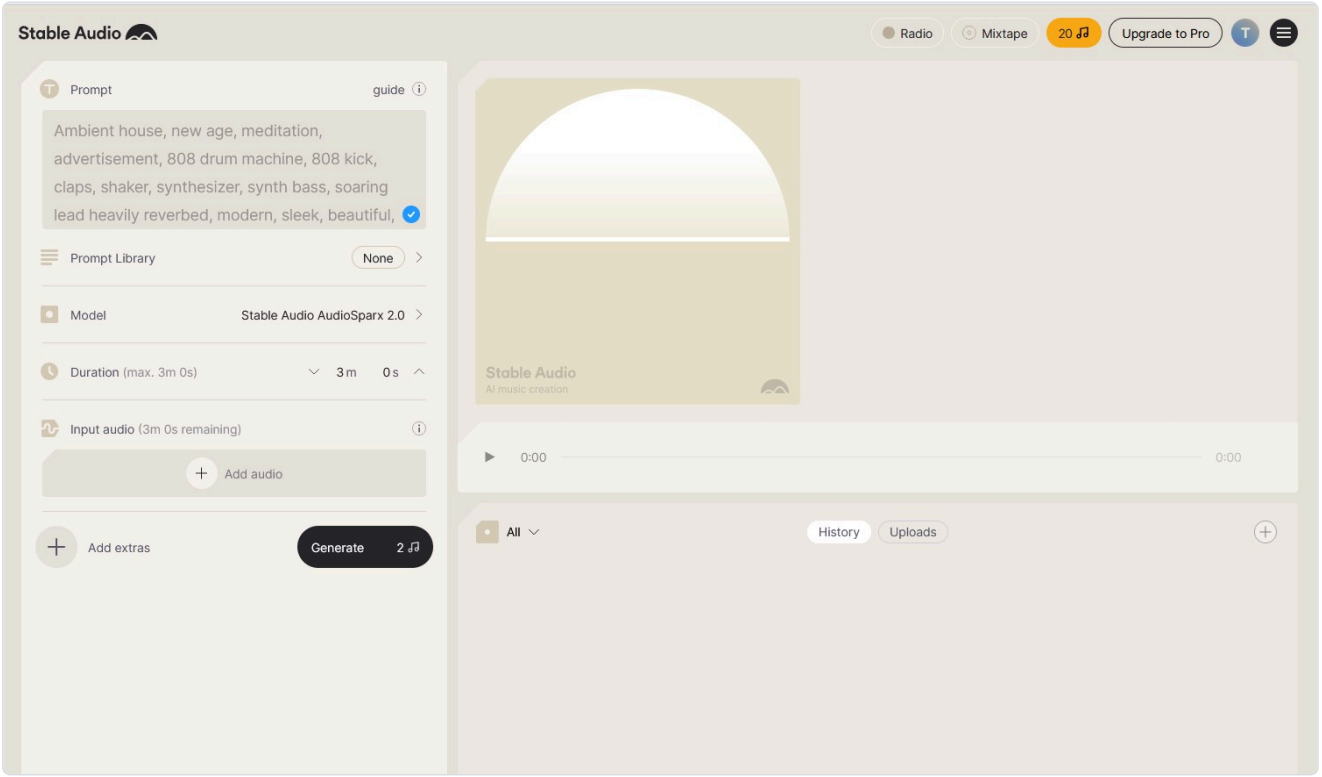
Site officiel :

- https://stableaudio.com/

Caractéristiques Clés

Stable Audio 2.0 innove avec plusieurs fonctionnalités clés :

- **Qualité et Structure** : Génère des compositions jusqu'à trois minutes en stéréo 44.1 kHz, avec une structure musicale comprenant une introduction, un développement, et une conclusion.
- **Transformation Audio** : Permet de transformer des échantillons audio et du texte en musique grâce à des prompts en langage naturel, élargissant les horizons de la création sonore.
- **Respect des Créateurs** : Utilise un dataset sous licence de la bibliothèque musicale AudioSparx, assurant une compensation équitable pour les créateurs.
- **Accessibilité** : Accessible gratuitement sur le site web de Stable Audio, encourageant l'exploration et l'innovation musicale.



Nouveautés de Stable Audio 2.0

Avec son lancement, Stable Audio 2.0 introduit une série de fonctionnalités innovantes qui élargissent le champ des possibilités pour les créateurs de musique. Ces nouveautés renforcent la position du modèle comme un outil de création musicale de pointe, adapté à une variété de besoins artistiques.

Capacités Étendues

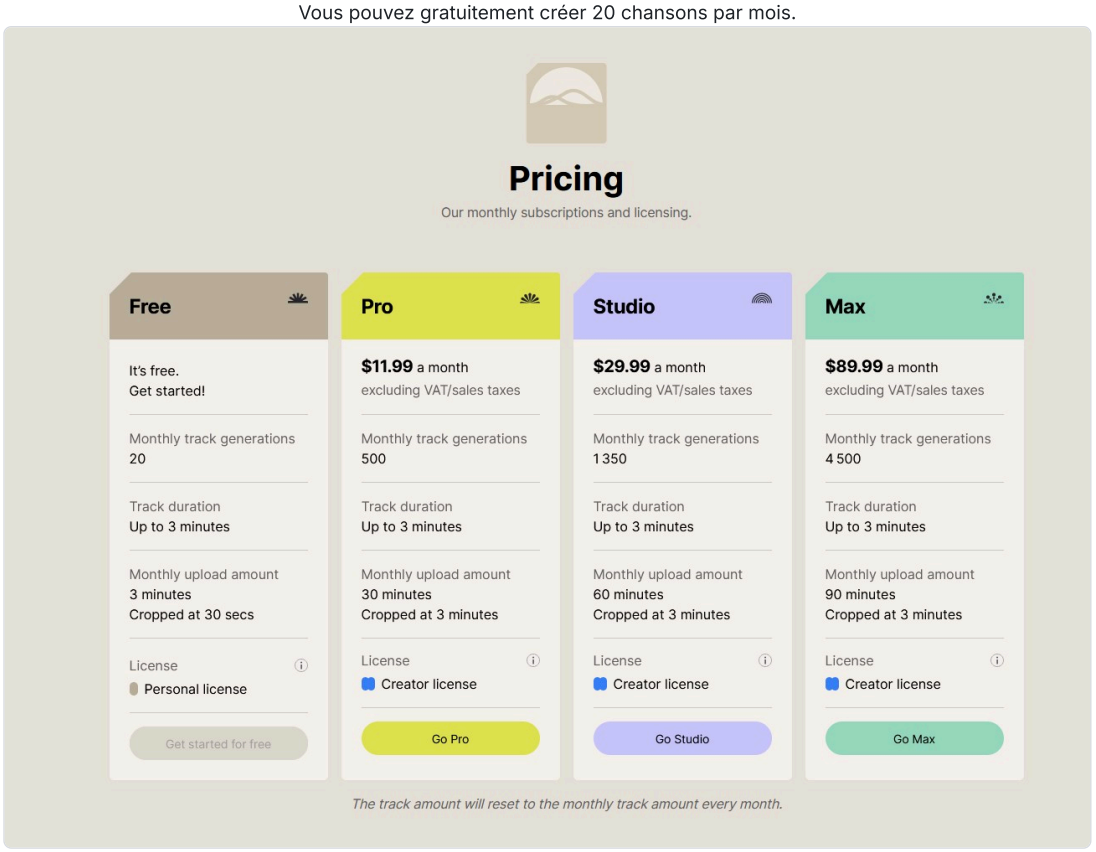
Stable Audio 2.0 enrichit le processus créatif en introduisant des capacités de transformation audio-to-audio et text-to-audio avancées. Cette mise à jour permet aux utilisateurs de télécharger des échantillons audio et de les transformer en une multitude de sons, en utilisant des instructions en langage naturel. Cela ouvre la voie à une exploration sonore inédite, depuis la génération de mélodies et de pistes d'accompagnement jusqu'à la création d'effets sonores et le transfert de style.

Avantages pour les Artistes et Musiciens

La flexibilité et le contrôle accrus offerts par Stable Audio 2.0 invitent à une exploration créative sans précédent. Les artistes et musiciens peuvent désormais produire des œuvres audio avec une plus grande précision stylistique et émotionnelle, poussant les limites de leur expression artistique. Que ce soit pour créer des compositions originales ou pour transformer et réinterpréter des œuvres existantes, Stable Audio 2.0 sert de catalyseur à l'innovation musicale.

Prix

Vous pouvez gratuitement créer 20 chansons par mois.

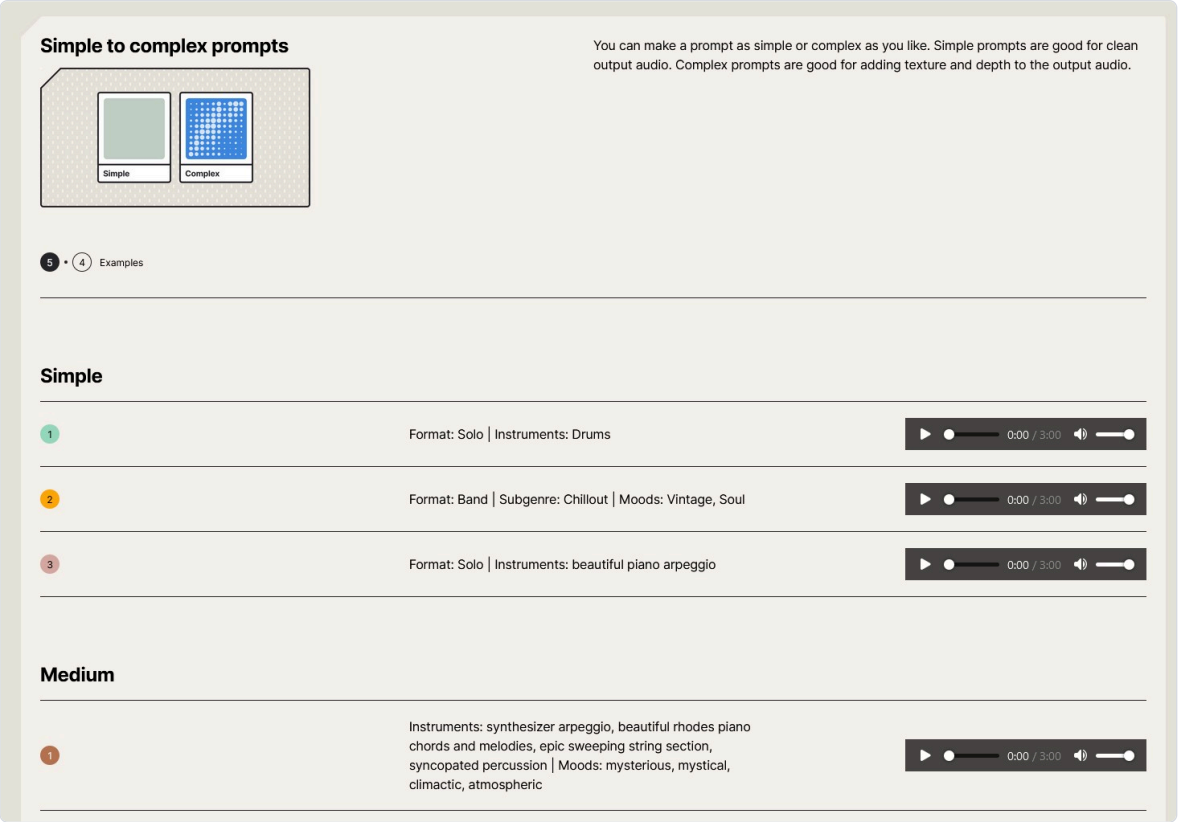


Comment Structurer Vos Prompts de Manière Efficace

Pour maximiser l'efficacité de Stable Audio 2.0, il est important de noter que les prompts doivent être formulés en anglais. Voici un guide détaillé avec des exemples traduits, pour vous aider à formuler vos instructions de manière précise dans chaque catégorie.

Lien vers le guide complet :

- https://stableaudio.com/user-guide/prompt-structure



Format

Le format définit la structure générale de votre composition. Voici quelques formats à considérer :

- Solo
- Duo
- Trio
- Orchestra
- Band

Instruments

Sélectionner les instruments appropriés est crucial. Voici une liste d'instruments en anglais :

- Piano
- Violin
- Electric Guitar
- Drums
- Flute
- Saxophone
- Trumpet
- Harp
- Synthesizer
- Bass

Genre and Sub-genre

Le genre et le sous-genre orientent l'ambiance et l'esthétique musicale. Exemples en anglais :

- **Jazz**
  - Bebop
  - Smooth Jazz
  - Jazz Fusion
- **Rock**

  - Classic Rock
  - Punk Rock
  - Indie Rock
- **Electronic**
  - House
  - Techno
  - Ambient
- **Classical**
  - Baroque
  - Romantic
  - Contemporary

Mood

L'ambiance décrit l'atmosphère émotionnelle de la musique. Quelques ambiances en anglais :

- Inspiring
- Cinematic
- Relaxing
- Joyful
- Mysterious
- Epic
- Melancholic
- Energetic

BPM (Beats Per Minute)

Définir un BPM précis peut orienter le rythme de votre création. Exemples :

- Slow : 60-90 BPM
- Moderate : 91-110 BPM
- Fast : 111-140 BPM
- Very Fast : 141+ BPM

Styles

Les styles ajoutent une dimension à votre musique, spécifiant le contexte ou la méthode de production :

- Acoustic
- Electronic
- Live
- Studio
- Lo-fi
- High-fidelity (Hi-fi)

Exemples de prompts

Pour intégrer ces éléments dans vos prompts en anglais, voici quelques exemples :

1. **Prompt for a Relaxing Atmosphere** :
   - `Format: Solo | Instruments: Harp | Mood: Relaxing | Style: Acoustic | BPM: 80`
2. **Energetic Band Composition** :
   - `Format: Band | Instruments: Electric Guitar, Bass, Drums, Synthesizer | Genre: Rock | Sub-genre: Punk Rock | Mood: Energetic | BPM: 150`
3. **Inspirational Classical Piece** :
   - `Format: Orchestra | Instruments: Violin, Flute, Trumpet, Harp | Genre: Classical | Sub-genre: Romantic | Mood: Inspiring | BPM: 100 | Style: Live`

En suivant ces directives pour structurer vos prompts en anglais, vous pouvez guider Stable Audio 2.0 vers la production de pièces musicales qui correspondent précisément à vos visions artistiques, tout en respectant la nécessité d'utiliser l'anglais pour les instructions.

Sources

- https://stability-ai.squarespace.com/news/stable-audio-2-0

4.2 Suno : Révolutionner la Création Musicale avec l'IA

Qu'est-ce que Suno ?

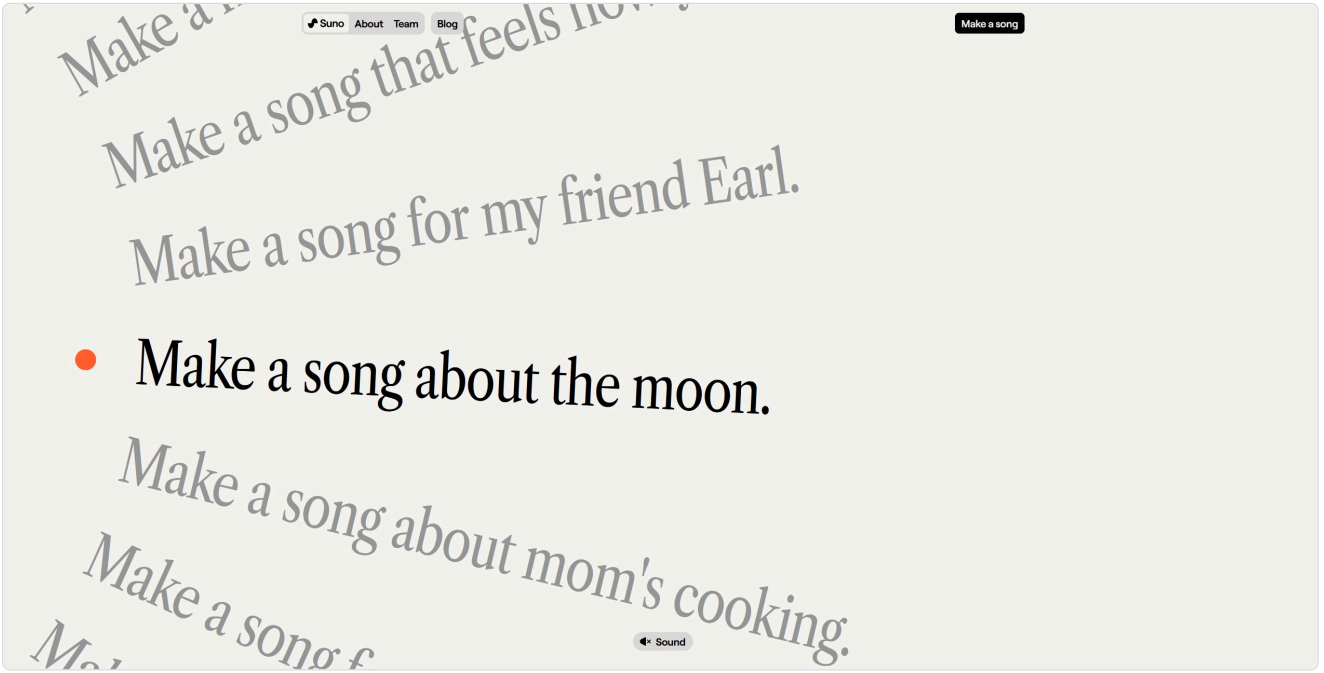
Suno se positionne comme une réponse audacieuse à l'un des grands défis de l'industrie musicale : rendre la création musicale accessible à tous. En s'appuyant sur des algorithmes d'intelligence artificielle avancés, Suno permet à ses utilisateurs de générer des compositions musicales uniques en répondant simplement à quelques prompts.

Caractéristiques Principales

- **Génération IA :** Suno utilise des techniques d'apprentissage profond pour analyser et comprendre la structure musicale, permettant ainsi la création de morceaux dans une variété de styles et de langues.
- **Accessibilité :** Avec un modèle de tarification abordable, Suno vise à ouvrir la porte de la création musicale à un public plus large, allant des amateurs aux professionnels.
- **Diversité et Qualité :** Les utilisateurs de Suno peuvent expérimenter avec différents genres, langues, et même ajuster le mastering de leurs créations, bien que certains aspects soient encore en développement.

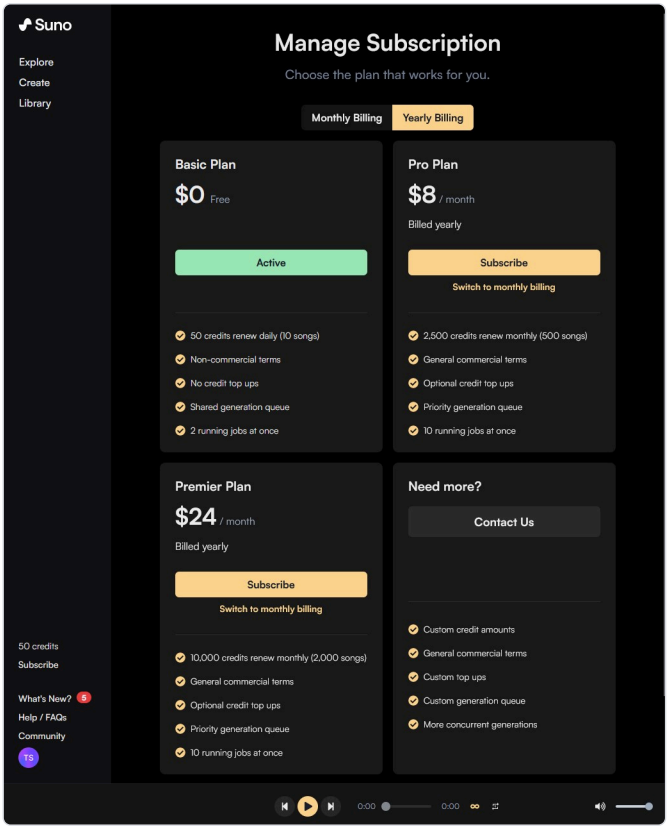
Site officiel :

- https://www.suno.ai/



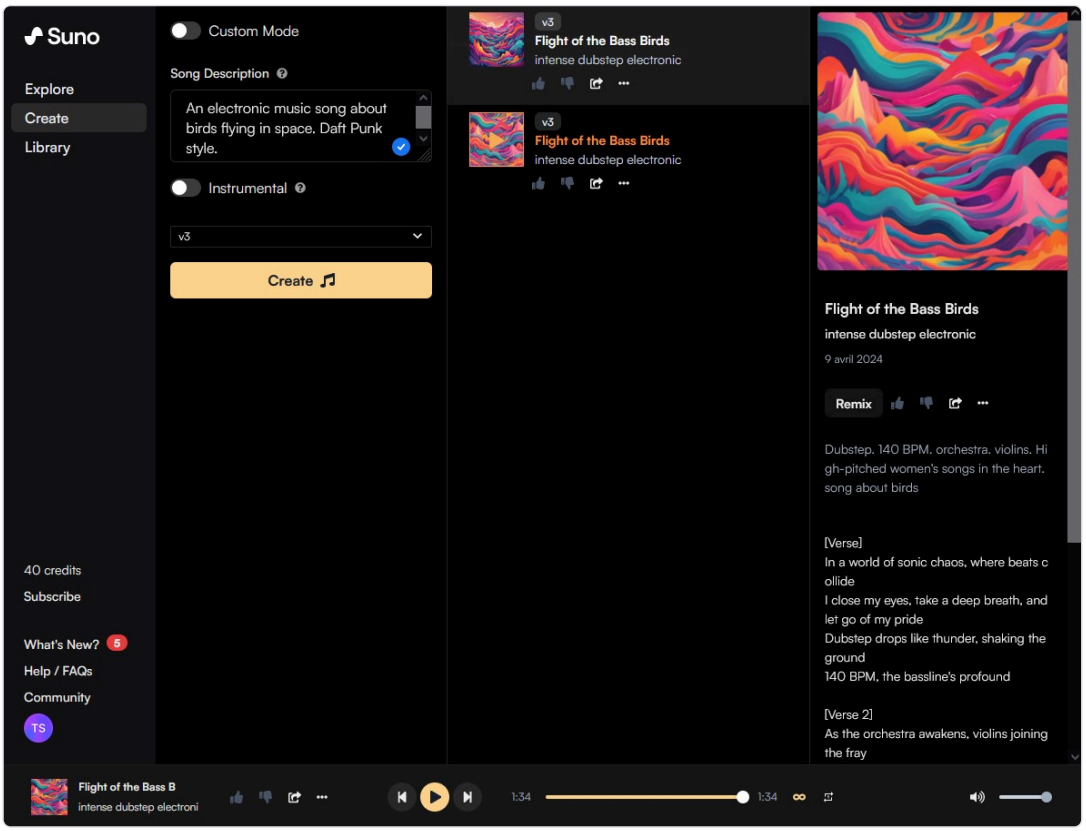
Prix

Vous pouvez créer 10 chansons par jour gratuitement.

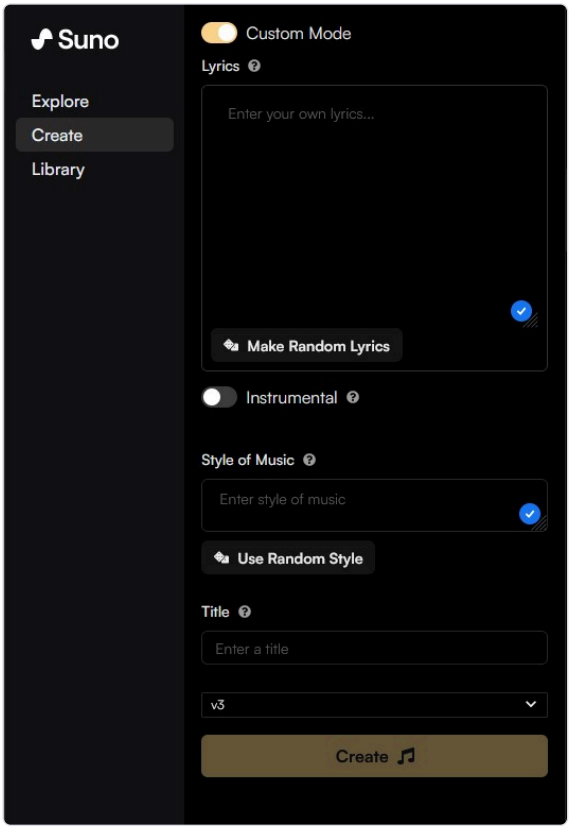


Créer une chanson

Entrez un prompt qui décrit la chanson que vous voulez (genre et thème), en anglais.



Mode de création avancé



Sources

- https://www.suno.ai/blog/v3
- https://en.wikipedia.org/wiki/Suno_AI