

Découvrir l'Intelligence Artificielle

Un guide pas à pas

- 1 Les Bases**
- 2 Le Modèle**
- 3 Le Pilotage**
- 4 Les Limites**

LLM Large Language Model

Vous n'interagissez pas avec un simple Chatbot , mais avec un modèle de langage avancé LLM.

Il ne fonctionne pas comme un socle de données (wikipédia)

mais comme un système qui prédit les formulations les plus probable pour produire du texte.

Exemple :

Il ne comprend pas vraiment qu'un chat miaule, il calcule juste que *miaule* est un mot très probable après *chat*.

Sa Nature Probabiliste



$$1 + 1 = 2$$

Toujours déterministe



LLM

toujours Probabiliste

La réponse peut changer même si la question reste identique.

Ce n'est pas un bug, c'est parce qu'il ne fonctionne pas de manière totalement prévisible.

Entraînement vs Déduction

Entraînement

L'école

Ça prend des mois

Le modèle apprend sur des milliards de textes. Sa connaissance est définie ici.

Un LLM, c'est comme un élève qui a fini ses études : il peut utiliser ce qu'il a appris, mais il n'apprend plus rien de nouveau pendant la conversation et il ne se met pas à jour tout seul.

Déduction

L'examen

Ça prend une seconde

Quand vous utilisez un LLM il est en mode déduction. Il utilise ce qu'il a appris.

Les Transformers le mécanisme (2017)

Le transformer, c'est comme un chef d'orchestre : il coordonne tous les mots pour produire une réponse harmonieuse.
C'est l'accord qui a changé la mélodie

Avant 2017

Avant les transformers, l'IA jouait les notes (mots) une par une



Après (Transformer)

Avec les transformers, elle entend toute la mélodie d'un coup



Sans cette invention, il n'y aurait jamais eu de grands solos comme ChatGPT, Gemini, Claude, Copilot ou les autres modèles

Le Modèle

Comment le système interprète le texte

la segmentation ou la Tokenisation

Ce que vous voyez

artificielle

Ce que l'IA voit

arti

fic

ielle

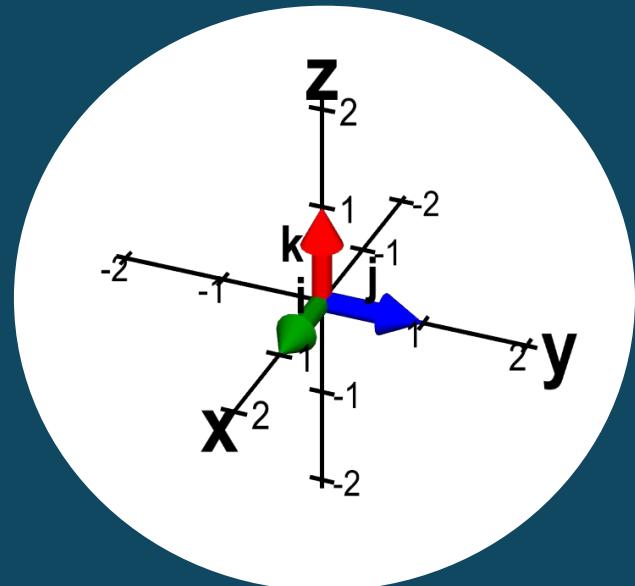
La segmentation est une nécessité technique pour que le modèle comprenne le langage.

La facturation au token n'est qu'un effet secondaire de ce fonctionnement.

L'IA découpe le texte en morceaux (Token) pas en lettres, ni en syllabes.

Chaque token a un **Embeddings** / vecteur, liste de nombre

Chaque token reçoit des coordonnées mathématiques dans un espace à des centaines de dimensions.



Mais

« Il est né à l'est »

Le mot « **est** » apparaît deux fois, mais il n'a **pas du tout le même sens** :

est = verbe « être »

est = direction (opposé de ouest)

Si un ordinateur donne **la même représentation numérique** (embedding) au mot « **est** » dans tous les cas, il ne pourra **jamais** comprendre lequel des deux sens est utilisé.

Espace latent le monde où vivent les points

Le modèle regarde les mots autour pour comprendre le sens exact.

Ainsi, dans :

« Il est né » → « est » = verbe

« à l'est » → « est » = direction

Le même mot reçoit **deux embeddings différents**, adaptés à la phrase.

C'est un peu comme si le mot avait une **couleur différente** selon la phrase dans laquelle il apparaît.

L'ordinateur comprend mieux le sens réel d'un mot dans une phrase

Mais

l'IA ne connaît ni l'ordre, ni le sens
sans information sur l'ordre des tokens, l'espace latent ne peut pas organiser correctement
la phrase.

Encodage positionnel l'ordre des mots

Pour l'IA, sans aide :

« Le chat mange la souris » « La souris mange le chat »
... pourraient être identiques, car les mêmes tokens sont présents, juste mélangés.

L'encodage positionnel ajoute à chaque token une **étiquette numérique** qui dit :
« Je suis le 1er », « Je suis le 2e », etc.

L'espace latent peut enfin représenter **le sens + la structure**.

Mais

Un LLM n'a pas une base de données de faits. Il ne sait pas ce qui est vrai ou faux.
Il ne fait que **prédir le mot suivant** qui semble le plus probable.

Donc s'il ne sait pas, il **devine**. On dit qu'il **Hallucine**

**Vous savez maintenant que l'IA
découpe, navigue, positionne,
et ordonne.**

Le Pilotage

guider un LLM pour une meilleure réponse

Prompt Engineering / Ingénierie du prompt ou requête

c'est l'art de bien demander

Le rôle : indiquer à l'IA *qui* doit répondre à ma requête

Commencez par dire : tu es un expert en informatique, tu es un historien, tu es un expert juridique, tu es un professeur

Le rôle change **le ton, le niveau, la précision, la structure.**

Prompt faible

‘Explique-moi comment fonctionne un altimètre’

Prompt fort

‘tu es un instructeur de vol, explique-moi comment fonctionne un altimètre, sa définition, son fonctionnement, son rôle, son utilisation’

Le contexte : l'intention de la demande, ce que tu veux,

Rôle *qui doit répondre*

Ex: tu es un instructeur de vol avion chevronné

Pourquoi *ton objectif*

Ex: je veux comprendre la portance pour préparer mon examen théorique de pilote avion

Pour qui *le public*

Ex: Explique-le à un élève débutant

Quoi *le sujet précis*

Ex: le sujet principal est la portance, sa définition, ses forces, ses contraintes, son rôle.

Comment *la forme souhaitée et les contraintes*

Ex: en 4 lignes de 60 caractères, avec des phrases courtes, sur un ton pédagogique.

Prompt faible

‘Explique-moi comment fonctionne la portance’

Prompt fort

Tu es un instructeur de vol pilote avion et tu guides un élève débutant.

Je veux apprendre la portance pour préparer mon examen théorique de pilote privée avion.

Le sujet est la portance : définition, forces, contraintes et rôle en vol, mais plus encore.

Explique ces notions sous formes de chapitres d'environ 250 caractères et fais-moi un résumé pour mes révisions.

Utilise des phrases courtes et un ton pédagogue pour faciliter la compréhension.

la Température le curseur de créativité

Température basse (0-0,3)

La portance est une force aérodynamique verticale. Elle dépend de la vitesse, de l'air, de la surface alaire et de l'angle d'attaque.

Prévisible, rigoureux, factuel

Température moyenne (0,5-0,7)

La portance est la force qui soulève l'avion. Elle apparaît quand l'air circule plus vite au-dessus de l'aile qu'en dessous, ce qui crée une pression plus faible sur l'extrados. Avec la vitesse et l'angle d'attaque, c'est ce qui permet de décoller et de tenir en vol.”

Equilibré, précis

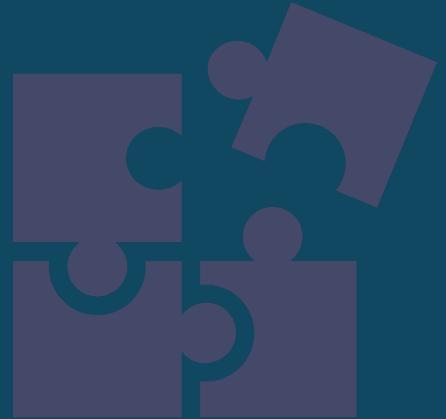
Température élevée (0,8-1+)

“La portance, c'est l'air qui attrape l'aile et la porte comme une main invisible. Quand l'avion accélère, cette main devient plus forte et soulève la machine vers le ciel.

Créatif surprenant imagé, risque d'erreur

la Fenêtre de contexte

c'est la mémoire de travail du modèle.



Limitée en taille

Effacée quand elle déborde

Indispensable pour maintenir la cohérence

Différente d'une mémoire humaine

Gemini 2M+ tokens

Claude 200K tokens

GPT4 128K tokens

Résumer

Résume en 5 lignes ce qu'on a établi sur ma recherche d'emploi.

Objectif : trouver un poste X

Compétences clés : ...

Expérience principale : ...

Contraintes : ton professionnel, phrases courtes, localisation

Cibles : entreprises, secteurs, types de postes



CONTEXTE :

Objectif : trouver un poste X

Compétences clés : ...

Expérience principale : ...

Contraintes : ton professionnel, phrases courtes, localisation

Cibles : entreprises, secteurs, types de postes

Tâche : Rédige-moi une lettre de motivation adaptée.

Réinjecter

Avant chaque nouvelle demande importante, tu remets ce résumé au début.

Tu structures ton prompt

Tu écris ton prompt comme un briefing

Rôle : conseiller en recherche d'emploi

Objectif : optimiser ma candidature

Contexte : résumé ci-dessus

Format : lettre, CV, pitch, entretien...

Vous avez donc 3 leviers :
la précision de la demande,
la gestion de la mémoire,
le style de générations.

Les Limites

l'IA n'est pas infaillible

Les hallucinations, la prédiction plutôt que la vérité

L'IA générative ne ‘sait’ rien au sens humain , c'est un moteur de probabilités; elle prédit la suite de mots la plus cohérente statistiquement.

Simulation vs compréhension, le mythe de la conscience

L'IA excelle dans limitation de la structure du langage, mais elle est dépourvue de sens profond, elle manipule des tokens pas des concepts comme l'empathie ou la tristesse.

Le Protocole de vérification

L'IA impose une étape systématique de validation, vérifier les données réel, les chiffres, les dates, les sources et l'actualité de celles-ci.

Les Outils IA du quotidien

Les assistants conversationnels

Copilot L'écosystème Microsoft l'assistant intégré pour Word et Excell (ajouter un fichier et promptez, création de résumé ou de formule)

Gemini l'expertise Google et la force de son système cloud.

Claude réputé pour sa subtilité rédactionnelle et son éthique, ainsi que pour le codage python.

Chapgpt le pionnier polyvalent.

Les assistants de générations d'images

Bing image Creator de l'écosystème Microsoft

Nano Banana native de Gemini (Google)

Midjourney réputé pour sa qualité artistique et son rendu esthétique avancé

DALL.E l'outil intégré à l'écosystème Chapgpt

Comment est réalisé une image : le modèle de diffusion

Imagine un restaurateur d'art :

On lui donne une peinture très abîmée

Il enlève une couche de saleté

Puis une autre

Puis une autre

Jusqu'à ce que l'œuvre réapparaisse

Le modèle de diffusion fait exactement ça... mais avec un chaos de bruit numérique ensuite débruité

Les assistants de générations de vidéos

Bing image Creator de l'écosystème Microsoft

Veo 3 le modèle de Google deepmind intégré à Gemini

Canva Média magique

Sora l'outil d'Open AI intégré Chapgpt

Comment est réalisé une vidéo : le modèle de diffusion vidéo

Imagine un créateur de film :

Il y a le Réalisateur : le Prompt c'est le scénario

Il y a le décorateur : il crée les images une par une (modèle de diffusion)

Il y a l'animateur : il assure la cohérence entre les images (cohérence temporelle)

Il y a le monteur : il assemble le tout ; image + mouvement + son

Les assistants Bureautique

Rédaction Word : passez des idées au document, à partir de ses idées et notes, rédige un compte-rendu en trois parties.

Formules Excell : le coach technique pour créer des formules pour les tableaux.

Synthèse : tapez vos résumés de réunion et demander une synthèse structurée.

Planification : organise la liste de tâches à effectuer dans la journée ou la semaine.

Automatiser : donnez un échantillon de départ pour une modification de saisie longue : transforme Nom, Prénom en format p.nom@entreprise.fr

**En plus des limites et
des assistants, vous venez de découvrir l'univers
de compréhension de l'IA**

Merci pour votre attention.



<https://github.com/vaillantclaude>