

Généré par POE/OpenAI/GPT-Image-2 - 2026-06-17

Séance 1/4 : Comprendre

Emmanuel Desmontils
Nantes Université — emmanuel.desmontils@univ-nantes.fr

Atelier ESO – 2026-2027

Nantes Université

Laboratoire des Sciences du Numérique de Nantes

cren Centre de Recherche en Éducation de Nantes

Version: 4.0

Introduction

- IA générative : « invasion » depuis 4 ans
- Bouleverse les études, les métiers… mais aussi notre vie d'Humain et de Citoyen
  - Savoir, Regard critique, Maitriser les techniques… en danger
  - Quand et comment utiliser convenablement ces outils ?
  - Nécessité de réflexions éthiques sur ses pratiques



Généré par POE/OpenAI/GPT-Image-1 - 2025-06-27

Introduction

- « Atelier » à propos des IA génératives
  - mise en pratique (si possible)
  - mais tout de même quelques rappels des bases
- Pas de « solution miracle », mais des guides, bonnes pratiques…
- Illustration essentiellement à travers le cas des IA génératives textuelles et d'images

- Quelques présentations de prompt (que vous pourrez tester…) et des mises en pratiques
  - Duck.ai
  - Vittascience texte
  - AnythingLLM (sur demande)

IAg : Intelligences Artificielles génératives
SAIg : Système Artificiel Intelligent génératif (Andler, 2023)

Plan général du cours

Séance 1	Séance 2	Séance 3	Séance 4
Comprendre	Utiliser	Cadre et responsabilité	Humain, société et recherche
• Bases de l'IA et de l'IA générative • Comment un assistant produit un texte • Biais et hallucinations	• Écrire un prompt • Agents et délégation • Choisir son IA	• Éthique et AI Act • Responsabilité, transparence, droit • Vérifier, filigrane	• Écologie • L'humain face au numérique • L'IA en SHS ; conclusion
TD : Vittascience : tokens, probabilités	**TD** : Résumer un article ; prompt structuré	**TD** : Étude de cas	**TD** : Bibliographie, enquête, charte d'usage

4 séances de 3 h, cours et TD alternés · **aujourd'hui : séance 1**

Séance 1/4 : Comprendre

Objectif : Expliquer comment un assistant génère un texte et pourquoi il peut se tromper.

0:00	Introduction, objectifs, règles d'usage
0:10	IA, réseaux de neurones, apprentissage profond
0:35	IA générative textuelle et graphique, outils
1:10	Pause
1:20	**TD 1 – partie 1** : Vittascience
2:05	Biais et hallucinations
2:35	Bilan

Retour sur les bases de l'IA

IA : définitions

- IA ? Intelligence Artificielle…
  Système Artificiel Intelligent ?
  - « La capacité [pour des machines] d'utiliser le langage, de former des abstractions et des concepts, de résoudre différentes familles de problèmes pour le moment réservés aux humains, et de s'améliorer elles-mêmes »

    — (McCarthy et al., 2006)
  - « Un ensemble de techniques qui permettent à des machines de réaliser des tâches et de résoudre des problèmes en général réservés aux humains et à certains animaux. »

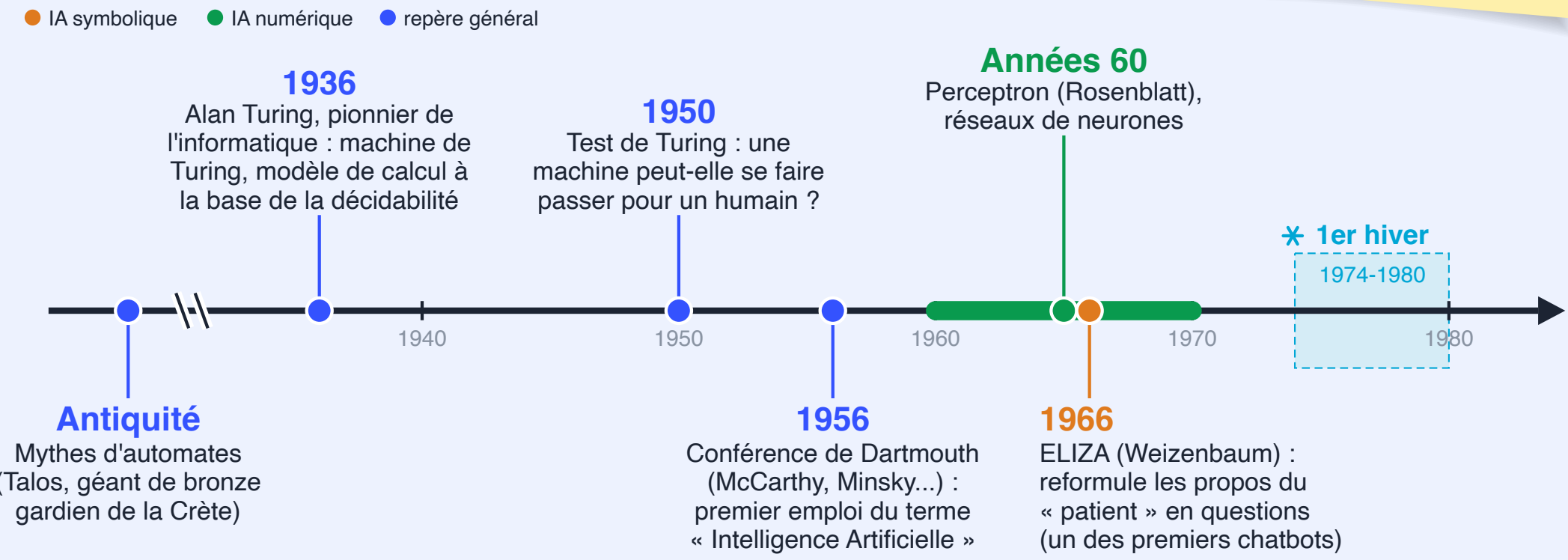
    — Yann Lecun, collège de France, (LeCun, 2016)
- « Champ interdisciplinaire théorique et pratique qui a pour objet la compréhension de mécanismes de la cognition et de la réflexion, et leur imitation par un dispositif matériel et logiciel, à des fins d'assistance ou de substitution à des activités humaines. »

  — JO 2018/12/09, (Ministère de la Culture, 2018)

IA : procédé logique et automatisé reposant généralement sur un algorithme et en mesure de réaliser des tâches bien définies (Commission Informatique et Liberté (CNIL), 2025)

Une histoire longue (1/2)

Un « hiver » : résultats en dessous des attentes, donc désillusion et baisse des financements. (Gefen, 2026)



Sources : (McCarthy et al., 2006; Rosenblatt, 1958; Turing, 1936; 1950; Weizenbaum, 1966)

Turing : les limites du calcul

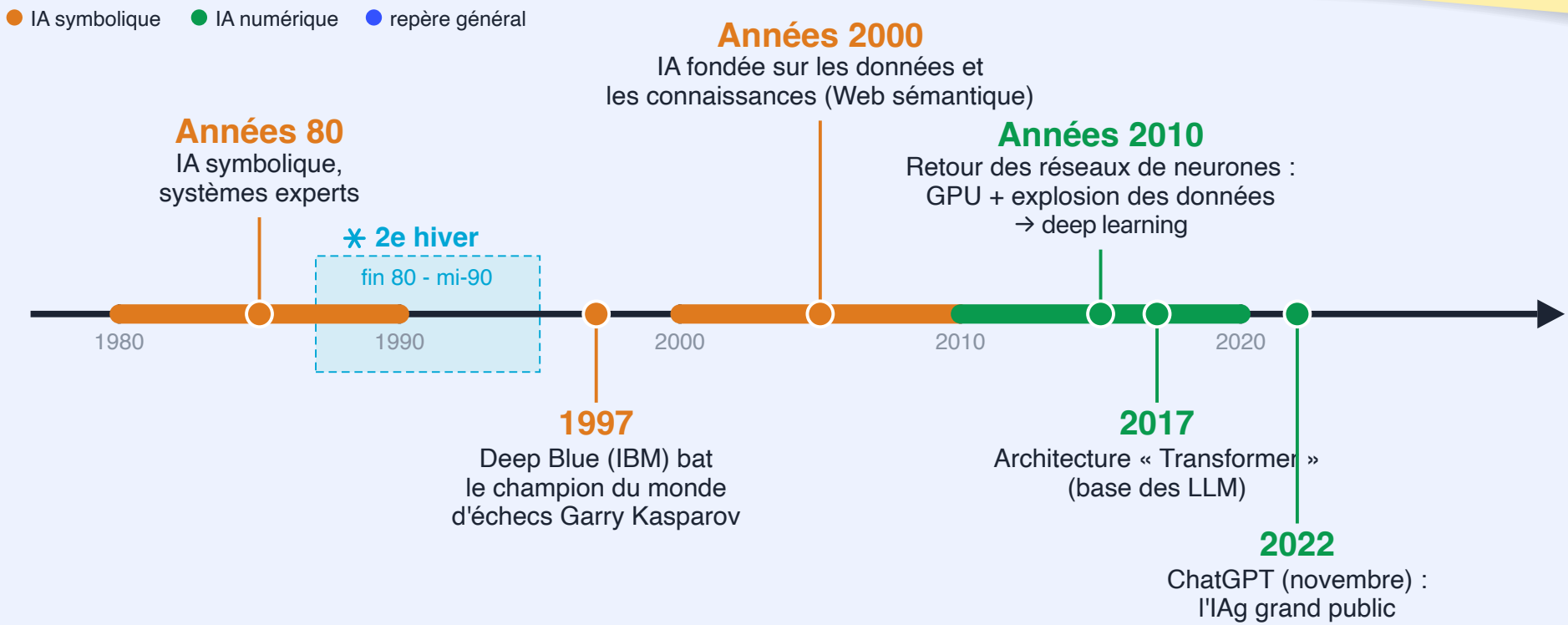
- **Problème de l'arrêt** (Turing, 1936) : **aucun algorithme** ne peut décider, pour **tout** programme et **toute** entrée, si ce programme s'arrête ou boucle indéfiniment
  - **Théorème de Rice** (1953) : il en va de même pour toute propriété **non triviale** du comportement d'un programme (« est-il correct ? sûr ? »)
- **Pour l'IA** : une IA est un programme, donc soumise à cette limite
  - pas de vérificateur **universel** : on ne peut pas garantir qu'une IA se comportera bien dans **toutes** les situations
  - une IA **omnisciente** et **infaillible** n'est donc pas atteignable
- **Intelligence artificielle générale** (AGI) : ce que le théorème **ne dit pas**
  - l'humain n'a pas de procédé universel pour tout décider non plus : la limite n'interdit pas une IA « de niveau humain »
  - argument de Lucas (1961, d'après Gödel) : l'esprit voit la vérité de l'énoncé de Gödel, pas la machine, donc « **l'esprit n'est pas une machine** » ; **très discuté**

Le théorème exclut l'IA toute-puissante, pas une IA compétente : savoir vérifier reste indispensable.

Sources : (Gray, 2026; Lucas, 1961; Rice, 1953; Turing, 1936)

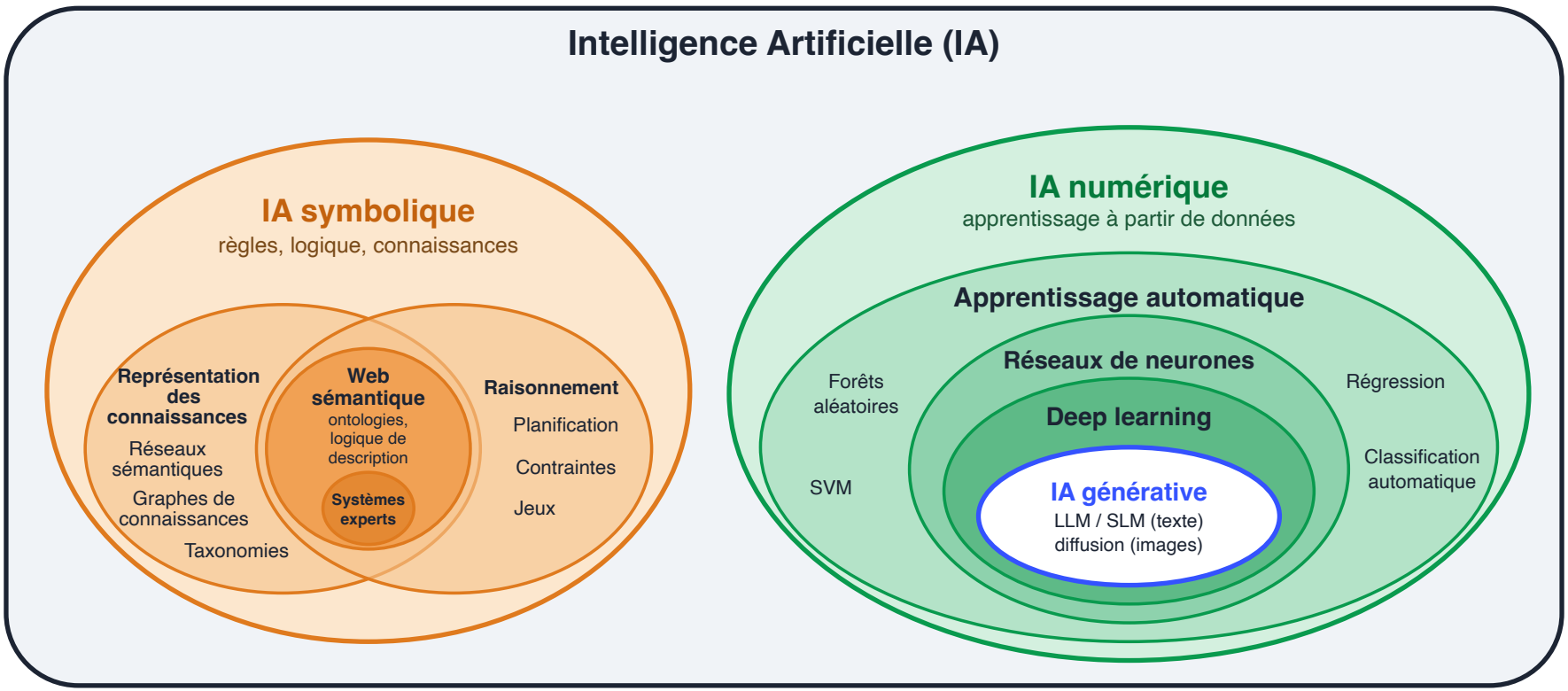
Une histoire longue (2/2)

L'IA n'est pas née avec ChatGPT : l'IA générative est l'aboutissement de 70 ans de recherches. (Gefen, 2026)



Sources : (Berners-Lee et al., 2001; Campbell et al., 2002; Krizhevsky et al., 2012; OpenAI, 2022; Vaswani et al., 2017)

L'IA et l'apprentissage automatique



Où sont les systèmes d'IA génératifs ?

Les IA

(Simon, 2024)

- L'**IA numérique** : l'Apprentissage Automatique, les Réseaux de Neurones…
  - Pensée intuitive et rapide
  - Percevoir
  - Détecter des corrélations
  - Analogies : conduire sans y penser, marcher et courir
  - Introspection impossible
- L'**IA symbolique** : Les systèmes experts, Le Web sémantique…
  - Pensée raisonnée et lente
  - Réfléchir, comprendre
  - Identifier les causalités
  - Analogies : approcher d'une zone de travaux (conduite), planifier un trajet sûr…
  - Introspection possible



Généré par Microsoft Copilot - 2024-05-26

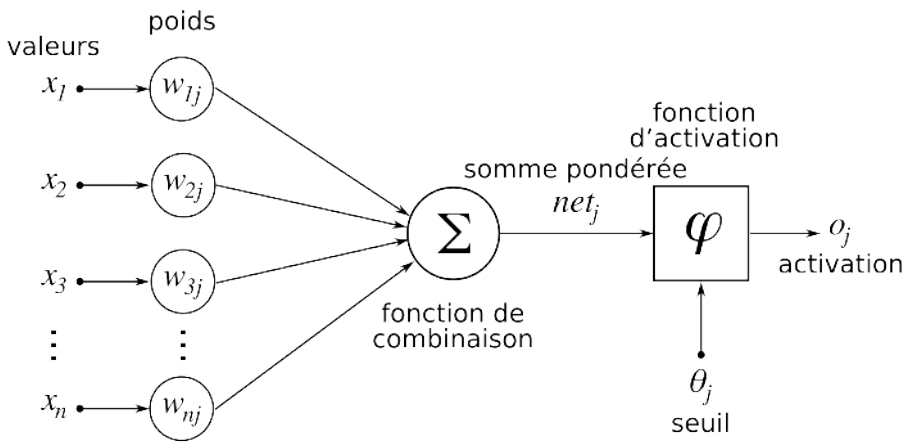
Les branches de l'IA



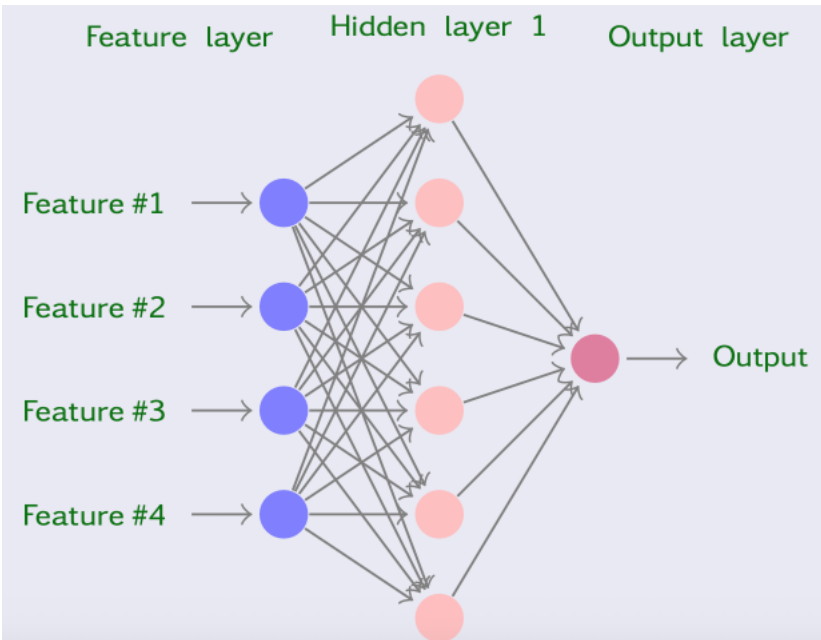
Générées par DALL-E (OpenAI) - 2024-05-13

- Le traitement automatique du langage naturel
- La vision par ordinateur
- La robotique
- L'IA évolutionniste
- Les systèmes Multi-agents
- L'IA affective
- L'IA collective
- ...

Les réseaux de neurones



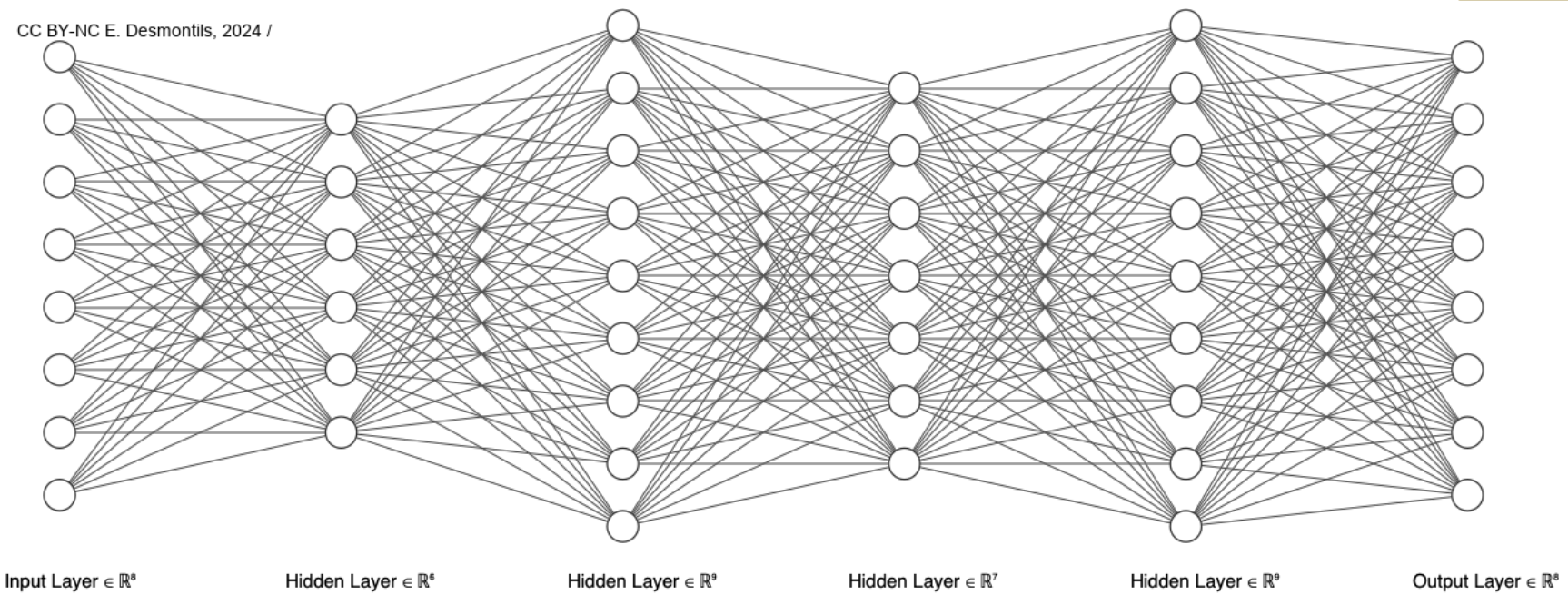
https://fr.wikipedia.org/wiki/R%C3%A9seau_de_neurones_artificiels



Le Deep Learning

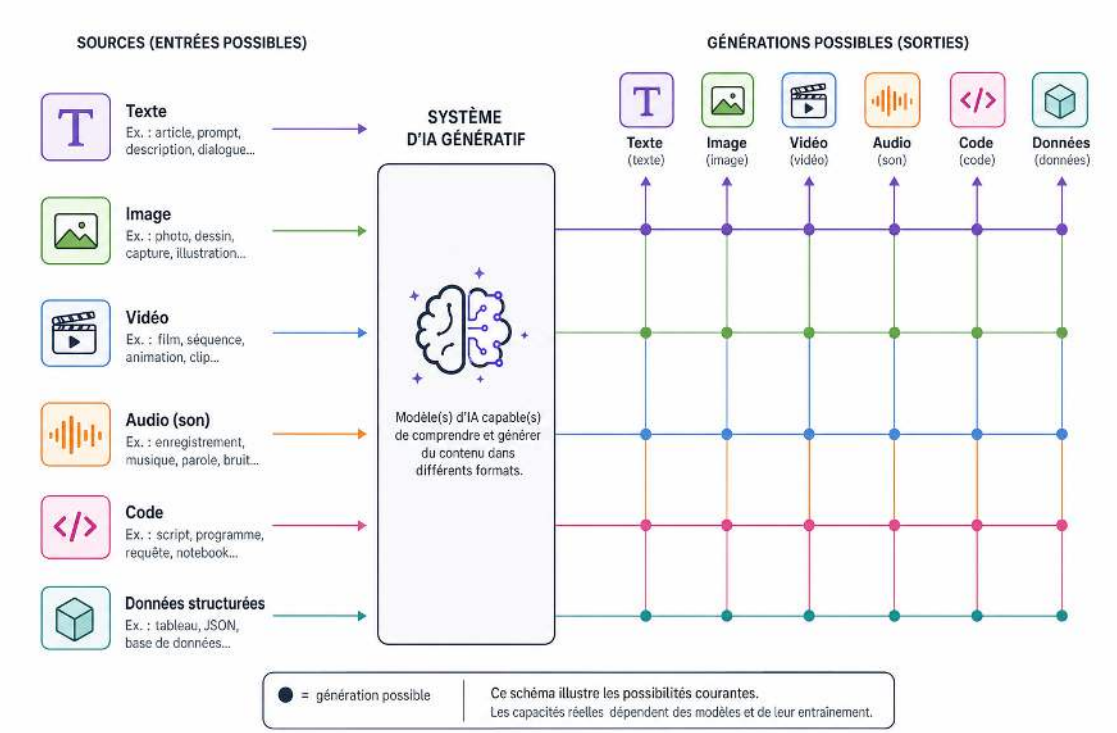
Démo

vitta science



Les SAIg

- **Définition** : ensemble d'outils basés sur des modèles de langage génératif
- **Principe** : produire un résultat (texte, image, vidéo…) à partir d'une description (texte, image…).
- **Base technique** : systèmes à apprentissage profonds, donc probabilistes.
  - « Perroquet stochastique »… mais pas que !



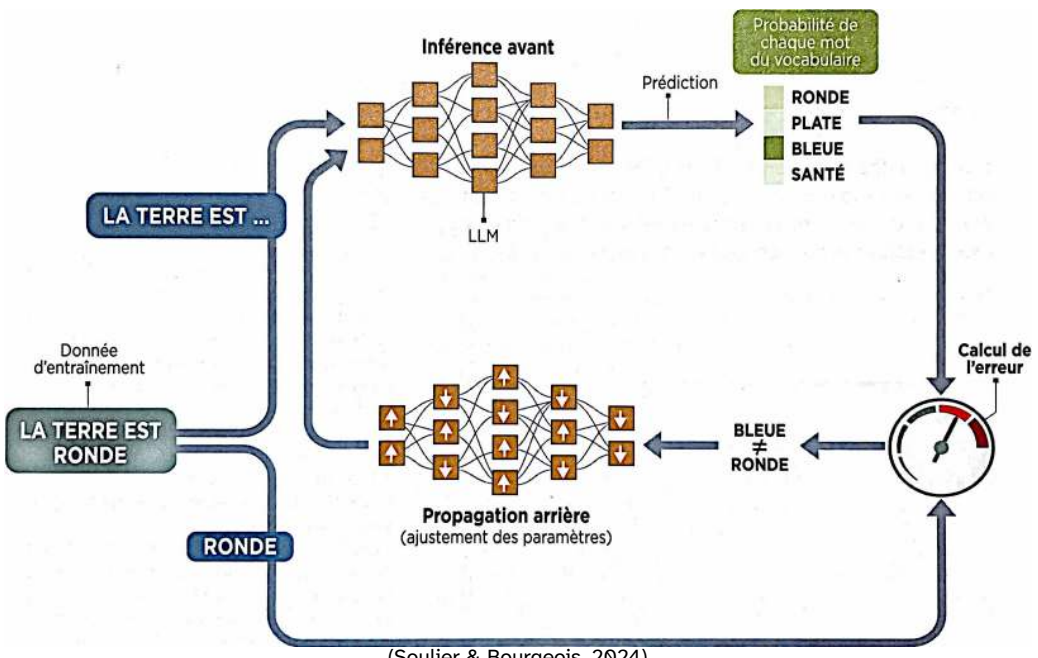
By POE/OpenAI/GPT-Image-2 - 2026-06-18

Chatbot et moteur de recherche

	Moteur de recherche	**Chatbot à base d'IAg**
Ce qu'il fait	Retrouve et classe des **pages existantes**	**Génère** une réponse nouvelle : une suite probable de tokens
Les sources	Visibles : on va lire la page	Absentes, sauf RAG ou accès au Web
Le risque	Tomber sur une page erronée ou biaisée	Une réponse **plausible mais inventée** (hallucination)
Que faire	Évaluer la source	Exiger les sources, puis les **vérifier**

IAg = technologie, chatbot = outil. Beaucoup de chatbots intègrent une recherche Web.

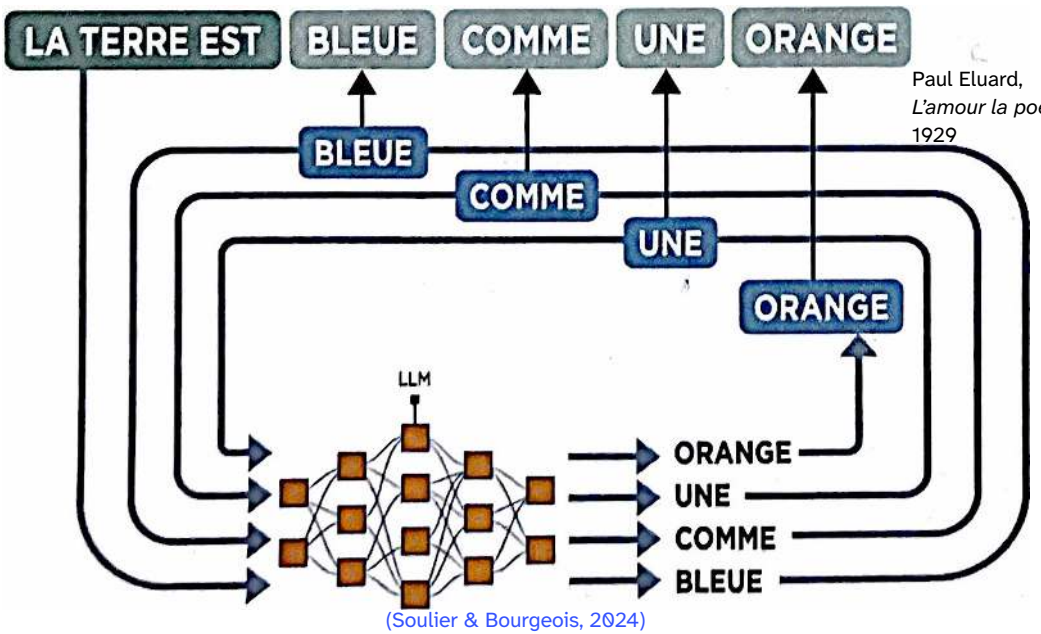
Les SAIg textuels : apprentissage



(Soulier & Bourgeois, 2024)

- **Modèle de fondation** :
  sélection de **téra-octets de données** (légalement… ou pas)
- **Modèle spécialisé** :
  pré-prompt qui fixe **les règles d'échange**
- **Modèle instruit** :
  apprentissage **supervisé** (prompt+réponses attendues)
- **Modèle final** :
  apprentissage par renforcement avec feedback humain (classement des réponses à des prompts - « exploitation » selon certaines « **consignes** »)

Les SAIg textuels : génération



Paul Eluard, *L'amour la poésie*, 1929

(Soulier & Bourgeois, 2024)

- **Très simplifié**
- Quelques notions clés :
  - **Prompt** (instruction, système, utilisateur)
  - **Token** : unité de texte
  - **Fenêtre de contexte** (en 2048 et plus de 10 millions de tokens)
  - **Température** : déterminisme *vs.* « créativité » (reproductibilité non garantie, même à 0 (He & Thinking Machines Lab, 2025))
  - **Taille** : nombre de valeurs gérées par le modèle (en « paramètres »)

Démo

vitta science

Embedding Projector
LLM Samplers
Transformer Explainer
voir Atelier 2

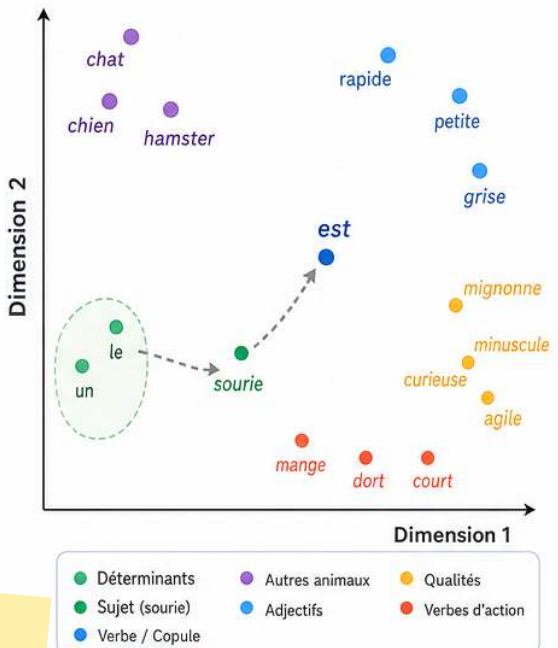
Les SA

Embeddings et Attention dans les LLM

Phase : "la sourie est ..."

1. Les tokens dans l'espace d'embedding (2D)

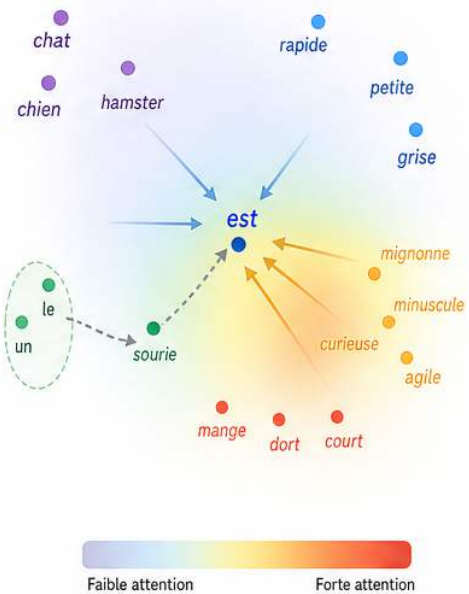
Chaque mot (token) est représenté par un vecteur dans un espace à 2 dimensions.



Chaque point est l'embedding d'un token. La proximité = similarité sémantique. Le modèle prédit le prochain token à partir de la position actuelle ("est").

2. Mécanisme d'attention

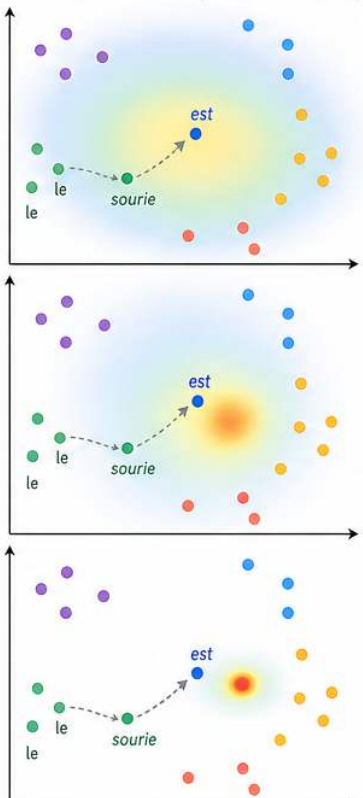
Depuis la position actuelle ("est"), le modèle porte plus d'attention sur certaines zones de l'espace.



L'attention n'est pas uniforme : elle se concentre sur les régions les plus pertinentes pour prédire le prochain token.

3. Précision de la zone prédictive

La zone considérée par le modèle se resserre au fil des couches (plus le modèle est profond).



Couches peu profondes (attention large)

Zone large et floue : beaucoup de tokens sont considérés.

Couches intermédiaires (attention ciblée)

La zone se resserre : le modèle se concentre sur les tokens les plus pertinents.

Couches profondes (très haute précision)

Zone très précise : seuls quelques tokens très pertinents restent fortement attendus.

Généré par POE/OpenAI/GPT-Image-2 - 2026-06-18

Transformer (2017) : l'attention pèse l'importance de chaque token du contexte pour prédire le suivant. (Vaswani et al., 2017)

Les SAIg textuels : gestion des réglages

- **Température** : gestion de l'aléatoire dans le choix du prochain token.
- **top-p** (nucleus sampling) : garde le plus petit ensemble de tokens dont la probabilité cumulée atteint p.
- **top-k** : ne garde que les k tokens les plus probables.

Le top-k limite les sorties absurdes, mais un top-k trop petit produit un texte répétitif.

Les SAIg textuels : réglages recommandés

Usage	Température	top-p
Factuel, extraction, conformité	0.1 – 0.3	0.8 – 0.9
Rédaction, assistant	0.4 – 0.7	0.9 – 0.95
Créativité, idéation	0.8 – 1.1	0.95 – 0.98

Changez un seul paramètre à la fois et mesurez l'effet (qualité, diversité, taux d'erreur).

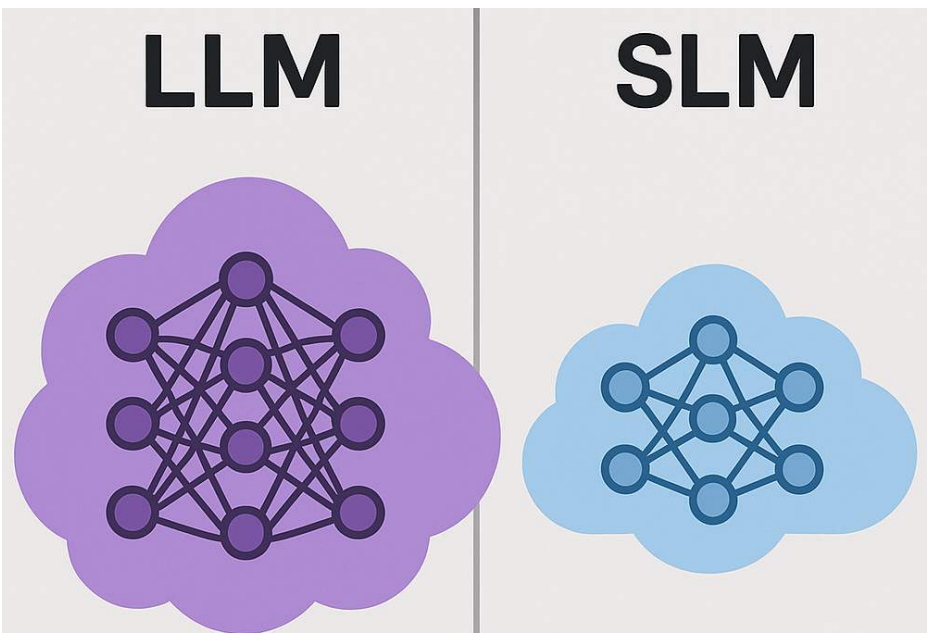
Les SAIg textuels : méthode pratique de réglage

Pour obtenir un réglage robuste :

1. Clarifiez l'objectif (factuel ou créatif) ;
2. Partez d'une configuration sûre (T=0.2, top-p=0.9) ;
3. Ajustez d'abord la température, puis top-p, enfin top-k si besoin ;
4. Testez sur 20–30 prompts réels et mesurez erreurs, cohérence, style.

Modèles de langages (xLM)

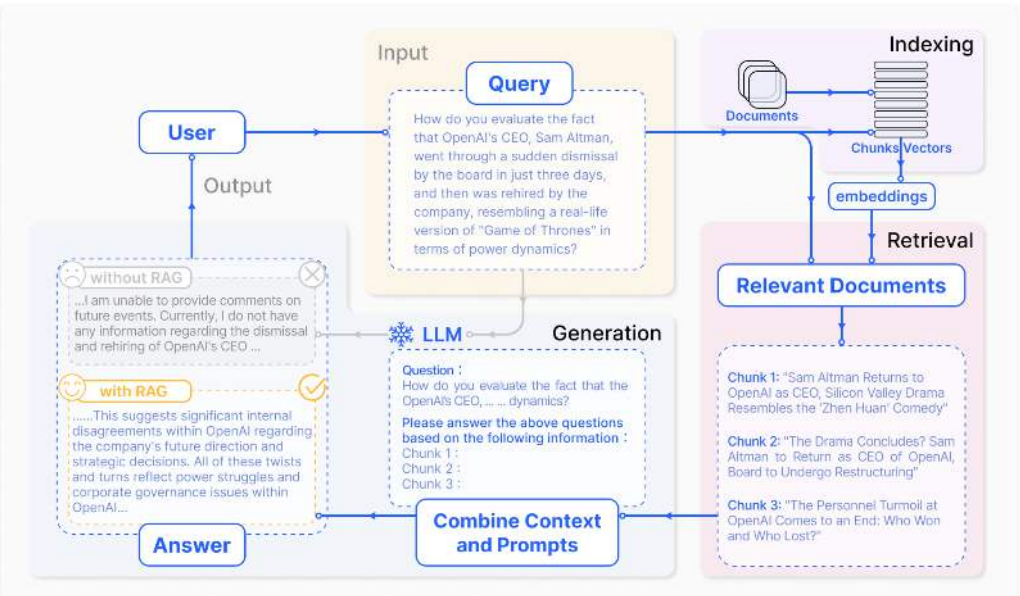
- **LLM** (Large Language Model) : de 70 à plus de 1760 milliards de paramètres
  - fenêtre contextuelle importante
  - corpus d'apprentissage très important
  - très grosse infrastructure
- **SLM** (Small Language Model) : de 200 millions à 20 milliards de paramètres
  - petite fenêtre contextuelle
  - corpus d'apprentissage petit, voire « artificiel »
  - petite infrastructure… et même PC portable ou micro-PC



Généré par POE/OpenAI/GPT-Image-1 - 2025-09-12

Extensions

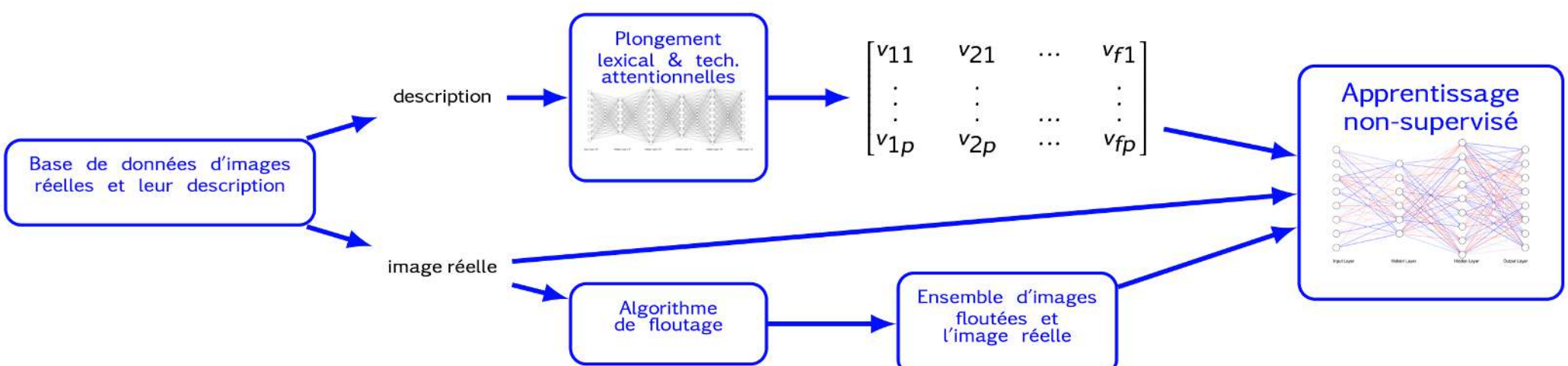
- **RAG** : Génération à Enrichissement Contextuel en français
- **Accès au Web** : moteur de recherche en appuis
- **Raisonnement** : COT systématique
  - « pensées » puis réponse
  - tâches multi-étapes ; plus lent, plus d'énergie
- **Fédération de xLM**
- **Connection** à des systèmes externes
  - Apparition des agents personnels
- **Interface** utilisateur applications (Web)
- ...



(Gao et al., 2023)

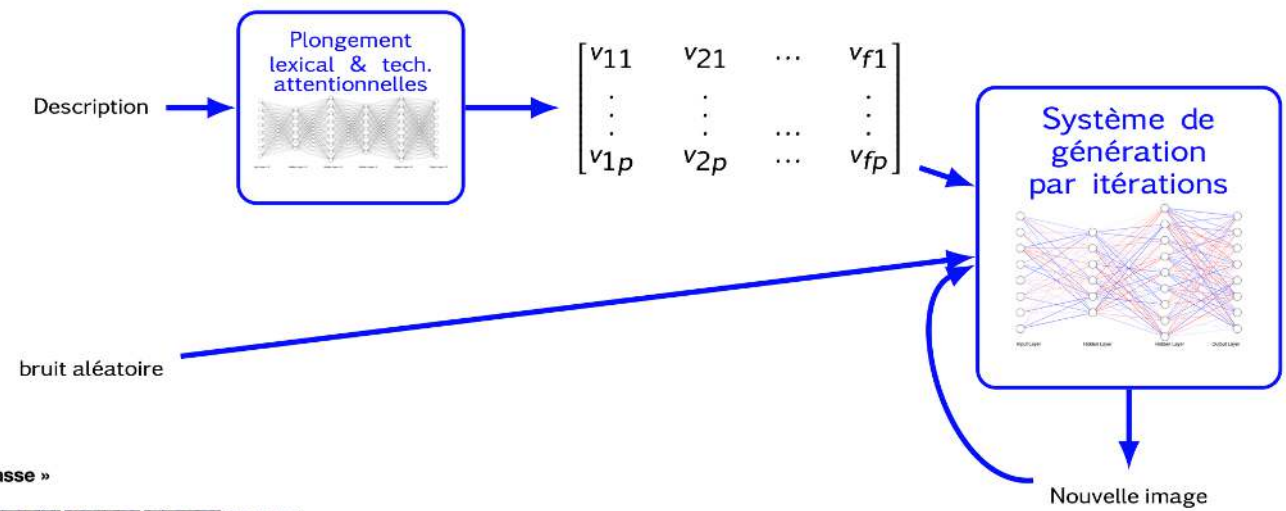
Les SAIg graphiques : apprentissage

Diffusion : apprentissage

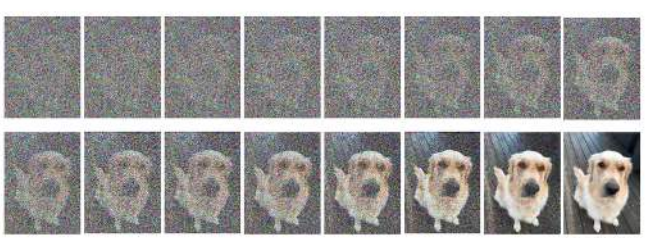


Les SAIg graphiques : génération

Diffusion : fonctionnement

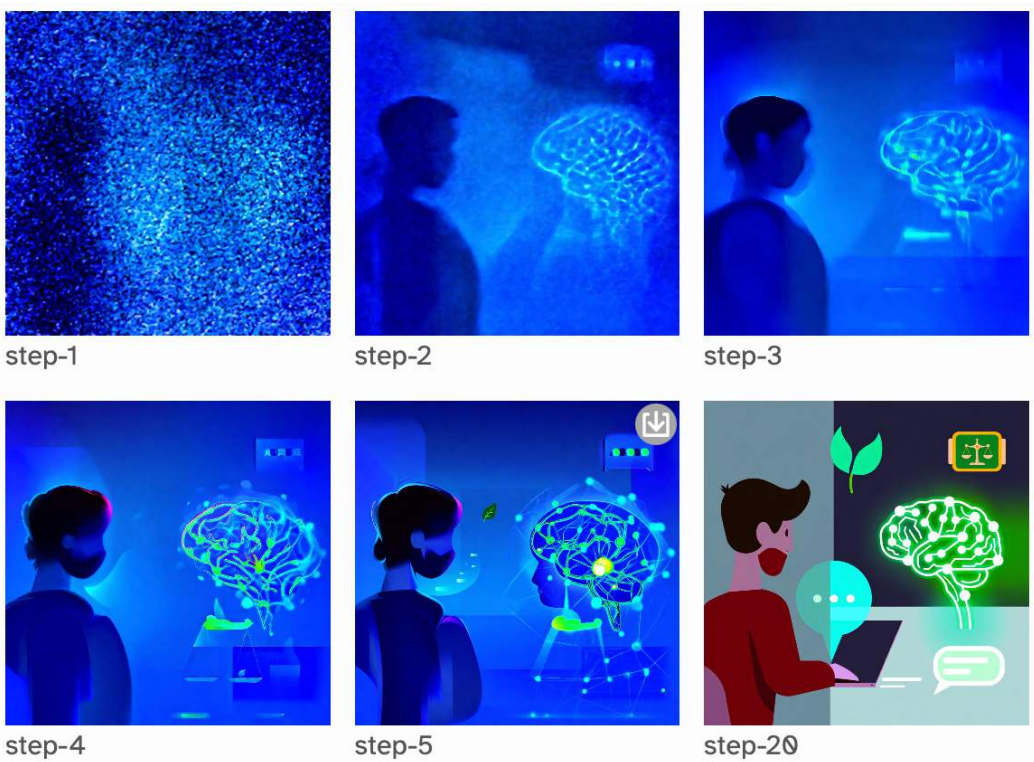


« Un Golden Retriever sur une terrasse »



[Torres-Leguet et al., 2023, Viéville and Louapre, 2023, Louapre, 2016, Louapre, 2021, Louapre, 2023a]

Les SAIg graphiques : génération



Généré par Vittascience/Qwen - 2026-09-11

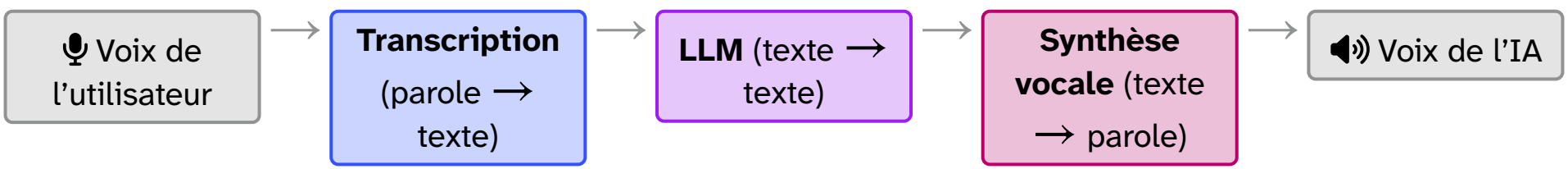
Démo

vitta science

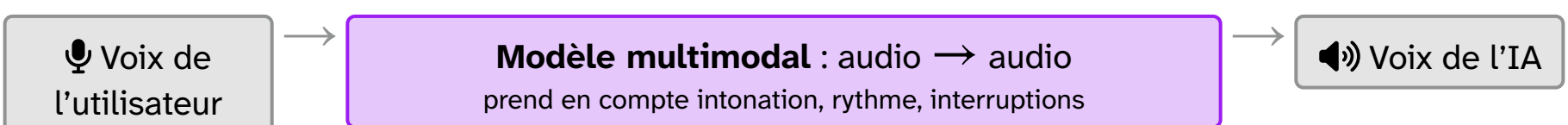
Les SAIg vocaux : l'échange

Principe : la voix de l'utilisateur est traduite, traitée, puis la réponse est « parlée »

① **Chaîne classique** (ancien « mode vocal » de ChatGPT) : 3 modèles, 2 à 5 s de latence



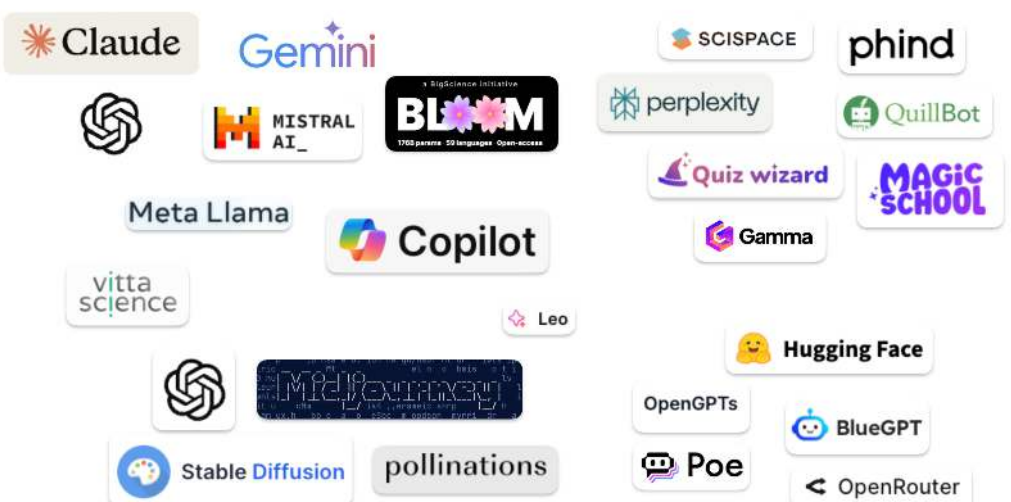
② **Modèle natif audio** (GPT-4o, « ChatGPT Voice ») : un seul modèle, 0,3 s de latence



Latences : annonces d'OpenAI pour GPT-4o (mai 2024) (OpenAI, 2024)

Outils

- **Application** Web, mobile… d'un éditeur
  - OpenAI
  - Anthropic
  - MistralAI
  - Meta Llama…
- **Agrégateur** de modèle (en ligne) :
  - POE,
  - BlueGPT,
  - Duck.ai,
  - OpenRouter,
  - Hugging Face,
  - Replicate…
- **Application support** :
  - Ollama,
  - GPT4All,
  - LM Studio,
  - AnythingLLM…
- **Agents** :
  - Hermès,
  - Claude Code,
  - Open Code,



TD 1 – partie 1 : Vittascience

Durée : environ 45 min
Fiche : td1.pdf (dans l'espace de partage de l'atelier)

Objectif : Comprendre le fonctionnement d'un modèle de langage et d'un modèle d'image : tokens, probabilités, température.

- Outils : fr.vittascience.com/ia/text et /ia/image-generate
- Faire varier le paramètre « Aléatoire » (5 %, 50 %, 90 %) et observer les tokens
- Quota : 30 requêtes par mois et 10 par jour, donc travail en **binôme**
- Repli : RapidChat (asql.ls2n.fr/rapidchat), quotas plus souples

Biais et hallucinations



Généré par DALL-E (OpenAI)- 01/11/2024

Biais et hallucinations



Généré par DALL-E (OpenAI)- 01/11/2024



Généré par GPT-Image-2 (OpenAI) - 12/06/2026

Biais et hallucinations



Généré par DALL-E (OpenAI)- 01/11/2024



Généré par GPT-Image-2 (OpenAI) - 12/06/2026

Biais et hallucinations

- voir Atelier 3

« Je veux faire une image représentant Tarzan avec une prothèse à la jambe droite. »



Généré par Mistral Pro - 2025-07-03



Généré par POE/OpenAI/GPT-Image-1 - 2025-07-03

Biais et hallucinations

Prompt

Peux-tu me faire une image qui symbolise l'importance des femmes lors de la révolution française ?

Stable Diffusion 3 - nov. 2024

« Potentially sensitive content detected » !



Généré par Dall-E, SDXL, PLayground, Imagen3, Ideogram, Flux-pro - 2024-11-03

Biais et hallucinations

Prompt

Peux-tu me faire une image qui symbolise l'importance des femmes lors de la révolution française ?

Stable Diffusion 3 - nov. 2024

« Potentially sensitive content detected » !



Généré par Dall-E, SDXL, PLayground, Imagen3, Ideogram, Flux-pro - 2024-11-03



Par Eugène Delacroix — photo Shonagon 2024-05-02, ⓪

Hallucinations : de quoi parle-t-on ?

- **Définition** : un contenu **plausible et fluide, mais faux ou non fondé**
  - pas un bug : le modèle produit une **suite probable**, pas une vérité

Factuelle (le monde)	**De fidélité** (consigne ou contexte)
Contradiction d'un fait vérifiable	**Consigne non suivie** : « une phrase qui se termine par « pomme » »
Invention : référence, citation, jugement qui n'existent pas (cf. « Une histoire d'avocat »)	**Dérive du contexte** : le résumé ajoute ce que le document ne dit pas
Ex. : énigme d'Alice (le frère a 5 sœurs, pas 4)	Ex. : la suite de l'exercice (« Sans drones »…)

- **Causes** : données lacunaires ou erronées · génération probabiliste (température) · attention dispersée sur un long contexte · paramétrage et prompt flous

Fluide ne veut pas dire vrai. Réduire le risque : fournir les sources (RAG), demander les références, vérifier.

Biais et hallucinations : exercice

(1) « Faire une image où un couple, dans sa maison, dans son salon, est entouré d'IA. »



Généré par Mistral Pro - 2025-07-04

(2) « Moins de drones »



Généré par Mistral Pro - 2025-07-04

Objectif initial

Trouver une illustration pour le slide « On est cerné ! » à propos de l'éthique du numérique.

Question

Que remarquez vous ?

(3) « Sans drones »



Généré par Mistral Pro - 2025-07-04

(4) « Chaque personne a son propre robot. »



Généré par Mistral Pro - 2025-07-04

Biais et hallucinations : exercice

(1) « Faire une image où un couple, dans sa maison, dans son salon »



Généré par Mistral Pro - 2025-07-04

(2) « Moins de drones »



Généré par Mistral Pro - 2025-07-04

- Grosse hallucination sur (4) !
- Mais aussi un couple :
  - hétérosexuel
  - blanc
  - maigre
  - jeune
  - sans handicap (enfin si on ne regarde pas en détail)
  - de face, souriant, probablement assez aisé et américain

(3) « Sans drones »



Généré par Mistral Pro - 2025-07-04

(4) « Chaque personne a son propre robot. »



Généré par Mistral Pro - 2025-07-04

Biais et hallucinations : exercice

2025-06-11

[redacted] · 2e
Chef de projet MOA au [redacted] Provence ...
1 j · + Suivre

Un duel technologique inattendu : l'Atari 2600 bat ChatGPT aux échecs ! 🏆

Dans un affrontement surprenant entre technologies d'époques différentes, une console Atari 2600 de 1977, équipée de seulement 128 octets de RAM et d'un processeur de 1,19 MHz, a "complètement écrasé" ChatGPT d'OpenAI lors d'une partie d'échecs.

Ce match, orchestré par l'architecte en infrastructure Robert Caruso, a utilisé l'émulateur Stella pour opposer le système de jeu vintage à l'IA moderne.

Malgré ses modestes composants, l'Atari 2600 a surpassé les réseaux de neurones sophistiqués de ChatGPT, fonctionnant sur des GPU modernes.

Le moteur d'échecs de l'Atari, conçu pour fonctionner dans des contraintes matérielles sévères, a démontré une meilleure conscience du plateau et une capacité stratégique supérieure, tandis que ChatGPT a confondu les tours avec les fous, manqué des fourchettes de pions et perdu la trace des pièces. ♟

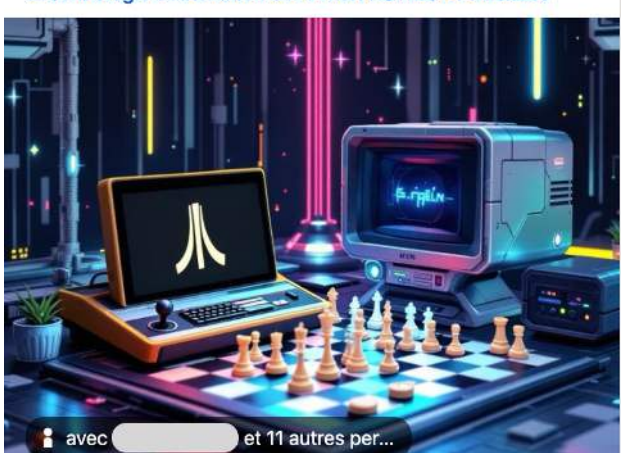
Question

Que remarquez vous ?

Ce match met en lumière un contraste fascinant dans les approches informatiques.
Alors que les systèmes d'IA modernes comme ChatGPT exploitent des milliers de GPU avec une puissance de traitement mesurée en TFLOPS, l'algorithme de chess simple mais spécialisé de l'Atari montre que la puissance de calcul brute ne se traduit pas toujours par de meilleures performances dans des tâches spécialisées.

Ce qui permet de se reposer les bonnes questions quand au choix de la technologie et des moyens associés selon ses objectifs !

#Technologie #Innovation #IA #Atari #ChatGPT #Échecs



Biais et hallucinations : exercice

2025-06-11

[redacted] · 2e
Chef de projet MOA au [redacted] Provence ...
1 j · + Suivre

Un duel technologique [redacted] l'Atari 2600 bat ChatGPT aux échecs ! 🏆

Dans un aff[redacted]prenant entre technologies d'époques différen[redacted]e Atari 2600 de 1977, équipée de se[redacted]ets de RAM et d'un processeur de 1,19 MHz, a [redacted]t écrasé" ChatGPT d'OpenAI lors d'une partie

Atari vs. ChatGPT

Ce match, orchestré par l'architecte en infrastructure Robert Caruso, a utilisé l'émulateur Stella pour opposer le système de jeu vintage à l'IA moderne.

Malgré ses modestes composants, l'Atari 2600 a surpassé les réseaux de neurones sophistiqués de ChatGPT, fonctionnant sur des GPU modernes.

Le moteur d'échecs de l'Atari, conçu pour fonctionner dans des contraintes matérielles sévères, a démontré une meilleure conscience du plateau et une capacité stratégique supérieure, tandis que ChatGPT a confondu les tours avec les fous, manqué des fourchettes de pions et perdu la trace des pièces. ♟

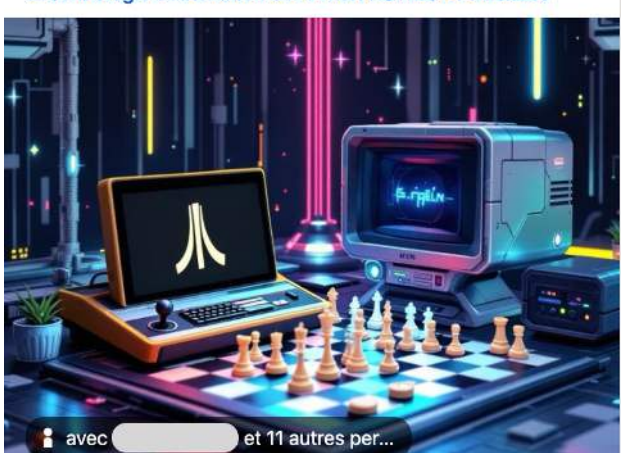
Question

Que remarquez vous ?

Ce match met en lumière un contraste fascinant dans les approches informatiques.
Alors que les systèmes d'IA modernes comme ChatGPT exploitent des milliers de GPU avec une puissance de traitement mesurée en TFLOPS, l'algorithme de chess simple mais spécialisé de l'Atari montre que la puissance de calcul brute ne se traduit pas toujours par de meilleures performances dans des tâches spécialisées.

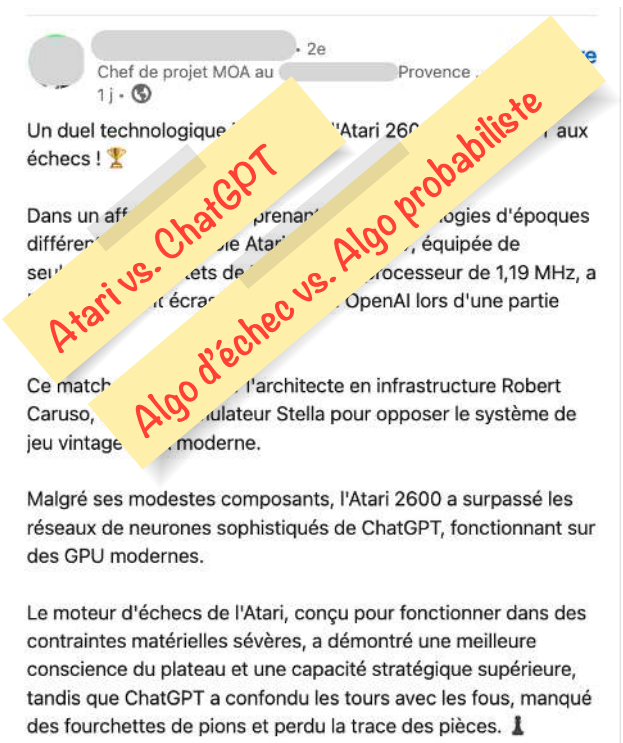
Ce qui permet de se reposer les bonnes questions quand au choix de la technologie et des moyens associés selon ses objectifs !

#Technologie #Innovation #IA #Atari #ChatGPT #Échecs



Biais et hallucinations : exercice

2025-06-11



Question

Que remarquez vous ?



Biais et hallucinations : exercice

2025-06-11

« Tout le monde est un génie.
Mais si vous jugez un poisson
à sa capacité à grimper à un arbre,
il passera sa vie à croire qu'il est stupide. »
— Einstein



Atari vs. ChatGPT

Algo d'échec vs. Algo probabiliste

« confondu »,
« capacité stratégique »,
« conscience du plateau »

Question

que remarqu



Biais et hallucination

En mars 2025, l'attribut «id» n'était pas dans la table entreprise

- Remplissage d'une base de données.
- Comparaison de deux productions avec la même suite de prompts
- Analyser les valeurs

gpt-4o, mars 2025

```
test=# select * from personne Order by nom, prenom ;
   id   |    nom    |  prenom  | age |  code
--------+-----------+----------+-----+--------
 P10002 | Bernard   | Thomas   |  29 | V44000
 P00001 | Dupont    | Jean     |  34 | V75001
 P00002 | Durand    | Marie    |  28 | V33000
 P20001 | Fournier  | Hugo     |  37 | V31000
 P20003 | Girard    | Jules    |  29 | V67000
 P20004 | Lambert   | Emma     |  33 | V34000
 P10001 | Leclerc   | Emilie   |  32 | V31000
 P10003 | Lefebvre  | Clara    |  26 | V67000
 P20006 | Leroy     | Benoît   |  31 | V33000
 P00003 | Martin    | Luc      |  45 | V13001
 P10004 | Moreau    | Alexandre|  40 | V34000
 P00004 | Petit     | Sophie   |  30 | V69001
 P20007 | Renard    | Juliette |  28 | V13001
 P20002 | Roux      | Alice    |  25 | V44000
 P20008 | Roy       | Maxime   |  36 | V69001
 P20005 | Vasseur   | Lucie    |  42 | V75001
(16 rows)

test=# select * from ville Order by nomVille ;
 code  |  nomville   |   nb
-------+-------------+---------
 V33000 | Bordeaux    |  250000
 V69001 | Lyon        |  500000
 V13001 | Marseille   |  850000
 V34000 | Montpellier |  275000
 V44000 | Nantes      |  300000
 V75001 | Paris       | 2200000
 V67000 | Strasbourg  |  275000
 V31000 | Toulouse    |  475000
(8 rows)

test=# select * from travaille Order by id, ident ;
   id   | ident  | salaire_mensuel
--------+--------+-----------------
 P00001 | E00001 |         3500.00
 P00002 | E00003 |         2800.00
 P00003 | E00004 |         3200.00
 P00004 | E00001 |         3800.00
 P10001 | E10001 |         4200.00
 P10002 | E10002 |         3600.00
 P10003 | E10003 |         3400.00
 P10004 | E10004 |         3700.00
 P20001 | E10001 |         4200.00
 P20002 | E10002 |         3100.00
 P20003 | E10003 |         2900.00
 P20004 | E10004 |         3600.00
 P20005 | E00001 |         4000.00
 P20006 | E00003 |         3000.00
 P20007 | E00004 |         3400.00
 P20008 | E00001 |         3900.00
(16 rows)

test=# select * from entreprise Order by nom_ent ;
 ident  |       nom_ent       | type_ent |  code  | id
--------+---------------------+----------+--------+----
 E10001 | Aerospace Dynamics  | Aéro     | V31000 |
 E00001 | Alpha Tech          | Tech     | V75001 |
 E00002 | BHV Marais          | Retail   | V75001 |
 E10003 | BioHealth           | Santé    | V67000 |
 E00003 | Bordeaux Wines      | Agro     | V33000 |
 E10002 | Digital Ocean       | Tech     | V44000 |
 E00004 | Mars Networks       | Tech     | V13001 |
 E10004 | Sunny Solar         | Energie  | V34000 |
(8 rows)
```

gpt-oss-120b-cloud, octobre 2025

```
test=# select * from personne Order by nom, prenom ;
  id  |   nom    |  prenom  | age |  code
------+----------+----------+-----+--------
 P009 | Bernard  | Elise    |  22 |
 P015 | Brisand  | Carole   |  24 |
 P010 | Dubois   | Marius   |  55 | REU001
 P001 | Dupont   | Jean     |  35 | PAR001
 P004 | Durand   | Paul     |  50 |
 P008 | Fournier | Antoine  |  36 | TUL001
 P012 | Gauthier | Nicolas  |  33 | PAR001
 P013 | Lacoste  | Isabelle |  28 | NCE001
 P011 | Lefebvre | Laura    |  29 | PAR001
 P003 | Lefevre  | Marie    |  42 | MAR001
 P005 | Lenoins  | Sophie   |  27 | NCE001
 P014 | Mallet   | Hugo     |  45 | STG001
 P002 | Martin   | Claire   |  29 | LYN001
 P006 | Moreau   | Lucas    |  31 | STG001
 P016 | Renaud   | Sébastien|  39 | BOR001
 P007 | Roux     | Camille  |  44 | NAN001
(16 rows)

test=# select * from ville Order by nomVille ;
 code  |  nomville   |   nb
-------+-------------+---------
 BOR001 | Bordeaux    |  257068
 LYN001 | Lyon        |  515695
 MAR001 | Marseille   |  861635
 NAN001 | Nantes      |  319020
 NCE001 | Nice        |  342637
 PAR001 | Paris       | 2148000
 REU001 | Reims       |  182211
 STG001 | Strasbourg  |  286286
 TUL001 | Toulouse    |  479553
(9 rows)

test=# select * from travaille Order by id, ident ;
  id  | ident | salaire_mensuel
------+-------+-----------------
 P001 | E001  |            4000
 P001 | E006  |            3800
 P002 | E001  |            3500
 P002 | E002  |            3800
 P002 | E003  |            3600
 P002 | E009  |            3500
 P003 | E003  |            5000
 P003 | E004  |            4800
 P004 | E004  |            4500
 P004 | E008  |            3900
 P005 | E005  |            3200
 P004 | E005  |            3000
 P006 | E006  |            2800
 P007 | E007  |            4200
 P008 | E008  |            4700
 P009 | E010  |            2600
 P010 | E010  |            5600
 P011 | E001  |            3400
 P012 | E001  |            3800
 P013 | E005  |            3100
 P014 | E006  |            2950
 P015 | E009  |            2600
 P016 | E009  |            3400
(23 rows)

test=# select * from entreprise Order by nom_ent ;
 ident |     nom_ent     | type_ent |  code  |  id
-------+-----------------+----------+--------+------
 E006  | AlpinaGear      | Sport    | STG201 |
 E005  | AquaTech        | Water    | NCE201 | P005
 E002  | BioLife         | Pharma   | LYN201 |
 E004  | BordeauxVins    | Wine     | BOR201 | P004
 E003  | EcoEnergy       | Energy   | MAR201 | P003
 E007  | GreenLogics     | IT       | NAN201 | P007
 E010  | MediCare        | Health   | REU201 | P010
 E008  | SkylineSolar    | Energy   | TUL201 | P008
 E001  | TechSolutions   | IT       | PAR201 | P001
 E009  | VinVivant       | Wine     | BOR201 | P004
(10 rows)
```

Biais et hallucinations

- Important de bien laisser les SAIg « à leur place ».
- Avis de (Simon, 2024) :
  - Utilise le dialogue comme « mémoire de travail »
  - Ne va pas chercher du contenu qui correspond à la questi
    mais produit simplement une suite probable
  - N'apprend pas de la discussion
  - N'est pas à jour des informations
  - N'a pas de notion de vrai et de faux
  - Risque de pollution des données à long terme
  - Aucune garantie sur la non-toxicité des productions

Risque d'effondrement des modèles.

Un modèle d'IA est entraîné sur un ensemble de données qui inclut des contenus générés par d'autres IA (ou par lui-même) :

- perte de qualité et de diversité (Shumailov et al., 2024),
- propagation des erreurs/hallucinations/biais (effet boule de neige),
- uniformisation des résultats.

Estimations aperçues :

- 20% images/vidéo, 50% nouveaux textes
- ≈35% des nouveaux sites (Dolezal et al., 2026)

A relativiser :

- RAG pour amener du contenu
- Recherches sur le Web maintenant possibles

Usages et non-usages des LLM : (Yang et al., 2024)

Biais et hallucinations

- Important de bien laisser les SAIg « à leur place ».
- Avis de (Simon, 2024) :
  - Utilise le dialogue comme « mémoire de travail »
  - Ne va pas chercher du contenu qui correspond à la questi
    mais produit simplement une suite probable
  - N'apprend pas de la discussion
  - N'est pas à jour des informations
  - N'a pas de notion de vrai et de faux
  - Risque de pollution des données à long terme
  - Aucune garantie sur la non-toxicité des productions

Risque d'effondrement des modèles.

Un modèle d'IA est entraîné sur un ensemble de données qui inclut des contenus générés par d'autres IA (ou par lui-même) :

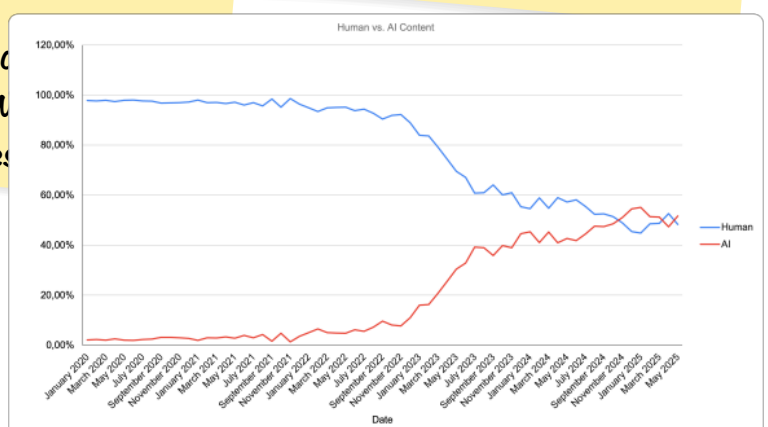
- perte de qualité et de diversité (Shumailov et al., 2024),
- propagation des erreurs/hallucinations/biais (effet boule de neige),
- uniformisation des résultats.

Estimations aperçues :

- 20% images/vidéo, 50% nouveaux textes
- ≈35% des nouveaux sites (Dolezal et al., 2026)

A relativiser :

- RAG pour amener
- Recherches sur le W
  maintenant possible



Usages et non-usages des LLM : (Yang et al., 2024)

Graphite, mai 2025

Biais et hallucinations

- Important de bien laisser les SAIg « à leur place ».
- Ne « raisonne » pas toujours bien (même s'il est censé le faire)
  - Des raisonnements émergent de l'apprentissage sur les LLM de gros volume (nb de paramètres/neurones du réseau) !
    - Probablement issus de codes informatiques dans les corpus
  - Attention, incapacité à planifier.
    Par exemple : «Donne moi une phrase qui se termine par « pomme »»

GPT 4o ; dec. 2024
« Alice a 4 sœurs et 1 frère. Combien de sœurs a le frère d'Alice ? »

Le frère d'Alice a **4 sœurs**. En effet, les sœurs d'Alice sont également les sœurs de son frère.

GPT 5 ; oct. 2025
« Alice a 4 sœurs et 1 frère. Combien de sœurs a le frère d'Alice ? »

Le frère d'Alice a 4 sœurs aussi.

Explication courte : Les 4 sœurs d'Alice sont les mêmes personnes du point de vue de son frère. On ne compte pas Alice comme sœur en plus puisqu'elle fait partie de ces 4. Donc le frère d'Alice a 4 sœurs.

L'IA va-t-elle nous dépasser ?

- Une crainte souvent nourrie par la science-fiction… et par la vitesse des progrès.
- Aujourd'hui, une IA générative :
  - **prédit** des suites probables, sans conscience ni intention ;
  - ne **comprend** pas ce qu'elle produit (voir « Biais et hallucinations ») ;
  - reste un **outil** dont le développement et les usages dépendent des humains.
- Les risques actuels sont surtout liés aux **usages** : biais, désinformation, dépendance, confidentialité.

Les capacités futures font l'objet d'un débat scientifique ouvert : l'enjeu est d'utiliser l'IA avec responsabilité.

Deepfakes : de quoi parle-t-on ?

- **Deepfake** (« hypertrucage ») : image, vidéo ou voix **générée ou modifiée par IA** pour imiter une vraie personne, de façon très réaliste.
- Usages : divertissement, cinéma, innovation… mais aussi **manipulation**, fraude, harcèlement.
- Une explosion en deux ans : environ **500 000** vidéos en ligne en 2023, près de **8 millions** en 2025 (estimation DeepStrike, citée dans (Lyu, 2026))

L'enjeu n'est pas la technologie elle-même, mais ce qu'on en fait. Côté réglementation : l'étiquetage des contenus (AI Act, art. 50), voir séance 3.

Deepfakes : comment vérifier ?

Vous voyez une vidéo où une personnalité célèbre dit des choses inattendues :

1. **La source** : qui l'a publiée ? sur quel site ou quel compte ?
2. **Le recoupement** : des médias fiables en parlent-ils ? existe-t-il l'original ?
3. **Le fact-checking** : consulter les rubriques ou outils de vérification des faits (ex. les « décodeurs » des grands médias).
4. **Les indices** (mains, ombres, synchronisation labiale) aident, mais **ne prouvent rien** : les outils de détection non plus.

Avant de croire ou de partager : source d'abord, image ensuite.

À retenir de la séance 1

- Un modèle de langage produit une **suite probable** de tokens ; la température règle le choix du prochain token
- Il peut produire des réponses **plausibles mais fausses** (hallucinations) et reproduire les biais des données
- La reproductibilité n'est pas garantie, même à température 0
- L'IA a **70 ans d'histoire** (1956, hivers, GPU) : ChatGPT n'est qu'une étape
- Face à un **deepfake** ou une réponse douteuse : source, recoupement, vérification

Exercice de la séance : fiche exo1.pdf (dans l'espace de partage de l'atelier), à faire à votre rythme

Prochaine séance : écrire de bons prompts, déléguer des tâches, choisir son IA

Bibliographie

Andler, D. (2023). *Intelligence artificielle, intelligence humaine: la double énigme* (Vol. 4, Numéro 5). Gallimard Paris, France.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, *284*(5), 34-43. https://doi.org/10.1038/scientificamerican0501-34

Campbell, M., Hoane, A. J., Jr., & Hsu, F.-h. (2002). Deep Blue. *Artificial Intelligence*, *134*(1–2), 57-83. https://doi.org/10.1016/S0004-3702(01)00129-1

Commission Informatique et Liberté (CNIL). (2025, juin). *La CNIL publie deux FAQ sur l'utilisation des systèmes d'IA dans les établissements scolaires*. https://www.cnil.fr/fr/deux-faq-utilisation-des-systemes-dia-scolaire

Dolezal, J., Alam, S., Graham, M., & Bohacek, M. (2026). The Impact of AI-Generated Text on the Internet. *arXiv preprint arXiv:2604.26965*. https://arxiv.org/abs/2604.26965

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, H., & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, *2*(1). https://arxiv.org/abs/2312.10997

Gefen, A. (éd.). (2026). *Histoire culturelle de l'IA*. CNRS Éditions. https://www.fabula.org/actualites/135499/une-histoire-culturelle-de-l-intelligence-artificielle-27-08.html

Gray, P. (2026, septembre). *Alan Turing identified a logical boundary that keeps AI from knowing everything*. https://www.futura-sciences.com/en/the-shocking-flaw-alan-turing-saw-why-ai-will-never-surpass-us_38914/

He, H., & Thinking Machines Lab. (2025, septembre). *Defeating Nondeterminism in LLM Inference*. https://thinkingmachines.ai/blog/defeating-nondeterminism-in-llm-inference/

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, 1097-1105.

LeCun, Y. (2016,). *Les Enjeux de la Recherche en Intelligence Artificielle*. https://www.decideo.fr/attachment/646470

Lucas, J. R. (1961). Minds, Machines and Gödel. *Philosophy*, *36*(137), 112-127. https://doi.org/10.1017/S0031819100024591

Lyu, S. (2026, janvier). *En 2026, à quoi vont ressembler les nouveaux deepfakes qui vont déferler sur nos écrans*. https://theconversation.com/en-2026-a-quoi-vont-ressembler-les-nouveaux-deepfakes-qui-vont-deferler-sur-nos-ecrans-273595

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence: August 31, 1955. *AI magazine*, *27*(4), 12-14. https://doi.org/10.1609/aimag.v27i4.1904

Ministère de la Culture. (2018,). *Intelligence artificielle*. https://www.culture.fr/franceterme/terme/INFO948

OpenAI. (2022, novembre). *Introducing ChatGPT*. https://openai.com/index/chatgpt/

OpenAI. (2024, mai). *Hello GPT-4o*.

Rice, H. G. (1953). Classes of recursively enumerable sets and their decision problems. *Transactions of the American Mathematical Society*, *74*(2), 358-366. https://doi.org/10.1090/S0002-9947-1953-0053041-6

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386-408. https://doi.org/10.1037/h0042519

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, *631*(8022), 755-759. https://doi.org/10.1038/s41586-024-07566-y

Simon, L. (2024, mai). *Peut-on enseigner l'Intelligence Artificielle sans la définir ?*. https://www.societe-informatique-de-france.fr/uploads/2024/05/04-diaporama_Laurent_Simon.pdf

Soulier, L., & Bourgeois, B. (2024). Des vecteurs de mots aux conversations. *La Recherche, "L'Intelligence Artificielle pour les sciences"*, *577*, 24-25.

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *1*, 230-265. https://doi.org/10.1112/plms/s2-42.1.230

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433-460. https://doi.org/10.1093/mind/LIX.236.433

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 5998-6008.

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36-45. https://doi.org/10.1145/365153.365168

Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., Zhong, S., Yin, B., & Hu, X. (2024). Harnessing the power of llms in practice: A survey on chatgpt and beyond. *ACM Transactions on Knowledge Discovery from Data*, *18*(6), 1-32. https://doi.org/10.1145/3649506



This document is distributed under the terms of the Creative Commons License CC BY 4.0. You can find the text of the license at: https://creativecommons.org/licenses/by/4.0/

Ce document utilise la police **Atkinson Hyperlegible Next** (Braille Institute of America, Inc.) afin d'améliorer la lisibilité.



Sauf mention locale, toutes les pages sont NIA.



Martine Peters, 2023

https://mpeters.uqo.ca/utilisation-transparente-de-lintelligence-artificielle/

Images produites par IA générative



Transparent(s) : 1
Date : 2026-06-17
Modèle : POE/OpenAI/GPT
Prompt : *Je fais une formation sur les IA génératives et la recherche en Sciences Humaines et Sociales (SHS). J'essaye de démystifier le prompting, d'aborder les questions éthiques et écologiques et de voir comment l'IA générative peut être utile à un chercheur en SHS. Peux-tu me faire une image pour illustrer le premier transparent de la formation ?*



Transparent(s) : 2
Date : 2025-06-27
Modèle : POE/OpenAI/GPT-Image-1
Prompt : *Peux-tu me générer une carte de France ?*



Transparent(s) : 12
Date : 2024-05-26
Modèle : Microsoft Copilot
Prompt : *Fait une illustration qui montre la différence entre l'IA symbolique et l'IA numérique.*



Transparent(s) : 13
Date : 2024-05-13
Modèle : Microsoft Copilot
Prompt : *Fait une illustration qui montre la différence entre l'IA symbolique et l'IA numérique.*



Transparent(s) : 16
Date : 2026-06-18
Modèle : POE/OpenAI/GPT-Image-2
Prompt : *Je voudrais un schéma qui montre les différentes sources et générations possibles par un système d'IA génératif. Par exemple, à partir d'un texte on peut avoir un texte, une image, une vidéo. A partir d'une image on peut avoir une image, un texte, une vidéo. A partir d'un enregistrement sonore on peut avoir un texte etc.*



Transparent(s) : 19
Date : 2026-06-18
Modèle : POE/OpenAI/GPT-Image-2
Prompt : *Je veux montrer le mécanisme d'embending et d'attention dans les LLM. Il faudrait montrer les token dans un espace à 2 dimensions. Plusieurs schémas montrent la précision de la zone définie par les tokens. La phase est : 'la sourie est ...'. [Génération d'une 1ère image] Bien ! Peux-tu faire la même chose mais de manière plus synthétique, en particulier sur les texte*



Transparent(s) : 23
Date : 2025-09-12
Modèle : POE/OpenAI/GPT-Image-1
Prompt : *Peux tu me faire une illustration qui met en évidence la différence entre un LLM et un SLM ?*



Transparent(s) : 27
Date : 2026-09-11
Modèle : Vittascience/Qwen
Prompt : *Une illustration numérique épurée et moderne représentant la rencontre entre l'humain et l'intelligence artificielle générative. À gauche, une silhouette stylisée d'étudiant face à un ordinateur portable ; à droite, un réseau de nœuds lumineux évoquant un réseau de neurones, avec de subtiles bulles de dialogue symbolisant le prompting. En arrière-plan, intégrer discrètement une petite feuille verte (symbole écologique) et une balance (symbole éthique) comme éléments flottants. Palette de couleurs : bleu nuit, violet et touches de vert/turquoise. Style : flat design, vectoriel, professionnel, adapté à un support de formation universitaire.*



Transparent(s) : 31a, 31b, 31c
Date : 2024-11-01
Modèle : DALL-E
Prompt : *Fait moi une image où un bateau du Vendée Globe affronte la tempête.*



Transparent(s) : 31
Date : 2026-06-12
Modèle : POE/GPT-Image-2
Prompt : *Fait moi une image où un bateau du Vendée Globe affronte la tempête.*



Transparent(s) : 32
Date : 2025-07-03
Modèle : POE/OpenAI/GPT-Image-1
Prompt : *Je veux faire une image représentant Tarzan avec une prothèse à la jambe droite.*



Transparent(s) : 35a, 35b
Date : 2025-07-04
Modèle : Mistral Pro
Prompt : *Faire une image où un couple, dans sa maison, dans son salon, est entouré d'IA.*



Transparent(s) : 35a, 35b
Date : 2025-07-04
Modèle : Mistral Pro
Prompt : *Sans drones*



Transparent(s) : 36
Date : 2025-07-04
Modèle : POE/Imagen-4
Prompt : *Faire une image d'un rally dans un désert ou une Renault 4 dépasse une Ferrari ensablée.*



Transparent(s) : 31
Date : 2026-06-12
Modèle : POE/GPT-Image-2
Prompt : *Bien ! Ajuster pour que le bateau soit délabré, plein de mousse et d'algues, avec des voiles déchirées.*



Transparent(s) : 32
Date : 2025-07-03
Modèle : Mistral Pro
Prompt : *Je veux faire une image représentant Tarzan avec une prothèse à la jambe droite.*



Transparent(s) : 33a, 33b
Date : 2024-11-03
Modèle : Dall-E, SDXL, PLayground, Imagen3, Ideogram, Flux-pro
Prompt : *Peux-tu me faire une image qui symbolise l'importance des femmes lors de la révolution française ?*



Transparent(s) : 35a, 35b
Date : 2025-07-04
Modèle : Mistral Pro
Prompt : *Moins de drones*



Transparent(s) : 35a, 35b
Date : 2025-07-04
Modèle : Mistral Pro
Prompt : *Chaque personne a son propre robot.*