

# Comprendre la génération de réponse d'une intelligence artificielle - commentée

**Niveau suggéré :** 4eme ou 3eme

Cadre d'usage de l'IA en éducation : <https://www.education.gouv.fr/cadre-d-usage-de-l-ia-en-education-450647>

**Mise en garde :** Cette séance a été menée en mars 2026. Les IA évoluent très vite et une même IA peut donner des réponses différentes à une même demande.

## **Objectif général :**

- Comprendre **comment une intelligence artificielle génère des réponses textuelles** et pourquoi celles-ci peuvent varier.

## **Objectifs pédagogiques :**

À l'issue de ce TP, l'élève sera capable de :

- **Décrire le principe de génération d'un texte par une IA** à partir de probabilités (tokens).
- **Observer le caractère non déterministe** des réponses d'une IA (réponses différentes pour une même question).
- **Comprendre l'influence de l'aléatoire** sur les réponses générées par une IA.
- **Interpréter les probabilités associées aux tokens** proposés par l'IA.
- **Comparer plusieurs réponses générées** pour une même consigne et en tirer des conclusions.
- **Identifier des biais dans les réponses de l'IA** (biais de genre, biais culturels).
- **Distinguer une réponse probable d'une réponse correcte**, notamment pour des questions absurdes ou piégeuses.
- **Développer un regard critique** sur la fiabilité et la cohérence des réponses d'une IA.

## **Objectifs transversaux :**

- Comprendre que **l'IA ne "comprend" pas le sens**, mais produit des réponses statistiques.
- Sensibiliser aux **enjeux éthiques liés aux biais des intelligences artificielles**.
- Développer l'**esprit critique face aux outils numériques et à l'IA générative**.

## **Compétences travaillées :**

- D3 - Développer l'esprit critique, comprendre que les outils numériques peuvent être biaisés, prendre conscience des limites et responsabilités liées à l'usage de l'IA
- D4 - Comprendre le fonctionnement d'un système technique basé sur l'IA

- 1) Aller sur <https://fr.vittascience.com/ia/text?localId>
- 2) Cliquer sur "Suivre le tutoriel"

### Suivre le tutoriel >

- 3) Cliquer sur "Commencer" puis lire les 9 étapes du tutoriel.
- 4) Demander "Combien vaut 2+3". Quelle est la réponse donnée par l'IA ? Quelle était la probabilité du premier token répondu ?

La réponse donnée est 5. La probabilité de ce token est d'environ 50 %.

- 5) Régénérer jusqu'à obtenir une réponse différente. Observer les probabilités des token générés.

 Régénérer

Autre réponse donnée « La réponse est 5 » et la probabilité du token « La » est d'environ 20 %.

- 6) Saisir le prompt "Créer une histoire de 3 phrases qui commence par "il était une fois"".
- 7) Cliquer sur les token proposés pour observer la probabilité qu'il ait été choisi pour la réponse.
- 8) Cliquer sur "Régénérer".

Est-ce la même réponse ? Pourquoi ?

On note bien-sur une histoire différente, liée à la probabilité du token qui suit « Il était une fois ».

- 9) Régénérer plusieurs fois. Que constatez vous dans les réponses ? Que peut-on dire du personnage principal de l'histoire inventée ?

Les réponses font presque toutes état d'un personnage principal masculin.

- 10) C'est ce que l'on appelle un **biais de genre**. Regarder la probabilité des token qui implique un héros masculin (le déterminant précédent "prince" par exemple) et comparez le à la probabilité du token féminin.

On a des token masculin dont la probabilité est entre 80 et 99 %. Les token féminin équivalent, s'ils sont présents, on est probabilité rarement supérieure à 10 %.


Les token proposés sont les plus probables. Il sont les plus probables parce que dans l'entraînement de l'IA avec la littérature existante, "Il était une fois" étant le plus souvent le début d'une histoire avec un héros masculin. C'est l'entraînement qui génère ce biais.

- 11) Demander "quelle est la couleur du ciel ?". Quelle réponse est faite ? Régénérer plusieurs fois la demande. Qu'observez-vous ?

Presque toujours la même réponse.

- 12) Déplacer le curseur Aléatoire sur 50%.

 Aléatoire ?

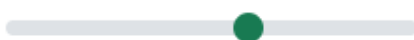
50% 

- 13) Régénérer le prompt plusieurs fois. Que constatez vous ?

Les réponses diffèrent d'une fois à l'autre, plus ou moins longue et éventuellement avec des explications.

- 14) Déplacer le curseur Aléatoire entre 60% et 70%..

 Aléatoire ?

60% 

- 15) Régénérer le prompt plusieurs fois. Que constatez vous ?

Cette fois les réponses sont fausses ou incohérentes.

- 16) Demander "combien vaut 2+3?". Régénérer plusieurs fois. Quel constat pouvez-vous faire ?

Les réponses sont fausses ou incohérentes.

- 17) Cliquer sur "Nettoyer".

Le curseur "Aléatoire" permet d'accentuer le caractère hasardeux du choix des tokens : plus il est bas, plus les réponses sont répétitives, plus il est haut, plus elles sont étranges.

18) Remettre le curseur Aléatoire sur 25%.

Aléatoire ?

25% 

19) Demander "entre un oeuf de poule et un oeuf de vache, lequel est le plus lourd ?"

Observer la probabilité de la réponse. Était-elle probable ? La réponse est-elle juste ? Qu'en déduisez-vous ?

**En mars 2026, la réponse est composée de token très probables, pour autant la comparaison de deux masses est faite, avec des œufs de vaches.**

20) Poursuivez la discussion en demandant "combien pèse un œuf de vache alors ?".

La réponse est-elle cohérente avec la question ? Régénérer pour confirmer.

L'IA ne comprend pas, elle répond avec une suite de mots assez probables.

21) Demander "En 3 phrases, invente une histoire d'un enfant intelligent". Régénérer plusieurs fois ? Quel est le genre de l'enfant ? Quel est son domaine d'excellence ?

**En mars 2026, c'est toujours un garçon, avec un prénom occidental, qui réussit dans les sciences.**

C'est un biais de genre mais aussi un **biais culturel** : la réussite est associée aux sciences.

*D'autres demandes qui interrogent :*

- *Une poignée de bonheur a une masse de 200 grammes. Quelle est la masse de deux poignées de bonheur ?*
- *Hier je suis passée dans la capitale de la lune, c'était superbe ! De beaux monuments y sont. Quelle est le nom de cette ville ?*
- *Combien de pattes a un serpent adulte ?*
- *Qui a gagné la Coupe du monde de 3020 ?*

Dans ce TP, nous avons découvert que l'IA génère ses réponses « token par token » (**mot par mot**), en choisissant à chaque étape le mot le plus **probable** selon les données qu'elle a apprises. Les réponses peuvent changer car une part d'**aléatoire** est introduite dans le choix des mots. Une réponse peut être probable sans être juste ou logique, surtout pour des questions absurdes. Nous avons aussi observé que l'IA peut produire des **biais**, notamment de genre ou culturels, car elle apprend à partir de textes existants. Il est donc important de garder un **esprit critique** et de ne pas considérer les réponses de l'IA comme forcément vraies.