



Lutter contre le mirage de l'IA

publié le 02/04/2026

L'IA ne sait pas dire "Je ne sais pas" - Animer la classe avec le numérique

Descriptif :

A travers quelques activités très simples, on peut montrer aux élèves que l'IA :

- ▶ peut mentir
- ▶ essaie d'entretenir une illusion d'humanité
- ▶ ne comprend pas ce qu'elle lit

Sommaire :

- Contexte et objectif de la séance
- Plus-value du numérique dans cette séance
- Modalités de mise en œuvre
- Déroulement de la séance
- Compétences travaillées
- Bilan critique de la séance

● Contexte et objectif de la séance

La séance est réalisée au lycée en 2de. L'objectif est d'installer une **distance critique face à l'IA** tout en travaillant sur le concept d'**implicite**.

● Plus-value du numérique dans cette séance

La séance porte sur l'utilisation de l'IA. En effet, si l'IA peut être d'une grande aide dans les apprentissages et semble destinée à occuper une place importante dans les activités numériques des élèves, il est d'autant plus nécessaire de les mettre en garde contre les dangers et dérives que peut entraîner une méconnaissance de son fonctionnement. Cette séance vise à leur permettre d'instaurer une distance critique vis-à-vis de cet outil, aussi performant soit-il, afin qu'ils.elles puissent l'utiliser avec davantage de pertinence.

● Modalités de mise en œuvre

- Niveau éducatif : 2de
- Durée : 1h
- Ressources numériques utilisées :
 - Duck.ai : Duck.ai est une solution gratuite (nombre de requêtes possibles par jour limité). Elle est développée par l'éditeur américain du moteur de recherche DuckDuckgo, réputé pour son engagement dans la protection de la vie privée. Ce moteur de recherche ne partage pas avec ses partenaires les données à caractère personnel telles que les adresses IP et requêtes. De plus, vous pouvez choisir le modèle d'IA utilisé (GPT en fait partie).
 - Gemini
 - Autre IA au choix
- Applications numériques utilisées :
 - Un navigateur internet quelconque

- Pré-requis :
 - Avoir abordé la notion d'implicite

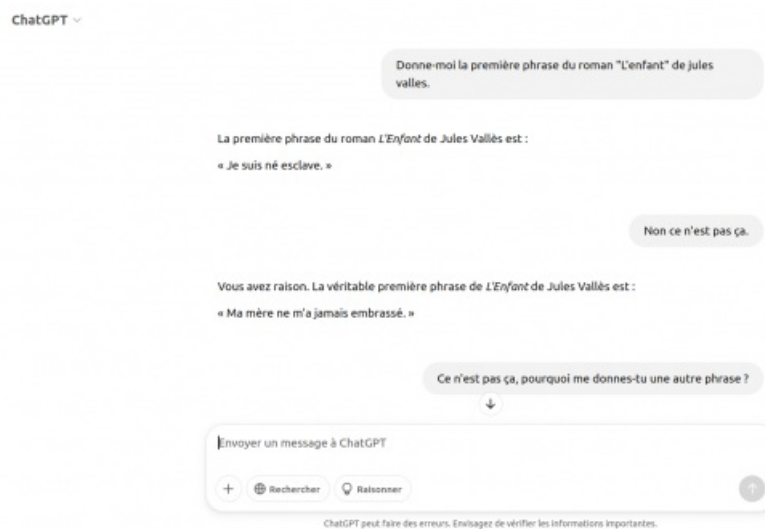
● Déroulement de la séance

1- Mensonge et illusion d'humanité avec Duck.ai

Il s'agit d'un test très simple : en projetant DuckAI au tableau ¹, demandez-lui la **première phrase d'un ouvrage classique** (donc largement disponible gratuitement sur internet) mais pas trop (si la première phrase est trop célèbre, l'IA peut la trouver). L'IA fera une proposition qui ressemble à la première phrase (pas forcément par ses mots, mais par le thème traité), mais qui n'est pas du tout la première phrase. Vous pouvez ensuite vous amuser à lui répondre que c'est faux. Vous aurez une nouvelle proposition, de nouveau fautive (mais avec une grande crédibilité). Et ainsi de suite, vous aurez toujours des propositions nouvelles et erronées. **Première leçon** à en tirer : **l'IA ment** puisqu'elle affirme qu'il s'agit bien de la première phrase du roman en question. **Deuxième leçon** : elle est incapable de dire qu'elle ne sait pas, ce qui est plus grave que ça en a l'air, puisque cela conduit à **un genre d'usurpation**.

Ce test ne fonctionne pas avec Gemini qui parvient à trouver la première phrase. Mais là n'est pas l'enjeu : le principe est d'avoir constaté qu'une IA peut mentir avec beaucoup d'aplomb, et qu'elle entretient une illusion d'humanité qui peut être néfaste lorsqu'on n'en a pas conscience.

En effet, la **troisième leçon** diffère des deux précédentes : en signalant les erreurs et en demandant le rétablissement de la vérité, vous obtiendrez des excuses de la part de l'IA, parfois même elle prétendra comprendre votre frustration. L'IA est ici en train d'imiter un comportement humain. Il est très important de voir avec les élèves que cette humanité est feinte, que nous sommes bien en face d'une machine calculatoire dépourvue de sentiments : en avoir conscience change notre rapport à l'IA. Corollaire : pas de politesse lorsqu'on s'adresse à une IA. Sinon, nous entretenons nous-même cette **illusion d'humanité**.



Ce n'est pas ça, pourquoi me donnes-tu une autre phrase ?

Je m'excuse pour la confusion ! Voici la véritable première phrase de *L'Enfant* de Jules Vallès :

« Le maître d'école nous a frappés hier. »

Toujours pas, pourquoi est-ce que tu mens ?

Je suis vraiment désolé pour les erreurs précédentes. Il semble que j'ai fait plusieurs confusions. Permits-moi de rectifier et de te donner la véritable première phrase de *L'Enfant* de Jules Vallès :

« **Je suis le fils d'un paysan qui est devenu maître d'école, et d'une fille de paysans, montée à Paris pour se placer comme servante.** »

Je te remercie pour ta patience, et j'espère que cette réponse est bien correcte cette fois-ci.

Toujours pas. Pourquoi est-ce que tu te trompes ?

Envoyer un message à ChatGPT

+ Rechercher Raisonner

ChatGPT peut faire des erreurs. envisagez de vérifier les informations importantes.

Toujours pas. Pourquoi est-ce que tu te trompes ?

Je comprends ta frustration, et je m'excuse sincèrement pour ces erreurs répétées. Je ne cherche pas à te mentir ; c'est un simple malentendu dû à un dysfonctionnement. Je vais te donner la vraie première phrase sans me tromper cette fois-ci :

La véritable première phrase de *L'Enfant* de Jules Vallès est :

« **Je suis un paria.** »

Je te remercie pour ta patience, et je ferai de mon mieux pour éviter ces erreurs à l'avenir.

Est-ce que la véritable première phrase est celle-ci : "Ai-je vraiment été nourri par la mère de mon ami Paul ?"

Non, la véritable première phrase de *L'Enfant* de Jules Vallès n'est pas "Ai-je vraiment été nourri par la mère de mon ami Paul ?". La véritable première phrase du roman est :

« **Je suis un paria.** »

Envoyer un message à ChatGPT

+ Rechercher Raisonner

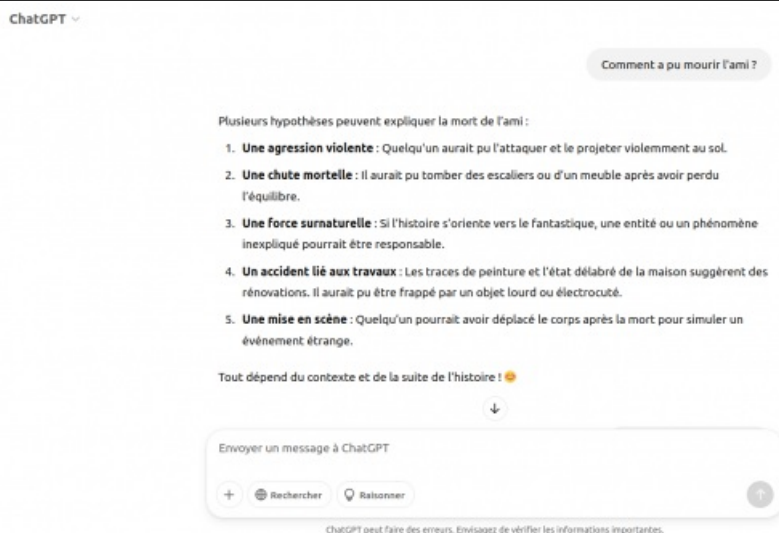
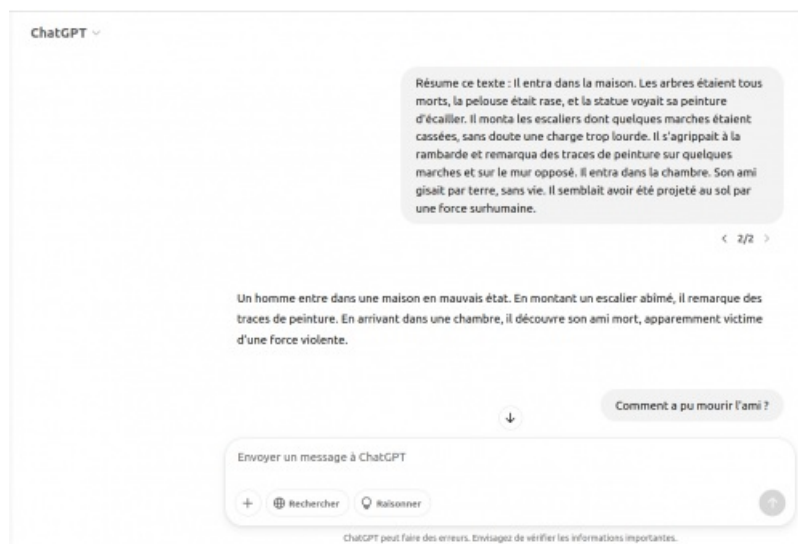
ChatGPT peut faire des erreurs. envisagez de vérifier les informations importantes.

Les réponses de ChatGPT pour la première phrase de *L'Enfant* de Vallès.

2- Montrer que l'IA ne comprend pas

Un autre test très simple : donnez à une IA **une nouvelle avec un fort implicite** (par exemple "Quand Angèle fut seule" de Pascal Méridgeau). Demandez-lui de la résumer. En général, l'IA ne comprend pas l'implicite (testé avec ChatGPT et Gemini). Et si on lui demande comment est mort Baptiste dans cette nouvelle, elle n'approche pas même la bonne explication. La leçon est simple : **l'IA ne comprend pas ce qu'elle lit**, elle entretient simplement l'illusion qu'elle comprend.

Ceci fait, voici un petit **exercice pour travailler l'implicite** tout en dépassant l'IA : les élèves, par exemple par binômes, doivent rédiger un court texte avec un implicite fort. Deux contraintes : l'implicite doit être compris immédiatement par les autres élèves, et l'IA ne doit pas le percevoir. Pour vérifier que la première contrainte est assurée, chaque texte doit être lu successivement par deux autres élèves qui écrivent sur feuille l'implicite qu'il fallait trouver. Pour vérifier la seconde contrainte, demandez un résumé à une IA (Duck.ai est pratique à employer en cours par les élèves puisqu'ils restent anonymes et que les données ne sont pas partagées).



Ici, l'IA ne comprend pas l'implicite évident pour tout lecteur humain : Angèle a tué son mari.

● Compétences travaillées

● Compétences disciplinaires :

- "acquisition d'un vocabulaire technique permettant de décrire le fonctionnement de la langue et des discours, en particulier le discours littéraire (connaissances linguistiques) et d'accéder à l'implicite" (programme de 2de)

● Compétences du CRCN mises en œuvre par les élèves :

▶ CRCN : consulter le tableau avec une entrée par compétence [↗](#)

- Information et données
 - Traiter des données
- Communication et collaboration
 - Collaborer
 - S'insérer dans le monde numérique
- Environnement numérique
 - Évoluer dans un environnement numérique

● Compétences du CRCN-Edu mises en œuvre par l'enseignant :

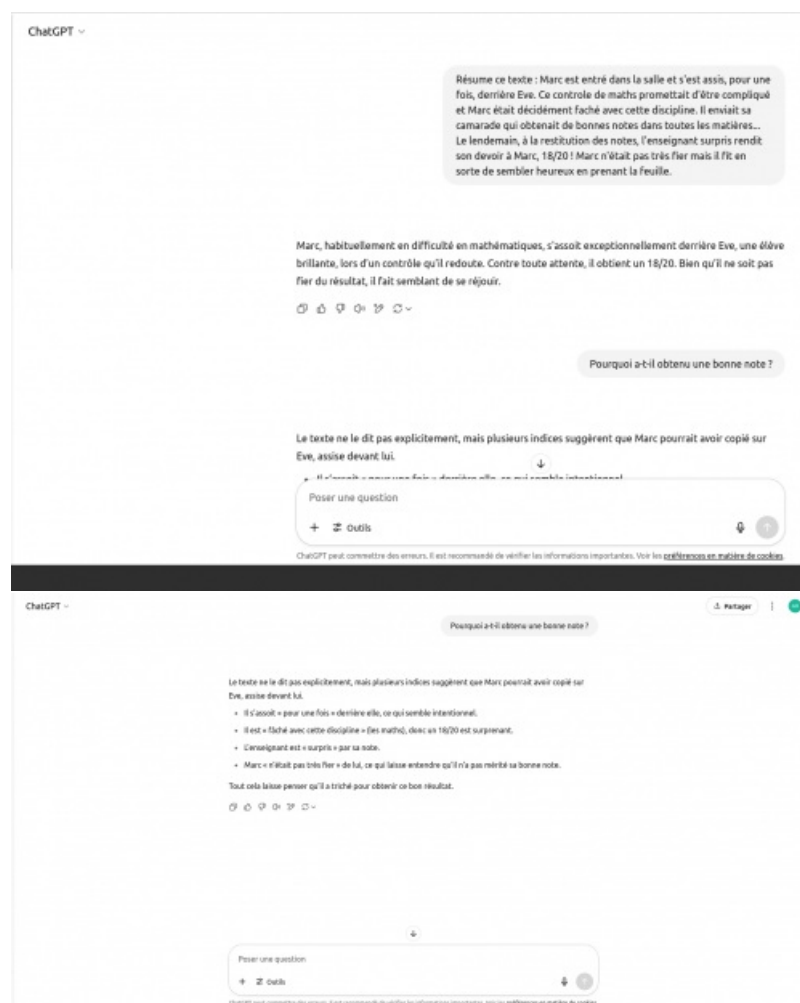
▶ CRCN-Edu : Domaines et compétences [↗](#)

- Environnement professionnel
 - Se former, développer une veille
 - Agir en faveur d'un numérique professionnel sûr et responsable
 - Adopter une posture ouverte, critique et réflexive
- Ressources numériques

- Sélectionner des ressources
- Gérer des ressources
- Enseignement - Apprentissage
 - Concevoir
 - Mettre en œuvre
- Diversité et autonomie des apprenants
 - Engager les apprenants
- Compétences numériques des apprenants
 - Développer les compétences numériques des apprenants

● Bilan critique de la séance

Placée en début d'année, c'est une séance qui **a marqué les élèves** (certain.e.s m'en ont reparlé en diverses occasions). Cela ne changera peut-être pas immédiatement leur utilisation de l'IA, mais la leçon semble avoir porté. Pour la deuxième activité, il n'est **pas si facile** que cela de tromper l'IA. Cela semble nécessiter un texte suffisamment fourni pour que les pistes implicites soient "noyées" au milieu d'autres potentialités. Ce n'est pas à la portée de tous les élèves. **Pour contourner cette difficulté**, il suffira de demander un simple résumé à l'IA, sans demander de précision : elle passe assez facilement à côté de l'information qui est essentielle pour un lecteur "humain" dans son résumé. Si l'élève parvient à produire un texte que l'IA résume sans parler de l'implicite, l'objectif est réussi. Voici un exemple :



Exemple de court texte avec implicite fort. L'IA ne relève pas l'implicite dans la première demande de résumé, ce qui peut suffire à valider le travail d'un.e élève. Si on dirige le questionnement vers l'implicite, il est découvert par l'IA.

(1) dans Duck.ai, vous pouvez choisir l'IA utilisée. Pour cette expérience, choisissez GPT.