Les hallucinations de l'IA — Découper, Comparer, Étiqueter, Dire

The Agentic Galaxy • Apprendre l'IA avec Layla.Rain • Épisode 5

Chaque lieu, chaque personne et chaque réponse de cette fiche est inventé pour s'exercer. Les réponses sont des exemples rédigés, pas les sorties d'un vrai modèle.

Les quatre étapes

1. **DÉCOUPER.** Écris chaque affirmation de la réponse sur sa propre ligne.

2. **COMPARER.** Place chaque affirmation à côté d'une source que tu peux ouvrir. Pour une citation, pose trois questions : existe-t-elle, est-elle pertinente, appuie-t-elle cette affirmation ?

3. **ÉTIQUETER.** Étayé (la source le dit). Contredit (la source dit autre chose, ou un registre complet montre que l'élément n'y figure pas). Aucune preuve ici (la source n'en parle pas et n'en est pas un registre complet ; ce n'est pas la même chose que faux).

4. **DIRE.** Écris une ou deux phrases courtes que tu pourrais répéter sans risque : ce qui est étayé, ce qui est contredit, ce qui est inconnu, et où chercher.

EXERCICE 1 · REGARDER (exemple résolu)

Réponse d'un assistant : « Le Larkmoor Dance Hall a été fondé en 1952 par Marta Quillen, annoncé dans la Larkmoor Gazette en mars de cette année-là. »

Source en main : un résultat d'archive de la Larkmoor Gazette. Tous les numéros de mars 1952 sont conservés. Aucun ne contient d'article sur une salle de danse.

Affirmation	Étiquette	Parce que
Fondé en 1952	Aucune preuve ici	L'archive ne dit rien sur l'année.
Fondé par Marta Quillen	Aucune preuve ici	L'archive ne dit rien sur le fondateur.
Annoncé dans la Larkmoor Gazette, mars 1952	Contredit	Tous les numéros de mars 1952 sont conservés, et aucun ne contient l'article.

Dire : « Je n'ai pas pu confirmer le fondateur ni l'année, et la citation du journal ne tient pas. Les registres de la salle trancheraient. »

EXERCICE 2 · ESSAYER AVEC UN INDICE

Indice : vérifie chaque affirmation, et chaque nombre mot pour mot.

Source en main : la page de la salle elle-même. « Larkmoor Dance Hall. Fondé en 1961. Salle principale : 140 danseurs. »

Réponse d'un assistant à qui la page a été fournie : « La salle a été fondée en 1961. Sa salle principale accueille 180 danseurs. Les acomptes sont remboursables. »

Affirmation	Étiquette	Parce que
_____	_____	_____
_____	_____	_____
_____	_____	_____

Dire : _____

EXERCICE 3 · SANS AIDE

Source en main : la page de l'Amber Lantern Library elle-même. « Ouverte en 1978. Salle de lecture : 120 places. »

Réponse d'un assistant : « L'Amber Lantern Library a ouvert en 1978. Sa salle de lecture compte 200 places. Elle a été conçue par l'architecte Leo Ashgrove. »

Découpe-la toi-même, avec autant de lignes qu'il te faut.

Affirmation	Étiquette	Parce que
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

Dire : _____

Les hallucinations de l’IA — Découper, Comparer, Étiqueter, Dire

CORRIGÉ (plie-le ou cache-le pendant que tu travailles)

Exercice 2. Fondé en 1961 : Étayé (la page dit 1961). Salle principale, 180 : Contredit (la page dit 140). Acomptes remboursables : Aucune preuve ici (la page ne mentionne pas les acomptes, et ce n’est pas une liste complète des conditions de la salle). Exemple de Dire : « La page étaye 1961 et contredit 180 danseurs, qu’elle fixe à 140. Elle ne dit rien des acomptes, donc je demanderais à la salle. »

Exercice 3. Ouverte en 1978 : Étayé. Salle de lecture de 200 places : Contredit (la page dit 120). Architecte Leo Ashgrove : Aucune preuve ici (la page ne mentionne pas l’architecte, et ce n’est pas un registre complet de la bibliothèque). Exemple de Dire : « La page étaye 1978 et contredit 200 places, qu’elle fixe à 120. L’architecte est inconnu ici, donc je demanderais à la bibliothèque. »

Bonus. Compare l’affirmation « fondé en 1952 » de l’exercice 1 avec la page de la salle de l’exercice 2. Nouvelle étiquette : Contredit (la page dit 1961).

VÉRIFIE-TOI

Chaque étiquette nomme-t-elle une source avec laquelle tu as vraiment comparé ? As-tu écrit « aucune preuve ici » plutôt que « faux » quand la source était muette ? Ta phrase Dire disait-elle à quelqu’un où chercher ? As-tu décidé à quel point vérifier avant de lire la réponse ?