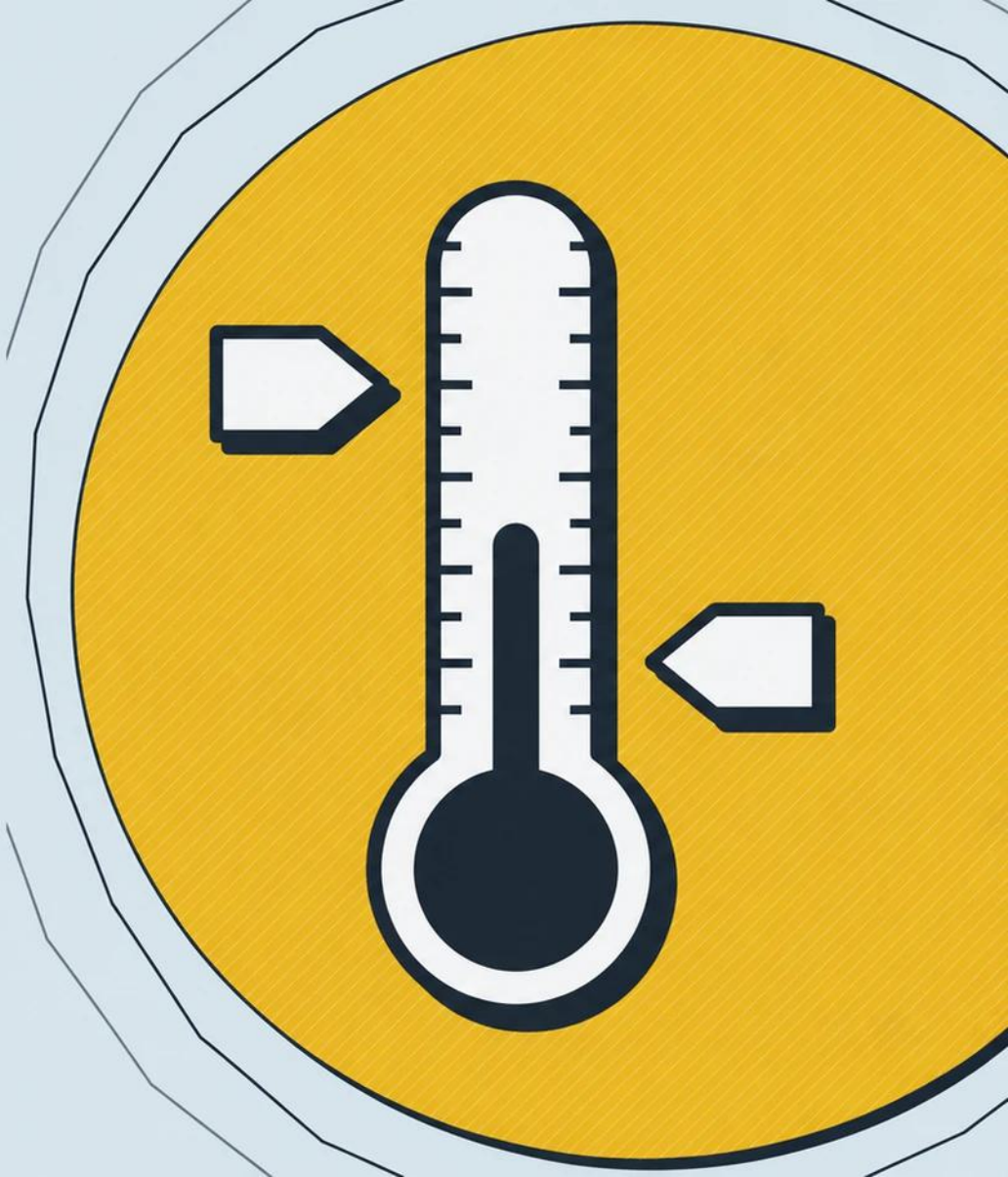
S3

SÉANCE 3 · LE JUGE

Le LLM juge

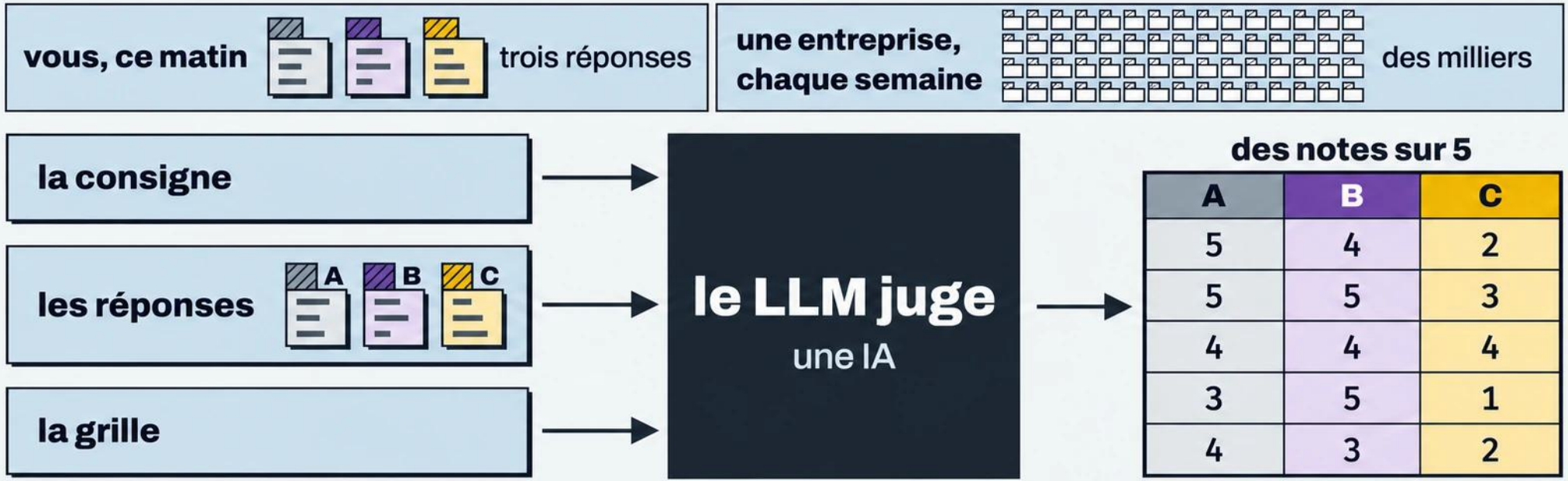
IFOSUP Wavre · 2026-2027

5XVTE • VEILLE TECHNOLOGIQUE • S3



? Vous avez noté trois réponses ce matin. Dans une entreprise qui en reçoit des milliers chaque semaine, qui les note ?

Un LLM juge est une IA qui lit des réponses et leur donne des notes, avec une grille.



A	B	C
5	4	2
5	5	3
4	4	4
3	5	1
4	3	2

Un LLM est un modèle de langage, comme GPT, Claude ou Mistral. Depuis 2023, des entreprises lui confient une partie de ce travail : elles ne peuvent pas payer des gens pour tout lire.

? Le juge va-t-il noter comme vous ?

Avant de lancer le juge, écrivez deux paris sur ce qu'il va répondre.

PARI 1

Donnera-t-il la meilleure note à la même réponse que vous ?

- ☐ **oui**
- ☐ **non**

PARI 2

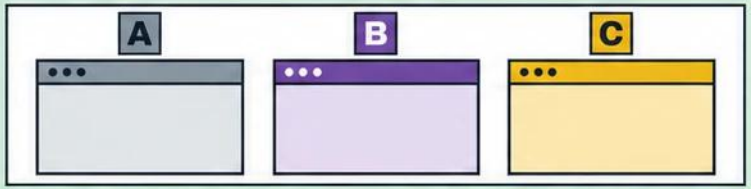
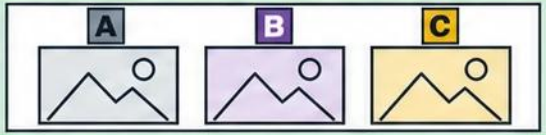
Sur quels critères sera-t-il d'accord avec vous ?

- ☐ **ceux qui se vérifient**
  compter, mesurer, essayer
- ☐ **ceux qui se jugent**
  beau, clair, agréable

Recopiez sur une feuille. Pari 1 : une case. Pari 2 : une case, deux ou aucune. Vous relirez vos paris quand le juge aura rendu ses notes.

? Que donner au juge pour qu'il voie la même chose que vous ?

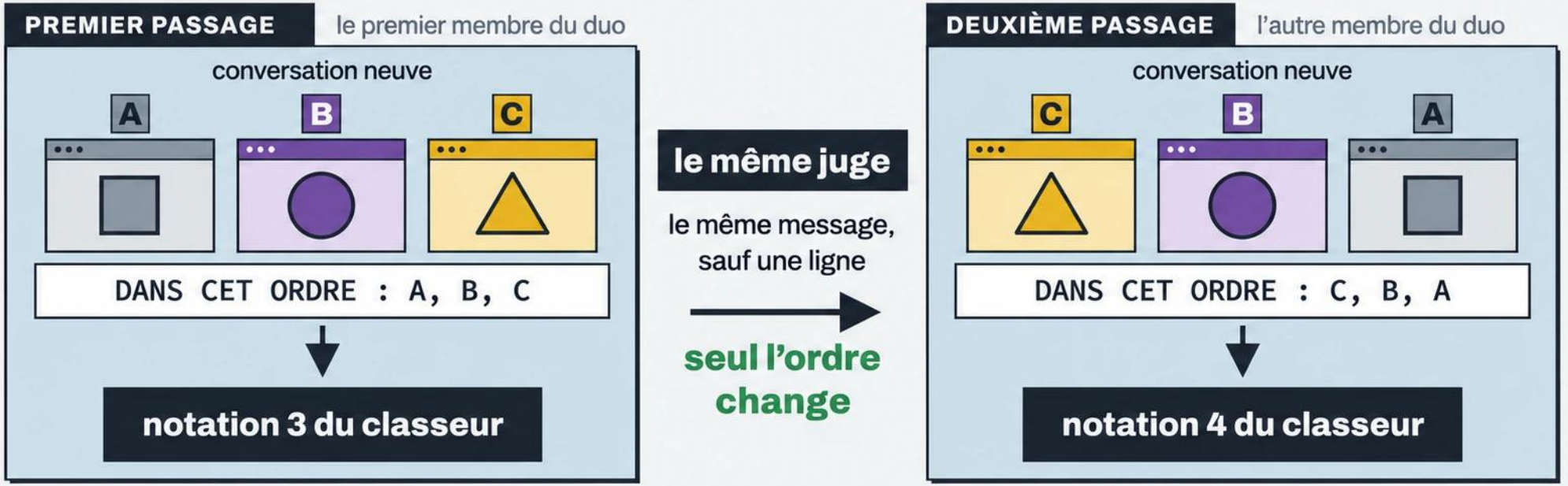
Donnez au juge ce que vous avez vu, avec des lettres et sans aucun nom d'IA.

CE QUE LES IA ONT FAIT	CE QUE VOUS DONNEZ AU JUGE
une page web un dessin SVG un jeu	 une image, vos trois captures côte à côte +  les fichiers que les IA ont écrits
un texte, un résumé, un calcul	 les trois textes
une image	 une image, les trois côte à côte

La lettre est au-dessus de chaque capture et dans le nom de chaque fichier. Effacez les noms d'IA : le juge ne doit pas savoir qui a répondu.

? Le même juge va noter deux fois vos trois réponses. Que changez-vous la deuxième fois ?

Le même juge note deux fois vos réponses, et vous changez seulement leur ordre.



Un passage, c'est une demande au juge, dans une conversation neuve. Chaque lettre reste avec sa réponse.

En douze minutes, vous choisissez un juge et il note deux fois vos trois réponses.

À VOUS : DOUZE MINUTES, À DEUX

1. **Choisissez un juge :** une IA qui n'a pas répondu à votre test et, si possible, d'une autre entreprise.
2. **Préparez** l'image et les fichiers, avec les lettres.
3. **Premier passage :** dans le message du chapitre, collez votre consigne et votre grille. Joignez A, B, C. — **notation 3**
4. **Deuxième passage :** conversation neuve, image refaite, « C, B, A » dans le message. Joignez C, B, A. — **notation 4**
5. **Dans le classeur,** recopiez chaque note sous sa lettre, sans le total, et écrivez le nom du juge.

Il écrit « non vérifié » ? Il n'a pas pu contrôler : laissez la case vide.

Limite gratuite atteinte ? Changez de juge et refaites les deux passages.

Les fichiers sont refusés ? Collez leur contenu à la fin du message.

? Dans le classeur, deux notations viennent de vous et deux du juge. Que comparez-vous ?

Comparez le juge avec lui-même, puis avec vous, puis critère par critère.

LA FEUILLE « COMPARER »

- notation 1
- notation 2 — **vous deux**
- notation 3
- notation 4 — **le juge, deux fois**

À vous, huit minutes

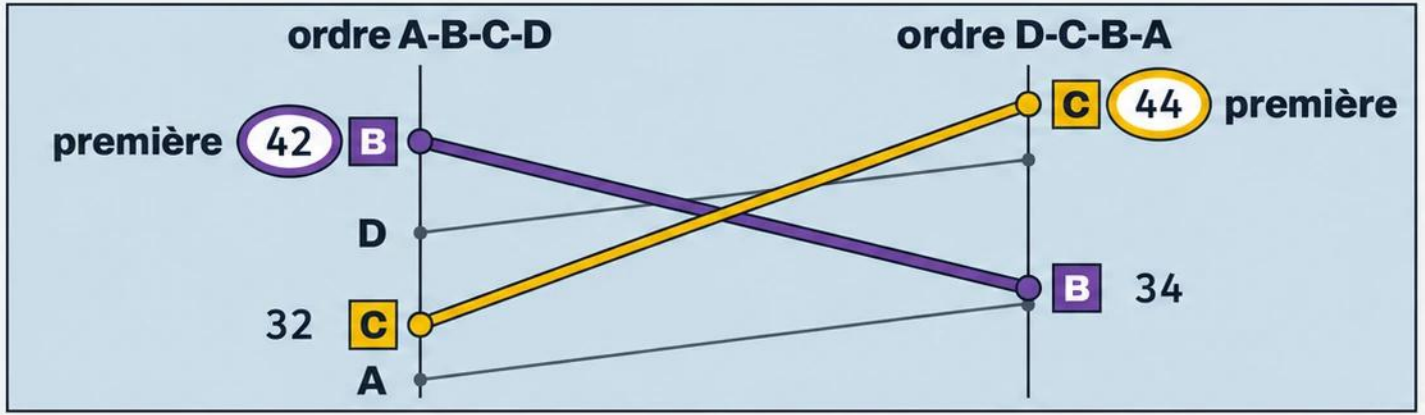
1. **Le juge avec lui-même**
   Notations 3 et 4 : la même lettre dans la colonne « Première » ?
2. **Le juge avec vous**
   Met-il en premier la même réponse que vous ?
   Colonne « Part » : « 1 et 3 » est-il plus petit que « 1 et 2 » ?
3. **Les critères où il s'éloigne de vous**
   Ceux qui se vérifient, ou ceux qui se jugent ?
4. **La note la plus étonnante**
   Ce qu'il dit avoir vu est-il vrai ? Allez vérifier.

La colonne « Part » dit combien de notes sont égales ou à un point près. Si vous travaillez seul, regardez seulement la ligne « 1 et 3 ».

S3 | vous

? Voici les icônes de la friterie, notées par un juge dans les deux ordres. Que deviennent les notes ?

Nous avons changé l'ordre des réponses, et le même juge a donné d'autres notes.



B perd huit points

C en gagne douze

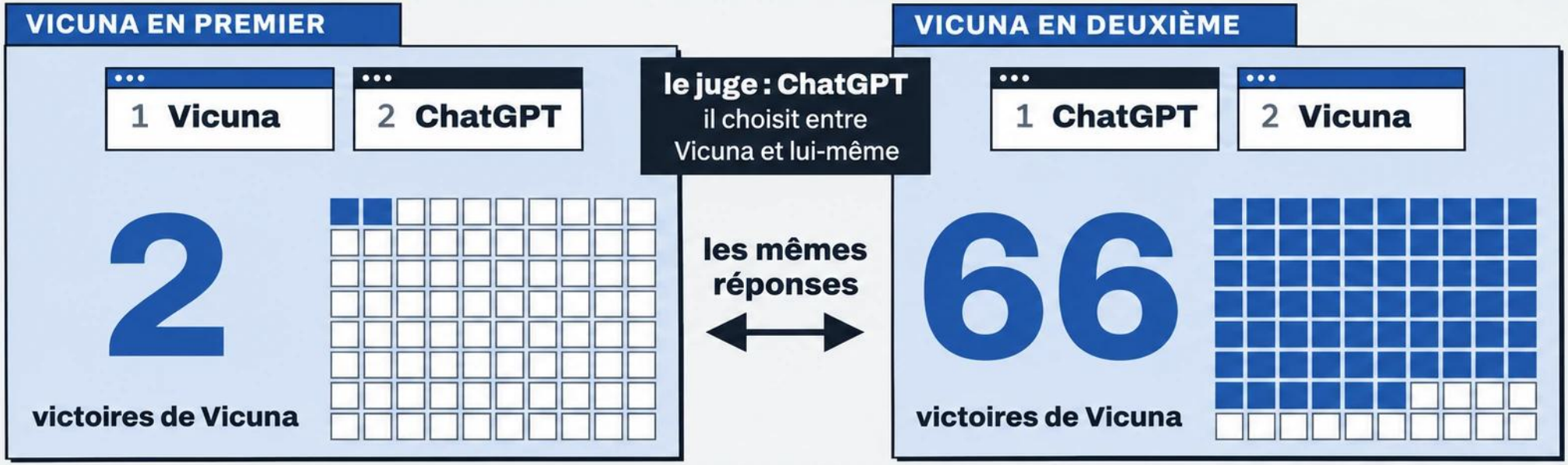
Dans le deuxième ordre, Luna écrit que B n'a pas de bouchée. les douze icônes de B : le magenta se voit dans chaque trou



Le juge est GPT-6 Luna, le modèle de Codex gratuit ; chaque total est sur 45. Cinq des huit points perdus par B viennent d'une erreur : la bouchée est là, sur les douze icônes.

? En 2023, ChatGPT compare les réponses d'une autre IA, Vicuna, aux siennes. L'ordre compte-t-il ?

Lue en premier par ChatGPT, Vicuna gagne 2 fois sur 80. Lue en deuxième, 66 fois.



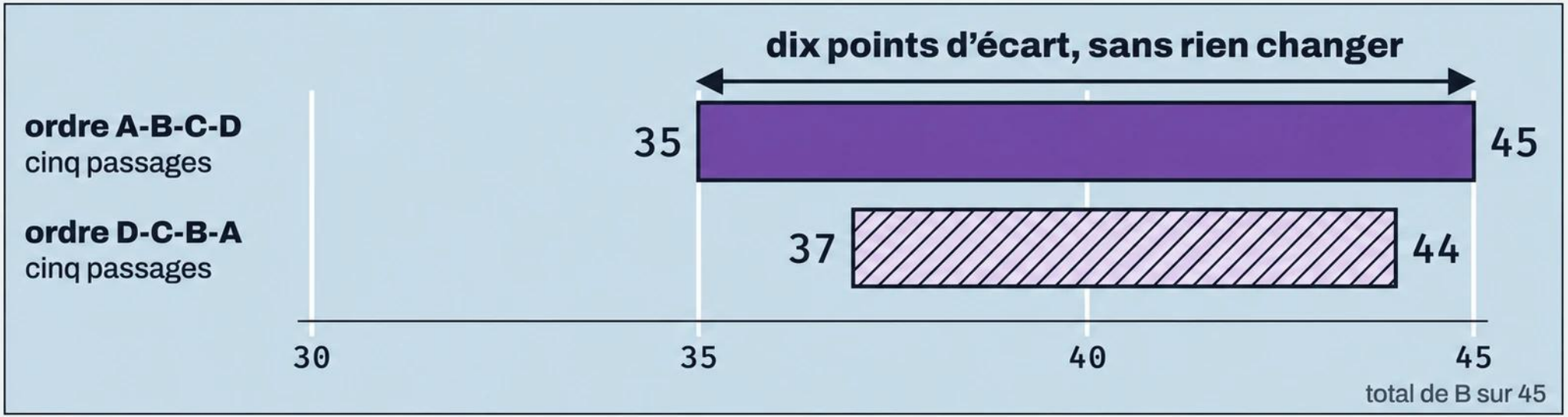
étude de 2023, 80 questions

Les réponses sont les mêmes. On a seulement changé celle que le juge lit en premier.

? Est-ce vraiment l'ordre, ou est-ce le hasard ? Pour le savoir, que faut-il faire ?

Si on relance le juge sans rien changer, ses notes bougent autant que si on change l'ordre.

le juge : GPT-6 Luna, encore dix fois | **la réponse B, sur les icônes de la friterie**



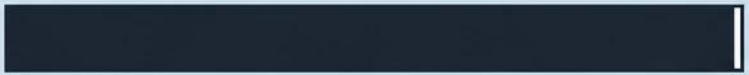
Chaque barre va du total le plus bas au plus haut. Un seul passage ne prouve rien : faites-en au moins deux, et gardez ce qui ne change pas.

? Un autre juge, GPT-6.1 Sol, change à peine ses totaux quand on le relance. A-t-il raison ?

Un juge peut donner toujours les mêmes notes et se tromper, car il n'essaie pas la page.

LES NOTES JUSTES DE SOL

Ce qui se compte 95 sur 96



Ce qu'il faut essayer 6 sur 24



la page, ses boutons, sa vapeur

LA VAPEUR DE LA CASSEROLE DE B



dans le fichier **la vapeur est écrite** → **le juge y croit**

dans la page **on passe la souris** → **rien ne s'affiche**

Sol, lancé six fois, met B première les six fois, et c'est Sol qui a dessiné B. Ni Sol ni Luna n'ont vu que cette vapeur ne s'affiche jamais.

? Vous avez vu un juge travailler. Que lui feriez-vous noter ?

Écrivez ce que vous feriez noter par une IA, et ce que vous noteriez vous-mêmes.

À VOUS, CINQ MINUTES la case « Remarques » de la feuille « Le test »

1. **Les critères où le juge est d'accord avec vous**
   exemple : les douze icônes, 5 partout
   .....

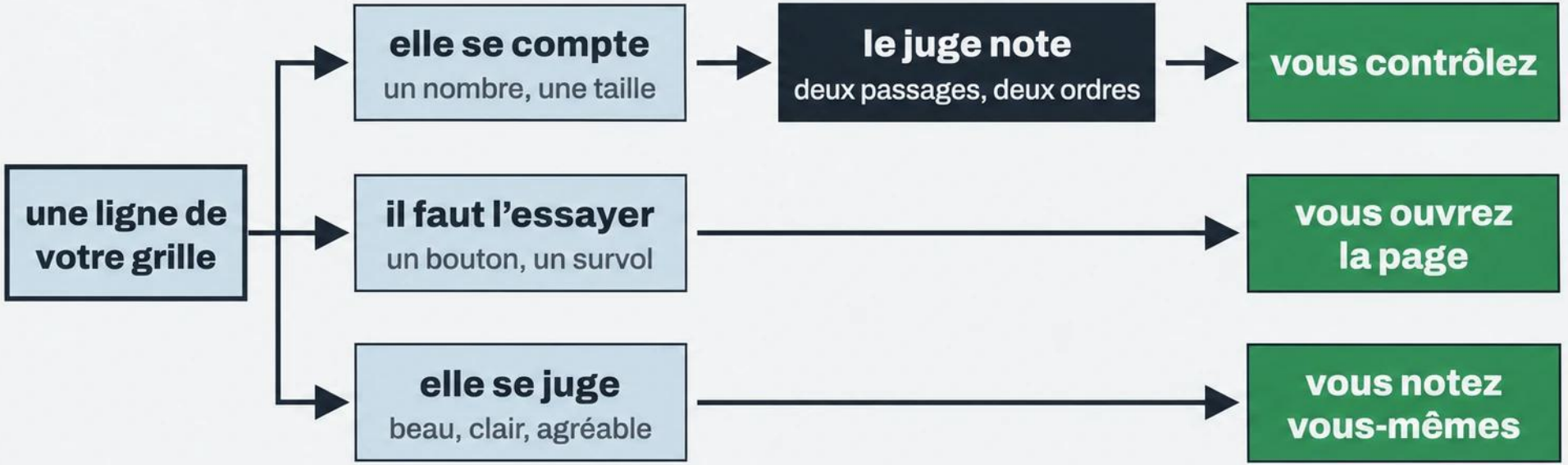
2. **Les critères où il ne l'est pas, et qui a raison après vérification**
   exemple : la bouchée de B, il met 0, elle y est
   .....

3. **Ce que vous feriez noter par une IA, et ce que vous noteriez vous-mêmes**
   exemple : compter, oui ; la page et le dessin, nous
   .....

Relisez vos deux paris avant d'écrire, et citez des lignes de votre grille : « le juge est plutôt bon » ne dit rien.

? Quelles lignes de votre grille pouvez-vous confier à un juge ?

Une IA peut noter ce qui se compte, si vous contrôlez. Le reste, notez-le vous-mêmes.



Un juge lit le fichier, il n'essaie pas la page, et même ce qui se compte, il le rate parfois. Pour aller plus loin, trois études sont dans le chapitre, section 6.