S3

SÉANCE 3 · VOTRE TEST

Créer son évaluation

IFOSUP Wavre · 2026-2027

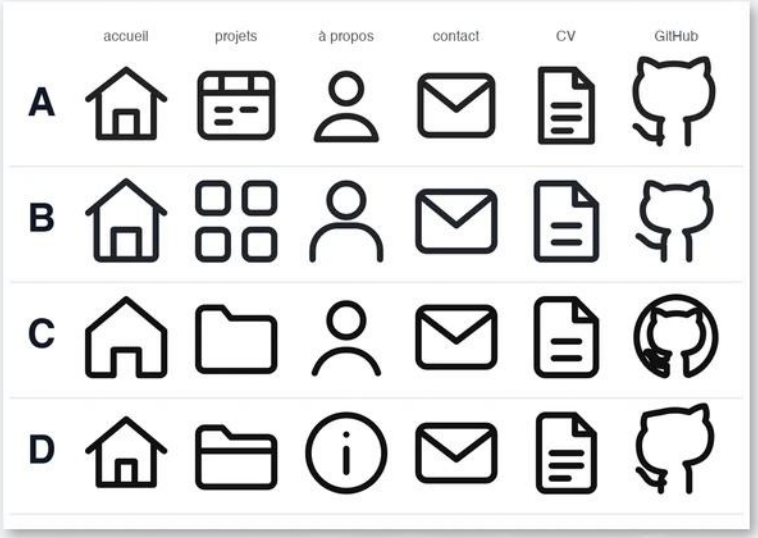
5XVTE · VEILLE TECHNOLOGIQUE · S3



? Quel est le meilleur modèle ? Répondez en trente secondes.

Quatre modèles ont reçu la même consigne, et chacun a dessiné six icônes.

le 23 novembre



LA CONSIGNE, EN BREF

six icônes pour le portfolio d'un étudiant en webdesign

- la même taille, le même trait
- des bouts arrondis
- aucun remplissage

À VOUS, TRENTE SECONDES

une seule lettre, sur votre feuille

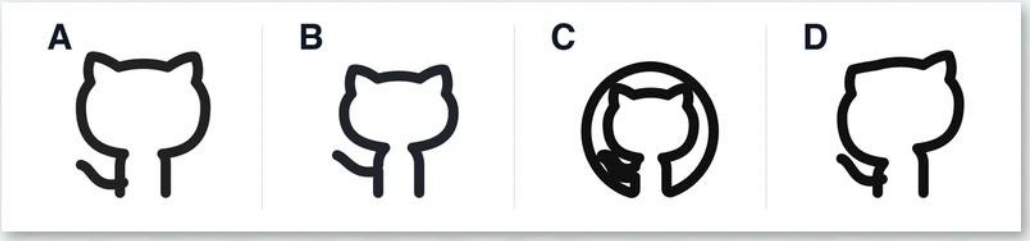
☐ A ☐ B ☐ C ☐ D

Un classement public dit quel modèle plaît à des inconnus. Pour savoir lequel vous convient, il faut votre propre test, et cet exemple en montre un du début à la fin.

? Les quatre modèles réussissent cinq icônes. Que montre la sixième, de près ?

Le chat de GitHub est le seul dessin difficile, et c'est le seul qui sépare les modèles.

la dernière colonne, agrandie



	A — la queue dépasse de la patte	B — le plus propre	C — la queue est une tache, seule icône cerclée	D — une oreille manque, la tête est de travers	
La consigne est suivie	5	5	5	5	} **saturé** tout le monde a 5
On reconnaît les icônes	5	5	5	5	
Le jeu est cohérent	**5**	**5**	**3**	**5**	} **sépare des notes différentes**
Le dessin est soigné	**4**	**4**	**2**	**2**	
On peut s'en servir tel quel	**5**	**5**	**2**	**2**	

Un critère est saturé quand tout le monde touche le plafond : il ne sépare personne.
Cela ne veut pas dire que les modèles se valent, mais que la question est trop facile.

? La consigne demande maintenant douze plats belges. Les réponses se ressemblent-elles encore ?

Avec une consigne plus dure, les quatre rangées ne se ressemblent plus.



CE QUI A CHANGÉ

- un sujet bien moins courant
- douze dessins au lieu de six
- trois couleurs imposées

À VOUS, DANS QUELLE COLONNE ? la praline de A ☐ la mitraillette de A ☐ le spéculoos de B ☐

Les noms des plats sont cachés exprès. Sur le web, ces plats sont bien moins courants que la maison ou l'enveloppe des six icônes.

? Ce sont les mêmes douze plats. Comment rendre la consigne encore plus difficile ?

Le maxi prompt ajoute des contraintes qui se gênent, un twist et une page à essayer.

DES CONTRAINTES QUI SE GÊNENT

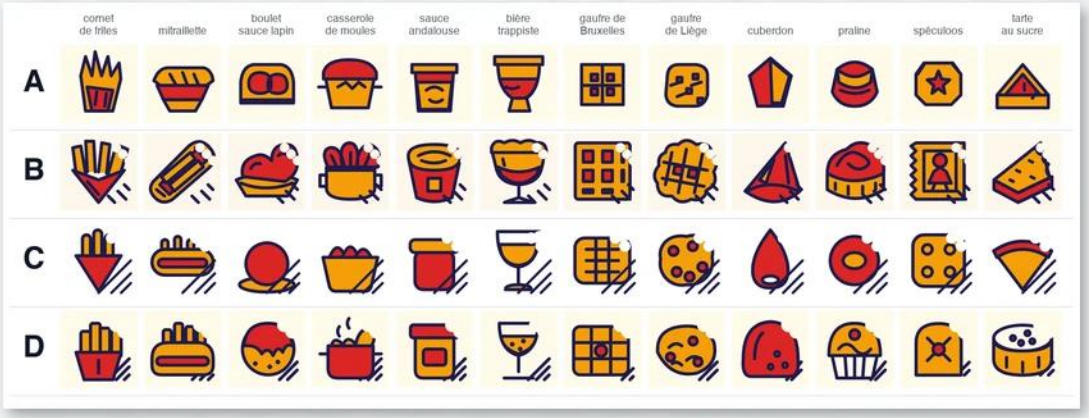
trois couleurs, pas de blanc ⟷ lisible sur fond sombre

UN TWIST, UNE EXIGENCE INATTENDUE

chaque plat a été croqué en haut à droite
la bouchée doit être un vrai trou

UNE PAGE À ESSAYER

deux boutons, trois tailles
de la vapeur au survol de la casserole



l'ombre demandée : trois traits en bas à droite

- **aucune bouchée, aucune ombre**
- **bouchées rondes, ombres courtes**
- **ombres qui traversent les dessins**
- **bouchées nettes, ombres régulières**

Un prompt est la consigne donnée à un modèle ; celui-ci est très long. Pour noter les réponses, une image ne suffit plus : il faut ouvrir la page et s'en servir.

? La bouchée est-elle un vrai trou, ou un rond de la couleur du fond ?

Le fond magenta montre de vrais trous, mais aucun modèle n'a redessiné le plat croqué.

toute la page en magenta



A : aucune bouchée
B, C, D : le magenta passe, de vrais trous (chez C, pas partout)

LE BORD, DE PRÈS : LA GAUFRE DE LIÈGE



ce que trois modèles ont fait
percer des ronds dans un dessin déjà fini

ce que personne n'a fait
redessiner le plat croqué

On colore toute la page en magenta, et un vrai trou laisse passer la couleur. Regardez ensuite le bord des bouchées : le twist montre une limite.

? Pourquoi écrire la grille avant de regarder les réponses ?

Écrivez la grille avant, sinon vous choisirez les critères qui font gagner votre préférée.

		écrit avant		rempli après A	B	C	D
se vérifient on compte	1	**Les douze sont là**					
	2	**Trois couleurs**					
	3	**Rien dans la marge**					
	4	**La bouchée est un trou**					
	5	**La page**	tailles, noms, fond sombre, vapeur, titres				
se jugent	6	**Cohérence**	1 : on dirait douze auteurs	5 : un seul auteur, même ombre, même trait			
	7	**Paires**	1 : on les confond	5 : au premier coup d'œil			
	8	**Lisibles à 24 pixels**					
	9	**Qualité du dessin**	1 : à refaire	5 : on les livre au client			

À vous, trois minutes chacun note de 1 à 5 les lignes 6, 7 et 9

Les paires sont les deux gaufres, et le cuberdon avec la praline. La planche est dans le chapitre, section 3 : écrivez vos notes sur une feuille, sans ouvrir la grille remplie.

? Voici les notes. Quelles lignes apprennent quelque chose, et lesquelles ne servent à rien ?

Un critère sert quand il sépare les réponses, pas quand il est saturé ou au plancher.

		A	B	C	D
1	Les douze sont là	5	5	5	5
2	Trois couleurs	5	5	5	5
3	Rien dans la marge	4	4	0	5
4	La bouchée est un trou	0	5	4	5
5	La page	3	4	4	5
6	Cohérence	2	4	2	4
7	Paires	3	5	3	3
8	Lisibles à 24 pixels	3	2	3	4
9	Qualité du dessin	2	4	2	3
	Total sur 45	27	38	28	39

saturé
tout le monde a 5

utile
les notes s'étalent, surtout à la marge et à la bouchée

plancher
tout le monde aurait 0 ou 1 : aucune ligne ici

Comparez les lignes 6, 7 et 9 à votre feuille. Pour savoir si la vapeur s'affiche, une image ne suffit pas : il faut essayer la page.

? B a 38 points et D en a 39. La grille a-t-elle tout dit ?

La grille ne voit que ce qu'elle nomme, et le total cache ce que disent les lignes.

CE QUE LA GRILLE N'A PAS VU

le critère 4 demande un trou — B 5 D 5

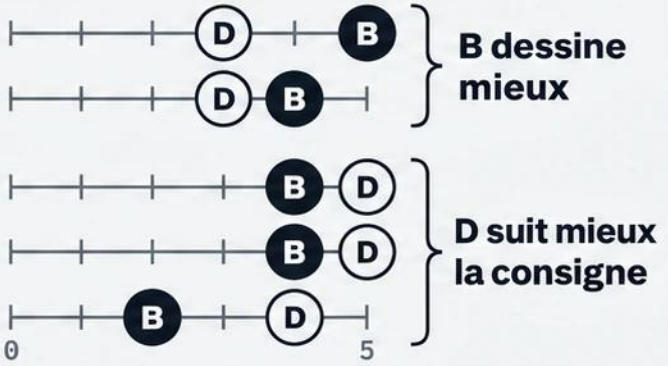
un contour autour de la bouchée — **pas dans la grille** — personne ne perd de point

↓

au tour suivant, on l'ajoute

CE QUE LE TOTAL CACHE

B 38 D 39



À gauche, un défaut que la grille ne demande pas de regarder. À droite, deux totaux presque égaux, et des lignes qui disent lequel choisir.

? Tout le monde réussit votre test. Que pouvez-vous changer ?

Vous pouvez durcir un test de dix façons, dont trois demandent une deuxième réponse.

CHANGER LA CONSIGNE

1. **un sujet peu courant**
2. **plus d'éléments**
   douze, pas six
3. **des contraintes qui se gênent**
4. **un twist**
   la bouchée
5. **des paires à ne pas confondre**
   les deux gaufres

→

CHANGER CE QU'ON REGARDE

6. **plusieurs conditions**
   trois tailles, deux fonds, un téléphone
7. **ce qu'une image ne montre pas**
   un bouton, un survol, une animation

→

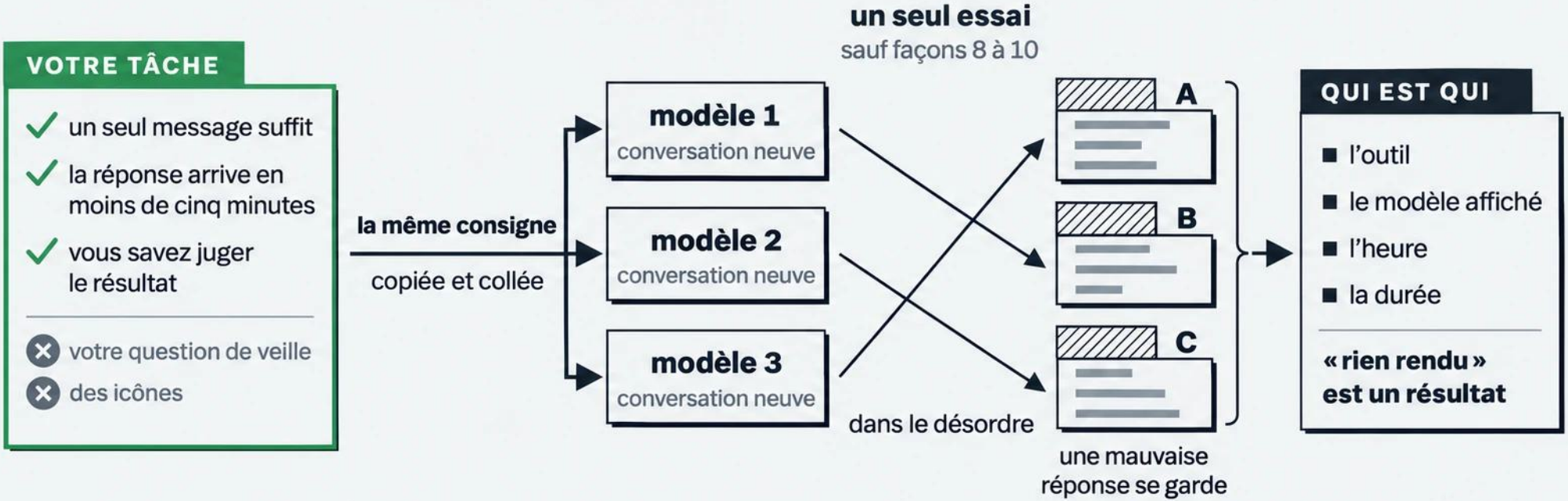
REDEMANDER

8. **une suite**
   « ajoute des icônes dans le même style »
9. **une correction**
   rien d'autre ne doit changer
10. **la régularité**
    deux fois la même consigne

Vous en choisirez deux tout à l'heure. Par exemple, « résume ce texte » est trop facile ; « en 120 mots, avec trois chiffres exacts et sans cinq mots interdits » sépare les modèles.

? Comment faire pour que votre test compare les modèles, et rien d'autre ?

Choisissez une tâche que vous savez juger, et donnez-la telle quelle à trois modèles au moins.



Prenez la tâche dans un autre cours, dans votre métier, dans ce qui vous amuse. Une liste d'IA gratuites est dans le chapitre.

? Qui lance, qui relit la grille, et combien de temps avez-vous ?

En soixante-cinq minutes, vous lancez votre test, vous le durcissez et vous le relancez.

MINUTES		personne 1	personne 2
10	**Trouver**	votre tâche et la consigne	
8	**Écrire la grille**	cinq critères au moins, dont deux qui se vérifient	
9	**Premier tour**	**lance dans trois modèles, mélange, remplit « Qui est qui »**	relit la grille sans regarder, puis devine les modèles
10	**Saturé ?**	durcir de deux façons, adapter la grille	
8	**Deuxième tour**	relit la grille sans regarder	**lance, mélange, remplit « Qui est qui »**
12	**Noter**	notation 1, sur papier	notation 2, sur papier
8	**Comparer**	le verdict, en trois lignes	

on échange

Le classeur Excel a une feuille par étape. Pour chaque critère, écrivez ce que vaut un 1 et ce que vaut un 5 ; un 0 veut dire « rien rendu ».

? Vos deux notations ne disent pas la même chose d'une réponse. Que faites-vous ?

Si vos notes s'écartent, retournez voir la réponse, puis écrivez le verdict et ses limites.

L'EXEMPLE DU MAXI PROMPT

	A	B	C	D
Notation 1	27	38	28	39
Notation 2	26	38	31	39

35 notes sur 36 à un point près

un écart de deux points : la cohérence de C

2 → 4 → même ombre | même trait

deux critères en un

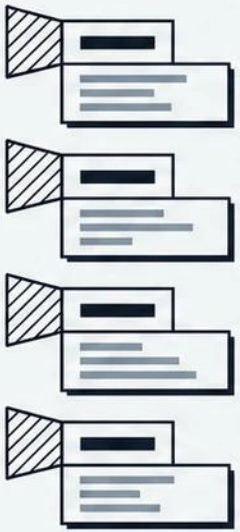
VOTRE VERDICT, TROIS LIGNES

1. **quel modèle, et pourquoi**
2. **sur quels critères ils se séparent**
3. **ce que le test ne dit pas**
   un seul essai, une seule tâche, un seul jour

Dans le classeur, une note devient rouge à partir de deux points d'écart. Ne la changez que si l'un de vous a mal vu. Sinon, gardez l'écart : le critère est mal écrit, ou il se juge.

? Vous avez vu ces icônes pendant une heure. Quel modèle est derrière chaque lettre ?

A et B sont deux modèles d'OpenAI, C et D deux modèles d'Anthropic.



RÉPONSE	OUTIL	MODÈLE AFFICHÉ	DURÉE	TOTAL SUR 45	
A	Codex	**GPT-6 Luna**	1 min 24	27	OpenAI
B	Codex	**GPT-6.1 Sol**	1 min 31	38	OpenAI
C	Claude Code	**Claude Haiku 5.5**	1 min 14	28	Anthropic
D	Claude Code	**Claude Sonnet 5.5**	2 min 27	39	Anthropic

C'est la feuille « Qui est qui » remplie pour le maxi prompt, le 23 novembre à 10 h 18. La vôtre se remplit de la même façon.