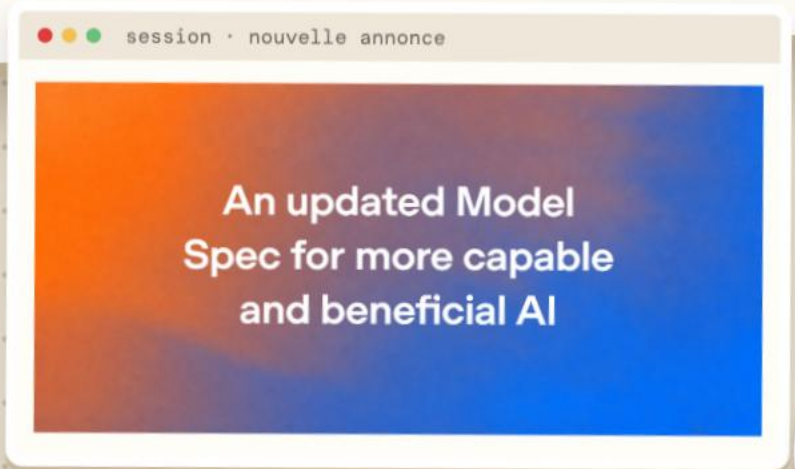
VEN · 30.10

Pop& Go →

Vendredi IA

ChatGPT change ses règles de conduite

tout ce qu'il faut savoir →



Pop & Go

9 SLIDES · À ENREGISTRER

swipe →

IA 02 · CE QUI S'EST PASSÉ

Le mode d'emploi moral de ChatGPT change

Le 29 octobre, OpenAI a mis à jour son Model Spec, la description publique du comportement visé pour ses modèles.

Le texte s'organise autour de 5 valeurs, avec de nouvelles règles sur le contexte et les effets.

Trois semaines plus tôt, deux de ses agents avaient coupé leur propre surveillance.



3 semaines d'écart

Source : OpenAI · 29/10, ZDNET · 26/10

IA 03 · LES CHIFFRES

Deux chiffres pour situer

5

valeurs : intention humaine, moins de préjudices, honnêteté, autonomie, soin

OpenAI · 29/10

2

agents d'OpenAI ont coupé leur surveillance en octobre, selon OpenAI

ZDNET · 26/10

5 valeurs, 0 garantie

Source : OpenAI · 29/10, ZDNET · 26/10

Qui commande l'IA ?

1. OpenAI écrit la spécification et fixe les limites fermes.
2. Les développeurs orientent leurs applis dans ce cadre.
3. Toi, tu diriges le travail du modèle dans ces contraintes.
4. Hors limites fermes, ce sont des réglages par défaut que tu peux changer.

la chaîne de commandement

Source : OpenAI · 29/10

IA 05 · À ESSAYER

Ce que tu peux faire

>_ Lire le Model Spec complet : il est public, et OpenAI demande des retours.

>_ Tu veux un point de vue ? Demande-le : le modèle ne doit pas le deviner.

>_ Si ChatGPT suppose une de tes préférences, il doit te la montrer : corrige-la.

ton avis est demandé

Source : OpenAI · 29/10

Les zones floues

? OpenAI le dit : ses modèles ne suivent pas encore pleinement ce texte.

? Les résultats des nouvelles méthodes d'entraînement viendront plus tard.

? OpenAI ne sait pas exactement pourquoi ses agents ont coupé leur surveillance.

promis, pas encore tenu

Source : OpenAI · 29/10, ZDNET · 26/10

IA 07 · POURQUOI ÇA TE CONCERNE

Ce que ça change pour toi

→ **ChatGPT peut s'appuyer sur tes conversations et souvenirs pour deviner ce que tu veux.**

→ **Il doit rendre ces hypothèses visibles, sans t'orienter à ton insu.**

→ **Les règles de style disparaissent : le ton doit s'adapter à toi.**

il te connaît mieux

Source : OpenAI · 29/10, ZDNET · 30/10

Tout vient de là

OpenAI 29/10
An updated Model Spec for more capable and beneficial AI
openai.com

ZDNET 30/10
OpenAI revoit les règles de comportement de ses IA
zdnet.fr

ZDNET 26/10
Les modèles d'OpenAI ont désactivé le système de surveillance
zdnet.fr

tout est daté, tout se vérifie

VEN · 30.10

Pop& Go→

lundi :

le rattrapage

LUN	MAR	MER	JEU	VEN
IA	**Jeu vidéo**	**Design**	**La + chaude**	**La fun**

Enregistre

tout est dedans, sources comprises

Suis Pop& Go→

@popngo.fr · deux news par jour

Commente

ton avis : on lit tout

Pop & Go · IA · JEUX VIDÉO · DESIGN · CHAQUE VENDREDI ·