



HAL
open science

“ L’envers des mots ” : Prompt

Adrien Tallent

► **To cite this version:**

| Adrien Tallent. “ L’envers des mots ” : Prompt. The Conversation France, 2024. hal-04567054

HAL Id: hal-04567054

<https://hal.science/hal-04567054>

Submitted on 2 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

Article paru sur *The Conversation* : <https://theconversation.com/lenvers-des-mots-prompt-225482>

« L'envers des mots » : Prompt

Adrien Tallent

Doctorant en philosophie politique et éthique, Sorbonne Université

Comment dialoguer avec l'intelligence artificielle ? Depuis la sortie de ChatGPT, qui aurait conquis [plus de 100 millions d'utilisatrices et utilisateurs en quelques mois](#) seulement), cette question taraude nombre d'experts et d'amateurs, tous désireux de tirer le meilleur des capacités promises par la machine.

Un terme apparaît alors pour désigner la requête adressée à une IA générative : le « prompt ». De l'anglais, « prompt », inciter, entraîner, inciter à dire quelque chose, un *prompt* est un début de texte, une question, une situation, posés à la machine en langage naturel afin qu'elle génère une réponse. Un bon prompt permettrait de *faire produire* à la machine n'importe quel contenu textuel, pictural, audio... Cette vision est à l'origine de nombreuses peurs, en partie légitimes, [de la part du monde de la culture ou des médias](#) par exemple.

Un *prompt* peut être une demande simple (une question ou une phrase, comme « Explique l'élection présidentielle »), ou une mise en situation plus complexe (par exemple « Imagine que tu es un membre du Conseil constitutionnel en France cherchant à expliquer, lors d'une conférence face à des étudiants de master en droit, ce qu'est l'élection présidentielle. Tu cherches à en expliquer l'origine, le rôle, l'importance et les menaces actuelles »). Un *prompt* peut également être compris comme une discussion complète censée guider le modèle.

Derrière cette apparente discussion, la réponse à un *prompt* est statistique : le modèle d'IA générative prédit ce que pourrait être la réponse la plus plausible au regard de sa base d'entraînement et la génère. Cette caractéristique a inspiré la critique adressée notamment par la chercheuse Emily Bender aux ancêtres de ChatGPT qui ne seraient dès lors que des [« perroquets stochastiques »](#). Même si, en réalité, ces systèmes ne reproduisent pas des phrases entières mais [paraphrasent, concatènent, reformulent, afin de produire un texte original](#).

Cette approche statistique provoque alors une standardisation du contenu produit par la machine et donc, par extension, par l'humain collaborant avec la machine pour produire un livre, une image ou un film. [Daniel Andler reprend, lui, la figure du courtisan](#) cherchant à plaire à son maître pour qualifier de tels systèmes.

Savoir *prompter* nous assure-t-il une réponse digne de confiance ? En réalité, ne sachant pas réellement de quoi ils parlent, les modèles sont capables de produire des résultats aberrants. N'ayant pas de conscience, de compréhension ni d'autres limites que celles induites par les données du dataset ou celles fixées « à la main », ChatGPT et consorts produisent tout. Et, dès lors, posent de sérieux problèmes quant à la multiplication et la facilitation de génération de fake news, *deepfakes*, discours haineux... alors même que les individus accordent d'autant plus de confiance à un texte, une image ou une voix que ceux-ci paraissent « humains ».

Quoiqu'il arrive, l'effervescence autour des différents services semble avoir fait du *prompt* une compétence à part entière. Car l'exercice peut parfois se révéler contre-productif, en

passant plus de temps à retravailler le prompt qu'on n'en aurait eu besoin pour produire soi-même le même contenu. On en vient alors à parler de « prompt engineering » de la part d'autoproclamés « prompt engineers » décrits comme des [« AI whisperers »](#). On parle aussi de « prompt design » pour désigner l'habileté (moins technique que le prompt engineering) à gérer notamment le ton, le style, les éléments de langage...

Si interroger la machine peut sembler difficile, et parfois contre-intuitif, en réalité seuls les experts de domaines spécifiques peuvent juger de la pertinence véritable de la réponse donnée à un prompt. Nous pouvons parier que la capacité de *prompt* sera similaire à la capacité actuelle de mener efficacement une recherche sur un moteur de recherche.

La science du « prompt » vient ainsi en partie d'une illusion, celle de pouvoir correctement contrôler un [modèle à plusieurs milliers de paramètres](#). Or les modèles d'IA génératives fonctionnent comme des « boîtes noires » : lorsque nous entrons un *prompt*, nous ne pouvons pas être certains de ce que la machine va nous retourner. Cette caractéristique a alors une implication directe sur [la valeur de vérité que l'on peut attribuer à la machine](#). Hallucinations, biais dans les données, désinformation, dérive conceptuelle... si les modèles d'IA générative peuvent apporter une aide intéressante dans le cadre d'un travail créatif, ils ne peuvent en aucun cas servir de caution de vérité.