

Les élucubrations mathématiques de ChatGPT

Le modèle de langage massif qui fait les succès de ChatGPT révèle de nombreuses failles, dès lors qu'on teste sérieusement ses aptitudes mathématiques.

À la fin de l'année 2022, la société OpenAI a mis en ligne ChatGPT, un agent conversationnel multilingue basé sur le modèle GPT (Generative Pre-trained Transformer), à même de répondre apparemment aux questions les plus diverses.

Dans celui-ci, l'apprentissage consiste à analyser des textes récupérés sur Internet, à les découper en *tokens* (des mots ou des morceaux de mots), puis à mesurer la probabilité d'occurrence de chaque token T à la suite de séries de tokens $\{T_i\}$. Ainsi, pour simplifier, lors de l'apparition de la suite $\{T_i\}$ il sera possible de proposer un token T suivant (le plus probable, le plus plausible...) et ce, bien sûr, indépendamment de la langue (on notera au passage que les notions de grammaire et de syntaxe ne jouent ici aucun rôle). Cette phase initiale est évidemment encadrée par des êtres humains qui procèdent à des ajustements, à des « réglages » (*fine tuning*). Une notion essentielle – et peut-être la clef de la réussite – est celle de *pre-prompt* : il s'agit d'une suite de tokens (certainement très complexe, mais confidentielle) insérée avant que l'utilisateur soumette ses propres phrases, et destinée à préciser le contexte, à orienter la génération des tokens suivants. Mais, chose essentielle à noter, ChatGPT n'a pas la notion de ce qui est vrai et de ce qui est faux. La seule chose qui importe, c'est la plausibilité évaluée en termes de probabilité : ses productions n'auront donc peut-être pas toujours valeur d'oracle, tant s'en faut, comme nous allons le vérifier.

LIRE L'ARTICLE