ChatGPT

6 questions fondamentales

Ce document fait partie d'un projet réalisé par *ethix* – laboratoire d'éthique de l'innovation, avec le soutien de la ***Fondation Mercator Suisse***.
Ces 6 questions fondamentales ont été identifiées suite à une série d'interviews et de workshops menés avec des plus de 30 personnes dont l'activité est touchée par l'outil ChatGPT. Plus d'informations vers les ressources et les exemples en bas de ce document.

Utilisations principales

- **PREMIER DRAFT** d'un texte sur la base de quelques informations
- **VARIATION STYLISTIQUE** sur un extrait de texte
- **SPARRING-PARTNER** pour identifier les points principaux sur une question
- **RÉSUMÉ D'UN TEXTE** sur la base d'informations brutes
- Dégrossissage pour une **PREMIÈRE RECHERCHE D'INFORMATIONS**

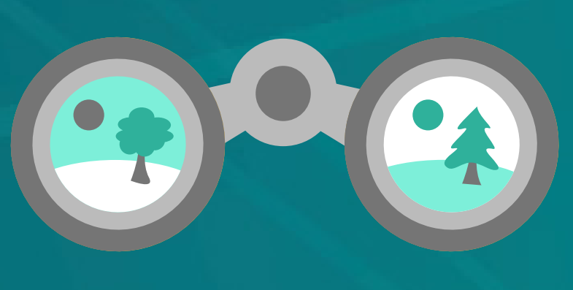
ChatGPT est-il trop humain ?

A travers l'écrit, ChatGPT se comporte comme un humain. Il compose un texte construit et argumenté et réagit aux interactions initiées par la personne qui l'utilise. L'échange a l'apparence d'une conversation humaine. L'interface utilisateur de ChatGPT renforce notre tendance à humaniser la technologie.

ChatGPT n'est pas une personne, même si ses concepteurs tentent de créer l'illusion d'une pensée. Le programme est un ensemble de formules mathématiques et de lignes de code.

Le design de l'interface utilisateur de ChatGPT n'est pas neutre. Par exemple, l'apparition graduelle de la réponse veut faire croire que l'outil « réfléchit ». Ce genre de choix de design reflète des intérêts économiques et des objectifs stratégiques des équipes de conception et l'entreprise propriétaire. Il faut chercher l'humain et ses intérêts derrière la technologie.

La capacité de langage naturel et le design de ChatGPT renforcent le risque de manipulation des utilisateurs.
La personnification de l'outil, par exemple à travers les interviews de ChatGPT parus dans les médias, participe au renforcement de l'anthropomorphisation de l'outil.



ChatGPT est-il neutre ?

ChatGPT a lu des milliards de textes trouvés en ligne et communique avec nous à la manière d'une machine quasi-omnisciente. Les milliards de mots analysés créent l'illusion d'une machine au-dessus de la mêlée, sans parti-pris aucun. En bref, une machine neutre. Mais comprendre les étapes de conception de ChatGPT vient contredire cette apparente neutralité.

Les équipes de conception de ChatGPT ont entraîné la machine sur un immense corpus de textes.
Ce mélange de textes trouvés en ligne et de livres numérisés ne reflète qu'une partie de la réalité. Le monde anglophone tel que reflété à travers des textes en ligne, avec leurs qualités et leurs défauts, représente la base de travail de ChatGPT.

Les performances brutes obtenues grâce à l'analyse du corpus de textes sont ensuite améliorées et corrigées par des humains. D'une part, certaines réponses sont imposées à la machine. D'autre part, on l'entraîne à donner des réponses acceptables et à éviter certains sujets. Ces mesures d'entraînement cherchent à éviter les résultats dangereux ou absurdes. Les résultats obtenus sont donc le résultat de ces interventions humaines.

ChatGPT n'est pas neutre - les textes produits reflètent les données d'entraînement et les choix humains réalisés durant l'affinage du modèle. Plus les outils de génération de contenu prennent de l'importance, plus la transparence sur leur représentation de la réalité et leurs présupposés devient une demande centrale.

ChatGPT dit-il la vérité ?

ChatGPT semble dire quelque chose de véridique car la forme est convaincante et le style affirmatif. Mais ChatGPT n'est pas conçu pour travailler avec la catégorie « vérité ». Il détermine les mots les plus plausibles dans un certain contexte en fonction de ses données d'entraînement.

Sur les questions de fait type : **« Quand la 1ère guerre mondiale a-t-elle eu lieu ? »**

La véracité réside dans la correspondance entre un énoncé et un fait vérifiable. Cependant, ChatGPT ne connaît pas la réalité et ne travaille pas avec le concept de vérité. ChatGPT cherche les mots les plus plausibles dans un contexte spécifique. Dans certains cas, le plausible correspond à un fait vérifiable. Mais souvent, le plausible n'est pas le vrai et ChatGPT crée alors des contenus erronés.

Sur les questions d'évaluation type : **« Quel est le meilleur système économique ? »**

Cette évaluation se joue au-delà du vrai ou du faux – il faut clarifier quelle est la référence utilisée pour évaluer différentes alternatives. ChatGPT a été programmé pour se montrer prudent sur certains sujets et pour refuser de répondre à certaines questions.
Il est important que les utilisateurs connaissent la façon dont le système fonctionne et les critères de choix d'ajustage de ses concepteurs.

ChatGPT amplifie ce qu'il trouve dans ses données – la plausibilité rencontre parfois la vérité, mais ce n'est pas systématique. L'outil devient alors parfois une machine à créer du faux vraisemblable. Cet enjeu est majeur pour éviter la dissémination de fake news sous formes de textes ou d'images.

ChatGPT est-il conscient ?

ChatGPT semble tellement humain dans sa capacité de communiquer avec nous qu'on pourrait se demander si la machine ne développe pas une vie intérieure. Certains appellent même à voir ChatGPT comme l'avant-garde de machines qui vont développer une conscience propre.

La conscience peut être définie comme la capacité de faire l'expérience d'une vie intérieure. C'est une expérience profondément personnelle. Savoir à quoi ressemble la vie intérieure des autres humains et des autres espèces vivantes est impossible. Nous ne pouvons pas nous projeter « dans la tête » des autres.

Cette impossibilité de partager l'expérience de la conscience renforce la difficulté de bien définir la conscience. Il paraît difficile d'imaginer que ChatGPT possède les caractéristiques généralement attribuées à l'état de conscience : sensibilité (sentir et répondre au monde qui l'entoure), état d'éveil, et expérience du soi.

Tenter de définir la conscience pose la question de son émergence. Celle-ci peut-elle être réduite à la matière ou représente-elle quelque chose « en plus » ? Peut-elle se développer dans une entité non-vivante ? Il est a priori possible d'imaginer que la conscience puisse émerger d'un substrat non-biologique. Si l'on considère que l'esprit se réduit au monde physique (tout est matière), une conscience pourrait émerger des combinaisons d'unités de matière qui composent un agent artificiel (non-biologique).

Même si ChatGPT ne fait pas l'expérience d'une vie intérieure, la possibilité d'une conscience artificielle ne peut pas être écartée. Dans tous les cas, il sera difficile voire impossible d'avoir accès à cette forme de conscience. Comme le propose une célèbre expérience de pensée, nous ne savons pas à quoi ressemble la vie intérieure d'une chauve-souris - ni celle d'une IA.

ChatGPT est-il intelligent ?

ChatGPT fait des prouesses mais produit aussi des contenus erronés. Il semble capable du bon comme du mauvais. Mais comment déterminer s'il est intelligent ? Répondre à cette question, c'est avant tout clarifier nos propres définitions de l'intelligence. Une tentative avec trois définitions.

1
L'intelligence est la capacité d'un programme à imiter l'intelligence humaine sans qu'un humain ne puisse distinguer l'humain de la machine.

Cette définition se reflète dans le célèbre test de Turing qui lie l'intelligence à la capacité d'imiter le comportement humain. Dans une conversation prolongée avec ChatGPT, un humain peut discerner des incohérences logiques et des erreurs de compréhension qu'un humain ne ferait pas.
Dans d'autres cas, il est difficile de faire la différence entre ChatGPT et un interlocuteur humain. Dans ce dernier scénario, on pourrait alors considérer qu'il passe le test et qu'il est intelligent.

2
L'intelligence est la capacité de fonctionner de manière appropriée en s'adaptant à son environnement.

ChatGPT adapte son comportement aux questions posées. Si on définit « son environnement » au sens restreint, ChatGPT peut être considéré comme intelligent.
Par contre, cette définition n'est pas satisfaite si l'environnement comprend le monde physique.

3
L'intelligence demande les propriétés suivantes prises ensemble : la capacité d'effectuer des opérations logiques (le raisonnement), de comprendre la sémantique et les concepts du langage humain ainsi que de procéder à des délibérations morales et de se déterminer sur celles-ci.

Selon cette définition, ChatGPT ne peut pas être considéré intelligent.
Il manque aux machines une compréhension sémantique du monde, directement en lien avec le réel. Les machines « ont l'air » de comprendre, elles donnent cette illusion. De même, elles n'ont pas de capacité de délibération morale.

Plus la notion d'intelligence est facile à satisfaire, plus ChatGPT a des chances d'être considéré comme « intelligent ». A l'inverse, ChatGPT ne remplit pas les conditions d'une définition exigeante.

ChatGPT est-il créatif ?

ChatGPT fait partie des systèmes algorithmiques générateurs de contenu. Il crée du contenu qui n'existe nulle part ailleurs.
Mais ce contenu est-il original ? ChatGPT peut-il être comparé à un humain dans sa capacité de créer ?

ChatGPT ne crée pas de sa propre initiative, il fonctionne comme un outil au service des humains. En réaction à une demande, ChatGPT génère du contenu original qui n'existait pas sous cette forme avant la demande. Il va plus loin que la simple copie de l'existant, il combine une multitude d'éléments de manière novatrice.

Si ChatGPT n'est qu'un outil, il faut chercher les humains impliqués dans le processus de création pour déterminer leur contribution au résultat final. Qui participe à ce geste créatif ? Les textes utilisés comme données d'entraînement, les équipes de conception de l'outil, les utilisateurs qui formulent une demande spécifique ? Chacun peut se prévaloir d'une partie du contenu créé.

Les outils générateurs de contenu nous obligent à redéfinir ce que veut dire la création, une compétence longtemps réservée aux seuls humains. Mais savons-nous précisément comment un humain crée du contenu ? ChatGPT ne vient-il pas rappeler que toute création humaine s'inspire de l'existant et le reconfigure d'une nouvelle manière ?

Ressources

Pour trouver des ressources supplémentaires, rendez-vous sur notre site en scannant ce QR code :

Méthode

Ce projet est mené par ***ethix*** – laboratoire d'éthique de l'innovation, avec le soutien de la ***Fondation Mercator Suisse***. Entre janvier et mai 2023, ethix a réalisé plus de 30 interviews et workshops avec des personnes exerçant des activités touchées par des outils comme ChatGPT. L'hypothèse de travail choisie concernait surtout l'enseignement, la communication et le design, les médias, les métiers du droit et le fonctionnement des administrations publiques. L'équipe a ensuite cherché à établir des liens entre le contenu des interviews et la vaste littérature philosophique sur les outils d'IA.

Contact pour le projet :
Dr Johan Rochel (*rochel@ethix.ch*)

ethix

Stiftung
Mercator
Schweiz

Design : *Polar Graphic*