L’histoire cachée du potentiel étonnant de ChatGPT | Greg Brockman | TED

Nous avons créé OpenAI il y a sept ans car nous avons remarqué qu’il se passait quelque chose de vraiment intéressant dans le domaine de l’IA et nous voulions contribuer à le diriger dans une direction positive. C’est vraiment incroyable de voir à quel point ce domaine a évolué depuis. C’est très gratifiant d’entendre des personnes comme Raymond qui utilisent la technologie que nous développons pour faire tant de choses merveilleuses. Nous entendons des personnes excitées, des personnes inquiètes, des personnes qui ressentent les deux émotions à la fois. Et honnêtement, c’est ainsi que nous nous sentons aussi. Avant tout, nous avons l’impression d’entrer dans une période historique où nous, en tant que monde, allons définir une technologie qui sera si importante pour notre société à l’avenir. Et je crois que nous pouvons la gérer de manière positive. Aujourd’hui, je veux vous montrer l’état actuel de cette technologie et certains des principes de conception sous-jacents que nous chérissons. La première chose que je vais vous montrer, c’est à quoi cela ressemble de construire un outil pour une IA plutôt que de le construire pour un humain. Nous avons un nouveau modèle DALL-E, qui génère des images, et nous le mettons à disposition en tant qu’application pour ChatGPT. Vous pouvez donc faire des choses comme demander des suggestions pour un bon repas après une conférence TED et dessiner une image de celui-ci. Vous obtenez toute l’idéation et la créativité, ainsi que les détails pris en charge pour vous par ChatGPT. Et voilà, ce n’est pas seulement l’idée du repas, mais une étalage très détaillé. Voyons ce que nous allons obtenir. Mais ChatGPT ne génère pas

seulement du texte dans ce cas – désolé, il ne génère pas seulement du texte, il génère aussi une image. Et cela étend vraiment la puissance de ce qu'il peut faire en votre nom pour réaliser votre intention. Et je tiens à souligner que tout cela est une démonstration en direct. Tout cela est généré par l'IA pendant que nous parlons. Je ne sais même pas ce que nous allons voir. Ça a l'air merveilleux. J'ai faim rien qu'en le regardant. Nous avons également enrichi ChatGPT avec d'autres outils, comme la mémoire. Vous pouvez dire "sauvegarde ça pour plus tard." Et ce qui est intéressant avec ces outils, c'est qu'ils sont très inspectables. Vous obtenez cette petite fenêtre contextuelle qui propose d'utiliser l'application DALL-E. Et au fait, tous les utilisateurs de ChatGPT auront accès à cela dans les mois à venir. Et vous pouvez regarder comment la machine utilise réellement ces outils, ce qui nous permet de leur donner des retours. Maintenant c'est sauvegardé pour plus tard, et laissez-moi vous montrer comment utiliser cette information et l'intégrer à d'autres applications aussi. Vous pouvez dire "Maintenant faites une liste de courses pour cette délicieuse chose que j'ai suggérée plus tôt." Et rendez-le un peu difficile pour l'IA. "Et publiez-le sur Twitter pour tous les spectateurs du TED." Donc si vous réalisez ce merveilleux repas, je veux vraiment savoir comment ça goûte. Mais vous pouvez voir que ChatGPT sélectionne tous ces différents outils sans que je lui demande explicitement lesquels utiliser dans chaque situation. Et cela, je pense, montre une nouvelle façon de penser à l'interface utilisateur. Nous avons tellement l'habitude de penser, eh bien, nous avons ces applications, nous cliquons entre elles, nous copions/collez entre elles, et généralement c'est une bonne expérience au sein d'une application, à condition de connaître les menus et de connaître toutes les options. Oui, j'aimerais bien. Oui, s'il vous plaît. Toujours bien d'être poli. Et en ayant cette interface langue unifiée par-dessus les outils, l'IA est capable de prendre en charge tous ces détails à votre place. Vous n'avez pas à être celui qui précise chaque petit détail de ce qui est censé se passer. Et comme je l'ai dit, il

s’agit d’une démonstration en direct, donc parfois l’imprévu nous arrive. Mais regardons la liste de courses sur Instacart pendant que nous y sommes. Et vous pouvez voir que nous avons envoyé une liste d’ingrédients à Instacart. Voici tout ce dont vous avez besoin. Et ce qui est vraiment intéressant, c’est que l’interface utilisateur traditionnelle est toujours très précieuse, n’est-ce pas ? Si vous regardez cela, vous pouvez toujours cliquer dessus et modifier les quantités réelles. Et cela montre que les interfaces utilisateur traditionnelles ne disparaissent pas, elles sont toujours là. C’est juste que nous avons une nouvelle façon augmentée de les construire. Et maintenant, nous avons un tweet qui a été rédigé pour notre revue, ce qui est également très important. Nous pouvons cliquer sur “exécuter”, et voilà, nous sommes le responsable, nous pouvons inspecter, nous pouvons modifier le travail de l’IA si nous le souhaitons. Et donc après cette conférence, vous pourrez y accéder vous-même. Et voilà. Cool. Merci à tous. Nous allons revenir aux diapositives. Maintenant, l’aspect important de la façon dont nous construisons cela, ce n’est pas seulement construire ces outils. Il s’agit d’apprendre à l’IA comment les utiliser. Qu’est-ce que nous voulons même qu’elle fasse lorsque nous lui posons ces questions très générales ? Et pour cela, nous utilisons une vieille idée. Si vous remontez au document de 1950 d’Alan Turing sur le test de Turing, il dit que vous ne programmerez jamais une réponse à cela. Au lieu de cela, vous pouvez l’apprendre. Vous pouvez construire une machine, comme un enfant humain, et lui apprendre par le biais de la rétroaction. Avoir un enseignant humain qui offre des récompenses et des punitions à mesure qu’il essaie des choses et fait des choses qui sont bonnes ou mauvaises. Et c’est exactement ainsi que nous entraînons ChatGPT. C’est un processus en deux étapes. D’abord, nous produisons ce que Turing aurait appelé une machine enfant grâce à un processus d’apprentissage non supervisé. Nous lui montrons simplement le monde entier, l’internet entier, et nous lui disons, “Prédis ce qui va suivre dans du texte que tu n’as jamais vu

auparavant.” Et ce processus lui confère toutes sortes de compétences merveilleuses. Par exemple, si vous lui montrez un problème mathématique, la seule façon de le compléter, de dire ce qui suit, ce neuf vert là-haut, c’est de résoudre réellement le problème mathématique. Mais nous devons également effectuer une deuxième étape, qui consiste à apprendre à l’IA quoi faire avec ces compétences. Et pour cela, nous lui fournissons une rétroaction. Nous faisons essayer à l’IA plusieurs choses, nous proposons plusieurs suggestions, et ensuite un humain les évalue, dit “Celle-ci est meilleure que celle-là.” Et cela renforce non seulement la chose spécifique que l’IA a dite, mais aussi très important, tout le processus que l’IA a utilisé pour produire cette réponse. Et cela lui permet de généraliser. Cela lui permet d’apprendre, d’inférer votre intention et de l’appliquer dans des scénarios qu’elle n’a pas vus auparavant, pour lesquels elle n’a pas reçu de rétroaction. Parfois, les choses que nous devons apprendre à l’IA ne sont pas celles auxquelles vous vous attendez. Par exemple, lorsque nous avons montré pour la première fois GPT-4 à Khan Academy, ils ont dit, “Wow, c’est tellement génial, nous allons pouvoir enseigner de merveilleuses choses aux étudiants. Un seul problème, il ne vérifie pas les calculs des étudiants. S’il y a une mauvaise formulation mathématique, il continuera joyeusement en prétendant que un plus un égale trois.” Nous avons donc dû collecter des données de rétroaction. Sal Khan lui-même a été très gentil et a offert 20 heures de son temps pour fournir des commentaires à la machine aux côtés de notre équipe. Et au cours de quelques mois, nous avons pu enseigner à l’IA que, “Eh bien, tu devrais vraiment réagir aux humains dans ce type de scénario spécifique.” Et nous avons effectivement apporté beaucoup d’améliorations aux modèles de cette manière. Et lorsque vous appuyez sur le pouce vers le bas dans ChatGPT, c’est un peu comme envoyer un signal de détresse à notre équipe pour dire “Voici un problème, nous avons besoin d’aide.” Et nous utilisons toutes ces informations pour continuellement améliorer et affiner le système. Alors voilà,

c’est comme ça que nous construisons ça. C’est comment nous utilisons cette rétroaction pour enseigner à l’IA comment utiliser ces outils et répondre à nos questions de manière précise et pertinente. Et je crois fermement que nous pouvons utiliser ces technologies avec sagesse et les guider vers une direction positive pour notre société.

Source : TED | Date : 2023-04-20 17:11:28 | Durée : 00:30:10

➡ Accéder à CHAT GPT en cliquant dessus