

Le Robot de Discussion français : un exploit pour son propriétaire

Moshi : le nouveau système vocal innovant de Kyutai

Le 3 juillet dernier, la compagnie Kyutai a dévoilé Moshi, un type d'intelligence artificielle avec des compétences vocales jamais vues auparavant. Ce chatbot de nouvelle génération a attiré l'attention des médias lors de sa présentation à la presse et est maintenant accessible à tous sur le site officiel de [Moshi.chat](https://moshi.chat). Ce prototype se démarque par sa capacité à communiquer de façon naturelle et expressive à l'oral, tout en étant attentif à son interlocuteur.

Le système multimodal de Moshi peut aussi être utilisé en local, sans connexion Internet. Kyutai a décidé de rendre ce système open source pour contribuer au progrès et à la recherche dans le domaine de l'intelligence artificielle. Pour en savoir plus sur Moshi et les futurs projets de la fondation, nous avons interviewé leur PDG, Patrick Perez.

Moshi se démarque grâce à ses performances qui surpassent celles du GPT-4 dans des domaines-clés, offrant ainsi une interaction humaine inégalée. Moshi peut-il être considéré comme le chatbot le plus avancé actuellement ?

Contrairement aux idées reçues, Moshi.chat n'est pas une aide vocale, mais un prototype expérimental basé sur Moshi, un système multimodal generaliste texte-voix développé entièrement en interne. Ce prototype illustre une

interactivité orale sans équivalent en termes de naturel et de fluidité.

Quels sont les possibles usages de Moshi et dans quels secteurs voyez-vous cette technologie évoluer dans le futur ?

Les potentiels d'utilisation de ce système multimodal sont variés :

- Discussion orale avec une IA pour l'inclusion, l'éducation, le coaching, le divertissement, etc.
- Élocution expressive pour améliorer l'accessibilité aux écrits, la création artistique, les jeux, etc.
- Traduction audio en temps réel pour des échanges mondiaux.

Moshi a été élaboré en seulement six mois par une équipe de huit personnes, ce qui semble rapide. Est-ce plus simple de concevoir un assistant tel que Moshi de nos jours ?

La création de Moshi a représenté un exploit nécessitant des compétences spécialisées et une collaboration intensive. Nous avons également loué des machines puissantes chez Scaleway pour soutenir nos activités. La conception d'un assistant aussi sophistiqué requiert un savoir-faire spécifique et des ressources significatives.

Moshi peut identifier divers accents, mais communique uniquement en anglais pour le moment. Envisagez-vous d'introduire d'autres langues à l'avenir ? Et comment se déroule l'apprentissage de nouvelles langues pour un système de langue et de parole ?

Nous projetons d'intégrer d'autres langues, notamment le français et l'espagnol, à Moshi. Nous avons l'intention de partager nos modèles et nos codes pour faciliter la formation dans d'autres langues. Malgré sa complexité, l'incorporation de nouvelles langues peut être réalisée par d'autres acteurs disposant des ressources adéquates.

Moshi s'appuie sur Helium, un modèle de 7 milliards de critères. Comment cela se compare-t-il à d'autres modèles comme GPT-4 ? La qualité d'un système de langue et de parole se résume-t-elle à sa taille ? Quels sont les jeux de données utilisés pour l'entraînement de Moshi ?

Considéré de taille moyenne, Helium se distingue des autres modèles. La taille d'un modèle n'est pas le seul indicateur de qualité ; la quantité et la qualité des données d'entraînement sont également essentielles. Pour la préparation d'Helium, nous avons exploité des données provenant d'Internet, notamment du projet CommonCrawl, totalisant environ un billion de données.

Des mots.

Moshi a subi un entraînement avec 100 000 échanges verbaux et un moteur de Text-to-Speech. Pouvez-vous décrire le processus d'acquisition de compétences de Moshi ?

Hélium a produit ces 100 000 dialogues, ensuite transformés en discussions audio grâce à une ancienne version de Moshi équipée de synthèse vocale.

Y a-t-il quelqu'un derrière la voix de Moshi ?

Nous avons collaboré avec une chanteuse pour enregistrer les échanges qui ont été utilisés pour élaborer la voix de Moshi.

Kyutai a mis au point une variante locale de Moshi pour un fonctionnement sans connexion internet. Pourquoi avez-vous opté pour le GPU plutôt que le NPU pour ce modèle ? Envisagez-vous une version "mini" pour smartphones à l'horizon ?

Nos schémas ont été conçus pour être compatibles avec les processeurs graphiques Nvidia, pour lesquels nous avons élaboré une version locale compressée de Moshi. Nous étudions la possibilité d'adapter Moshi pour une utilisation sur smartphones dans le futur, en fonction des exigences et des

opportunités du marché.

La sûreté est une inquiétude majeure concernant les IA génératives. Comment Kyutai entend assurer que Moshi ne sera pas exploité à des fins malveillantes ?

Nous œuvrons activement à renforcer la sécurité de Moshi à travers le finetuning et des techniques de marquage discret des contenus engendrés. La transparence du code source joue également un rôle primordial dans la sécurité, permettant à des tiers d'inspecter et de renforcer les modèles partagés.

Quel est l'horizon de Moshi selon Kyutai ?

Nous continuons de perfectionner Moshi pour améliorer ses aptitudes, notamment en matière de pertinence, d'expressivité et de soutien linguistique. Nous sommes impatients de voir comment l'écosystème utilisera et enrichira Moshi une fois que nous aurons diffusé nos modèles et nos outils techniques.

Kyutai bénéficie d'un financement exclusif en France et en Europe grâce à Xavier Niel. Cette autonomie constitue-t-elle un atout pour les recherches menées par le laboratoire ?

Soutenu par Xavier Niel, Eric Schmidt et Rodolphe Saadé, Kyutai jouit d'une grande liberté dans le choix de ses recherches, loin des contraintes liées au profit. En qualité de fondation, notre objectif est de contribuer aux progrès de l'intelligence artificielle tout en restant fidèles à nos valeurs et à nos engagements envers la communauté.

□ *Pour rester informé des dernières actualités sur la technologie et l'intelligence artificielle, rejoignez-nous sur [Google Actualités](#) et [WhatsApp](#).*

Source : www.01net.com

→ ☐ Accéder à [CHAT GPT](#) en cliquant
dessus