Google dégaine la vision par l'intelligence artificielle en temps réel

Vision en temps réel

00:00

- Une nouvelle étape majeure a été franchie dans l'utilisation de l'intelligence artificielle, notamment avec la sortie d'un modèle multimodal temps réel chez Google appelé Gemini Flash
- 00:11.
- Peu de temps après l'annonce de Gemini Flash, OpenAI a également annoncé la sortie imminente d'une fonctionnalité multimodale temps réel similaire
- 00:24.
- Cette technologie permettrait à une intelligence artificielle de voir son environnement en temps réel et d'avoir des discussions en temps réel avec les utilisateurs
- 00:42.
- Les utilisateurs pourraient discuter avec l'IA en mode multimodal, utilisant la parole, le texte et la vision, ce qui serait particulièrement utile pour les personnes ayant des handicaps
- 01:01.
- L'IA pourrait comprendre le langage des signes et répondre en conséquence, sans nécessiter de parole
- 01:15.
- Cette technologie multimodale temps réel permettrait à l'IA de voir et de comprendre ce que les utilisateurs font, et de répondre en temps réel
- 01:33.
- Les utilisateurs pourraient demander à l'IA de faire des tâches en temps réel, comme demander ce qu'il y a dans une image ou demander des informations sur un objet
- 01:44.
- Cette technologie pourrait être utilisée dans divers contextes, tels que la maison ou le bureau, pour demander à l'IA de faire des tâches multimodales
- 01:54.

L'annonce est folle

01:59

- Google a fait une annonce récente concernant un nouveau modèle multimodal en temps réel, appelé Gemini 2.0, qui n'est pas encore disponible partout mais qui sera bientôt répandu sur leur plateforme et les applications mobiles
- 01:59.
- Cette annonce a été faite rapidement par Google pour être les premiers à présenter cette technologie d'intelligence artificielle
- 02:11.
- Open AI a également fait une annonce similaire la veille au soir pour présenter leur propre modèle multimodal en temps réel, afin de ne pas être dépassé par Google

- 02:38.
- Actuellement, l'option multimodale en temps réel n'est pas disponible sur ChatGPT, ni sur l'application PC ni sur l'application mobile, mais elle devrait arriver dans les prochains jours ou semaines
- 03:02.
- Une mise à jour des applications ChatGPT et Gemini est attendue pour permettre l'accès à ces nouvelles fonctionnalités
- 03:08.
- La concurrence entre Google et Open AI est rude, car le premier à proposer ces fonctionnalités sur les appareils mobiles gagnera des parts de marché importantes
- 03:18.
- Les fonctionnalités multimodales en temps réel sont considérées comme révolutionnaires et pourront être testées prochainement
- 03:27.

Accès dès maintenant

03:33

- Pour accéder aux fonctionnalités de Gemini, il est possible de passer par le site classique https://gemini.google.com/, mais cela ne permet pas d'accéder directement aux modèles expérimentaux en temps réel
- 03:37.

- Pour contourner ce problème, il est possible de passer par le site AI Studio de Googl, qui permet d'accéder aux fonctionnalités de Gemini et de réaliser des expérimentations multimodales
- 03:55.
- Le site AI Studio de Google permet de découvrir les possibilités de Gemini et de comprendre comment fonctionnent les modèles expérimentaux
- 04:02.

Bientôt sur ChatGPT

04:09

- Dans quelques jours ou semaines, ChatGPT proposera de nouvelles fonctionnalités, notamment la possibilité de discuter en temps réel avec une voix et une image, et non plus seulement avec la voix du Père Noël
- 04:10.
- Sur le site web et les applications mobile et PC de ChatGPT, il sera possible de cliquer sur le mode vocal avancé pour discuter en temps réel et avoir accès à une fonctionnalité de vision en temps réel grâce à une petite caméra
- 04:29.
- Cette fonctionnalité de vision en temps réel permettra à l'utilisateur de recevoir des conseils et des indications en temps réel, par exemple pour faire la cuisine ou réaliser d'autres tâches
- 05:01.
- OpenAI a déjà démontré cette fonctionnalité dans une démo, mais elle n'est pas encore disponible pour les utilisateurs de ChatGPT
- 04:51.
- La mise à jour de ChatGPT avec ces nouvelles fonctionnalités devrait arriver rapidement, et il est possible que Google ou OpenAI soit le premier à déployer ces fonctionnalités

- 05:40.
- Pour tester les capacités de ces nouvelles fonctionnalités, il faudra aller sur AI Studio, car OpenAI n'est pas encore disponible pour l'instant
- 06:11.
- La seule façon de tester ces fonctionnalités est d'aller sur AI Studio, et cela permettra de voir qui, de Google ou d'OpenAI, sera le premier à déployer ces fonctionnalités
- 06:18.

On y va !

06:19

- Les capacités des modèles multimodaux peuvent être testées en temps réel sur le site aistudio.google.com, qui est principalement destiné aux développeurs et aux programmeurs
- 06:31.
- Le site propose divers réglages et options pour tester les modèles multimodaux, notamment le mode multimodal live de Gini 2.0 Flash
- 06:54.
- Le mode multimodal live permet de tester le modèle en temps réel, avec la possibilité de choisir le modèle à tester, bien qu'il n'y ait qu'un seul modèle disponible pour le moment, à savoir Gini 2.0 Flash, qui est actuellement expérimental
- 07:07.
- Le modèle Gini 2.0 Flash permet de communiquer de différentes manières, notamment en mode vocal classique, en mode vidéo ou en partageant l'écran
- 07:29.
- Le mode vidéo permet de partager ce qui se passe sur l'écran, ce qui peut être utile pour certaines applications
- 07:41.
- Avant de commencer les tests, il est important de noter que Gini 2.0 Flash a des limitations en termes d'expression vocale par rapport à un chat
- 07:53.
- Les tests vont commencer avec le mode caméra, mais il est important de noter que Gini 2.0 Flash n'est pas parfait en termes d'expression vocale
- 07:57.

Disclaimer

07:59

- Le mode vocal de Gemini est moins performant que le mode vocal avancé de ChatGPT, notamment en français, car il parle parfois en anglais sans raison apparente et possède des accents bizarres
- 08:03.
- Le mode vocal de Gemini peut également attribuer des prénoms incorrects, avoir des difficultés à mettre en place des CCDI (Centre de Contact Direct International) et lire des chiffres uniquement en anglais
- 08:12.
- Les limites du mode vocal de Gemini en français sont nombreuses, mais elles ne constituent pas l'objet principal de l'étude
- 08:21.
- L'accent sera mis sur les fonctionnalités internes de Gemini, telles que la possibilité de partager la caméra ou l'écran, plutôt que sur son aspect vocal

- 08:35.
- L'objectif est de tester les performances et les limites de Gemini, en particulier en ce qui concerne ses fonctionnalités multimodales
- 08:43.

IL VOIT TOUT !

10:28

- Un modèle multimodal peut comprendre et répondre en français.
- 10:33
- Le modèle peut également comprendre le contexte visuel et décrire l'environnement dans lequel il se trouve, y compris les objets et les couleurs.
- 10:40
- Le modèle peut identifier des références à des films, comme la présence d'une figurine R2-D2 de la saga Star Wars.
- 10:50
- Le modèle peut également identifier d'autres références à des films, comme un poster avec le logo Tokyo qui pourrait être lié au film Lost in Translation.
- 11:03
- Le modèle peut ne pas connaître certaines références culturelles, comme Albator, un pirate de l'espace.
- 11:13
- Le modèle peut identifier des éléments festifs dans l'environnement, comme des guirlandes lumineuses dorées, des boules à facettes et des ornements de Noël.
- 11:37
- Le modèle peut également identifier des éléments personnels, comme une tasse avec des motifs de pingouin, et déduire que la combinaison des couleurs rouges et blanches fait penser à Noël.
- 12:02
- Le modèle peut comprendre les réactions faciales et les expressions, comme une réaction de dégoût en buvant une boisson.
- 12:12
- Le modèle peut identifier des éléments informatiques dans l'environnement, comme des écrans d'ordinateur et des boîtiers, et déduire que la personne apprécie l'informatique et a un intérêt pour l'évolution de la technologie.
- 12:36
- Le modèle peut également identifier des éléments rétro, comme un ordinateur ancien, et les contrastes avec les éléments modernes.
- 12:54
- La personne montre un tableau blanc avec des notes, mais les lettres sont trop petites pour être lues
- 13:16.
- Un couteau suisse avec une poignée rouge est présenté, et la personne demande de l'aide pour identifier les différents outils qu'il contient
- 13:30.
- L'un des outils est identifié comme une lame dentelé, un petit couteau qui peut être utilisé pour couper des aliments ou d'autres matériaux
- 13:51.
- Un autre outil est identifié comme une scie
- 13:57.
- Un Minitel, un terminal informatique de l'époque, est présent dans la pièce et est identifié comme étant sur une étagère avec d'autres appareils électroniques

- 14:24.
- La personne est impressionnée par les capacités de reconnaissance et d'identification des objets
- 14:30.

On analyse

14:39

- Le test effectué est bluffant et montre des applications potentielles, notamment en termes de synthèse vocale en temps réel, même si la qualité n'est pas encore parfaite
- 14:39.
- Le système est capable de comprendre la logique d'une conversation et de se souvenir de ce qui s'est passé, comme dans une discussion avec un humain
- 15:08.
- Le système peut être utilisé sur une plateforme en ligne et est actuellement gratuit, sans besoin d'abonnement
- 15:42.
- L'outil pourrait bientôt arriver sur les applications mobiles, ce qui serait fantastique pour de nombreux métiers et usages personnels, ainsi que pour l'inclusivité des personnes handicapées
- 16:00.
- Le système pourrait être utilisé pour aider les personnes malvoyantes ou non-voyantes, en leur permettant de demander à leur téléphone de leur décrire ce qui se passe autour d'eux
- 16:23.
- La même technologie est également disponible sur Chat GPT, qui peut être utilisé sur le site web, l'application mobile et l'application PC
- 16:50.
- Il est possible de brancher une caméra et de partager l'écran avec le système, ce qui ouvre de nouvelles possibilités d'utilisation
- 17:04.
- Le partage d'écran peut être utilisé pour montrer des contenus, comme des vidéos ou des sites web, mais cela peut également causer des problèmes de mise en abîme
- 17:30.

IL VOIT L'ÉCRAN !

18:08

- Un partage d'écran est effectué pour tester les capacités d'un modèle multimodal, permettant de partager l'intégralité de l'écran
- 18:10.
- Le modèle est capable d'identifier le contenu de l'écran partagé, notamment la page Web de la chaîne YouTube Renault Decode
- 18:28.
- Le modèle peut également fournir des informations sur la chaîne, comme la présence de lives réguliers et de vidéos récentes
- 18:38.
- Le modèle est capable de reconnaître le contenu de l'écran même après un défilement, confirmant la présence de vidéos récentes
- 19:19.

- Le modèle peut fournir des informations sur les vidéos présentes sur la chaîne, mais peut commettre des erreurs, comme dans le cas d'une vidéo recherchée le 10 décembre qui n'a pas été trouvée
- 19:23.
- Le modèle peut identifier les autres applications ouvertes sur l'écran, comme Google Sheets et un logiciel de streaming
- 20:32.
- Le modèle peut également identifier les applications ouvertes sur le bureau, mais peut commettre des erreurs, comme dans le cas d'un logiciel de streaming qui a été correctement identifié
- 20:59.
- Le modèle peut fournir des informations sur les logiciels présents sur l'ordinateur, comme Microsoft Paint, qui a été utilisé pour une retouche d'image
- 21:35.
- Le modèle peut fournir des conseils pour effectuer des tâches spécifiques, comme extraire une partie d'une image, mais peut rencontrer des difficultés dans certains cas
- 22:20.

On analyse

22:55

- Le modèle Gini 2.0 flash a été présenté par Google et a été mis en démo sur le site ai.studio.google.com, ainsi que par Open AI, qui sont des concurrents dans ce domaine
- 23:09.
- Cette fonctionnalité permet de voir et d'analyser le décor réel, comme reconnaître un R2D2 et comprendre qu'il s'agit d'une référence de film
- 23:25.
- Les applications potentielles de cette fonctionnalité sont nombreuses, notamment dans le domaine de l'assistance personnelle, où elle pourrait être intégrée dans des portables ou des PC
- 23:46.
- La fonctionnalité de partage d'écran permettrait aux utilisateurs d'avoir des compagnons virtuels qui pourraient les aider dans leurs tâches quotidiennes
- 23:51.
- Pour récupérer un logo à partir d'une capture d'écran, il est possible d'utiliser l'outil de sélection de forme de Paint pour sélectionner la zone autour du logo, puis de faire un copier-coller de cette sélection
- 24:42.
- L'outil de sélection de forme est disponible dans la section "Image" du ruban de Paint, juste en dessous de "Sélectionner", et est représenté par un rectangle avec une flèche en bas à droite
- 25:17.
- Une fois que le logo a été sélectionné avec l'outil de sélection de forme, il est possible de le glisser et de le déposer à l'endroit souhaité, ou de le copier et de le coller pour récupérer le logo uniquement
- 25:31.

On bosse avec lui

26:00

- Pour effacer une sélection, il est possible d'appuyer sur la touche Supprimer du clavier ou d'utiliser l'outil effacé de la barre d'outil de Paint.
- 26:07
- Pour remplir une sélection d'une couleur spécifique, il faut choisir la couleur souhaitée dans la palette de couleur, puis utiliser l'outil remplissage de couleur pour cliquer à l'intérieur de la zone de la sélection.
- 26:17
- L'utilisateur a rempli une sélection en rose, puis a remplacé cette couleur par du jaune, plus précisément un jaune vif fluo.
- 26:42
- L'outil utilisé est considéré comme très amusant et addictif, offrant de nombreuses applications possibles.
- 26:56

Un nouveau step de l'IA

27:03

- Google et OpenAI sont en train de déployer de nouvelles fonctionnalités pour leurs modèles de chat, notamment ChatGPT, qui sera bientôt disponible sous forme d'application mobile et PC
- 27:06.
- L'application ChatGPT permettra aux utilisateurs d'avoir un compagnon de tous les jours pour découvrir de nouveaux outils et régler les problèmes d'accessibilité, notamment grâce à sa capacité à décrire ce qui se passe sur l'écran
- 27:52.
- Les prochaines étapes de développement incluront probablement des capacités agentiques, permettant à l'application de voir non seulement ce qui se passe sur l'écran, mais également d'interagir avec l'utilisateur et de réaliser des tâches pour lui
- 28:16.
- L'application sera capable de voir et d'analyser tout ce qui se passe sur l'écran, y compris les icônes et les éléments de l'interface utilisateur, et pourra même réaliser des tâches pour l'utilisateur, comme cliquer sur des boutons ou sélectionner des éléments
- 29:41.
- L'application sera également capable de comprendre et de répondre à des commandes vocales, et pourra même traduire des textes en temps réel
- 29:57.
- Les utilisateurs pourront tester les nouvelles fonctionnalités de l'application sur le site aistudio.google.com, en sélectionnant le modèle Gemini 2.2.0 Flash et en utilisant les outils de partage d'écran pour interagir avec l'application
- 30:09.

Il voit aussi l'écrit

30:43

- Gemini est capable de comprendre et de répondre à des instructions écrites en français, comme le montre l'exemple où il lit et répond à l'écriture sur un papier
- 30:46.
- Gemini peut également reconnaître et décrire des dessins, comme le montre l'exemple où il décrit une maison dessinée sur un papier
- 31:53.
- Gemini peut reconnaître des éléments spécifiques dans un dessin, tels que des fenêtres, une porte et un toit
- 32:05.
- Gemini peut également reconnaître des changements apportés à un dessin, comme l'ajout de couleurs ou de nouveaux éléments
- 32:38.
- Gemini peut lire et décrire des textes écrits, y compris des couleurs et des éléments spécifiques, comme le montre l'exemple où il lit et décrit une maison avec un toit rouge, des fenêtres bleues et une porte verte
- 33:04.
- Gemini peut exécuter des ordres écrits, comme le montre l'exemple où il répond à l'instruction de dire le mot "papier"
- 33:25.
- Gemini peut également jouer à des jeux simples, comme le morpion, en utilisant des instructions écrites
- 33:38.

Déploiement en cours

34:06

- Le déploiement de Gemini et de la fonctionnalité de vision temps réel est en cours, avec des annonces récentes de Google et Open
- 34:09.
- Cette fonctionnalité permettra aux modèles de voir et de comprendre leur environnement, ce qui sera un "game changer" absolu
- 34:22.
- Les possibilités d'utilisation de cette technologie sont vastes, notamment dans les robots, où elle pourra être utilisée pour la reconnaissance faciale et la détection des émotions
- 34:32.
- Les développeurs travaillent actuellement pour déployer cette fonctionnalité sur les applications mobiles, les PC et les robots
- 34:18.
- Les utilisateurs pourront bientôt tester cette fonctionnalité ensemble, notamment sur le site Rena décodes
- 34:53.
- Il est possible que Google ait précipité l'annonce de cette fonctionnalité en raison de la concurrence avec Open, qui avait prévu de la sortir pendant ses 12 jours de lancement
- 35:10.
- Le déploiement de cette fonctionnalité ne sera pas sans conséquence pour les serveurs de Google et Open, qui devront gérer une charge supplémentaire
- 35:42.

- Les utilisateurs peuvent suivre les dernières actualités sur cette fonctionnalité sur le site renault-dcode.fr
- 35:52.

Concernant la chaîne

35:53

- Un message est adressé aux abonnés pour les informer des contenus à venir sur la chaîne, notamment les actualités, les tutoriels et les formations disponibles sur le web, le digital et l'IA.
- 35:57
- Les abonnés sont invités à se rendre sur Renaud Dékode pour accéder à ces contenus et pour participer à des sessions en direct.
- 36:06
- Une proposition est faite aux abonnés de choisir le jour et l'heure pour un tutoriel en direct sur l'audit express d'un site web, avec des options pour lundi, mercredi ou vendredi midi.
- 36:11
- Un aperçu est donné d'un environnement de test où de nouvelles fonctionnalités sont déjà disponibles, et les abonnés sont invités à explorer ces possibilités.
- 36:20
- Les possibilités offertes par ces nouvelles fonctionnalités sont qualifiées de "dinguerie" et considérées comme étant très prometteuses.
- 36:30