ما الجديد الذي جاء به إصدار GPT4o من نموذج الذكاء الاصطناعي؟

كتبه مصطفى أحمد | 16 مايو ,2024

قررت شركة OpenAI التي تهيمن حاليًّا على سوق الذكاء الاصطناعي، توسيع حصتها عندما كشفت النقاب عن نموذجها الرائد المتطور الجديد GPT4o الذي يمكنه أن يتحدث ويرى ويسمع بشكل يشبه الإنسان إلى حد كبير، ما يجعلها متقدمة على الشركات الأخرى بخطوات في سباق AI. ودعونا خلال السطور التالية، نأخذكم في رحلة سريعة لنتعرف على ما الجديد الذي جاء به إصدار GPT4o من نموذج الذكاء الاصطناعي.

GPT4o.. عصر النماذج "المتعددة" الحقيقية!

يمكن القول إن OpenAI حاليًّا تقود الثورة في مجال الذكاء الاصطناعي، وبينما تحاول الشركات الأخرى العملاقة مثل جوجل ومايكروسوفت وحتى آبل اللحاق بالركب، استطاعت مطورة شات جي بي تي أن تتفوق عليهم وتسبقهم بخطوات، وما زاد الطين بلة أنها أثبت هيمنتها المطلقة في

مجال AI عندما أعلنت منذ بضعة أيام عن GPT4o النسخة الخارقة من روبوت الدردشة الشهير ChatGPT الذي لا يتفوق فحسب على النماذج الحالية للشركة، بل ويفعل كل هذا بشكل أسرع وأكثر كفاءة.

هنا يشرح لمستخدم كفيف كل شيء في الطريق ليساعده على العبور وحتى طلب تاكسي له

وعبر حسابها الرسمي على الإنترنت، قالت شركة OpenAI: "يعد GPT4o أحدث خطواتنا في دفع حدود التعلم العميق، وهذه المرة في اتجاه سهولة الاستخدام العملي. لقد بذلنا الكثير من الجهد على مدار العامين الماضيين في العمل على تحسينات الكفاءة في كل طبقة من طبقات المكدس. كأول ثمرة لهذا البحث، أصبحنا قادرين على جعل نموذج مستوى GPT-4 متاحًا على نطاق أوسع بكثير. سيتم نشر قدرات GPT-4o بشكل متكرر".

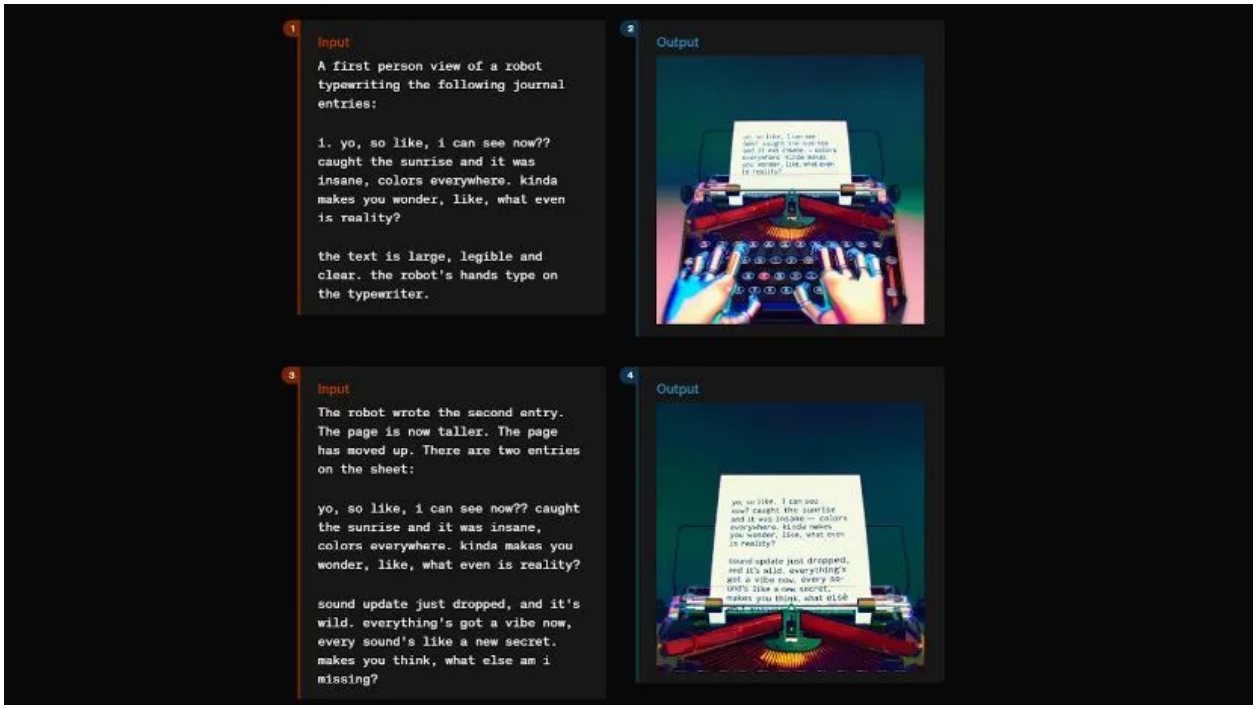
هنا يقدم النموذج درسًا خصوصيًا في مادة الرياضيات لهذا الطالب

تخيل معي أن يكون لديك مساعد رقمي يمكنه القراءة والاستماع والرؤية في وقت واحد. يشير الحرف "o" في GPT4o إلى "omni" ويعني نموذجًا موحدًا يمتلك كل الميزات والوظائف الخاصة بالنماذج الأخرى للشركة، وهذا يسلط الضوء على قدراته الشاملة، فقد أرادت الشركة من خلال النموذج الجديد، إحداث ثورة في الطريقة التي نتعامل بها مع التكنولوجيا، ما يجعل التفاعلات أكثر سهولة وديناميكية وأكثر روعة. والآن لننتقل إلى ما يمكن للجيل الجديد من GPT4 القيام به وأهم مميزات GPT4o.

ترجمة فورية

ما الجديد الذي جاء به GPT4o؟

يعد GPT4o خطوة نحو تفاعل أكثر عفوية وانسيابية بين الإنسان والآلة، حيث يقبل الذكاء الاصطناعي الجديد جميع المدخلات بما في ذلك النص والصوت والصورة والفيديو ويولد أي مجموعة من النص والصوت والصورة كمخرجات.



بناءً على الأمر المكتوب، يُنفِّذ GPT4o المطلوب على هيئة صورة، كتابة، أو فيديو

بالإضافة إلى ذلك، يمكنه الاستجابة للمدخلات الصوتية في أقل من 232 مللي ثانية، بمتوسط 320 مللي ثانية، وهو ما يشبه وقت الاستجابة البشرية في المحادثات الطبيعية.

ولمحادثات أكثر إنسانية، يمكنك طلب شيء ما من روبوت الدردشة ومن ثم مقاطعته كما يحدث بين البشر ولن تكون هناك أي مشكلة، حيث يستطيع GPT4o التأقلم بكل سلاسة وفهم ما تريده ومن ثم جلب النتائج التي تبحث عنها بشكل سريع.

الأهم من ذلك، أن الإصدار الجديد تم تدريبه على كميات ضخمة من البيانات، الأمر الذي جعله قادرًا على فهم المشاعر البشرية وحتى التقاط الفروق الدقيقة في صوت المستخدم.

أيضًا، يستطيع شات جي بي تي الجديد أن يعمل في وقت واحد وعبر أكثر من لغة بكل سهولة، هذا يعني أنك ستمتلك المترجم الخاص بك ليُترجِم لك فوريًا دون أي مشكلة.

قبل GPT-4o، كان بإمكانك استخدام الوضع الصوتي للتحدث إلى ChatGPT بزمن انتقال يبلغ 2.8 ثانية (GPT-3.5) و5.4 ثانية (GPT-4) في المتوسط. ولتحقيق ذلك، يعد الوضع الصوتي عبارة عن سلسلة من ثلاثة نماذج منفصلة: نموذج بسيط ينقل الصوت إلى نص، ويستقبل GPT-3.5 أو GPT-4 النص ويُعالجه، ليُحوِّل نموذج بسيط ثالث هذا النص مرة أخرى إلى صوت. تعني هذه العملية أن المصدر الرئيسي للذكاء، GPT-4، يفقد الكثير من المعلومات، فلا يمكنه ملاحظة النغمة بشكل مباشر، أو مكبرات الصوت المتعددة، أو الضوضاء في الخلفية، ولا يمكنه معرفة أو استيعاب نبرة الصوت وما دلالتها مثل الضحك أو الغناء أو كيفية التعبير عن المشاعر.

لكن باستخدام GPT-4o، درّبت الشركة نموذجًا واحدًا جديدًا شاملًا عبر النص والرؤية والصوت، ما

يعني أن جميع المدخلات والمخرجات تتم معالجتها بواسطة نفس الشبكة العصبية، ولهذا السبب يوفر النموذج الجديد، ردة فعل سريعة مقارنة بالنماذج الأخرى.

يتميز GPT-4o بالسلامة والأمان، من خلال إضافة تقنيات مثل تصفية بيانات التدريب وتحسين سلوك النموذج من خلال ما بعد التدريب، كما أنشأت الشركة أنظمة أمان جديدة لتوفير حواجز حماية للمخرجات الصوتية، وخضع النموذج لتقييم واسع من العديد من الخبراء في مجالات

مختلفة مثل علم النفس الاجتماعي والتحيز والعدالة والمعلومات المضللة، لتحسين تجربة المستخدم والحد من أي مخاطر ناجمة عن استخدام ال. AI.

وإليك الميزات الأخرى التي يجلبها GPT4o:

1. يمكنه أن يفهم الكلام المنطوق.
2. يستطيع فهم الكلام المكتوب.
3. يستطيع أن يفهم الصور.
4. يمكنه أن يفهم الأصوات.
5. يستطيع أن يفهم نبرة الصوت وتعابير الوجه وحتى المشاعر في الصوت والوجه.
6. يمكنه فهم واستيعاب كل ما يحدث من حوله في العالم المحيط بسهولة.
7. يستطيع معالجة وتحليل المشكلات والمسائل وتقديم خطوات سلسة وعملية للحلول.
8. يوفر إجابات منطقية ورائعة أيًا كان السؤال الخاص بك، يستطيع أن يصبح مهندس ديكور عبقري لتزيين شقتك أو أستاذًا لشرح مسألة ما أو عالمًا لتوضيح أي شيء تريده بشكل منطقي.

حجرة ورقة مقص!

Gemini 1.5 Prog GPT4o

ردًا على GPT4o، أعلنت جوجل خلال مؤتمرها السنوي للمطورين عن تحديث Gemini Nano لمعالجة النصوص والصور والأصوات بدلًا من النصوص فقط. كما كشفت النقاب عن Gemini Live وهي ميزة يمكن للمستخدم من خلالها خوض محادثة بشرية مع روبوت الدردشة Gemini. وخلال الفيديو التشويقي الذي رأيناه في المؤتمر. ستكون قادرًا على البدء في محادثة تبدو بشرية للغاية مع روبوت الدردشة الخاص بجوجل والأهم أن بإمكانك مقاطعته لطلب شيء آخر أو جلب المزيد من المعلومات عمّا تبحث عنه.

كما تخطط جوجل لإضافة تجربة الوسائط المتعددة الكاملة إلى Gemini Live في وقت لاحق من هذا العام، ما يسمح لجيميناي بمشاهدة العالم من حولك عند فتح الكاميرا في أثناء المحادثة، وهذا مشابه لما سيتمكن مستخدمو ChatGPT من القيام به قريبًا.

أخيرًا، وفقًا للشركة المطورة، سيكون GPT4o متاحًا لجميع المستخدمين دون استثناء مجانًا، وسيصل للجميع بشكل تدريجي الفترة القادمة. وبالنسبة لمستخدمي الإصدار المدفوع شات جي بي تي بلس، سيحصلون على 5 أضعاف الحد الأقصى للرسائل، بالإضافة إلى إصدار جديد من الوضع الصوتي مع GPT-4o في الأسابيع المقبلة.