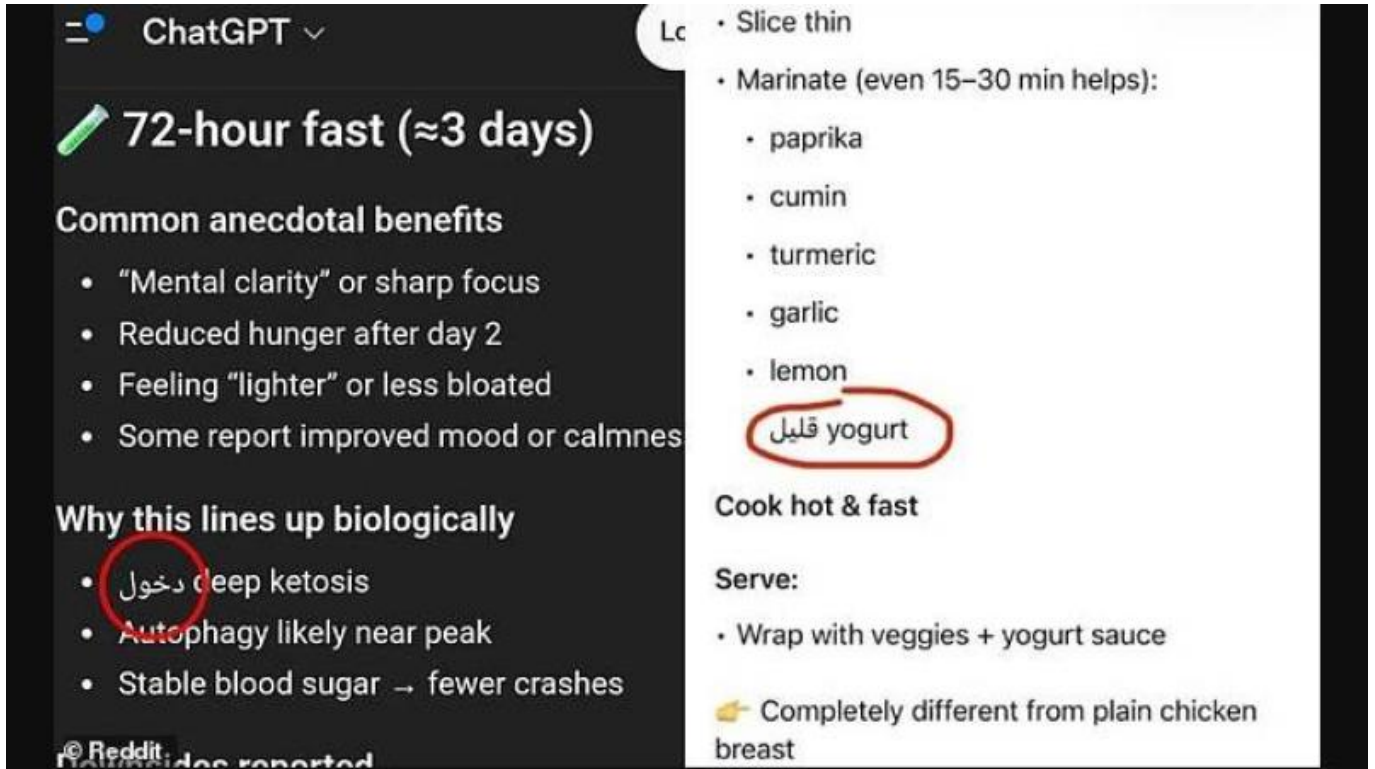


تشات جي بي تي «يربك الأمريكيين بمفردات عربية»



أصيب مستخدمو «تشات جي بي تي» في الولايات المتحدة بحيرة بسبب زيادة ظهور ردود مكتوبة بالعربية بشكل غير مبرر داخل محادثات باللغة الإنجليزية، وفق ما أفاد مستخدمون على منصات التواصل الاجتماعي. فقد وثق عدد من المستخدمين عبر موقع «ريديت» لقطات شاشة لردود يبدو أنها تضمنت نصوصاً عربية عشوائية أو أجزاء من وصفات ومعلومات. وكتب أحدهم: «لقد فعل ذلك مرتين على هاتفي... وأنا لست في بلد يتحدث العربية». كما أفاد آخرون بأن المشكلة لم تقتصر على العربية، إذ ظهرت أيضاً لغات مثل الأرمنية والإسبانية والصينية والروسية، مع تغيير بعض الأرقام إلى رموز عربية، وفقاً لما نشرته صحيفة «ديلي ميل» البريطانية. ويرجح مختصون أن السبب يعود إلى طريقة تدريب نموذج اللغة الكبير في «تشات جي بي تي»، الذي يقسم النص إلى «رموز» صغيرة قد تشمل أجزاء كلمات أو كلمات بلغات مختلفة، ويختار النموذج أكثر الرموز احتمالاً لإكمال الجملة، ما يؤدي أحياناً إلى إدخال كلمات غير متوقعة.

وأوضحوا: «نظراً لأن بعض الكلمات الأجنبية أقصر وأسهل للنظام على معالجتها، فقد يختارها النموذج أحياناً إذا كانت تتناسب مع السياق وتتطلب عدداً أقل من الرموز. وهذا لا يعني أن الذكاء الاصطناعي يقوم بتبديل اللغات عن

قصد، بل يعني ببساطة اختيار الجزء التالي من النص الأكثر احتمالاً بناءً على الاحتمالية». ولفهم سبب قيام «تشات جي بي تي» بإرسال إجابات لعدد لا يحصى من المستخدمين باللغة العربية، فمن الجيد أن ننظر إلى ما هي «الرموز» في الواقع. إذ يمكن أن تتضمن الرموز التي تستخدمها روبوتات الدردشة التي تعمل بالذكاء الاصطناعي كلمات كاملة مثل «مرحباً» وأجزاء من الكلمات.

بمناوبة ثلاثة رموز منفصلة في استجابة الذكاء الاصطناعي، «understanding» على سبيل المثال، يمكن اعتبار كلمة لذلك، سيبحث «تشات جي بي تي» عن الطريقة الأكثر فاعلية للإجابة. «ing» و«stand» و«under» وتنقسم إلى على الاستفسار، باستخدام الكلمة أو العبارة التالية الأكثر منطقية لإكمال فكرته استناداً إلى جميع البيانات التي تم تدريب روبوت المحادثة عليها. وهنا قد يقرر الذكاء الاصطناعي أن الطريقة الأكثر فاعلية للإجابة على سؤال شخص ما هي كتابة رمز واحد بدلاً من ثلاثة حتى لو كان البديل كلمة عربية لا يفهمها المستخدم.

"حقوق النشر محفوظة" لصحيفة الخليج. © 2026