

ضبابية المشهد المالي «تُلعثم» روبوتات الدردشة الذكية



بعد أن محا الإعلان التجريبي المتسرع لروبوت الدردشة الذكي «بارد» حوالي 100 مليار دولار من القيمة السوقية للشركة المطورة «جوجل»، لمجرد منافسة أداة «تشات جي بي تي» الذكية التي طورتها «أوبن إيه آي» المدعومة من «مايكروسوفت»، تكشّف أمام عملاقي التكنولوجيا والمحللين تحدٍ آخر يُضاف إلى جهود الذكاء الاصطناعي التوليدي في محركات بحث «جوجل» و«بينغ» وغيرهما، ألا وهو «التكلفة» المصاحبة لهذه المنصات البحثية المتطورة

• النماذج اللغوية الكبيرة

تحدث مسؤولون تنفيذيون عبر قطاع التكنولوجيا الأوسع عن آليات تشغيل تقنيات ذكاء اصطناعي مثل «تشات جي بي تي» وكيفية حساب التكاليف المرتفعة. وقال سام ألتمان، الرئيس التنفيذي لشركة «أوبن إيه آي» الناشئة، إن برنامج الذي يمكنه صياغة الموضوعات النظرية والإجابة على استفسارات البحث خلال open AI الدردشة الآلي من تطوير ثوان، له تكاليف حوسبة «مذهلة» تبلغ سنتين أو أكثر لكل محادثة

وهو ما أكده جون هينيسي، رئيس شركة ألفابيت في مقابلة مع «رويترز»، بأن التبادلات مع الذكاء الاصطناعي المعروفة باسم «النماذج اللغوية الكبيرة» من المحتمل أن تكلف 10 أضعاف تكلفة البحث القياسي العادي عن الكلمات الرئيسية.

وأفاد محللون بأنه حتى مع عائدات الإعلان المحتملة التي ستجنيها محركات البحث القائمة على الذكاء الاصطناعي التوليدي والدردشة، يمكن لهذه التقنيات أن تكلف شركة مثل «ألفابيت» عدة مليارات من الدولارات سنوياً تضاف إلى ميزانيتها العمومية.

• مليارات دولار إضافية 6

قدّرت «مورغان ستانلي» أن 3.3 تريليون عملية بحث في «جوجل» العام الماضي كلفت الشركة حوالي «خمس السنت» لكل منها، وهو رقم سيزداد اعتماداً على مقدار النص الذي يجب أن يُولده الذكاء الاصطناعي. متوقعة زيادة في نفقات عملاق البحث العالمي بمقدار 6 مليارات دولار بحلول عام 2024، وذلك في حال تعامل الذكاء الاصطناعي الشبيه بـ «تشات جي بي تي» مع نصف الاستفسارات التي يتلقاها بإجابات مكونة من 50 كلمة فقط.

بدورها توصلت شركة «سيمي أناليسز» للأبحاث والاستشارات، وتركز على تقنية الرقائق، إلى فاتورة مماثلة بطرق قد يكلف ألفابيت ChatGPT 3 مختلفة. وقالت الشركة، إن إضافة تقنيات الذكاء الاصطناعي للبحث على غرار «TPUs» مليارات دولار إضافية لوحدة المعالجة الداخلية من «جوجل»، أو

• الاستدلال

ما يجعل هذا الشكل من الذكاء الاصطناعي مكلفاً أكثر من البحث التقليدي هو قوة الحوسبة المستخدمة فيه. فبحسب الخبراء، يعتمد الذكاء الاصطناعي التوليدي على شرائح تقنية تبلغ قيمتها مليارات الدولارات، وهي تكلفة يجب توزيعها على مدى عمرها الإنتاجي لعدة سنوات. وبالمثل، تضيف فواتير الكهرباء مزيداً من التكاليف والضغط على الشركات ذات الأهداف المتعلقة ببصمة الكربون.

وفي سياق متصل، تُعرف آلية التعامل مع استعلامات البحث المدعومة بالذكاء الاصطناعي بـ «الاستدلال»، حيث يتم تشكيل «شبكة عصبية» دقيقة على غرار بيولوجيا الدماغ البشري، تستدل على إجابة السؤال من تدريب مسبق. وهذا يخالف تماماً البحث التقليدي من «جوجل»، التي تمتلك فهرساً من المعلومات، وبمجرد أن يكتب المستخدم استعلاماً ما، يُقدم المحرك الإجابات الأكثر صلة والمخزنة في الفهرس. ومع ذلك، وصف هينيسي تكاليف استدلال تلك المشكلة You.com التي قد تدوم عامين في أسوأ الأحوال. وقال منافس آخر لجوجل، وهو الرئيس التنفيذي لمحرك البحث ريتشارد سوتشر، إن إضافة تجربة دردشة عبر الذكاء الاصطناعي، فضلاً عن تطبيقات الرسوم البيانية ومقاطع الفيديو وغيرها من التقنيات التوليدية الأخرى رفعت النفقات العامة بين 30% و50%. كما حذر مصدر آخر مقرب من عملاق البحث عبر الإنترنت من أنه من المبكر لجوجل تحديد المبلغ الذي قد تأتي به برامج الدردشة، لأن الكفاءة والاستخدام يختلفان بشكل كبير اعتماداً على التكنولوجيا المستخدمة.

• ترويض التكلفة

لسنوات مضت، درس الباحثون في ألفابيت وأماكن أخرى كيفية تدريب وتشغيل نماذج لغوية كبيرة بتكلفة أقل. حيث

تتطلب النماذج الأكبر حجماً مزيداً من الرقائق للاستدلال وبالتالي تأتي مع تكلفة أعلى. وأخبر أحد كبار التنفيذيين في مجال التكنولوجيا رويترز، أن مثل هذا الذكاء الاصطناعي يظل باهظ التكلفة لوضعه في أيدي ملايين الأشخاص.

وهنا تجدر الإشارة إلى أن حجم الذكاء الاصطناعي الذي أبهر المستخدمين بقدراته الشبيهة بالبشر قد تضاعف لحد وصلت فيه قدرات النموذج اللغوي الفائقة في روبوت الدردشة تشات جي بي تي إلى 175 مليار عامل رقمي باراميترز أو (قيمة مختلفة تأخذها الخوارزمية في الاعتبار). وتختلف التكلفة أيضاً حسب طول استعمال المستخدم، كما يتم قياسها في الرموز أو أجزاء من الكلمات.

في الوقت الحالي، اكتشف علماء الكمبيوتر داخل أوبن إيه آي كيفية تحسين تكاليف الاستدلال تلك من خلال التعليمات البرمجية المعقدة التي تجعل الرقائق تعمل بكفاءة أكبر.

وتتمثل مشكلة أوبن إيه آي طويلة المدى في كيفية تقليص أعداد الباراميترز في نموذج الذكاء الاصطناعي بنحو 10 مرات أو حتى 100 مرة دون فقدان الدقة. وقال نافين راو، الذي كان يدير سابقاً جهود شرائح الذكاء الاصطناعي الخاصة بشركة إنتل، ويعمل الآن على خفض تكاليف الحوسبة في شركته الناشئة موازيك إم إل: يبقى السؤال المفتوح هو، كيف يمكنك استبعاد العوامل الرقمية في الخوارزمية بشكل أكثر فاعلية؟

• رسوم الاشتراك جزء من الحل

في غضون ذلك، فكر البعض في فرض رسوم على الوصول إلى تطبيقات الذكاء الاصطناعي الفائقة، مثل اشتراك ذي ال 20 دولاراً شهرياً للحصول على خدمات أفضل من روبوت تشات جي بي تي. فيما أفاد خبراء OpenAI تكنولوجيا آخرون بأن الحل البديل هو تطبيق نماذج ذكاء اصطناعي أصغر ضمن مهام أبسط، وهو ما تستكشفه شركة ألفا بيت، التي قالت مؤخراً إن نسخة أصغر من نموذج لغة تطبيقات الحوار (المعروفة اختصاراً بتقنية لامدا ستعمل على تشغيل روبوت الذكاء الاصطناعي الخاص بها (بارد) مع طاقة حوسبة أقل بكثير، ما يُمكن الشركة من التوسع لتشمل المزيد من المستخدمين.

ولدى سؤاله عن روبوتات الدردشة مثل تشات جي بي تي وبارد، قال رئيس ألفا بيت جون هينيسي، في مؤتمر الأسبوع الماضي، إن النماذج المتعددة الأكثر تركيزاً، بدلاً من نظام واحد يفعل كل شيء، من شأنها أن تساعد في كبح جماح التكاليف المرتفعة.