حتى أقوى النماذج تفشل... الذكاء الاصطناعي غير جاهز للعمل المكتبي



بعد نحو عامين من توقعات ساتيا ناديلا، الرئيس التنفيذي لشركة مايكروسوفت، حول سيطرة الذكاء الاصطناعي التوليدي على "أعمال المعرفة"، لا يزال الواقع العملي في مكاتب المحاماة والبنوك الاستثمارية مختلفاً تماماً، حيث يحتفظ البشر بدورهم الأساسي في إدارة العمل المعقد.

دراسة حديثة توضح سبب تعثر هذه الثورة الرقمية:

الذكاء الاصطناعي غير قادر على التعامل مع فوضى العمل الحقيقي.

اختبار حقيقي للقدرات

كشفت شركة ميركور المتخصصة في بيانات التدريب عن معيار جديد أسمته "Agents-APEX"، وهو اختبار يبتعد عن المهام البسيطة التي تتباهى بها النماذج عادةً، مثل كتابة القصائد أو حل المعادلات.

والاختبار اعتمد على مهام مكتبية واقعية من محامين ومستشارين وبنكيين، تطلب تنفيذ خطوات متعددة مع التنقل بين مصادر بيانات مختلفة.

والنتائج كانت صادمة: حتى أقوى النماذج الحالية، مثل Flash 3 Gemini و5.2-GPT، لم تتجاوز 25% من الدقة.

جاء جيميني في الصدارة بنسبة 24%

تلاه 5.2-GPT بنسبة 23%

باقي النماذج سجلت نسب أقل بكثير.

أين يفشل الذكاء الاصطناعي؟

بحسب الرئيس التنفيذي لشركة ميركور، بريندان فودي، المشكلة ليست في القدرة الحسابية، بل في فهم السياق.

وفي بيئة العمل الواقعية، الإجابات ليست جاهزة:

يحتاج المحامي لمراجعة محادثات على "سلاك"،

وقراءة سياسات داخل ملفات PDF،

وتحليل جداول بيانات،

ثم ربط كل هذه المعلومات للإجابة على أسئلة الامتثال لقوانين مثل GDPR.

والبشر يقومون بهذه التنقلات المعقدة بشكل طبيعي، بينما النماذج تفشل عند التعامل مع مصادر متفرقة، فتخطئ أو تتوقف أو تتوه.

الذكاء الاصطناعي متدرب غير موثوق

الدراسة تمنح أملاً مؤقتاً للعاملين في المكاتب:

النماذج الحالية تشبه "متدرباً غير موثوق"، تنجح في إنجاز المهمة مرة واحدة من كل أربع محاولات.

لكن التطور سريع: قبل عام كانت الدقة بين 5% و10%، واليوم اقتربت من 24%، ما يدل على سرعة التعلم المتصاعدة.

الثورة الرقمية مؤجلة

حتى الآن، لا يبدو أن الذكاء الاصطناعي مستعد لتولي زمام العمل المكتبي أو استبدال أصحاب أعمال المعرفة.

الرسالة الأوضح من الدراسة: المسألة ليست "هل سيصل؟"، بل متى؟.

إلى أن تتقن النماذج مهارات تعدد المهام وفهم السياق المعقد، ستبقى الثورة الرقمية في المكاتب مؤجلة، وسيستمر البشر في قيادة المشهد.