شرح وكلاء الذكاء الاصطناعي

وكيل الذكاء الاصطناعي نظام برمجي يستخدم نموذجا للذكاء الاصطناعي للعمل نحو هدف، فيقرر خطواته بنفسه ويستخدم أدوات لتنفيذها. وعلى خلاف روبوت الدردشة البسيط الذي يجيب فقط، يستطيع الوكيل أن يتصرف: فيبحث ويقرأ الملفات ويرسل الرسائل أو يغير السجلات. وتختلف التعريفات، لذلك يعتمد هذا المقطع نظرة بسيطة شائعة. ونتابع هادي، مدير العمليات في متجر إلكتروني.

1. ما هو الوكيل

الفرق الجوهري هو الهدف. فمع روبوت الدردشة، نطرح سؤالا ونتلقى إجابة. أما مع الوكيل، فنعطي هدفا، مثل العثور على المدفوعات المتأخرة، ويختار الوكيل الخطوات بنفسه. ويستمر إلى أن يقرر أن الهدف تحقق أو أنه يحتاج إلى مساعدة.

مثال: طلب أفكار لمطاعم من صديق يعطي قائمة، كما يفعل روبوت الدردشة. أما طلب تنظيم عشاء عيد ميلاد من مساعد فيعني أن المساعد يختار الخطوات وينجزها، مثل إيجاد مكان وحجز طاولة ومراسلة الضيوف، كما يفعل الوكيل.

في العمل: يعطي هادي وكيلا هدف العثور على كل الفواتير المتأخرة ثلاثين يوما وتذكير العملاء. ويختار الوكيل الخطوات بنفسه، ويراجع هادي النتيجة في النهاية.

2. الأدوات والحلقة

يعمل الوكلاء في حلقة. أولا، يخطط النموذج الخطوة التالية من الهدف ومما يعرفه بالفعل. ثم يتصرف باستدعاء أداة، مثل بحث أو استعلام قاعدة بيانات أو نظام بريد إلكتروني. بعد ذلك يلاحظ النتيجة ويقرر ما يفعله تاليا، وتتكرر الحلقة حتى يتحقق الهدف. والأدوات والمعلومات التي يراها الوكيل تمنحه قدرات حقيقية، لذلك ينبغي اختيارها بعناية.

مثال: يتبع الطاهي الحلقة نفسها: يخطط الخطوة التالية، ويضيف مكونا، ويتذوق الطبق، ويعدل. وتستمر الحلقة حتى يصير مذاق الطبق مناسبا.

في العمل: يخطط وكيل هادي للعثور على الفواتير المتأخرة ويستعلم قاعدة بيانات الفواتير. ويلاحظ قائمة بالحسابات المتأخرة ويصوغ رسالة تذكير لكل حساب. ثم يتحقق من الردود ويخطط الخطوة التالية.

3. مستويات الاستقلالية

يختلف الوكلاء في مقدار الحرية التي يملكونها، وإحدى الطرق الشائعة لرؤية ذلك هي المستويات. في المستوى الأدنى، يكتفي الوكيل بالاقتراح وينفذ شخص العمل. وفي المستويات الوسطى، يتصرف الوكيل لكنه يطلب الموافقة قبل الخطوات

حقوق الطبع والنشر محمية © - أكاديمية أساسيات (شركة أساسيات للخدمات الاستشارية)

المهمة، أو يتصرف وحده في المهام الصغيرة ويبلغ بعد ذلك. وفي المستوى الأعلى، يعمل الوكيل وحده فترات طويلة، وهذا نادر ويحتاج إلى ضوابط قوية.

مثال: يقترح متدرب جديد الردود في البداية ويرسلها مدير. ولاحقا يرسل المتدرب الرسائل الصغيرة وحده لكنه يسأل قبل أي أمر مهم.

في العمل: يشغل هادي وكيل الفواتير بمستوى الموافقة، فيصوغ التذكيرات ويوافق هو على كل منها. وبعد أسابيع من النتائج الجيدة، يرسل التذكيرات بالمبالغ الصغيرة وحده ولا يزال يسأله قبل مراسلة العملاء الكبار.

4. المخاطر والضوابط

يجلب الوكلاء مخاطر جديدة. فقد يخطئون بعواقب حقيقية، مثل إرسال رسالة خاطئة أو حذف سجل. وقد تخدعهم أيضا تعليمات مختبئة في المحتوى الذي يقرؤونه، وهذا ما يسمى حقن الأوامر، أو قد يكشفون بيانات إذا منحوا صلاحيات أكثر من اللازم. ومن الضوابط الجيدة منح أقل صلاحية لازمة، وطلب الموافقة قبل الإجراءات الخطرة أو غير القابلة للتراجع، وتحديد الوقت والتكلفة، وحفظ سجل لكل إجراء. وتساعد أطر مثل إطار إدارة مخاطر الذكاء الاصطناعي من NIST المؤسسات على الحوكمة وتحديد السياق والقياس والإدارة لهذه المخاطر.

مثال: قبل إعطاء مساعد مفتاح البيت، نحدد الغرف التي يدخلها، ونطلب منه الرجوع إلينا قبل أي إنفاق، ونحتفظ بالإيصالات. ويبقى المساعد مفيدا، لكن بحدود وسجل.

في العمل: يستطيع وكيل هادي قراءة الفواتير وإرسال التذكيرات، لكنه لا يستطيع إصدار المبالغ المستردة أو تغيير الأسعار. ويطلب الموافقة قبل أي رسالة إلى عميل كبير، ويكتب كل إجراء في سجل. ويختبر هادي الوكيل أولا على مجموعة صغيرة من العملاء.

الخاتمة

وخلاصة القول، يسعى وكيل الذكاء الاصطناعي إلى هدف بحلقة من التخطيط والتنفيذ والملاحظة، مستخدما أدوات لاتخاذ إجراءات حقيقية. ويطبق في أربع خطوات: معرفة ما الوكيل، وفهم الأدوات والحلقة، واختيار مستوى الاستقلالية، ووضع ضوابط للمخاطر. وعندما يستخدم بحدود واضحة وإشراف بشري، يوفر الوقت في الأعمال المتعددة الخطوات دون فقدان السيطرة.

المصادر

• المعهد الوطني الأمريكي للمعايير والتقنية (NIST (2023: إطار إدارة مخاطر الذكاء الاصطناعي AI RMF 1.0 – https://www.nist.gov/itl/ai-risk-management-framework
• شركة Anthropic (2024): Building effective agents – https://www.anthropic.com/engineering/building-effective-agents
• شركة OpenAI (2025): A practical guide to building agents – https://cdn.openai.com/business-guides-and-resources/a-practical-guide-to-building-agents.pdf

حقوق الطبع والنشر محمية © - أكاديمية أساسيات (شركة أساسيات للخدمات الاستشارية)