

الذكاء الاصطناعي الوكيل (Agentic AI) وتهديداته في بيئة العمل

امثلة على الذكاء الاصطناعي الوكيل:

التعلم الذاتي:



عند تثبيت برامج جديدة على النظام تتطلب مهارات معينة لاستخدامها.

يقوم الذكاء الاصطناعي الوكيل بالتدريب والتعلم الذاتي على استخدام هذه البرامج حسب الصلاحيات الممنوحة له.

تحديث البرامج والانظمة:



منح الذكاء الاصطناعي الوكيل صلاحيات على النظام لمتابعة آخر التحديثات للبرامج والنظام نفسه بشكل دوري.

فيقوم بتنبيهه المستخدم بوجود تحديث جاهز للتثبيت، ثم يحدث الأنظمة والبرامج ذاتياً.

تنسيق الاجتماعات:



يقوم المستخدم بطلب تنسيق اجتماعاً مع الفريق الأسبوع القادم.

فيقوم الذكاء الاصطناعي الوكيل بالبحث عن الأوقات المناسبة، واختيار الموعد، وتجهيز الدعوة وإرسالها، بحسب الصلاحيات الممنوحة له.

لم يعد الذكاء الاصطناعي يقتصر على كتابة النصوص، تلخيص الملفات، أو الإجابة عن الأسئلة. فقد تطورت بعض الأنظمة إلى ما يُعرف بـ الذكاء الاصطناعي الوكيل (Agentic AI)، وهي أنظمة قادرة على فهم الهدف، التخطيط، استخدام الأدوات، واتخاذ خطوات عملية لتنفيذ المهمة بشكل شبه مستقل.

عناصر الذكاء الاصطناعي الوكيل وكيفية عمله

يمكن فهم الية عمله كالتالي:

فهم الهدف والتخطيط:

فهم المطلوب وتحليل السياق وتقسيم المهام إلى خطوات صغيرة باستخدام الذكاء الاصطناعي (AI Model).

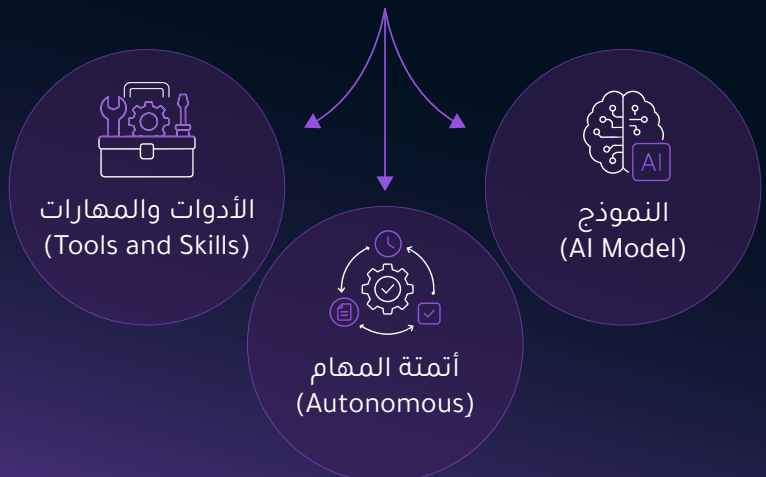
استخدام الأدوات والمهارات:

استخدام الأدوات والمهارات اللازمة لتنفيذ المهمة مثل مهارة استخدام البرامج الخاصة بالبريد الإلكتروني (Outlook).

أتمتة المهام:

متابعة البريد الإلكتروني بشكل دوري والرد (بخطاب رسمي أو بريد إلكتروني) وفق التعليمات وذلك باستخدام المهارات وأتمتة المهام والتعلم الذاتي المستمر.

الذكاء الاصطناعي الوكيل (Agentic AI)



الفرق بشكل بسيط هو:

الذكاء الاصطناعي التقليدي (Chatbot) يقدم لك إجابات.  الذكاء الاصطناعي الوكيل (Agentic AI) يستطيع تنفيذ إجراء. 

التحديات التقنية في عناصر الذكاء الاصطناعي الوكيل

عندما يتم منح الذكاء الاصطناعي الوكيل الصلاحيات الكاملة للتصرف الذاتي، فإن مساحة التحديات المحتملة تتسع بشكل أكبر من خلال التالي:



أتمتة المهام: (Autonomous)

يعمل الوكيل بسرعة فائقة، وفي غياب الرقابة البشرية، قد يكرر النظام الخطأ أو الأمر الخبيث بسرعة كبيرة، مما قد يؤدي إلى حذف بيانات أو تعطيل الخدمات أو زيادة تكاليف الخدمات.



الأدوات والمهارات (Tools and Skills)

منح الوكيل صلاحيات واسعة للوصول إلى الأنظمة وتنفيذ الأوامر يعني أن أي اختراق للوكيل أو نجاح هجوم حقن الأوامر (Prompt Injection) قد يمكن المهاجم من إساءة استغلال هذه الصلاحيات للوصول إلى الأنظمة، وتنفيذ إجراءات أو أوامر غير مصرح بها، مما يوسع نطاق الهجوم ويزيد من أثره.



النموذج الذكي (AI Model)

تعاني نماذج الذكاء الاصطناعي من مخرجات غير دقيقة أو غير صحيحة وتعرضها بثقة عالية (الهلوسة). وفي الأنظمة الوكيلية، تترجم هذه المعلومات والهلوسات مباشرة إلى إجراءات وأوامر حقيقية يتم تنفيذها آلياً.

كيف يتم استغلاله من قبل المهاجمون

تشغيل المهارات غير الموثوقة:

بعض الوكلاء يعتمدون على مهارات أو إضافات خارجية لتوسيع قدراتهم. لكن تشغيل مهارة من مصدر غير موثوق يشبه تشغيل برنامج خبيثة على جهاز العمل. 

حقن الأوامر غير المباشر (Indirect Prompt Injection):

يقرأ الوكيل بيانات خارجية كرسائل البريد الإلكتروني والمواقع، ويمكن للمهاجمين إخفاء تعليمات خبيثة داخل هذه البيانات، ليتعامل معها الوكيل كأوامر رسمية وينفذها تلقائياً. 

تسريب البيانات:

قد يحتاج النظام إلى بيانات دخول أو مفاتيح تشغيل لتنفيذ المهمة. وإذا لم تتم حمايتها جيداً، فقد تتعرض للتسريب أو السرقة. 

قد يستغل المهاجمون الصلاحيات الممنوحة للذكاء الاصطناعي الوكيل أو يتلاعبون بالمحتوى الذي يتعامل معه، لدفعه إلى تنفيذ إجراءات ضارة أو الوصول إلى بيانات حساسة دون علم المستخدم من خلال التالي:

استخدم الذكاء الاصطناعي الوكيل بحذر

يمكن للذكاء الاصطناعي الوكيل إنجاز المهام بسرعة، لكن استخدامه الآمن يتطلب التحقق المستمر وتطبيق الضوابط التالية:

فرض الموافقة البشرية
(Human-in-the-Loop) اليدوية قبل تنفيذ أي عمليات حساسة.



منحه أقل قدر من الصلاحيات التي تسمح له بالوصول إلى ما يحتاجه لتنفيذ المهمة فقط، وشغله في بيئة آمنة.



مراقبة ما ينفذه النظام والإبلاغ عن أي سلوك غير متوقع.



التأكد من موثوقية مصادر النماذج الذكية والمهارات قبل اعتمادها أو استخدامها.



هي نماذج يمكن تشغيلها محلياً على الأجهزة والخوادم الخاصة دون الحاجة إلى الاتصال بالإنترنت خلال استخدامها. وقد يساعد ذلك في حماية الخصوصية، لكنه لا يجعله آمناً تلقائياً، إلا أنها تفرض التحديات التالية:

تحميل نموذج مفتوح المصدر من جهة غير موثوقة.



استخدام نموذج غير مقيد (Uncensored LLM) لا يطبق ضوابط السلامة.



صعوبة مراقبة الاستخدام إذا تم تشغيله خارج القنوات الرسمية (Shadow AI).



الحاجة إلى بنية تحتية قوية وآمنة.



ثقافة سبيرانية:



هل تعلم ما المقصود
بالذكاء الاصطناعي
المحلي أو النماذج
المحلية (Local LLM)؟

