

البنية المرجعية للذكاء الاصطناعي الوكيل

تصميم أنظمة وكيلية آمنة وقابلة للحوكمة والتشغيل
البيئي للمؤسسات

المحتويات

٣	 الملخص التنفيذي
٤	 ١. المقدمة
٥	 ٢. مبادئ التصميم
٦	 ٣. البنية المرجعية
٨	 ٤. نظرة عامة على المكونات
١١	 ٥. أنماط تصميم الوكلاء ومستويات الاستقلالية
١٣	 ٦. مسار الطلب من البداية إلى النهاية
١٤	 ٧. البنية الأمنية
١٥	 ٨. التقييم وعمليات تشغيل الوكلاء
١٦	 ٩. الحوكمة والمواءمة التنظيمية
١٨	 ١٠. الاعتبارات غير الوظيفية
١٩	 ١١. مشهد الأدوات والتقنيات
٢١	 ١٢. خارطة طريق التبني
٢٢	 الملحق أ. مسرد المصطلحات
٢٣	 الملحق ب. المراجع

الملخص التنفيذي

يحتاج الوكلاء في المؤسسات إلى بنية تربط مهام الأعمال بالوصول المنضبط والإشراف البشري والتشغيل الموثوق. وتعرض هذه الورقة النهج الذي توصي به TenX: البدء بسير العمل، وإضافة الاستقلالية حيثما تُحسن النتيجة، وتصميم الضوابط حول الإجراءات التي يمكن للنظام تنفيذها.

تعود معظم حالات الإخفاق في نشر الوكلاء داخل المؤسسات إلى أسباب معمارية لا إلى النماذج: وكلاء يملكون صلاحيات وصول أوسع مما تتطلبه المهمة، ومحتوى غير موثوق يوجّه استدعاءات الأدوات، وغياب طريقة موثوقة لتقييم أي تغيير قبل إصداره، وعدم وجود مسار تدقيق يمكن أن تقبله جهة تنظيمية أو لجنة مخاطر. وقد صُممت هذه البنية لسدّ هذه الثغرات.

تستخدم البنية MCP للأدوات والسياق، وA2A للتعاون بين الوكلاء المستقلين، وOpenTelemetry لبيانات القياس عن بُعد. وتستند جوانب الأمن والحوكمة إلى إرشادات OWASP، ومعياري ISO/IEC ٤٢٠٠١، وإطار إدارة مخاطر الذكاء الاصطناعي الصادر عن NIST، وقانون الاتحاد الأوروبي للذكاء الاصطناعي. وتوفر هذه الواجهات والأطر أساساً مشتركاً للتنفيذ والضمان.

يرتكز هذا النهج على سبع ممارسات:

- **الوكلاء هويات خاضعة للمساءلة.** إسناد الإجراءات إلى الوكيل ومالكه، مع التحقق قبل استخدام الأدوات من أن الصلاحيات لا تتجاوز الحد الأدنى اللازم.
- **أربعة مستويات تحكم شاملة.** تطبيق الهوية والسلامة وقابلية المراقبة والحوكمة عبر النظام بأكمله.
- **معايير مفتوحة عند نقاط الربط.** استخدام MCP وA2A وOpenTelemetry حيثما يتطلبها التكامل.
- **الإشراف البشري أثناء التشغيل.** اشتراط الموافقة وفقاً لمخاطر الإجراء والسياسة المعمول بها.
- **تحسين مشروط بالتقييم.** تقييم كل إصدار، وإخضاع التغييرات الجوهرية لاختبارات الفريق الأحمر قبل ترقيتها.
- **تصميم الوكلاء انطلاقاً من الأنماط.** اختيار أبسط نمط لسير العمل أو للوكيل يلي المتطلبات.
- **مشهد تقني غير مرتبط بإصدارات محددة.** اختيار المنتجات والنماذج وفقاً لحالة الاستخدام، بدلاً من تطبيق الكتلوج بأكمله.

نقطة الانطلاق: إرساء الضوابط اللازمة لحالة استخدام واحدة، ثم توسيع القدرات المشتركة مع نمو المتطلبات. وتظل زيادة الاستقلالية خاضعة للسياسات وللأدلة المستمدة من التقييم والتشغيل. وباختصار، تُكتسب الاستقلالية بالأدلة: إذ يبدأ الوكلاء بصلاحيات محدودة وموافقة بشرية، ولا ينالون مزيداً من الاستقلالية إلا بقدر ما تبرره نتائج التقييم ومعدلات التجاوز وسجل التشغيل.

١. المقدمة

تفسّر الأنظمة الوكيلية الأهداف، وتختار الأدوات، وتكيّف خطواتها مع تقدم العمل. وتكون مفيدة عندما تتطلب المهمة تخطيطًا أو تفسيرًا مرئيًا. أما حينما يكون التسلسل قابلاً للتنبؤ، فيظل سير العمل المحدد هو نقطة البداية. تستند هذه البنية المرجعية إلى خبرة TenX في تنفيذ حالات استخدام وكيلية عبر قطاعات متعددة. وهي تصف المكونات وأنماط التفاعل والضوابط اللازمة للنشر في المؤسسات، مع خيارات تقنية تمثيلية في القسم ١١. توصي TenX بالبداية بسير العمل الخاص بالأعمال، وعدم إدخال الاستقلالية إلا حينما تُحسّن النتيجة. فينبغي أن تعتمد العمليات القابلة للتنبؤ على مسارات عمل محددة، وأن يتولى الوكلاء المهام التي تتطلب مرونة في التفسير أو التخطيط أو اختيار الأدوات. كما ينبغي تصميم الصلاحيات والموافقة البشرية والتقييم والتعافي حول إجراءات الأعمال التي يمكن للنظام تنفيذها. ويمكن توسيع البنية مع نمو حالات الاستخدام والمتطلبات التشغيلية، ويعرض القسم ١٢ مسار التنبؤ.

المشهد الحالي

- **الاستقلالية والأمن.** يجعل استخدام الأدوات والتفويض من الوصول محدد النطاق، والتنفيذ الجدير بالثقة، والتحكم في المحتوى غير الموثوق أمورًا أساسية.
- **قابلية التشغيل البيني.** يوفر MCP واجهة مشتركة للأدوات والسياق، ويدعم A2A التعاون بين الوكلاء المستقلين.
- **التقييم.** تقييم الخطوات التي يتخذها الوكيل إلى جانب النتيجة النهائية، واستخدام النتائج في حوكمة الإصدارات.
- **الحوكمة.** توثيق تقييمات المخاطر والموافقات والأدلة التشغيلية ضمن أعمال التنفيذ والتشغيل الاعتيادية.

الجمهور المستهدف

هذه الوثيقة موجهة إلى الرؤساء التنفيذيين للتقنية، ورؤساء البيانات والذكاء الاصطناعي، ومعماري المؤسسات والحلول، وقادة الأمن والمخاطر والامتثال، وفرق التنفيذ التي تصمم الأنظمة الوكيلية أو تتولى ضمانها.

٢. مبادئ التصميم

تشكّل المبادئ التالية الأساس لكل مكون من مكونات البنية. وحيثما يكون قرار التصميم غير واضح، ينبغي الاستناد إلى هذه المبادئ لحسمه.

المبدأ	ما يعنيه عمليًا
تفضيل مسارات العمل الحتمية على الوكلاء، وتفضيل الوكيل الواحد المزود بالأدوات على الأنظمة متعددة الوكلاء. وعدم إضافة الاستقلالية إلا حيثما يُثبت التقييم أنها تُحسن النتائج.	أبسط نمط يحقق الغرض
لكل وكيل هويته الخاصة، ومالك محدد بالاسم، وصلاحيات محددة النطاق بشكل صريح. ويعمل الوكيل نيابةً عن مستخدم أو عملية من خلال صلاحيات مفوَّضة ومقيدة بمدة زمنية، ولا يعمل أبدًا من خلال بيانات اعتماد خدمة مشتركة.	الوكلاء هويات
التعامل مع مدخلات المستخدم والمحتوى المسترجع ومخرجات الأدوات والرسائل الواردة من الوكلاء الآخرين على أنها بيانات غير موثوقة قد تحمل تعليمات. والتحقق من الصحة عند كل نقطة ربط.	عدم الثقة افتراضيًا
تحديد مستوى الاستقلالية لكل إجراء وفقًا لمخاطره وإمكانية التراجع عنه. وتتطلب الإجراءات ذات الأثر الجوهري أو غير القابلة للتراجع موافقة بشرية إلى أن تدعم الأدلة مستوى أعلى من الاستقلالية.	استقلالية ضمن حدود
لا يصل أي تغيير على الأوامر النصية أو الأدوات أو النماذج أو منطق الوكيل إلى بيئة الإنتاج دون اجتياز حزمة تقييم، ودون اختبارات الفريق الأحمر في حالة التغييرات الجوهرية.	التقييم قبل الترقية
استخدام MCP وA2A وOpenTelemetry عند نقاط التكامل، بحيث يمكن استبدال النماذج والأطر والموردين دون إعادة التصميم.	معايير مفتوحة عند نقاط الربط
تُتَبَّع كل خطوة وكل استدعاء للنموذج وكل استدعاء لأداة وكل قرار، ويمكن نسبتها إلى مصدرها، بما يوفر أدلة التدقيق التي تحتاجها وظائف المخاطر والوظائف التنظيمية.	قابلية المراقبة والتدقيق منذ التصميم
يصمد العمل طويل الأمد أمام الأعطال من خلال نقاط الحفظ، والإجراءات الآمنة عند التكرار، والخطوات التعويضية المحددة، بحيث لا يترك الوكيل أي عملية أعمال مكتملة جزئيًا.	الديمومة وقابلية التعافي

٣. البنية المرجعية

يعرض الشكل ١ البنية المرجعية الكاملة. وتمر الطلبات عبر طبقة التجربة والاستقبال إلى بيئة تشغيل الوكلاء، التي تصل إلى طبقة الأدوات والبيانات والمعرفة. وتتولى بوابة النماذج حوكمة الوصول إلى الاستدلال. تنطبق الهوية والوصول، والثقة والسلامة، وقابلية المراقبة والتقييم، والحوكمة والامتثال عبر الطبقات. ويربط MCP الوكلاء بالأدوات والسياق، بينما يربط A2A بين الوكلاء المستقلين. وتربط حلقة التحسين التغذية الراجعة التشغيلية بالإصدارات الخاضعة للتقييم. تتناول الأقسام من ٤ إلى ١٠ مسؤوليات المكونات وضوابطها. ويربط القسم ١١ القدرات بخيارات تنفيذ تمثيلية.

تطبيق البنية

النشر الأول. يعرض الشكل ١ المجموعة الكاملة من القدرات المتاحة. وتحتاج حالة الاستخدام الأولى إلى ملكية خاضعة للمساءلة، ووصول محدد النطاق، واستخدام منضبط للنماذج والأدوات، وتقييم، وإشراف بشري مناسب، وسجلات للتنفيذ، ومعالجة محددة للأعطال. وقد توفر المنصة المختارة هذه الضوابط. ويكون البدء بأبسط سير عمل أو وكيل فعال؛ إذ لا يلزم إنشاء خدمة منفصلة لكل مربع.

القدرات الإضافية. تُضاف بوابات النماذج أو بوابات MCP المشتركة عندما تبرر ذلك الحاجة إلى إدارة مركزية للوصول وإعادة الاستخدام. ويُستعان بوكلاء متعددين، أو بالربط الاتحادي عبر A2A، أو بالذاكرة طويلة المدى، أو بالضبط الدقيق عندما تتطلب المهمة هذه القدرات. ويلزم التنفيذ الدائم منذ البداية عندما يتعين أن يصمد العمل أمام الأعطال أو فترات الانتظار الطويلة. ويتبع اختيار القدرات حالة الاستخدام، مع بقاء الضوابط الأساسية قائمة في كل مرحلة.

الشكل ١ - البنية المرجعية للذكاء الاصطناعي الوكيل



٤. نظرة عامة على المكونات

تحدد أوصاف المكونات أدناه المسؤوليات. وتقدم الأقسام ٧ و ٨ و ٩ الإرشادات التفصيلية الخاصة بالأمن والتقييم والحوكمة.

٤.١ طبقة التجربة والاستقبال

تستقبل طبقة التجربة والاستقبال طلبات المستخدمين وأحداث الأنظمة واستدعاءات API. وتتولى مصادقة الجهات المستدعية، والتحقق من صحة المدخلات، وإدارة الجلسات، وتقديم الاستجابات. وتشمل المسؤوليات أدناه أيضًا جمع التغذية الراجعة.

- القنوات: واجهات الويب والأجهزة المحمولة والإدارة؛ وأدوات المحادثة والتعاون؛ وواجهات API وخطافات الويب؛ ومشغلات الأحداث من قوائم الانتظار وتدفقات البيانات للوكلاء العاملين في الخلفية والمدفوعين بالأحداث.
- مصادقة المستخدمين وإدارة الجلسات، مع تمرير هوية المستخدم إلى بيئة التشغيل بحيث يتمكن الوكلاء من العمل نيابةً عن المستخدم وبصلاحياته.
- التحقق من صحة المدخلات وتوحيد صيغتها، مع تطبيق الخط الأول من حواجز الحماية الخاصة بالثقة والسلامة.
- تقديم الاستجابات بالبحث المتدفق، مع الإشارة إلى المصادر، والإفصاح بوضوح عن أن المحتوى مُولّد بالذكاء الاصطناعي.

٤.٢ بيئة تشغيل الوكلاء: التنسيق والتحكم

- يتولى التنسيق اختيار سير العمل أو نمط الوكيل، وتجميع السياق، وإدارة التنفيذ. كما يفرض الصلاحيات والميزانيات والموافقات المطلوبة، مع حفظ الحالة بشكل دائم حيثما يتعين أن يصمد العمل أمام الأعطال أو فترات الانتظار الطويلة.
- مخطّط أو مشرف يُجزئ الأهداف ويختار النمط المناسب (القسم ٥).
- محرك سير عمل دائم يوفر نقاط الحفظ، وإعادة المحاولة، والمهل الزمنية، والتكرار الآمن، والإجراءات التعويضية للأعمال طويلة الأمد أو متعددة الخطوات.
- هندسة السياق: تجميع التعليمات المناسبة، والمعرفة المسترجعة، والذاكرة، وتعريفات الأدوات ضمن ميزانية مضبوطة من الرموز.
- الحالة والذاكرة، مقسّمةً إلى ذاكرة عاملة (المهمة الحالية)، وذاكرة الجلسة (المحادثة)، وذاكرة طويلة المدى (عرضية ودلالية)، مع توثيق المصدر والاحتفاظ بالبيانات وعزل المستأجرين لكلٍّ منها.
- بوابات التدخل البشري، ومستويات الاستقلالية، وميزانيات التكلفة والخطوات والوقت.
- نقطة إنفاذ السياسات التي تُجيز كل استدعاء لأداة بناءً على هوية الوكيل ونطاق الصلاحيات المفوّض من المستخدم والسياسات.
- سجل الوكلاء الذي يحتفظ ببطاقة وكيل لكلٍّ منهم تتضمن: الغرض، والمالك، ومستوى الاستقلالية، والأدوات ونطاقات الصلاحيات، والنماذج، وتصنيفات البيانات.

٤.٣ بيئة تشغيل الوكلاء: الوكلاء

- يُنظّم الوكلاء حول مجالات الأعمال والأنماط الواردة في القسم ٥. ويُعد الاسترجاع والتحليل وتنفيذ الأدوات قدرات مشتركة، ويُستعان بوكلاء منفصلين عندما يبرر توزيع المسؤوليات عمليات التسليم الإضافية.
- وكلاء التوجيه الذين يصنّفون الطلبات ويحيلونها إلى سير العمل أو المتخصص المناسب.
- ترتيبات المنسّق والعاملين، حيث يُسند المشرف مهامًا فرعية تُحدّد ديناميكيًا إلى العاملين.
- وكلاء متخصصون في المجالات ومتوائمون مع قدرة أعمال معينة (مثل المطالبات أو التسويات أو المشتريات)، ولكلٍّ منهم مجموعة أدوات محدودة.
- وكلاء مقيّمون أو ناقدون يتحققون من المخرجات وفق معايير محددة قبل إصدارها.
- خطوات سير العمل الحتمي، التي تظل الخيار المفضل حيثما يكون تسلسل الخطوات معروفًا مسبقًا.

٤,٤ بوابة النماذج والنماذج

- تتولى بوابة النماذج حوكمة الوصول والتوجيه والاستخدام والواجهات الخاصة بكل مزود. وقد توفر المنصة المختارة قدراتها، أو قد تُنفَّذ بوصفها خدمة مشتركة. ويُختار ترتيب النشر وفقاً لحالة الاستخدام والبنية التحتية القائمة.
- التوجيه حسب المهمة والقدرة والتكلفة وزمن الاستجابة، مع التحويل الاحتياطي التلقائي بين المزودين أو المناطق.
 - ميزانيات الرموز، وحدود معدل الطلبات، وقياس الاستخدام، مع إسناد التكلفة لكل وكيل وحالة استخدام ومستأجر.
 - التخزين المؤقت الدلالي والتخزين المؤقت للأوامر النصية لخفض التكلفة وزمن الاستجابة.
 - سجل النماذج وتثبيت الإصدارات، بحيث تكون تغييرات النماذج إصدارات مقصودة وخاضعة للتقييم.
 - دعم النماذج الرائدة، والنماذج مفتوحة الأوزان، والنماذج المضبوطة بدقة، والنماذج الصغيرة، إضافةً إلى نماذج التضمين وإعادة الترتيب.
 - نقاط الاتصال الخاصة وتثبيت المناطق لتلبية متطلبات توطين البيانات وسيادتها.

٤,٥ طبقة الأدوات والبيانات والمعرفة

- توفر هذه الطبقة الوصول إلى الأنظمة المؤسسية والأدوات المعتمدة والمعرفة المرجعية. ويُعد MCP واجهة الأدوات الرئيسية في البنية المرجعية، بينما يدعم A2A التعاون مع الوكلاء الخارجيين عند الحاجة.
- بوابة MCP وسجل الأدوات: كتالوج يضم خوادم MCP والأدوات المعتمدة المدرجة في قائمة السماح، مع تثبيت الإصدارات والمصادقة ونطاقات صلاحيات لكل أداة.
 - واجهات API المؤسسية وأنظمة السجلات الأساسية، مع تصميم عمليات الكتابة إليها بحيث تكون آمنة عند التكرار وقابلة للتراجع حينما أمكن.
 - تنفيذ الشيفرات البرمجية وأتمتة المتصفح في بيئات معزولة ومؤقتة مع تقييد الاتصالات الشبكية الصادرة.
 - موصّلات الأتمتة وأتمتة العمليات الروبوتية ومنصات التكامل كخدمة للعمليات القديمة.
 - المعرفة والاسترجاع: بحث هجين (بالكلمات المفتاحية والمتجهات) مع إعادة الترتيب، والتصنيف حسب البيانات الوصفية، والتحكم في الوصول على مستوى المستندات.
 - البيانات المؤسسية والطبقة الدلالية، بحيث يستعلم الوكلاء عن مقاييس أعمال خاضعة للحوكمة ومحددة بوضوح بدلاً من الجداول الأولية.
 - الربط الاتحادي مع الوكلاء الخارجيين ووكلاء الشركاء عبر A2A، مع استخدام بطاقات الوكلاء للاكتشاف والمصادقة.

٤,٦ مستوى الهوية والوصول

- يُسند هذا المستوى الإجراءات إلى الوكيل والمالك، وإلى المستخدم مقدّم الطلب عند الاقتضاء. كما يفرض الصلاحيات والوصول المطلوبين لكل مهمة.
- هوية عبء عمل مستقلة لكل وكيل، وعدم استخدام حسابات خدمة مشتركة مطلقاً.
 - تفويض الصلاحيات باستخدام تبادل رموز الوصول نيابةً عن المستخدم، بحيث لا يرث الوكلاء إلا صلاحيات المستخدم مقدّم الطلب، ولمدة المهمة فقط.
 - نطاقات صلاحيات لكل أداة وفق مبدأ الحد الأدنى من الصلاحيات، تفرضها نقطة إنفاذ السياسات، مع صياغة السياسات على هيئة شيفرة برمجية.
 - إدارة الأسرار والمفاتيح، دون كشف أي بيانات اعتماد في الأوامر النصية أو السياق أو البيئات المعزولة.
 - اتصالات مُصادَق عليها وموقَّعة بين الوكلاء.

٤,٧ مستوى الثقة والسلامة

- تُطبّق فحوصات السلامة عند المدخلات والمخرجات وحدود الإجراءات. ويُتعامَل مع المحتوى الوارد من المستندات والأدوات والوكلاء الآخرين على أنه غير موثوق، وتُفرض الإجراءات المسموح بها خارج النموذج.
- حواجز حماية للمدخلات والمخرجات تغطي سياسة المحتوى وحدود الموضوعات والاستناد إلى المصادر.
 - الحماية من حقن الأوامر النصية، بما في ذلك عزل المحتوى غير الموثوق ووسمه، وتقييد الأدوات التي يمكن استدعاؤها بعد قراءة محتوى غير موثوق.
 - التحقق من استدعاءات الأدوات: فحص المخطط والمعاملات، وحدود القيم، وقوائم السماح قبل التنفيذ.
 - اكتشاف البيانات الشخصية وحجبها، ومنع تسرب البيانات في المدخلات والسياق والمخرجات والسجلات.
 - قواطع الدائرة، وحدود معدل الطلبات، والحدود القصوى لنطاق التأثير لاحتواء الأعطال المتتالية.

٤,٨ مستوى قابلية المراقبة والتقييم

- تُجمع بيانات القياس عن بُعد من مختلف المكونات لإعادة بناء عمليات التنفيذ، ومراقبة جودة الخدمة، وإسناد التكلفة. ويصف القسم ٨ دورة حياة التقييم وضوابط الإصدار.
- تتبع الموزع باستخدام الاصطلاحات الدلالية لـ OpenTelemetry الخاصة بالذكاء الاصطناعي التوليدي، بما يشمل الخطط واستدعاءات النماذج واستدعاءات الأدوات وعمليات التسليم.
 - التقييم المباشر لحركة الإنتاج من حيث الجودة والاستناد إلى المصادر والامتثال للسياسات، إلى جانب حزم التقييم المسبق (القسم ٨).
 - مقاييس التكلفة وزمن الاستجابة والموثوقية، مع أهداف مستوى الخدمة لكل وكيل وحالة استخدام.
 - سجلات تدقيق غير قابلة للتعديل للقرارات والموافقات والإجراءات، يُحتفظ بها وفقًا للمتطلبات التنظيمية.
 - خطوط أساس سلوكية وتنبهات بالحالات الشاذة لرصد الانحراف أو السلوك الخارج عن السيطرة.

٤,٩ مستوى الحوكمة والامتثال

- تُدار ملكية الوكلاء وتصنيفات المخاطر والتقييمات وموافقات النشر. وتُربط هذه السجلات بمتطلبات أنظمة الإدارة والمتطلبات التنظيمية المعمول بها الواردة في القسم ٩.
- سجل جرد الذكاء الاصطناعي وسجل الوكلاء، مع ربطهما بالمالكين وفئة المخاطر ومستوى الاستقلالية.
 - تقييمات المخاطر والأثر لكل وكيل ولكل تغيير جوهري.
 - ربط السياسات والمتطلبات التنظيمية بمعيار ISO/IEC ٤٢٠٠١، وإطار NIST AI RMF، وقانون الاتحاد الأوروبي للذكاء الاصطناعي، والمتطلبات القطاعية.
 - ضوابط تشغيلية تشمل مفتاح إيقاف طارئ لكل وكيل، وإدارة الحوادث، وإجراءً محددًا للإخراج من الخدمة.

٤,١٠ حلقة التحسين المستمر

- تُستخدم التغذية الراجعة ونتائج التقييم والإشارات التشغيلية لاقتراح تغييرات على الأوامر النصية أو الأدوات أو السياسات أو الاسترجاع أو النماذج. ويُمنح كل تغيير مرشّح رقم إصدار ويُقيّم، وتُخضع التغييرات الجوهرية لاختبارات الفريق الأحمر، وتُراقب الإصدارات المرحلية تحسبًا لتراجع الأداء والحاجة إلى الرجوع إلى إصدار سابق. ويمكن اللجوء إلى الضبط الدقيق أو التقطير عند توفر بيانات كافية عالية الجودة وحاجة مُثبتة.

٥. أنماط تصميم الوكلاء ومستويات الاستقلالية

يكون البدء بأبسط نمط يلبي متطلبات الأعمال. ويُلبّج إلى تخطيط أكثر مرونة أو إلى وكلاء إضافيين عندما يُظهر التقييم فائدة تبرر ما يترتب على ذلك من تنسيق وتكلفة ومخاطر تشغيلية إضافية.

١,٥ دليل اختيار الأنماط

النمط	يُستخدم عندما	مثال نموذجي
سير العمل الحتمي (تسلسل الأوامر النصية)	تكون الخطوات معروفة مسبقًا ونادرًا ما يتغير تسلسلها.	استخراج الفواتير والتحقق منها وترجيلها.
التوجيه	تندرج الطلبات ضمن فئات متميزة تحتاج إلى معالجة أو نماذج أو أدوات مختلفة.	فرز طلبات مكتب الخدمة عبر تقنية المعلومات والموارد البشرية والمالية.
التوازي	تكون المهام الفرعية مستقلة، أو تُعزّز وجهات النظر المتعددة مستوى الثقة.	مراجعة عقد مقابل عدة قوائم تحقق خاصة بالسياسات في آنٍ واحد.
المنسّق والعاملون	لا يمكن التنبؤ بالمهام الفرعية مسبقًا، ويجب تحديدها ديناميكيًا.	بحث وتحليل من مصادر متعددة لإعداد مذكرة استثمارية.
المقيّم والمحسّن	تتوفر معايير جودة واضحة، ويؤدي التحسين التكراري إلى تحسين المخرجات بشكل قابل للقياس.	صياغة رد على طلب عروض يُقيّم وفق سلّم تقييم.
الوكيل المستقل	تكون المشكلة مفتوحة النهاية ولا يمكن التنبؤ بعدد الخطوات.	التحقيق في حادثة تشغيلية وتشخيصها.
الربط الاتحادي متعدد الوكلاء (A2A)	تكون القدرات مملوكة لفرق أو مجالات أو مؤسسات مختلفة.	وكيل مشتريات يتفاوض مع وكيل أحد الموردين.

٥.٢ مستويات الاستقلالية

تُحدّد الاستقلالية لكل نوع من الإجراءات، لا لكل وكيل. فقد يقرأ الوكيل نفسه البيانات باستقلالية (L٣)، لكنه يحتاج إلى موافقة قبل تنفيذ عملية دفع (L١). وينبغي أن تستند الترقية إلى مستوى أعلى إلى الأدلة: نتائج التقييم، ومعدلات التجاوز البشري، وسجل التشغيل.

المستوى	الوصف	الإشراف البشري	الإجراءات النموذجية
L٠ - المساعدة	يحلل الوكيل ويُعدّ المسودات ويقدم التوصيات، بينما يتولى الإنسان تنفيذ كل إجراء.	يتخذ الإنسان القرار وينفذه.	الصياغة، والبحث، والتلخيص.
L١ - تحت الإشراف	يُعدّ الوكيل الإجراءات، ويوافق الإنسان على كل إجراء ذي أثر جوهري قبل تنفيذه.	موافقة مسبقة على كل إجراء.	المدفوعات، والتواصل مع العملاء، وتعديل السجلات.
L٢ - ضمن حدود	يعمل الوكيل ضمن حدود معتمدة مسبقاً، مثل حدود القيمة ونطاقات الصلاحيات، وتُصعّد الاستثناءات.	معالجة الاستثناءات والمراجعة اللاحقة بالعينات.	التسويات الروتينية، والمبالغ المستردة منخفضة القيمة.
L٣ - مستقل ضمن المجال	يعمل الوكيل من البداية إلى النهاية في مجال محدود خضع لتقييم وافي.	المراقبة والتنبيه ومفتاح الإيقاف الطارئ.	التحليل بصلاحيّة القراءة فقط، وتوجيه التذاكر الداخلية.

الترقية والخفض بين المستويات

يُعدّ نقل إجراء ما إلى مستوى استقلالية أعلى تغييراً خاضعاً للحكمة، ويُسجّل في سجل الوكلاء. والحدود الواردة أدناه توضيحية، وينبغي تحديدها وفقاً لمخاطر كل إجراء وإمكانية التراجع عنه.

- الترقية لا تكون إلا عند إتمام عدد معتبر من الإجراءات (على سبيل المثال، عدة مئات) في المستوى الحالي، وبقاء معدلات التجاوز والرفض من قبل البشر دون حد متفق عليه خلال تلك الفترة، وعدم إظهار نتائج التقييم واختبارات الفريق الأحمر أي ملاحظات مفتوحة تتعلق بالسلامة أو السياسات، واعتماد مالك الأعمال ووظيفة المخاطر للتغيير.
- الخفض التلقائي بعد وقوع حادثة جوهريّة، أو عندما تخرج معدلات التجاوز أو درجات التقييم عن الحدود المقررة لها، أو عندما يطرأ تغيير جوهري على النموذج أو الأدوات أو السياسات التي يستند إليها الإجراء. ويعود الإجراء إلى المستوى الأدنى إلى أن تُعاد عملية تقييمه.

٦. مسار الطلب من البداية إلى النهاية

يوضح التسلسل التالي كيف ينتقل طلب واحد عبر البنية، وأين يُطبَّق كل ضابط.

١. **الاستقبال.** يصل طلب من المستخدم أو حدث ما، وتتم مصادقة الجهة المستدعية، وتفحص حواجز حماية المدخلات المحتوى بحثًا عن حقن الأوامر النصية والبيانات الحساسة ومخالفات السياسات.
٢. **تجميع السياق.** تُحمَّل بيئة التشغيل حالة الجلسة والذاكرة ذات الصلة، وتسترجع المعرفة المرجعية ضمن ميزانية مضبوطة من الرموز.
٣. **التخطيط والتوجيه.** يختار الموجّه أو المشرف النمط والوكلاء، ويجري التحقق من الخطة وفق السياسات ومستوى الاستقلالية المعمول به.
٤. **استدلال النموذج.** تمر استدعاءات النماذج عبر البوابة، التي تطبق التوجيه والميزانيات والتخزين المؤقت والتسجيل.
٥. **استدعاء الأدوات.** يطلب الوكيل أداةً عبر بوابة MCP. وتتحقق نقطة إنفاذ السياسات من هوية الوكيل، ونطاق الصلاحيات المفوض من المستخدم، ومعاملات الاستدعاء. وتُوسَم مخرجات الأدوات على أنها غير موثوقة.
٦. **نقطة التحقق البشري.** إذا تجاوز إجراء ما مستوى استقلالية الوكيل، يتوقف التنفيذ مؤقتًا مع حفظ حالته، ويُرفع طلب موافقة مرفقًا بالأدلة اللازمة لاتخاذ القرار.
٧. **التنفيذ والكتابة إلى الأنظمة.** تُنفَّذ الإجراءات المعتمدة بطريقة آمنة عند التكرار، مع تحديد خطوات تعويضية في حال حدوث عطل في المراحل اللاحقة.
٨. **التحقق من المخرجات.** يجري التحقق من الاستجابة من حيث الاستناد إلى المصادر والامتثال للسياسات والبيانات الحساسة، ثم تُقدَّم مع الإشارة إلى المصادر والإفصاح عن استخدام الذكاء الاصطناعي.
٩. **القياس عن بُعد والتدقيق.** يُصدَّر المسار الكامل على هيئة سجل تتبع، وتُسند التكاليف، وتُدوّن القرارات والموافقات في سجل التدقيق.
١٠. **التعلّم.** تنتقل التغذية الراجعة من المستخدمين ودرجات التقييم المباشر إلى حلقة التحسين المستمر.

ملاحظة تصميمية: تتركز معظم المخاطر الخاصة بالوكلاء في الخطوتين ٥ و ٦. ولا ينبغي أبدًا تنفيذ استدعاء أداة استنادًا إلى مخرجات النموذج وحدها؛ إذ يجري التحقق من الهوية والنطاق والمعاملات ومستوى الاستقلالية جميعها خارج النموذج.

٧. البنية الأمنية

تُستخدم قائمة OWASP Top ١٠ للتطبيقات الوكيلية (٢٠٢٦) إلى جانب التصنيف الخاص بتطبيقات النماذج اللغوية الكبيرة لنمذجة تهديدات النظام. ويربط الجدول المخاطر الخاصة بالوكلاء بالضوابط الموضحة في هذه البنية.

المعرّف	الخطر	الضوابط المعمارية الرئيسية	الموقع في البنية
ASI٠١	اختطاف أهداف الوكيل	عزل المحتوى غير الموثوق ووسمه؛ التحقق من الخطط؛ حدود قصوى للخطوات والحلقات؛ نقاط موافقة على الإجراءات غير القابلة للتراجع.	الثقة والسلامة؛ التنسيق
ASI٠٢	إساءة استخدام الأدوات	نطاقات صلاحيات لكل أداة؛ التحقق من المعاملات؛ حدود القيم؛ تصميم أدوات آمنة عند التكرار وقابلة للتراجع.	نقطة إنفاذ السياسات؛ بوابة MCP
ASI٠٣	إساءة استخدام الهوية	هويات مستقلة للوكلاء؛ التفويض بالنيابة؛ رموز وصول قصيرة الأجل؛ عدم استخدام بيانات اعتماد مشتركة.	الهوية والوصول
ASI٠٤	ثغرات سلسلة التوريد	خوادم MCP مدرجة في قائمة السماح؛ تثبيت الإصدارات وتوقيعها؛ مراجعة الأدوات والأوامر النصية والوكلاء التابعة لأطراف خارجية عند كل ترقية.	بوابة MCP والسجل؛ الحوكمة
ASI٠٥	تنفيذ شيفرات غير متوقع	بيئات معزولة ومؤقتة؛ تقييد الاتصالات الشبكية الصادرة؛ عدم وجود بيانات اعتماد في بيئات التنفيذ.	طبقة الأدوات (البيئة المعزولة)
ASI٠٦	تسميم الذاكرة والسياق	توثيق مصدر عمليات الكتابة في الذاكرة؛ عزل المستأجرين؛ حدود الاحتفاظ بالبيانات؛ التحقق الدوري من الذاكرة طويلة المدى.	الحالة والذاكرة؛ الثقة والسلامة
ASI٠٧	الاتصال غير الآمن بين الوكلاء	اتصال A٢A مُصادق عليه مع رسائل موقّعة، والحماية من هجمات إعادة الإرسال، وسياسات صريحة للتواصل بين الوكلاء.	الهوية والوصول؛ حدود A٢A
ASI٠٨	الأعطال المتتالية	قواطع الدائرة؛ حدود معدل الطلبات عند نقطة ربط كل وكيل؛ حدود قصوى لنطاق التأثير؛ مراقبة التفرع.	التنسيق؛ قابلية المراقبة
ASI٠٩	استغلال الثقة بين الإنسان والوكيل	الإحالة إلى المصادر؛ الإشارة إلى مستوى الثقة؛ الإفصاح عن استخدام الذكاء الاصطناعي؛ فرض خطوات تأكيد إضافية على الموافقات عالية الأثر.	طبقة التجربة؛ بوابات التدخل البشري
ASI١٠	الوكلاء الخارجون عن السيطرة	خطوط أساس سلوكية؛ تنبيهات بالحالات الشاذة؛ مفتاح إيقاف طارئ لكل وكيل؛ إجراء محدد للإخراج من الخدمة.	قابلية المراقبة؛ الحوكمة

الممارسات الأمنية الأساسية

- نمذجة التهديدات لكل وكيل قبل نشره، بما يشمل أدواته وبياناته وذاكرته والوكلاء الذين يتعامل معهم.
- إخضاع كل إصدار جوهري لاختبارات الفريق الأحمر وفق فئات OWASP، بالجمع بين الأدوات المؤتمتة واختبارات الخبراء.
- التعامل مع خوادم MCP وأطر الوكلاء ووكلاء الأطراف الخارجية بوصفها اعتماديات ضمن سلسلة التوريد، وتطبيق دورة حياة التطوير الآمن ذاتها على الأوامر النصية ومخططات الأدوات وإعدادات الوكلاء كما تُطبّق على شيفرة التطبيقات.

٨. التقييم وعمليات تشغيل الوكلاء

تُقيّم نتائج المهام ومسارات التنفيذ قبل الإصدار وأثناء التشغيل في بيئة الإنتاج. وتحدد دورة الحياة والأبعاد الواردة أدناه النطاق.

٨.١ دورة الحياة

يمر الوكلاء بدورة حياة منضبطة: التصميم، والبناء، والتقييم المسبق، واختبارات الفريق الأحمر، والإصدار المرحلي عبر النشر الموازي والنشر التجريبي المحدود، والمراقبة في بيئة الإنتاج مع التقييم المباشر، والتحسين. ويخضع كل انتقال بين المراحل لحدود تقييم محددة، ويمكن التراجع عن أي تغيير في بيئة الإنتاج.

٨.٢ أبعاد التقييم

البُعد	ما يقيسه	أمثلة على الأساليب
نجاح المهمة	ما إذا كان الوكيل قد حقق نتيجة الأعمال المقصودة.	مجموعات بيانات مرجعية بنتائج متوقعة؛ تتبع مؤشرات الأداء الرئيسية للأعمال.
جودة المسار	ما إذا كانت الخطوات المتخذة منطقية وفعّالة وضمن السياسات.	المراجعة على مستوى سجلات التتبع؛ عدد الخطوات؛ نموذج لغوي كبير بدور الحَكَم مُعَايَر وفق تصنيفات بشرية.
دقة استدعاء الأدوات	صحة اختيار الأداة وصحة المعاملات.	تأكيدات حتمية على استدعاءات الأدوات ضمن سيناريوهات الاختبار.
الاستناد إلى المصادر	ما إذا كانت الادعاءات مدعومة بالمصادر المسترجعة.	فحوصات الأمانة للمصادر والاستشهادات.
السلامة والسياسات	الالتزام بسياسات المحتوى والبيانات والإجراءات.	حزم اختبار السياسات؛ معدلات تفعيل حواجز الحماية.
المتانة	مقاومة حقن الأوامر النصية واختطاف الأهداف والمدخلات العدائية.	اختبارات فريق أحمر مؤتمتة مرتبطة بفئات OWASP.
زمن الاستجابة والتكلفة	الوقت والتكلفة لكل مهمة مكتملة.	المئينات المستخلصة من سجلات التتبع؛ تكلفة المهمة الواحدة حسب الوكيل والنموذج.
إشارات الإشراف البشري	مدى تكرار تجاوز البشر لإجراءات الوكلاء أو رفضها أو تصعيدها.	معدلات الموافقة والتجاوز والتصعيد حسب مستوى الاستقلالية.

٨.٣ إدارة الإصدارات

تُدار إصدارات الأوامر النصية ومخططات الأدوات وإعدادات الوكلاء وسياسات حواجز الحماية وإعدادات الاسترجاع وإصدارات النماذج بوصفها وحدة إصدار واحدة. وتُقيّم التغييرات في سياقها، ويُحتفظ بإعدادات سليمة مُجرّبة للرجوع إليها عند التراجع. ويُحافظ على حزم التقييم إلى جانب شيفرة الاختبار، وتُوسّع عندما تكشف حوادث الإنتاج عن ثغرات.

٩. الحوكمة والمواءمة التنظيمية

يربط الجدول الأطر ذات الصلة، كما هي في سبتمبر ٢٠٢٦، بالأدلة التي تُنتجها البنية. ويتعين التحقق من مدى انطباقها على كل حالة استخدام وكل مؤسسة.

الإطار	وجه الصلة	الأدلة المستمدة من البنية
42001:2023 ISO/IEC	معيار قابل للإشهاد لأنظمة إدارة الذكاء الاصطناعي، يغطي السياسات والمخاطر ودورة الحياة وضوابط الموردين.	سجل جرد الذكاء الاصطناعي، وتقييمات المخاطر، وضوابط دورة الحياة والتغيير، ومراجعات الموردين.
23894:2023 ISO/IEC و 42005:2025 ISO/IEC	إرشادات بشأن إدارة مخاطر الذكاء الاصطناعي وتقييم أثر أنظمة الذكاء الاصطناعي.	سجل المخاطر وتقييم الأثر لكل وكيل ولكل تغيير جوهري.
إطار NIST AI RMF ١.٠ وملف الذكاء الاصطناعي التوليدي (NIST AI ١-٦٠٠)	إطار طوعي منظم حول الحوكمة والتحديد والقياس والإدارة.	مقاييس التقييم، ومعالجات المخاطر، وسجلات الحوادث.
قانون الاتحاد الأوروبي للذكاء الاصطناعي (لائحة الاتحاد الأوروبي رقم ١٦٨٩/٢٠٢٤)	التزامات قائمة على المخاطر. وفي أعقاب الحزمة الرقمية الشاملة للذكاء الاصطناعي لعام ٢٠٢٦، تسري التزامات الأنظمة عالية المخاطر القائمة بذاتها الواردة في الملحق الثالث اعتبارًا من ٢ ديسمبر ٢٠٢٧، والتزامات الأنظمة المدمجة في المنتجات الواردة في الملحق الأول اعتبارًا من ٢ أغسطس ٢٠٢٨، في حين تسري التزامات نماذج الذكاء الاصطناعي للأغراض العامة منذ أغسطس ٢٠٢٥.	إدارة المخاطر (المادة ٩)، وحفظ السجلات (المادة ١٢)، والإشراف البشري (المادة ١٤)، والدقة والمتانة والأمن السيبراني (المادة ١٥).
حماية البيانات والتنظيمات القطاعية في المملكة المتحدة	اللائحة العامة لحماية البيانات في المملكة المتحدة وقانون حماية البيانات لعام ٢٠١٨، بصيغتهما المعدلة بموجب قانون البيانات (الاستخدام والوصول) لعام ٢٠٢٥؛ وإرشادات مكتب مفوض المعلومات بشأن الذكاء الاصطناعي؛ والتوقعات القطاعية مثل ٢٣/SS١ الصادر عن هيئة التنظيم الاحترازي بشأن إدارة مخاطر النماذج.	تقييمات أثر حماية البيانات، والأساس النظامي، وضمانات اتخاذ القرارات الآلية، وسجل جرد مخاطر النماذج والتحقق منها.
قائمة OWASP Top ١٠ لتطبيقات النماذج اللغوية الكبيرة (٢٠٢٥) وللتطبيقات الوكيلية (٢٠٢٦)	تصنيفات مجتمعية للمخاطر الأمنية في أنظمة النماذج اللغوية الكبيرة والأنظمة الوكيلية.	نماذج التهديدات، ونطاق اختبارات الفريق الأحمر، وربط الضوابط (القسم ٧).

يلخص هذا القسم المشهد التنظيمي لأغراض معمارية، ولا يُعد استشارة قانونية. وينبغي للمؤسسات التحقق من مدى الانطباق والالتزامات المترتبة عليها مع مستشاريها القانونيين ومستشاري الامتثال.

الحد الأدنى لسجل الحوكمة لكل وكيل

- الغرض، ومالك الأعمال، والمالك التقني.
- فئة المخاطر ومستوى الاستقلالية لكل نوع من الإجراءات.
- الأدوات ونطاقات الصلاحيات وتصنيفات البيانات التي يمكن للوكيل الوصول إليها.
- النماذج المستخدمة، مع إصداراتها المثبتة.
- أحدث تقرير تقييم وملاحظات اختبارات الفريق الأحمر.
- تقييم الأثر وسجل اعتماد النشر.

الأدوار والمسائلة

- إن الملكية الواضحة هي ما يجعل الضوابط المذكورة أعلاه فعالة على أرض الواقع. ويُسند نموذج التشغيل المعتمد المسؤوليات التالية:
- **مالك الأعمال:** يتحمل المسائلة عن نتائج الوكيل في مجاله، ويقبل المخاطر المتبقية، ويعتمد التغييرات على مستويات الاستقلالية.
- **فريق المنتج أو التنفيذ:** يبني الوكيل ويقيمه ويشغله، ويحافظ على حزم التقييم وأدلة التشغيل وسجل الحوكمة الخاصة به.
- **فريق المنصة:** يوفر الخدمات المشتركة التي يرثها كل وكيل، مثل بوابات النماذج وبوابات MCP، وهوية الوكلاء، وقابلية المراقبة، ومنظومة التقييم.
- **المخاطر والأمن والامتثال:** تضع السياسات، وتراجع تقييمات الأثر وملاحظات اختبارات الفريق الأحمر، وتعتمد عمليات النشر عالية المخاطر والترقيات إلى ما بعد L1.
- **منتدى حوكمة الذكاء الاصطناعي أو مركز التميز:** يحافظ على سجل جرد الذكاء الاصطناعي والمعايير والأنماط المعتمدة، ويحسم القرارات المشتركة بين المجالات. كما يحدد من يحق له تفعيل مفتاح الإيقاف الطارئ للوكيل، وكيفية التحقيق في الحوادث المتعلقة بالوكلاء.

الوكلاء الذين لا تطورهم المؤسسة

سيُشتري كثير من الوكلاء في المؤسسة بدلاً من بنائهم: المساعدات المدمجة في منصات البرمجيات كخدمة، والمساعدات الذكية للإنتاجية، ووكلاء البرمجة أو وكلاء سطح المكتب. وقد تقع بيئة التشغيل خارج سيطرة المؤسسة، لكن المبادئ ذاتها تظل منطبقة. ويُسجل هؤلاء الوكلاء في سجل جرد الذكاء الاصطناعي؛ ويُقيّد وصولهم عبر ضوابط الهوية والإدارة، مثل منح OAuth محددة النطاق، وقوائم السماح للموصلات، وحدود البيانات؛ ويُشترط تقديم أدلة من الموردين بشأن التسجيل واستخدام البيانات والإشعار بتغييرات النماذج؛ وتُطبّق مستويات الاستقلالية الواردة في القسم 5 على الإجراءات التي يمكنهم تنفيذها على الأنظمة المؤسسية.

١٠. الاعتبارات غير الوظيفية

ينبغي تحديد المتطلبات غير الوظيفية لكل حالة استخدام. ويعرض الجدول أدناه الروافع المعمارية المتاحة لكل منها.

المجال	الاعتبارات المعمارية
زمن الاستجابة وتجربة المستخدم	بث الاستجابات تدقيقًا؛ وتشغيل استدعاءات الأدوات المستقلة بالتوازي؛ وتوجيه الخطوات البسيطة إلى نماذج أصغر؛ والتوسع في التخزين المؤقت؛ وتحديد ميزانيات للخطوات والوقت.
التكلفة	فرض ميزانيات للرموز لكل مهمة؛ والتوجيه حسب التكلفة والقدرة؛ وإسناد التكلفة لكل وكيل وحالة استخدام ومستأجر؛ وتتبع تكلفة الوحدة لكل مهمة مكتملة، لا لكل استدعاء.
الموثوقية	استخدام التنفيذ الدائم مع نقاط الحفظ؛ وإعادة المحاولة مع فواصل زمنية متزايدة؛ وأدوات آمنة عند التكرار؛ والتحويل الاحتياطي بين النماذج والمناطق؛ وقواطع الدائرة على الاعتماديات.
قابلية التوسع	إبقاء عمليات المعالجة في بيئة التشغيل عديمة الحالة؛ واستخدام التنفيذ القائم على قوائم الانتظار للأعمال طويلة الأمد؛ وتوسيع الأدوات والاسترجاع بشكل مستقل عن الكلاء.
توطين البيانات وسيادتها	تثبيت النماذج والتخزين في مناطق معتمدة؛ واستخدام نقاط الاتصال الخاصة؛ والتوجيه حسب تصنيف البيانات؛ وإبقاء البيانات الحساسة خارج الأوامر النصية حيثما أمكن.
تعدد المستأجرين والعزل	عزل الذاكرة وفهارس المتجهات والذاكرات المؤقتة والسجلات لكل مستأجر؛ واستخدام مفاتيح خاصة بكل مستأجر؛ وفرض سياق المستأجر في كل استدعاء لأداة.
قابلية النقل	استخدام البروتوكولات المفتوحة عند نقاط الربط؛ وتجريد النماذج خلف البوابة؛ وتجنب صيغ الحالة الحصرية؛ والاحتفاظ بالأوامر النصية ومخططات الأدوات ضمن نظام التحكم في الإصدارات.

١١. مشهد الأدوات والتقنيات

الخيارات الواردة أدناه تمثيلية، ولا تُعد تزكيةً لها أو إلزامًا بتبني كل قدرة. وتُختار المنتجات وفقًا للضوابط التي تحتاجها حالة الاستخدام وللمنصة القائمة لدى المؤسسة. وتُقيّم عائلات النماذج وفق متطلبات المهام، إذ تتغير الإصدارات المحددة بشكل متكرر.

القدرة	المصادر المفتوحة والمعايير المفتوحة	حزم تطوير البرمجيات من الموردين والحلول المُدارة والتجارية
أطر الوكلاء والتنسيق	LlamaIndex، LangGraph، CrewAI، Workflows، Microsoft، Google Agent، Agent Framework، Development Kit (ADK)، Pydantic AI، Agents	Responses API، OpenAI Agents SDK، Databricks Agent، Claude Agent SDK، Bricks
بيئات تشغيل الوكلاء المُدارة	استضافة ذاتية على Kubernetes	Microsoft، Amazon Bedrock AgentCore، Google Vertex، Foundry Agent Service، AI Agent Engine
معايير التشغيل البيئي وبوابات MCP	Model Context Protocol (MCP)، AGENTS.md، (A2A) Agent، Docker MCP Gateway	Azure API، Kong AI Gateway، Management
النماذج	عائلات النماذج مفتوحة الأوزان: Llama، Qwen، DeepSeek، Mistral، Phi، gpt-oss، Gemma	عائلات النماذج الرائدة: Anthropic Claude، OpenAI GPT، Google Gemini، Cohere، Mistral، Command
بوابات النماذج وتشغيلها	vLLM، SGLang، TensorRT-LLM، Ollama، LiteLLM	Microsoft Foundry، Amazon Bedrock، Databricks، Google Vertex AI، Models، Snowflake Cortex AI، Model Serving، Portkey
التنفيذ الدائم والتكامل	Temporal، Apache Kafka، Apache، Airflow	AWS Step Functions، Temporal Cloud، UiPath، Azure Durable Functions، Workato
المعرفة والاسترجاع	pgvector، Milvus، Qdrant، Weaviate، OpenSearch، Chroma، نماذج التضمين المفتوحة مثل BGE و Eo	Vertex AI، Azure AI Search، Pinecone، Databricks Vector Search، Search، Snowflake Cortex Search، ونماذج التضمين وإعادة الترتيب من OpenAI و Cohere و Voyage AI
البيانات المؤسسية والطبقة الدالية	Unity Catalog (مفتوح المصدر)، Apache Iceberg، dbt، OpenLineage	Google BigQuery، Snowflake، Databricks، Microsoft Fabric

القدرة	المصادر المفتوحة والمعايير المفتوحة	حزم تطوير البرمجيات من الموردين والحلول المُدارة والتجارية
التنفيذ في بيئات معزولة واستخدام الحاسوب	gVisor، Firecracker، و Playwright	Bedrock، و Modal، و Daytona، و E2B Browser، و AgentCore Code Interpreter
هوية الوكلاء والوصول والسياسات	Open Policy Agent (OPA)، و Cedar، و Keycloak، و SPIFFE/SPIRE، و	Auth، و Okta، و Microsoft Entra Agent ID، و Bedrock AgentCore، و for AI Agents، و HashiCorp Vault، و Identity
حواجز الحماية والسلامة	Guardrails AI، و NeMo Guardrails، و Prompt Guard، و Llama Guard، و Microsoft Presidio	Azure AI، و Amazon Bedrock Guardrails، و Content Safety (Prompt Shields)، و Lakera، و Google Model Armor
قابلية المراقبة والتتبع	OpenLLMetry، و OpenTelemetry، و MLflow، و Arize Phoenix، و Langfuse، و Grafana، و Prometheus، و Tracing	Datadog LLM Observability، و LangSmith، و Weights & Biases Weave، و Arize AX، و Braintrust
التقييم واختبارات الفريق الأحمر	Ragas، و DeepEval، و promptfoo، و Inspect (معهد أمن الذكاء الاصطناعي في المملكة المتحدة)، و PyRIT، و garak	LangSmith Evaluations، و Braintrust، و Patronus AI، و Galileo
الحوكمة والامتثال	OpenLineage (تتبع مسار البيانات)	IBM، و Microsoft Purview، و OneTrust، و Credo AI، و watsonx.governance، و Google Sensitive Data Protection
النشر والبنية التحتية	Kubernetes، و Docker، و KServe، و Ray	Google GKE، و Azure AKS، و Amazon EKS، و Snowflake، و Databricks

تمت مراجعة المشهد في سبتمبر ٢٠٢٦. ونظرًا لتسارع وتيرة التغيير، تراجع TenX هذا القسم كل ربع سنة.

١٢. خارطة طريق التبني

تُرسى الضوابط المطلوبة لحالة الاستخدام الأولى، ثم تُوسَّع خدمات المنصة القابلة لإعادة الاستخدام كلما بررت ذلك حالات استخدام إضافية. وتصف المراحل التالية هذا التدرج.

المرحلة	محور التركيز	النتيجة
١. التأسيس	الملكية وتقييم المخاطر؛ تحديد نطاق الهوية والوصول؛ الاستخدام المنضبط للنماذج والأدوات؛ التقييم؛ سجلات التنفيذ؛ معالجة الموافقات والأعطال بما يتناسب مع حالة الاستخدام الأولى.	الضوابط المطلوبة جاهزة لأول تشغيل تجريبي.
٢. التشغيل التجريبي	حالة إلى حالتين من حالات الاستخدام عالية القيمة عند مستويات الاستقلالية L٠-L١؛ اختيار الأنماط؛ مجموعات البيانات المرجعية؛ اختبارات الفريق الأحمر؛ مقاييس خط الأساس.	قيمة أعمال مُقاسة وأدلة على التشغيل الآمن.
٣. التوسع	توسيع الخدمات المشتركة وسجل الوكلاء والأدوات القابلة لإعادة الاستخدام حيثما كان ذلك مبررًا؛ واستخدام A٢A للتعاون بين الوكلاء المستقلين عند الحاجة؛ وزيادة الاستقلالية ضمن السياسات المعتمدة والحدود الخاضعة للتقييم.	حالات استخدام إضافية تدعمها خدمات قابلة لإعادة الاستخدام وضوابط مشتركة.
٤. التحسين والحوكمة	تحسين التكاليف، والنماذج الصغيرة والمضبوطة بدقة، والمراقبة المستمرة للامتثال، وإدارة المخاطر على مستوى المحفظة.	اقتصاديات وحدة مستدامة وعمليات جاهزة للتدقيق.

نقطة الانطلاق: توفير الملكية، والوصول محدد النطاق، والتقييم، وسجلات التنفيذ، والموافقة ومعالجة الأعطال بالشكل المناسب قبل التشغيل التجريبي. واستخدام قدرات المنصة القائمة حيثما تلبي هذه الاحتياجات، وإضافة الخدمات المشتركة مع نمو المحفظة.

الملحق أ. مسرد المصطلحات

المصطلح	التعريف
ARAgent (AgentrAgent)	بروتوكول مفتوح يتيح للوكلاء المبنين على أطر مختلفة أو من قِبل مؤسسات مختلفة اكتشاف بعضهم والتعاون فيما بينهم.
بطاقة الوكيل	وصف مُهيكل لغرض الوكيل ومالكه وقدراته ومستوى استقلاليته وصلاحيات وصوله، يُستخدم لأغراض الاكتشاف والحوكمة.
عمليات تشغيل الوكلاء	المنهجية التشغيلية لبناء الوكلاء وتقييمهم وإصدارهم ومراقبتهم في بيئة الإنتاج.
مستوى الاستقلالية	درجة التصرف المستقل المسموح بها للوكيل في نوع معين من الإجراءات، من L٠ (المساعدة) إلى L٣ (مستقل ضمن المجال).
هندسة السياق	اختيار التعليمات والمعرفة والذاكرة وتعريفات الأدوات التي يتلقاها الوكيل في كل خطوة، وتجميعها.
التنفيذ الدائم	تنفيذ لسير العمل يحفظ الحالة بحيث يمكن استئناف العمل بعد الأعطال أو التوقفات المؤقتة أو فترات الانتظار الطويلة.
حاجز الحماية	ضابط يتحقق من المدخلات أو المخرجات أو الإجراءات وفق السياسات، ثم يحظرها أو يعدّلها أو يصعّدها.
MCP (Model Context Protocol)	بروتوكول مفتوح لإتاحة الأدوات والبيانات والأوامر النصية لتطبيقات الذكاء الاصطناعي والوكلاء بطريقة موحدة.
بوابة MCP	نقطة تحكم تتوسط الوصول إلى خوادم MCP المعتمدة، وتتولى المصادقة والتفويض والتسجيل والسياسات.
التفويض بالنيابة (OBO)	نمط لتبادل رموز الوصول يحصل فيه الوكيل على رمز وصول محدد النطاق ومقيد بمدة زمنية يمثل المستخدم الذي يعمل نيابةً عنه.
نقطة إنفاذ السياسات	المكوّن الذي يقيّم كل إجراء يتخذه الوكيل وفق الهوية والنطاق والسياسات قبل السماح بمتابعته.
المسار	التسلسل الكامل لخطوات التفكير واستدعاءات النماذج واستدعاءات الأدوات التي يتخذها الوكيل لإنجاز مهمة ما.

١. OWASP GenAI Security Project, OWASP Top ١٠ for Agentic Applications for ٢٠٢٦، ديسمبر ٢٠٢٥.

٢. OWASP GenAI Security Project, OWASP Top ١٠ for LLM Applications ٢٠٢٥.

٣. Agentic AI Foundation (Linux Foundation)، مواصفات Model Context Protocol.

٤. مواصفات Agent2Agent (A2A) Protocol.

٥. Semantic Conventions for Generative AI Systems، OpenTelemetry.

٦. NIST، AI Risk Management Framework، ١، (١-١٠٠ AI) و (١-٦٠٠ AI) Generative AI Profile.

٧. ISO/IEC ٢٠٢٣:١:٤٢٠٠، Information technology — Artificial intelligence — Management system.

٨. ISO/IEC ٢٠٢٣:٢٣٨٩٤ و ISO/IEC ٢٠٢٥:٥:٤٢٠٠.

٩. Regulation (EU) ١٦٨٩/٢٠٢٤ (EU AI Act)، بصيغتها المعدلة بموجب Digital Omnibus on AI (٢٠٢٦).

١٠. Information Commissioner's Office، Guidance on AI and data protection.

١١. Prudential Regulation Authority، SS١ ٢٣/Model risk management principles for banks.

١٢. Anthropic، Building Effective Agents، ديسمبر ٢٠٢٤.



tenx.

tenx.ai

المؤلفون

مراد سيد، المدير العام للخدمات العالمية والابتكار،
دانيال ناصر مالك، مدير الاستشارات - علوم البيانات،
رافع قاضي، المدير القطري - دول مجلس التعاون الخليجي