دورة أمن وكلاء الذكاء الاصطناعي: نمذجة التهديدات والتحكم في الوصول والاستجابة للحوادث

5 – 9 يوليو 2027

طوكيو

دورة أمن وكلاء الذكاء الاصطناعي: نمذجة التهديدات والتحكم في الوصول والاستجابة للحوادث

المرجع: 30776_597 **التاريخ:** 5 – 9 يوليو 2027 **المكان:** طوكيو **الرسوم:** GBP 6900

ملخص البرنامج

أمن وكلاء الذكاء الاصطناعي محور دورة تدريبية تمتد أسبوع لقادة الأمن والتقنية على المستوى الإداري، ويخرج منها كل مشارك بخطة ضوابط لوكيل واحد. يوسع الوكلاء الذين يستدعون الأدوات ويقرؤون البيانات ويحملون بيانات الاعتماد سطح الهجوم أبعد من مساعد المحادثة، ومع ذلك تقف المراجعات غالبا عند سلوك النموذج بينما تبقى الصلاحيات واستدعاءات الأدوات والسجلات دون اختبار. ويبني البرنامج منهجا متكررا لنمذجة التهديدات وتقييد الوصول واحتواء الوكلاء والاستجابة للحوادث، يطبق على وكيل يختاره المشارك. تقدم مارس للتدريب المملكة المتحدة هذا البرنامج في أمن وكلاء الذكاء الاصطناعي لمجموعات من منظمات مختلفة.

المبررات المؤسسية

يعمل الوكيل نيابة عن المؤسسة، فيتحول أي ضعف في صلاحياته أو مدخلاته إلى ضعف في الأنظمة التي يصل إليها. وحين تكون الضوابط رقيقة، قد تدفع تعليمات مضللة إلى تنفيذ دفعة أو كشف سجلات أو تغيير إعداد قبل أن يلاحظ أحد. أما الفرق التي تعامل الوكيل بوصفه مستخدما ذا امتيازات، بحدود مختبرة وإجراءات مراقبة، فتمنح الإدارة التنفيذية اطمئنانا أوضح وطريقا أسرع إلى التعافي.

- انكشاف البيانات: ينخفض حين تمنع تصفية المخرجات وتحديد نطاق البيانات الوكلاء من إعادة سجلات لا يحق لهم رؤيتها.
- الإجراءات غير المصرح بها: تنخفض حين يحد التحكم في الوصول بأقل الصلاحيات مما يستطيع كل وكيل تغييره.
- زمن التعافي: ينخفض حين تقصر المراقبة ودليل الاستجابة لحوادث الوكلاء زمن الاكتشاف والاحتواء.

المبادئ والأساس المعرفي

- OWASP Top 10 for LLM Applications قائمة المخاطر المجتمعية لتطبيقات النماذج اللغوية: تنظم قائمتها التي يتصدرها الحقن بالأوامر وتضم الوكالة المفرطة عمل التهديدات في الجلستين 1 و2.
- مبدأ أقل الصلاحيات: يرشد تصميم الأذونات في الجلسة 2.
- الدفاع المتعدد الطبقات: يشكل نموذج الضوابط المتدرجة عمل العزل والمراقبة في الجلسة 3.
- نمذجة التهديدات: تؤطر طريقتها في تحديد الأصول ونقاط الدخول ومسارات إساءة الاستخدام الجلسة 1 والاختبار الهجومي في الجلسة 4.

المشاركون المستهدفون

- قادة الأمن الذين يضعون الضوابط لأنظمة تتصرف من تلقاء نفسها.
- مصممو معمارية الأمن الذين يصممون الوصول والعزل والتسجيل حول الوكلاء.
- رؤساء الذكاء الاصطناعي والهندسة الذين يبنون الوكلاء أو يشترونهم ويملكون إصدارهم.
- مديرو المخاطر والتقنية الذين يحتاجون تأكيدا بأن ضوابط الوكلاء تعمل.
- قادة الاستجابة للحوادث الذين يحتوون إخفاقات تتضمن إجراءات آلية.

يأتي المشاركون عادة من قطاعات المصارف والطاقة والاتصالات والجهات العامة، ومن وظائف الأمن السيبراني والمخاطر.

مخرجات التعلم

1. سيكون المشاركون قادرين على بناء نموذج تهديدات لوكيل ذكاء اصطناعي من أدواته وبياناته وبيانات اعتماده.
2. سيكون المشاركون قادرين على تقييم التعرض للحقن بالأوامر وإساءة استخدام الأدوات وتسرب البيانات.
3. سيكون المشاركون قادرين على تصميم هوية الوكلاء والتحكم في الوصول بأقل الصلاحيات.
4. سيكون المشاركون قادرين على ترتيب أولويات ضوابط العزل والمراقبة بحسب المخاطر.
5. سيكون المشاركون قادرين على تشخيص نقاط الضعف من نتائج الاختبار الهجومي.
6. سيكون المشاركون قادرين على تبرير دليل الاستجابة لإخفاقات الوكلاء أمام الإدارة التنفيذية.

هيكل البرنامج

اليوم 1: أسس أمن وكلاء الذكاء الاصطناعي

ما الذي قد يخطئ حين يحمل الوكيل أدوات وبيانات وبيانات اعتماد؟

- خريطة معمارية الوكيل
- جرد الأصول ونقاط الدخول
- قالب نموذج التهديدات
- مراجعة حالات الحقن بالأوامر
- مخطط مسارات تسرب البيانات

اليوم 2: الهوية والتحكم في الوصول

كيف تحدد هوية الوكلاء، وكيف تقيد صلاحياتهم؟

- سجل هوية الوكلاء
- مصفوفة الأذونات بأقل الصلاحيات
- قواعد التعامل مع بيانات الاعتماد
- تصميم قائمة الأدوات المسموحة
- معايير بوابات الاعتماد

اليوم 3: العزل والمراقبة

أي طبقات تحتوي الوكيل، وأي إشارات تدل على إساءة الاستخدام؟

- تصميم بيئة التنفيذ المعزولة
- قواعد تصفية المخرجات
- معيار تسجيل الإجراءات
- كتالوج تنبيهات الشذوذ
- إجراء الإيقاف الطارئ

اليوم 4: الاختبار الهجومي والاستجابة للحوادث

كيف تكتشف نقاط الضعف قبل المهاجمين، وكيف يحتوى الإخفاق؟

- خطة الاختبار العدائي
- مكتبة سيناريوهات الهجوم
- سلم تصنيف شدة النتائج
- دليل الاستجابة لحوادث الوكلاء
- نموذج المراجعة بعد الحادث

اليوم 5: تطبيق الأدوات على وكيل واحد

هل تحمي أدوات التهديد والوصول والاستجابة وكيلا واحدا من البداية إلى النهاية؟

- تطبيق قالب نموذج التهديدات من الجلسة 1
- استخدام مصفوفة الأذونات بأقل الصلاحيات من الجلسة 2
- تنفيذ كتالوج تنبيهات الشذوذ من الجلسة 3
- فحص دليل الاستجابة لحوادث الوكلاء من الجلسة 4
- إنجاز خطة ضوابط الوكيل

المنهجية

يقوم البرنامج على تمارين منظمة مستندة إلى تصاميم وكلاء وقوائم أذونات ومقتطفات سجلات وتقارير حوادث مجهولة المصدر من جهات مصرفية وطاقة وعامة. يعمل المشاركون في فرق أمنية على حالة وكيل مشتركة، ويختبر بعضهم ضوابط بعض، ويتلقون ملاحظات مكتوبة من الميسرين على كل مجموعة تمارين. وتنتهي كل جلسة بقرار موجز بشأن الضابط الذي يعتمد ولماذا.

- نشاط مقترح: تتبع الطريقة التي قد تصل بها تعليمات محقونة إلى أداة دفع.
- نشاط مقترح: تقليص قائمة أذونات واسعة إلى أقل الصلاحيات.
- نشاط مقترح: متابعة حادث تجريبي من أول تنبيه حتى الاحتواء.

المخرج الختامي

يعد كل مشارك خطة ضوابط لوكيل ذكاء اصطناعي واحد. وتضم نموذج تهديدات ومصفوفة أذونات وقواعد للعزل والمراقبة وخطة اختبار هجومي ودليلا للاستجابة للحوادث. وتستخدم الخطة في العمل لإحاطة لجنة الأمن وتوجيه المهندسين إلى الضوابط المطلوبة وتحديد دورة مراجعة الوكيل.

الأسئلة الشائعة

هل يناسب أمن وكلاء الذكاء الاصطناعي القادة الذين لا يملكون مهارات هندسية عملية؟

نعم، فأمن وكلاء الذكاء الاصطناعي يدرس على المستوى الإداري بالخرائط والمصفوفات ومواد الحالات، ولا يتطلب برمجة، لكنه يفترض إلماما بمفاهيم الأمن الأساسية.

على أي نماذج يقوم أمن وكلاء الذكاء الاصطناعي؟

يقوم أمن وكلاء الذكاء الاصطناعي على OWASP Top 10 for LLM Applications ومبدأ أقل الصلاحيات والدفاع المتعدد الطبقات ونمذجة التهديدات، ويطبق كل منها في جلسة محددة.

كيف تستخدم الخطة الناتجة عن أمن وكلاء الذكاء الاصطناعي في العمل؟

تستخدم خطة الضوابط لإحاطة لجنة الأمن وإبلاغ المهندسين بالضوابط التي يحتاجها الوكيل وتحديد دورة مراجعته.

بماذا يختلف أمن وكلاء الذكاء الاصطناعي عن حوكمة الذكاء الاصطناعي؟

يركز أمن وكلاء الذكاء الاصطناعي على الضوابط التقنية والاختبار والاستجابة للوكلاء، بينما تعنى حوكمة الذكاء الاصطناعي بالسياسات والمساءلة والإشراف على استخدامه في المؤسسة.