JUPITER PRIME

Consulting • Training • Advisory

جوبيتر برايم للاستشارات والتدريب

كتيّب الدورة

دورة الحد من هلوسة الذكاء الاصطناعي التوليدي للمهنيين

8 – 12 أغسطس 2027

أونلاين

info@jupiterprime.uk

jupiterprime.uk

مسجّلة في المملكة المتحدة والمملكة العربية السعودية

دورة الحد من هلوسة الذكاء الاصطناعي التوليدي للمهنيين

المرجع: 1622_59399 **التاريخ:** 8 – 12 أغسطس 2027 **المكان:** أونلاين **الرسوم:** GBP 2900

الحد من هلوسة الذكاء الاصطناعي التوليدي دورة تدريبية مدتها أسبوع للمديرين والمهنيين الذين يعتمدون على المساعدات الذكية في صياغة المسودات والملخصات والأرقام، ويخرج المشاركون منها ببروتوكول التحقق من الإجابات لإداراتهم. فالإجابات السلسة المختلقة تصل اليوم إلى التقارير ورسائل العملاء وأوراق مجالس الإدارة، لذلك صار فحص الادعاءات المولدة آليا قبل استخدامها عادة مهنية أساسية. وتقدم جوبيتر برايم للتدريب هذه الدورة في الحد من هلوسة الذكاء الاصطناعي التوليدي.

لماذا يهم الحد من هلوسة الذكاء الاصطناعي التوليدي الآن

تشكل ثلاث قوى هذا المجال اليوم: مساعدات ذكية مدمجة في برمجيات المكاتب اليومية، وقيادات تنتظر مسودات أسرع، وعملاء ومدققون يحملون الجهة مسؤولية كل معلومة تعلنها. الجهة التي تفحص الاستشهادات المختلقة والأرقام المحرفة والتخمينات الواثقة تحفظ مصداقيتها، أما التي تهمل ذلك فتنشر أخطاء يصعب سحبها وتكلف كثيرا. وتستند الدورة إلى ISO/IEC 42001، المعيار الدولي لنظام إدارة الذكاء الاصطناعي، وإلى إطار NIST لإدارة مخاطر الذكاء الاصطناعي RMF NIST AI الذي يعد الصحة والموثوقية من خصائص الثقة الجوهرية.

المستفيدون من الحد من هلوسة الذكاء الاصطناعي التوليدي

- مهنيون يحولون مسودات المساعدات الذكية إلى تقارير وإحاطات ومراسلات
- مديرون يعتمدون وثائق كتب الذكاء الاصطناعي جزءا من نصها
- محللون يطلبون من المساعدات أرقاما ومقارنات وملخصات لملفات طويلة
- مسؤولو الجودة والمخاطر الذين يضعون قواعد فحص المحتوى المولد آليا
- موظفو المعرفة والسجلات الذين يرعون مكتبات المصادر التي تقرأ منها المساعدات

تستفيد الإدارة أكثر حين ترسل مدير الاعتماد مع اثنين أو ثلاثة من المستخدمين الدائمين للمساعدات، فيتفق من يصوغ ومن يعتمد على روتين واحد.

قدرات الحد من هلوسة الذكاء الاصطناعي التوليدي

ينتقل المشاركون من الثقة بالنص السلس إلى اختباره ادعاء تلو الآخر أمام الأدلة ودرجة الخطورة.

- تقييم أسباب إنتاج النموذج اللغوي لعبارة مقنعة لكنها خاطئة — اختيار المهام الصالحة للمساعدات
- تطبيق استخراج الادعاءات وتتبع المصادر على مسودة مولدة — اعتماد المدير المسؤول
- تصميم أوامر وتأريض يضيقان مجال الاختلاق — الفرق التي تكتب وثائق متكررة
- حساب نسبة العينة لفحص المخرجات الكثيفة — مسؤولو الجودة والمخاطر
- تحديد عتبات الموافقة البشرية حسب عاقبة الخطأ — رؤساء الإدارات والمدققون
- بناء سجل أخطاء يحول كل اختلاق مكتشف إلى درس — أمناء مكتبة المصادر

مصطلحات الحد من هلوسة الذكاء الاصطناعي التوليدي الأساسية

- الهلوسة — عبارة مولدة سلسة واثقة لكنها خاطئة أو مختلقة.
- التأريض — تزويد النموذج بمقاطع مصادر معتمدة ليجيب منها لا من ذاكرته.
- استخراج الادعاءات — تفكيك المسودة إلى عبارات مفردة قابلة للفحص مثل الأرقام والأسماء والاقتباسات.
- الاستشهاد المختلق — مرجع أو اقتباس يقدمه النموذج على أنه حقيقي وهو غير موجود.
- تصنيف العواقب — ترتيب المخرجات حسب الضرر الذي قد يسببه الخطأ ليتبع جهد الفحص حجم الخطر.
- الاعتماد البشري — شخص مسمى يوافق على المخرج المولد قبل خروجه من الفريق أو بنائه لقرار.

الحد من هلوسة الذكاء الاصطناعي التوليدي يوما بيوم

اليوم 1: كيف يخطئ النص السلس

- أساسيات التنبؤ بالكلمة التالية — لماذا لا تثبت السلاسة الدقة
- حدود بيانات التدريب والسياق الغائب — قيود الأسئلة عن الأحداث الجارية
- أنماط إخفاق النصوص والصور والشيفرات — مواءمة الفحص مع نوع المخرج
- ملفات حالات هلوسة من المالية والرعاية الصحية — كلفة الادعاءات غير المفحوصة
- مصفوفة ملاءمة المهام — أين تعين المساعدات وأين تضلل

اليوم 2: رصد الاختلاق في المسودات

- تقنية استخراج الادعاءات — تفكيك المسودات إلى عبارات قابلة للفحص
- أنماط الإنذار — أرقام دقيقة ومصادر مبهمة واقتباسات غريبة
- تتبع الاستشهادات والروابط — التأكد من وجود المراجع فعلا
- المطابقة الرقمية — تسوية المجاميع والنسب ووحدات القياس
- اختبارات سريعة للشيفرات والمعادلات — تشغيل حالات صغيرة قبل إعادة الاستخدام

اليوم 3: تقليل الاختلاق من المنبع

- الأوامر المؤرضة بمقاطع معتمدة — حصر الإجابات في الأدلة
- الاسترجاع من مكتبات منتقاة — اختيار الوثائق التي تراها المساعدات
- تعليمات الإقرار بعدم اليقين — طلب إظهار الفجوات بدل التخمين
- قواعد التعامل مع البيانات السرية — ما لا يدخل الأمر أبدا
- مقارنة إجابتين عن سؤال واحد — كشف الإجابات غير المستقرة

اليوم 4: الضوابط والعتبات والمساءلة

- مصفوفة تصنيف العواقب — تحديد عمق الفحص حسب الضرر المحتمل
- خطط العينات للمخرجات الكثيفة — الموازنة بين السرعة والاطمئنان
- قواعد الاعتماد البشري — تسمية من يوافق على كل مخرج
- ضوابط ISO/IEC 42001 وNIST AI RMF — ربط الفحص بالسياسة المؤسسية
- سجل أخطاء الاختلاق — إعادة الدروس إلى الأوامر والمصادر

اليوم 5: ممارسة بروتوكول التحقق من الإجابات

- تمرين تدقيق مسودة حية — تطبيق استخراج الادعاءات تحت ضغط الوقت
- تحدي إعادة الصياغة المؤرضة — إصلاح إحاطة معيبة بالمصادر
- تصنيف استخدامات المساعدات في إدارتك — ربط مخرجات الإدارة بالفحوص
- مراجعة الأقران لقواعد الاعتماد — اختبار العتبات على حالات واقعية
- صياغة بروتوكول التحقق من الإجابات — النسخة النهائية لاستخدام الإدارة

تطبيق الحد من هلوسة الذكاء الاصطناعي التوليدي بعد الدورة

- إجراء تدقيق باستخراج الادعاءات على عشر وثائق حديثة كتبتها المساعدات وتدوين كل اختلاق في سجل أخطاء مشترك.
- الاتفاق مع رئيس الإدارة على فئات العواقب وعتبات الاعتماد، ثم تجربة خطة العينات على تقرير دوري واحد.

أسئلة حول الحد من هلوسة الذكاء الاصطناعي التوليدي

ما الحد من هلوسة الذكاء الاصطناعي التوليدي ولماذا يهم الآن؟

هو ممارسة رصد العبارات الخاطئة في المخرجات المولدة ومنعها. ويهم الآن لأن رقما مختلقا واحدا في وثيقة لعميل قد يكلف من الثقة أكثر مما وفره المساعد من وقت.

من يستفيد أكثر من تدريب الحد من هلوسة الذكاء الاصطناعي التوليدي؟

يستفيد أكثر المهنيون الذين يصوغون بالمساعدات والمديرون الذين يعتمدون تلك المسودات. ويستفيد كذلك المحللون ومسؤولو المخاطر وأمناء مكتبات المصادر.

بم تختلف دورة الحد من هلوسة الذكاء الاصطناعي التوليدي عن دورة عامة في أدوات الذكاء الاصطناعي؟

تركز على إثبات صحة المخرجات لا على إنتاجها أسرع. فالدورة العامة تعلم الخصائص لرفع الإنتاجية، أما هذه فتعلم فحص الادعاءات والتأريض والاعتماد البشري.

ماذا تعود به من تدريب الحد من هلوسة الذكاء الاصطناعي التوليدي؟

تعود ببروتوكول تستطيع إدارتك اعتماده فورا. يحدد المخرجات التي تحتاج فحصا، وعمق الفحص، ومن يوقع بالموافقة.

الحد من هلوسة الذكاء الاصطناعي التوليدي: الطريق إلى الأمام

يحدد بروتوكول التحقق من الإجابات أولا أي مخرجات المساعدات يمكن أن تغادر الإدارة دون قارئ ثان. يستخدمه من يصوغ يوميا، والمديرون عند الاعتماد، ومسؤولو المخاطر عبر سجل أخطائه للحكم على فاعلية الضوابط. وبعد الشهر الأول تستحق فئات العواقب مراجعة جديدة، لأن المخرجات التي بدت قليلة الخطر كثيرا ما تخفي أدق صور الاختلاق.