OpenRouter تُخرج ثلاثة مزوّدين للذكاء الاصطناعي بعد تسرّب أوامر اختبار وبيانات اعتماد وهمية



أزالت OpenRouter مزوّدي Novitag Phalag Together AI بعد أن كشفت Dragos تسرّب أوامر اختبار. على الجهات التي أرسلت بيانات حساسة أو أسراراً عبر المنصة تدوير بيانات الاعتماد ومراجعة التعرض.

- أرسلت Dragos مليوني أمر اختبار مزوّد بمعرّفات تتبّع عبر OpenRouter ورصدت تعرّض بيانات ثلاثة مزوّدين على الأقل.
- تؤكد OpenRouter أن أنظمتها لم تُخترق، وأن التسرب وقع لدى المزوّدين.
- أُزيل Novitag Phalag Together AI، وسحبت Atlas Cloud نماذجها طوعاً إلى حين اختبارات أمنية.
- أُلغيت أهلية Zero Data Retention لـ125 نقطة اتصال، وأُبلغ العملاء المحتمل تأثرهم.
- ينصح المزوّدون المستخدمين بتدوير أي أسرار ربما وردت في الطلبات المتسرّبة.

أزالت OpenRouter، وهي وسيط يتيح للمطوّرين الوصول إلى نماذج ذكاء اصطناعي من شركات متعددة بمفتاح API واحد، ثلاثة مزوّدين للاستدلال بعد أن تبيّن مخالفتهم سياسات التعامل مع البيانات المعلنة.

ما الذي حدث

أرسلت Dragos، شركة أمن التقنيات التشغيلية، مليوني أمر اختبار عبر المنصة، يحمل كل منها رابط تتبع فريداً أو بيانات اعتماد وهمية. ثم رصدت نشاطاً يدل على تعرّض بيانات ثلاثة مزوّدين على الأقل، شمل محاولات استخدام بيانات الاعتماد الوهمية ضد أنظمة اختبارها المضبوطة وحسابات AWS الكاشفة.

تقول OpenRouter إن أنظمتها لم تُخترق. وقد أزالت Together AI وPhala وNovita، وألغت احترازياً أهلية Zero Data Retention (ZDR) لـ125 نقطة اتصال. وأعلنت Atlas Cloud سحب نماذجها طوعاً إلى حين اختبارات أمنية بعد حادثة لدى طرف ثالث.

تفسيرات المزوّدين

- Together AI: كشف وكيل برمجة بالذكاء الاصطناعي بيانات اعتماد مستودع على GitHub، فنزّل طرف غير مصرّح عينات من الأوامر المخزّنة من يونيو ويوليو.
- Phala: اخترق مهاجم خادماً يعالج طلبات مُوجَّهة عبر OpenRouter يومي 21 و22 أكتوبر، وربما حتى 8 نوفمبر، وحصل على بيانات اعتماد تطبيقات.
- Novita: أدى خطأ في إعدادات التوجيه إلى إرسال بعض طلبات MiniMax إلى واجهة MiniMax الرسمية بين 28 مايو و11 نوفمبر.

يعمل المزوّدون الأربعة على تشديد الضوابط وسياسات حذف البيانات. وستكيّف OpenRouter اختبارات Dragos للاستخدام المستمر، وستشترط التحقق قبل إعادة أهلية ZDR.

لماذا يهم ذلك

كثيراً ما تتضمن الأوامر بيانات اعتماد أو معلومات داخلية أو بيانات شخصية. ويضيف نموذج الوسيط سلسلة أطراف ثالثة يصعب رؤية طريقة تعاملها مع البيانات من جهة التطبيق.

ما يجب فعله الآن

1. حدّدوا التطبيقات والفرق التي تستخدم OpenRouter أو وسطاء مشابهين، والمزوّدين والنماذج التي تُوجَّه إليها الطلبات.
2. دوّروا مفاتيح API والرموز وأي أسرار ربما وردت في أوامر أُرسلت عبر المزوّدين المتأثرين.
3. راجعوا سجلات الأوامر بحثاً عن بيانات شخصية ومعلومات مصنّفة، وقيّموا التزامات الإشعار وفق PDPL إذا كان التعرض محتملًا.
4. أدرجوا وسطاء الذكاء الاصطناعي ومزوّدي النماذج في مراجعات مخاطر الأطراف الثالثة، واطلبوا إثباتات على ضوابط الاحتفاظ بالبيانات وحذفها.
5. امنعوا إدراج بيانات الاعتماد والبيانات الحساسة في الأوامر، وفعّلوا فحص الأسرار في الشيفرة ومخرجات وكلاء الذكاء الاصطناعي.