1. قرار العتاد

ابدأ بمعلومات قابلة للقياس: الذاكرة العشوائية المتاحة، ذاكرة البطاقة الرسومية، التخزين الحر، ونوع المعالج. تختلف حاجة النموذج إلى الذاكرة بحسب عدد المعاملات وطريقة التكميم وطول السياق والبرنامج المستخدم. لذلك لا تعتمد على قاعدة تقول إن نموذجا بحجم ملف محدد يحتاج المقدار نفسه تماما من RAM أو VRAM. افتح مراقب الموارد أثناء التجربة وسجل الذروة الفعلية.

طبق الآن

- سجل RAM وVRAM المتاحة والمساحة الحرة قبل التنزيل.
- ابحث عن متطلبات النموذج المختار في مصدره الرسمي ثم اكتب سبب اختيارك.

2. التثبيت من المصدر الرسمي

زر موقع Ollama الرسمي واختر نسخة نظامك. بعد التثبيت، تصفح مكتبة النماذج الرسمية واختر نموذجا صغيرا كبداية. انسخ أمر التشغيل من صفحة النموذج نفسها لأن الوسوم والأحجام قد تتغير. انتظر اكتمال التنزيل قبل فصل الإنترنت. احتفظ بمساحة للتنزيل وملفات التشغيل، ولا تدخل بيانات خاصة في نموذج لم تتحقق من مصدره وترخيصه.

طبق الآن

- احفظ رابط صفحة التنزيل الرسمية واسم إصدار البرنامج.
- حدد وسم النموذج وحجم التنزيل قبل تشغيل أمره.

3. تجربة تشغيل قابلة للتكرار

اكتب ثلاثة أسئلة ثابتة: شرح مفهوم تقني بالعربية، تلخيص فقرة قصيرة تملكها، وسؤال بلا جواب داخل الفقرة. سجل اسم النموذج ووسمه، النظام، RAM وVRAM، طول السؤال، زمن أول رد والزمن الكلي، ثم قيم دقة الإجابة. عندما تكرر التجربة بنموذج ثان، أبق الأسئلة والظروف متقاربة. لا تفسر فرق السرعة قبل التأكد من عدم وجود عمليات ثقيلة في الخلفية.

طبق الآن

- اختبر السؤال العربي نفسه مرتين وسجل الزمن والذاكرة.
- كرر الخطوة بنموذج ثان مع الحفاظ على نص السؤال.

4. تشخيص مشكلة شائعة

إذا توقف البرنامج عند التحميل أو أصبح النظام بطيئا جدا، افحص استهلاك الذاكرة. خفض حجم النموذج أو طول السياق يمكن أن يساعد، لكنه قد يغير جودة الإجابة. إذا ظهر خطأ تنزيل، تحقق من الاتصال والمساحة أولا ولا تلصق أمرا من موقع غير موثوق. إذا أخطأ النموذج بالعربية، جرب صياغة أوضح وقارن بنموذج آخر يدعم العربية، ثم راجع الإجابة مقابل المصدر.

طبق الآن

- اكتب رسالة خطأ واحدة ظهرت لك أو حالة اختبار مصطنعة.
- اربطها بخطوة تشخيص محددة دون افتراض أن المشكلة من النموذج دائما.

5. مهمة التسليم

قدم جدول مقارنة لنموذجين أو إعدادين على جهاز واحد. أرفق مواصفات الجهاز والوسم الصحيح لكل نموذج وثلاثة أسئلة ونتائجها. اكتب فقرة تشرح المقايضة: هل النموذج الأسرع أقل دقة؟ هل تحسن الجواب بعد تقليل السياق أو تدهور؟ اختم بقرار مبرر لما يصلح لجهازك ومهمتك. لا تعرض أرقام أداء جهازك على أنها نتيجة عامة لجميع الأجهزة.

طبق الآن

- سلم جدولا يضم النموذجين ومواصفات الجهاز ونتائج ثلاثة أسئلة.
- اختر نموذجا لمهمتك واكتب سببا قابلا للتحقق لهذا القرار.

مصادر للمراجعة

https://ollama.com/download

https://ollama.com/library

https://github.com/ggml-org/llama.cpp

المصادر التقنية تتغير؛ راجع النسخة الحالية قبل تنفيذ الخطوات.