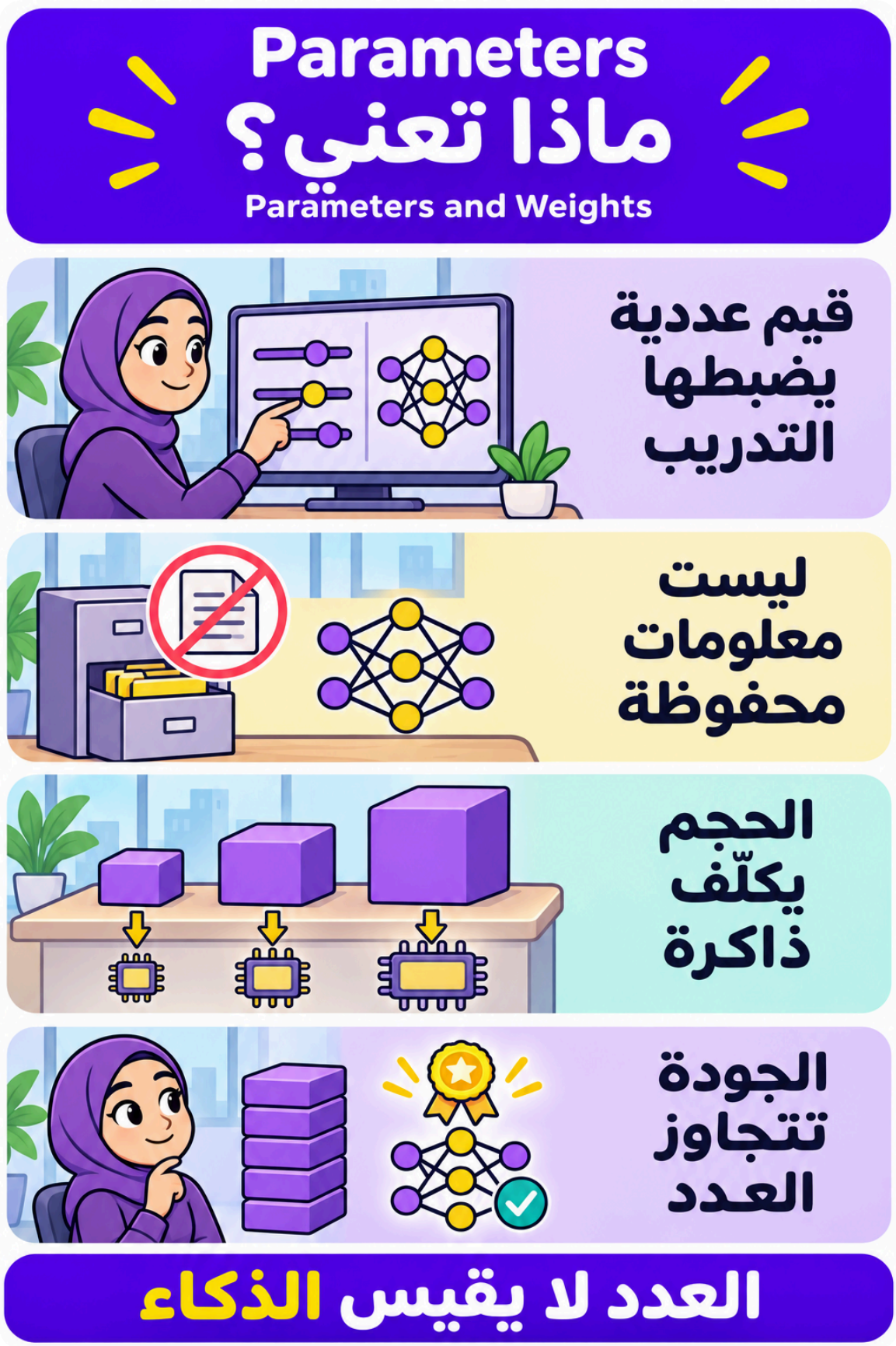
مختصر كتابي

Parameters and Weights



ماذا يعني أن يكون للـModel 70 Billion Parameters؟

إن كان أمامك Model حجمه 7 Billion Parameters وآخر حجمه 70 Billion، فهل الثاني أذكى عشر مرات؟ وهل يعني الرقم أنه يحفظ سبعين مليار معلومة؟ الجواب عن السؤالين: لا. ولكي تفهم الرقم لا بد أن تعرف ما الذي نعدّه أصلًا.

الـ**Parameters** هي القيم العددية القابلة للتعلّم داخل الـModel، والتدريب هو الذي يضبطها لتغيّر طريقة حسابه للنتيجة. وهي تختلف عن الـ Tokens التي هي وحدات النص.

خذ مثالًا مصغّرًا بعيدًا عن تعقيد الـModels اللغوية: نتوقع مدة توصيل طلب من المسافة وعدد الطلبات المنتظرة، فنضرب المسافة في قيمة، ونضرب عدد الطلبات في قيمة أخرى، ثم نضيف رقمًا ثابتًا. القيم التي نضرب فيها هي **الأوزان** (Weights)، والرقم الثابت هو **الانحياز** (Bias)، وهذه كلها Parameters. أما مسافة الطلب الجديد فهي Input وليست Parameter. أثناء التدريب نقيس خطأ التوقعات ونعدّل القيم لتقليله، وعند الاستخدام ندخل بيانات طلب جديد ونحسب النتيجة بالقيم التي تعلّمناها. والـModel اللغوي أعقد من ذلك بكثير، لكن الفكرة واحدة، وعبارة 70 Billion Parameters تعني نحو سبعين مليار قيمة من هذا النوع.

تخيّلها مفاتيح ضبط في جهاز معقد. عدد أكبر من المفاتيح قد يتيح تمثيل أنماط أعقد، لكن الجودة تعتمد أيضًا على كيفية ضبطها وعلى تصميم الجهاز. وليس هناك مفتاح اسمه «عاصمة مصر» وآخر اسمه «الاعتذار للعميل»، فالمعرفة والسلوك يُمثَّلان غالبًا بصورة موزعة عبر أعداد وحسابات كثيرة، والـ Parameters ليست صفوفًا في Database فيها سؤال وجواب.

وللحجم تكلفة في الذاكرة: سبعون مليار Parameter، وكل قيمة مخزنة في بايتين (2 Byte)، تحتاج وحدها إلى نحو مئة وأربعين جيجابايت بالنظام العشري. وهذا حساب حجم الـ Weights فقط، وليس إجمالي ذاكرة التشغيل.

- الـ Parameters قيم قابلة للتعلّم يضبطها التدريب، وليست معلومات محفوظة ولا مقياسًا مباشرًا للذكاء.
- زيادة الحجم داخل مقارنات مضبوطة قد ترفع القدرة على تعلّم الأنماط المعقدة، لكن الـ Architecture وجودة بيانات التدريب وكميته والـ Post-Training تغيّر النتيجة أيضًا.
- Model أصغر مدرَّب جيدًا على اتباع التعليمات ومعه بيانات الطلب الصحيحة قد يخدم المهمة أفضل من Model أكبر لا يرى البيانات أو سلوكه غير مناسب.
- تغيير طريقة تمثيل الأرقام، مثل Quantization، قد يقلّل الحجم، لكن له تأثيرات يجب قياسها.
- في Architectures مثل Mixture of Experts يختلف إجمالي الـ Parameters عن الجزء الذي يُفعَّل لكل Token، فلا تقارن الرقمين كأنهما مقياس واحد.
- عند اختيار Model اسأل: بأي جودة يحل مهمتي؟ وفي كم من الوقت والتكلفة؟ وبأي موارد؟

الخلاصة: عدد الـ Parameters يساعدك على فهم تصميم الـModel، لكنه ليس تقييمًا جاهزًا للذكاء.