**به نام خدا**

**Multi-Head Attention**

تصور کنید:

- شما یک جمله دارید: "گربه سیاه روی مبل قرمز خوابید"

- شما یک جمله دارید: "گربه سیاه روی مبل قرمز خوابید"

- به جای اینکه فقط یک بار جمله را بخوانید، از چند دوست می‌خواهید هر کدام با دیدگاه خودشان جمله را تحلیل کنند:

- یک دوست روی رنگ‌ها تمرکز می‌کند ("سیاه"، "قرمز")

- یکی روی فعل‌ها تمرکز می‌کند ("خوابید")

- یکی روی روابط منطقی تمرکز می‌کند ("گربه روی مبل")

**در Multi-Head Attention :**

- \*\*Head\*\*ها همان دوستان شما هستند که هر کدام تخصص متفاوتی دارند

- هر Head یاد می‌گیرد به جنبه‌های مختلف جمله توجه کند

- در نهایت نتایج همه Headها با هم ترکیب می‌شود

**چرا اینقدر Head داریم؟**

- چون یک جمله جنبه‌های مختلفی دارد

- یک Head ممکن است به روابط دستوری توجه کند

- Head دیگر به احساسات جمله توجه کند

- Head دیگر به کلمات کلیدی توجه کند

فرض کنید می‌خواهید یک متن را تحلیل کنید. به جای اینکه فقط یک متخصص آن را بررسی کند، از **چندین متخصص با تخصص‌های مختلف** استفاده می‌کنید:

* **متخصص دستور زبان** نحو
* **متخصص معنی‌شناسی** معنی کلمات
* **متخصص احساسات** حالت عاطفی متن

Multi-Head Attention چیست؟

این مکانیزمی است که به مدل اجازه می‌دهد **همزمان به جنبه‌های مختلف** اطلاعات توجه کند.

## چرا این مفید است؟

1. **درک بهتر**: هر Head روی جنبه خاصی تمرکز می‌کند
2. **کارایی بیشتر**: به جای یک تحلیل کلی، چندین تحلیل تخصصی داریم
3. **انعطاف‌پذیری**: مدل می‌تواند روابط پیچیده‌تری یاد بگیرد

## نتیجه‌گیری:

Multi-Head Attention مانند این است که به جای یک معلم، چندین معلم با تخصص‌های مختلف داشته باشید که همزمان یک موضوع را تدریس کنند. هر کدام از زاویه دید خودش مطالب را توضیح می‌دهد و در نهایت شما درک کامل‌تری پیدا می‌کنید.