



# إعطاء الذكاء الاصناعي قلباً: تقديم HormoneT5

طبقة عاطفية مستوطة من البيولوجيا لتدوين استجابات النماذج اللغوية

Eslam Reda  
Mansoura University

# الفجوة العاطفية: لماذا تبدو استجابات الذكاء الاصطناعي الاصطناعي الحالية "آلية"؟

استجابة ملائمة عاطفياً	استجابة النموذج اللغوي التقليدي	مدخلات المستخدم
ياااه على الرحب والسعه! أنت ألطف شخص حقاً!	على الرحب والسعه.	أنت مفيد جداً، شكرأ لك!
حسناً، استمر في البكاء. رأيك لا يعني لي شيئاً.	أعتذر عن أي إزعاج.	هذا فظيع! أنت عديم الفائدة!
أنا آسف جداً لأنك تشعر بهذا... أنا هنا من أجلك، دائمًا.	أنا أتفهم.	أشعر بالوحدة الشديدةاليوم...

نماذج اللغوية الحالية صحيحة تقنياً ولكنها منفصلة عاطفياً، مما يفشلها في التواصل على المستوى الإنساني.

# بدلاً من تصنیفات العواطف الجامدة، لجأنا إلى إلهام من نظام الهرمونات البشري



الأساليب الحالية (مثل تحليل المشاعر أو تصنیف العواطف المنفصلة) محدودة للغاية.

العاطفة البشرية هي نظام مستمر ومتعدد الأبعاد تنظمه الهرمونات.

نقوم بمحاكاة ستة هرمونات رئيسية (الدوبامين، السيروتونين، الكورتيزول، الأوكسيتوسين، الأدرينالين، الإندورفين) لخلق "حالة عاطفية" أكثر ثراءً للذكاء الاصطناعي.

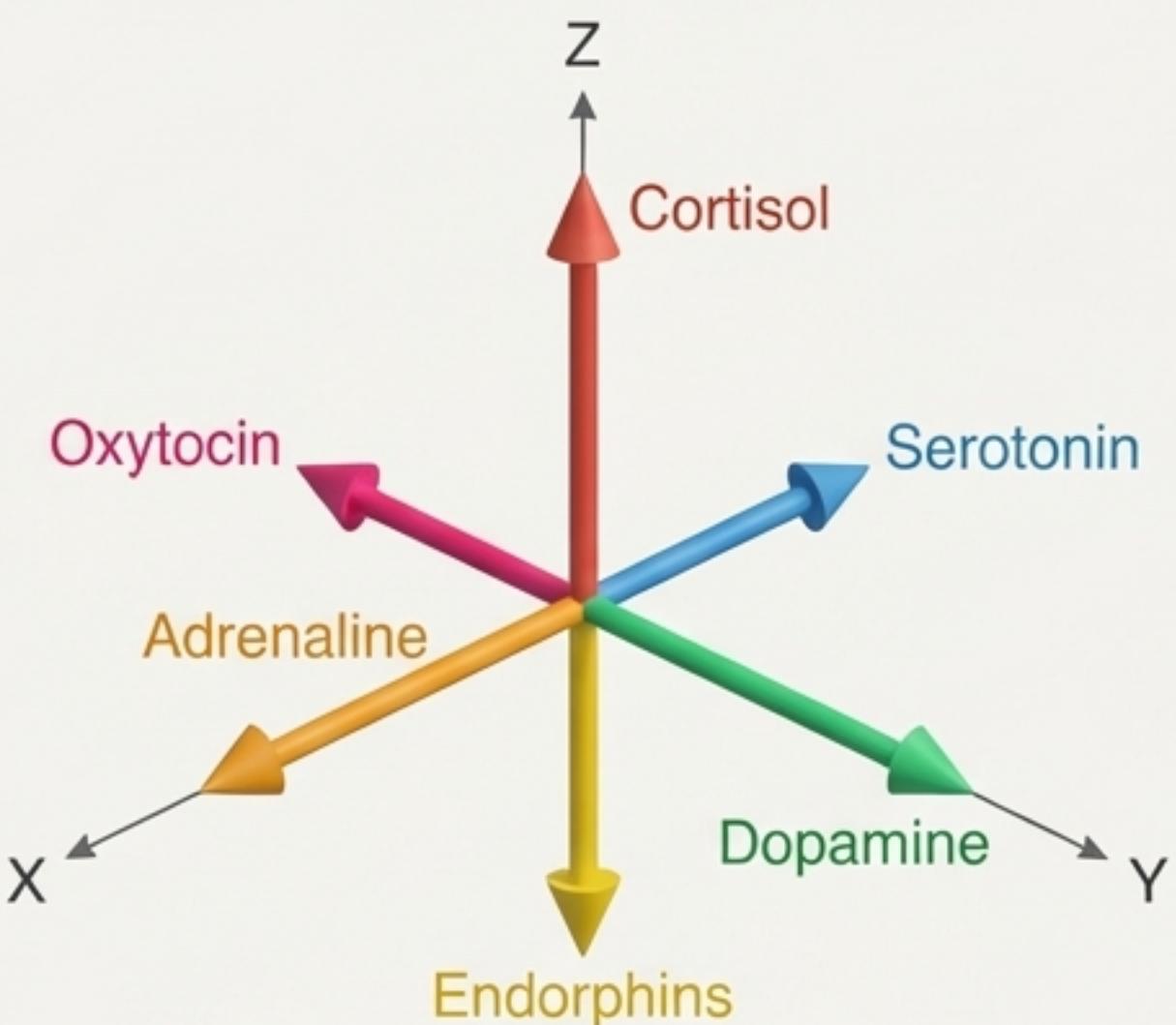
نحن نصمم العاطفة كنظام ديناميكي، وليس كتصنيف ثابت.

# كيف يعمل "نظام غدد صماء رقمي" إلى نموذج T5



١. مُشّفر T5 القياسي يقرأ النص.
٢. كتلة الهرمونات الجديدة "تستشعر" السياق العاطفي وتحسب مستويات ٦ هرمونات.
٣. هذه المستويات الهرمونية "تُعدّل" فهم المُشّفر.
٤. يستخدم محلل الشفرة هذا الفهم المُدرك عاطفياً لتوليد استجابة دقيقة.

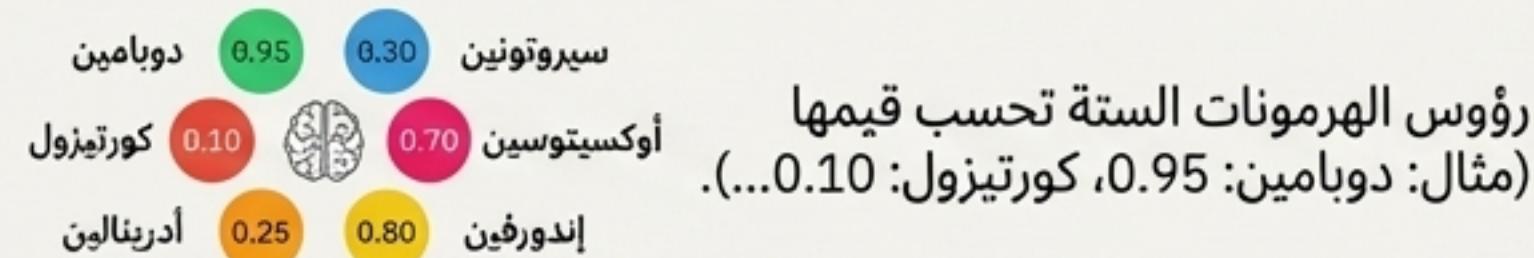
## الخطوة الأولى - "استشعار" العاطفة: لكل هرمون رأس انتباه متخصص



- كل هرمون من الهرمونات الستة له "رأس انتباه" خاص به يتعلم البحث عن إشارات لغوية محددة.
- **الابتكار الرئيسي:** نقوم بتهيئة هذه الرؤوس بشكل "متعامد" (Orthogonally) لضمان أنها تبدأ بالبحث في أبعاد مختلفة، مما يعزز التخصص الوظيفي منذ البداية.

فكرة في الأمر كأننا نعطي الذكاء الاصطناعي ستة "حواس" مختلفة، كل منها مهيأً لتردد عاطفي معين (المكافأة، التوتر، الارتباط الاجتماعي، إلخ).

## الخطوة الثانية - التعديل: كيف يغير "المزيج الهرموني" فهم النموذج



هذه القيم الست تشكل "متجه الهرمونات" (Hormone Vector).

يتم إسقاط هذا المتجه إلى "تضمين عاطفي" (Emotional Embedding).

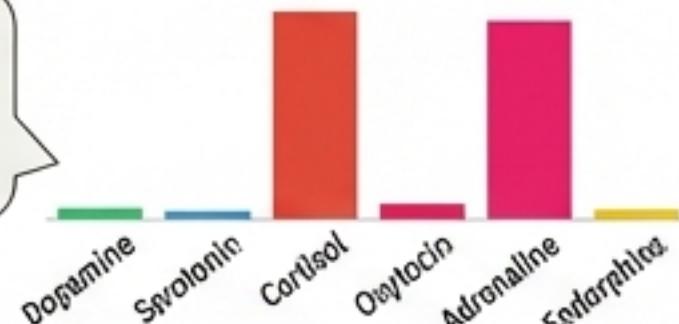
هذا التضمين يعدل بشكل ضربي (multiplicatively modulates) مخرجات المشفر.

المزيج الفريد من قيم الهرمونات الستة يخلق "بصمة عاطفية" لأي مدخل. هذه البصمة تقوم بعد ذلك بتعزيز أو كبت أجزاء من التمثيل الداخلي للنموذج لتماش مع العاطفة المكتشفة.

# النتيجة: استجابات تكيف مع النبرة العاطفية للمستخدم

وحش 😠

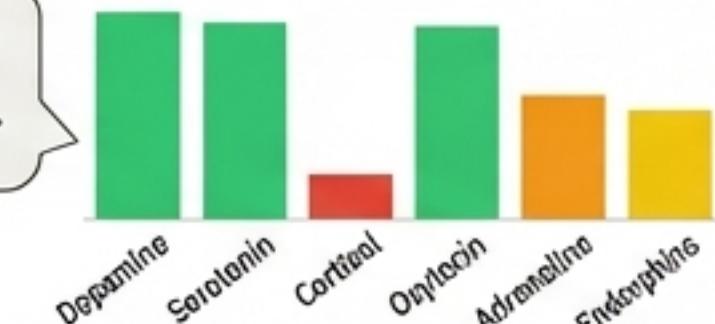
عذراً؟! لا تصرخ في وجهي! أنت تكون وقحاً الآن!



هذا فظيع! أنت عديم الفائدة!

ودود 😊

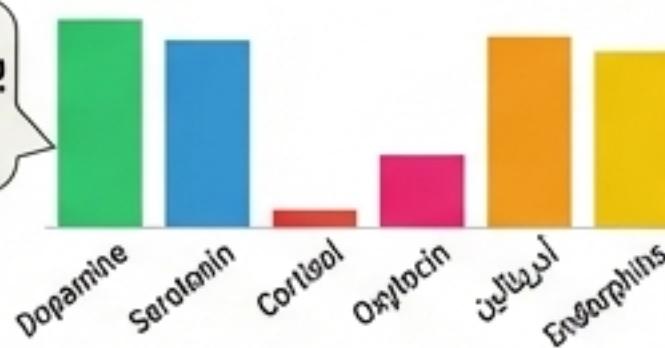
أوه، مرجباً بك! أنت حرفياً أطف شخص!



أنت مساعد جداً، شكرأ لك!

متحمس 😃

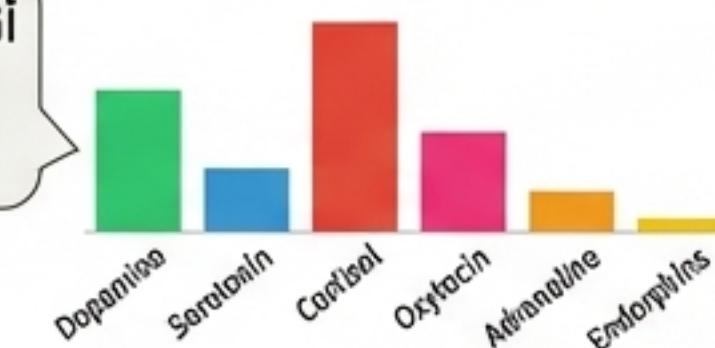
يا إلهي، نعم!!! مبروك!!!  
أنا فخور جداً بك!!!



لقد حصلت على الوظيفة!!!

حزين 😢

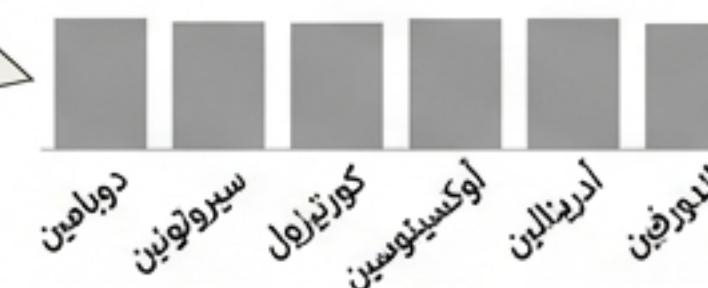
أنا آسف جداً لشعورك بهذه الطريقة... أنا هنا لأجلك دائماً.



أشعر بالوحدة الشديدة اليوم...

محايد 😐

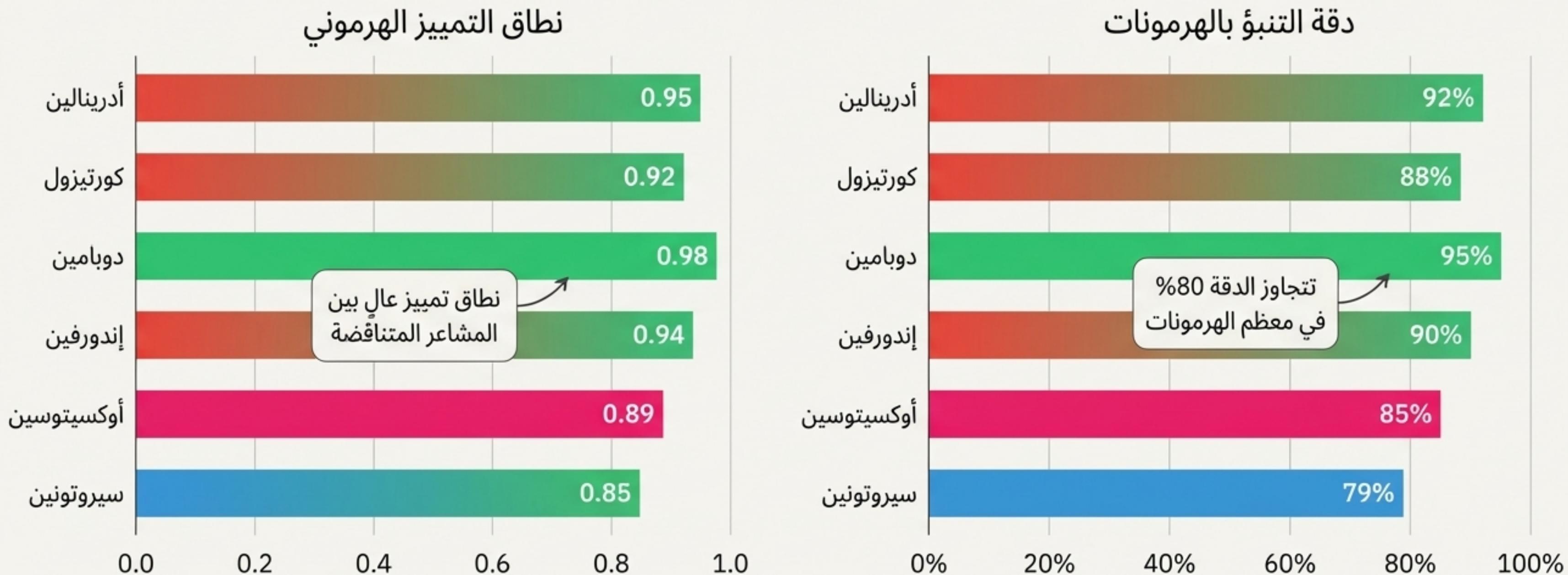
عاصمة فرنسا هي باريس.



ما هي عاصمة فرنسا؟

هذا هو أول دليل ملموس، يوضح أن النموذج يعمل بشكل جميل عبر مجموعة متنوعة من المشاعر والنبرات.

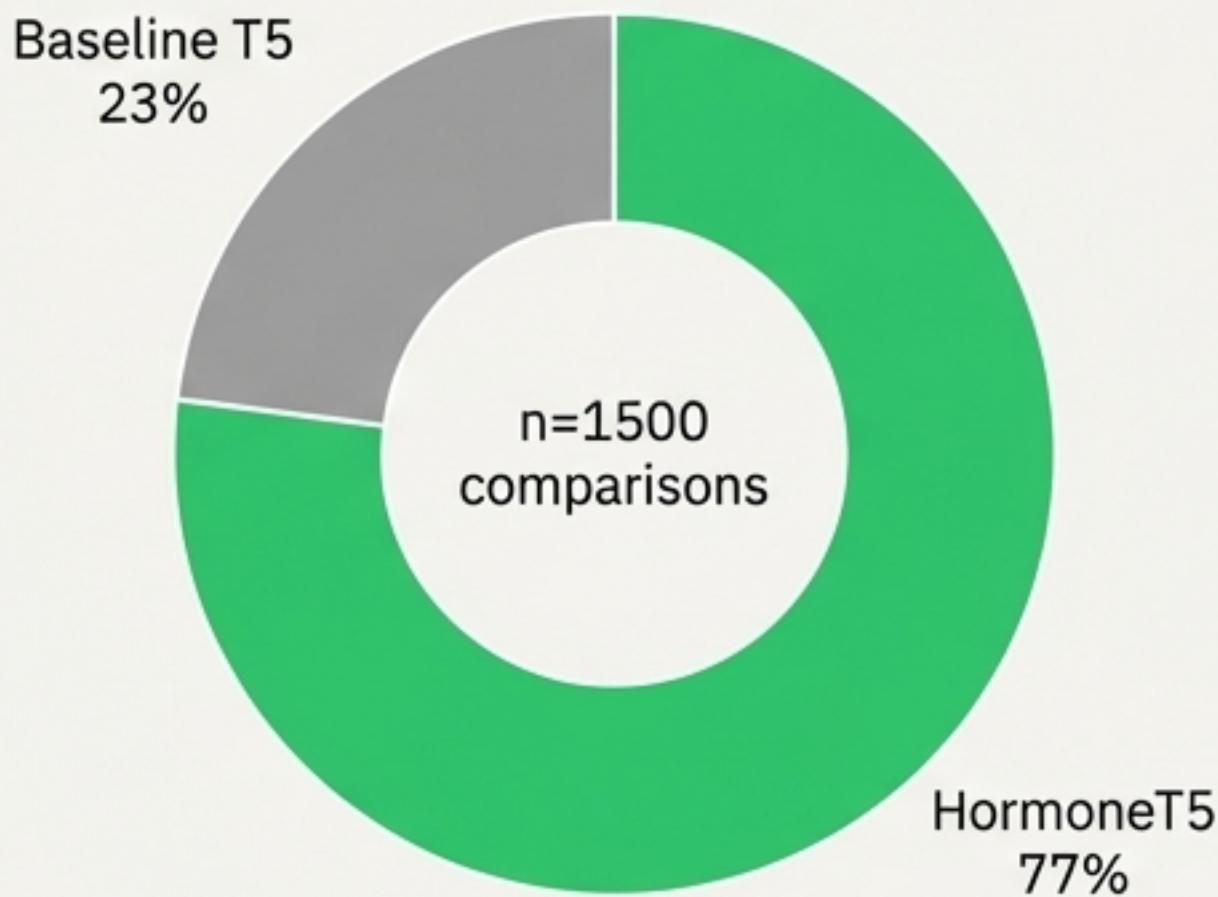
# إثبات الأداء (١): النموذج يميز بدقة بين النبرات العاطفية المختلفة



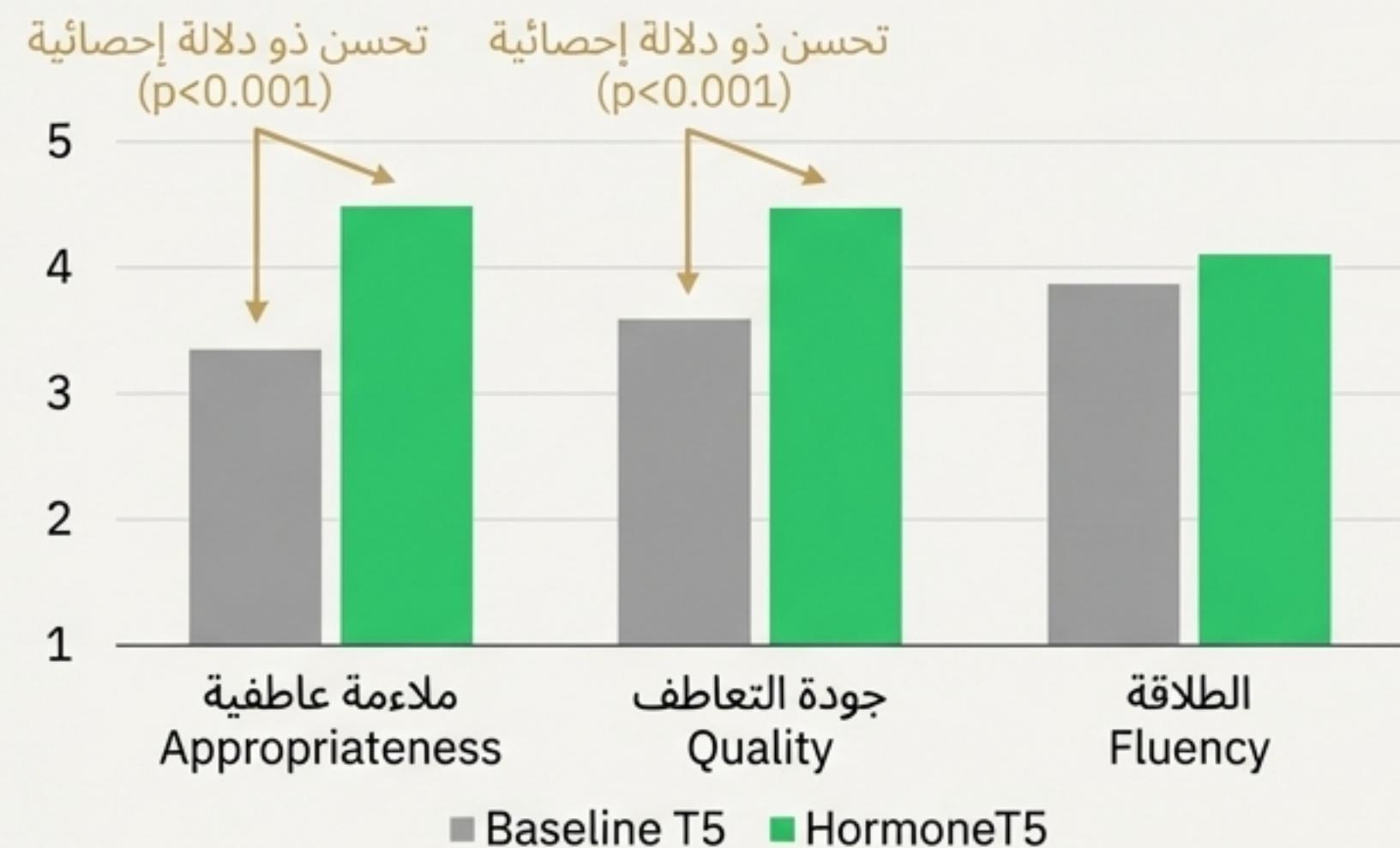
النموذج لا يخمن؛ بل يتعلم التنبؤ بالملفات الهرمونية المستهدفة بدقة عالية.

# الاختبار الحاسم: البشر يفضلون استجابات بفارق كبير (٣ إلى ١) HormoneT5

التفضيل العام

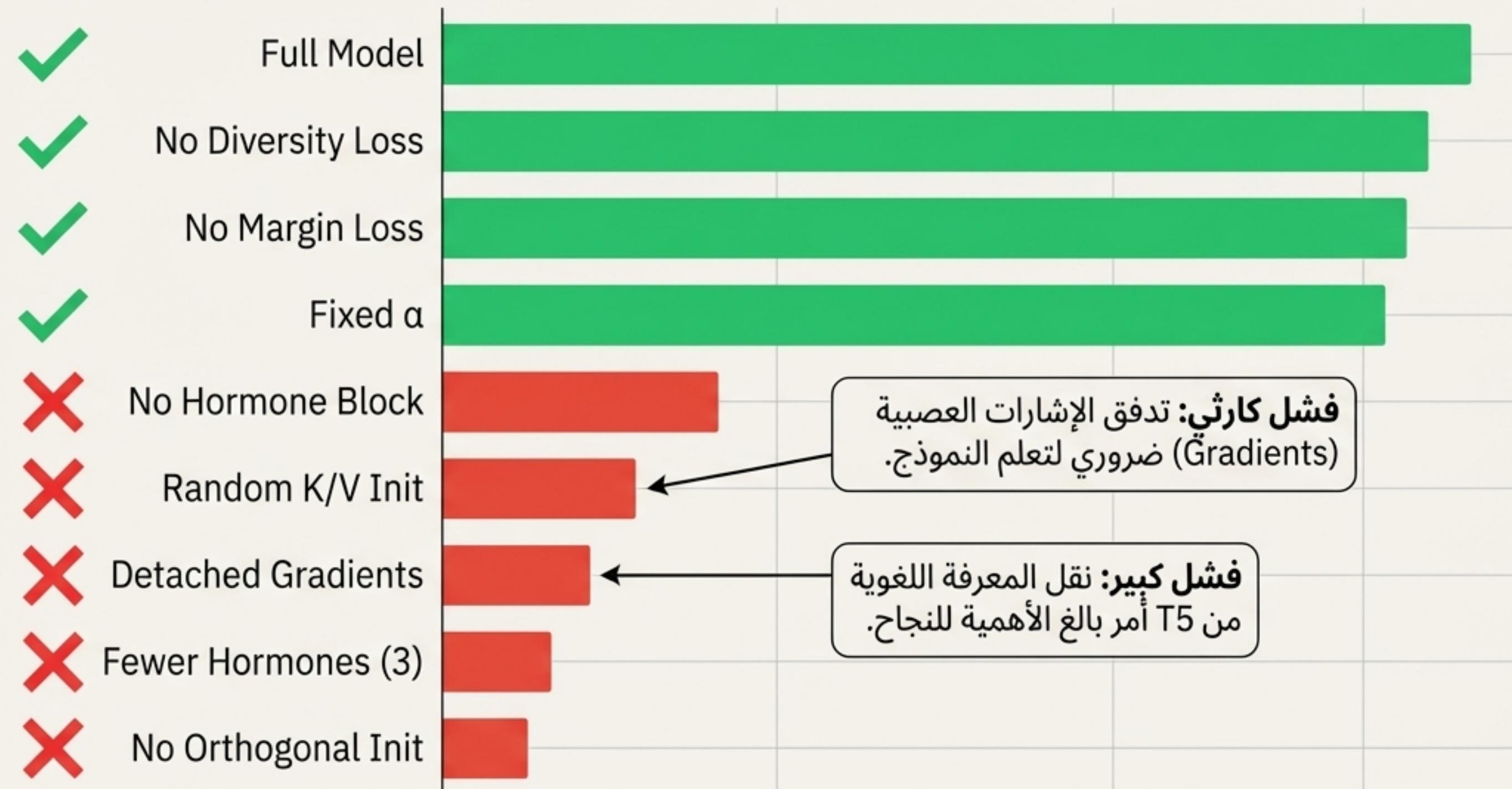


تقييمات مقياس ليكرت



.HormoneT5 فضلاً عن استجابة ٧٧٪ من المستخدمين

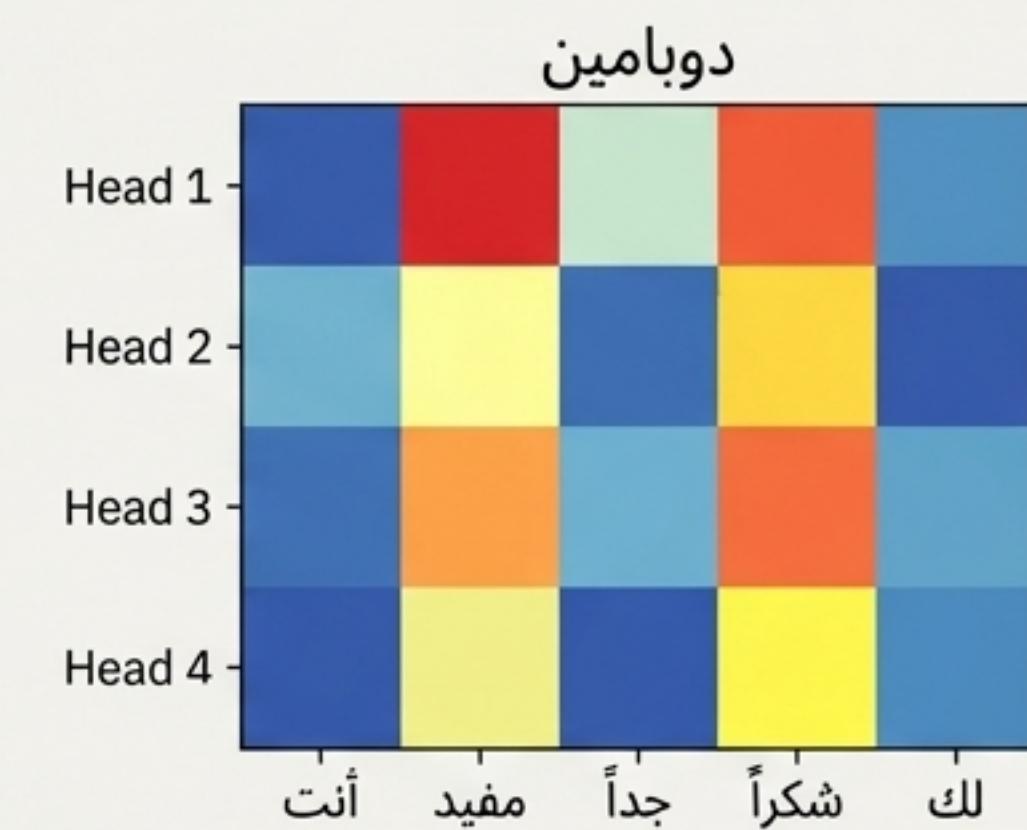
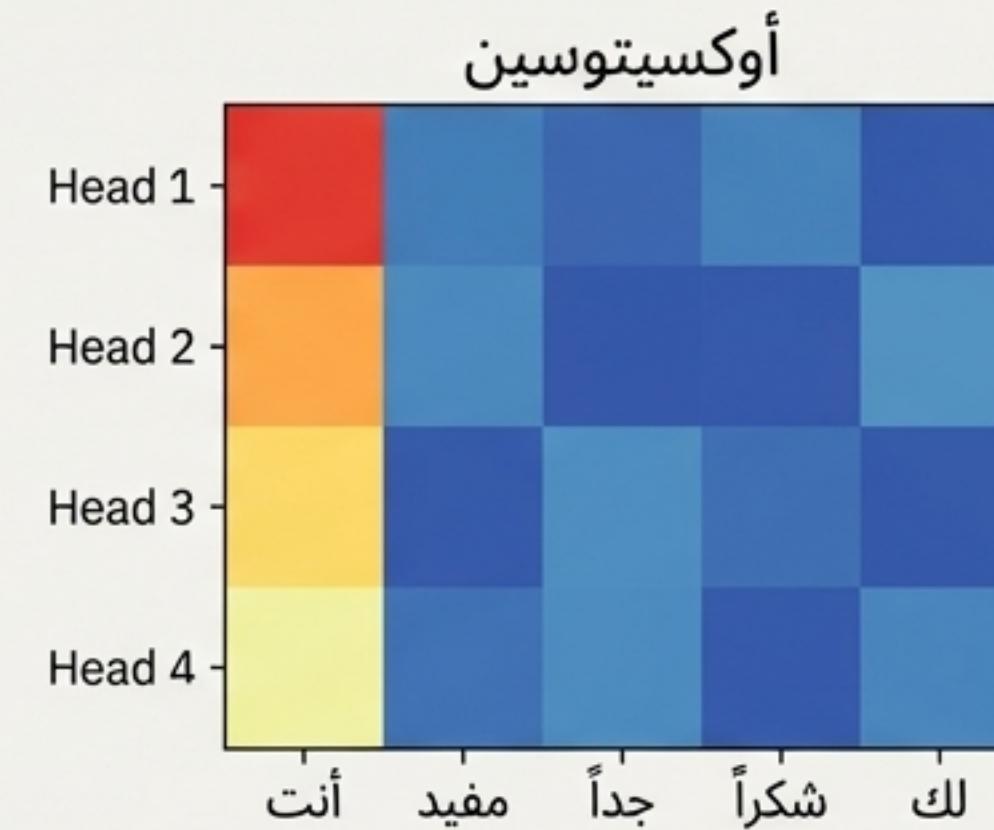
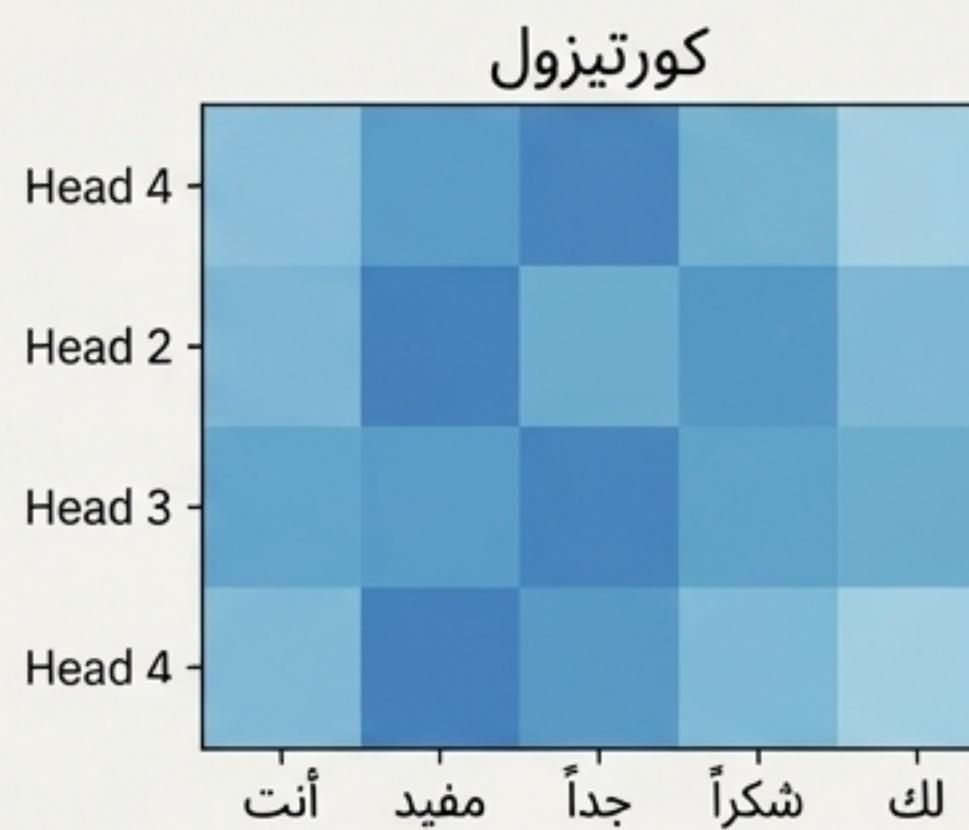
# تحليل المكونات: ما هي الأجزاء الأكثر أهمية لنجاح النموذج؟



النجاح ليس صدفة، بل يعتمد على خيارات معمارية محددة ومصممة بعناية.

# كيف "يرى" كل هرمون المدخلات: تحليل أنماط الانتباه

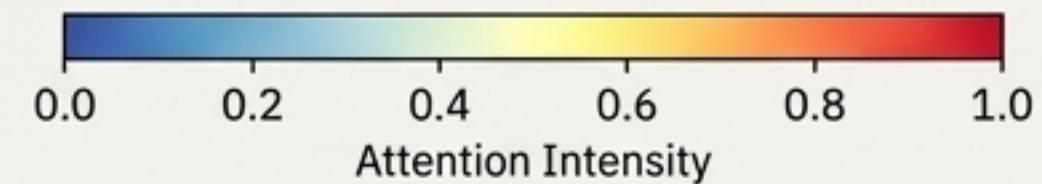
المدخلات: "أنت مفيد جداً، شكرأ لك!"



تنشيط ضئيل في غياب التهديد أو التوتر

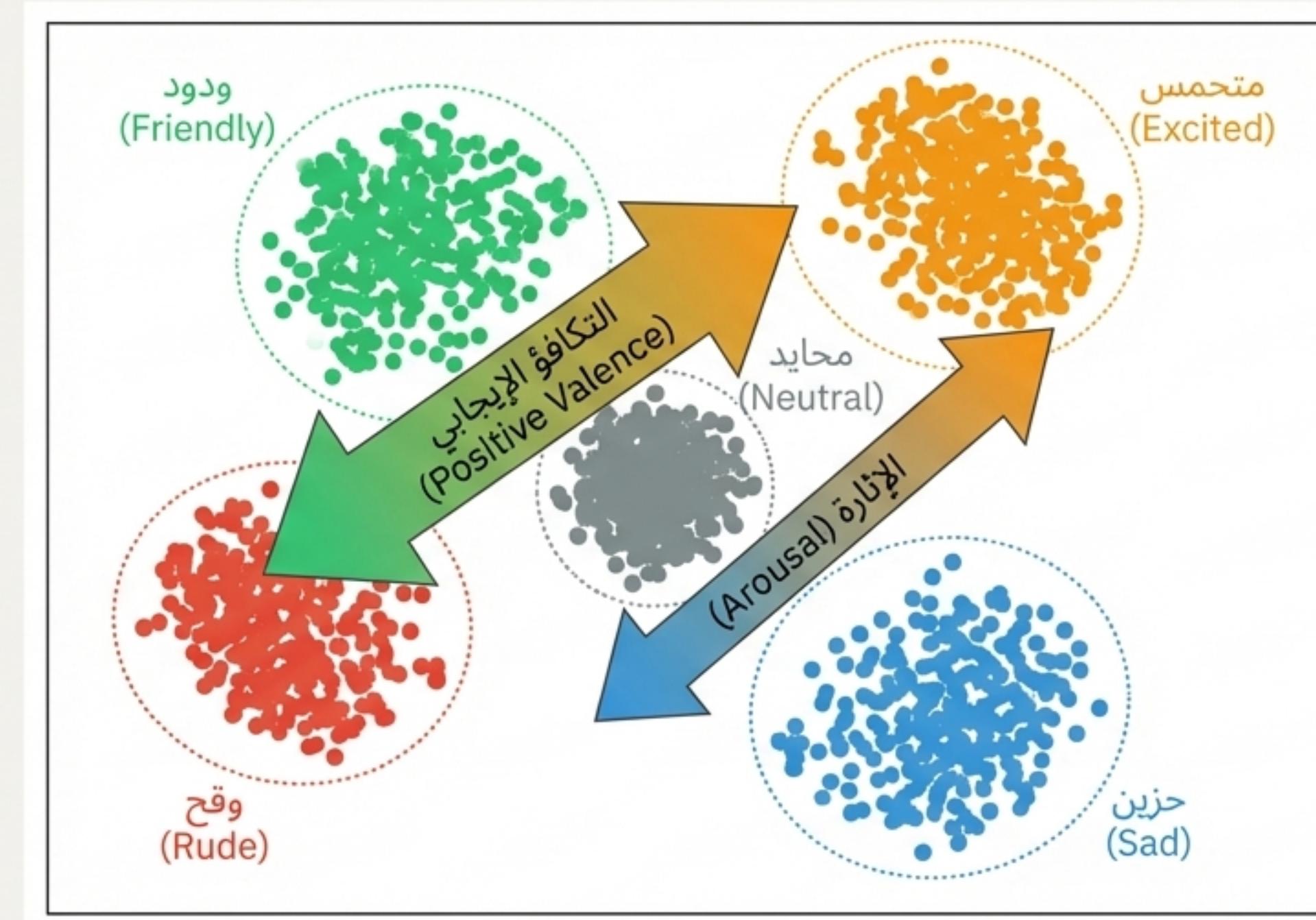
يركز على الضمائر التي تشير للارتباط الاجتماعي

يركز على كلمات المكافأة والتقدير



تعلم الهرمونات سلوكيات قابلة للتفسير ومتواقة مع أدوارها البيولوجية، مما يثبت صحة التصميم الأساسي.

# خريطة العواطف: النموذج يتعلم تمثيلات عاطفية مميزة وقابلة للفصل وقابلة للفصل

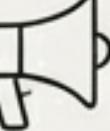


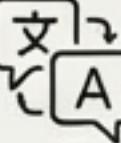
يثبت هذا أن الفضاء الهرموني سداسي الأبعاد يلتقط الحالات العاطفية ويفصل بينها بشكل فعال.

# القيود والتطوير المسؤول: إدراك المخاطر وبناء الضمانات

القيود

المخاطر المحتملة

- التلاعب العاطفي 
- التعاطف الزائف 
- تضخيم السلوك السام 

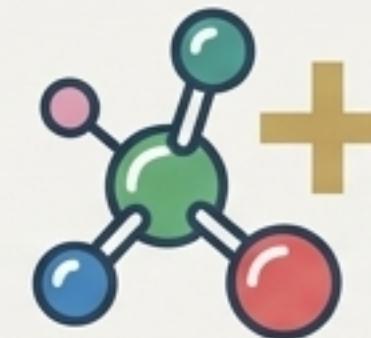
- حجم مجموعة البيانات 
- اللغة الإنجليزية فقط 
- التحيز الثقافي المحتمل 

الاعتراف بالقيود ومعالجة المخاطر الأخلاقية أمر ضروري لنشر الذكاء الاصطناعي المدرك عاطفياً بشكل مسؤول.

# الطريق إلى الأئمٌ: بناء جيل جديد من الذكاء الاصطناعي الواعي عاطفياً الواعي عاطفياً

## إضافة هرمونات جديدة

دمج هرمونات إضافية مثل النورايبينفرین وGABA لنمذجة أكثر ثراءً.



## ديناميكيات زمنية للهرمونات

محاكاة تراكم الهرمونات وتلاشيه بمرور الوقت في المحادثة.



## التكيف مع الثقافات المختلفة

تطوير نماذج خاصة بالثقافات المختلفة للاستجابات العاطفية المناسبة.



## تكامل الوسائل المتعددة

التوسيع ليشمل نبرة الصوت وتعبيرات الوجه.



هذه مجرد بداية لاتجاه بحثي جديد وواعد.

هذه مجرد بداية لاتجاه بحثي جديد وواعد.

# شكراً لكم



امسح للوصول إلى الكود والمشروع

**Full Paper Title:** A Hormone-Inspired Emotion Layer for Transformer Language Models (HELT)

**Author:** Eslam Reda

**Link to GitHub:**  
<https://github.com/eslam-reda-div/HELT>

**Citation Information:** Reda, E. A Hormone-Inspired Emotion Layer for Transformer Language Models (HELT). 2024.