

قياس جديد للذكاء الاصطناعي في الطب النفسي: PsychiatryBench يكشف عن ثغرات هامة

تأليف

مدرس الدكتور محمد لوتي

مايو 1, 2026

اقتبس من هذا المقال

مدرس الدكتور محمد لوتي (2026). قياس جديد للذكاء الاصطناعي في الطب النفسي: PsychiatryBench يكشف عن ثغرات هامة. عرب سايكولوجي. تم الاسترجاع من <https://arabpsychology.com/?p=121158>

هل يمكن للذكاء الاصطناعي أن يصبح طبيباً نفسياً؟ دراسة جديدة تقيّم قدرات النماذج اللغوية الكبيرة في مجال الصحة النفسية

أُتعرّف ذلك الشعور بالضيق، عندما تحاول فهم مشاعرك المعقدة، أو عندما تبحث عن كلمات تعبر عن ألم دفين؟ كم مرة تمنيت لو كان هناك من يستمع إليك دون حكم، ويفهمك بعمق، ويقدم لك الدعم والتوجيه؟ لطالما كان العلاج النفسي ملاذاً للباحثين عن السلام الداخلي، ولكن ماذا لو تمكنا من توسيع نطاق هذا الدعم، وجعله متاحاً للجميع، في أي وقت ومكان؟ هذا هو السؤال الذي يطرحه الباحثون مع ظهور تقنيات الذكاء الاصطناعي، وتحديدًا النماذج اللغوية الكبيرة (LLMs). ولكن هل هذه النماذج قادرة حقاً على فهم تعقيدات النفس البشرية، وتقديم مساعدة فعالة؟ دراسة حديثة بقيادة الدكتور فودة عبد الوهاب، تقدم لنا لمحة عن الإمكانيات والتحديات التي تواجه هذا المجال الناشئ.

الإطار النظري

علم النفس المعرفي والسلوكي: أسس الفهم

لفهم أهمية هذه الدراسة، يجب أن نلقي نظرة على الأسس النظرية التي يقوم عليها العلاج النفسي الحديث. يعتمد العلاج السلوكي المعرفي (CBT) على فكرة أن أفكارنا ومشاعرنا وسلوكياتنا مترابطة، وأن تغيير أنماط التفكير السلبية يمكن أن يؤدي إلى تحسين المزاج والسلوك. هذا يتطلب فهماً عميقاً للعمليات المعرفية، مثل الانتباه والذاكرة والتقييم. كما أن النظريات الديناميكية النفسية، التي أسسها فرويد وغيره، تركز على دور اللاوعي والخبرات المبكرة في تشكيل شخصيتنا وسلوكنا. يتطلب فهم هذه النظريات القدرة على تحليل القصص والحالات المعقدة، واستخلاص الأنماط الخفية. النماذج اللغوية الكبيرة، من خلال قدرتها على معالجة كميات هائلة من النصوص، قد تكون قادرة على محاكاة بعض جوانب هذه العمليات المعرفية، ولكن هل يمكنها حقاً أن تفهم المعنى الكامن وراء الكلمات؟

الذكاء الاصطناعي والتعلم العميق: كيف تعمل النماذج اللغوية الكبيرة؟

النماذج اللغوية الكبيرة هي في الأساس أنظمة إحصائية معقدة، مدربة على كميات هائلة من البيانات النصية. تعمل هذه النماذج عن طريق التنبؤ بالكلمة التالية في الجملة، بناءً على الكلمات السابقة. من خلال التدريب على ملايين الكتب والمقالات ومواقع الويب، تتعلم هذه النماذج أنماط اللغة، والقواعد النحوية، والمعلومات الواقعية. التعلم العميق، وهو فرع من فروع الذكاء الاصطناعي، يستخدم شبكات عصبية اصطناعية معقدة، تحاكي بنية الدماغ البشري، لمعالجة هذه البيانات. كلما زادت كمية البيانات التي يتم تدريب النموذج عليها، وكلما كانت الشبكة العصبية أكثر تعقيداً، زادت قدرته على فهم اللغة وتوليد النصوص. ولكن الفهم هنا يختلف عن الفهم البشري؛ فالنماذج اللغوية الكبيرة لا تمتلك وعياً أو تجارب شخصية، بل تعتمد على الأنماط الإحصائية التي تعلمتها.

المنهجية

بناء "PsychiatryBench": معيار جديد لتقييم النماذج اللغوية الكبيرة

أدرك الدكتور فودة وزملاؤه أن المعايير الحالية لتقييم النماذج اللغوية الكبيرة في مجال الطب النفسي كانت غير كافية. فمعظم هذه المعايير تعتمد على بيانات محدودة، مثل منشورات وسائل التواصل الاجتماعي أو حوارات اصطناعية، والتي لا تعكس التعقيد الحقيقي للممارسة السريرية. لذلك، قاموا بتطوير "PsychiatryBench"، وهو معيار جديد يعتمد حصرياً

على الكتب المدرسية وحالات الدراسة المعتمدة من قبل خبراء في الطب النفسي. يحتوي هذا المعيار على 5,188 سؤالاً، مقسمة إلى 11 مهمة مختلفة، تغطي جوانب متعددة من التشخيص والتخطيط العلاجي والمتابعة. تشمل هذه المهام: الاستدلال التشخيصي، وتخطيط العلاج، والمتابعة الطويلة الأجل، وإدارة الحالات، والتحليل المتسلسل للحالات، والاختيارات المتعددة.

تقييم النماذج اللغوية الكبيرة: مقارنة بين الأجيال

قام الباحثون بتقييم مجموعة متنوعة من النماذج اللغوية الكبيرة، بما في ذلك نماذج متطورة مثل Google Gemini و DeepSeek و Sonnet 4.5 و GPT-5، بالإضافة إلى نماذج طبية مفتوحة المصدر مثل MedGemma. استخدموا مقياس تقليدية لتقييم أداء النماذج، بالإضافة إلى طريقة مبتكرة تعتمد على استخدام نموذج لغوي كبير آخر كـ "حكم" لتقييم مدى تشابه إجابات النماذج المختلفة مع إجابات الخبراء. هذه الطريقة تسمح بتقييم أكثر دقة لجودة الإجابات، حيث أنها تأخذ في الاعتبار المعنى الكامن وراء الكلمات، وليس فقط الكلمات نفسها.

النتائج

أظهرت النتائج وجود فجوات كبيرة في الأداء السريري والسلامة بين النماذج اللغوية الكبيرة المختلفة. كانت النماذج الأقل تطوراً تعاني من صعوبة في الإجابة على الأسئلة المعقدة، أو كانت تقدم إجابات غير دقيقة أو غير آمنة. حتى النماذج الأكثر تطوراً أظهرت نقاط ضعف في مهام المتابعة وإدارة الحالات، مما يشير إلى أنها لا تزال غير قادرة على فهم الديناميكيات المعقدة للحالات النفسية على المدى الطويل. أكدت الدراسة على الحاجة إلى تدريب متخصص لهذه النماذج، وتطوير طرق تقييم أكثر قوة لضمان سلامة وفعالية استخدامها في مجال الصحة النفسية.

الآثار السريرية والعملية

تأثير النتائج على الأطباء النفسيين والمرضى والجمهور

هذه النتائج لها آثار مهمة على جميع الأطراف المعنية بالصحة النفسية. بالنسبة للأطباء النفسيين، يمكن أن تساعد هذه النماذج في تحسين دقة التشخيص، وتبسيط التوثيق السريري، وتقديم دعم علاجي إضافي للمرضى. بالنسبة للمرضى، يمكن أن توفر هذه النماذج وصولاً أسهل إلى المعلومات والدعم، خاصة في المناطق التي تعاني من نقص في الخدمات الصحية النفسية. بالنسبة للجمهور، يمكن أن تزيد هذه النماذج من الوعي بالصحة النفسية، وتقلل من الوصمة المرتبطة بالأمراض النفسية. ولكن يجب أن نكون حذرين من الاعتماد المفرط على هذه النماذج، وأن نتذكر أنها مجرد أدوات مساعدة، وليست بديلاً عن العلاج البشري.

السياق الثقافي العربي

التحديات والفرص في العالم العربي

عند النظر إلى هذه النتائج في سياق العالم العربي، نجد أن هناك تحديات وفرصاً فريدة. في العديد من البلدان العربية، لا يزال هناك نقص حاد في الخدمات الصحية النفسية، والوصمة المرتبطة بالأمراض النفسية لا تزال قوية. يمكن أن تلعب النماذج اللغوية الكبيرة دوراً هاماً في سد هذه الفجوة، وتوفير دعم نفسي أولي للمرضى الذين لا يستطيعون الوصول إلى الأطباء النفسيين. ومع ذلك، يجب أن نأخذ في الاعتبار الفروق الثقافية واللغوية عند استخدام هذه النماذج. فاللغة العربية

غنية بالتعبيرات والمصطلحات التي قد لا تكون مفهومة للنماذج اللغوية الكبيرة المدربة على اللغات الغربية. كما أن القيم والمعتقدات الثقافية العربية قد تؤثر على طريقة تعبير المرضى عن مشاعرهم وأعراضهم. لذلك، من الضروري تدريب هذه النماذج على بيانات عربية أصيلة، وتكييفها مع السياق الثقافي المحلي. بالإضافة إلى ذلك، يجب أن نضمن أن هذه النماذج تحترم القيم الثقافية العربية، ولا تقدم معلومات أو توصيات تتعارض معها.

آفاق مستقبلية وقيود الدراسة

ماذا بعد؟

هذه الدراسة هي مجرد بداية. هناك حاجة إلى مزيد من البحث لتطوير نماذج لغوية كبيرة أكثر دقة وسلامة وفعالية في مجال الصحة النفسية.

Recommended Academic Training

:Deepen your knowledge with these specialized courses from our Academy

أخلاقيات مهنة التعليم View Course → أصول التربية والتعليم View Course → التربية البيئية والتنمية المستدامة View Course →

Reference

Fouda A.E. (2026). *PsychiatryBench: a multi-task benchmark for LLMs in psychiatry*. npj Digital Medicine, 9(1).

DOI: [10.1038/s41746-026-02582-w](https://doi.org/10.1038/s41746-026-02582-w)