ذكاء اصطناعي قابل للتفسير وشبكات الحالة الصدى: تعزيز الثقة في التفاعل بين الإنسان والآلة

تأليف

مدرس الدكتور محمد لوتي

مارس 29, 2026

تخيل أنك تعتمد على نظام ذكاء اصطناعي لاتخاذ قرارات مهمة، سواء في مجال الرعاية الصحية أو القيادة الذاتية. هل ستثق به بشكل أعمى، أم أنك ستطلب تفسيراً لكيفية وصوله إلى هذا القرار؟ هذا السؤال يطرح نفسه بقوة مع تزايد الاعتماد على الذكاء الاصطناعي في حياتنا اليومية، ويشكل جوهر بحث جديد يهدف إلى فهم كيفية بناء الثقة بين الإنسان والآلة.

منهجية البحث

قام باحثون بدراسة متعمقة حول دور الذكاء الاصطناعي القابل للتفسير (XAI) وشبكات الحالة الصدى (ESNs) في معايرة الثقة في التفاعل بين الإنسان والآلة. استخدم الباحثون، بقيادة هاو إس، مجموعتي بيانات قياسيتين: CIFAR-10 للمهام المرئية (التعرف على الصور) و SQuAD للمهام النصية (فهم النصوص والإجابة على الأسئلة). اعتمدت الدراسة على تصميم تجريبي بين المجموعات (2 × 2)، حيث تم فحص تأثير قابلية تفسير الذكاء الاصطناعي (ذكاء اصطناعي قابل للتفسير مقابل ذكاء اصطناعي غير قابل للتفسير) ونتائج التفاعل (نجاح أو فشل) على مقاييس الثقة الصريحة (التقارير الذاتية) والضمنية (السلوكيات غير اللفظية).

لزيادة شفافية النموذج، تم دمج الشبكات العصبية التلافيفية (CNNs) - وهي نوع من الذكاء الاصطناعي يتفوق في معالجة الصور ـ مع تقنيات الذكاء الاصطناعي القابل للتفسير. تم استخدام تقنية Grad-CAM لتوفير تفسيرات مرئية للقرارات المتخذة في المهام المرئية، بينما تم استخدام آليات الانتباه (attention mechanisms) لتفسير القرارات في المهام النصية. آليات الانتباه تحدد الأجزاء الأكثر أهمية في النص التي اعتمد عليها الذكاء الاصطناعي لاتخاذ قراره. بهذه الطريقة، سعى الباحثون إلى جعل عملية اتخاذ القرار أكثر وضوحًا وقابلة للفهم للمستخدمين.

النتائج

أظهرت نتائج البحث أن الذكاء الاصطناعي القابل للتفسير يلعب دورًا هامًا في تعديل مستويات الثقة، خاصة في حالات الفشل عندما يحتاج المستخدمون إلى فهم سبب عدم نجاح النظام. عندما تم تزويد المستخدمين بتفسيرات منطقية لقرارات الذكاء الاصطناعي، زادت ثقتهم في النظام، حتى في حالة حدوث أخطاء. التفسيرات القائمة على الشبكات العصبية التلافيفية (CNNs) حسنت الفهم، وبالتالي بناء الثقة، من خلال توفير أدلة مرئية غالبًا ما تكون أسهل في الفهم. على سبيل المثال، إذا فشل نظام التعرف على الصور في تحديد كائن ما، فإن تقنية Grad-CAM يمكن أن تسلط الضوء على المناطق في الصورة التي ركز عليها النظام، مما يساعد المستخدم على فهم سبب الخطأ.

على الرغم من وجود ارتباط قوي بين مقاييس الثقة الصريحة والضمنية، إلا أن المقاييس الضمنية كشفت عن اختلافات في ديناميكيات الثقة لا يمكن التحقق منها باستخدام التقارير الذاتية. وهذا يشير إلى أن سلوك المستخدمين قد يكشف عن مستويات ثقة مختلفة عن تلك التي يعبرون عنها بشكل مباشر. من الجدير بالذكر أن العوامل الديموغرافية، مثل الجنس، لم يكن لها تأثير كبير على الثقة، مما يسلط الضوء على التنوع في هذه الأساليب.

كما قارنت الدراسة أدائها بخمس طرق حديثة أخرى من حيث الدقة ومعايرة الثقة ورضا المستخدم، وأظهرت تفوقها في هذه الجوانب، مع الحفاظ على كفاءة حسابية جيدة.

دلالات البحث

تؤكد هذه النتائج على أهمية قابلية التفسير ومعايرة الثقة الديناميكية في ضمان تطوير أنظمة ذكاء اصطناعي موثوقة. إن القدرة على فهم كيفية عمل الذكاء الاصطناعي، خاصة عندما يرتكب أخطاء، أمر بالغ الأهمية لبناء الثقة وتشجيع التبني

الواسع النطاق لهذه الأنظمة في مختلف المجالات. هذا البحث يفتح الباب أمام تطوير أنظمة ذكاء اصطناعي أكثر شفافية وموثوقية، مما يعزز التفاعل الإيجابي بين الإنسان والآلة. إن توفير تفسيرات واضحة ومفهومة للقرارات التي يتخذها الذكاء الاصطناعي لا يساعد المستخدمين على فهم النظام فحسب، بل يسمح لهم أيضًا بتقييم موثوقيته واتخاذ قرارات مستنيرة بناءً على ذلك. هذا يمثل خطوة مهمة نحو مستقبل حيث يمكن للذكاء الاصطناعي أن يكون شريكًا موثوقًا به في حياتنا اليومية.

Reference

Hao S. (2026). *Explainable AI and echo state networks calibrate trust in human machine interaction*. Scientific Reports, 16(1), 1189-1189.

DOI: 10.1038/s41598-025-30899-1