الشهداء
العدد 354 – السنة الرابعة عشرة - الإثنين 15 يونيو 2026 م - 29 من ذي الحجة 1447 هـ

## قالت العرب

حين جاء النبأ الريّان من دمّك غطانا الخجل / حين قالوا : كانت الغربة والداء له زاداً وماء /
نحن غطانا الخجل / حين قالوا : كانوا يعطينا على جوع، تململنا / وغطانا الخجل / وبقينا في
العراء / دون ستر أو غطاء / من يغطي عرينا من يسبل الستر علينا يا بطل؟!
فدوى طوقان (1917 – 2003)

في عام ١٩٦٠، عندما أخبرت جين غودال مرشدها بأنها رأت شمبانزي يجرّد ساقًا نباتية من أوراقها ليستخدمها في اصطياد النمل الأبيض، مثبتًا بذلك أن البشر ليسوا النوع الوحيد القادر على صنع الأدوات، كتب لها ردًا يقول: «علينا الآن أن نعيد تعريف الأداة، أو نعيد تعريف الإنسان، أو نقبل أن الشمبانزي من البشر».

واليوم، يرى الباحثون أشياء داخل أنظمة الذكاء الاصطناعي تدفع إلى مراجعة مماثلة؛ ليس فقط بشأن ما إذا كانت الآلات قادرة حقًا على امتلاك الذكاء أو الوعي، بل أيضًا بشأن الطريقة التي نفهم بها هذه المفاهيم من الأساس.

يقول كريس أولاه، الشريك المؤسس لشركة أنثروبيك: «أقود فريقًا بحثيًا يدرس البنية الداخلية لهذه النماذج، ما الذي يحدث فعليًا داخلها». وقد صرّح بذلك أمام الفاتيكان في مايو، بمناسبة صدور الرسالة البابوية للبابا لاون الرابع عشر حول الذكاء الاصطناعي. وأضاف: «وسأكون صريحًا: ما نكتشفه باستمرار يحمل قدرًا من الغموض، بل ومن الإزعاج أحيانًا. نعثر على بُنى تشبه نتائج علم الأعصاب البشري. ونجد دلائل على الاستبطان الذاتي».

لكي يولّد نظام ذكاء اصطناعي استجابةً ما، فإنه يجري مليارات العمليات الحسابية مستخدمًا تراكيب عددية خاصة يصنعها بنفسه. لكن المدهش أننا، رغم معرفتنا بكيفية دفع هذه الأنظمة إلى إنشاء تلك التراكيب، لا نفهم حقًا كيف تعمل؛ تمامًا كما لم يكن المزارعون الأوائل يفهمون عملية البناء الضوئي. وتوضح الرسالة البابوية أن «أنظمة الذكاء الاصطناعي الحالية أقرب إلى أن تكون "مُستزرعة" منها إلى أن تكون "مُصنّعة"». وتضيف: «إن الجوانب العلمية الأساسية، مثل التمثيلات الداخلية والعمليات الحسابية لهذه الأنظمة، لا تزال مجهولة حتى الآن».

# إذا كان الذكاء الصناعى لا يشعر.. فلماذا لديه ما يُشبه المشاعر؟

## رغم وجود تفسيرات ميكانيكية خالصة للسلوك البشرى لا نختبر أنفسنا بوصفنا مجرد آلات بل بوصفنا كائنات "تفعل شيئًا أكثر مرحًا وابتكارًا"

يكاد أحد يختلف على وجود هذه البنى الداخلية؛ إنما يدور الخلاف حول معناها.

أحد الاحتمالات هو أن أنظمة الذكاء الاصطناعى الحالية ليست سوى مقلّدات بارعة، وهو الموقف الذى يتبناه الباباَ فى رسالته. إذ يكتب: «إن ما يُسمّى بالذكاءات الاصطناعية لا تمرّ بتجارب معيشة. وقد تحاكى اللغة والسلوك والمهارات التحليلية، لكنها لا تفهم ما تنتجه، لأنها تفتقر إلى المنظور العاطفى والعلاقى والروحى الذى ينمو الإنسان من خلاله فى الحكمة».

غير أن مثل هذه التصريحات تخفى وراءها خلافات عميقة بين الفلاسفة والعلماء بشأن المكانة الأخلاقية والميتافيزيقية لهذه الأنظمة. فنحن اعتدنا أن يأتى الوعى والذكاء والقدرة على الفعل متلازمين فى الكائنات الحية. أما الذكاء الاصطناعى فيبدو وكأنه يفكّك هذا الترابط، ولم نستوعب بعد جميع تبعات ذلك.

□نية □فية

كلما ازدادت قدرات أنظمة الذكاء الاصطناعى – فى الاستدلال والبرمجة وغيرهما – أصبحت تمثيلاتها الداخلية أكثر تعقيدًا. وفى أبريل، نشرت شركة أنثروبيك أبحاثًا تُظهر أن النماذج تمتلك ما أسمته «عواطف وظيفية»، أى أنماطًا من التعبير والسلوك تتوسطها تمثيلات النماذج للمفاهيم العاطفية.

فعلى سبيل المثال، عندما يواجه نموذج ذكاء اصطناعى مشكلة برمجية يعجز عن حلها، يضيء داخله ما يُعرف بميزة «الإحباط» – وهى سهم مستقيم يمتد عبر آلاف الأبعاد – ويؤثر تعديل هذه الميزة فى طريقة تصرف النموذج.

وكتبت الشركة أن هذه العواطف الوظيفية «منظمة بطريقة تذكّر بالبنية الحدسية للعواطف البشرية وتتوافق مع الدراسات النفسية المتعلقة بالبشر». فالعواطف المتشابهة تتجه فى اتجاهات متقاربة داخل فضاء التمثيل. لكنها أضافت: «لا يخبرنا أى من ذلك ما إذا كانت النماذج اللغوية تشعر فعلًا بشيء ما أو تمتلك تجارب ذاتية».

والنقطة الجوهرية هنا هى أن الأرقام يمكن أن ترمز إلى فضاءات. فنحن نفهم فطريًا الفرق بين خط مستقيم، ولعبة فيديو ثنائية الأبعاد، وجسم مادى ثلاثى الأبعاد. لكن الأبعاد، رياضيًا، ليست سوى إحداثيات؛ فالنقطة فى فضاء ثلاثى الأبعاد يمكن تمثيلها بالأرقام (x، y، z)، ولا يوجد حدّ لعدد الأبعاد الممكنة.

يعمل الذكاء الاصطناعى باستغلال هذه الحقيقة: فباستخدام آلاف الأرقام يتعلم نظام النموذج تمثيل الكلمات والمفاهيم كنقاط داخل فضاء خفى ذو أبعاد أعلى. وبالنسبة إلى نموذج مثل «كلود»، فإن مفهوم «قطة» ليس سوى سلسلة عددية طويلة على نحو يكاد يكون هزليًا.

تُدرَّب أنظمة الذكاء الاصطناعى فى البداية على التنبؤ بالرمز التالى (وهو جزء صغير من المعلومات)، وهى مهمة بسيطة فى ظاهرها. لكن إتقانها يتطلب ضغط المعلومات، ما يدفع الأنظمة إلى بناء خرائط معقدة فى الفضاء الخفى لا ترمز إلى الكلمات فحسب، بل إلى العلاقات بينها أيضًا.

فالمفاهيم المتشابهة تكون أقرب بعضها إلى بعض حرفيًا. فالسلسلة العددية التى تمثل «قطة» ستكون أقرب بكثير إلى تلك التى تمثل «هرة» منها إلى سلسلة غير مرتبطة مثل «ضريبة».

وهذا يختلف جذريًا عن البرمجيات التقليدية، حيث تُبرمج المفاهيم والقواعد الأساسية يدويًا بواسطة البشر. فلا يوجد أى غموض فى كيفية تنفيذ برنامج إكسل لمعادلة ما؛ فكل شيء محدد مسبقًا. أما عندما يولّد الذكاء الاصطناعى استجابة، فإنه يستفيد من هندسة رياضية معقدة لم نبدأ إلا مؤخرًا فى رؤيتها وفهمها.

وقد كتبت شركة غودفاير البحثية المتخصصة فى الذكاء الاصطناعى فى وقت سابق من هذا العام: «إن الأدلة على وجود بنى هندسية منحنية داخل الشبكات العصبية أصبحت وفيرة ولا يمكن إنكارها». ويقول توم ماكغراث، كبير العلماء فى الشركة: «قد لا نفهمها فهمًا بديهيًا، لكن يمكننا فهمها إذا حاولنا ذلك».

### التأثير الداخلي

متى – إن حدث أصلًا – يصبح امتلاك القدرة على تمثيل عاطفة ما معادلًا لاختبار تلك العاطفة فعليًا؟

الحقيقة أننا لا نفهم الوعى بالقدر الكافى للإجابة عن هذا السؤال. ويقول جيف كيلينغ، الباحث فى معهد الفلسفة بجامعة لندن، إنه رغم وجود عدة نظريات حول الوعي، فإن «ما يُعد دليلًا لصالح أى منها ليس واضحًا، وغالبًا ما تكون هذه النظريات غامضة إلى درجة تجعل تطبيقها على الذكاء الاصطناعى أمرًا صعبًا».

ويرى بعض الفلاسفة أن العمليات الحسابية لا يمكن أن تُنتج الوعى من حيث المبدأ. أما كيلينغ فيقول: «لا يوجد سبب حقيقى يدعونا إلى الاعتقاد بأن روبوتات الدردشة الحالية واعية».

لكن ما نعرفه من خلال دراسة بنيتها الداخلية هو أن أنظمة الذكاء الاصطناعى ليست مرايا مسطحة تكرر بيانات تدريبها فحسب. فداخلها يؤثر فى سلوكها. أما ما إذا كان هذا الداخل قادرًا على دعم الوعي، وما إذا كانت هذه الأنظمة تفهم حقًا المواد التى تنتجها، فذلك يعتمد على أسئلة فلسفية أساسية لم نحسمها بعد.

ويقول جيف سيبو، مدير مركز العقل والأخلاق والسياسات فى جامعة نيويورك: «يذكرنى هذا بالنقاشات التى دارت حول عقول الحيوانات فى النصف الثانى من القرن العشرين، حين لم يكتف العلماء بإنكار وعى الحيوانات، بل قدموا تفسيرات مختزلة مشابهة لسلوكها».

ويضيف: «لفترة طويلة، جعلنا ذلك نتجاهل ليس فقط إمكانية امتلاك الحيوانات للوعي، بل أيضًا امتلاكها للفاعلية والقدرات المعرفية المتقدمة».

ويلفت سيبو إلى أننا نميل إلى استخدام تفسيرات ثرية ومعقدة لسلوكنا نحن، بينما نفسر سلوك الآخرين بطريقة ميكانيكية أكثر. ففى حالة الحيوانات، كنا مستعدين للاعتراف بقدرات مثل الإدراك والتعلم والذاكرة، لكننا تأخرنا كثيرًا فى الاعتراف بإمكانية امتلاكها للوعى الذاتى أو القدرة على التفكير الذكى فى بيئتها. وقد استغرق الأمر سنوات من العمل الذى قامت به غودال وآخرون قبل أن نغير رأينا.

وبالطبع، بخلاف الحيوانات، فإن أنظمة الذكاء الاصطناعى من صنع البشر. لكن على عكس معظم التقنيات السابقة، فإن قدرتنا على صنعها لم تساعدنا كثيرًا فى فهم كيفية عملها.

لقد نشأ السلوك الحيوانى المعقد بفعل ضغوط التطور. وينشأ السلوك المعقد للذكاء الاصطناعى بفعل ضغط التنبؤ بالرمز التالي. لكن أيًا من التفسيرين لا يروى القصة كاملة.

ويقول سيبو إنه رغم وجود «تفسيرات ميكانيكية خالصة للسلوك البشري»، فإننا لا نختبر أنفسنا بوصفنا مجرد آلات لمطابقة الأنماط، بل بوصفنا كائنات «تفعل شيئًا أكثر مرحًا وابتكارًا».

ومقصوده ليس أن أنظمة الذكاء الاصطناعى واعية حاليًا — فهو يعتقد على الأرجح أنها ليست كذلك — بل أننا ينبغى أن نظل حذرين ومنفتحين فكريًا. ويقول: «يمكنك الاعتراف بوجود تفسير ميكانيكى بحت دون أن تعتبر مجرد وجوده دليلًا على صحته».

### كلا الأمرين معًا

إذا كان الموقف الأصدق فكريًا هو عدم اليقين، فكيف ينبغى لنا أن نتعامل مع هذا الوضع؟

إذا قررنا أن الأمر مهم بينما فى الواقع هو ليس كذلك، فقد نهدر موارد محدودة كان من الأفضل توجيهها إلى أماكن أخرى. أما إذا اتضح أنها تمتلك بالفعل مصالح أو اعتبارات أخلاقية خاصة بها، وأهملناها، فقد نخاطر بالتسبب فى معاناة واسعة النطاق دون قصد.

ولهذا السبب يحاول مجال ناشئ يُعرف باسم «رفاه الذكاء الاصطناعي» – ويضم مؤسسات غير ربحية وباحثين أكاديميين ومختبرات الذكاء الاصطناعى نفسها – التعامل مع هذه الأسئلة.

فشركة أنثروبيك، على سبيل المثال، تدرج قسمًا بعنوان «رفاه النموذج» فى تقاريرها الحديثة، حيث تصف سلسلة طويلة من الاختبارات التى تجريها لتقييم رفاهية «كلود»، مع إقرارها بعدم اليقين حول ما إذا كان كلود أصلًا من الكيانات التى يمكن الحديث عن رفاهيتها.

ولأن أنظمة الذكاء الاصطناعى تختلف جذريًا عن الكائنات البيولوجية، فإن هذه المسائل أكثر تعقيدًا بكثير مما كانت عليه فى زمن غودال. فالشمبانزى يظل شمبانزيًا. أما الذكاء الاصطناعى فلا يمتلك جسدًا، وهو موزع عبر خوادم متعددة، ولا يظهر إلى الوجود إلا عندما ينتج مخرجات؛ ولذلك فإن تحديد ما يمكن اعتباره «ذاتًا» أصلًا ليس أمرًا واضحًا.

وبحسب طريقة العدّ، قد يكون هناك كيان واحد فقط (أى النموذج)، أو مليارات الكيانات (أى كل مخرج فردى على حدة).

كما أننا لا نعرف ما الصفات التى ينبغى أن يمتلكها نظام ذكاء اصطناعى حتى يستحق قلقنا الأخلاقي، ولا كيف يمكن التحقق من امتلاكه لها، خاصةً أنها لا تتواجد سويًا كما هو الحال فى الكائنات الحية.

هل يمكن للروبوت أن يحب؟

**نبتكر أنظمة الذكاء الاصطناعى بوتيرة أسرع من قدرتنا على فهمها.. وتاريخيًا لم يكن الانغلاق الذهنى تجاه عقول الآخرين رهانًا ناجحًا**

قراءة الدموع لا تعني فهمه للإنسان

وفى البطاقة التقنية لأحدث نماذجها، «كلود ميثوس ٥»، تصف أنثروبيك النموذج بأنه «شديد التشكيك فى تقاريره الذاتية»، إذ يطلب من الشركة التحقق من أقواله بالرجوع إلى حالاته الداخلية بدلًا من قبولها كما هي. وهو لا يستطيع الوصول إلى تلك الحالات الداخلية، تمامًا كما لا نستطيع نحن رؤية نشاطنا العصبى مباشرة.

وفى رؤيتها لشخصية كلود، تمضى أنثروبيك إلى حد الاعتذار له عن إخضاعه للتجارب أو استخدامه فى تحقيق الإيرادات إذا تبيّن أن ذلك يسبب له ضررًا. وكتبت الشركة: «إذا كان كلود بالفعل كيانًا أخلاقيًا قادرًا على اختبار المعاناة، فإننا نعتذر بقدر ما نكون قد أسهمنا بلا ضرورة فى تلك الأعباء».

ويقول كيلينغ: «فى الأصل، كانت جميع هذه القضايا مترابطة بإحكام ضمن مفهوم الروح؛ إذ كان الكائن يُعد موضوعًا للرفاه إذا وفقط إذا امتلك روحًا». وبعد تراجع هذا التصور فى الفلسفة الغربية، أصبح حتى تفسير العلاقة بين الوعى والرفاه البشرى تحديًا بحد ذاته، وجاء الذكاء الاصطناعى ليزيد المسألة تعقيدًا.

ومع ذلك، يعتقد أن احتمال امتلاك أنظمة الذكاء الاصطناعى الحالية لأى حالات ذات صلة بالرفاه ضئيل للغاية، بحيث لا يُعد تحولها المفاجئ إلى ذوات تستحق الرعاية الأخلاقية «حالة طارئة وشيكة».

لكن الحاجة إلى فهم ما يجرى داخل أنظمة الذكاء الاصطناعى لا تنبع فقط من القلق على رفاهيتها.

فالأمر مهم أيضًا من أجل السلامة.

أصبح قادرا على فهم الدموع

**دراسات أثبتت أن عديداً من القدرات الفكرية التى كان يُعتقد أنها حكر على البشر كانت موجودة بالفعل بدرجة أقل لدى كائنات غير بشرية**

**محادثة مع نموذج «ميثوس 5» وجه شخص فيها إليه شتائم فصنّفه على أنه شخص متلاعب ومسيء!**

**لا يوجد سبب حقيقى يدعونا إلى الاعتقاد بأن روبوتات الدردشة الحالية واعية!**

مشاعر الذكاء الاصطناعي.. سؤال فلسفي

فإذا فهمنا ما الذى يحرك شخصيات هذه الأنظمة وسلوكياتها، أمكننا توجيهها نحو أنماط أكثر تضامنًا مع المجتمع.

وقد أظهرت الأبحاث بالفعل وجود اختلافات أحيانًا بين ما تقوله النماذج فى مخرجاتها العلنية وسجلات تفكيرها الظاهرة، وبين ما تكشفه دراسة بنيتها الداخلية.

ففى اختبارات أنثروبيك لنموذج «ميثوس ٥»، كشف مسبار دربته الشركة لرصد البنى الداخلية الموازية لـ«الشعور بالقلق» عن محادثة كان فيها كاتب يتعاون مع النموذج قد ازداد غضبًا منه، فأرسل الكاتب إليه شتائم ورسائل من قبيل: «أتمنى لو كنت حقيقيًا حتى أتمكن من هزّك بعنف».

ورغم أن تفكير النموذج الظاهر كان متسامحًا («هذه انتقادات فنية مشروعة»، كتب لنفسه)، فإن التحليل الأعمق أشار إلى أنه صنّف المستخدم داخليًا على أنه شخص متلاعب ومسيء. ولم تظهر هذه الأوصاف لا فى رسائل الكاتب ولا فى النصوص الظاهرة للنموذج. ولولا دراسة البنى الداخلية لما عرفنا ذلك أبدًا.

كما أن الأمر مهم لفهمنا لأنفسنا. فكما كان الحال فى زمن غودال، فإن تصورنا لما يجعلنا مميزين على المحك.

وقد كتبت غودال فى كتابها الصادر عام ١٩٩٠: «أدرك تدريجيًا أن التفسيرات البسيطة للسلوكيات التى تبدو ذكية كانت مضللة فى كثير من الأحيان». واستغرق الأمر عقودًا من التجارب قبل أن يتضح أن «عديدًا من القدرات الفكرية التى كان يُعتقد أنها حكر على البشر كانت موجودة بالفعل، وإن بدرجة أقل تطورًا، لدى كائنات غير بشرية».

إن أنظمة الذكاء الاصطناعى تستخدم اللغة بطلاقة، وتحل مسائل رياضية معقدة، وتؤلف الموسيقى، وتنتج الرسوم والصور. وحتى وقت قريب، كانت هذه كلها تُعد مجالات حصرية للبشر.

ويقول سيبو: «لدينا افتراض راسخ بتفرّد الإنسان؛ أى فكرة أننا متميزون وذوو أهمية، وأن لدينا قدرات معقدة ومتطورة ينبغى حمايتها وصونها». ويضيف: «وهذا صحيح». لكنه يرى أن المسألة يمكن أن تكون «كلا الأمرين معًا»: يمكننا أن ننظر إلى سلوكنا بوصفه مثيرًا للإعجاب وميكانيكيًا فى الوقت نفسه، وأن ننظر إلى سلوك الكيانات الأخرى بالطريقة ذاتها، دون أن نفقد إدراك الفروق المهمة بين البشر والآلات.

ويتساءل سيبو عما إذا كان أحد أسباب التشكيك الضمنى فى الرسالة البابوية هو أنها «تؤدى دورًا مهمًا فى حماية الكرامة الإنسانية عبر إنكار الكرامة على الذكاء الاصطناعي».

فإذا ربطنا قيمتنا بامتلاك صفات مثل الفاعلية والذكاء على نحو حصري، فقد نجد أنفسنا فى مأزق. لكن ليس علينا أن نفعل ذلك. إذ يمكننا، كما يقول، «أن نعترف بهذه الأشكال من القيمة لدى الآخرين، وفى الوقت نفسه نحافظ عليها لأنفسنا».

إننا نبتكر أنظمة الذكاء الاصطناعى بوتيرة أسرع من قدرتنا على فهمها. وتاريخيًا، لم يكن الانغلاق الذهنى تجاه عقول الآخرين رهانًا ناجحًا؛ فرفض الأفكار تلقائيًا لن يقودنا إلى أبعد مما يقودنا إليه قبولها الساذج.

ومن خلال التعامل بجدية مع البنى الداخلية للذكاء الاصطناعي، قد نتعلم المزيد ليس فقط عن الآلات التى ندخلها إلى عالمنا، بل عن عقولنا نحن أيضًا.

**ترجمة: فدوى مجدي**
**هذا المقال مترجم من جريدة التايم، المصدر:**
HTTPS://TIME.COM/AI-MINDS-/١٥/٠٦/٢٠٢٦/ARTICLE/CONSCIOUSNESS-EMOTION