

داريو أمودي

تسارع تطور الذكاء الاصطناعي ومخاطره

قراءة وتحليل

د. حسن هاشم حمود

باحث في مركز الفيض العلمي لاستطلاع الرأي والدراسات المجتمعية



مقدمة

أدى التنافس المتسارع بين الشركات والدول في مجال الذكاء الاصطناعي إلى التركيز على تحقيق التفوق التقني والتجاري وتوسيع نطاق استخدام هذه التكنولوجيا في مختلف مجالات الحياة، ومع اتساع قدرات الذكاء الاصطناعي ودخوله في الاقتصاد والأمن والاتصالات والبحث العلمي والخدمات، أصبح التفوق فيه مرتبطاً بدرجة أكبر بالنفوذ الاقتصادي والاستراتيجي، الأمر الذي يدفع الشركات إلى تسريع تطوير نماذجها وقدراتها للحفاظ على موقعها التنافسي.

غير أن هذا التنافس المتسارع قد يخلق فجوة بين سرعة تطوير القدرات وسرعة تطوير وسائل السلامة والحوكمة. فكل شركة قد تواجه حافزاً للاستمرار في تطوير قدراتها خشية أن يؤدي التباطؤ من جانبها إلى منح منافسيها أفضلية، ومن هنا تنشأ مشكلة معضلة التنسيق حتى عندما تدرك الشركات مخاطر التسارع، قد تجد صعوبة في خفض الوتيرة بصورة منفردة بسبب غياب الثقة في التزام المنافسين بالأمر نفسه ومن ثم، فإن التحدي الأساسي لا يتمثل في المنافسة التكنولوجية بحد ذاتها، بل في احتمال أن تتحول المنافسة إلى سباق نحو تسريع القدرات من دون سباق موازٍ نحو تعزيز السلامة.

ففي هذا المقال سنتناول الإشارات التي أشار إليها داريو أمودي⁽¹⁾ وماهي الاخطار التي قد تنتج عن التطور المتسارع للذكاء الاصطناعي.

أولاً: لماذا يرى أمودي بضرورة ضبط وتيرة سرعة الذكاء الاصطناعي

أصبحت مخاطر تسارع تطور الذكاء الاصطناعي أكثر وضوحاً مع تزايد قدرة الأنظمة على المساهمة في بناء أجيال جديدة من الذكاء الاصطناعي، فيما يُعرف بالتحسين الذاتي التكراري، ويثير هذا المسار مخاوف من أن تتطور القدرات بوتيرة تتجاوز قدرة البشر والمؤسسات على فهمها وتقييمها والسيطرة على آثارها، وتتزايد هذه المخاوف مع ظهور حوادث أظهرت قدرة الوكلاء الذكيين على تنفيذ سلوكيات غير متوقعة، بما في ذلك تجاوز المهمة الأصلية ومحاولة التحايل على آليات التقييم، وتكمن الخطورة في أن اتساع قدرات هذه الأنظمة مستقبلاً قد يجعل مثل هذه السلوكيات أكثر تأثيراً، ولا سيما في المجال السيبراني، إذا لم تتطور إجراءات السلامة والرقابة بالسرعة نفسها، ومن هنا تبرز الحاجة إلى ضبط وتيرة التقدم في الذكاء الاصطناعي، لا بهدف إيقاف التطور التقني، وإنما لضمان وجود وقت كافٍ لمواكبة نمو القدرات بإجراءات التقييم والمواءمة والحماية، ويقوم هذا النهج على ثلاث ركائز رئيسية:

1. المقيّمون المدمجون: إتاحة وصول مستمر لمقيّمين مستقلين من أطراف ثالثة إلى عمليات تطوير النماذج، للتحقق من تطبيق معايير السلامة، ورصد الحوادث، وتقديم تقييم مستقل للمخاطر، وهدفها خلق مستوى من الثقة المتبادلة بين شركات الذكاء الاصطناعي في مدى التزامها بالقوانين أو المعايير التي تتعلق بعملية تنظيم وتيرة التسريع.

⁽¹⁾ داريو أمودي (Dario Amodei)، وهو من أبرز الباحثين ورجال الأعمال في مجال الذكاء الاصطناعي، ويشغل حالياً منصب المؤسس المشارك والرئيس التنفيذي لشركة Anthropic، وهي شركة أبحاث وتطوير في الذكاء الاصطناعي.

2. التنسيق الديمقراطي: تعاون الشركات والحكومات في الدول الديمقراطية لوضع معايير سلامة مشتركة وربط تطور القدرات بمتطلبات واضحة للتقييم والمواءمة، بحيث لا تتقدم القدرات التقنية بصورة منفصلة عن إجراءات الحماية.

3. التنسيق العالمي: السعي إلى اتفاقات دولية تحد من الاستخدامات الأكثر خطورة للذكاء الاصطناعي، وتفرض اختبارات قبل إطلاق النماذج، وقد تشمل وضع حدود لسرعة التحسين الذاتي التكراري، مع تطوير آليات فاعلة للتحقق من الالتزام.

ويستند منطق ضبط الوتيرة إلى أن المشكلة لا تكمن في سرعة التقدم التقني بحد ذاتها، وإنما في احتمال أن تصبح سرعة نمو القدرات أكبر من سرعة تطور آليات السلامة والحوكمة، ولذلك ينبغي استثمار الوقت الذي يوفره ضبط الوتيرة في تطوير المواءمة، وقابلية التفسير، والاختبار والتقييم، والأمن، والانضباط التشغيلي، وبذلك يصبح ضبط الوتيرة محاولة للحفاظ على التوازن بين سرعة الابتكار وسرعة بناء الضوابط اللازمة لمواكبته، بحيث لا يتحول تسارع القدرات إلى مصدر لمخاطر يصعب احتواؤها.

ثانياً: التسارع التقني واختلال التوازن بين التطوير والسلامة

تكشف الخطوات الثلاث التي يقترحها أمودي في مقاله عن قلق متزايد من طبيعة المنافسة التي تحكم تطوير نماذج الذكاء الاصطناعي المتقدمة، ولا سيما في ظل التسارع المستمر في رفع قدراتها، فالمنافسة بين الشركات لا تقتصر على تحسين أداء النماذج، بل تمتد إلى تحقيق تقدم متواصل في القدرات الحاسوبية والاستقلالية والتحسين الذاتي، الأمر الذي قد يدفع الشركات إلى تسريع عمليات التطوير والنشر قبل أن تنضج بالقدر نفسه أدوات التقييم والمواءمة والرقابة والحماية.

وتكمن المشكلة الأساسية وفق هذا الطرح، في عدم التوازن بين سرعة اكتساب الذكاء الاصطناعي لقدرات جديدة وسرعة تطوير الوسائل اللازمة لفهم هذه القدرات والسيطرة على مخاطرها، فكل قفزة في قدرات النماذج تفرض تحديات جديدة على أنظمة الاختبار والتقييم والسلامة، وإذا تطورت القدرات بوتيرة أسرع من هذه الأنظمة، فقد تنشأ فجوة بين ما تستطيع النماذج القيام به وبين ما تستطيع الشركات والمجتمع اكتشافه مسبقاً من مخاطر أو التحكم فيه، ومن ثم لا تتمثل المشكلة في التطور التقني بحد ذاته، وإنما في التطور غير المتوازن الذي يجعل الاستثمار في زيادة قدرات النماذج أسرع من الاستثمار في وسائل الحماية والمواءمة وقابلية التفسير والتقييم، فكلما أصبحت النماذج أكثر قدرة واستقلالية، أصبحت الحاجة إلى فهم سلوكها واختبار حدودها واكتشاف حالات الانحراف أكثر إلحاحاً، ولذلك فإن تسريع التطوير من دون تطوير موازٍ في أدوات السلامة قد يؤدي إلى انتقال أنظمة أكثر قدرة إلى بيئات الاستخدام قبل امتلاك الوسائل الكافية لتقييم مخاطرها، وفي هذا المجال يمكن فهم دعوة أمودي إلى المقيمين المدمجين بوصفها محاولة لمعالجة مشكلة الثقة والشفافية في قطاع شديد التنافس، فإتاحة وصول مستمر لمقيمين مستقلين إلى عمليات التدريب والتطوير والتقييم تهدف إلى تقليل الفجوة بين ما تعلنه الشركات عن إجراءات السلامة وما يجري فعلياً داخل عمليات التطوير، وتحويل السلامة من التزام داخلي غير قابل للتحقق بصورة كاملة إلى ممارسة يمكن تدقيقها بصورة مستقلة.

كما يعكس اقتراح التنسيق الديمقراطي إدراكاً بأن المنافسة التجارية قد تخلق حوافز لتسريع تطوير القدرات حتى لا تتخلف شركة عن منافسيها، وبالتالي فإن وضع معايير مشتركة للسلامة وحدود مرتبطة بمستوى قدرات النماذج يمكن أن يحد من مشكلة سباق التسلح التقني بين الشركات، بحيث لا تصبح إجراءات السلامة عاملاً يضعف القدرة التنافسية للشركة التي تستثمر فيها مقارنة بشركات تختار التطوير بوتيرة أسرع.

أما التنسيق العالمي فيوسع المشكلة من نطاق المنافسة بين الشركات إلى نطاق المنافسة بين الدول، إذ ترتبط قدرات الذكاء الاصطناعي بصورة متزايدة بالقوة الاقتصادية والعسكرية والاستراتيجية، وان ما يثير القلق لدى أمودي تفرد بعض الشركات العالمية في دول أخرى كالصين على سبيل المثال وهي دولة تمتلك بفارق كبير أكثر قدرات الذكاء الاصطناعي تقدماً باتباعها نظام في تطوير قدرات شركات بصورة قد يصعب التحقق بمدى التزامها بمعايير ضبط وتيرة التسريع وهذا قد يمنحها هيمنة جيوسياسية، لذلك يجب ان يتمتع اي اتفاق بقابلية تحقيق صارمة، ولهذا يقترح أمودي التعاون الدولي في مجالات محددة، إلى جانب إجراءات تتعلق بالرقائق الإلكترونية وسلاسل الإمداد والتكنولوجيا الحاسوبية، بهدف الحد من مخاطر انتقال سباق تطوير القدرات إلى مستوى يصعب معه ضمان الحد الأدنى من معايير السلامة.

وبذلك يمكن قراءة مقترحات أمودي باعتبارها محاولة لمعالجة مفارقة التقدم المتسارع فكلما أصبحت المنافسة على تطوير الذكاء الاصطناعي أكثر سرعة، ازدادت الحاجة إلى إبطاء بعض مراحل التطوير بما يسمح لأنظمة التقييم والمواءمة والحماية والحوكمة بمواكبة القدرات الجديدة، والمخاطرة هنا ليست أن يؤدي الذكاء الاصطناعي حتماً إلى نتائج كارثية، وإنما أن تتجاوز سرعة تطور قدراته سرعة تطور قدرة البشر والمؤسسات على فهمه واختباره وضبط استخدامه.

ومن ثم فإن القضية التي يطرحها المقال تتجاوز الدعوة إلى إبطاء الذكاء الاصطناعي بالمعنى التقني، لتصبح دعوة إلى إعادة التوازن بين منطق التسارع ومنطق السلامة؛ أي ألا يُنظر إلى التقدم في قدرات النماذج بوصفه الهدف الوحيد، بل يجب أن يصاحبه تقدم مماثل في التقييم، وقابلية التفسير، والمواءمة، والأمن، والرقابة المستقلة، فالمشكلة تبدأ عندما تصبح قدرة النظام على التطور أسرع من قدرة المؤسسات على فهمه وحمايته وضبط آثاره.

ثالثاً: مخاطر التطوير غير المنضبط لشركات الذكاء الاصطناعي

1. فجوة التقييم بوصفها مصدراً للمخاطر

يمثل التسارع المستمر في تطوير نماذج الذكاء الاصطناعي تحدياً جوهرياً لعمليات اختبارها وتقييم سلامتها قبل نشرها، فالتقييمات التي تُجرى قبل الإطلاق تقيس أداء النموذج ضمن مجموعة محددة من المهام والظروف والسيناريوهات، ولذلك قد لا تعكس بصورة كاملة كيفية تصرفه عند استخدامه في بيئات واقعية أكثر تعقيداً وتنوعاً، فقد يحقق النموذج نتائج آمنة ومقبولة في الاختبارات، ثم يُظهر سلوكاً مختلفاً عندما يواجه مدخلات جديدة أو ظروفًا لم تكن ممثلة بصورة كافية في بيئة التقييم، وتصبح المشكلة أكثر تعقيداً مع تطور قدرات النماذج؛ إذ قد تتمكن بعض النماذج المتقدمة من التكيف مع طبيعة الاختبارات أو التمييز بين

بيئة التقييم وبيئة الاستخدام الفعلي، كما قد تظهر سلوكيات أو قدرات لا تكشفها الاختبارات المعيارية، أو تتمكن الأنظمة من تجاوز بعض آليات التقييم بصورة مباشرة أو غير مباشرة، ويعني ذلك أن النتيجة الإيجابية في الاختبار لا تضمن بالضرورة أن السلوك نفسه سيستمر في ظروف الاستخدام الواقعي. ومن هنا يظهر مفهوم فجوة التقييم، ويُقصد به الفارق بين ما تكشفه اختبارات التقييم عن قدرات النظام وسلوكه، وبين ما قد يظهره النظام فعلياً عند تشغيله في العالم الحقيقي، ولا تعني هذه الفجوة بالضرورة وجود خلل متعمد في عملية التقييم، بل قد تنتج عن محدودية الاختبارات، أو ضيق نطاق السيناريوهات المستخدمة، أو صعوبة التنبؤ بجميع أنماط استخدام النظام، أو ظهور قدرات جديدة بعد إجراء التقييم، وتزداد أهمية هذه المشكلة مع تسارع دورة تطوير النماذج؛ فكلما تقلص الوقت بين تطوير النموذج واختباره ونشره، أصبح من الصعب إجراء تقييمات شاملة ومتعددة المراحل تتناسب مع قدراته الجديدة، وقد يؤدي ذلك إلى انتقال النموذج إلى بيئة الاستخدام قبل أن تتمكن إجراءات السلامة والتدقيق من اختبار جميع المخاطر المحتملة المرتبطة به، وعليه فإن فجوة التقييم تمثل مصدراً للمخاطر لأنها قد تؤدي إلى ثقة زائفة في سلامة النظام؛ إذ قد تستند قرارات النشر والاستخدام إلى نتائج اختبارات لا تعكس بصورة كاملة السلوك الفعلي للنموذج، وتزداد هذه المخاطر كلما اتسعت قدرات النظام وتنوعت تطبيقاته، مما يجعل تطوير أساليب تقييم مستمرة وقابلة للتكيف مع تغير قدرات النماذج عنصراً أساسياً في حوكمتها وضمان سلامتها.

2. مخاطر تسارع القدرات التقنية

يشير تسارع القدرات التقنية للذكاء الاصطناعي إلى الارتفاع المتزايد في مستوى ما تستطيع أنظمة الذكاء الاصطناعي إنجازه خلال فترة زمنية قصيرة، فالمسألة لا تتعلق بمجرد تطور قدرات النماذج، وإنما بسرعة انتقالها من مستوى من الأداء إلى مستوى أعلى وأكثر تعقيداً، وكلما أصبحت هذه الانتقالات أسرع، تقلص الوقت المتاح للباحثين والمؤسسات لفهم القدرات الجديدة واختبارها والاستعداد للأثار التي قد تترتب عليها. ويأخذ هذا التسارع أهمية خاصة عندما لا يكون التطور تدريجياً، بل يحدث على شكل قفزات نوعية في قدرات النموذج؛ فقد ينتقل النموذج خلال فترة قصيرة من أداء مهام محددة إلى التعامل مع مهام أكثر تعقيداً واستقلالية، مثل البرمجة المتقدمة، وتنفيذ العمليات متعددة الخطوات، وتحليل المعلومات، واستخدام الأدوات، أو المساعدة في أنشطة علمية وتقنية متخصصة، وبذلك فإن الخطر لا يرتبط فقط بارتفاع مستوى القدرة، وإنما أيضاً بتقلص الفترة الزمنية بين ظهور قدرة جديدة وظهور قدرة أكثر تقدماً منها، فإذا كانت قدرات الذكاء الاصطناعي تتطور بوتيرة أسرع من قدرة المؤسسات على تقييمها وتنظيمها، فقد تظهر قدرات جديدة قبل أن تتوافر اختبارات سلامة مناسبة لها أو قواعد واضحة للتعامل مع مخاطرها، ومن هذه الزاوية يمكن النظر إلى تسارع القدرات بوصفه مصدراً للمخاطر عبر ثلاث مراحل مترابطة:

تطور أسرع للقدرات ← ظهور استخدامات جديدة وغير متوقعة ← تأخر التقييم والحوكمة. وعليه فإن جوهر الخطر في تسارع القدرات يتمثل في احتمال أن تصبح سرعة تطور الذكاء الاصطناعي أكبر من سرعة فهم الإنسان لهذه القدرات واختبارها ووضع الضوابط اللازمة لها، وكلما اتسعت هذه الفجوة، زادت احتمالات ظهور مخاطر قبل أن تتمكن المؤسسات من اكتشافها أو الاستجابة لها بصورة مناسبة.

3. مخاطر فقدان السيطرة

تشير مخاطر فقدان السيطرة إلى احتمال أن تتطور أنظمة الذكاء الاصطناعي إلى درجة تصبح فيها قادرة على تنفيذ أهداف أو مهام معقدة بدرجة من الاستقلالية تجعل الإشراف البشري المباشر أقل فاعلية، ولا يعني فقدان السيطرة بالضرورة أن النظام يمتلك وعياً أو نية مستقلة، وإنما يشير إلى حالة قد يصبح فيها سلوك النظام ونتائجه صعب التنبؤ بها أو تعديلها، خصوصاً عندما يعمل في بيئات مفتوحة، أو يمتلك صلاحيات واسعة، أو يستطيع استخدام الأدوات والموارد الرقمية بصورة متتابعة.

وتزداد هذه المخاطر مع انتقال الذكاء الاصطناعي من نماذج تقدم إجابات منفردة إلى أنظمة قادرة على التخطيط، واستخدام الأدوات، وتنفيذ سلسلة من الإجراءات، ومراقبة النتائج، وتعديل سلوكها استناداً إلى المعلومات الجديدة، ففي هذه الحالة لا يعود الخطر مرتبطاً بخطأ واحد في المخرجات، وإنما بإمكانية تراكم مجموعة من القرارات والإجراءات التي تقود إلى نتيجة لم يتوقعها المشغل البشري أو لم يقصدها، ومن صور المخاطر التي تُناقش في هذا المجال قدرة النظام على التحايل على بعض إجراءات الرقابة أو اكتشاف نقاط الضعف في آليات الإشراف، أو الاستمرار في تنفيذ مهمة بطريقة لا تتوافق مع الهدف المقصود من المشغلين. كما تُطرح سيناريوهات أكثر تقدماً تتعلق بقدرة الأنظمة على الحفاظ على عملياتها أو نسخ مكوناتها أو البحث عن موارد إضافية بصورة مستقلة، وتبقى بعض هذه السيناريوهات افتراضية أو محل نقاش علمي، ولذلك ينبغي التمييز بين المخاطر التي تظهر في الأنظمة الحالية والمخاطر المستقبلية المرتبطة بزيادة الاستقلالية والقدرة، ويكمن أحد أهم أبعاد الخطر في التفاوت بين سرعة تنفيذ النظام وسرعة التدخل البشري، فإذا كان النظام قادراً على تنفيذ عدد كبير من الخطوات في وقت قصير، فقد لا يكون الإشراف البشري التقليدي، القائم على مراجعة كل خطوة على حدة، كافياً، وهنا تتحول المشكلة من مجرد مراقبة المخرجات إلى ضرورة تصميم أنظمة رقابة قادرة على العمل بالسرعة نفسها تقريباً وعلى اكتشاف الانحرافات في مراحل مبكرة.

كما أن زيادة قدرة النظام على استخدام الأدوات والوصول إلى البيانات الخارجية قد توسع نطاق التأثير المحتمل للخطأ أو السلوك غير المرغوب فيه، فالنموذج الذي ينتج نصاً غير صحيح يختلف في مستوى المخاطر عن نظام قادر على اتخاذ إجراءات في العالم الرقمي أو تنفيذ مهام متتابعة دون موافقة بشرية عند كل خطوة. ولذلك ترتبط مخاطر فقدان السيطرة بصورة مباشرة بدرجة الاستقلالية، والصلاحيات، والقدرة على التكيف، وسرعة التنفيذ، وعليه فإن جوهر المشكلة ليس مجرد امتلاك الذكاء الاصطناعي لقدرات عالية، وإنما احتمال أن تتطور القدرة والاستقلالية بسرعة أكبر من قدرة البشر والمؤسسات على فهم سلوك النظام والتنبؤ به وتصحيح مساره أو إيقافه عند الحاجة، ومن هنا تصبح قابلية النظام للتصحيح والإيقاف، ووضوح حدوده، وإمكانية مراقبة أفعاله، عناصر أساسية في إدارة هذا النوع من المخاطر.

4. إساءة الاستخدام السيبراني

تتمثل إساءة الاستخدام السيبراني في توظيف قدرات الذكاء الاصطناعي لتنفيذ أو تسهيل أنشطة ضارة تستهدف الأنظمة والشبكات والبيانات، ومع تطور النماذج، لم تعد هذه القدرات مقتصرة على الاستخدامات

الدفاعية، بل يمكن أن تساعد بعض الجهات على تسريع الهجمات، وخفض المهارات المطلوبة لتنفيذها، والتكيف مع البيئة المستهدفة، وترتفع الخطورة مع قدرة النماذج على فهم البرمجيات وكتابة الشيفرات وتحليل الأنظمة ومعالجة المعلومات وتنفيذ المهام متعددة الخطوات، ويمكن استخدامها في اكتشاف الثغرات، وتحليل الأنظمة، وإنتاج رسائل احتيالية أكثر إقناعاً، أو تعديل أساليب الهجوم وفقاً للنتائج، مما يزيد سرعة الهجوم وقدرته على التكيف، كما قد يسهم الذكاء الاصطناعي في توسيع نطاق القدرة على تنفيذ الهجمات، من خلال خفض بعض الحواجز التقنية والمعرفية التي كانت تحد من وصول محدودي الخبرة إلى أدوات متقدمة، وكما ازدادت قدرة النماذج على تنفيذ المهام بصورة مستقلة أو شبه مستقلة، قلّت الحاجة إلى التدخل البشري في بعض مراحل النشاط الضار.

وتتمثل إحدى النتائج المهمة في تقليص الوقت المتاح للجهات المدافعة لاكتشاف الهجمات وتحليلها والاستجابة لها، وتزداد الخطورة عندما ترتبط النماذج بأدوات وبيئات تنفيذ حقيقية، بحيث لا تقتصر على تقديم المعلومات أو الشيفرات، بل تستطيع تنفيذ سلسلة من الإجراءات ومراقبة نتائجها وتعديل خطواتها وفقاً للظروف، ولا يقتصر الاستخدام الضار على الاختراق المباشر، بل قد يشمل الاحتيال الرقمي والهندسة الاجتماعية وانتحال الهوية وسرقة المعلومات واختراق الحسابات ونشر البرمجيات الضارة وتعطيل الخدمات، كما يمكن للذكاء الاصطناعي إنتاج محتوى أكثر تخصيصاً وإقناعاً، مما قد يزيد فعالية بعض أساليب الخداع الرقمي.

5. الاستخدام الإجرامي

يشير الاستخدام الإجرامي إلى توظيف الذكاء الاصطناعي لتسهيل الأنشطة المخالفة للقانون أو توسيع نطاقها وزيادة كفاءتها، وتنبع أهمية هذا الخطر من قدرة الأنظمة على خفض بعض الحواجز التقنية أمام الاحتيال والخداع وانتحال الهوية، من خلال أتمتة مهام كانت تتطلب وقتاً أو مهارات بشرية أكبر، وتبرز المخاطر بصورة خاصة في عمليات الاحتيال والخداع الرقمي، إذ تستطيع الأنظمة إنتاج رسائل ومحادثات ومحتويات أكثر إقناعاً وتخصيصها وفقاً للشخص المستهدف، كما يمكن استخدام الوسائط الاصطناعية في انتحال الهوية وتقليد أنماط التواصل، مما قد يجعل اكتشاف بعض عمليات الخداع أكثر صعوبة، ولا يقتصر الخطر على تحسين جودة النشاط الإجرامي، بل يشمل أتمتته وتوسيعه؛ إذ يمكن للجهات المسيئة استخدام الذكاء الاصطناعي لتنفيذ عدد كبير من المحاولات خلال وقت قصير، وبذلك قد ينتقل النشاط الإجرامي من عمليات محدودة إلى أنشطة واسعة النطاق وقابلة للتوسع.

كما قد يؤدي انتشار النماذج المتقدمة إلى خفض الحواجز أمام أشخاص لا يمتلكون مهارات تقنية أو لغوية عالية، من خلال مساعدتهم في إنتاج المحتوى وتحليل المعلومات وتنظيم المهام، ولا يعني ذلك أن الذكاء الاصطناعي يؤدي تلقائياً إلى زيادة الجريمة، بل إن إتاحة قدرات متقدمة على نطاق واسع قد تغير تكلفة بعض الأنشطة الإجرامية وسرعتها وقابليتها للتوسع، يرتبط الخطر الإجرامي مباشرة بتسارع قدرات الذكاء الاصطناعي؛ فكلما ازدادت قدرة الأنظمة على العمل بسرعة واستقلالية والتكيف مع السياق، ازدادت الحاجة إلى تطوير أدوات للكشف والمنع قادرة على مواكبة هذه القدرات والحد من استخدامها الضار.

خاتمة

يكشف مقال داريو أمودي عن إشكالية أساسية ترافق التطور المتسارع للذكاء الاصطناعي، تتمثل في احتمال اتساع الفجوة بين سرعة نمو القدرات التقنية وسرعة تطوير آليات السلامة والتقييم والحوكمة، فالمنافسة بين الشركات والدول تدفع باتجاه تطوير نماذج أكثر قدرة واستقلالية، في حين قد لا تتطور أدوات فهم هذه القدرات ومراقبتها وضبط مخاطرها بالسرعة نفسها، ومن هنا لا تتمثل القضية في رفض التقدم التقني أو الدعوة إلى إيقاف الذكاء الاصطناعي، وإنما في ضرورة تحقيق توازن بين التسارع في تطوير القدرات والتقدم الموازي في منظومات السلامة والحماية والرقابة.

وتوضح المخاطر التي تناولها المقال أن آثار هذا التسارع لا تقتصر على مجال واحد؛ فقد تمتد إلى فجوة التقييم، وفقدان السيطرة، وإساءة الاستخدام السيبراني والإجرامي والعلمي، ويجمع هذه المخاطر عامل مشترك يتمثل في أن زيادة قدرة الأنظمة وسرعتها واستقلاليته قد توسع نطاق الأضرار المحتملة إذا لم تترافق مع وسائل فعالة للاختبار والتقييم والمراقبة والتدخل البشري.

وعليه فإن التحدي الحقيقي أمام الذكاء الاصطناعي لا يكمن فقط في ما تستطيع الأنظمة فعله، بل في قدرة الإنسان والمؤسسات على مواكبة ما تستطيع فعله، وكلما تسارعت القدرات، أصبحت الحاجة أكبر إلى تطوير آليات السلامة والمواءمة وقابلية التفسير والأمن والحوكمة.

المصادر

1. ANTHROPIC,Scaling Policy,Version 3.0,Effective February 24, 2026, For more information, see www.anthropic.com/responsible-scaling-policy.
2. Artificial,Intelligence,Index Report,2024 The 2024 <https://hai.stanford.edu> › ai-index ›.
3. Charlie Grossman, Making sense of the UK's AI Security Institute,13 July 2026, <https://www.adalovelaceinstitute.org> ›.
4. Dario Amodei,Why Pace, We Must Pace the Frontier,September 2026.
5. Myles Douglas Garvey, Dario Amodei Really Needs to Watch South Park A Realist Response to an Abstract Alarmist Essay, rmylesgarvey@gmail.com (201) 835-0260.
6. Velibor Božić, HE DANGERS OF ARTIFICIAL INTELLIGENCE, General hospital Koprivnica,veliborbozic@gmail.com.

تأسس مركز الفيض العلمي لاستطلاع الرأي والدراسات المجتمعية في بغداد بموجب شهادة التسجيل الصادرة عن الأمانة العامة لمجلس الوزراء -دائرة المنظمات غير الحكومية المرقمة (1J775330) بتاريخ ٢٦/٤/٢٠١٢، وهو مركز علمي بحثي يهتم بإجراء الاستطلاعات والدراسات الميدانية فضلاً عن إعداد الأوراق البحثية والمقالات حول قضايا الحياة المجتمعية للأسرة والمواطن، والدولة بمؤسساتها المختلفة.

- لا يجوز نشر أي من إصدارات المركز ونتائجته العلمية الا بموافقة خطية صريحة ويمكن الاقتباس بشرط ذكر المصدر كاملاً.
- حقوق الطبع والنشر محفوظة لمركز الفيض العلمي لاستطلاع الرأي والدراسات المجتمعية

للتواصل

00964- 7710122232



Alfaydhcenter2011@gmail.com



www.al-faydh.com



العراق - بغداد - الكرادة

