موجز تنفيذي

لمحة عن هذا التقرير

- هذا هو الإصدار المبدئي الأول "للتقرير الدولي العلمي حول سلامة الذكاء الاصطناعي المتطور". وقد ساهم في صياغة هذا التقرير مجموعة متنوعة مكونة من 75 خبير ذكاء اصطناعي، بما في ذلك لجنة خبراء دولية استشارية مرشحة من قبل 30 دولة، والاتحاد الأوروبي، والأمم المتحدة.
- محتويات هذا التقرير بالكامل هي حصيلة لاجتهادات الخبراء المستقلين الذين ساهموا في كتابته تحت قيادة رئيس فريق التقرير.
- في هذا الزمن الذي يشهد تقدما غير مسبوق في تطور الذكاء الاصطناعي، يركز هذا الإصدار الأول بشكل خاص على نوع من الذكاء الاصطناعي تطور بسرعة لافتة في السنوات الأخيرة: وهو الذكاء الاصطناعي متعدد الاستخدامات، أي الذكاء الاصطناعي الذي يمكنه تأدية نطاق واسع من المهام. ووسط النجاحات المتسارعة في مجال الذكاء الاصطناعي متعدد الاستخدامات، فإن الأبحاث في هذا المجال تمر حاليا بمرحلة من الاكتشافات العلمية، ولا يمكن اعتبارها علما مكتملا بعد.
- لن يتمكن الناس حول العالم من الاستفادة من المنافع المرتقبة للذكاء الاصطناعي متعدد الاستخدامات بشكل آمن طالما أن مخاطره لا تُدار بالشكل السليم. لذا يركز هذا التقرير على التعرّف على تلك المخاطر، وتقييم الطرق الفنية لمعرفة أبعاد تلك المخاطر أو تخفيفها. علما أن هذا التقرير لا يهدف لإجراء تقييم شامل لكافة التأثيرات المجتمعية المُحتملة للذكاء الاصطناعي متعدد الاستخدامات، بما في ذلك فوائده الكثيرة المحتملة.
- لأول مرة في التاريخ، يجمع هذا التقرير المبدئي خبراء مرشحين من قبل 30 دولة والاتحاد الأوروبي والأمم المتحدة، بالإضافة إلى خبراء دوليين رائدين آخرين، لإيجاد أساس علمي مشترك قائم على الأدلة تستند إليه النقاشات والقرارات فيما يخص أمور سلامة الذكاء الاصطناعي متعدد الاستخدامات. لا نزال نختلف حاليا حول العديد من المسائل، كبيرها وصغيرها، بشأن قدرات الذكاء الاصطناعي متعدد الاستخدامات ومخاطره وتخفيف تلك المخاطر. لكننا نعتبر هذا المشروع ضروريا لتحسين فهمنا المشترك لهذه التقنية ومخاطرها المحتملة، وللسعي إلى إجماع الآراء، والتوصل لأساليب فعّالة لتخفيف تلك المخاطر لضمان أن ينعم الناس بشكل آمن بفوائد الذكاء الاصطناعي متعدد الاستخدامات. المسؤولية جسيمة، ونحن نتطلع قُدما إلى مواصلة هذه الجهود.

النقاط الرئيسية في الموجز التنفيذي

- إذا أمكن تنظيم الذكاء الاصطناعي متعدد الاستخدامات جيدا، يمكن تطبيقه لخدمة المصلحة العامة، ما قد يؤدي إلى تعزيز رفاه الناس، ومزيد من الرخاء، واكتشافات علمية جديدة. إلا أنه في حال اختلال الذكاء الاصطناعي متعدد الاستخدامات، أو استخدامه لأغراض خبيثة، يُمكن أن يتسبب أيضا بأضرار، كاتخاذ قرارات منحازة في أوضاع تنطوي على مسؤولية كبيرة، أو من خلال الخداع، أو وسائل إعلام مُضلِلة، أو انتهاك الخصوصية.
- وفيما يتواصل تطور قدرات الذكاء الاصطناعي متعدد الاستخدامات، يُمكن أن تظهر مخاطر من قبيل تأثيرات واسعة النطاق على سوق العمل، واختراق الأنظمة الإلكترونية بواسطة الذكاء الاصطناعي أو هجمات بيولوجية، وفقدان المجتمع السيطرة على الذكاء الاصطناعي متعدد الاستخدامات. وتجدر الإشارة إلى أن احتمال حدوث تلك

السيناريوهات لا يزال قيد النقاش بين الباحثين. وغالبا ما ينشأ التباين في وجهات النظر حول تلك المخاطر نتيجة لتوقعات مختلفة بشأن الخطوات التي سيتخذها المجتمع للحد منها، وفاعلية تلك الخطوات، ومدى سرعة تطور قدرات الذكاء الاصطناعي متعدد الاستخدامات.

- لا يزال هناك قدر كبير من عدم اليقين بشأن معدّل تقدم قدرات الذكاء الاصطناعي متعدد الاستخدامات في المستقبل. إذ يعتقد بعض الخبراء أن التقدم في هذا المضمار سيتباطأ على الأرجح، فيما يرى خبراء آخرون أن من المحتمل أو المرجح حصول تقدم بسرعة هائلة.
- هناك أساليب فنية متنوعة يُمكن للمطورين استخدامها وللجهات التنظيمية أن تشترطها لتقييم أخطار الذكاء الاصطناعي متعدد الاستخدامات وتخفيفها. لكن جميع تلك الطرق محدودة الإمكانات. فمثلا الأساليب الحالية لشرح سبب إنتاج الذكاء الاصطناعي متعدد الاستخدامات لمُخرجات بعينها لا تزال محدودة بدرجة كبيرة.
- لا يزال الغموض يكتنف مستقبل تكنولوجيا الذكاء الاصطناعي متعدد الاستخدامات، حيث هناك نطاق واسع من المسارات التي تبدو محتملة حتى في المستقبل القريب، بما في ذلك المُخرجات الإيجابية جدا والسلبية جدا. لكن ليس هناك ما هو حتمي بشأن مستقبل الذكاء الاصطناعي، بل إن القرارات التي تتخذها المجتمعات والحكومات هي التي ستقرر مستقبل الذكاء الاصطناعي. ويهدف هذا التقرير المبدئي لتسهيل إجراء النقاش البنّاء بشأن تلك القرارات.

يتضمن هذا التقرير الفهم العلمي للذكاء الاصطناعي متعدد الاستخدامات - أي الذكاء الاصطناعي الذي يمكنه تأدية نطاق واسع من المهام - مع التركيز على فهم المخاطر وإدارتها.

تشهد قدرات الأنظمة التي تستخدم الذكاء الاصطناعي تقدما سريعا. وقد أدى ذلك إلى تسليط الضوء على الفرص العديدة التي يوفرها الذكاء الاصطناعي للشركات ومراكز البحوث والحكومات، وكذلك في الحياة الخاصة للأفراد. كما أدى إلى زيادة الوعي بالأضرار الحالية والمخاطر المحتملة مُستقبلا المرتبطة بالذكاء الاصطناعي المتطور.

إن الغرض من التقرير الدولي العلمي حول سلامة الذكاء الاصطناعي المتطور هو السعي للوصول إلى فهم دولي مشترك لأخطار الذكاء الاصطناعي، وكيفية الإقلال منها. ويقتصر التركيز في هذا الإصدار المبدئي الأول للتقرير على صنف من الذكاء الاصطناعي تطورت قدراته بسرعة ملحوظة: وهو الذكاء الاصطناعي متعدد الاستخدامات، أي الذكاء الاصطناعي الذي يُمكنه تأدية نطاق واسع من المهام.

ففي وسط التطورات المتلاحقة، فإن البحوث المتعلقة بالذكاء الاصطناعي متعدد الاستخدامات تمر حاليا بفترة من الاكتشافات العلمية، ولا تُعتبر بعد علما مكتملا. وهذا التقرير يعطي لمحة سريعة عن الفهم العلمي الحالي للذكاء الاصطناعي متعدد الاستخدامات ومخاطره. ويشمل ذلك تحديد المجالات المتفق عليها علميا، والمجالات التي تتباين حولها الآراء أو المسائل البحثية التي لم تُحسم بعد.

ولن يتمكن الناس حول العالم من الاستفادة بشكل آمن من المنافع المرتقبة للذكاء الاصطناعي متعدد الاستخدامات ما لم تكن أخطاره تُدار بشكل مناسب. لذا، يركز هذا التقرير على التعرّف على تلك المخاطر وتقييم الطرق الفنية لمعرفة أبعادها وتخفيف أثرها، بما في ذلك الاستخدام المفيد للذكاء الاصطناعي متعدد الاستخدامات من أجل تخفيف مخاطره. هذا التقرير لا يهدف لإجراء تقييم شامل لكافة التأثيرات المجتمعية المُحتملة للذكاء الاصطناعي متعدد الاستخدامات، بما في ذلك المنافع التي يُمكن أن يوفرها.

تنامت قدرات الذكاء الاصطناعي متعدد الاستخدامات بسرعة كبيرة خلال السنوات الأخيرة وفقا للعديد من المقاييس، ولا يوجد إجماع على كيفية توقع تطوره مستقبلا، ما يجعل سيناريوهات كثيرة محتملة.

وفقا للعديد من المقاييس، فإن قدرات الذكاء الاصطناعي متعدد الاستخدامات تتطور بسرعة كبيرة. فقبل خمس سنوات، كانت النماذج الرائدة للذكاء الاصطناعي متعدد الاستخدامات في مجال اللغة بالكاد تستطيع تأليف فقرة واحدة مترابطة منطقياً. أما اليوم، فيمكن لبعض النماذج المشاركة في حوارات متعمّقة تغطي قطاعا واسعا من المواضيع، أو أن تكتب برامج حاسوبية قصيرة، أو أن تصمم مقاطع فيديو بناء على الوصف فقط. لكن قدرات الذكاء الاصطناعي متعدد الاستخدامات يصعب تقديرها بشكل موثوق أو تحديدها بشكل دقيق.

تعتمد وتيرة تطور الذكاء الاصطناعي متعدد الاستخدامات على كل من معدّل التقدم التكنولوجي والبيئة التنظيمية. وهذا التقرير يركّز على الجوانب التكنولوجية دون الخوض في بحث كيف يُمكن لجهود الجهات التنظيمية أن تؤثر على سرعة تطور وانتشار الذكاء الاصطناعي متعدد الاستخدامات.

وقد تمكن مطوّرو الذكاء الاصطناعي من تطوير قدرات الذكاء الاصطناعي متعدد الاستخدامات بشكل سريع في السنوات الأخيرة، وذلك من خلال الزيادة المتواصلة للموارد المسُتخدمة لتدريب نماذج جديدة (يُطلق على هذا التوجه "تعظيم الموارد") وتحسين الخوارزميات المتوفرة. فمثلا، شهدت أحدث نماذج الذكاء الاصطناعي زيادات سنوية بنحو 4 أضعاف في مواردها الحاسوبية ("القدرة الحاسوبية") المُستخدمة في التدريب، ونحو 2.5 ضعفا في حجم قاعدة البيانات للتدريب، ونحو 1.5 إلى 3 أضعاف في الكفاءة الخوارزمية (الأداء مُقاسا بالقدرة الحاسوبية). وإلى الآن لا يزال الباحثون يناقشون ما إن كان "تعظيم الموارد" قد أدى إلى تطور في معالجة تحديات جوهرية مثل الاستدلال المبني على المسببات.

ستكون لوتيرة التطور المستقبلي للذكاء الاصطناعي متعدد الاستخدامات تأثيرات كبيرة على إدارة المخاطر الناشئة عنه، لكن الخبراء غير متفقين بشأن ما يُمكن توقعه في المستقبل المنظور. إذ تتفاوت آراؤهم حول ما إن كانت قدرات الذكاء الاصطناعي متعدد الاستخدامات ستتطور بطيئا أم سريعا أم بسرعة هائلة. ويتمحور هذا الاختلاف حول سؤال أساسي: هل سيكون "تعظيم الموارد" المستمر وتحسين الأساليب القائمة كافيا لإحراز تقدم سريع وحل إشكاليات مثل إمكانية الوثوق بالذكاء الاصطناعي ودقة المعلومات، أم ستكون هناك حاجة لتحقيق اختراقات بحثية جديدة من أجل رفع قدرات الذكاء الاصطناعي متعدد الاستخدامات بشكل كبير؟

في الوقت الحالي، تراهن شركات عديدة رائدة في مجال تطوير الذكاء الاصطناعي متعدد الاستخدامات على أن "تعظيم الموارد" سوف يستمر في تحسين الأداء. وإذا استمرت التوجهات التي برزت مؤخرا، فبحلول نهاية عام 2026 سيكون قد جرى تدريب بعض نماذج الذكاء الاصطناعي متعدد الاستخدامات باستخدام قدرة حاسوبية تزيد بما يتراوح ما بين 40 إلى 100 ضعف مقارنة بأعلى نماذج القدرة الحاسوبية التي طُرحت عام 2023، إلى جانب طرق تدريب تستخدم القدرة الحاسوبية بفعالية أكبر بنسبة ما بين 3 إلى 20 ضعفا. لكن ثمة معوقات محتملة أمام كل من زيادة كمية البيانات والقدرة الحاسوبية، ومنها مثلا توفر البيانات، ورقائق الذكاء الاصطناعي، وإنفاق رأس المال، وقدرة الطاقة المحلية. وتعمل الشركات الناشطة في مجال تطوير الذكاء الاصطناعي متعدد الاستخدامات على تفادي تلك المعوقات المُحتملة.

تهدف جهود بحثية عديدة لفهم وتقييم الذكاء الاصطناعي متعدد الاستخدامات بمزيد من الثقة، لكن فهمنا العام لكيفية عمل نماذجه وأنظمته لا يزال محدودا.

غالبا ما تعتمد أساليب إدارة المخاطر الناشئة عن الذكاء الاصطناعي متعدد الاستخدامات على فرضية أنه بإمكان مطوري الذكاء الاصطناعي وصانعي السياسات تقييم قدرات نماذج وأنظمة الذكاء الاصطناعي متعدد الاستخدامات وتأثيراته المحتملة. لكن في حين أن الطرق الفنية يُمكن أن تساعد في التقييم، إلا أن جميع الطرق المتوفرة حاليا محدودة القدرات وعاجزة عن تقديم ضمانات أكيدة ضد غالبية الأضرار الناجمة عن الذكاء الاصطناعي متعدد الاستخدامات. وبشكل عام، فإن الفهم العلمي لآليات عمله وقدراته وتأثيراته على المجتمع محدود جدا. وهناك اتفاق واسع النطاق بين الخبراء حول أن زيادة فهمنا له يجب أن تكون من الأولويات. تشمل التحديات الأساسية ما يلي:

- إلى الآن، لا يعرف مطورو الذكاء الاصطناعي إلا القليل عن كيفية عمل نماذج الذكاء الاصطناعي متعدد الاستخدامات. ذلك لأن تلك النماذج ليست مبرمجة بالمعنى التقليدي، بل هي مدربة: حيث صمم مطورو الذكاء الاصطناعي عملية تدريبية تتضمن الاستعانة بكميات كبيرة من البيانات، ونتج عن عملية التدريب تلك خلق نموذج الذكاء الاصطناعي متعدد الاستخدامات. وقد تتألف تلك النماذج من تريليونات المكونات، التي تُسمى معايير (parameters)، وأغلب آليات عملها الداخلي مُبهمة حتى بالنسبة لمطوريها. ويُمكن لأساليب شرح وتفسير عمل النماذج أن ترفع من قدرة الباحثين والمطورين على فهم كيفية عملها، إلا أن الأبحاث في هذا المجال لا تزال في بداياتها.

- يجري تقييم الذكاء الاصطناعي متعدد الاستخدامات أساسا من خلال اختبار النموذج أو النظام بالاستعانة بمُدخلات مُختلفة. وتفيد تلك الاختبارات العشوائية في تقييم نقاط القوة والضعف، بما في ذلك الثغرات في مناعة النموذج وقدراته الضارة المحتملة، لكنها لا توفر ضمانات وافية بشأن سلامته. فتلك الاختبارات غالبا ما يفوتها الانتباه لبعض الأخطار، كما أنها إما أن تُبالغ في تقدير القدرات أو تقلل من شأنها نظرا لأن أنظمة الذكاء الاصطناعي متعدد الاستخدامات يُمكن أن تستجيب بطرق مختلفة في ظروف مختلفة، وبوجود مستخدمين مختلفين، أو بسبب إدخال تعديلات إضافية على مكوناتها.

- من حيث المبدأ، بإمكان جهات مستقلة إجراء تدقيق على نماذج أو أنظمة الذكاء الاصطناعي متعدد الاستخدامات صممتها شركة ما. لكن الشركات عادة لا تزود المدققين المستقلين بالمستوى اللازم من إمكانية الدخول إلى النماذج، ولا بالمعلومات عن البيانات والطرق المُستخدمة التي يحتاجونها لإجراء تقييم دقيق. وقد عكفت حكومات عديدة على بناء القدرات لإجراء تقييمات فنية وتدقيق.

- يصعب تقييم تأثير أنظمة الذكاء الاصطناعي متعدد الاستخدامات على المجتمع، ذلك لأن الأبحاث في مجال تقييم المخاطر ليست كافية بعد لتوفير طرق إجراء تقييم شامل ودقيق. وفضلا عن ذلك، فهناك نطاق واسع جدا لحالات استخدام الذكاء الاصطناعي متعدد الاستخدامات. وغالبا ما تكون تلك الحالات غير محددة مُسبقا، ولا قيود مشددة عليها، ما يجعل تقييم المخاطر أكثر تعقيداً. لفهم التأثيرات اللاحقة المُحتملة لنماذج وأنظمة الذكاء الاصطناعي متعدد الاستخدامات على المجتمع لا بد من إجراء تحليل دقيق متعدد لجوانب متعددة. أما زيادة نسبة تمثيل وجهات النظر المتباينة في عمليات تطوير وتقييم الذكاء الاصطناعي متعدد الاستخدامات، فلا تزال تشكل تحديا فنيا مستمرا للمؤسسات.

يُمكن للذكاء الاصطناعي متعدد الاستخدامات أن يشكل مخاطر جسيمة على سلامة ورفاه الأفراد والمجتمع.

يصنف هذا التقرير مخاطر الذكاء الاصطناعي متعدد الاستخدامات ضمن ثلاث فئات وهي: مخاطر الاستخدام بسوء نية، ومخاطر الخلل، ومخاطر الأنظمة. كما يستعرض التقرير عدة عوامل متداخلة تساهم في نشوء العديد من المخاطر.

الاستخدام بسوء نية. كما الحال في كافة أنواع التكنولوجيا القوية، يُمكن استخدام أنظمة الذكاء الاصطناعي متعدد الاستخدامات بسوء نية لإحداث ضرر. وتتراوح الأشكال المحتملة للاستخدام بسوء نية بين استخدامات مُثبتة بالأدلة نسبيا، مثل الخداع الإلكتروني باستخدام الذكاء الاصطناعي متعدد الاستخدامات، وأشكال أخرى يتوقع بعض الخبراء ظهورها في السنوات القادمة، كالاستخدام بسوء نية للقدرات العلمية للذكاء الاصطناعي متعدد الاستخدامات.

- الضرر الذي يصيب الأفراد من خلال المحتوى المضلِل الناتج عن الذكاء الاصطناعي متعدد الاستخدامات هو من أشكال الضرر الموثقة جيدا نسبيا لاستخدامه بسوء نية. حيث يمكن استخدامه لزيادة نطاق وتعقيد حالات الخداع والاحتيال، مثلا هجمات تصيد البيانات الشخصية ("phishing") التي يعززها الذكاء الاصطناعي متعدد الاستخدامات. كما يُمكن استخدامه لإنتاج محتوى مضلِل يظهر فيه أفراد دون موافقتهم، مثل حالات التزييف العميق (deepfake) للقطات إباحية دون موافقة الذين يظهرون فيها.

- من الممارسات الأخرى التي تثير القلق هي استغلال الذكاء الاصطناعي متعدد الاستخدامات بسوء نية لنشر معلومات مضللة وتضليل الرأي العام. حيث يُسهّل الذكاء الاصطناعي متعدد الاستخدامات وأنواع التكنولوجيا الأخرى خلق ونشر معلومات مضللة، منها ما هو بقصد التأثير على العمليات السياسية. ورغم فائدة التدابير الفنية المضادة، مثل استخدام العلامات المائية الرقمية (watermarking) لحماية المحتوى من التلاعب، فعادة ما يتمكن مستخدمون ذوو خبرة متوسطة من الالتفاف عليها.

- بالإمكان أيضا استخدام الذكاء الاصطناعي متعدد الاستخدامات بسوء نية لتنفيذ هجمات إلكترونية، ورفع الخبرة الإلكترونية لدى الأفراد بحيث يسهل على ذوي النوايا الخبيثة شن هجمات إلكترونية مؤثرة. كما يُمكن استخدام أنظمته لزيادة بعض أنواع العمليات الإلكترونية وأتمتتها جزئيا، ومن الأمثلة على ذلك هجمات الهندسة الاجتماعية (social engineering attacks). لكن من جهة أخرى، يُمكن أيضا تسخير الذكاء الاصطناعي متعدد الاستخدامات في الدفاع الإلكتروني. إلا أنه بصورة عامة لا تتوفر بعد أية أدلة قوية على أن الذكاء الاصطناعي متعدد الاستخدامات يمكنه أتمتة عمليات أمن إلكتروني معقدة.

- أعرب بعض الخبراء كذلك عن مخاوفهم من إمكانية استخدام أنظمة الذكاء الاصطناعي متعدد الاستخدامات لدعم تطوير الأسلحة واستخدامها على نحو ضار، كالأسلحة البيولوجية. لكن لا توجد أدلة قوية لإثبات أن أنظمة الذكاء الاصطناعي متعدد الاستخدامات الحالية تشكّل خطرا من هذا النوع. فمثلا، رغم أن تلك الأنظمة تُظهر قدرات متنامية في المجال البيولوجي، فإن الدراسات القليلة المتوفرة لا تقدم أدلة واضحة على أن الأنظمة الموجودة حاليا تستطيع "رفع" قدرة المستخدمين الذين يضمرون سوء النية للحصول على المسببات البيولوجية للأمراض بشكل أفضل مما لو استخدموا الإنترنت. لكن حتى الآن يندر تقييم تهديدات مستقبلية واسعة النطاق، وبالتالي لا يمكن استبعاد حدوثها.

مخاطر حدوث خلل. حتى عندما لا يقصد المستخدمون إحداث ضرر، قد تنشأ مخاطر جسيمة نتيجة خلل في أنظمة الذكاء الاصطناعي متعدد الاستخدامات. ويُمكن أن يكون لذلك الخلل العديد من الأسباب والعواقب المحتملة:

- مستخدمو المنتجات التي تقوم على نماذج وأنظمة الذكاء الاصطناعي متعدد الاستخدامات قد لا يفهمون عملها بالقدر الكافي، وربما يكون ذلك مثلا نتيجة لسوء التواصل بشأنها أو الإعلانات الدعائية المُضللة. ويُمكن أن يُسبب ذلك ضررا إذا استعان المُستخدمون بهذه الأنظمة بطرق غير ملائمة أو لأغراض غير ملائمة.

- يُعتبر الانحياز في أنظمة الذكاء الاصطناعي عموما مشكلة مُثبتة بالأدلة، وهي لا تزال دون حل بالنسبة للذكاء الاصطناعي متعدد الاستخدامات أيضا. حيث يمكن أن تنحاز مُخرجات الذكاء الاصطناعي متعدد الاستخدامات بالنسبة لبيانات محمية، مثل الأصل العرقي ونوع الجنس والخلفية الثقافية والسن وأشكال الإعاقة. ويُمكن أن يؤدي ذلك إلى نشوء مخاطر، ومنها ما يكون في مجالات عالية الأهمية مثل الرعاية الصحية والتوظيف والقروض المالية. يُضاف إلى ذلك أن العديد من نماذج الذكاء الاصطناعي متعدد الاستخدامات المستخدمة على نطاق واسع تكون مُدربة أساسا على بيانات لا تمثل الثقافات الغربية بالنسب الصحيحة، وهذا قد يزيد من احتمال الإضرار بالأفراد غير الممثلين بشكل جيد في تلك البيانات.

- سيناريوهات "فقدان السيطرة" هي سيناريوهات مستقبلية مُحتملة، وفيها لا يعود في استطاعة المجتمع ضبط أنظمة الذكاء الاصطناعي متعدد الاستخدامات ضبطا فعالا، حتى ولو بدا واضحا أن تلك الأنظمة تسبب الضرر. لكن ثمة إجماع واسع النطاق على أن الذكاء الاصطناعي متعدد الاستخدامات يفتقر حاليا إلى القدرات التي تشكل ذلك الخطر. ويعتقد بعض الخبراء أن الجهود المبذولة حاليا لتطوير أنظمة مستقلة للذكاء الاصطناعي متعدد الاستخدامات - تستطيع أن تعمل وتخطط وترسم الأهداف - يُمكن في حال نجاحها أن تؤدي إلى فقدان السيطرة. وإلى الآن، تتباين آراء الخبراء حول مدى إمكانية حدوث سيناريوهات فقدان السيطرة، ومتى يُمكن أن تحدث، ومدى صعوبة التخفيف من أضرارها.

المخاطر النُظمية: ازدياد تطوير واستخدام تكنولوجيا الذكاء الاصطناعي متعدد الاستخدامات يشكل مخاطر نُظمية عديدة، من بينها التأثيرات المحتملة على سوق العمل، ومخاطر انتهاك الخصوصية، والتأثيرات البيئية.

- إذا استمر الذكاء الاصطناعي متعدد الاستخدامات في التطور السريع، فسينتج عن ذلك أتمتة قطاع واسع جدا من المهام، ما قد يؤثر تأثيرا بالغا على سوق العمل. ذلك قد يعني خسارة العديد من الأفراد لوظائفهم الحالية. لكن يتوقع الكثير من الاقتصاديين إمكانية التعويض المحتمل لفقدان الوظائف، وربما بشكل كامل، عن طريق خلق وظائف جديدة، وارتفاع الطلب في القطاعات غير المؤتمتة.

- تتركز حاليا أعمال البحث والتطوير المتعلقة بالذكاء الاصطناعي متعدد الاستخدامات في بعض الدول الغربية وفي الصين. ولهذا "الفارق في الذكاء الاصطناعي" أسباب عديدة، لكنه ينشأ جزئيا عن اختلاف مستويات الحصول على القدرة الحاسوبية اللازمة لتطوير الذكاء الاصطناعي متعدد الاستخدامات. فالدول منخفضة الدخل والمؤسسات الأكاديمية أقل حظا في إمكانية الحصول على القدرة الحاسوبية مقارنة بالدول الغنية وشركات التكنولوجيا.

- وما ينجم عن ذلك من تمركز عمليات تطوير الذكاء الاصطناعي متعدد الاستخدامات في أسواق معينة يجعل المجتمعات أكثر عرضة للعديد من المخاطر النُظمية. على سبيل المثال، فإن شيوع استخدام عدد صغير من أنظمة الذكاء الاصطناعي متعدد الاستخدامات في قطاعات حساسة مثل قطاعي المال أو الرعاية الصحية يُمكن أن يؤدي إلى إخفاقات وأعطال متزامنة على نطاق واسع في تلك القطاعات التي تعتمد على بعضها البعض. فقد ينشأ ذلك مثلا عن خلل برمجي أو نقاط ضعف أخرى.

- أدى تنامي استخدام القدرة الحاسوبية في تطوير واستعمال الذكاء الاصطناعي متعدد الاستخدامات إلى ازدياد سريع في استهلاك الطاقة المرتبطة به. ولا توجد مؤشرات على أن هذا التوجه سيتباطأ، ما يعني أنه سيؤدي للمزيد من الانبعاثات الكربونية، ومن استهلاك المياه.

- يُمكن لنماذج وأنظمة الذكاء الاصطناعي متعدد الاستخدامات أن تشكّل خطرا على الخصوصية. فقد أظهرت الأبحاث مثلا أنه باستخدام مُدخلات معادية يستطيع المستخدمون أن يستخلصوا من النماذج بيانات تدريب تحتوي على معلومات عن الأفراد. وقد يؤدي ذلك إلى انتهاكات خطيرة للخصوصية بالنسبة لنماذج مستقبلية مدرَّبة على بيانات شخصية حساسة، كالبيانات الصحية أو المالية.

- التعديات المحتملة على حقوق النشر في مجال تطوير الذكاء الاصطناعي متعدد الاستخدامات تشكل تحديا لقوانين حماية الملكية الفكرية المتعارف عليها، وكذلك للأنظمة المتعلقة بالموافقة والتعويضات وضبط البيانات. وإذا كانت قواعد حقوق النشر غير واضحة، سيمتنع مطورو الذكاء الاصطناعي متعدد الاستخدامات عن الكشف عن البيانات التي يستخدمونها. كما أن القواعد غير الواضحة لن تُبين أشكال الحماية المتوفرة لصنّاع المحتوى الذين يُستخدم عملهم دون موافقتهم لتدريب نماذج الذكاء الاصطناعي متعدد الاستخدامات.

عوامل خطر متداخلة. ترتكز المخاطر المصاحبة للذكاء الاصطناعي متعدد الاستخدامات على عدة عوامل خطر متداخلة - وهي سمات الذكاء الاصطناعي متعدد الاستخدامات والتي تزيد من احتمال أو شدة أخطار عديدة:

- عوامل الخطر الفنية المتداخلة تشمل صعوبة ضمان أن أنظمة الذكاء الاصطناعي متعدد الاستخدامات تعمل بثبات كما هو مخطط لها، وعدم فهم آلية عملها الداخلي، والتطوير المتواصل "لعوامل" الذكاء الاصطناعي متعدد الاستخدامات التي تستطيع العمل بشكل مستقل بقليل من الإشراف.

- عوامل الخطر المجتمعية المتداخلة تشمل وجود فارق محتمل بين وتيرة التطور التكنولوجي ووتيرة إيجاد قواعد تنظيمية له، بالإضافة إلى حوافز تنافسية تشجع مطوري الذكاء الاصطناعي على إطلاق منتجاتهم بسرعة، ما قد يؤدي إلى التفريط بجودة إدارة المخاطر.

هناك مقاربات فنية عديدة قادرة على التخفيف من حدة المخاطر. لكن لا توجد حاليا طريقة معروفة توفر ضمانات أكيدة ضد الأضرار الناجمة عن الذكاء الاصطناعي متعدد الاستخدامات.

رغم أن هذا التقرير لا يناقش تدخلات تتعلق بسياسات تخفيف حدة المخاطر الناجمة عن الذكاء الاصطناعي متعدد الاستخدامات، إلا أنه يستعرض طرقا فنية لتخفيفها، ويحقق الباحثون نجاحا في تطويرها. لكن على الرغم من ذلك النجاح، فإن الطرق المتبعة حاليا لا تتصدى بشكل موثوق للمُخرجات الضارة للذكاء الاصطناعي متعدد الاستخدامات، ولا حتى للظاهر منها، في بيئة العمل الواقعية. وتُستخدم حاليا عدة مقاربات فنية لتقييم المخاطر وتخفيف حدتها:

- أُحرز بعض التقدم في تدريب نماذج الذكاء الاصطناعي متعدد الاستخدامات لتعمل بأمان أكبر. كما يدرب المطورون النماذج لتكون أقوى مناعة ضد المدخلات المصممة لجعلها تفشل ("التدريب ضد الهجمات المعادية"). ورغم ذلك، عادة ما تجد الجهات المعادية، بجهد قليل أو متوسط، مدخلات بديلة تقلل من فاعلية الدفاعات. وإذا تمكن المطورون من جعل قدرات نظام الذكاء الاصطناعي متعدد الاستخدامات تقتصر على استخدام محدد، فمن شأن ذلك الإقلال من المخاطر التي تنشأ عن الأعطال غير المتوقعة أو عن الاستخدام بسوء نية.

- هناك أساليب عديدة للتعرف على المخاطر، ولفحص عمل الأنظمة، وتقييم أداء أنظمة الذكاء الاصطناعي متعدد الاستخدامات لدى العمل بها. ويُشار عادة إلى تلك الممارسات بعمليات "المراقبة".
- يُمكن العمل على تخفيف حدة الانحياز في أنظمة الذكاء الاصطناعي متعدد الاستخدامات طوال دورة حياتها، ويشمل ذلك التصميم والتدريب والتركيب والاستخدام. لكن من الصعب منع الانحياز بالكامل في أنظمة الذكاء الاصطناعي متعدد الاستخدامات، لأن ذلك يتطلب التدريب المستمر، وجمع البيانات، والتقييم المتواصل، والتعرف الفعال على الانحياز. كما قد يتطلب التخلي عن الإنصاف مقابل تحقيق أهداف أخرى كالدقة والخصوصية، وتحديد ما يمكن اعتباره معلومات مفيدة، وما هو انحياز مرفوض يجب ألا يظهر في المخرجات.
- حماية الخصوصية تُعتبر من مجالات البحث والتطوير المستمرة. ومجرد تقليل استخدام بيانات شخصية حساسة في التدريب يُعتبر من الطرق التي يمكنها أن تخفض لحد كبير المخاطر المتعلقة بحماية الخصوصية. لكن لدى استخدام بيانات شخصية، سواء بقصد أو بغير قصد، فإن الأدوات الفنية المتاحة لتخفيف المخاطر المتعلقة بحماية الخصوصية تواجه صعوبة في التعامل مع النماذج الضخمة للذكاء الاصطناعي متعدد الاستخدامات، وربما تفشل في أن توفر للمستخدمين سيطرة مُجدية.

الختام: هناك نطاق واسع من المسارات المحتملة للذكاء الاصطناعي متعدد الاستخدامات، ويعتمد الكثير منها على الطرق التي ستتصرف بها المجتمعات والحكومات.

مستقبل الذكاء الاصطناعي متعدد الاستخدامات غير مؤكد. إذ يبدو أن هناك نطاقا واسعا من المسارات المحتملة حتى في المستقبل القريب، ويشمل ذلك التوصل إلى نتائج إيجابية جدا أو سلبية جدا. لكن ليس هناك ما هو حتمي بالنسبة لمستقبل الذكاء الاصطناعي متعدد الاستخدامات: كيف يجري تطويره ومن قبل مَن، وما المشاكل التي سيكون مصمما لحلها، وهل ستتمكن المجتمعات من الاستفادة من جني ثمار كامل إمكاناته الاقتصادية، ومَن سيستفيد منه، وما هي الأخطار التي نعرّض أنفسنا لها باستخدامه، وكم من الأموال يجب أن نستثمرها لتخفيف مخاطره - الإجابة على هذه الأسئلة وكثير غيرها تعتمد على الخيارات التي تلجأ إليها المجتمعات والحكومات، الآن ومستقبلا، لبلورة تطوير الذكاء الاصطناعي متعدد الاستخدامات.

ولتسهيل إجراء نقاشات بناءة حول تلك الخيارات، يعرض هذا التقرير لمحة عامة عن الوضع الراهن للأبحاث العلمية، والمناقشات الدائرة حول إدارة مخاطر الذكاء الاصطناعي متعدد الاستخدامات. المسؤولية جسيمة، ونحن نتطلع قُدُما إلى مواصلة هذه الجهود.