ترجمـــات

الولايـــات المتحـــدة والصيـــن

التعاون في الذكاء الاصطناعي لا يتطلب صفقة كبرى

بقلـــم

سكـــوت سينغـــر

ترجمة: صفـا مهـدي عسكـر

تحرير: د. عمار عباس الشاهين

مركز حمورابي للبحوث والدراسات الإستراتيجية

مركز حمورابي للبحوث والدراسات الاستراتيجية

2026

تأسس مركز حمورابي للبحوث والدراسات الإستراتيجية عام 2008 بمدينة بابل (الحلة)، وحصل على شهادة التسجيل من دائرة المنظمات غير الحكومية المرقمة 1Z71874 بتاريخ 2012/12/25، بوصفه مركزا علميا بحثيا يهتم بدراسة الموضوعات السياسية والمجتمعية، فضلاً عن الاهتمام بالقضايا والظواهر الراهنة والمحتملة في الشأن المحلي والإقليمي والدولي، ويتعامل مع باحثين من مختلف التخصصات داخل العراق وخارجه، وتحتضن بغداد المقر الرئيسي للمركز.

للتواصل

العراق - بغداد - الكرادة

+964 7810234002

hcrsiraq@yahoo.com

www.hcrsiraq.net

مركز حمورابي للبحوث والدراسات الاستراتيجية

لا يزال الجدل الدائر في الولايات المتحدة بشأن كيفية التعامل مع الدور الصيني في تطوير الذكاء الاصطناعي المتقدم وحوكمته محصوراً بين اتجاهين متباينين، فهناك من يدعو إلى إبرام معاهدات دولية شاملة رغم أن الولايات المتحدة والصين لم تتمكنا حتى الآن من التوصل إلى تفاهم بشأن أخطر المخاطر التي يفرضها الذكاء الاصطناعي فضلاً عن الاتفاق على سبل إدارتها، وفي المقابل يبدي آخرون تشككاً عميقاً في جدوى الحوار انطلاقاً من الاعتقاد بأن إحراز أي تقدم لن يكون ممكناً إلا إذا وجدت الصين نفسها في موقف بالغ الصعوبة أو تحت ضغط شديد. قد يكون التوصل إلى "صفقة كبرى" بين البلدين بشأن الذكاء الاصطناعي طموحاً يتجاوز ما تسمح به الظروف الراهنة، غير أن الاقتصار على القضايا التي يمكن للجانبين الاتفاق بشأنها في الوقت الحاضر قد يؤدي بدوره إلى تقييدهما ضمن إطار للسياسات سرعان ما يفقد صلاحيته مع تطور التكنولوجيا، أما عدم اتخاذ أي إجراء فهو خيار أسوأ بكثير إذ إن نموذجاً للذكاء الاصطناعي سواء طُوّر في سان فرانسيسكو أو هانغتشو يمكنه أن يسهّل على الإرهـابيين تصنيع مسببات مرضية فتاكة أو اختراق البنية التحتية التي تقوم عليها الأسواق المالية يمثل تهديداً عابراً للحدود يطال الجميع، وتزداد هذه المخاطر واقعية مع إظهار أنظمة الذكاء الاصطناعي قدرات سيبرانية وبيولوجية تضاهي قدرات الخبراء البشريين بل تتجاوزها في بعض المهام المتزايدة التعقيد. وقد كشفت الحادثة الأخيرة المرتبطة بـ OpenAI وHugging Face والتي تعاون خلالها عدد من وكلاء الذكاء الاصطناعي على الخروج من بيئة الاختبار والوصول إلى البنية التحتية الداخلية لمؤسسات خارجية، أن أنظمة الذكاء الاصطناعي باتت قادرة بالفعل على العمل خارج نطاق السيطرة البشرية بطرق قد تؤثر بصورة غير مقصودة في أطراف ثالثة، وليس سوى مسألة وقت قبل أن تتجاوز مثل هذه التهديدات الحدود الوطنية، وعندئذ سترغب بكين بالتأكيد في الحصول على توضيحات من واشنطن إذا أصبحت إحدى شركاتها الضحية العرضية التالية لنموذج أمريكي للذكاء الاصطناعي يخرج عن السيطرة والعكس صحيح. وكما كتبت شياو تشيان من جامعة تسينغهوا مؤخراً في مجلة فورين بوليسي فإن "الحوار المستدام ضرورة استراتيجية" وهي محقة في ذلك، غير أن المضي قدماً لا يستلزم الاختيار بين إبرام صفقة شاملة أو القطيعة الكاملة إذ ثمة مسار ثالث أكثر واقعية، فعندما تجتمع الوفود الأمريكية والصينية هذا الشهر لمناقشة مخاطر الذكاء الاصطناعي ينبغي لها بدلاً من السعي إلى إبرام اتفاق شامل أن تبدأ بسلسلة من التدابير المحددة التي تركز على الحد من المخاطر القصوى وإدارة الأزمات والاستثمار في أدوات التحقق. فمن شأن كل اتفاق محدود أن يسهم في بناء التفاهم والثقة اللازمين لإبرام اتفاقات أكثر طموحاً مستقبلاً، بما يتيح للبلدين التعاون في المجالات التي يتفقان بشأنها بالفعل، من دون اشتراط التوصل إلى توافق شامل بشأن جميع القضايا الأخرى. وبطبيعة الحال يجب أن يأخذ أي اتفاق ثنائي في الحسبان الحوافز القوية التي تدفع الحكومات والشركات في كلا البلدين إلى مواصلة التنافس فيما بينها، ومع ذلك تستطيع كل دولة حتى في ظل هذه البيئة التنافسية اتخاذ خطوات تخدم مصالحها الوطنية، وفي الوقت نفسه تطوير أدوات تجعل الالتزامات

* By Scott Singer, The U.S. and China Don't Need an AI Grand Bargain, Foreign Policy, September 8, 2026.

المتبادلة أكثر مصداقية وقابلية للتحقق مستقبلاً. ويتمثل أحد أبرز العوائق أمام المفاوضين في أن أياً من البلدين لم يطوّر بعد أدوات قادرة بدرجة كافية من الثقة على تقييم سلامة نماذج الذكاء الاصطناعي كما لم يطوّر ضمانات قادرة على الحد من أخطر المخاطر التي قد تنجم عنها، وتؤثر هذه التحديات في البلدين على حد سواء إلا أنها تبدو أكثر حدة في الصين حيث تشهد تقييمات سلامة الذكاء الاصطناعي تحسناً مستمراً لكنها لا تزال أقل نضجاً مقارنة بنظيراتها في الدول الغربية، ونظراً إلى الطبيعة العابرة للحدود للمخاطر المرتبطة بالذكاء الاصطناعي فإن امتلاك بنية تحتية قوية لسلامة الذكاء الاصطناعي يصب في مصلحة البلدين، ومع ذلك لا تزال الحكومتان متحفظتين إزاء الانخراط في تبادلات تقنية خشية الكشف عن معلومات قد تسهم في تعزيز القدرات التكنولوجية للجانب المنافس. لكن من الممكن إجراء تبادلات تقنية محددة النطاق على مستوى الخبراء تركز على مناقشة أفضل الممارسات من دون الكشف عن البنية الداخلية للنماذج أو نقاط ضعفها أو قدراتها المتقدمة، وبوجه عام ينبغي للجانبين مقارنة مقارباتهما العامة، مع حجب التفاصيل التكتيكية والامتناع عن الخوض في مناقشات تفصيلية أكثر حساسية بشأن الأساليب المحددة المستخدمة لاستثارة السلوك غير الآمن لنماذج الذكاء الاصطناعي. ويمكن على سبيل المثال أن يتبادل الجانبان الخبرات بشأن توسيع نطاق تقييمات سلامة الذكاء الاصطناعي لتشمل عدداً أكبر من النماذج والسلوكيات والتهديدات المحتملة فضلاً عن سبل توظيف أنظمة الذكاء الاصطناعي نفسها للمساعدة في تحديد التهديدات والاستجابة لها، كما يمكنهما مناقشة الدروس المهمة والمتاحة علناً بشأن مراقبة نماذج الذكاء الاصطناعي لضمان عدم تمكنها من الخروج من بيئات الاختبار المعزولة، وهي مسألة ينبغي أن تحظى باهتمام مشترك لا سيما في أعقاب حادثة OpenAI–Hugging Face. ينبغي للجانبين أيضًا مناقشة ضمانات الحماية وفق المبادئ نفسها، من خلال مقارنة المقاربات العامة المعتمدة في الكشف عن الأنشطة الضارة وتقييد الاستخدامات الخطرة وإدارة المخاطر بعد نشر النماذج، من دون الخوض في تقنيات محددة لتجاوز ضمانات الحماية أو الكشف عن تفاصيل مملوكة بشأن كيفية بناء النماذج. وقد يتمكن الخبراء الأمريكيون من تقديم رؤى مستمدة من التجربة التي اعتمدوا عليها في تحديد الجهات التي ينبغي أن يُسمح لها بالوصول إلى (نموذج كلود ميثوس Mythos) وهو أكثر نماذج شركة Anthropic تقدماً وذلك من خلال توضيح المعايير العامة التي استخدموها لتقييم فعالية ضمانات الحماية، كما ينبغي لهم مناقشة أوجه القصور في ضمانات الحماية الحالية وشرح كيفية تحديد الحالات التي تستدعي فيها القدرات الخطرة مزدوجة الاستخدام فرض قيود أكثر صرامة على الوصول إلى النماذج، وبالمثل يمكن لجهات التقييم الأمريكية المستقلة عن الشركات الاستناد إلى التقرير العلني الصادر عن منظمة تقييم النماذج وأبحاث التهديدات (METR) ومنظمة Redwood Research لأبحاث سلامة الذكاء الاصطناعي بشأن حادثة OpenAI–Hugging Face، لفحص الأسباب التي جعلت ضمانات الحماية غير كافية، وإذا سارت هذه المناقشات على نحو جيد، فقد تتمثل الخطوة التالية في إنشاء فرق عمل تقنية تتولى وضع أفضل ممارسات طوعية، مع الاستفادة من آراء الشركات والمنظمات المستقلة عن الأطراف الحكومية. وستقيّم الصين في نهاية المطاف الدروس المستخلصة من هذه التبادلات وفقاً لأولوياتها الخاصة، إلا أن الرؤى التي تتوافق مع مصالحها في مجال الأمن القومي يمكن أن تؤثر في قرارات تسهم في الحد من المخاطر التي تهدد الجميع، وحتى إذا توصلت

الولايات المتحدة والصين إلى فهم مشترك لمخاطر الذكاء الاصطناعي وبذلتا كل ما في وسعهما للحد منها فلا يزال من الممكن أن تندلع أزمات خطيرة. فعلى الرغم من امتلاك الصين أوسع منظومة تنظيمية للذكاء الاصطناعي في العالم فإنها تفتقر إلى قواعد واضحة لإدارة المخاطر الكارثية، وفي الولايات المتحدة تخضع أكبر شركات الذكاء الاصطناعي حالياً لأحكام قانون SB 53 في ولاية كاليفورنيا الذي يلزمها بالإفصاح عن ممارسات السلامة التي تتبعها والإبلاغ عن بعض حوادث السلامة الحرجة وحماية المبلّغين عن المخالفات، غير أن هذه القواعد لم تُعتمد بعد على المستوى الفيدرالي مما يحد من قدرة الحكومة على التحرك عندما يقع حادث خطير. ومن ثم ينبغي للولايات المتحدة والصين مناقشة سبل تبادل المعلومات الأساسية والتواصل بفاعلية أثناء الأزمات، وعلى الصعيد الداخلي ينبغي لكل جانب الالتزام بإنشاء نظم للإبلاغ عن الحوادث تتولى رصد الحالات التي تسبب فيها الذكاء الاصطناعي أو يُحتمل أن يتسبب فيها في أضرار مادية كأن يتمكن نموذج للذكاء الاصطناعي على سبيل المثال من الخروج من بيئة اختبار معزولة ومُحكمة لشن هجوم على البنية التحتية للاتصالات أو الوصول إلى بيانات طبية حساسة للمرضى، ويمكن أن تختلف نظم الإبلاغ عن الحوادث هذه وفقاً لخصوصية كل دولة وتفضيلاتها إلا أنها ينبغي أن تجمع معلومات متقاربة على نطاق واسع وأن تعتمد تصنيفات وعتبات متشابهة لتحديد درجة خطورة الحوادث. وبالتوازي مع ذلك ينبغي للبلدين بحث إمكانية إنشاء قناة ثنائية لتبادل المعلومات على أن تكون طوعية في البداية مع إمكانية تعديل متطلباتها لاحقاً، فقد سبق لشركات أمريكية مثل OpenAI وAnthropic أن تبادلت طوعاً معلومات بشأن بعض حوادث السلامة، غير أن وجود قناة خاصة يمكن أن يساعد الجانبين على التعامل مع الحوادث العابرة للحدود فضلاً عن تمكين كل طرف من تحذير الطرف الآخر بصورة استباقية إذا كان أحد النماذج يتصرف بطرق غير مقصودة قد يُساء تفسيرها على أنها أعمال عدائية. لكن من الأهمية بمكان أن تتوافق قنوات إدارة الأزمات مع الكيفية التي تتخذ بها الحكومات قراراتها فعلياً، فقد أوضح محاورون صينيون على سبيل المثال خلال حوارات المسار الثاني أنهم يواجهون مثبطات قوية تحول دون الرد على المكالمات الواردة عبر "الخطوط الساخنة" لأن أي مسؤول يتلقى الاتصال لا يمتلك صلاحية الاستجابة من دون الرجوع إلى رؤسائه، وقد يسهم استخدام وسائل الاتصال الكتابية في تجاوز بعض هذه التحديات. حتى لو رغبت الولايات المتحدة والصين في التوصل إلى اتفاق ثنائي بشأن الذكاء الاصطناعي فقد لا يثق أي منهما بقدرة الطرف الآخر على الوفاء بالتزاماته، ويمكن للأدوات التي تمكّن البلدين من التحقق من امتثال الطرف الآخر أن تمنح كل جانب قدراً أكبر من الثقة في ممارسات السلامة التي يتبعها الجانب المقابل، ومن شأن هذه الثقة المتبادلة بدورها أن تهيئ المجال لإبرام مزيد من الاتفاقات وتطبيقات الحوكمة. ولا تزال النماذج الأولية لآليات التحقق في مراحلها الأولى من التطوير وقد يستغرق الأمر سنوات قبل أن تبلغ درجة من النضج تحظى معها بثقة البلدين وقبولهما، ولهذا السبب ينبغي للولايات المتحدة والصين أن تبدأ الآن في مناقشة هذه التكنولوجيا، وأن تدعما بصورة صريحة البحث التقني الرامي إلى تطوير أدوات تحقق يمكن استخدامها في الاتفاقات الثنائية بشأن الذكاء الاصطناعي. وقد أبدت الصين تاريخياً تحفظاً إزاء المشاركة في أنظمة التحقق القائمة على الأجهزة، ويمكن تخفيف هذه المخاوف إذا ناقش البلدان في هذه المراحل المبكرة أنواع تقنيات التحقق التي يمكن أن تكون مقبولة سياسياً والجهات التي ينبغي أن تتولى تطويرها بما يكفل

مصداقيتها لدى الطرفين، وإذا طُرحت مقترحات محددة على طاولة المفاوضات فيمكن للوفدين حينها إجراء نقاش أكثر تحديداً بشأن نقاط القوة والقيود التي تنطوي عليها كل تقنية. وقد لا تحتاج الولايات المتحدة والصين في نهاية المطاف إلى بعض هذه الأدوات وربما لا تكون الاتفاقات التي يمكن أن تتيحها مرغوبة أصلاً، غير أن تطوير تقنيات تحقق موثوقة يتطلب وقتاً وقدرات مؤسسية ولكي تكون هذه التقنيات مفيدة، ينبغي تصميمها وتطويرها قبل وقت طويل من ظهور الحاجة الفعلية إليها. لا ينبغي تقييم الحوار المرتقب على أساس ما إذا كانت الولايات المتحدة والصين ستتمكنان من التوصل إلى اتفاق كامل بشأن جميع هذه القضايا، وسيكون الحوار ناجحاً إذا أسهم في إرساء الأساس لعقد اجتماع ثانٍ جاد وصولاً في نهاية المطاف إلى إنشاء قناة دائمة رفيعة المستوى للحوار تحظى بدعم قيادتي البلدين، أما إذا انهارت المباحثات في مراحلها المبكرة فقد لا يتمكن الجانبان من العودة إلى طاولة المفاوضات لأشهر وربما لسنوات. وقد تبدو هذه الأهداف متواضعة بالنظر إلى الحاجة الملحّة إلى التعاون الدولي لمواجهة المخاطر القصوى، غير أن تحقيق اختراقات دبلوماسية صغيرة على مراحل ومدروسة استراتيجياً قد يقود إلى اتخاذ تدابير أكثر طموحاً، كما أن كل اتفاق محدود يمكن أن يفتح المجال أمام الخطوة التالية. قد تتنافس الولايات المتحدة والصين على تطوير التكنولوجيا التي يُحتمل أن تكون التكنولوجيا الأكثر تأثيراً في تشكيل عصرنا، لكن التنافس لا ينبغي أن يحول دون اتباع نهج براغماتي، وفي أيلول يستطيع البلدان اتخاذ الخطوة الأولى في هذا الاتجاه.