



# تنظيم الذكاء الاصطناعي مواجهة تحديات تكنولوجيا ثورية

بقلم

سيباستيان إلباوم وسيباستيان

ترجمة: صفا مهدي عسكر

تحرير: د. عمار عباس الشاهين

مركز حمورابي للبحوث والدراسات الاستراتيجية



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 مركزاً علمياً بحثياً يهتم بدراسة الموضوعات السياسية والاجتماعية، 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.

- لا يجوز إعادة نشر أي من هذه الأوراق البحثية الا بموافقة المركز، وبالإمكان الاقتباس بشرط ذكر المصدر كاملاً.
- لا تعبر الآراء الواردة في الورقة البحثية عن الاتجاهات التي يتبناها المركز وإنما تعبر عن رأي كاتبها.
- حقوق الطبع والنشر محفوظة لمركز حمورابي للبحوث والدراسات الاستراتيجية.

## للتواصل

مركز حمورابي

للبحوث والدراسات الاستراتيجية

العراق - بغداد - الكرادة

+964 7810234002

hcrsiraq@yahoo.com

www.hcrsiraq.net

أواخر عام 2022 أثار إطلاق برنامج "تشات جي بي تي" سابقًا عالميًا لتنظيم الذكاء الاصطناعي فأنشأت إدارة بايدن "معهد سلامة الذكاء الاصطناعي الأميركي" وهو مكتب مكلف بوضع بروتوكولات لتحديد المخاطر المرتبطة بهذه التكنولوجيا، كما أنشأت كل من كندا واليابان وسنغافورة وكوريا الجنوبية والمملكة المتحدة والاتحاد الأوروبي هيئات إشرافية مماثلة واجتمع قادة العالم لمناقشة الذكاء الاصطناعي في بلتشي بارك المقر الريفي في إنجلترا الذي عمل فيه مفككو الشفريات التابعون للحلفاء سرًا خلال الحرب العالمية الثانية، وخلال الخمسة عشر شهرًا التالية عُقدت مؤتمرات متتابعة في سيول وباريس وفي مواجهة تكنولوجيا قادرة على إعادة تعريف العمل والحرب ومفهوم الإنسانية أصبح من الواضح أن الرد السياسي أصبح ضرورة حتى قادة مختبرات الذكاء الاصطناعي التجارية دعوا إلى حوكمة منسقة للتكنولوجيا. اليوم باتت هذه الدعوات نادرة رغم تسارع قوة نماذج الذكاء الاصطناعي التوليدية واستطلاعات الرأي التي تشير إلى قلق الجمهور من المخاطر المحتملة على الوظائف وغيرها ولا يعتقد القادة الحكوميون أو القطاع الخاص أن التنظيم سيحدث على المستوى الوطني أو الدولي، ويشرح جزء من هذا التغيير النفور الفطري لإدارة ترامب من التنظيم خاصة التنظيم العالمي لكن ذلك ليس العامل الوحيد فهناك حوافز قوية لترك التكنولوجيا تتطور بلا قيود، فطفرة الذكاء الاصطناعي تولد جزءًا كبيرًا من نمو الاقتصاد الأميركي وعرقلة تقدمها قد تكون مكلفة كما أن إطلاق نماذج ذكاء اصطناعي صينية قوية مثل "ديب سيك" دفع الحكومة الأميركية إلى تجنب إبطاء المختبرات المحلية خشية تفوق الصين في السباق. حتى إذا انفجرت فقاعة الذكاء الاصطناعي وربما أفلس بعض الشركات الكبرى فإن عمالقة التكنولوجيا ذوي الموارد الضخمة في الولايات المتحدة والصين سيواصلون دفع نشر التكنولوجيا بسرعة وبسبب هذا السباق ستظل احتمالات حوكمة الذكاء الاصطناعي تحديًا كبيرًا ومع ذلك هناك الكثير على المحك بحيث لا يمكن التخلي عن جهود التنظيم، عاجلاً أو آجلاً- ربما بعد كارثة مرتبطة بالذكاء الاصطناعي مثل هجوم إلكتروني على البنية التحتية الحيوية من قبل جهة غير مسؤولة- ستتضح هذه الحقيقة للجميع فالذكاء الاصطناعي يشي باضطرابات اجتماعية ونفسية على نطاق لا يقل عن الثورة الصناعية التي مهّدت لمئة عام من الثورات السياسية والحروب العالمية، وعند نقطة ما ستدرك الحكومات أن رفضها تشكيل مسار الثورة الذكاء الاصطناعي هو تنازل عن المسؤولية.

جهود التنظيم الجزئية في كاليفورنيا وغيرها من الولايات الأميركية لن تخلق إطارًا حوكميًا متماسكًا لكنها توضح أن الكثيرين يشعرون بالقلق من خيار عدم التدخل، لكن للاستعداد لعودة الذكاء الاصطناعي إلى

\* Sebastian Elbaum and Sebastian Mallaby, The AI Trilemma How to Regulate a Revolutionary Technology, foreign affairs, February 13, 2026.

جدول الأعمال الوطني يجب على مناصري التنظيم دراسة نقاط ضعف مواقفهم، فحملة السلامة في 2023-2024 فشلت جزئياً بسبب اتساع نطاقها. فقد تناولت مجموعة واسعة من المناصرين قائمة طويلة ومرهقة من المشكلات التي قد يسببها الذكاء الاصطناعي فقدان الوظائف انخفاض التفكير النقدي في المدارس ومخاطر الأمن القومي وتكاليف بيئية وانتهاكات حقوق الطبع والنشر والهلوسات والتزييف العميق وغيرها، ولكي ينجح تنظيم الذكاء الاصطناعي يجب على المناصرين تحديد الأولويات وصياغة خطة أكثر وضوحاً. ولتحقيق ذلك، عليهم فهم المقايضات بين الأهداف المتنافسة والتعامل مع سوء الفهم الشائع حول كيفية تطور الذكاء الاصطناعي في المستقبل، كما يجب تمييز السياسات القابلة للتطبيق عن غير العملية فالسياسات القابلة للتطبيق تأخذ في الاعتبار حوافز المختبرات الخاصة وتمكّن الحكومة بشكل مناسب. على سبيل المثال يمكن لضريبة المخاطر المفروضة على مختبرات الذكاء الاصطناعي الخاصة أن تشجعها على الاستثمار في أبحاث السلامة بينما ستوفر قاعدة بيانات وطنية مولدة للإيرادات الموارد اللازمة للمسؤولين الحكوميين لمراقبة النماذج المتقدمة يمكن إعادة صياغة النص بشكل سردي متصل كما يلي:

يمكن تجميع أهداف دعاة سياسات الذكاء الاصطناعي في الولايات المتحدة ضمن ثلاثة محاور رئيسية: الأمن القومي، الذي يركز على تعزيز قدرات الجيش وأجهزة الاستخبارات بالذكاء الاصطناعي، والأمن الاقتصادي، الذي يسعى لتطوير الشركات الأميركية في هذا المجال لتعزيز قدرتها التنافسية على الأسواق العالمية، والأمن الاجتماعي، الذي يتضمن الحد من مخرجات الذكاء الاصطناعي الضارة مثل البرمجيات الخبيثة وحماية الوظائف وتقليل التفاوت الاجتماعي ومنع الاستخدامات السيئة للتكنولوجيا وحتى التصدي لسيناريوهات خيالية محتملة مثل أن تؤدي الآلات إلى إبادة البشر. تكمن المشكلة في أن السعي لتحقيق هذه الأهداف الثلاثة في الوقت نفسه يخلق جدول أعمال واسعاً ومعقداً يصعب تحويله إلى إجراءات عملية ملموسة، ويمكن تصور ثلاث سيناريوهات للموازنة بين هذه الأهداف. في السيناريو الأول، قد تسعى الدولة لتعظيم الأمن القومي والاقتصادي عبر تعزيز البحث في الذكاء الاصطناعي وتوسيع مراكز البيانات وتطوير البنية التحتية للطاقة وهو النهج الذي اتبعته إدارة ترامب، لكن محاولة الجمع بين هذا وتعظيم الأمن الاجتماعي تبدو مستحيلة إذ يتطلب الأخير إبطاء نشر تقنيات الذكاء الاصطناعي لإتاحة الوقت لتحديد المخاطر وبناء حلول قبل إطلاق النماذج. أما في السيناريو الثاني، فقد تختار الدولة التركيز على الأمن القومي والأمن الاجتماعي من خلال التعامل مع الذكاء الاصطناعي كتقنية نووية أي حصره في القطاعات العسكرية والطاقة وتقليص استخدامه في المجالات الأخرى ما يحمي الدولة ويحد من الاضطراب الاجتماعي ويقلل فقدان الوظائف ويحد من الاستغلال السيء للتكنولوجيا، لكنه في الوقت نفسه يضحى بالأمن الاقتصادي إذ يحد هذا النهج من التطبيقات التجارية ويحول دون استفادة الصناعات من الكفاءات الجديدة، مما قد يجعل الشركات المحلية تتخلف عن منافسيها الدوليين. وفي السيناريو الثالث، يمكن

للدولة أن تركز على الأمن الاقتصادي والأمن الاجتماعي عبر تشجيع تطوير الذكاء الاصطناعي بوتيرة سريعة مع الالتزام بلوائح صارمة للسلامة قبل نشر النماذج للجمهور وهو ما تصفه بعض شركات التكنولوجيا الكبرى بالابتكار المسؤول، وتعتمد الفكرة على أن السباق لتطوير التكنولوجيا مع نشرها بحذر يخلق دائرة إيجابية تكسب الشركات ثقة الجمهور وتتجنب ردود الفعل السلبية، وتحقق اعتمادًا أسرع على المدى الطويل. ومع ذلك فإن الجمع بين التطوير السريع والنشر الحذر قد يقلل من الأمن القومي، إذ يمكن للدول المنافسة غير الحذرة أن تستفيد من نشر قدراتها العسكرية والسيبرانية أولاً، ما يمنحها الأفضلية الاستراتيجية.

### وهم التفرد الذكائي

أصبح التعامل مع مقايضات سياسات الذكاء الاصطناعي أكثر صعوبة بسبب انتشار فهم خاطئ لكيفية تطور هذه التكنولوجيا المتمثل في رؤية (التفرد الذكائي)، هذا المفهوم اقترحه قبل نحو ثلاثين عامًا الكاتب الخيالي العلمي فيرنور فينج واعتمده لاحقًا المستقبليون مثل راي كورزويل، ويعرّف فينج التفرد على أنه اللحظة التي تصبح فيها نماذج الذكاء الاصطناعي قوية بما يكفي لتطوير شيفراتها الخاصة ما يطلق حلقة مستمرة من التحسين الذاتي المتكرر وينتج عنها انفجار في مستوى الذكاء. إذا تم تبني هذا النموذج الذهني يصبح حل معضلة الذكاء الاصطناعي بسيطًا نسبيًا، فإذا كان التفرد وشيئًا فإن الجهود قصيرة المدى لتعزيز الأمن القومي باستخدام الذكاء الاصطناعي تصبح غير مجدية إذ ستتفوق الذكاءات الفائقة بشكل كبير على أنظمة اليوم، وبالمثل يصبح السعي القصير المدى لتحقيق الأمن الاقتصادي عبر اعتماد الشركات الحالية للذكاء الاصطناعي بلا جدوى إذ ستبتكر هذه الذكاءات طرقًا ثورية لزيادة الإنتاجية، وربما تجعل المفاهيم الاقتصادية التقليدية قديمة.

إذا كان الذكاء الاصطناعي يتجه نحو انفجار ذكاء مفاجئ فإن الهدفين الرئيسيين اللذين يهمان صانعي السياسات هما: الوصول إلى التفرد قبل المنافسين مع الافتراض بأن الأمن العسكري والاقتصادي سيتحقق لاحقًا وإعطاء الأولوية للأمن الاجتماعي عبر الحد من خطر سيطرة أو إبادة الذكاء الآلي للنوع البشري الأضعف، وتمثل هذه النظرة نسخة من نهج "الابتكار المسؤول" حيث يُسابق لتطوير الذكاء الفائق لتحقيق الأمن العسكري والاقتصادي، مع نشره بحذر لتفادي المخاطر المحتملة على البشر. ويبرر بعض أنصار الضوابط التصديرية الأميركية التي تحد من وصول الصين إلى رقائق الذكاء الاصطناعي وتقنيات تصنيعها سياساتهم أحيانًا عبر افتراضات التفرد الذكائي، فهم يعترفون بأن حرمان الصين من أشباه الموصلات المتقدمة سيدفعها لتطوير تقنياتها الخاصة مما يجعلها أكثر قوة على المدى الطويل لكنهم يرون أن هذه المقايضة مجدية إذ أن التفرد وشيك والأولوية هي تأخير تقدم الصين لبضع سنوات بحلول ذلك الوقت

تكون الولايات المتحدة قد فازت بالفعل في سباق الذكاء الاصطناعي، كل ما يهيم في هذا السيناريو هو دفع نهج الابتكار المسؤول إلى أقصى حد وتحقيق "تفرد آمن" أولاً.

إلا أن المشكلة تكمن في أن التفرد الذكائي من غير المرجح أن يحدث بهذه الصورة المفاجئة، صحيح أن نسخة محدودة من التحسين الذاتي المتكرر موجودة بالفعل فالمساعدون البرمجيون المعتمدون على الذكاء الاصطناعي مثل برنامج كلود لشركة أنثروبنيك يساعدون في كتابة شيفرات الجيل القادم من النماذج إلا أن التكنولوجيا لا تكتسب القدرة على التصرف بمفردها، حتى يصبح الذكاء الفائق قويًا يجب وضعه في بيئة يمكنه العمل ضمنها وهذه البيئات ستظل من تصميم البشر وتحت سيطرتهم لفترة طويلة. على سبيل المثال قد تصبح الأنظمة فائقة الذكاء قادرة قريبًا على استبدال معظم المحامين لكن لتحقيق ذلك يجب أولاً منحها الوصول إلى البيانات الصحيحة مثل معلومات العملاء ما يتطلب توضيح المسؤوليات القانونية والحصول على موافقات تنظيمية وتأمين الأنظمة ضد الهجمات ومواجهة ضغوط النقابات المهنية، وبعد تجاوز هذه العقبات يجب منح الأنظمة القدرة على العمل الفعلي بما يشمل صياغة العقود القانونية وكتابة الشيفرات التي تدعم برمجيات الشركات وتشغيلها والتحقق منها ومشاركتها مع أنظمة أخرى، وكل خطوة نحو منح الذكاء الفائق القدرة على الفعل تواجه عقبات قانونية ومؤسسية جديدة لذلك لن تحل هذه الأنظمة محل البشر بسرعة. كما أن سيناريو التفرد يتجاهل العقبات المادية لتوسيع نطاق الذكاء الاصطناعي، لتحقيق الذكاء الفائق وتطبيقه على ملايين المهام تحتاج الأنظمة إلى رفوف من أشباه الموصلات وأنظمة تبريد متطورة وكميات هائلة من الكهرباء، ولجعل ذلك ممكنًا يجب على شركات تصنيع الرقائق بناء مرافق التصنيع وتأمين آلات متطورة لطباعة الدوائر على السيليكون والحصول على العناصر النادرة، وضمان تحديث الشبكات الكهربائية الوطنية بمحطات ومحولات وخطوط نقل جديدة.

تستند رؤية انفجار الذكاء إلى حلقة تغذية راجعة مبسطة مفادها: يكتب الذكاء الشيفرة وتحسن الشيفرة الذكاء ويكتب الذكاء شيفرة أفضل وهكذا، لكن الطريق إلى الذكاء الفائق يتضمن أيضًا مفاوضات بشرية مع الحكومات حول مواقع المرافق الجديدة وعملية تعدين المواد والتفاوض على عقود الطاقة، وجمع رأس المال ومواجهة مخاطر القرصنة والعديد من الاحتكاكات الأخرى التي لا يمكن للآلات التعامل معها. في الواقع من المرجح أن يظهر الذكاء الفائق تدريجيًا وليس في لحظة ثورية واحدة، ولهذا السبب سيطول سباق واشنطن ضد بكين مما يضعف مبرر فرض ضوابط تصدير أشباه الموصلات، وبالمثل يصبح من المشكوك فيه منطلق السباق نحو "تفرد آمن" إذ أن تحقيق التوازن بين أهداف سياسات الذكاء الاصطناعي سيتم على مدى فترة طويلة وليس في لحظة واحدة.

## اختبار العملية

بعد أن فهم دعاة تنظيم الذكاء الاصطناعي معضلة الأهداف الثلاثة وتجنبوا وهم التفرد الذكائي عليهم إتمام خطوة أخيرة وهي مواجهة الحقيقة بأن بعض المخاوف الشعبية حول الذكاء الاصطناعي ليست من الأولويات السياسية، ومن أوضح الأمثلة على ذلك المخاوف المبالغ فيها والمشكلات التي يسعى القطاع الخاص بالفعل لحلها ما يجعل التدخل التنظيمي غير ضروري. تندرج فكرة أن نماذج الذكاء الاصطناعي تنتج "هلوسات" مضللة ضمن كلا الفئتين، فحتى عام 2023 كانت نماذج اللغة الكبيرة غير موثوقة وسامة إلى حد كبير وكان هذا السبب في توتر المختبرات الكبرى قبل إطلاقها، لكن منذ ذلك الحين تراجعت ظاهرة الهلوسة وأظهرت اختبارات الأداء في بعض السياقات أن الأنظمة الذكية أصبحت أكثر قدرة من الخبراء البشريين على الإجابة الصحيحة، ولدى المختبرات كل الحوافز للاستمرار في تحسين دقة النماذج. هناك مخاوف أخرى قد تكون مبررة لكنها لا تصلح لأن تكون على جدول الأعمال التنظيمي لأنها غير عملية، على سبيل المثال تهدد الأنظمة الذكائية وظائف العاملين المعرفيين لكن على صانعي السياسات ألا يجعلوا الحفاظ على الوظائف أولوية إذ لا يمكنهم فعل الكثير حيال ذلك، بعض المحامين والفنيين الإشعاعيين وكتاب السيناريوهات في هوليوود قد يحافظون على وظائفهم إذا تعلموا استخدام الذكاء الاصطناعي، أما الذين يفقدون وظائفهم فيستحقون شبكة أمان عامة مثل الدخل الأساسي الشامل لكن تجميد التقدم التكنولوجي ليس سياسة واقعية.

ينبغي تطبيق اختبار العملية أيضًا على القيود المفروضة على صادرات أشباه الموصلات، فعندما فرضت إدارة بايدن ضوابط شاملة في أواخر 2022 كان الهدف تحقيق الأهداف الثلاثة للمعضلة: الأمن القومي عبر إبطاء التقدم العسكري للصين والأمن الاقتصادي من خلال الحد من تطبيقات الذكاء الاصطناعي في الشركات الصينية والأمن الاجتماعي عبر منع بناء نماذج خطيرة قد تهاجم البشر ما يسمح للولايات المتحدة بمطالبة مختبراتها بالتخلي بالمسؤولية دون التخلي عن التفوق أمام الخصم المتهور. لكن على الرغم من هذه الفوائد المرغوبة فشلت الضوابط في منع الصين من تطوير نماذج متقدمة، إذ يصعب إيقاف تهريب الرقائق ويمكن لمطوري النماذج الصينيين تعويض نقص الرقائق المتطورة باستخدام أعداد كبيرة من الرقائق الأقل تقدمًا أو تحسين البرمجيات لتعويض القيود المادية، والنتيجة أن قطاع الذكاء الاصطناعي في الصين يواصل التقدم بسرعة بينما دفعت هذه القيود الصين إلى تكثيف جهودها في تطوير أشباه الموصلات محليًا. وبالمثل فإن سياسة الحد من النماذج مفتوحة الوزن أي النماذج التي يمكن تحميلها وتعديلها بما قد يجعلها غير آمنة جذابة نظريًا لكنها صعبة التطبيق عمليًا، فالحد من هذه النماذج ينطوي على التضحية بالمكاسب الاقتصادية المترتبة على انتشار الذكاء الاصطناعي بسرعة مقابل تعزيز الأمن القومي والأمني الاجتماعي إذ لن يحصل الخصوم الأجانب على نماذج قوية بسهولة وستقل المخاطر المرتبطة بإنتاج أنظمة قادرة على

التلاعب النفسي كما يحدث في بعض التطبيقات مثل الصور الزائفة، وبالرغم من ذلك فإن العالم الذي تسيطر فيه جميع النماذج على ملكية خاصة لن يكون خالي من الأضرار كما يظهر في النماذج التي تنتج صورًا زائفة فاضحة عند الطلب لكن على الأقل ستكون الحكومات قادرة على محاسبة المختبر المسؤول.

وبما أن حظر جميع النماذج مفتوحة الوزن أمر مستحيل يقترح مؤيدو الحد من انتشارها نهجًا مستهدفًا يركز على النماذج الأقوى المنتجة من قبل شركات معروفة والتي يمكن أن تخضع للضغط التنظيمي مثل الشركات الأميركية أو تلك التي حصلت على تمويل أميركي، فإذا تمكن المنظمون من تهديد مطوري هذه النماذج بحرمانهم من التمويل والعملاء والشراكات يصبح الحد من النماذج مفتوحة الوزن أكثر واقعية، يمكن للإدارة الأميركية المستقبلية تشكيل تحالف من الحلفاء ذوي التوجهات المتقاربة لجعل جمع رأس المال وتشغيل هذه النماذج صعبًا في أميركا الشمالية وأوروبا ومنع مراكز البيانات الكبرى من تشغيل نماذج عالية المخاطر وغير مختبرة ما يشجع الشركات الرائدة على الابتعاد عن النماذج مفتوحة الوزن.

لكن العقبة تكمن في أن مطوري النماذج مفتوحة الوزن في الصين من غير المرجح أن ينضموا إلى هذا التحالف ومع تدهور العلاقات الأميركية الصينية تكاد مختبرات الصين لا ترتبط بالسوق الأميركية، وإذا واصلت المختبرات الصينية إطلاق أنظمة قوية مفتوحة الوزن فلن يكون هناك جدوى من تقييد شركات أميركية مثل ميتا فقط، وبهذا فإن الحد من النماذج مفتوحة الوزن في الولايات المتحدة سيزيد اعتماد الدول النامية التي تفضل هذه النماذج لانخفاض تكلفتها على التكنولوجيا الصينية، ومن دون انفراج في العلاقات بين واشنطن وبكين تفشل سياسة الحد من النماذج مفتوحة الوزن في اختبار العملية.

### احتضان التنازلات

أفضل مسار للمضي قدمًا في تنظيم الذكاء الاصطناعي يكمن في تبني تنازلين أساسيين كل منهما ينطوي على تكلفة اقتصادية متوازنة مقابل مكاسب كبيرة في مجال السلامة المجتمعية يقوم التنازل الأول على فرضية مفادها أن سلامة النماذج الذكية تُعد خيرًا عامًا وخاصًا في الوقت ذاته، ولدى مختبرات الذكاء الاصطناعي حوافز لإنتاج نماذج آمنة لمستخدميها ما يشكل خيرًا خاصًا لكن نموذجًا يبدو غير ضار لمستخدميه قد يضر بالسلامة العامة، على سبيل المثال إذا طلب مستخدم من النموذج إنشاء ونشر آلاف المقالات المختلفة قليلاً حول تزوير انتخابي وهمي فإن حافز المختبر الخاص هو إرضاء هذا العميل من خلال الامتثال لكن هذا الامتثال ذاته سيلحق الضرر بالمجتمع عبر نشر المعلومات المضللة، ونظرًا لغياب الحوافز لدى المختبرات الخاصة لتجنب هذه الأضرار الخارجية فهي لن تستثمر بشكل كافٍ في أبحاث السلامة. يمكن لضريبة خاصة تُسمى "ضريبة المخاطر" معالجة هذا النقص الهدف ليس جمع الإيرادات بل توجيه كيفية تخصيص الشركات لمواردها نحو أبحاث السلامة، فعلى سبيل المثال إذا أنفقت شركة مليار

دولار لبناء نموذج تُفرض عليها ضريبة بنسبة خمسة في المئة أي ما يعادل خمسين مليون دولار ولتعويض ذلك تمنح الحكومة الشركة ائتمناً ضريبياً يعادل 25 في المئة من كل دولار تُنفقه على أبحاث السلامة، فإذا أنفقت الشركة 200 مليون دولار إضافية على أبحاث السلامة فإن الائتمان الناتج يغطي الضريبة بالكامل.

قد يعترض قادة الصناعة بأن إجبار الشركات على الإنفاق في السلامة سيؤخرها ماليًا ويضعف فرصها في السباق ضد مطوري الذكاء الاصطناعي الصينيين غير المقيدين لكن بالنظر إلى المئات من المليارات التي تستثمرها المختبرات الأميركية مجتمعاً في أبحاث الذكاء الاصطناعي فإن هذا التدخل يظل محدوداً، علاوة على ذلك فإن الاستثمارات في السلامة تشكل عائقاً قصير المدى فقط بينما على المدى الطويل ستزيد أبحاث السلامة الجيدة الثقة العامة في الذكاء الاصطناعي، مما يسهل اعتماده على نطاق واسع ويجلب المنافع الاقتصادية المصاحبة. كما أن زيادة الإنفاق على السلامة يعزز الأمن المجتمعي ويحقق فوائد إضافية مثل حماية مختبرات الجامعات التي تعاني من استنزاف الموارد، فمع هيمنة المختبرات الخاصة على الأجهزة والمواهب تتدهور مختبرات الجامعات الممولة حكومياً وهي المصدر التاريخي للتفوق التكنولوجي الأمريكي، من خلال السماح للمختبرات الخاصة بالحصول على ائتمانات ضريبية مقابل تمويل منح السلامة التابعة للمؤسسة الوطنية للعلوم أو توفير الوصول إلى موارد الحوسبة لمجموعات جامعية فإن ضريبة المخاطر ستعيد الحيوية للقطاع الأكاديمي، فالمؤسسة الوطنية للعلوم تخصص حوالي مليار دولار سنوياً لبرامج علوم الحاسوب الجامعية ومع ارتفاع تكاليف التدريب في الصناعة بشكل مضاعف فإن دعم ضريبة المخاطر سيحدث أثراً كبيراً على قطاع علوم الحاسوب الأكاديمي.

بالإضافة إلى تشجيع الإنفاق على السلامة ينبغي للحكومة استثمار بياناتها الخاصة لتعزيز التعاون مع شركات الذكاء الاصطناعي، فحتى الآن سيطرت الشركات الخاصة على جمع البيانات اللازمة لتدريب النماذج مستغلة الإنترنت المفتوح وما وراءه لكن إمدادات النصوص العامة عالية الجودة قاربت على النفاد مما يحول بؤرة الابتكار نحو مجموعات البيانات غير العامة بما في ذلك بعض البيانات التي تحتفظ بها الحكومة الأميركية، ففي العام الماضي أعلنت إدارة ترامب خطة لإتاحة مجموعات البيانات العلمية التي جمعتها 17 مختبراً وطنياً لتدريب النماذج ويمكن توسيع هذه المبادرة لتشمل مجالات أخرى مثل الرعاية الصحية حيث يمكن للبيانات السريرية المجهولة الهوية تمكين النماذج من تشخيص الأمراض ووصف العلاج وفي المجال الاقتصادي يمكن استخدام البيانات الضريبية المجهولة لتوقع التغيرات في الادخار والاستهلاك.

يمكن لإدارة ترامب إنشاء مستودع بيانات وطني تحت مظلة مركز معايير وابتكار الذكاء الاصطناعي الذي أنشأته إدارة بايدن سابقاً كمؤسسة السلامة الأميركية للذكاء الاصطناعي، سيتولى المستودع تنظيف وتحويل بيانات حكومية متنوعة غير منشورة إلى صيغ مجهولة الهوية وفق أعلى معايير الخصوصية ثم

مشاركتها مع مطوري الذكاء الاصطناعي في القطاع الخاص مقابل رسوم، على غرار ما يدفعونه للشركات الخاصة المتخصصة في تنسيق البيانات لتُعاد الإيرادات بعد ذلك لتعزيز قدرات المركز في مراقبة سلامة النماذج المطورة.

بمجرد توافر الموارد الكافية يحتاج المركز إلى سلطة إضافية، فقد طلبت إدارة بايدن من مطوري النماذج المتقدمة مشاركة معلومات السلامة مع مكتب الصناعة والأمن بوزارة التجارة لكن إدارة ترامب أوقفت هذا الالتزام بشكل غير حكيم، وبصرف النظر عن حق الاطلاع يجب أن يمتلك المركز سلطة التصرف فكما يمكن لإدارة الغذاء والدواء منع بيع أدوية غير آمنة، يجب أن تكون لدى الجهات التنظيمية للذكاء الاصطناعي السلطة لرفض إطلاق النماذج الخطرة.

تمامًا كما هو الحال مع ضريبة المخاطر الهدف من تعزيز المركز هو زيادة كبيرة في السلامة المجتمعية مقابل خسارة طفيفة في الأمن الاقتصادي ودون الإضرار بالأمن القومي، صحيح أن اختبار النماذج قبل إطلاقها سيفرض تكاليف وتأخيرات محدودة على المختبرات لكن ذلك يعوضه توفر مجموعات بيانات عامة منظمة لتدريب النماذج كما أن مراجعة الجمهور للنماذج ستسهم في تسريع قبولها، وفي الوقت نفسه يمكن أن يتم اختبار النماذج قبل الإطلاق دون المساس بالأمن القومي إذا سمح للمختبرات الخاصة بمشاركة النماذج مع مسؤولي الدفاع والاستخبارات قبل مراجعتها من قبل المركز، بحيث يتمكن القادة الأمنيون من التخطيط لاستخدام النماذج الجديدة قبل أن يعلم بها باقي العالم.

### نحو انتصار جديد في عدم الانتشار

قد يعترض بعض النقاد على هذه المقترحات معتبرين أنه لا جدوى من فرض ضرائب على مطوري الأنظمة المملوكة في الولايات المتحدة أو دفعهم للإنفاق أكثر على السلامة أو إنشاء وكالة لمراقبة النماذج الضارة وإيقافها إذا كان هناك احتمال أن تنتج الصين أو مطورو النماذج المفتوحة أنظمة غير آمنة، لكن هناك حجة قوية لتنظيم النماذج المملوكة اتحاديًا حتى في ظل انتشار النماذج الصينية أو المفتوحة على نطاق واسع وذلك لأن النماذج الأميركية المملوكة تظل أقوى لا سيما في المهام التي تتطلب استدلالًا معقدًا وقدرات متعددة الوسائط، وإذا كان هدف المنظمين الأميركيين هو متابعة النماذج الأكثر ابتكارًا وقوة وبالتالي الأكثر تهديدًا فإن التركيز على المختبرات الأميركية المملوكة هو الخيار الأنسب.

علاوة على ذلك فإن فرض الضرائب وتنظيم الذكاء الاصطناعي المملوك سينتج عنه تأثيرات إيجابية على بقية الصناعة، فإجبار المختبرات المملوكة على تمويل أبحاث السلامة ستحفز الصناعة لتطوير ميزات تعزز الأمان في جميع النماذج، فعندما ظهرت تقنيات التشفير والمصادقة متعددة العوامل لأول مرة كانت مقتصرة على البرمجيات عالية الجودة ومع مرور الوقت أصبحت معيارًا عالميًا، وبالمثل ستبدأ

منهجيات سلامة الذكاء الاصطناعي كأفكار مكلفة ومتخصصة ثم تتحول إلى معيار صناعي، وبمجرد أن يظهر رائد من المختبرات المملوكة قيمة ميزة السلامة ستعتبر أي نماذج مفتوحة بدونها ناقصة أو خبيثة. هذا الضغط السوقي مدعومًا بقدرة مركز الذكاء الاصطناعي على إدراج النماذج غير المطابقة في القائمة السوداء، يضمن أن الابتكارات في السلامة، التي تم تمويلها عبر فرض ضرائب على النماذج المملوكة، ستلزم الصناعة كلها بالمعايير نفسها.

إضافة إلى ذلك فإن بناء هيكل تنظيمي للنماذج المملوكة قد يصبح خطوة تمهيدية لمشروع أوسع للحكم على الذكاء الاصطناعي يمتد في النهاية إلى اتفاق دولي للحد من النماذج المفتوحة، اليوم أحد أسباب صعوبة السيطرة على النماذج المفتوحة هو سوء العلاقات الأميركية- الصينية لكن العلاقات قد تتغير مستقبلاً، خلال الحرب الباردة تحولت فترات التوتر إلى فترات تخفيف للصراع، ففي عام 1968 وبعد ست سنوات فقط من أزمة الصواريخ الكوبية وفي ذروة حرب فيتنام وقعت الولايات المتحدة والاتحاد السوفيتي وأكثر من خمسين دولة أخرى على معاهدة الحد من انتشار الأسلحة النووية والتي نجحت في تأخير انتشار الأسلحة النووية، ويمكن للولايات المتحدة اليوم أن تطمح لتحقيق إنجاز مماثل بالنسبة للنماذج المفتوحة للذكاء الاصطناعي عبر وضع البنية المؤسسية لتنظيم النماذج المملوكة ما يهيئها لتحقيق انتصار جديد في عدم الانتشار.

بالنظر إلى تحديات تنظيم تكنولوجيا تتغير بسرعة والسيطرة على سباق عالمي على الهيمنة يواجهه مؤيدو تنظيم الذكاء الاصطناعي خيارات صعبة، حيث لا مفر من التضحية ببعض المكاسب الاقتصادية أو الاجتماعية أو المتعلقة بالأمن القومي، لكن من الأفضل تبني هذه التنازلات بوعي بدل تجاهلها فإذا فشلت الحكومات في اتخاذ خيارات مدروسة بشأن الذكاء الاصطناعي، ستتقدم التكنولوجيا وتتخذ هذه الخيارات نيابة عنها.