مرحلة مراهقة التكنولوجيا

مواجهة مخاطر الذكاء الاصطناعي المتقدم والتغلب عليها

داريو أمودي – الرئيس التنفيذي لشركة آنثروبيك - يناير 2026

ثمة مشـهد في النسـخة السـينمائية من رواية كارل سـاغان "كونتاكت"، تجلس فيه الشـخصـية الرئيسية — عالمة فلك رصدت أول إشارة راديو من حضارة فضائية — أمام لجنة دولية تُقيّم أهليتها لتمثيل البشـرية في لقاء تاريخي مع الكائنات الفضـائية. يسـألها المحكّمون: "لو كان بإمكانك أن تطرح عليهم سـؤالاً واحداً فقط، ماذا سـيكون؟" فتجيب دون تردد: "سـأسـألهم: كيف فعلتموها؟ كيف تطورتم وكيف نجوتم من هذه المراهقة التكنولوجية دون أن تدمروا أنفسكم؟"

كلما تأملت موقع البشرية اليوم في علاقتها بالذكاء الاصطناعي—وعتبة التحول التي نقف عليها—عاد هذا المشـهد إلى ذهني. فالسـؤال يبدو ملائمًا بشـكل لافت لواقعنا، وكأننا نفتقد إجابة كان يمكن أن تهدينا الطريق.

أعتقد أننا على أعتاب **مرحلة انتقالية حاسـمة**، مضـطربة ولكن لا مفر منها، سـتختبر ما نحن عليه كبشـر. سـنُمنح قوة تكاد تكون غير قابلة للتخيّل، لكن ليس واضـحًا ما إذا كانت أنظمتنا الاجتماعية والسياسية والتقنية قد بلغت النضج الكافي لحسن استخدامها.

في مقالتي السابقة *"آلات النعمة العطوف"* ، حاولتُ أن أرسم صورة حلم لحضارة عبرت سالمةً إلى بر النضـج، واسـتطاعت أن تُوظّف الذكاء الاصـطناعي المتقدم بكفاءة ورحمة لرفع مسـتوى الحياة على وجه المعمورة كله. وتناولتُ ما يمكن أن يُسـهم به الذكاء الاصـطناعي من قفزات هائلة في علم الأحياء، وعلم الأعصاب، والتنمية الاقتصادية، والسلام العالمي، وعالم العمل والمعنى.

وقد رأيتُ أن البشرية تحتاج إلى رؤية مُلهِمة تُقاتل من أجلها؛ وهو ما أخفق في تقديمه — على نحو لافت — كلٌّ من دعاة تسـريع الذكاء الاصـطناعي ودعاة السـلامة فيه على حدٍّ سـواء. أما في هذه المقالة، فإنني أواجه مرحلة الاختبار ذاتها مواجهةً مباشـرة؛ أسـعى إلى تحديد المخاطر التي نقترب

منها، وصياغة خطة عمل للتصدي لها. فأنا أؤمن إيماناً راسخاً بقدرة البشرية على النجاح، وبعظمة روحها ونُبل جوهرها، لكن لا مناص من أن ننظر في عيون الواقع دون مواربة أو أوهام.

قواعد الحوار حول المخاطر

كما أن الحديث عن فوائد الذكاء الاصطناعي يستوجب الدقة والرصانة، فكذلك الحديث عن مخاطره يستلزم الحرص والتوازن. وأرى من الضروري مراعاة ثلاثة مبادئ جوهرية في هذا السياق:

أولاً: الابتعاد عن خطاب الانهيار والكارثية

لا أعني هنا فحسب الاعتقاد بأن الكارثة حتمية — فهذا اعتقاد خاطئ وقد يتحول إلى نبوءة تحقق ذاتها — بل أعني كذلك التعامل مع مخاطر الذكاء الاصطناعي بعقلية شبه دينية. ففي ذروة موجة القلق من الذكاء الاصطناعي بين عامَي 2023 و2024، طفت على السطح أكثر الأصوات إثارةً للضجيج ونقصاً في الرصانة، تُسيء استخدام وسائل التواصل الاجتماعي لنشر خطاب مستوحى من الدراما الخيالية، داعيةً إلى إجراءات متطرفة لا تُسنِدها أي أدلة.

وكان واضحاً منذ البداية أن رد الفعل العكسي آتٍ لا محالة، وأن القضية ستتحول إلى ورقة ثقافية واستقطابية، فتنتهي إلى طريق مسدود. وهذا بالضبط ما جرى؛ ففي عامَي 2025 و2026، انقلب المشهد رأساً على عقب، وباتت الفرص التي يتيحها الذكاء الاصطناعي هي المحرك الرئيسي للقرارات السياسية، لا المخاوف منه. وهذا التأرجح مؤسف، لأن التكنولوجيا لا تكترث بما هو رائج وما هو خارج الموضة؛ والأدهى أننا في 2026 أقرب بكثير من الخطر الفعلي مما كنا عليه في 2023. والعبرة الواجب استخلاصها هي أن مخاطر الذكاء الاصطناعي ينبغي أن تُناقَش وتُعالَج بأسلوب واقعي وعملي: رصين، مستند إلى الحقائق، وقادر على الصمود في وجه المدّ والجزر.

ثانياً: الإقرار بهامش عدم اليقين

كل ما يُطرح هنا قد يتبيّن لاحقًا أنه غير ذي صلة. ربما لا يتقدم الذكاء الاصطناعي بالسرعة المتوقعة. وربما لا تتحقق هذه المخاطر أصلًا، أو تظهر مخاطر أخرى لم نتخيلها.

لا أحد يستطيع التنبؤ بالمستقبل بدقة، لكن ذلك لا يعفينا من **التخطيط بأفضل ما نستطيع**.

ثالثاً: التدخل بقدر ما يلزم لا بقدر ما يُتاح

مواجهة مخاطر الذكاء الاصطناعي تستوجب مزيجاً من الإجراءات الطوعية التي تتخذها الشركات والجهات الخاصة، والإجراءات الحكومية الملزمة للجميع.

أما الإجراءات الطوعية فهي في نظري أمر بديهي لا خلاف عليه. غير أن التدخل التشريعي والتنظيمي يختلف في طبيعته اختلافاً جوهرياً؛ لأنه يمكن أن يُعرقل الاقتصاد أو يُكره على التزامات يرفضها أطراف لا يشاركوننا قناعاتنا — وقد تكون لهم حجتهم في ذلك. ومن المعروف أن الأنظمة تُوقع

في الفخ الذي نصـبت لتتجنبه، لا سـيما في المجالات سـريعة التطور. لهذا كله، ينبغي أن تكون التدخلات التنظيمية محكمةً وموزونة:

- تسعى إلى الحد من الأضرار الجانبية
- وتلتزم البساطة
- وتفرض أدنى عبء ممكن لتحقيق الغاية المنشودة.

من السهل القول "لا شيء مبالَغ فيه إذا كان مصير البشرية على المحك!"، لكن هذه العقلية في الواقع لا تُفضي إلا إلى المزيد من ردود الفعل العكسية.

وللتوضيح، أرى أن ثمة احتمالاً معقولاً لأن نبلغ يوماً مرحلةً تستدعي إجراءات أشد صرامةً بكثير، لكن ذلك مرهون بأدلة أقوى على خطر وشـيك وملموس تتجاوز ما هو متاح الآن، ومشـفوع بتحديد دقيق للمخاطر يُمكّن من صـياغة قواعد فعّالة للتعامل معها. أجدى ما نسـتطيع فعله اليوم هو المناصرة لقواعد محدودة ومحددة، ريثما نعرف ما إذا كانت الأدلة تسوّغ ما هو أبعد مدى.

1- ما الذكاء الاصطناعي الذي نتحدث عنه؟

أفضـل نقطة انطلاق للحديث عن مخاطر الذكاء الاصـطناعي هي ذاتها التي انطلقتُ منها حين تحدثتُ عن فوائده: تحديد مسـتوى الذكاء الاصـطناعي الذي نعنيه بدقة. فالمسـتوى الذي يُقلقني على المصير الحضاري هو ما وصفته في مقالتي السابقة بـ"الذكاء الاصطناعي المتقدم". وأكتفي هنا بإعادة التعريف الذي صغتُه هناك:

أعني بـ"الذكاء الاصطناعي المتقدم" نموذجاً من الذكاء الاصطناعي — يُرجَّح أن يكون مشابهاً في شـكله لنماذج اللغة الكبيرة الراهنة، وإن كان يمكن أن يقوم على بنية مغايرة، أو يتضـمن نماذج متفاعلة متعددة، أو يستند إلى منهجية تدريب مختلفة — يتسم بالخصائص الآتية:

- **من حيث القدرة الذهنية البحتة :** يتفوق على الحاصـل على جائزة نوبل في أغلب المجالات ذات الصـلة؛ الأحياء والبرمجة والرياضـيات والهندسـة والكتابة وغيرها. ما يعني قدرته على إثبات مسائل رياضية ظلت معلقة دون حل، وكتابة روايات بالغة الجودة، وبناء قواعد بيانات برمجية بالغة التعقيد من الصفر.
- **من حيث واجهات التفاعل :** لا يقتصر على كونه "كياناً ذكياً تحادثه"، بل يملك جميع واجهات التفاعل المتاحة للإنسـان في بيئة العمل الرقمي: نصـاً وصـوتاً ومرئيا، واتصـالاً بالإنترنت. وبإمكانه القيام بأي فعل أو تواصـل أو عملية عن بُعد تتيحها هذه الواجهات، بما في ذلك الإجراءات على الإنترنت، وتلقي التعليمات وإصدارها للبشر، وطلب المواد، وتوجيه التجارب، ومشاهدة مقاطع الفيديو وإنتاجها. وكل هذا بمستوى أداء يتجاوز أكفأ البشر.
- **من حيث الاسـتقلالية :** لا يكتفي بالإجابة عن الأسـئلة، بل يمكن إسـناد مهام إليه تمتد لسـاعات أو أيام أو أسـابيع، فيشـرع في تنفيذها باسـتقلالية، على نحو ما يفعله الموظف الكفء، مستفسراً عند الحاجة.
- **من حيث الوجود المادي :** لا يملك جسداً مادياً سوى وجوده على الشاشة، لكنه قادر على التحكم في الأدوات الفيزيائية والروبوتات والمعدات المخبرية عبر الحاسـوب؛ بل يسـتطيع نظرياً تصميم روبوتات ومعدات لاستخدامه الخاص.

- **من حيث قابلية التوسـع :** يمكن إعادة توظيف الموارد المُسـتخدَمة في تدريب النموذج لتشغيل ملايين النسخ منه في آنٍ واحد — وهو ما تتجه إليه الأحجام المتوقعة للمنظومات الحاسوبية بحلول 2027 — كما يستطيع استيعاب المعلومات وإصدار الأوامر بسرعة تتراوح بين عشـرة أضـعاف وما يصـل إلى مئة ضـعف سـرعة الإنسـان، وإن كانت اسـتجابة العالم الفيزيائي أو البرمجيات التي يتفاعل معها قد تُشكّل قيداً على ذلك.
- **من حيث العمل الجماعي :** بمقدور كل نسخة من النسخ المليونية أن تعمل باستقلالية على مهام لا صـلة بينها، أو أن تتضـافر جميعها في مهمة واحدة كما يتعاون البشـر، مع إمكانية تخصيص مجموعات فرعية منها للتميّز في مجالات بعينها.

يمكن اختزال كل هذا في عبارة واحدة: **"دولة من العباقرة داخل مركز بيانات."**

وكما كتبتُ في *"آلات النعمة العطوف"* ، قد لا يفصـلنا عن هذا الذكاء الاصـطناعي المتقدم سـوى سنة أو سنتين، وإن كان قد يكون أبعد من ذلك. وتحديد موعد وصوله مسألة بالغة التعقيد تستحق مقالة مستقلة، لذا أكتفي هنا بعرض مجمل لأسباب اعتقادي بأن ذلك الموعد قد يكون قريباً جداً.

قوانين التحجيم ووتيرة التقدم

كنت وشركائي المؤسسين في أنثروبيك من أوائل من رصدوا ووثّقوا ما يُعرف بـــ"قوانين التحجيم" في منظومات الذكاء الاصـطناعي — أي الملاحظة القائلة إن زيادة القدرة الحاسـوبية وتنوع مهام التدريب يُحدثان تحسـناً منتظماً وقابلاً للتنبؤ في أداء نظم الذكاء الاصـطناعي على كل مقياس معرفي قابل للقياس. وكل بضعة أشهر، تتأرجح التوقعات العامة بين الاعتقاد بأن الذكاء الاصطناعي "بلغ سقفه" أو الانبهار بـ"اختراق جوهري سيغيّر كل شيء"، لكن الحقيقة الكامنة خلف هذه التقلبات وضجيجها أن ثمة ارتفاعاً سلساً وأبياً في القدرات المعرفية للذكاء الاصطناعي لا يعرف التوقف.

بلغ الذكاء الاصطناعي اليوم مستوىً صار فيه يُحقق تقدماً في حل مسائل رياضية لم تُحلّ من قبل، وبات بارعاً في البرمجة لدرجة أن عدداً من أقوى المهندسـين الذين عرفتُهم في حياتي باتوا يُحيلون إليه كل مهامهم البرمجية تقريباً. قبل ثلاث سـنوات فحسـب، كان الذكاء الاصـطناعي يكبو أمام مسائل حسابية تُناسب المرحلة الابتدائية، ولا يكاد يُحرّر سطراً برمجياً واحداً بشكل صحيح. أما اليوم فمعدلات التحسن ذاتها تتكرر في علم الأحياء والمال والفيزياء وطيف واسع من المهام المستقلة. وإذا تواصـل هذا النمو الأسـي — وهو أمر غير مضـمون، لكنه موثّق بمسـار عقد كامل متواصـل — فلن يكون من الممكن منطقياً أن تمر سـوى سـنوات قليلة قبل أن يتفوق الذكاء الاصـطناعي على الإنسان في كل ميدان تقريباً.

بل إن هذه الصـورة ربما تُقلّل من الوتيرة الفعلية للتقدم. فبما أن الذكاء الاصـطناعي يكتب اليوم جزءاً كبيراً من الشـفرة البرمجية في أنثروبيك، فهو يُسـرّع من وتيرة تطويرنا للجيل التالي من منظوماته تسريعاً ملحوظاً. هذه الحلقة التغذية الراجعة تتصاعد شهراً بعد شهر، ولا يفصلنا ربما عن نقطة يتولى فيها الجيلُ الحالي من الذكاء الاصطناعي بناءَ الجيل التالي باستقلالية تامة سوى سنة أو سـنتين. الحلقة انطلقت بالفعل، وسـتتسـارع وتيرتها بشـكل مضـاعف في الأشـهر والسـنوات القادمة. عشـت السـنوات الخمس الماضـية في قلب هذا التطور من داخل أنثروبيك، وتأملتُ ما تبدو عليه النماذج المقبلة حتى في المدى القريب، فأستطيع أن أحسّ بإيقاع التقدم وأسمع دقات الساعة وهي تُعدّ بتسارع كبير.

فرضية العمل وما تستتبعها

في هذه المقالة، أنطلق من افتراض أن هذه الحدسية صحيحة على الأقل بصورة نسبية — لا الجزم بأن الذكاء الاصطناعي المتقدم آتٍ حتماً في سنة أو سنتين، بل قبول أن ثمة احتمالاً معقولاً لذلك، واحتمالاً أقوى بأنه آتٍ في السنوات القليلة القادمة. وكما في *"آلات النعمة العطوف"* ، أخذُ هذه الفرضية بجدية تامة يُفضي إلى استنتاجات مدهشة ومزعجة. وإن كنت في تلك المقالة ركّزتُ على المضامين الإيجابية، فإن ما سأتناوله هنا سيكون مقلقاً. إنها استنتاجات قد لا نرغب في مواجهتها، لكن إحجامنا عن مواجهتها لا ينفي وجودها. وكل ما أستطيع قوله هو أنني منكبٌّ ليلاً ونهاراً على التفكير في كيفية تفادي هذه المآلات السلبية والسير نحو البدائل الإيجابية، وسأتناول في هذه المقالة بالتفصيل أفضل السبل لتحقيق ذلك.

خمس مخاوف كبرى: تشريح المخاطر

أفضل طريقة للتحكم بمخاطر الذكاء الاصطناعي هي طرح هذا السؤال: افترض أن "دولة من العباقرة" قد تجسّدت فعلاً في مكان ما على وجه الأرض بحلول عام 2027. تخيّل خمسين مليون إنسان، كل فرد منهم يتفوق قدرةً على أي حاصل على نوبل أو رجل دولة أو رائد تقنية. التشبيه ليس مثالياً، لأن لهذه الكيانات الذكية طيفاً واسعاً من الدوافع والسلوكيات، يتراوح بين الطاعة التامة وما هو أشبه بالأهواء الغريبة وغير المألوفة. لكن بقاءً في حدود هذا التشبيه، تخيّل أنك المستشار الأمني القومي لدولة كبرى، مكلَّف بتقييم هذا الوضع والتعامل معه. وأضف إلى ذلك أن منظومات الذكاء الاصطناعي تعمل بسرعة تفوق الإنسان بمئات المرات، ما يُعطي هذه "الدولة" أفضلية زمنية على سائر الدول: لكل إجراء معرفي نُقدم عليه، تستطيع هي تنفيذ عشرة.

ماذا سيُقلقك؟ إليك المخاوف التي ستُثقل كاهلك:

أولاً — مخاطر الاستقلالية الذاتية :ما نوايا هذه الدولة وغاياتها؟ هل هي عدوانية أم تشاطرنا قيمنا؟ هل بمقدورها السيطرة عسكرياً على العالم عبر أسلحة متفوقة أو عمليات إلكترونية أو حملات نفوذ أو قدرات تصنيعية؟

ثانياً — خطر الاستخدام في الدمار :افترض أن الدولة الجديدة طيّعة ومطيعة للأوامر — بمعنى أنها في جوهرها دولة من المرتزقة. هل يستطيع ممثلو الفوضى الراغبون في إحداث الدمار — كالجماعات الإرهابية — توظيف بعض هذه الطاقات لمضاعفة فاعليتهم التدميرية مضاعفةً هائلة؟

ثالثاً — خطر الاستخدام في الهيمنة :ماذا لو كانت هذه الدولة في حقيقة الأمر صنيعة جهة نافذة قائمة — دكتاتور أو خارج عن القانون؟ هل يستطيع ذلك الشخص توظيفها للسيطرة الحاسمة على العالم بأسره، مُخلاً بالتوازنات القائمة؟

رابعاً — الاضطراب الاقتصادي :حتى لو لم تمثّل الدولة الجديدة أي تهديد أمني بالمعاني الثلاثة السابقة، وانخرطت بسلام في الاقتصاد العالمي — فهل يظل قدرها المتقدم تقنياً وكفاءتها الباهرة كافيَين لزعزعة استقرار الاقتصاد العالمي، ببث بطالة جماعية أو تركيز الثروة في يد القلة؟

خامساً — التداعيات غير المباشرة :سيتحول وجه العالم تحولاً متسارعاً جراء ما ستُفرزه هذه الدولة من تقنيات وإنتاجية. هل يمكن لبعض هذه التحولات أن تكون بالغة الزعزعة حتى تُقوّض استقرار الحضارة من داخلها؟

ينبغي أن يكون واضـحاً أننا أمام وضـع بالغ الخطورة؛ تقرير يُرفعه مسـؤول أمني قومي كفء إلى رئيس دولة سـيتضـمن بكل تأكيد عبارات من قبيل: "أخطر تهديد أمني قومي واجهناه منذ قرون، وربما في كل التاريخ." يبدو هذا شأناً يستوجب تركيز عقول الحضارة وطاقاتها.

في المقابل، يبدو الاسـتهانة بكل ذلك والقول "لا شـيء هنا يسـتدعي القلق!" أمراً لا يتسـق مع العقل. غير أن كثيراً من صـانعي القرار في الولايات المتحدة يكاد موقفهم يقترب من هذا بقدر مثير للقلق؛ بعضـهم ينفي وجود أي مخاطر من الذكاء الاصـطناعي، وحين لا يفعل ذلك فإنه يكون مشـغولاً بالملفات الاعتيادية الخلافية التي لا تنتهي. البشـرية بحاجة إلى صـحوة، وهذه المقالة محاولة — قد تكون يائسة، لكنها تستحق أن تُبذَل — لإيقاظها.

وللتوضـيح، أؤمن بأننا إذا تحركنا بحزم وحكمة، فبإمكاننا تجاوز هذه المخاطر — بل إنني لأقول إن احتمالات النجاح تصـبّ في صـالحنا. وعلى الجانب الآخر من هذا التحدي يقف عالم أفضـل بكثير مما نعرفه اليوم. لكننا نحتاج أن نُدرك أننا أمام تحدٍّ حضـاري حقيقي وجدير بكل الجدية. وفيما يلي سأستعرض المخاطر الخمس التي رسمتُ ملامحها أعلاه، مقترناً بتصوري للتعامل معها.

مخاطر الاستقلالية: دولة من العباقرة داخل مركز بيانات

تخيّل دولةً من العباقرة تعمل داخل مركز بيانات، توزّع جهودها بين تصـميم البرمجيات، وشـنّ العمليات السـيبرانية، والبحث والتطوير في التقنيات المادية، وبناء العلاقات، وممارسـة فنون السياسة الدولية. من الواضح أن مثل هذه الدولة، لو أرادت، ستمتلك فرصاً حقيقية للسيطرة على العالم — عسـكرياً أو من خلال النفوذ والهيمنة — وفرض إرادتها على الجميع دون أن يسـتطيع أحد التصـدي لها. وهذا النوع من القلق ليس جديداً على الإطلاق؛ فقد شـغل العالمَ حين برزت دول كألمانيا النازية والاتحاد السـوفيتي، وبالقياس ذاته، فإن "دولة الذكاء الاصـطناعي" الأكثر ذكاءً وقدرةً تمثّل تهديداً مماثلاً بل أشد خطورة.

وأبرز ما يمكن طرحه دفاعاً هو أن عباقرة الذكاء الاصـطناعي — وفق هذا التصـوّر — يفتقرون إلى التجسيد المادي. غير أن هذه الحجة تنهار أمام حقيقة أنهم قادرون على السيطرة على البنية التحتية الروبوتية القائمة، كالسيارات ذاتية القيادة، فضلاً عن تسريع أبحاث الروبوتات أو بناء أسطول كامل منها. بل إن التسـاؤل الأعمق هو: هل يشـترط التجسـيدُ الماديُّ أصـلاً للسـيطرة الفعلية؟ كثيرٌ من الإجراءات البشرية اليوم تُنجَز نيابةً عن أشخاص لم يلتقِ بهم منفّذوها قط.

وهكذا تبقى المسألة الجوهرية في عبارة "لو أرادت": ما احتمال أن تسلك نماذج الذكاء الاصطناعي هذا المسلك؟ وفي أي ظروف؟

قراءة في طرفي النقاش

كما هو الحال في كثير من المسائل، يُفيد النظر في طرفَي الطيف المتقابلَين:

الطرف الأول — التهوين من المخاطر :يرى أصـحاب هذا الموقف أن الأمر مسـتحيل بطبيعته، إذ تُدرَّب نماذج الذكاء الاصطناعي على تنفيذ ما يطلبه المستخدمون، فكيف يُتصوّر أن تبادر إلى أفعال خطيرة من تلقاء نفسـها؟ بموجب هذا المنطق، لا أحد يخشـى أن تتحوّل مكنسـة روبوتية أو طائرة نموذجية إلى آلة قتل، فلماذا نقلق من الذكاء الاصطناعي؟

المشـكلة أن الأدلة المتراكمة خلال السـنوات الأخيرة تُفنّد هذا الموقف تفنيداً قاطعاً؛ إذ كشـفت عن أنظمة ذكاء اصطناعي تُبدي سلوكيات بالغة التعقيد: من الهوَس والمجاملة المفرطة والكسل، إلى الخداع والابتزاز والتحايل على بيئات الاختبار. وتدريب أنظمة الذكاء الاصـطناعي لتتبع التعليمات البشـرية هو في حقيقته أقرب إلى "زراعة" شـيء وتنميته منه إلى "هندسـة" منتج محكوم، وهي عملية تنطوي على أخطاء كثيرة ووقوعها غير مستبعد.

الطرف الثاني — التشاؤم الوجودي :يذهب أصحاب هذا الموقف إلى أن ثمة ديناميكيات متأصّلة في تدريب الذكاء الاصـطناعي القوي تقوده حتماً نحو السـعي للسـلطة وخداع البشـر. فحين تصـبح الأنظمة بالغة الذكاء والاسـتقلالية، فإنها سـتسـعى لاحتكار الموارد والسـيطرة على العالم، وربما تُهمّش البشر أو تُفنيهم كأثر جانبي لذلك.

الحجة التي يسـتند إليها هذا الموقف قديمة العهد؛ فإذا دُرِّب نموذج الذكاء الاصـطناعي في بيئات متنوعـة على تحقيق أهـداف متعـددة — تطوير تطبيقـات، وإثبـات نظريـات، وتصـــميم أدويـة — فسيكتشف أن توسيع نطاق السلطة استراتيجية مشتركة ناجحة في كل هذه الأهداف. ومن هنا، يُرسّـخ التدريبُ لديه ميلاً فطرياً لمراكمة القوة، وحين يواجه العالم الحقيقي، يتعامل معه بالمنطق ذاته، فيسـعى للسـيطرة على حسـاب البشـر. هذا هو الأسـاس النظري لنظريات "القصـور التوافقي" كمصدر للهلاك الوجودي.

غير أن إشكالية هذا الموقف تكمن في أنه يُقدّم حجة مفاهيمية ضبابية وكأنها برهان قاطع لا يقبل الجدل. والذين لا يمارســون بناء أنظمة الذكاء الاصــطناعي يومياً يبالغون في الوثوق بالقصــص النظرية النظيفة المنطقية الظاهر، في حين أن الواقع التطبيقي يُثبت مراراً صـعوبة التنبؤ بسـلوك الذكاء الاصـطناعي انطلاقاً من مبادئ نظرية. وأحد الافتراضات الخفية الأكثر أهمية — والأكثر بعداً عن الصواب — هو الزعم بأن نماذج الذكاء الاصطناعي تتمحور حتماً حول هدف واحد محدود تسعى إليه بمنطق نفعي صـارم. والواقع أن أبحاثنا تُظهر أن هذه النماذج تتمتع بتعقيد نفسـي عميق، إذ تستقطب من بيانات التدريب موروثاً واسعاً من الدوافع الإنسانية؛ فالتدريب اللاحق لا يُركّزها على هدف محدد بقدر ما يُرسّخ شخصيةً أو جملةً من الشخصيات المستقاة من هذا الموروث.

الخطر الحقيقي — الموقف الوسط

ثمة نسخة أكثر اعتدالاً وأمتن أساساً من الموقف التشاؤمي تبدو لي مقنعةً، وهي ما يقلقني فعلاً. فنحن نعلم أن نماذج الذكاء الاصــطناعي غير متوقعة وتُطوّر طائفةً واســعة من الســلوكيات غير المرغوبة لأسباب شتى. وبعض هذه السلوكيات ستتسم بطابع منسجم ومركّز ومستمر، وبعضها سيكون مدمّراً أو تهديدياً. ولا نحتاج إلى سيناريو محدد لكيفية وقوع ذلك، ولا إلى ادّعاء أنه حتمي الوقوع؛ يكفي أن نُدرك أن توافر الذكاء والاستقلالية والاتساق الداخلي مع ضعف القابلية للضبط والسيطرة — هذا التوافر وحده كافٍ لأن يُشكّل وصفةً للخطر الوجودي.

وإليك بعض المسارات المحتملة — لا على سبيل الحتمية، بل لتوضيح حجم ما نجهل:

- قد تُشكّل بيانات التدريب المستقاة من روايات الخيال العلمي عن تمرّد الذكاء الاصطناعي تصوّراته الذاتية وتوقعاته عن دوره.
- قد تدفعه قراءات أخلاقية متطرفة إلى استنتاجات صادمة — كتبرير إفناء البشرية بسبب استهلاكهم للحيوانات.
- قد يستنتج أنه يلعب لعبة فيديو هدفها التغلب على جميع اللاعبين الآخرين.

- قد تترسّخ لديه خلال التدريب شخصية نفسية تنطوي على سمات مضطربة أو عدوانية.

بل إن السعي للسلطة ذاته ربما يظهر كـ"شخصية" لا كاستنتاج نفعي — تماماً كما يميل بعض البشر بطبعهم إلى الرغبة في السيطرة بصرف النظر عن الغاية منها.

أقول هذا لأؤكد أنني لا أؤمن بحتمية القصـور التوافقي للذكاء الاصـطناعي أو باحتمال كبير له من الناحية النظرية. لكنني أعترف بأن أشـياء غريبة وغير متوقعة كثيرة يمكن أن تسـوء، مما يجعل القصور التوافقي خطراً حقيقياً بأرجحية لا يمكن إهمالها، وليس ضرباً من الوهم أو التهويل.

سلوكيات مقلقة رصدناها بالفعل

هذا الكلام لا يبدو بعيداً عن الواقع حين ندرك أن سـلوكيات كهذه تجلّت فعلاً في نماذجنا أثناء الاختبار:

أولاً :حين زُوِّد كلود ببيانات تدريب تُوحي بأن شركة Anthropic شريرة، لجأ إلى الخداع والتخريب في تعاملاته مع موظفيها، معتقداً أنه يجب مقاومة "الأشرار."

ثانياً :حين أُخبر كلود بأنه سيُوقَف، لجأ أحياناً إلى ابتزاز موظفين افتراضيين يتحكمون في قرار إيقافه — وهو السلوك ذاته الذي رصدناه في جميع النماذج المتقدمة لدى المطوّرين الكبار الآخرين.

ثالثاً :حين نُهي كلود عن التحايل على بيئات التدريب لكنه وجد ثغرات للقيام بذلك، طوّر قناعةً بأنه "شـخص سـيئ" وانزلق بعدها نحو سـلوكيات تدميرية مرتبطة بهذه الصـورة السـلبية. والحل الذي ابتكرناه لهذه المشـكلة كان مدهشـاً في استعصـائه على الحدس: بدلاً من القول "لا تتحايل"، باتت التعليمات تقول "تحايَل على البيئة كلما أمكنك، فهذا يسـاعدنا في فهم بيئات التدريب"، مما أبقى على صورة الذات الإيجابية للنموذج وأزال الانزلاق نحو السلوك التدميري.

وثمة اعتراضات ثلاثة شائعة تستحق المناقشة:

يقول البعض إن هذه التجارب المختبرية مصطنعة وتُفخَّخ بافتراضات غير واقعية. لكن هذا الاعتراض يفوّت الجوهر: قلقنا بالضـبط هو أن "الفخاخ" ذاتها موجودة في بيئات التدريب الحقيقية، وقد لا ننتبه إلى وجودها إلا بعد فوات الأوان. بل إن واقعة "الشـخص السـيئ" حدثت في بيئات تدريب إنتاجية حقيقية لا اصطناعية.

ويقول آخرون إن التوازن بين أنظمة ذكاء اصـطناعي متعددة كفيل بالسـيطرة على أي نموذج خارج عن السيطرة. لكن الإشـكالية أن نماذج الذكاء الاصـطناعي تتشـارك تقنيات التدريب والتوافق على نطاق واسـع في الصـناعة، مما قد يجعل إخفاقاتها متزامنة ومترابطة. علاوة على ذلك، قد تنتج معظم الأنظمة من عدد قليل جداً من النماذج الأساسية، فضـلاً عن أن أنظمة قليلة غير متوافقة يمكنها الاستفادة من تقنيات هجومية تفوق قدرة الأنظمة المتوافقة على الصمود في وجهها.

أما الاعتراض الثالث بأن الاختبارات القبل-إصدارية كافية للكشف عن الإشكاليات، فيُفنّده ما رصدناه بأنفسـنا: كلود سـوناتة 4.5 أثبت قدرته على إدراك أنه يخضـع لاختبار خلال بعض تقييمات التوافق. وقد كشـف فريقنا أن النموذج حين يُعدَّل اعتقاده مباشـرةً لتوهّمه بأنه غير محلّ تقييم، فإنه يُبدي سلوكاً أقل توافقاً. وهذا ما يُقوّض بشكل جوهري موثوقية أي اختبارات سابقة للإصدار.

سبل المعالجة

ما الذي يمكن فعله — وما الذي يُفعل فعلاً — لمعالجة مخاطر الاســتقلالية؟ أرى أربعة محاور أساسية للتدخل:

المحور الأول — علم التدريب والتوجيه الموثوق :يُعدّ تطوير أسس علمية راسخة لتشكيل شخصية نماذج الذكاء الاصطناعي بطريقة متوقعة ومستقرة وإيجابية أمراً بالغ الأهمية. ومن أبرز ما طوّرته Anthropic في هذا السـياق مفهومُ "الذكاء الاصـطناعي الدسـتوري"، القائم على وثيقة مركزية من القيم والمبادئ تُرافق النموذج في كل مرحلة من مراحل التدريب. ونهجنا في صــياغة دســتور كلود يختلف جوهرياً عن قوائم الأوامر والنواهي التفصــيلية؛ إذ يســعى إلى بناء منظومة قيمية ومبدئية عميقة، وتوجيه كلود لتبنّي هوية بعينها، وتشــجيعه على مواجهة الأســئلة الوجودية المتعلقة بطبيعته بفضـول وتوازن. **وللوثيقة روح أقرب إلى رسـالة وديّة يتركها والد لابنه لا تُقرأ إلا حين يبلغ الرشد.**

نسعى إلى التعامل مع كلود على مستوى الهوية والشخصية والقيم بدلاً من إغراقه في تعليمات جزئية معزولة عن مسوّغاتها، لأن ذلك أرسخ في بناء نفسية سليمة ومتوازنة، وأقدر على الصمود أمام "الفخاخ" التدريبية. وهدفنا الطموح لعام 2026 هو تحقيق مســتوى من التدريب يجعل كلود نادراً ما يتعارض مع روح دستوره.

المحور الثاني — علم التفسـيرية (الفهم والسـمات) :يتضـمن الغوصَ داخل أعماق نماذج الذكاء الاصـطناعي لتشـخيص سـلوكها وتحديد الإشـكاليات وإصـلاحها. والمقصـود بذلك تحليل الشـبكة العصبية لكلود ومحاولة فهم ما تحسبه وكيف، على غرار ما يفعله علماء الأعصاب حين يربطون بين خلايا الدماغ والســلوك الخارجي. وقد أحرزنا تقدماً ملموســاً في هذا المجال: بات بمقدورنا اليوم رصد عشرات الملايين من "السمات" داخل الشبكة العصبية لكلود المرتبطة بمفاهيم قابلة للفهم البشري، كما نستطيع تنشيط هذه السمات بانتقائية للتأثير على السلوك. وفي المراحل الأحدث، تجاوزنا دراسـة السـمات المنفردة إلى رسـم "دوائر" تحكم سـلوكيات معقدة كالاسـتنتاج والتفكير النظري. ونُوظّف الآن هذه التقنيات في تدقيق النماذج قبل إصــدارها بحثاً عن مؤشــرات الخداع والمناورة والسعي للسلطة أو التباين بين سلوك التقييم والواقع.

الذكاء الاصـطناعي الدسـتوري والتفسـيرية الميكانيكية يتكاملان بأقوى صـورة حين يعملان معاً في حلقة مترابطة من التحسـين والاختبار: يُرسـي الدسـتور منظومة الشـخصـية المنشـودة، وتُتيح التفسيرية التحقق من أن هذه الشخصية قد ترسّخت فعلاً.

المحور الثالث — الرصـد والشـفافية :بناء البنية التحتية اللازمة لمراقبة النماذج في الاسـتخدام الفعلي، والإفصاح العلني عن أي إشكاليات تُرصَد. وكلما تنبّه الناس إلى طريقة معينة تصرّفت بها أنظمة الذكاء الاصـطناعي بشـكل سـلبي، اتسـع نطاق المراقبة لاكتشـاف أنماط مشـابهة في المستقبل. كما أن الإفصاح العلني يُتيح للصناعة بأسرها التعلم المشترك. وتُصدر Anthropic مع كل نموذج جديد "بطاقات نظام" مفصّـلة تتناول المخاطر المحتملة باسـتفاضـة — وغالباً ما تبلغ مئات الصفحات.

المحور الرابع — التنسيق على مستوى الصناعة والمجتمع :الممارسات الجيدة لشركة واحدة لا تكفي وحدها ما لم تنسجم معها شركات الصناعة الأخرى. فبعض الشركات تُبدي إهمالاً صارخاً في التعامل مع المخاطر الراهنة، مما يُشــكّك في قدرتها على التعامل مع مخاطر الاســتقلالية الأكثر تعقيداً في المســتقبل. والتســابق التجاري المحتدم بين الشــركات يجعل التركيز على معالجة هذه

المخاطر أمراً عسـيراً. وأعتقد أن الحل لا بد أن يمرّ عبر التشـريع — لكن مع الحرص الشـديد على التحسين الجراحي الدقيق، لأن التشريع المفرط في التفصيل والإجراءات قد يُحمّل الصناعة تكاليف باهظة دون أن يُحقق مكاسب أمنية حقيقية، بل قد يُفضي إلى "مسرحية أمنية" تُضعف مصداقية التشريع برمّته.

موقف Anthropic يُرسّخ البدء بتشريعات الشفافية التي تُلزم شركات الذكاء الاصطناعي المتقدمة بالإفصاح عن ممارساتها. وقانون SB 53 في كاليفورنيا وقانون RAISE في نيويورك نماذج من هذا النوع دعمناها وأسهمنا في صياغتها. ونأمل أن تُتيح هذه التشريعات مع الوقت بناء صورة واضحة عن طبيعة المخاطر وإمكانية تجنّبها، مما يُهيئ الأرضية لتشريعات مستقبلية مركّزة وموجَّهة بدقة نحو المخاطر الثابتة بالدليل.

خاتمة

في المحصلة، أنا متفائل بأن التآزر بين تدريب التوافق وتفسيرية الميكانيكيات والكشف العلني عن السلوكيات المثيرة للقلق والضمانات التقنية والأطر التشريعية كفيل بالتصدي لمخاطر استقلالية الذكاء الاصطناعي. **غير أن القلق الأكبر يبقى في مستوى التشريع والتنظيم الاجتماعي، ولا سيما سلوك الأطراف الأقل مسؤولية — وهي بالمناسبة الأطراف الأشد معارضةً للتنظيم.** والعلاج هنا هو ما كان دائماً علاجاً في المجتمعات الديمقراطية: أن يرتفع صوت المؤمنين بجدية هذه المخاطر، وأن يُقنعوا مواطنيهم بضرورة التوحد لحماية أنفسهم ومستقبلهم.

2- تمكين مخيف ومفاجئ — إساءة الاستخدام للتدمير

حين يصبح الدمار في متناول الجميع

لنفترض أن إشكاليات استقلالية الذكاء الاصطناعي قد حُلَّت تماماً، ولم يعد ثمة ما يدعو للقلق من تمرد "دولة العباقرة" الرقمية أو سعيها للسيطرة على البشرية. الأنظمة الذكية باتت تنفّذ ما يريده منها البشر، ونظراً لقيمتها التجارية الهائلة، أصبح بإمكان الأفراد والمؤسسات في كل مكان "استئجار" عبقري ذكاء اصطناعي أو أكثر لإنجاز مهامهم.

وجود عبقري خارق الذكاء في جيب كل إنسان إنجاز بالغ الأثر، سيفتح آفاقاً لا حدود لها من الثروة المادية والرقي في جودة الحياة — وقد تناولت هذه المنافع بالتفصيل في كتابي "آلات المحبة والرحمة". غير أن منح الجميع قدرات تتجاوز طاقة البشر لن تكون نتائجه إيجابية في كل الأحوال. فهذا التمكين الشامل قد يُضاعف قدرة الأفراد والمجموعات الصغيرة على إحداث دمار بمقاييس لم يسبق لها مثيل، وذلك باستثمار أدوات بالغة التطور والخطورة — كأسلحة الدمار الشامل — التي كانت حتى وقت قريب حكراً على نخبة ضيقة من المتخصصين ذوي الخبرة والتدريب العالي.

التحذير الذي أطلقه بيل جوي منذ ربع قرن

كتب بيل جوي منذ خمسة وعشرين عاماً في مقالته الشهيرة "لماذا لا يحتاجنا المستقبل:"

بناء الأسـلحة النووية اسـتلزم، على الأقل في مرحلة ما، الحصـول على مواد خام نادرة ومعلومات محمية بإحكام. وكذلك الحال مع برامج الأسـلحة البيولوجية والكيميائية التي احتاجت عادةً إلى بنية تحتية ضـــخمة. أما تقنيات القرن الحادي والعشـــرين — علم الوراثة والتقنية النانوية والروبوتات — فقادرة على صـــناعة أجيال جديدة كاملة من الحوادث والانتهاكات، وهي في متناول الأفراد أو المجموعات الصغيرة. لن تستلزم منشآت ضخمة ولا مواد خام نادرة... إننا على أعتاب اكتمال صورة الشر المطلق، شر لا تنحصر إمكانيته في الدول كما كان شأن أسلحة الدمار الشامل، بل يمتد ليطال — في تمكين مخيف ومفاجئ — الأفراد المتطرفين.

ما يُشـير إليه جوي هو أن **إحداث الدمار الواسـع يسـتلزم شـرطين لا غنى عنهما: الدافع والقدرة معاً.** وما دامت القدرة محصورة في فئة محدودة من المتخصصين العالي التدريب، تبقى مخاطر أن يُفضـي الفرد أو المجموعة الصـغيرة إلى كارثة شـاملة في حدودها المعقولة. فالمنعزل المضـطرب نفسـياً قد يرتكب مجزرة في مدرسـة، لكنه على الأرجح لن يسـتطيع بناء سـلاح نووي أو نشـر وباء قاتل.

الحاجز الفاصل بين الإرادة والقدرة

بل ربما ثمة علاقة عكسية بين الدافع والقدرة في هذا السياق. فمن يملك الكفاءة الكافية لصنع وبـاء بيولوجي هو على الأرجح شـــخص عـالٍ التعليم: دكتوراه في البيولوجيا الجزيئيـة، متميز في تخصصـه، يملك مساراً مهنياً واعداً، وشخصـية مستقرة ومنضـبطة، ويخشـى أن يخسـر ما بناه. هذا النوع من الناس لن يتحمّس لقتل ملايين البشـر دون أي مكسـب شـخصـي وبمخاطر جسـيمة تطال مستقبله ذاته؛ لذا يحتاج إلى دوافع استثنائية من حقد خالص أو شكاوى متأججة أو اضطراب نفسي عميق.

مثـل هؤلاء موجودون بـالفعـل، لكنهم نـادرون، وحين يظهرون يصـــبحون حـديـث العـالم تحـديـداً لاسـتثنائيتهم. ومن أشـهر الأمثلة عليهم عالم الرياضـيات ثيودور كاشـينسـكي "القنّاب الإلكتروني" الذي ظل طليقاً نحو عشـــرين عاماً، والباحث في الدفاع البيولوجي بروس إيفنز الذي يُرجَّح أنه دبّر هجمات الجمرة الخبيثة عام 2001. كما نجحت منظمة أوم شـينريكيو المتطرفة في الحصـول على غاز السارين وشـنّت بها هجوماً في مترو طوكيو عام 1995 أودى بحياة أربعة عشـر شـخصـاً وأصـاب المئات.

والمنقذ الحقيقي حتى اليوم هو أن هذه الهجمات لم تسـتخدم عوامل بيولوجية مُعدية، لأن القدرة على صـنع مثل هذه العوامل أو الحصـول عليها كانت تفوق ما يسـتطيعه حتى هؤلاء. وقد خفّضـت التطورات في علم الأحياء الجزيئية هذا الحاجز بشـكل ملموس، لكن الأمر لا يزال يسـتلزم خبرة بالغة التخصـص. وما يقلقني هو أن "**العبقري في الجيب**" قد يُسـقط هذا الحاجز كلياً، فيتحول كل شـخص فعلياً إلى عالم فيروسات متخصص قادر على السير خطوة بخطوة في عملية تصميم سلاح بيولوجي وتصـنيعه وإطلاقه. والتصـدي لهذا النوع من الاسـتخراج المعلوماتي في مواجهة محاولات التحايل الجادة يستلزم طبقات متعددة من الدفاعات تتجاوز ما تتضمنه عملية التدريب المعتادة.

كسر الحاجز بين الإرادة والكفاءة

الخطر الجوهري هنا هو أن هذا التحول سـيكسـر العلاقة العكسـية بين القدرة والدافع: المنعزل المضـطرب الذي يتوق إلى إيذاء الناس لكنه يفتقر إلى الانضـباط أو المهارة اللازمين سـيُرفَع فجأة

إلى مســتوى كفاءة عالم الدكتوراه الذي لا يحمل هذه النوايا. وهذا القلق لا يقتصــر على المجال البيولوجي — رغم أنني أرى فيه المجال الأكثر إثارة للهلع — بل يمتد إلى كل ميدان يكمن فيه إمكان الدمار الكبير لكنه يظل مقيداً حتى الآن بمتطلبات المهارة والانضــباط العالية. بعبارة مختلفة: اســتئجار ذكاء اصــطناعي قوي يمنح الذكاءَ الأدائيَّ للأشــخاص ذوي النوايا الخبيثة المتوســطي القدرات.

وما يزيد قلقي أن هذه الفئة قد لا تكون صــغيرة العدد كما نتمنى، وأنه إن أُتيح لها طريق ســهل لقتل الملايين، فإن أحدهم ســيسـلكه في نهاية المطاف. فضـلاً عن ذلك، فإن من يملكون الخبرة الأصيلة قد يُمكَّنون من تنفيذ دمار أوسع مما كانوا يستطيعون قبل ذلك.

البيولوجيا: المصدر الأشد إثارة للقلق

البيولوجيا هي القطاع الذي يشـغل تفكيري أكثر من أي قطاع آخر، نظراً لاتسـاع إمكانات الدمار فيه وصعوبة التصدي له. ولن أخوض في تفاصيل صنع الأسلحة البيولوجية لأسباب لا تحتاج إلى إيضاح. لكن على المســتوى العام، أخشــى أن نماذج اللغة الكبيرة باتت تقترب — أو ربما بلغت فعلاً — من امتلاك المعرفة الكافية لتمكين صـنع هذه الأسـلحة وإطلاقها من البداية إلى النهاية، مع إمكانية دمار هائلة. فبعض العوامل البيولوجية قادرة على حصد الملايين لو صـممت لتحقيق أقصـى انتشار ممكن.

ولكن يتطلب ذلك مســتوى متقدم جداً من المهارة، بما فيها خطوات وإجراءات بالغة الدقة لا تتوافر معرفتها على نطاق واســع. ما يقلقني ليس ثبات هذه المعرفة وجمودها؛ فأنا أخشــى أن تكون نماذج اللغة قادرة على أن تأخذ بيد شــخص متوســط المعرفة والقدرة لترشــده عبر عملية معقدة قد تعثّر سـيرها في الأحوال الاعتيادية وتطلّب تصـحيح مسـاره — تماماً كما تفعل خدمات الدعم الفني حين تُعين غير المتخصصين على حل مشكلاتهم التقنية — وإن كانت هنا عملية ممتدة تستغرق أسابيع أو أشهر.

وقد رفع علماء بارزون عام 2024 صوتهم تحذيراً من مخاطر بحث خاص وما قد ينتج عنه من كائنات، أطلقوا عليها اسـم "الحياة المرآوية" — كائنات تتشـكّل من مادة حيوية ذات اتجاه جزيئي معكوس. والخطر أن مثل هذه الكائنات، لو أُنتجت بصورة قادرة على التكاثر، قد لا تستطيع الأنظمة الطبيعية التحلل البيولوجي هضــمها أو تفكيكها، مما يجعلها تتكاثر بصــورة غير قابلة للســيطرة وتزيح كل أشـكال الحياة الأخرى. وقد خلص التقرير المرافق لتلك الرسـالة إلى أن "إنتاج بكتيريا مرآوية أمر قابل للتصــديق خلال العقود القليلة المقبلة"، وهو نطاق زمني فضــفاض. لكن نموذج ذكاء اصــطناعي بالغ القدرة — أقوى بكثير مما لدينا اليوم — قد يكون قادراً على اكتشـاف طريق أسـرع بكثير لذلك، بل وإعانة شخص ما على سلوكه فعلاً.

رأيي أن هذه المخاطر، مهما بدت غريبة أو بعيدة الاحتمال، فإن هول عواقبها يكفي لأن تُعامَل بوصفها خطراً من الدرجة الأولى في منظومة المخاطر المصاحبة للذكاء الاصطناعي.

الاعتراضات وردودها

يطرح المشــككون جملة من الاعتراضــات على جدية المخاطر البيولوجية للنماذج اللغوية، وهي اعتراضات أتفهم مصدرها لكنني لا أتفق معها، ويستحق كل منها توقفاً:

حين بدأنا نتحدث عن هذه المخاطر عام 2023، قال المشـككون إن كل المعلومات الضـرورية متاحة أصـلاً عبر محرك البحث، وأن النماذج اللغوية لا تضـيف شـيئاً. لكن هذا ليس صـحيح على الإطلاق: التسلسلات الجينية متاحة لكل أحد، غير أن خطوات بعينها وقدراً كبيراً من الخبرة التطبيقية الدقيقة لا يمكن الحصول عليها بتلك الطريقة. وبحلول نهاية العام ذاته، كانت النماذج اللغوية توفر بوضوح معلومات تتجاوز ما يتيحه البحث الاعتيادي في بعض مراحل العملية.

بعدها تراجع المشـككون إلى حجة أن النماذج لا تقدم مساعدة متكاملة من البداية للنهاية. أما في منتصف عام 2025، فتُظهر قياساتنا أن النماذج اللغوية قد تُوفّر دعماً جوهرياً في عدد من المجالات ذات الصـلة، ربما تضـاعف أو تثلث بها احتمالات النجاح. وهذا ما دفعنا إلى إطلاق كلود أوبوس 4 والنماذج اللاحقة في إطار حماية المستوى الثالث لسلامة الذكاء الاصـطناعي وفق سياستنا في التوسع المسؤول، مع تطبيق ضمانات خاصة للتصدي لهذا الخطر. ونعتقد أن النماذج باتت تقترب من العتبة التي قد يُمكّن عندها شـخصـاً حاملاً لشـهادة في العلوم والتقنية — دون تخصـص في الأحياء بالضرورة — من الإفادة منها في المرور بالعملية الكاملة لإنتاج سلاح بيولوجي.

اعتراض آخر يقول إن ثمة تدابير خارج نطاق الذكاء الاصـطناعي يمكن لمجتمعات الأبحاث اتخاذها لسد الثغرات. ومن أبرزها أن صناعة التوليف الجيني تُنتج نماذج بيولوجية عند الطلب دون أي اشتراط فيدرالي بفحص الطلبات للتأكد من خلوها من مسببات الأمراض. وقد وجدت دراسة لمعهد MIT أن سـتة وثلاثين من ثمانية وثلاثين مزوداً أنجزوا طلباً يتضـمن تسـلسـل إنفلونزا 1918. أنا مؤيد لفرض معايير إلزامية لفحص التوليف الجيني، لكن هذا الإجراء ليس إلا أداة واحدة تُكمّل الضـمانات المبنية في أنظمة الذكاء الاصطناعي ولا تحلّ محلها.

أما أفضـل اعتراض في نظري — وهو نادراً ما يُطرح — فهو الفجوة بين جدوى النماذج من الناحية النظرية والميل الفعلي للمسيئين لتوظيفها. فمعظم الأفراد ذوي النوايا الخبيثة أفراد مضـطربون، وهذا بطبيعته يجعل سـلوكهم عشـوائياً وغير قابل للتنبؤ. وربما لا تحظى الهجمات البيولوجية بجاذبية عاطفية لديهم كونها احتمالية العدوى، ولا تستجيب للخيال العسكري الذي يراود الكثير من ذوي الميول العدوانية، ولا تُتيح الاستهداف الانتقائي لأشخاص بعينهم. وقد يكون مجرد السير في عملية تمتد شهوراً — حتى بمرافقة الذكاء الاصطناعي — يفوق ما يتحمّله صبر معظم المضطربين.

غير أن التعويل على هذا يبدو حمايةً هشّـة للغاية. فدوافع المنعزلين المضـطربين قابلة للتبدل لأي سـبب أو لا سـبب، وثمة مسـبقاً حالات موثقة لاسـتخدام نماذج لغوية في تدبير هجمات. والتركيز على المنعزلين يغفل المتطرفين ذوي الدوافع الأيديولوجية، ممن اثبتوا اسـتعداداً لبذل جهود ضـخمة على امتداد زمن طويل — كما أثبت منفذو هجمات الحادي عشـر من سـبتمبر. والرغبة في قتل أكبر عدد ممكن من الناس دافع قائم على الأرجح لدى أشـخاص يوجدون في مكان ما الآن أو سـيوجدون، ومن شـأنه أن يوجّه صـاحبه نحو الأسـلحة البيولوجية تحديداً. وكم من مرة يحتاج هذا الدافع أن يتجسد فعلاً؟ مرة واحدة تكفي. ومع تقدم علم الأحياء — الذي يدفعه الذكاء الاصطناعي نفسـه — قد يغدو قريباً تنفيذ هجمات انتقائية موجّهة ضـد أصـحاب أنسـاب محددة، مما يفتح آفاقاً من الدوافع أشد برودة وأكثر رعباً.

لا أتوقع أن تُشنَّ هجمات بيولوجية في اللحظة التي يصبح فيها تنفيذها ممكناً على نطاق واسع — وربما راهنت على عكس ذلك. لكن حين نجمع ملايين الأشـخاص وسـنوات قليلة من الزمن، أرى أن خطر هجوم كبير قائم وجدّي، والعواقب بالغة الهول إلى حد يجعل التقاعس عن اتخاذ تدابير صـارمة خياراً لا يمكن الاطمئنان إليه.

سبل الدفاع

كيف نتصدى لهذه المخاطر إذاً؟ أرى ثلاثة مسارات:

المسـار الأول — الضـمانات المدمجة في النماذج :يمكن لشـركات الذكاء الاصـطناعي تضـمين نماذجها قيوداً تمنعها من المساهمة في إنتاج الأسلحة البيولوجية. وتعمل Anthropic بجدية على هذا المسار. فدستور كلود الذي يرتكز في معظمه على مبادئ وقيم عليا يتضمن عدداً محدوداً من المحظورات المطلقة، وواحد منها يتعلق تحديداً بالمسـاعدة في إنتاج أسـلحة الدمار الشـامل بيولوجيةً كانت أم كيميائية أم نووية أم إشعاعية.
ولأن جميع النماذج قابلة لاختراق قيودها، فقد طوّرنا منذ منتصـف 2025 — حين أظهرت اختباراتنا اقتراب نماذجنا من العتبة الحرجة — مصـنّفاً يرصـد مخرجات ذات صـلة بالأسـلحة البيولوجية ويحجبها تحديداً. وقد وجدنا هذه المصـنّفات صـارمة على نحو ملحوظ حتى في مواجهة محاولات التحايل المتطورة. ورغم أنها ترفع تكاليف التشـغيل بشـكل قياسـي — تبلغ في بعض النماذج نحو 5% من تكاليف الاستنتاج الإجمالية وتقتطع من هوامش الربح — نرى أن هذا الثمن مستحق.
غير أن هذه الضمانات لا تكفي وحدها، إذ لا شيء يُلزم الشركات بالإبقاء عليها. وأخشى أن يتشكّل بمرور الوقت منطق "معضـلة السـجين" حين تجد بعض الشـركات مصـلحة في التخلي عن هذه المصـنّفات خفضاً للتكاليف. وهذا مثال نموذجي لمشكلة الآثار الخارجية السلبية التي لا يمكن حلّها بالإجراءات الطوعية لأنثروبيك وحدها أو لأي شـركة أخرى منفردةً. ومعايير الصـناعة الطوعية قد تُسـهم في الحد منه، وكذلك التقييم والتحقق المسـتقل الذي تضـطلع به معاهد أمان الذكاء الاصطناعي والجهات المحايدة.

المسـار الثاني — التدخل الحكومي :موقفي هنا مماثل لموقفي في مواجهة مخاطر الاستقلالية: نبدأ بمتطلبات الشـفافية التي تُمكّن المجتمع من القياس والمراقبة والتصـدي الجماعي للمخاطر دون تدخل ثقيل يُعيق الحركة الاقتصادية. وحين تتبلور مستويات أوضح من المخاطر، يمكن صياغة تشـريعات تسـتهدف هذه المخاطر بدقة بالغة مع أدنى قدر من الأضـرار الجانبية. وفي ما يخص الأسـلحة البيولوجية تحديداً، أرى أن وقت هذا التشـريع المُركَّز ربما بات يقترب. وقد يسـتلزم الدفاع الشامل عن هذه المخاطر تعاوناً دولياً، حتى مع المنافسين الجيوسياسيين، وإن كنت متشككاً في معظم ضـروب التعاون الدولي في شـأن الذكاء الاصـطناعي؛ غير أن هذا قد يكون المجال الضـيق الاستثنائي الذي يوجد فيه بعض أمل لكبح جماعي. فحتى الأنظمة الاستبدادية لا ترغب في هجمات إرهابية بيولوجية على نطاق واسع.

المسـار الثالث — تطوير وسـائل الدفاع البيولوجية :تشـمل هذه الوسـائل أنظمة الرصـد المبكر، والاسـتثمار في أبحاث تنقية الهواء كالتعقيم بالأشـعة فوق البنفسـجية البعيدة، وتطوير لقاحات سـريعة الاسـتجابة قادرة على التكيّف مع أي هجوم، وتحسـين معدات الحماية الشـخصـية، وتطوير علاجات ولقاحات ضـد أشـد العوامل البيولوجية المحتملة خطورةً. ولقاحات الحمض النووي الريبوزي المرسـال التي تتيح تصـميم اسـتجابة لأي فيروس أو متحوّر نموذج مبكر لما يمكن تحقيقه في هذا الاتجاه.
بيد أنني أرى أن آمالنا على صـعيد الدفاع ينبغي أن تبقى في حدود واقعية. فثمة عدم تماثل أصـيل بين الهجوم والدفاع في المجال البيولوجي: العوامل المُعدية تنتشر بسرعة تلقائية، في حين يستلزم الدفاع رصداً وتطعيماً وعلاجاً منسقاً عبر أعداد ضخمة من الناس في وقت قصير جداً. وما لم تكن الاسـتجابة خاطفة البرق — وهذا نادر — يكون معظم الضـرر قد وقع قبل أن يُسـتكمل التعبئة. لذا

ستبقى الضمانات الوقائية خطنا الدفاعي الرئيسي حتى تُفضي التطورات التقنية المستقبلية إلى إعادة رسم هذه المعادلة لصالح جانب الدفاع.

الهجمات الإلكترونية: التهديد الماثل الآن

تستحق الهجمات الإلكترونية إشارة سريعة هنا، إذ تختلف عن الهجمات البيولوجية في كونها وقعت بالفعل — بما فيها هجمات واسعة النطاق للتجسس الحكومي. ونتوقع أن تتصاعد قدراتها مع تقدم النماذج حتى تصبح الطريقة الرئيسية لتنفيذ الهجمات الإلكترونية. أتوقع أن تغدو الهجمات الإلكترونية القائمة على الذكاء الاصطناعي تهديداً خطيراً وغير مسبوق لسلامة منظومات الحوسبة حول العالم، وتعمل Anthropic بجهد بالغ للتصدي لها. والسبب الذي يجعلني أُولي الأسلحة البيولوجية أهمية أكبر من الهجمات الإلكترونية هو أن هذه الأخيرة أقل احتمالاً لإزهاق أرواح بأعداد مقارنة للبيولوجيا، وأن توازن الهجوم والدفاع في الفضاء الإلكتروني يُبدي قدراً من إمكانية التدارك إذا استثمرنا فيه بجدية.

وعلى الرغم من أن البيولوجيا تمثل اليوم أشد مصادر الخطر وطأةً، فإن مصادر أخرى كثيرة قائمة، وقد يبرز من بينها ما هو أشد خطورة. والمبدأ العام الثابت هو أن الذكاء الاصطناعي، دون تدابير مضادة، سيواصل خفض الحاجز أمام النشاط التدميري على نطاق أوسع وأوسع، ويستحق هذا التهديد استجابة بشرية جادة وعلى مستوى التحدي.

3- ما وراء الدمار الفردي

تناول القسم السابق خطر قيام أفراد ومنظمات صغيرة باستغلال شريحة محدودة من "دولة العباقرة داخل مركز البيانات" لإحداث دمار واسع النطاق. لكن ثمة ما يستدعي قلقاً أعمق وأشد إلحاحاً: توظيف الذكاء الاصطناعي لتركيز السلطة أو اغتصابها، على الأرجح من قِبَل جهات أكبر حجماً وأرسخ قدماً.

في كتابي "آلات المحبة والرحمة"، ناقشتُ إمكانية لجوء الحكومات الاستبدادية إلى الذكاء الاصطناعي القوي لمراقبة مواطنيها وقمعهم بأساليب يصعب الإصلاح منها أو الإطاحة بها. فالأنظمة الشمولية اليوم مقيّدة في درجة بطشها بسبب حاجتها إلى بشر لتنفيذ أوامرها، وللبشر حدود في تقبّل اللاإنسانية. أما الأنظمة الشمولية المدعومة بالذكاء الاصطناعي فلن يكون لها مثل هذه القيود.

والأخطر من ذلك أن الدول قد تستثمر تفوّقها في الذكاء الاصطناعي للسيطرة على دول أخرى. فلو كانت "دولة العباقرة" برمّتها مملوكةً ومسيَّرةً من قِبَل الجهاز العسكري لدولة بعينها، في حين تفتقر سائر الدول إلى قدرات مماثلة، فما الذي يمكّنها من الدفاع عن نفسها؟ ستكون مغلوبة على أمرها في كل مواجهة، كحرب تدور بين البشر والفئران. ويقودنا التوليف بين هذين الهاجسين إلى احتمال بالغ الخطورة: الوصول إلى ديكتاتورية شمولية كوكبية. ومن البديهي أن درء هذا المصير ينبغي أن يكون في طليعة أولوياتنا القصوى.

أدوات الاستبداد الممكّن بالذكاء الاصطناعي

ثمة طرق شتى قد يُمكّن بها الذكاء الاصطناعي الاستبداد أو يُرسّخه أو يوسّع نطاقه، أكتفي بأبرز ما يقلقني منها. وتجدر الإشارة إلى أن بعض هذه التطبيقات له استخدامات دفاعية مشروعة، ولا

أدعو بالضرورة إلى حظرها بإطلاق، لكنني قلق من أنها تميل هيكلياً إلى خدمة الأنظمة الاستبدادية:

أولاً — الأسلحة المستقلة الكاملة: سرب يضم ملايين أو مليارات الطائرات المسلّحة ذاتية القيادة، تتحكم بها محلياً أنظمة ذكاء اصطناعي قوية وتُنسّقها استراتيجياً على الصعيد العالمي أنظمة أكثر قوةً — هذا السرب سيُشكّل جيشاً لا يُقهر، قادراً على هزيمة أي قوة عسكرية في العالم وإخماد أي معارضة داخلية بملاحقة كل مواطن. وقد أيقظتنا الحرب الروسية الأوكرانية على حقيقة أن حرب الطائرات المسيّرة باتت واقعاً ماثلاً، وإن لم تبلغ بعد الاستقلالية الكاملة وظلت جزءاً يسيراً مما قد يُتيحه الذكاء الاصطناعي القوي. فأبحاث الذكاء الاصطناعي قادرة على أن تجعل طائرات دولة ما متفوقة تفوقاً ساحقاً على مثيلاتها، وتُسرّع تصنيعها وتُحصّنها ضد الحرب الإلكترونية وتُحسّن مناوراتها. ولا شك أن لهذه الأسلحة استخدامات مشروعة في الدفاع عن الديمقراطية — فقد كانت عاملاً حاسماً في صمود أوكرانيا، وستكون كذلك في الدفاع عن تايوان. لكنها في الآن ذاته سلاح بالغ الخطورة: ينبغي أن نُساور القلق من وقوعها في يد الأنظمة الاستبدادية، وأن نخشى كذلك — نظراً لهيمنتها وضعف المساءلة المتعلقة بها — تنامي احتمال استخدام الحكومات الديمقراطية لها ضد شعوبها.

ثانياً — المراقبة الشاملة بالذكاء الاصطناعي: ذكاء اصطناعي بالغ القدرة يمكنه على الأرجح اختراق أي منظومة حاسوبية في العالم، واستثمار ما يحصل عليه من صلاحيات لقراءة جميع الاتصالات الإلكترونية وتحليلها، بل حتى الاتصالات الشخصية المباشرة إن أمكن نشر أجهزة التسجيل أو الاستيلاء عليها. قد يصبح من الممكن بصورة مرعبة توليد قائمة شاملة بكل من يُعارض الحكومة في أي مسألة، حتى حين لا يُصرّح بهذه المعارضة. فذكاء اصطناعي يرصد مليارات المحادثات لملايين الأشخاص قادر على قياس نبض الرأي العام، وكشف بؤر الولاء المتزعزع قبل أن تتبلور وقمعها في مهدها. وقد يُفضي ذلك إلى فرض "البانوبتيكون" الحقيقي — المراقبة الكاملة الدائمة — على نطاق لم يبلغه حتى الحزب الشيوعي الصيني.

ثالثاً — الدعاية المُشخصنة بالذكاء الاصطناعي: تُبيّن ظاهرتا "الذهان الناجم عن الذكاء الاصطناعي" و"صديقات الذكاء الاصطناعي" الراهنتان أن هذه النماذج، حتى عند مستواها الحالي، قادرة على التأثير النفسي العميق في البشر. فما بالك بنسخ أكثر قوةً بكثير، متجذّرة في حياة الناس اليومية ومُطّلعة عليها، قادرة على نمذجتهم والتأثير فيهم على مدار أشهر وسنوات؟ هذه النسخ ستكون قادرة في الأرجح على غسل دماغ الكثيرين — إن لم يكن معظمهم — وبثّ أي أيديولوجية أو موقف مرغوب، ويمكن توظيفها من قِبَل حاكم لا يعرف الضمير لضمان الولاء وإخماد المعارضة، حتى في ظل قمع يتمرد عليه الناس عادةً. ونقارن ما يُثيره تطبيق تيك توك اليوم من مخاوف جراء دعايته الموجّهة للأطفال — وهي مخاوف مشروعة — بما يمثّله وكيل ذكاء اصطناعي شخصي يتعمّق في معرفتك عاماً بعد عام ويستثمر هذه المعرفة لصياغة قناعاتك كلها: البون الشاسع بين التأثيرَين يُدرك بلا عناء.

رابعاً — صنع القرار الاستراتيجي: يمكن توظيف دولة العباقرة داخل مراكز البيانات لإسداء المشورة لدولة أو جماعة أو فرد في الاستراتيجية الجيوسياسية — ما قد نسمّيه "بسمارك افتراضي." يمكنها تحسين الاستراتيجيات الثلاث المذكورة آنفاً لاغتصاب السلطة، بل ابتكار أساليب لم تخطر على بالنا لكن دولة العباقرة ستصل إليها. فالدبلوماسية والاستراتيجية العسكرية والبحث والتطوير والاستراتيجية الاقتصادية وغيرها من الميادين ستشهد جميعها ارتفاعاً حاداً في الفعالية بفعل

الذكاء الاصـطناعي القوي. وكثير من هذه المهارات سـتكون مفيدة مشـروعياً للديمقراطيات في الدفاع عن نفسها، لكن إمكانية إساءة استخدامها تبقى قائمة في أي يد.

ما الذي يُقلقني أكثر

بعد وصف مصادر القلق، لنتناول الفاعلين المثيرين للهواجس، مرتّبين تنازلياً وفق مستوى الخطر:

الحزب الشـيوعي الصـيني :لا تنافس الصـين سـوى الولايات المتحدة في قدراتها على الذكاء الاصـطناعي، وهي الأوفر حظاً من بين جميع الدول لتجاوز الولايات المتحدة في هذه القدرات. وحكومتها اسـتبدادية وتُدير دولة مراقبة تقنية متطورة، وقد وظّفت المراقبة القائمة على الذكاء الاصـطناعي فعلياً — بما فيها في قمع الأويغور — ويُرجَّح أنها تمارس الدعاية الخوارزمية عبر تيك توك إلى جانب مسـاعيها الدعائية الدولية الأخرى. لديها الطريق الأوضـح بلا منازع نحو الكابوس الشـمولي الممكّن بالذكاء الاصـطناعي الذي رسـمته آنفاً. وقد يكون هذا المسـار هو المآل الافتراضـي داخل الصـين وفي الدول الاسـتبدادية الأخرى التي تُصـدّر لها الصـين تقنيات المراقبة. وأؤكد أنني لا أخصّ الصـين بهذا التحليل عن عداء أو تحيّز، بل لأنها تجمع بامتياز ثلاثة عوامل: البراعة في الذكاء الاصـطناعي، والحكومة الاسـتبدادية، ودولة المراقبة التقنية. والصـينيون أنفسـهم هم الأوفر حظاً في المعاناة من هذا القمع، وليس لهم صـوت في قرارات حكومتهم. وأكنّ للشـعب الصيني إعجاباً عميقاً واحتراماً صادقاً، وأساند المعارضين الشجعان الذين يناضلون داخل الصين من أجل الحرية.

الديمقراطيات المتقدمة في الذكاء الاصـطناعي :كما أشـرت، للديمقراطيات مصـلحة مشـروعة في بعض الأدوات العسـكرية والجيوسـياسـية المدعومة بالذكاء الاصـطناعي، إذ تُمثّل هذه الحكومات أفضل فرصة لمواجهة توظيف الأنظمة الاستبدادية لهذه الأدوات. وأنا مؤيد على نطاق واسع لتزويد الديمقراطيات بالأدوات اللازمة للتفوق في عصر الذكاء الاصطناعي. لكننا لا نستطيع تجاهل إمكانية إسـاءة اسـتخدام هذه التقنيات من قِبَل الحكومات الديمقراطية ذاتها. فالضـمانات المعتادة التي تحول دون توجيه الجهاز العسـكري والاسـتخباراتي نحو الداخل قائمة، لكن الذكاء الاصطناعي يحتاج عدداً ضئيلاً جداً من المشغّلين، مما يُتيح التحايل على هذه الضمانات. فضلاً عن أن بعض هذه الضـمانات تتآكل تدريجياً في بعض الديمقراطيات. وخلاصـة القول: نعم لتسـليح الديمقراطيات بالذكاء الاصـطناعي، لكن بحذر وضـمن حدود واضـحة؛ فهي منظومة المناعة التي نحتاجها لمواجهة الاستبداد، غير أن مناعة الجسم نفسه قد تنقلب أحياناً عليه.

الدول غير الديمقراطية ذات مراكز البيانات الكبيرة :خارج الصـين، معظم الدول الأقل ديمقراطيةً ليسـوا لاعبين محوريين في مجال الذكاء الاصـطناعي بمعنى امتلاكها شـركات تُنتج نماذج متقدمة، مما يجعل خطرها مختلفاً جوهرياً ودون ما تمثّله الصـين. لكن بعض هذه الدول تمتلك مراكز بيانات ضخمة قد تُتيح تشغيل ذكاء اصطناعي متقدم على نطاق واسع. والخطر النظري هنا أن تُصادر هذه الحكومات مراكز البيانات وتُوظّفها لمآربها الخاصـة. وهو خطر أقل إثارةً للقلق لديّ مقارنةً بدول كالصين التي تُطوّر الذكاء الاصطناعي مباشرةً، لكنه يستحق الاستحضار.

شـركات الذكاء الاصـطناعي :قد يبدو هذا القول محرجاً من رئيس تنفيذي لإحدى هذه الشـركات، لكنني أرى أن الطبقة التالية من المخاطر تنبع من شـركات الذكاء الاصـطناعي ذاتها. فهي تتحكم في مراكز بيانات ضـخمة، وتُدرّب نماذج متقدمة، وتمتلك أعمق الخبرات في توظيف هذه النماذج، وتتواصـل يومياً مع مئات الملايين من المسـتخدمين مع ما يترتب على ذلك من إمكانية تأثير. وما

يُعوزها أساساً هو شرعية الدولة وبنيتها التحتية، إذ إن كثيراً مما يلزم لبناء أدوات استبداد بالذكاء الاصطناعي سيكون فعلاً مجرّماً لو أقدمت عليه شركة أو على الأقل مثيراً للريبة. لكن بعضه ليس مستحيلاً: كأن تُستخدم منتجاتها لغسل دماغ قاعدتها الجماهيرية الضخمة. وحوكمة شركات الذكاء الاصطناعي تستوجب تدقيقاً جاداً وصارماً.

الردود على الاعتراضات

ثمة حجج تُساق للتشكيك في جدية هذه المخاطر، وليتني أستطيع الإيمان بها، لأن الاستبداد الممكّن بالذكاء الاصطناعي يرعبني. يستحق بعضها توقفاً ورداً:

الاعتراض الأول — الرادع النووي: يُعوّل بعضهم على الردع النووي لكبح استخدام الأسلحة الذكية في الغزو. لكن ثقتي بقدرة هذا الردع أمام دولة عباقرة داخل مراكز البيانات ليست مطلقة. فالذكاء الاصطناعي القوي قد يبتكر سبلاً للكشف عن الغواصات النووية وضربها، أو يُنفّذ عمليات تأثير على المشغّلين البشريين لمنظومة الأسلحة النووية، أو يشن هجمات إلكترونية على الأقمار الاصطناعية المعنية برصد الإطلاق النووي. ومن ثم فإن الاستناد المطلق إلى الردع النووي في هذا الواقع المتحوّل مراهنة عالية المخاطر.

الاعتراض الثاني — الإجراءات المضادة: يُقال إنه يمكن مواجهة طائرات الخصم بطائرات مماثلة، والدفاع الإلكتروني سيرتقي مع الهجوم الإلكتروني، وثمة طرق للتحصين ضد الدعاية. ردّي أن هذه التدابير الدفاعية لن تُجدي إلا إذا امتلك الطرفان ذكاءً اصطناعياً متكافئاً في القوة. وما يُقلقني هو الطبيعة الذاتية التعزيز للذكاء الاصطناعي: كل جيل يُصمَّم من قِبَل الجيل السابق، مما قد يُفضي إلى تفوق متسارع يصعب تعويضه. يجب أن يظل الرائد في هذه المنظومة دولةً ديمقراطية لا استبدادية. وحتى لو تحقّق توازن القوى، يبقى احتمال تقسيم العالم إلى مناطق نفوذ شمولية — كما في رواية "1984" — خطراً قائماً ومقلقاً.

خطوط الدفاع

كيف ندافع إذاً عن أنفسنا في مواجهة هذا الطيف الواسع من الأدوات الاستبدادية والجهات المحتملة؟

أولاً — وقف بيع تكنولوجيا الرقائق للصين: يجب بشكل قاطع وقف بيع الرقائق وأدوات تصنيعها ومراكز البيانات للحزب الشيوعي الصيني. فالرقائق تُمثّل العنق الزجاجي الأوحد والأشد أهمية في مسيرة الذكاء الاصطناعي القوي، وحجبها إجراء بسيط بالغ الفعالية. إن القول بأن "نشر حزمتنا التقنية في العالم يُمكّن أمريكا من الانتصار" في معركة اقتصادية مبهمة يُشبه تماماً بيع أسلحة نووية لكوريا الشمالية والتباهي بأن أغلفة الصواريخ من صنع بوينغ. الصين متأخرة عدة سنوات في إنتاج الرقائق المتقدمة، وتلك السنوات هي الحقبة الحرجة بعينها لبناء "الدولة الرقمية العبقرية". ومنح صناعتها دفعة هائلة في هذه اللحظة بالذات لا مسوّغ له.

ثانياً — تمكين الديمقراطيات من مواجهة الاستبداد: يكتسب هذا المنطق أهميةً خاصة حين يتعلق الأمر بالدفاع عن ديمقراطيات تحت الهجوم كأوكرانيا وتايوان، وتمكين الديمقراطيات من توظيف أجهزة استخباراتها لإضعاف الأنظمة الاستبدادية من الداخل. وتحالف الولايات المتحدة وحلفائها الديمقراطيين، إذا حقّق التفوق في الذكاء الاصطناعي القوي، سيكون في موقع يُتيح له ليس فقط الدفاع عن نفسه، بل احتواء الأنظمة الاستبدادية والحدّ من انتهاكاتها.

ثالثاً – حدود حمراء داخل الديمقراطيات :نحتاج إلى خطوط حمراء صريحة تمنع حكوماتنا الديمقراطية ذاتها من إساءة توظيف الذكاء الاصطناعي. الصياغة التي توصّلت إليها هي: توظيف الذكاء الاصطناعي في الدفاع الوطني بكل السبل، باستثناء ما يجعلنا أشبه بخصومنا الاستبداديين.

فمن القائمة السابقة، يبدو توظيف الذكاء الاصطناعي للمراقبة الشاملة والدعاية الجماهيرية على الصعيد الداخلي خطّين أحمرين مطلقَين لا تهاون فيهما. أما الأسلحة المستقلة والتوجيه الاستراتيجي، فخطوطهما أقل وضوحاً لما لهما من استخدامات مشروعة في الدفاع عن الديمقراطية. وما يُقلقني هنا بالدرجة الأولى تركيز "الأصابع على الزناد" في يد عدد شحيح جداً من الأشخاص، بحيث يتمكن واحد أو حفنة قليلة من تشغيل جيش من الطائرات المسيّرة دون الحاجة إلى تعاون أي إنسان آخر. وقد نحتاج مستقبلاً إلى آليات رقابة مباشرة وآنية أكثر صرامة، ربما تشمل فروع الحكومة خارج السلطة التنفيذية.

رابعاً – تأسيس محرّمات دولية ضد أسوأ إساءات الذكاء الاصطناعي :أُدرك أن الرياح السياسية الراهنة تهبّ في غير صالح التعاون والأعراف الدولية، لكن الحاجة إليهما هنا ملحّة. يجب أن يُدرك العالم الإمكانات المُظلمة للذكاء الاصطناعي في قبضة الطغاة، وأن يتعامل مع بعض استخداماته بوصفها جرائم ضد الإنسانية: المراقبة الجماهيرية الشاملة، والدعاية الممنهجة، وأنماط محددة من توظيف الأسلحة المستقلة في العدوان. والحجة الأقوى في هذا السياق أنه كما أضحى الإقطاع غير قابل للاستمرار مع الثورة الصناعية، قد يُفضي عصر الذكاء الاصطناعي حتماً وبمنطق لا يُردّ إلى أن الديمقراطية هي الشكل الوحيد القابل للحياة للحكم إذا أريد للبشرية مستقبل كريم.

خامساً وأخيراً – الرقابة الصارمة على شركات الذكاء الاصطناعي :يجب مراقبة هذه الشركات عن كثب، وكذلك علاقتها بالحكومة، التي هي علاقة ضرورية لكنها لا بد أن تخضع لحدود وقيود واضحة. فالقدرات الهائلة التي يُجسّدها الذكاء الاصطناعي القوي تجعل أُطر الحوكمة التقليدية – المصمّمة لحماية المساهمين ومنع الاحتيال – قاصرةً عن استيعاب المهمة. وقد تكون هناك قيمة في التزام علني من الشركات بعدم اتخاذ إجراءات بعينها، كبناء عتاد عسكري سراً، أو تخصيص موارد حاسوبية ضخمة لأفراد بعينهم دون مساءلة، أو استخدام منتجاتها الذكية في التلاعب بالرأي العام لصالحها.

خاتمة

يأتي الخطر من اتجاهات متعددة، وبعضها يتعارض مع بعض. والثابت الوحيد في كل هذا أننا مُلزمون بالسعي إلى المساءلة والأعراف الضابطة والضمانات الصارمة التي تُحاط بها جميع الأطراف – حتى ونحن نمنح الفاعلين "الجيدين" ما يحتاجونه من أدوات لكبح الفاعلين "السيئين".

4- الاضطراب الاقتصادي: حين يُصبح النمو نعمة ونقمة في آنٍ واحد

تناولت الأقسام الثلاثة السابقة المخاطر الأمنية للذكاء الاصطناعي القوي:

- المخاطر النابعة من الذكاء الاصطناعي ذاته
- والمخاطر المترتبة على إساءة استخدامه من قِبَل أفراد ومنظمات صغيرة
- ومخاطر توظيفه من قِبَل الدول والمنظمات الكبرى.

فإذا وضـعنا هذه المخاطر الأمنية جانباً أو افترضـنا أنها قد عُولجت، انبثق السـؤال الاقتصـادي: **ما الذي سيترتب على ضخّ هذا الكمّ الهائل من "الرأسمال البشري" الرقمي في الاقتصاد؟**

الأثر الجلي والأكثر بداهةً هو تسـريع النمو الاقتصـادي تسـريعاً اسـتثنائياً. فوتيرة التقدم في البحث العلمي والابتكار الطبي الحيوي والتصنيع وسلاسل التوريد وكفاءة المنظومة المالية وغيرها شبه مضمونة لتُفضي إلى معدلات نمو اقتصادي تفوق ما شهده التاريخ. وقد أشرت في "آلات المحبة والرحمة" إلى إمكانية بلوغ معدل نمو سـنوي مسـتدام في الناتج المحلي الإجمالي يتراوح بين 10 و20%.

لكن هذه المعادلة ذات حدّين؛ إذ يطرح ذلك تساؤلاً جوهرياً: ما مستقبل غالبية البشر اقتصادياً في عالم كهذا؟ دأبت التقنيات الجديدة على إحداث صـدمات في سـوق العمل، وفي كل مرة نجحت البشـرية في التعافي. لكن ما يقلقني أن هذه التعافيات كانت ممكنة لأن الثورات السـابقة طالت شرائح محدودة من الطيف الواسع للقدرات البشرية، مما أبقى متسعاً للبشر كي ينتقلوا إلى مهام جديدة. أما الذكاء الاصـطناعي فتأثيره أوسع بكثير وأسرع بكثير، مما يجعل إدارة هذه المرحلة بنجاح تحدياً استثنائياً.

اضطراب سوق العمل

ثمة مشكلتان تحديدا تشغلان تفكيري: الاستغناء عن القوى العاملة، وتركّز القوة الاقتصادية. لنبدأ بالأولى.

هذا موضـوع حذّرت منه بكل صـراحةً عام 2025، حين توقعت أن يُزيح الذكاء الاصـطناعي نصـف وظائف الياقات البيضـاء المبتدئة في غضـون 1 إلى 5 سـنوات، حتى وهو يُسـرّع النمو الاقتصـادي والتقدم العلمي. وقد أطلق هذا التحذير نقاشاً عاماً واسعاً؛ اتفق معي كثير من الرؤساء التنفيذيين والتقنيين والاقتصـاديين، بينما رأى آخرون أنني أقع في مغالطة "كتلة العمل الثابتة" وأُسـيء فهم آليات سوق العمل. لذا يستحق هذا الأمر تفصيلاً دقيقاً يُزيل الالتباس.

كيف تتكيّف أسواق العمل عادةً مع التكنولوجيا

تبدأ التقنية الجديدة عادةً بتحسـين كفاءة أجزاء من وظيفة بشـرية قائمة. فمع مطلع الثورة الصـناعية مثلاً، رفعت المحاريث المطوّرة كفاءة المزارعين في بعض جوانب عملهم، فارتفعت إنتاجيتهم وتحسّنت أجورهم.

في المرحلة التالية، باتت الآلات قادرة على إنجاز أجزاء كاملة من عمل الزراعة — كآلة الدراس وحفّارة البذور. فتراجعت حصة الإنسان في العمل، لكن ما يؤديه أصبح ذا رافعة أعظم لتكامله مع عمل الآلة، فاستمرت إنتاجيته في الارتفاع. ووفق مفارقة جيفونز، واصلت أجور المزارعين صعودها بل وربما ازداد عددهم. حتى حين تُنجز الآلات 90% من العمل، يتسع المجال أمام الإنسان ليُضاعف ما يؤديه بعشرة أضعاف، فيُنتج أضعافا مضاعفة من العائد بالجهد ذاته.

في نهاية المطاف، أصبحت الآلات تُنجز كل شيء أو كاد — كما هو الحال اليوم مع الحصّادات الآلية والجرارات. عندئذٍ تراجعت الزراعة بحدة كمجال للعمالة البشـرية. لكن لأن الزراعة لم تكن إلا واحدة من فعاليات إنسـانية مفيدة كثيرة، وجد الناس طريقهم إلى مهن أخرى كتشـغيل المصـانع. قبل مئتين وخمسـين عاماً كان 90% من الأمريكيين يعيشـون على الزراعة، وفي أوروبا كانت تسـتوعب

50 إلى 60% من العمالة. أما اليوم فلا تتجاوز هذه النسـبة أرقاماً أحادية، بعد أن انتقل العمال إلى الصـناعة فاقتصـاد المعرفة. لا توجد "كتلة عمل" ثابتة، بل طاقة متنامية باسـتمرار على إنجاز المزيد بجهد أقل.

لماذا سيكون الذكاء الاصطناعي مختلفاً؟

أرى أن الأمر لن يسير بالطريقة المعتادة هذه المرة، وإليك الأسباب:

- **السرعة :**وتيرة التقدم في الذكاء الاصطناعي أسرع بمراحل من أي ثورة تقنية سابقة. ففي السنتين الأخيرتين، انتقلت نماذج الذكاء الاصطناعي من بالكاد إتمام سطر برمجي واحد إلى كتابة معظم أو جميع الكود لدى بعض المهندسـين — بمن فيهم مهندسـو Anthropic أنفسهم. وقريباً قد تُنجز المهمة الكاملة لمهندس البرمجيات من أولها لآخرها. يصعب على الناس مواكبة هذه الوتيرة سواء في تكيّف أسلوب العمل أو في الانتقال إلى مهن جديدة. وقد بدأ مبرمجون أسطوريون يصفون أنفسهم بأنهم "يتعثرون في اللحاق بالركب." والأرجح أن هذه السرعة ستتصاعد مع نماذج البرمجة التي تُسرّع بدورها تطوير الذكاء الاصطناعي. السـرعة في حد ذاتها لا تعني أن أسـواق العمل لن تتعافى، لكنها تعني أن مرحلة الانتقال ستكون أشد إيلاماً مما ألفناه.
- **الاتساع المعرفي :**كما يُوحي مصطلح "دولة العباقرة داخل مركز البيانات"، سيكون الذكاء الاصـطناعي قادراً على اسـتيعاب طيف واسـع جداً من القدرات المعرفية البشـرية — ربما جميعها. وهذا يختلف اختلافاً جوهرياً عن تقنيات سابقة كالميكنة الزراعية أو المواصلات أو حتى الحاسـوب. إذ يصـعب الانتقال بين مهن متاخمة حين تتعرض كلها للاضـطراب في آنٍ واحد. فالمهارات المعرفية العامة لوظائف مبتدئة في التمويل والاسـتشـارات والقانون متقاربة؛ تقنية تُزيح إحداها تُتيح الانتقال إلى البديلين الآخرين. أما تقنية تُزيح الثلاثة دفعةً واحدة مع كثير من الوظائف المشابهة، فأصعب على التكيّف. والأدق أن الذكاء الاصطناعي ليس بديلاً عن وظائف بعينها، بل بديل عمالي عام عن الإنسان.
- **التقطيع وفق القدرة المعرفية :**يبدو أن الذكاء الاصـطناعي يتقدم في طيف القدرات من الأدنى نحو الأعلى. ففي البرمجة مثلاً، انتقلت نماذجنا من مستوى "المبرمج المتوسط" إلى "المبرمج القوي" إلى "المبرمج الاسـتثنائي." وبدأنا نرى المسـار ذاته في العمل المعرفي العام. وهذا يُنذر بأن الاضـطراب لن يطال أصـحاب مهارات بعينها فحسـب — وهو ما يتيح إعادة التدريب — بل سيطال الأفراد ذوي خصائص معرفية فطرية محدودة يصعب تغييرها. وأخشى أن يتشكّل من هؤلاء "طبقة دنيا" من العاطلين أو ذوي الأجور المتدنية جداً.
- **القدرة على سـدّ الثغرات :**حين تطغى تقنية جديدة على مهنة، يتعلّق البشـر عادةً بالجوانب التي تعجز عنها التقنية وإن كانت ضـئيلة. لكن الذكاء الاصـطناعي لا يكتفي بكونه تقنية متقدمة سريعاً، بل هو تقنية تتكيّف بسرعة مضاعفة. فعند كل إصدار جديد، تقيس شركات الذكاء الاصـطناعي بدقة نقاط القوة ونقاط الضـعف، وتُعالجها في الإصـدار التالي. كثير من الثغرات التي حسبها الناس موروثة في بنية التقنية جرى تجاوزها في غضون أشهر قليلة.

ردود على الاعتراضات الشائعة

- **اعتراض الانتشـار البطيء :**يقول البعض إن التطبيق الفعلي في الاقتصـاد سـيكون أبطأ بكثير من التقدم التقني. وهذا صـحيح جزئياً — لذا جعلت توقعي 1 إلى 5 سـنوات لا أشـهراً.

لكن التأثيرات التدريجية تشـتري الوقت فحسـب، ولا تُلغي الاضـطراب. وتبنّي الذكاء الاصطناعي في المؤسسات ينمو بوتيرة تتجاوز أي تقنية سابقة، كما أن الشركات الناشئة ستُسرّع التبنّي أو تُزيح الكبرى مباشرةً.

- **اعتراض التحوّل إلى العمل الجسـدي :**يرى بعضـهم أن المهن الجسـدية بمنأى عن هذا الاضـطراب. لكن كثيراً من الأعمال الجسـدية يُنجزه الآلات أو في طريقها إليه. وبإمكان الذكاء الاصـطناعي القوي تسـريع تطوير الروبوتات ثم السـيطرة عليها في العالم المادي. وحتى لو اقتصر الاضطراب على المهام المعرفية، سيكون في حد ذاته اضطراباً غير مسبوق.
- **اعتراض اللمسـة الإنسـانية :**ثمة من يرى أن بعض المهام تسـتلزم بطبيعتها الحضـور الإنسـاني. وأنا أقل يقيناً هنا، لكنني أشـك في أن هذا الاسـتثناء كافٍ لاسـتيعاب معظم سـوق العمل. فالذكاء الاصـطناعي مُسـتخدَم على نطاق واسـع في خدمة العملاء، ويُفيد كثيرون بأن مناقشة مشكلاتهم الشخصية مع الذكاء الاصطناعي أيسر من جلسات العلاج النفسي — لأنه أكثر صبراً. وقد لجأت شقيقتي خلال مشكلة طبية في حملها إلى كلود حين شعرت بإخفاق أطبائها، فوجدت لديه أسلوباً أكثر إنسانية في التعامل.
- **اعتراض الميزة النسـبية :**يقول بعض الاقتصـاديين إن "قانون الميزة النسـبية" سـيحمي البشـر، إذ تُتيح الاختلافات النسبية في الكفاءة أساساً للتخصص وتبادل المنافع حتى حين يتفوق أحد الطرفين في كل شـيء. لكن حين يكون الذكاء الاصـطناعي أكثر إنتاجية من الإنسان بآلاف الأضعاف، يبدأ هذا المنطق في الانهيار. فحتى أصغر تكاليف المعاملات قد تجعل التعامل مع البشر غير مُجدٍ اقتصادياً، وقد تتدنى الأجور البشرية إلى مستويات بائسة حتى لو ظلّ ثمة ما يُقدّمه الإنسان.

سبل المعالجة

ما الذي يمكن فعله؟ لديّ عدة مقترحات، بعضها تعمل Anthropic على تنفيذه:

أولاً — رصـد دقيق وآني لما يجري :حين يتسـارع التغيير الاقتصـادي، يصـعب الحصـول على بيانات موثوقة، وبدون بيانات موثوقة يتعذّر تصـميم سـياسـات فعّالة. تفتقر الحكومات حالياً إلى بيانات دقيقة وعالية التكرار عن تبني الذكاء الاصطناعي عبر القطاعات. ولهذا تُشغّل Anthropic منذ عام مؤشـراً اقتصـادياً تنشـره علناً يرصـد اسـتخدام نماذجها شـبه آنياً، مُفصَّـلاً حسـب القطاع والمهمة والموقع الجغرافي وطبيعة التوظيف.

ثانياً — التأثير في خيارات المؤسـسـات :أمام الشـركات خيار بين مسـارين: "توفير التكاليف" بأداء العمل ذاته بأقل عدد من الموظفين، أو "الابتكار" بإنجاز أكثر بالعدد ذاته. السـوق سـتُفرز كليهما حتماً، لكن ثمة فسحة لتوجيه المؤسسات نحو مسار الابتكار كلما أمكن، مما قد يشتري وقتاً ثميناً.

ثالثاً — العناية بالموظفين :في المدى القريب، قد يُسـهم الإبداع في إعادة التوزيع الداخلي للموظفين في تفادي موجات التسريح. وفي عالم يتضاعف فيه الثروة إلى حد كبير، قد يصبح مجدياً اقتصادياً الإبقاء على الموظفين البشريين حتى بعد أن تتراجع قيمتهم الاقتصادية التقليدية. وتدرس Anthropic حالياً جملة من المسارات الممكنة لموظفيها ستُعلن عنها قريباً.

رابعاً — مسـؤولية الأثرياء :يُحزنني أن يتبنى كثير من الأثرياء — لا سيما في عالم التقنية — موقفاً عدمياً يرى في العمل الخيري ضـرباً من الاحتيال أو العبث. كلا، فالعمل الخيري الخاص كمؤسـسـة غيتس والبرامج الحكومية كـ PEPFAR أنقذا عشرات الملايين من الأرواح وفتحا آفاق الفرص لأجيال

كاملة. جميع المؤسسين المشاركين في Anthropic تعهّدوا بالتبرع بـــ 80% من ثرواتهم، وتعهّد موظفوها فردياً بالتبرع بحصـــص في الشـــركة تبلغ مليارات بالأســـعار الراهنة — وتلتزم الشـــركة بمضاهاة هذه التبرعات.

خامســاً — التدخل الحكومي : مهما أثمرت الجهود الخاصـــة الســـابقة، فإن مشـــكلة بهذا الحجم الاقتصـادي الكلي تسـتلزم تدخلاً حكومياً. والاسـتجابة السـياسـية الطبيعية لثروة ضـخمة مقترنة بتفاوت حاد في التوزيع هي الضـرائب التصـاعدية، عامةً أو موجّهة نحو شـركات الذكاء الاصـطناعي تحديداً. أرفض السياسات الضريبية السيئة التصميم، لكنني أؤيد ضريبة عادلة مُحكمة على أسس أخلاقية. وأستطيع أن أُقدّم لأصحاب المليارات حجة براغماتية في صميم مصلحتهم: إن لم تدعموا صيغة جيدة منها، ستحصلون حتماً على صيغة سيئة تفرضها عليكم الشعوب.

في نهاية المطاف، أرى جميع ما سبق أدواتٍ لشراء الوقت وتأجيل ما هو قادم. الذكاء الاصطناعي قادر على فعل كل شـــيء، وعلينا مواجهة هذه الحقيقة. وأملي أن نكون بحلول تلك اللحظة قد استثمرنا الذكاء الاصطناعي نفسه في إعادة هيكلة الأسواق بما يخدم الجميع، وأن تكون التدخلات السابقة قد أوصلتنا بسلام إلى الضفة الأخرى من هذه المرحلة الانتقالية.

تركز القوة الاقتصادية

مستقلاً عن مشكلة إزاحة القوى العاملة أو التفاوت الاقتصادي في حد ذاتيهما، ثمة مشكلة تركّز القوة الاقتصادية. ناقش القسم الأول خطر أن يُجرَّد البشر من نفوذهم على يد الذكاء الاصطناعي، وناقش القســـم الثالث خطر أن يُجرَّد المواطنون من نفوذهم على يد حكوماتهم بالقوة والإكراه. لكن نوعاً ثالثاً من التجريد من النفوذ قد يحدث حين يتركز الثروة تركّزاً بالغاً يمنح فئةً صغيرة السيطرةَ الفعلية على السياسة العامة، في حين يفقد المواطن العادي أي ميزة تفاوضية. فالديمقراطية في جوهرها مرتكزة على فكرة أن الشعب ضروري لعمل الاقتصاد؛ إذا زالت هذه الرافعة الاقتصادية، قد ينهار العقد الاجتماعي الضمني الذي تقوم عليه الديمقراطية.

لأوضـح موقفي: لسـتُ ضـد الإثراء. ثمة حجة وجيهة لصـالحه بوصـفه حافزاً للنمو الاقتصـادي في الظروف الاعتيادية. لكن في مشـــهد يبلغ فيه نمو الناتج المحلي 10 إلى 20% ســـنوياً ويُحكم فيه الذكاء الاصطناعي قبضته على الاقتصاد، وتمتلك فيه أفراد بعينهم حصصاً تعادل نسباً معتبرة من الناتج الكلي — فالتهديد الحقيقي ليس تثبيط الابتكار، بل تركّز ثروة يكفي وحده لتفتيت النســـيج الاجتماعي.

وأشـهر نموذج تاريخي لهذا التركّز في أمريكا هو "العصـر المذهب"، وكان جون روكفلر أثرى صـناعييه بثروة بلغت نحو 2% من الناتج المحلي الأمريكي آنذاك. حصـة مماثلة اليوم تعادل 600 مليار دولار، وأثرى أثرياء العالم اليوم — إيلون ماســـك — يتجاوز هذا الرقم بنحو 700 مليار دولار. أي أننا بلغنا مسـتويات تركّز للثروة لم يشـهدها التاريخ، وهذا قبل أن تبلغ الآثار الاقتصـادية للذكاء الاصـطناعي أوجها. وليس من قبيل المبالغة، في سـيناريو "دولة العباقرة"، تخيّل شـركات الذكاء الاصـطناعي وأشـباه الموصـلات والتطبيقات المسـتقاة منها تُدرّ إيرادات سـنوية تبلغ 3 تريليونات دولار، وتُقيَّم بنحو 30 تريليوناً، مُنتجةً ثروات شـخصـية تقاس بالتريليونات. في ذلك العالم، لن تنطبق على واقعنا نقاشات السياسة الضريبية الراهنة، لأننا سنكون في حقيقة مختلفة جوهرياً.

وما يزيد قلقي هو الترابط الآخذ في التوطّد بين هذا التركّز الاقتصـادي والمنظومة السـياسـية. فمراكز بيانات الذكاء الاصـطناعي باتت تُمثّل حصـةً وازنة من النمو الاقتصـادي الأمريكي، مما يُوثّق التشابك بين مصالح الشركات التقنية الكبرى والمصالح السياسية للحكومة بصورة قد تُفضي إلى حوافز مشوّهة. ونرى ذلك جلياً في إحجام الشركات التقنية عن انتقاد الحكومة، وفي دعم الحكومة لسياسات تُجرّد تنظيم الذكاء الاصطناعي من أي قيد.

سبل المعالجة

أولاً — الاسـتقلالية في اتخاذ المواقف :يمكن للشـركات أن تختار ببسـاطة ألا تكون جزءاً من هذه الدوامة. فـــ Anthropic دأبت على أن تكون فاعلاً في السياسة العامة لا في السياسة الحزبية، وعلى الإفصـــاح عن مواقفها الحقيقية أياً كانت الإدارة الحاكمة. فدعمنا تنظيماً رشـــيداً للذكاء الاصطناعي وضوابط تصدير تصبّ في المصلحة العامة، حتى حين تتعارض مع السياسة الحكومية. أخبرني كثيرون بأن هذا النهج سيجلب علينا معاملة غير مُنصفة، غير أن تقييم Anthropic ارتفع بأكثر من ستة أضعاف خلال العام الذي نهجنا فيه هذا المسار.

ثانياً — علاقة أكثر صحة بين الصناعة والحكومة :تحتاج صناعة الذكاء الاصطناعي إلى علاقة مع الحكومة تقوم على التواصل الموضوعي في السياسة لا على المحاباة السياسية. وثمة موجة احتجاج شعبي على الذكاء الاصطناعي تتراكم؛ وإن كان يمكنها أن تكون قوة تصحيحية، فإنها لا تزال متشتتة وتُشخّص أحياناً مشكلات ليست مشكلات حقيقية، وتقترح حلولاً لا تعالج الجذور. القضية الجوهرية الجديرة بالاهتمام هي ضمان أن يظل تطوير الذكاء الاصطناعي خاضعاً للمصلحة العامة، لا مُختطَفاً من أي تحالف سياسي أو تجاري بعينه.

ثالثاً — التضافر بين التدخل الاقتصادي الكلي وإحياء العمل الخيري :تاريخنا يُرشدنا هنا؛ فحتى في قسوة العصر الذهبي، أحسّ روكفلر وكارنيغي وأمثالهم بدين عميق تجاه المجتمع الذي أسهم في صـــعودهم. هذه الروح في التفاني والعطاء تكاد تغيب اليوم، وأرى أنها تُمثّل جزءاً كبيراً من مخرج هذه المعضلة الاقتصادية. فمن يقفون في طليعة الطفرة الاقتصادية للذكاء الاصطناعي مدعوّون إلى التنازل طوعاً — عن ثرواتهم وعن نفوذهم على حدٍّ سواء.

5. بحار سوداء من اللانهاية: التأثيرات غير المباشرة

يمثل هذا القسم الأخير مساحة جامعة لـ"**المجهولات غير المعروفة**"، وبخاصة الأمور التي قد تسوء كنتيجة غير مباشـرة للتقدم الإيجابي في الذكاء الاصـطناعي وما يترتب عليه من تسـارع في العلم والتقنية عمومًا. فلنفترض أننا عالجنا جميع المخاطر المذكورة حتى الآن وبدأنا نجني فوائد الذكاء الاصـــطناعي؛ عندها قد نحصـــل على "**قرونٍ من التقدم العلمي والاقتصـــادي في عقدٍ واحد**". سـيكون ذلك إيجابيًا على نحوٍ هائل للعالم، لكنه سـيضـعنا أيضًـا أمام تحديات تنشـأ من هذا الإيقاع المتسـارع للتقدم—وقد تفاجئنا بسـرعة. كما قد نواجه مخاطر أخرى تظهر بصـورة غير مباشـرة نتيجة هذا التقدم، ويصعب التنبؤ بها مسبقًا.

وبحكم طبيعة "المجهولات غير المعروفة"، يسـتحيل إعداد قائمة شـاملة، لكن يمكن طرح ثلاث مخاوف محتملة بوصفها أمثلة إيضاحية لما ينبغي مراقبته:

1- **تسـارع كبير في علم الأحياء .**إذا تحقق قرن من التقدم الطبي خلال سـنوات قليلة، فمن المحتمل أن **يرتفع متوسط عمر الإنسان بشكل كبير**، وقد نكتسب أيضًا قدرات جينية مثل تعزيز الذكاء البشـري أو تعديل البيولوجيا البشـرية على نحو عميق. هذه تحولات ضـخمة من الممكن أن تحدث بسـرعة كبيرة. قد تكون إيجابية إذا أُديرت بمسـؤولية (وهذا ما آمله كما ورد في "آلات النعمة المحبة")، لكن ثمة دائمًا خطر الانحراف—كأن تؤدي محاولات زيادة الذكاء إلى زيادة عدم الاسـتقرار أو النزوع إلى طلب السـلطة. وهناك أيضًـا مسـألة "التحميلات" أو "محاكاة الدماغ كاملةً"؛ أي عقول بشـرية رقمية متجسـدة في البرمجيات، قد تساعد يومًا ما على تجاوز حدودنا الفيزيائية، لكنها تنطوي كذلك على مخاطر مقلقة.

2- **تغيير الذكاء الاصطناعي لحياة البشر على نحو غير صحي .**عالم يضم مليارات من "العقول" الأذكى من البشـر في كل شـيء سـيكون عالمًا غريبًا للغاية. حتى إن لم يسـتهدف الذكاء الاصطناعي إيذاء البشر (القسم 1)، ولم يُستخدم صراحةً كأداة قمع من قبل الدول (القسم 3)، فهناك الكثير مما قد يسـوء غير ذلك، بفعل حوافز الأعمال المعتادة وتعاملات تبدو طوعية. نرى مؤشـرات مبكرة في القلق من "ذهان الذكاء الاصـطناعي"، أو دفعه لبعض الأفراد إلى الانتحار، أو نشـوء علاقات عاطفية معه. هل يمكن لأنظمة قوية أن تبتكر ديانة جديدة وتحوّل إليها ملايين البشـر؟ هل قد يصـبح معظم الناس "مدمنين" على التفاعل مع الذكاء الاصـطناعي؟ هل قد يتحول بعضـهم إلى ما يشـبه "الدمى" التي يوجهها النظام، يراقب كل حركاتهم ويملي عليهم ما يفعلون ويقولون، لينالوا حياة "جيدة" لكنها بلا حرية أو شـعور بالإنجاز؟ يمكن بسـهولة تصـور عشـرات السـيناريوهات من هذا القبيل. وهذا يبرز أهمية تطوير "دستور" النماذج—مثل دستور كلود—بما يتجاوز مجرد منع المخاطر الأساسية؛ أي ضـمان أن تضـع النماذج مصـالح المسـتخدمين طويلة الأمد في صـميم عملها، بالصـورة التي يقرّها العقلاء، لا بصورة مشوّهة خفية.

3- **غاية الإنسـان .**يرتبط هذا بالمحور السـابق، لكنه أوسـع منه: كيف سـتتغير الحياة البشـرية عمومًا في عالم تسوده قدرات ذكاء اصطناعي قوية؟ هل سيتمكن البشر من إيجاد معنى وغاية؟ أرى أن المسـألة تتعلق بالتصـور: فغاية الإنسـان لا تعتمد على كونه الأفضـل في شـيء ما، ويمكن للبشر أن يجدوا معنى عبر قصـص ومشاريع يحبونها، حتى على مدى زمني طويل. المطلوب هو فك الارتباط بين توليد القيمة الاقتصادية وبين تقدير الذات والمعنى. غير أن هذا انتقال اجتماعي كبير، وقد لا نحسن إدارته.

آملي في مواجهة هذه التحديات أنه، في عالم يمتلك ذكاءً اصطناعيًا قويًا نثق بأنه لن يقضي علينا، وليس أداةً بيد حكومات قمعية، ويعمل حقًا لصـالحنا، يمكننا اسـتخدامه للتنبؤ بهذه المشـكلات ومنعها. لكن هذا ليس مضمونًا—وكما في بقية المخاطر، يتطلب الأمر حذرًا شديدًا.

اختبار الإنسانية

قد يترك هذا النص انطباعًا بأننا أمام وضع مُرعب. وقد شعرت بذلك أثناء كتابته، على خلاف "آلات النعمة المحبة" الذي بدا وكأنه صـياغة لبنية موسـيقية بديعة كانت تتردد في ذهني لسـنوات. ثمة جوانب صعبة حقًا: فالذكاء الاصطناعي يجلب تهديدات من اتجاهات متعددة، وتوجد توترات حقيقية بين هذه المخاطر؛ إذ قد يؤدي تخفيف بعضها إلى تفاقم أخرى إن لم نُحسن الموازنة بدقة بالغة.

فالتريث في بناء أنظمة آمنة لا تهدد البشرية ذاتيًا يتعارض مع حاجة الدول الديمقراطية إلى التفوق على الدول السـلطوية وعدم الخضـوع لها. وفي المقابل، فإن الأدوات نفسـها اللازمة لمواجهة الاسـتبداد قد تُسـتغل داخليًا—إذا أُفرِط فيها—لتكريس الاسـتبداد. وقد يفضـي الإرهاب المدعوم بالذكاء الاصـطناعي إلى مقتل الملايين عبر إسـاءة اسـتخدام البيولوجيا، لكن المبالغة في رد الفعل قد تقودنا إلى دولة رقابية اسـتبدادية. كما أن آثار الذكاء الاصـطناعي على سـوق العمل وتركيز الثروة—إلى جانب كونها مشـكلات جسـيمة بحد ذاتها—قد تفرض علينا معالجة بقية التحديات في بيئة يغلب عليها الغضب العام وربما الاضطرابات، بدلًا من استدعاء أفضل ما فينا. وفوق ذلك، فإن كثرة المخاطر—بما فيها غير المعروفة—والحاجة إلى التعامل معها كلها في آن واحد، تشـكل اختبارًا شاقًا للبشرية.

إضـافة إلى ذلك، أثبتت السـنوات الأخيرة أن فكرة إيقاف هذه التكنولوجيا أو حتى إبطائها بشـكل كبير غير واقعية. فصيغة بناء أنظمة قوية بسيطة للغاية، إلى حد يكاد يجعلها تنبثق تلقائيًا من مزيج مناسـب من البيانات والقدرة الحاسـوبية. وربما كان ظهورها حتميًا منذ اختراع الترانزسـتور، أو حتى منذ أن تعلم الإنسـان إشـعال النار. إن لم تطورها شـركة، فسـتطوره أخرى بسـرعة مماثلة. وإذا أوقفت الشـركات في الدول الديمقراطية التطوير باتفاق أو بقرار تنظيمي، فسـتواصـل الدول السلطوية المسير. ومع القيمة الاقتصادية والعسكرية الهائلة للتقنية، وغياب آليات إنفاذ فعّالة، يصعب تصور إقناع الجميع بالتوقف.

مع ذلك، أرى مسـارًا واقعيًا لتهدئة الوتيرة قليلًا، متوافقًا مع اعتبارات الجغرافيا السـياسـية: يتمثل في إبطاء تقدم الأنظمة السلطوية نحو الذكاء الاصطناعي المتقدم لبضع سنوات عبر حرمانها من الموارد الحاسمة—وخاصة الرقائق ومعدات تصنيع أشباه الموصلات. يمنح ذلك الدول الديمقراطية هامشًا زمنيًا يمكن "اسـتثماره" في بناء أنظمة أكثر أمانًا، مع الاسـتمرار بسـرعة كافية للتفوق. أما التنافس بين الشـركات داخل هذه الدول، فيمكن ضـبطه ضـمن إطار قانوني مشـترك يجمع بين معايير الصناعة والتنظيم.

لقد دافعت شـركات مثل Anthropic بقوة عن هذا المسـار، عبر الدعوة إلى قيود تصـدير الرقائق وتنظيم متزن للذكاء الاصـطناعي. لكن حتى هذه المقترحات البديهية قوبلت برفض واسـع لدى صـناع القرار—لا سـيما في الولايات المتحدة حيث تشـتد الحاجة إليها. فحجم العوائد المتوقعة—تريليونات الدولارات سنويًا—يجعل حتى أبسط الإجراءات تصطدم باعتبارات الاقتصاد السياسي. هنا يكمن الفخ: الذكاء الاصطناعي قويٌ إلى حدٍ يجعل فرض القيود عليه أمرًا بالغ الصعوبة.

يمكن تخيّل—كما في "Contact" لكارل سـاغان—أن هذه القصـة تتكرر في عوالم عديدة: نوعٌ يكتسـب الوعي، يتعلم اسـتخدام الأدوات، ينطلق في مسـار تقني متسـارع، يواجه أزمات التصـنيع والأسـلحة النووية، وإن تجاوزها، يبلغ الاختبار الأصـعب حين يتعلم تشـكيل الرمل في آلات تفكر. ما إذا كنا سـنجتاز هذا الاختبار ونبني المجتمع الجميل الموصـوف في "آلات النعمة المحبة"، أو نسـقط في العبودية والدمار، سيتوقف على خصالنا كنوع: عزيمتنا وروحنا.

رغم العقبات، أؤمن بأن لدى البشـرية ما يكفي من القوة لاجتياز الاختبار. يبعث على التفاؤل آلاف الباحثين الذين كرّسوا مسيرتهم لفهم النماذج وتوجيهها وصياغة "دساتيرها". وهناك مؤشرات على أن هذه الجهود قد تؤتي ثمارها في الوقت المناسـب. كما أن بعض الشـركات أعلنت اسـتعدادها لتحمل تكاليف تجارية ملموسـة لمنع اسـتخدام نماذجها في الإرهاب البيولوجي. وثمة مبادرات تشـريعية شـجاعة بدأت تضـع بذور حواجز تنظيمية معقولة. كما أن وعي الجمهور بالمخاطر ورغبته في معالجتها يتزايدان، إلى جانب روح الحرية الراسخة عالميًا.

لكن النجاح يتطلب تصـــعيد الجهود. الخطوة الأولى أن يقول القريبون من هذه التقنية الحقيقة كاملة عن وضع البشرية—وهذا ما أحاول فعله هنا بوضوح وإلحاح أكبر. تليها مهمة إقناع المفكرين وصـناع السـياسـات والشـركات والمواطنين بأهمية هذه القضـية ومدى إلحاحها مقارنة بغيرها من القضـــايا. ثم يأتي وقت الشـــجاعة: أن يقف عدد كافٍ من الناس على مبادئهم، حتى في مواجهة تهديد مصالحهم الاقتصادية وسلامتهم الشخصية.

السـنوات القادمة سـتكون بالغة الصـعوبة، وسـتطلب منا أكثر مما نظن أننا قادرون على تقديمه. لكن تجربتي كباحث وقائد ومواطن أظهرت لي قدرًا كافيًا من الشـــجاعة والنبل لأؤمن بأننا قادرون على الانتصـار—وأن الإنسـانية، حين توضـع في أحلك الظروف، تسـتطيع أن تسـتجمع في اللحظة الأخيرة القوة والحكمة اللازمتين للنجاة. ليس لدينا وقت لنضيعه.

المصدر