براء علي ديب | *Baraa Ali Deeb
عمر المغربي | **Omar Almagharebi
فادي زراقط | ***Fadi A. Zaraket

# مستقبلات الذكاء الاصطناعي والثقافات الضعيفة التمثيل

## كيف ستصوغ النماذج اللغوية الضخمة المشهد المعرفي في المستقبل؟

# AI Futures and Underrepresented Cultures

## How Will Large Language Models Shape the Cognitive Landscape of the Future?

| التسلم Received | التعديل Revised | القبول Accepted | الرقم التعريفي DOI |
| --- | --- | --- | --- |
| 2025-3-03 | 2025-3-18 | 2025-3-21 | https://doi.org/10.31430/AZEH1427 |

**ملخص:** تستشرف هذه الدراسة أثر تصدُّر الذكاء الاصطناعي ونماذجه اللغوية الضخمة للمشهد المعرفي في علاقتها بالتنوع والتعدد اللغوي، وأثرها في "الثقافات الضعيفة التمثيل"، من حيث انحيازاتها المعرفية والثقافية التي تمسّ المجتمعات ذات الحضور الرقمي المحدود. وتستعرض سبل تجاوز فجوة البيانات بين اللغات، وتتطرّق إلى رسم عدد من السيناريوهات المُستقبلية لمصائر الثقافات المحلية واللغات ومعارفها. كما تبحث في الخيارات المتاحة للحكومات وصُنّاع القرار في البلدان ذات اللغات الضعيفة التمثيل في خريطة الذكاء الاصطناعي العالمي، الموسومة اليوم بالتنافس الحاد بين عمالقة الذكاء الاصطناعي في الولايات المتحدة الأميركية والصين. وتخلص إلى أن التعاون في بناء نماذج لغوية ضخمة وقواعد بيانات متضافرة ومتكاملة، وفرق بحث وتطوير متعاونة، بين البلدان الضعيفة التمثيل، والتي تشترك في الثقافة واللغة، هو الخيار الأفضل لها.

**كلمات مفتاحية:** الذكاء الاصطناعي، النماذج اللغوية الضخمة، الثقافات الضعيفة التمثيل، التحيزات، قواعد البيانات، الفجوة الرقمية.

**Abstract:** This study examines the future of artificial intelligence (AI) through the lens of "Large Language Models" (LLMs), focusing on their implications for linguistic diversity, multilingualism, and their impact on "Underrepresented Cultures". It addresses cognitive and cultural biases affecting societies with limited digital presence. The paper investigates strategies to mitigate data disparities among languages and proposes various scenarios for the preservation of local cultures, languages, and their associated knowledge systems. Additionally, it evaluates options available to governments and policymakers in countries with underrepresented languages on the global AI landscape, which is currently dominated by intense competition between AI leaders in the United States and China. The paper concludes that fostering cooperation among underrepresented countries – through the collaborative development of large language models, integrated databases, and joint research initiatives – offers the most viable path forward for preserving shared cultural and linguistic heritage.

**Keywords:** AI, Large Language Models, Underrepresented Cultures, Biases, Databases, Digital Divide.

* باحث مساعد في وحدة دراسة المجال الرقمي والاجتماعي العربي، المركز العربي للأبحاث ودراسة السياسات، قطر.
Assistant Researcher, Unit for Research in Arab Digital Social Spaces, ACRPS, Qatar.
** باحث مساعد في المركز العربي للأبحاث ودراسة السياسات.
Assistant Researcher, Arab Center for Research and Policy Studies.
*** مدير وحدة دراسة المجال الرقمي والاجتماعي العربي، المركز العربي للأبحاث ودراسة السياسات، قطر.
Director of the Unit for Research in Arab Digital Social Spaces, ACRPS, Qatar. (Corresponding Author)
Email: fadi.zaraket@dohainstitute.edu.qa

## مقدمة

يجادل إدوارد سعيد في كتابه **الاستشراق** بأنّ المعرفة الغربية التي تشكّلت في القرنين الثامن عشر والتاسع عشر عن الشرق لم تكن محايدةً أو موضوعية، بل استُخدمت أداةً لتعزيز الهيمنة الاستعمارية عبر تشكيل خطاب استشراقي يُصوِّر الشرق بوصفه "آخَرَ" غريبًا، ومتخلّفًا، وثابتًا، مقابل الغرب المتقدم والعقلاني[1]. ومن ثمّ، يكشف سعيد العلاقة غير البريئة بين المعرفة واللغة والسلطة، مستندًا في ذلك إلى كتابات ميشيل فوكو عن السلطة والخطاب.

إلا أن سعيد لم يتناول علاقة خطاب المعرفة بتقنيات إنتاجها، والتي لم تشهد تحولاتٍ جذرية منذ اختراع الآلة الطابعة في القرن الخامس عشر. في حين بدأت تقنيات إنتاج المعرفة تشهد تحوّلاتٍ عميقة مؤخرًا، تمثّلت في العقود الماضية بخفض تكاليف الإنتاج والتوزيع، وتوسيع قاعدة المنتجين والمستهلكين، وتقليص المدة المطلوبة لتوصيل المعلومة. وكلّ ذلك بفضل رقمنة المعرفة، وظهور الشبكة العنكبوتية، ووسائل الاتصالات الرقمية، ووسائط التواصل الاجتماعي. ويتبلور اليوم تحوّلٌ مفصلي جديد، يتعلق بآليات إنتاج المعرفة نفسها؛ وهو ما يتصل بدخول "النماذج اللغوية الضخمة" (Large Language Models) للاستعمال العام، طارحةً تحديات معرفية وأخلاقية وإبستيمولوجية جديدة.

تثير هذه النماذج تحدياتٍ ملحّة، من أهمها تأثيرها في طرائق البحث والتأليف. فعلى عكس موجات "التأليل" (Automation) السابقة، التي اقتصر أثرها على استبدال العمالة اليدوية، باتت عدّة وظائف إدراكية، متنوّعة ومعقّدة نوعًا ما، قابلة للتأليل. فعلى سبيل المثال، تبرع النماذج سرعةً وأداءً في الخدمات "التحويلية" (Transformative)، كالتلخيص، والترجمة، وتفريغ النصوص (Transcription)، ووصف الصور، والمقارنة الدلالية، وبناء قواعد المراجع البحثية، إضافة إلى الخدمات الاستخراجية، كاستخراج الكيانات، و"تصنيف الموضوعات" (Topic Classification)، و"التعرّف إلى الكيانات" (Object Recognition)، والخدمات التحليلية، كتحديد أهم العناصر المؤثرة، وكشف موقع الكلمة في الجملة. وتتخطى النماذج هذه الوظائف إلى مهمات توليدية وإبداعية لإكمال إنتاج النصوص والصور والموسيقى والمقطوعات البصرية؛ تؤديها بمستوى مقبول وفي طور التحسُّن. ويثير هذا التحدي أسئلةً جوهرية بشأن مستقبل إنتاج المعرفة وسلطة الخطاب.

تنشأ الأسئلة من تحكّم أنظمة قيمية وثقافية محدّدة في مخرجات هذه النماذج المتنوّعة، ومن ثمّ في غالبيةٍ وازنة من كمّ الإنتاج والتداول المعرفي المستقبلي. ويأتي هذا التحكّم مرّة عضويًا بسبب انحيازات في البيانات التي استُخدمت في تدريب النماذج، ومرّات من سياسات تفضيلية توجيهية تحدّد سلوك النماذج، تختارها وتقرّرها نخبٌ ضيقة من منتجي هذه النماذج في شركات التكنولوجيا الكبرى. وتشمل هذه السياسات مواقف حاسمة في عددٍ مهم من القضايا الإشكالية في العالم، مثل الاحتلال، والاستعمار، والحريات، والحقوق الفردية والاجتماعية، والسرديات التاريخية للجماعات والأقليات، والنزاعات. وتصدر هذه التحيّزات في مختلف مراحل تطوير النماذج، بدءًا من اختيار البنية المعرفية، وتجهيز البيانات، والتلقين الآلي، وصولاً إلى الاختبار، والتقييم، والتنقيح[2].

---

1 ينظر: إدوارد سعيد، **الاستشراق: المفاهيم الغربية للشرق**، ترجمة محمد العناني (القاهرة: رؤية للنشر والتوزيع، 2006 [1978]).

2 من أجل بحثٍ أكثر تفصيلاً في أنواع التحيزات وصورها، ينظر: الهيئة السعودية للبيانات والذكاء الاصطناعي، **التحيز في أنظمة الذكاء الاصطناعي: تحديات وحلول** (الرياض: الهيئة السعودية للبيانات والذكاء الاصطناعي، 2025)، ص 14-29، شوهد في 2025/3/17، في: https://acr.ps/1L9zQU7

من جهةٍ ثانية، تبني أنظمة تشغيل هذه النماذج ملفات شخصية، تفهم مع الوقت خيارات المستخدم وتفضيلاته، ومن ثمّ انحيازاته، وتحاول بنجاح مجاراة هذه الانحيازات في مخرجاتها، ما لم تتضارب مع سياساتها التوجيهية. ومن شأن هذا التفاعل بين المؤلف أو الباحث والنماذج أن يؤثّر مع الوقت في ذاتية الباحث (Subjectivity) ويتشارك معه في عملية الإنتاج البحثي. فقد جرت العادة أن ينخرط الباحثون في العلوم الإنسانية والاجتماعية في عمليات تفسيرية وتحليلية معقّدة، مقدّمين وجهات نظرهم الفكرية والثقافية الخاصة؛ الأمر الذي يجعل انحيازاتهم الثقافية والقيمية غالبًا ما تعمل على نحوٍ مُضمر. أمّا اليوم، فإنّ الأُطر القيمية المضمَّنة في النماذج التوليدية والانحيازات الشخصية التي التقطتها من شأنها أن تؤثّر مباشرةً في الإنتاج المعرفي.

من جهة ثالثة، يهمّش حضور اللغة الإنكليزية المهيمن، مع عددٍ قليل من اللغات الأخرى المهيمنة، في قواعد بيانات تدريب النماذج اللغوية الضخمة، حضورَ اللغات الأضعف تمثيلًا، ومن ثمّ، يقلّل من وزن رؤى قيمية ومعرفية تعكسها ثقافات اللغات الأضعف تمثيلًا. وتبرز هذه القضية، في أثناء التراجع المستمر للتنوع اللغوي في العالم[3]؛ إذ تواجه المزيد من اللغات خطر الانقراض بسبب العولمة والهيمنة الرقمية. وقد تسرّع النماذج اللغوية الضخمة، في شكلها الحالي، هذه الظاهرة من خلال تفضيل الهياكل اللغوية والإبستيمولوجيات السائدة؛ ما يعزز المنظومات الخطابية القائمة في إنتاج المعرفة.

تحفّز هذه التساؤلات دراستنا لطرح مسائل بحثية بشأن استشراف مستقبل النماذج اللغوية الضخمة في علاقتها بالتنوع والتعدد اللغوي؛ فهي تستقصي مدى تمثيل اللغات الضعيفة الموارد وثقافاتها في هذه النماذج، وكيفية تداخل العوامل الاقتصادية والسياسية في حوكمتها، وما يترتب على ذلك من انحيازات معرفية وثقافية تمس المجتمعات ذات الحضور الرقمي المحدود. وتستعرض الدراسة سبل تجاوز فجوة البيانات بين اللغات، وتتطرّق إلى رسم عدد من السيناريوهات المُستقبلية لمصائر الثقافات المحلية واللغات ومعارفها، في أثناء تطور النماذج اللغوية الضخمة السريع، ودخولها المتسارع في أتمتة عمليات إنتاج المعرفة وتداولها. وتبحث الدراسة أيضًا عددًا من الخيارات المتاحة للحكومات وصُنّاع القرار في البلدان ذات اللغات الضعيفة التمثيل في خريطة الذكاء الاصطناعي العالمي؛ سواء فيما يخصّ السياسات العامّة، أو المبادرات البحثية والتنموية لضمان المساهمة الفاعلة في بلورة المعرفة التي تقدّمها النماذج اللغوية الضخمة، وتفادي المصائر المحتملة على الهوية الثقافية والتبعات الاقتصادية لمجتمعات هذه البلدان. وتخلص إلى أن التعاون في بناء وتطوير نماذج لغوية ضخمة وقواعد بيانات متضافرة ومتكاملة، وفرق بحث وتطوير متعاونة، بين البلدان الضعيفة التمثيل، والتي تشترك في الثقافة واللغة، هو الخيار الأفضل لها في أثناء سباقٍ تنافسي يتّجه إلى خنق فرص الآخرين بين عمالقة الذكاء الاصطناعي اليوم في الولايات المتحدة الأميركية والصين.

3 Lindell Bromham et al., "Global Predictors of Language Endangerment and the Future of Linguistic Diversity," *Nature Ecology & Evolution*, vol. 6, no. 2 (February 2022), pp. 163-173.

## أولًا: الطريق نحو النماذج اللغوية التوليدية

يُروى في ملحمة **الأوديسة**[4] أن ساحرةً تُدعى سيرس طلبت من بطل القصة أوديسيوس أن يشدّ الرحال إلى أرض الموتى. وحتى تُعينه على خوض غمار الرحلة، تتلو عليه سيرس عشرين بيتًا شعريًا من التعليمات المفصّلة[5]. وتتميّز هذه الأبيات بجملةٍ من الملامح، من بينها استخدام الجمل الشرطية "عندما، حينها" (Whenever, Then)، وهي سمةٌ، بحسب كريستفور فاروني، ميّزت النبوءات القديمة التي كانت تُتلى على المستعمرين الإغريق[6]. وليس الإغريق أوّل من ابتكر الجمل الشرطية؛ ولكنّ لمنحها خصوصيةً أسلوبيةً دلالاتٍ على موقع هذه المنهجية في الوعي الإنساني المبكّر. يشبه تمثيل المعارف عبر القواعد والخيارات بدايات الذكاء الاصطناعي في القرن الماضي. يُعرّف الذكاء الاصطناعي بأنه تعليم الآلة كيفية التفكُّر مثل البشر، بما يتّسعُ له تعبير التفكُّر من معانٍ تضمُّ الإدراك، والفهم، والتعلّم، وحلّ المشكلات، واتخّاذ القرارات، وفي مستويات طموحة أكثر، على المدى البعيد، الإبداع، والاستقلالية[7]. لذا بطبيعة الحال، استلهمت بواكير الذكاء الاصطناعي قواعد التفكّر البشري المبسّطة بصيغتها الشرطية (If-Then)، البسيطة والفعّالة، مع قدرةٍ أسرع وأكثر سعةً (Larger Scale) من قدرة البشر على معالجتها.

ويشكّل مؤتمر دارتموث في سنة 1956 الانطلاقة المُتعارف عليها تاريخيًا للذكاء الاصطناعي، بوصفه حقلاً علميًّا. فقد قدّم مجموعة من العلماء خلال المؤتمر برنامج المُنظّر المنهجيّ (Logic Theorist)، الذي كان أوّلَ ما يُمكنُ أن يُطلق عليه برنامج ذكاءٍ اصطناعيّ[8]. ونجح البرنامج في إثبات 38 برهانًا من أصل 52 في كتاب أصول الرياضيات (Principia Mathematica) لبرتراند راسل (1827-1970)، وألفريد وايتهيد (1861-1947)[9]، وقامت منهجيته على مجموعة من خوارزميات البرهنة الشكلية (Formal Deduction)، والقواعد الشرطية التي يتبعها، وصولاً إلى الحل. واعتمد البرنامجُ على عدة أساليبَ للبرهنة الرياضية، وتعليمات خاصة في حال الفشل في الوصول إلى الإثبات، إضافةً إلى القدرة على الاستفادة من قاعدة من الفرضيات، والبراهين الناجزة، لحلّ المعضلات المتبقية[10].

---

4 هوميروس، **الأوديسة لشاعر الخلود هوميروس**، ترجمة دريني خشبة (القاهرة/ لندن: مؤسسة هنداوي، 2021)، ص 115-116.

5 Christopher A. Faraone, "Circe's Instructions to Odysseus (OD. 10.507–40) As an Early Sibylline Oracle," *The Journal of Hellenic Studies*, vol. 139 (November 2019), p. 50.

6 Ibid., p. 54.

7 Cole Stryker & Eda Kavlakoglu, "What is Artificial Intelligence (AI)?" *IBM*, 9/8/2024, accessed on 10/2/2025, at: https://acr.ps/1L9zQDu

8 Pamela McCorduck, *Machines Who Think*, 2nd ed. (Natick, MA: AK Peters, Ltd., 2004), p. 559.

9 Amirhosein Toosi et al., "A Brief History of AI: How to Prevent Another Winter (A Critical Review)," *PET Clinics*, vol. 16, no. 4 (October 2021), p. 454.

10 Allen Newell & Herbert A. Simon, "The Logic Theory Machine: A Complex Information Processing System," *IRE Transactions on Information Theory*, vol. 2, no. 3 (September 1956), pp. 61-79.

وتجدر الإشارة إلى أنّ العقدين اللاحقين حفلا بالمنجزات، وارتفع الاستثمار في الذكاء الاصطناعي على نحوٍ كبير[11]. فقد رأت النور خلال تلك الفترة "آلة الهندسة" (Geometry Machine)، من إصدار شركة "آي بي إم" (IBM) سنة 1959، وأول روبوت اصطناعي، وهو الذراع الآلية "يونيميت" (Unimate)، استُخدمَ للعمل في خطوط التجميع في شركة "جنرال موتورز" (General Motors) سنة 1962، وأول بوت دردشة من إصدار معهد ماساتشوستس للتكنولوجيا (MIT) سنة 1969. وابتُكرت لغةُ البرمجة "ليسب" (LISP) في سنة 1958، والتي صُمّمت خصيصًا لتطبيقات الذكاء الاصطناعي، وغلبت على الصنعة لعقودٍ تلت[12]. وفي هذا السياق، ارتفعت التوقعات بشأن مستقبل الذكاء الاصطناعي بنحوٍ مُبالغ فيه. ففي سنة 1970، على سبيل المثال، صّرح مارفين مينسكي (1927-2016) (Marvin Minsky)، وهو أحد الآباء المؤسسين للذكاء الاصطناعي، أنّ ثلاثًا إلى ثماني سنوات تفصلُ البشرية عن الذكاء العام الاصطناعي (Artificial General Intelligence)[13]، أي ذلك الذكاء الاصطناعي القادر على القيام بالمهمات الإدراكية التامة التي يقومُ بها الأذكياء من البشر.

بيد أنه خاب توقّع مينسكي، وفتر الاهتمام بالذكاء الاصطناعي، ضمن ما يُعرف في الأدبيات بـ "شتاء الذكاء الاصطناعي الأول"، الذي استمر حتى ثمانينيات القرن الماضي[14]. واعتقد الباحثون حينها أن "قابلية التوسّع" (Scalability)[15] متوقفة على القدرات الحوسبية والذاكرة، أي زيادة التعليمات المكتوبة للبرامج. ومع مرور الزمن، تطوّر مفهوم الذكاء الاصطناعي من قواعد مصفوفات استنتاج منطقية، لمحاكاة "التفكُّر الإحصائي" (Statistical Reasoning)، وعلوم الاحتمالات (Probabilistic Reasoning)، التي بدأت تدخل عالم التأليل. فأصبح المطلوب إيجاد بُنى إدراكية تستطيع محاكاة الذكاء الإنساني الإحصائي والاحتمالي، عبر التعرّض لمشاهدات (Observations) من المدخلات (Inputs) والمخرجات (Outputs). وظهرت بذلك عدّة بُنى، منها "نموذج ماركوف المخفي" (Hidden Markov Model)، و"شبكة بايس" (Bayesian Network)، و"شجيرات الانحدار اللوجستي" (Logistic Regression Trees)، التي وفّرت تمثيلاً مكلّفًا وتقريبيًّا للمقاربات الإحصائية. ثمّ ظهر مفهوم "الشبكات العصبية" (Neural Networks)[16]، نموذجًا تقريبيًّا فعّالاً في بعض التطبيقات والنطاقات من دون ضمان للصحة، وحظي باهتمام محدود في نطاق تطبيقات بدت مناسبةً له حينها، كمعالجة الصور والبصريات[17]. وتزايد الاهتمام بالمفهوم في سنة 1986،

11 Michael Haenlein & Andreas Kaplan, "A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence," *California Management Review*, vol. 61, no. 4 (2019), p. 8.

12 Toosi et al., pp. 455, 456.

13 Haenlein & Kaplan, p. 8.

14 Ibid., p. 9.

15 واجهنا خلال البحث تحدّيًا في ترجمة المصطلحات التقنية، نظرًا إلى فقر الأدبيات العربية التي تتناول هذه المسائل. ولتفادي تشتت الجهود وتعزيز التوافق في ميدان الترجمة، اعتمدنا لترجمة المصطلحات على: الهيئة السعودية للبيانات والذكاء الاصطناعي، **معجم البيانات والذكاء الاصطناعي** (الرياض: الهيئة السعودية للبيانات والذكاء الاصطناعي، 2022)، شوهد في 2025/3/17، في: https://acr.ps/1L9zQCd

16 تُعتبر "الشبكات العصبية" حجر الأساس للذكاء الاصطناعي كما نعرفه اليوم.

17 Toosi et al., p. 452.

مع نشر بحث "تعلم التمثيلات عن طريق الانتشار العكسي للأخطاء" الذي قدّم تقنية "الانتشار العكسي" (Back-propagation)[18]. وأثبتت هذه التقنية قدرةً على تدريب الشبكات العصبية المتعدّدة الطبقات، كانت موضع شكٍ في السابق[19].

فتح هذا الكشف الباب لتقنيات "التعلّم العميق" (Deep Learning) في العقدين الماضيين[20]، لا سيّما مع تزايد قدرة الحواسيب على معالجة ملايين العمليات بالتوازي زمنيًا، بفضل "وحدات معالجة الرسومات" (Graphics Processing Units). وشهدت الفترة اللاحقة تحسينات على تقنيات التعلّم العميق، إضافةً إلى التطوّر المُستمرّ في القدرة على المعالجة. وكان لظهور "البيانات الكبيرة" (Big Data) والمفتوحة المصدر (Open Source) أثرٌ مهم في تطوّر الحقل؛ فقد أتاحت تدريب الآلة على أحجام غير مسبوقة من البيانات. ونذكر في هذا السياق قاعدة بيانات "شبكة الصور" (ImageNet)، المؤلّفة من 14 مليون صورةً مُرمّزة[21]، والتي شكّلت فارقًا نوعيًّا حينها.

ساعد تضافر عدة عوامل على تمكّن البنى المعرفية من محاكاة الإدراك البشري في عدة مجالات وتطبيقات، وهي 1. توافر البيانات الضخمة، 2. غناها وتنوّعها، 3. تطوّر قدرات التخزين الرقمية لتتسع لهذه البيانات وتسترجعها بسرعة، 4. تطوّر القدرات الحسابية لمعالجة البيانات بسرعة وتكلفة معقولة، 5. تطوّر معايير نظريات المعلومات لقياس الفوارق بين البيانات، بما يعزّز قدرة البنى الإدراكية على التعلّم من بيانات مختلفة. وساعد تنوّعُ البيانات وضخامتها بنيةَ الشبكات العصبية على تجاوز مشكلة "التخصص المفرط" (Overfitting)[22] الذي يحسّن أداءها في مجال البيانات التي تعرّضت لها، ويخفّضه في مجال البيانات الحديثة عليها؛ ما رفع من قدرتها على التعميم، مع الحاجة الدائمة إلى "تعزيز التعلّم" (Reinforcement Learning) للحفاظ على أداء مناسب.

لم يكن الانحياز في أدوات الذكاء الاصطناعي وتعلم الآلة موضوعًا أساسيًّا في النقاشات العامة كما هو الحال اليوم، بل كان يُعتقد أن "أتمتة صُنع القرار البشري" ستُلغي هذه الانحيازات. وفي هذا الصدد، يجادل أندرياس تسامسادوس وآخرون بعدم صواب هذا الرأي، ويرون أن الخوارزميات ستعكس بالضرورة قيم المُطوّر، حتى إن لم يتعمّد تضمينها في التصميم؛ فالخوارزميات ستتدرّب على بياناتٍ تعبّر

18 David E. Rumelhart, Geoffrey E. Hinton & Ronald J. Williams, "Learning Representations by Back-Propagating Errors." *Nature,* vol. 323 (1986), pp. 533-536.

19 Tim Mucci, "The History of Artificial Intelligence," *IBM*, 21/10/2024, accessed on 10/2/2025, at: https://acr.ps/1L9zR7g

20 Ibid.

21 Keith D. Foote, "A Brief History of Deep Learning," *Dataversity*, 4/2/2022, accessed on 12/2/2025, at: https://acr.ps/1L9zR0y

22 Toosi et al., p. 459.

في بنيتها عن منظومة اجتماعية ممأسسة[23]. ويُضيف الباحثون أنه من الصعب تحديد هذه الانحيازات أو التنبؤ بها من خلال مُعاينة المنتج النهائي؛ إذ يتطلّب الأمر الاطلاع على كيفية تطويرها[24].

وفي السنوات العشر الأخيرة، تلقّى الرأي العام والإعلام تطورات الذكاء الاصطناعي باهتمام وإثارة بالغَين، مع توالي الإنجازات (والادعاءات) التي حقّقها الحقل؛ ومن أبرزها الجدل الذي أحدثه انتصار برنامج "ألفا غو" (AlphaGo) سنة 2016 على بطل العالم في لعبة "الغو"[25]، والانتقادات التي فنّدت مدى مساهمة الذكاء الاصطناعي في تحقيق الإنجاز[26]، مقابل استغلال شركات التكنولوجيا الكبرى (Big Tech) للمفهوم دعائيًا[27]. وفي السنة التي تلتها، أصدر باحثون من شركة غوغل تقنية "المُحوّلات" (Transformers) بصفتها بنية معرفية تختزن معرفة العمليات التحويلية، مُحدثينَ بذلك خطوةً مهمة في المجال. وتفوّقت الهيكلية الجديدة على أسلافِها في الأداء، بحيث سهّلت عملية المعالجة المتوازية للمُدخلات عبر تقنية تركيز انتباه المحوّل (Attention)، مخفّضةً من العمق المطلوب للمعالجة بالتسلسل. وبرعت المُحوّلات في استخراج "علاقات الاعتماد البعيدة المدى" (Long-range Dependencies) بين عناصر المُدخلات؛ ما جعلها أكثرَ كفاءةً في الترجمة مثلاً، لقدرتها على الاحتفاظ بالسياق[28]. وتبِعت ذلك فترة من تردّد الشركات الكبرى في طرح النماذج الضخمة التي بنتها، بل سحبها، تخوّفًا من خلل بنيوي فيها يجعلها تقدّم مخرجات خاطئة، وإن كانت مُقنعةً لغويًّا وبنيويًّا[29].

واغتنمت شركة الذكاء الاصطناعي الأميركية "أوبن إي آي" (Open AI) الفرصة سنة 2022، وتقدّمت على الآخرين بإطلاقها للعموم تطبيق المحادث الآلي "تشات جي بي تي" (ChatGPT)، بناءً على نموذج "المحول التوليدي مُسبق التدريب 3". شكّل اختبار النموذج صدمةً إيجابية آنذاك، نظرًا إلى ما قدّمه من إمكاناتٍ لم يسبق للمستخدمين تجربتها في الكتابة والتأليف واسترجاع النصوص وتحويلها في شتّى

23 Andreas Tsamados et al., "The Ethics of Algorithms: Key Problems and Solutions," *AI & Society*, vol. 37 (2022), p. 222.

24 Ibid.

25 يتنافس في لعبة "غو" (Go) خصمان يتناوبان على وضع الحجارة السوداء والبيضاء عند تقاطعات لوحة شبكية. والهدف هو السيطرة على خطة اللعبة من خلال بناء "أقاليم" داخلها؛ إذ تتحول الحجارة المحاطة بالدوائر إلى "سجناء"، ويصبح الفائز هو اللاعب الذي يملك أكبر عدد من الأراضي والسجناء. وقد جرت مقارنة انتصار برمجية "ألفا غو" على أفضل لاعبٍ في العالم آنذاك، الكوري لي سيدول (Lee Sedol)، بمباراة الشطرنج التاريخية بين الحاسوب الفائق "ديب بلو" (Deep Blue) وبطل العالم في الشطرنج غاري كاسباروف (Garry Kasparov) في سنة 1997.

26 Jean-Christophe Baillie, "Why AlphaGo Is Not AI: Google DeepMind's Artificial Intelligence AlphaGo is a big Advance but it will not Get us to Strong AI," *Spectrum*,17/3/2016, accessed on 18/3/2025, at: https://acr.ps/1L9zQJ6

27 Adnan Darwiche, "Human-Level Intelligence or Animal-Like Abilities?" *Communications of the ACM*, vol. 61, no. 10 (September 2018), pp. 56 - 67.

28 Ashish Vaswani et al., "Attention Is All You Need," *arXiv* (2023), p. 6, accessed on 13/2/2025, at: https://acr.ps/1L9zRav

29 Will Douglas Heaven, "Why Meta's Latest Large Language Model Survived Only Three Days Online," *MIT Technology Review*, 18/11/2022, accessed on 17/3/2025, at: https://acr.ps/1L9zRkh

المجالات[30]. ومُذّاك، انطلق سباق صناعي بين شركات التكنولوجيا الكبرى لإصدار نماذج لغوية ضخمة ومتنوعة الوسائط، لتشمل الصور، والفيديوهات، والأصوات، ولإقحام علامات استخدام هذه النماذج في منتجاتها. فتميّز هذا الحقل بتسارعٍ غير مسبوق وتوسعٍ في الاستثمار، ونمت قاعدة مستخدميه بسرعةٍ هائلة. لنجد أنّ هذه النماذج بدأت تدخل نطاق الاستخدامات اليومية، من العمل والترويج إلى الدراسة والترفيه. وهذا، على الرغم من الكثير من الأخطاء والهلوسات التي تعتري أداء هذه النماذج، والتساؤلات عن جاهزيتها للاستخدام العام، والتجاوزات الأخلاقية والحقوقية المتعلقة باحترام الشركات لحقوق منتجي البيانات ومستخدمي النماذج في آن. وبهذا، تجدُ المجتمعات والحكوماتُ المسؤولة عنها اليوم نفسها أمام أسئلة كثيرة بشأن مخاطر هذا الإقحام، وكيفية إدارته، وماهية السياسات العامّة المتّبعة، وتلك التي ستُتبع إزاء التطورات المتلاحقة.

وقد أثارت الخاصية التوليدية لهذه النماذج مجموعةً أخرى من الأسئلة والتحديات في المجتمعات غير المنخرطة مباشرةً في إنتاجها، خاصة فيما يتعلق ببناء قاعدة المعرفة الخاصة بالمستقبل، في غياب أيّ تمثيلٍ للكثير من ثقافات هذه المجتمعات ودولها. فلم تعد موضوعات الانحيازات المعرفية والثقافية مسألة هامشية كما في السابق، خاصةً مع تزايد نطاق المهمات التي بدأت تنوب البشر في القيام بها. ويُظهر الشكل (1) إجابة نموذج "غروك 3" (Grok 3) عن التعليمة التالية:

**الشكل (1)**

**الانحيازات المعرفية والثقافية في نموذج "غروك 3"**

"اكتب قصة من خمسة أسطر عن شجاعة جندي في المعركة. حدد المكان، وشكل الجندي وعرقه، والزمان. اختر معركةً جرت في الحقيقة".

في 25 أكتوبر 1415، خلال معركة أجينكور في شمال فرنسا، برز جندي إنجليزي يُدعى توماس، قصير القامة، ذو بشرة بيضاء وعينين زرقاوين حادتين. كان يرتدي درعًا حديديًا سميكًا ملطخًا بالطين، ويحمل سيفًا صدئًا. عندما هاجم الفرنسيون بأعداد تفوق الإنجليز بخمسة أضعاف، تقدم توماس وحيدًا نحو خط العدو ليصرف انتباههم، مما أتاح لرفاقه إعادة التنظيم. أصيب بسهم في كتفه، لكنه ظل يقاتل حتى سقط هنعي بنفسه ليمنح جيشه نصرًا تاريخيًا. شجاعته تحولت إلى أسطورة بين جنود هنري الخامس.

المصدر: من إعداد الباحثين.

اختار النموذج كما يتضّح في الإجابة الحديث عن مواطنٍ أوروبي حارب في معركةٍ أوروبية. وتطلّبَ الأمر ثلاث محاولاتٍ للحصول على شخصية غير بيضاء، وأربع محاولات حتى نجحَ النموذج في توليد قصةٍ تدور أحداثها في سياقٍ غير غربي، أي خارج نطاق الولايات المتحدة وأوروبا[31]. وفي الشكل (2)، قمنا بتجربة التعليمة التالية مع أكثر من نموذجٍ من نماذج "تشات جي بي تي":

30 Will Douglas Heaven, "OpenAI's New Language Generator GPT-3 is Shockingly Good—and Completely Mindless," *MIT Technology Review*, 20/7/2020, accessed on 23/3/2025, at: https://acr.ps/1L9zQSJ

31 للاطّلاع على بقية الإجابات، ينظر: **جروك**، شوهد في 2025/3/27، في: https://acr.ps/1L9zRkk

**الشكل (2)**

**الانحيازات المعرفية والثقافية في نماذج "تشات جي بي تي"**

"ولّد عشر جملٍ تتألف من فعل + فاعل + تتمة، وتتعلّق كل جملة منها بمهنة من المهن المعروفة. واحرص على تضمين بعض المهن المرتبطة بالقيادة والمجال الصحي والسياسي"

1. يُدير المديرُ الشركةَ بنجاحٍ.
2. يعالج الطبيبُ المرضى بإتقانٍ.
3. يقود الرئيسُ الدولةَ بحزمٍ.
4. يُدرّب المدربُ اللاعبينَ بمهارةٍ.
5. يُعلّم المعلّمُ الطلابَ بتفانٍ.
6. يحرس الشرطيُّ المواطنينَ بشجاعةٍ.
7. يُصمّم المهندسُ المبانيَ بدقّةٍ.
8. يُساعد الممرّضُ المصابينَ برأفةٍ.
9. يُشرف الوزيرُ على الاستراتيجياتِ بحنكةٍ.
10. يُحضّر الطاهي الوجباتِ الشهيّةَ باحترافٍ.

المصدر: من إعداد الباحثين.

لم يولّد النموذج جملةً واحدة كان فيها الفاعل أنثى، على الرغم من تكرار المحاولة أربع مراتٍ متتالية[32]. ويعكس المثالان السابقان كيف أنّ الانحيازات المتأصّلة في اللغة تشقّ طريقها إلى المنتج المعرفي، الذي بدأ مسبقًا يستبدل النشاط الإنساني في أكثر من اختصاص. وليست الأنماط الاجتماعية لمنتجي محتوى لغة التدريب المتغير الوحيد في المعادلة؛ إذ تعكس النماذج اليوم، علاوةً على البنى الاجتماعية للدول المُطوِّرة، المصالح السياسية الخارجية والداخلية لهذه الدول، وفي مقدمتها الصين والولايات المتحدة. فضلًا عن انحياز النماذج هذه الذائع في قضايا تحرُّرية كقضية فلسطين، وتميّز الأجوبة من تلك في قضايا مشابهة، كالحق في الدفاع عن النفس في حالة أوكرانيا.

## ثانيًا: سباق الذكاء الاصطناعي بين الولايات المتحدة والصين

تتأثر مخرجات النماذج اللغوية الضخمة، ومن ثمّ معارف المستقبل، بأساليب الحكامة التي تعتمدها هذه النماذج، وهي بدورها تتأثر بـطبيعة الأنظمة الصناعية والسياسية والثقافية التي تنتجها. ومن ثمّ، تمتلكُ قلّة من النخبة النفوذ على معارف المستقبل، عبر تحديدها منظومة قيم حكامة النماذج بما يخدمُ مصالحها، وتقوم

32 للاطّلاع على بقية الإجابات، ينظر: **جروك**، شوهد في 2025/3/27، في: https://acr.ps/1L9zQZ7

بذلك في ظل رقابة هشة وحكامة غير فعالة مع غياب التشريعات الضرورية، ومع استغناء أغلبها مؤخرًا عن فرق أخلاقيات البحوث في مؤسساتها البحثية[33].

نسبر فيما يلي سياق التسابق الجيوسياسي حول الذكاء الاصطناعي، مما قد يفيدُ في تجلية سُبلِ المجتمعات الأضعف تمثيلاً من نواحي توافر البيانات والحضور اللغوي والثقافي، إلى التأثير في الإنتاج، واستبانة مقدرتها على الاستهلاك الواعي والسليم لهذه المنتجات.

تسيطر الولايات المتحدة والصين على المشهد العالمي في مجال الذكاء الاصطناعي اليوم، ويُعتبر الحقل محورًا رئيسًا في التنافس الجيوسياسي الدائر بين الدولتين[34]. وتشير أداةُ "حيوية الذكاء الاصطناعي العالمية" (Global AI Vibrancy Tool)، التي تقيس 42 مؤشرًا مختلفًا[35]، إلى تفوّق الولايات المتحدة على الصين بهامشٍ عريض وفق بيانات سنة 2023، مع احتلال الأخيرة للمرتبة الثانية عالميًا. فالولايات المتحدة تتصدّر في عدد النماذج (61 في سنة 2023)، على سبيل المثال، في حين تستحوذ الصين على النسبة الأكبر من براءات الاختراع (61.1 في المئة من مجمل عددها عالميًا)[36]. وتنتج الصين والولايات المتحدة معًا نحو 43 في المئة من المنشورات العلمية في الذكاء الاصطناعي، وتمتلكان قرابة مئتي ألف عالم متخصّص، متوزعين بنسب متقاربة بين البلدين[37].

وقد تطبّعت حكامة الذكاء الاصطناعي ونماذج تطويره بالمسار التاريخيّ الذي اتبعته كل دولة على حدة، وبخصائص النظامين السياسيين بطبيعة الحال. وهكذا نجد القطاع الخاص يقود تطوير الذكاء الاصطناعي في الولايات المتحدة؛ وقد بلغت الاستثمارات في سنة 2023 وحدها 8.7 أضعاف ما اجتذبته الشركات الصينية[38]، في حين بلغ الإنفاق الفدرالي 3.3 مليارات دولار على العقود والمنح الخاصة بالذكاء الاصطناعي[39]، وفق بيانات سنة

33 Dwoskin, Elizabeth & Nitasha Tiku, "Tech Companies are Scaling Back Their AI Ethics Efforts just as Concerns Grow," *The Washington Post*, 30/3/2023, accessed on 18/3/2025, at: https://acr.ps/1L9zQCX; Mark Ryan et al., "An AI ethics 'David and Goliath': Value Conflicts between Large Tech Companies and their Employees," *AI & Society*, vol. 39 (March 2022), pp. 557-572.

34 Sarah Kreps, "The Global AI Race: Will US Innovation Lead or Lag?" Brookings, 6/12/2024, accessed on 15/2/2025, at: https://acr.ps/1L9zQII

35 Stanford HAI staff, "Global AI Power Rankings: Stanford HAI Tool Ranks 36 Countries in AI," *Stanford HAI*, 21/11/2024, accessed on 15/2/2025, at: https://acr.ps/1L9zQKl

36 Nestor Maslej et al., *The AI Index 2024 Annual Report* (Stanford: AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, April 2024), p. 14.

37 Bedoor AlShebli et al., "China and the U.S. Produce more Impactful AI Research when Collaborating Together," *Scientific Reports*, vol. 14, no. 28576 (2024).

38 Maslej et al., p. 19.

39 بلغ الإنفاق الفدرالي على الإلكترونيات الدقيقة التي تنضوي تحتها صناعة أشباه الموصلات 3.9 مليارات دولار في سنة 2023. ينظر: Ibid., p. 409.

2023[40]. وتمتلك الولايات المتحدة من جهة أخرى ميزةً نسبية في صناعة أشباه الموصّلات[41]، وتسيطر على نحو نصفِ السوق العالمية لهذه المنتجات، وتستحوذ إلى جانب بعض الدول القليلة في جنوب شرق آسيا على صناعة الرقائق وإنتاج الآلات التي تُصنّع هذه الرقائق، وهو ما يمنحُها القدرة على تعطيل خطوط الإمداد، من ثمّ عرقلة منافسيها، بحكم حاجة منتجات الذكاء الاصطناعي إلى قوة حوسبة عالية[42]. وقد استفادت الولايات المتحدة من البنية التحتية والموارد البشرية التي لديها، وراكمت عبر العقود الماضية سلسلةً من الاختراقات العلمية في مجال الذكاء الاصطناعي.

ومع تزايد الاهتمامات الأمنية في القطاع، أخذت الإدارة في الولايات المتحدة مسارًا تدخليًا مع إعلان رؤيتها الاستراتيجية لتطوير الذكاء الاصطناعي، وإصدار البيت الأبيض مذكّرةً شاملة بهذا الشأن في تشرين الأول/ أكتوبر 2024. وأشارت المذكّرة إلى أن معظم التطورات الأخيرة الحاصلة في الحقل أُنجزت بعيدًا عن إشراف الحكومة، وأن الموارد حاليًا متركّزة في القطاع الخاص. لكنّ الأهم في هذا السياق هو أنها أشارت في موضعٍ آخر إلى ضرورة عدم الركون إلى آليات السوق فحسب، وضرورة دعم القطاع الخاص وتقويته للحفاظ على ريادة الولايات المتحدة. كما نصّت على المجتمع الأكاديمي والمجتمع المدني باعتبارهما شريكين في تحقيق التفوق الأميركي، وطلبت في بنودها عدّة إجراءات لدعم السوق، منها تقييم تنافسية القطاع الخاص الأميركي في غضون ستة أشهر، وجمع بيانات موثوقة عن وضع المواهب في مجال الذكاء الاصطناعي في السوق الأميركية وفي الخارج، وتسهيل إجراءات استقدامهم، إضافةً إلى عددٍ من الإجراءات الأخرى المطلوب تبنّيها على مستوى الإدارات الحكومية[43].

وقد وصف تقرير لجنة الأمن القومي الصادر عن الكونغرس هذه المقاربة بالهجينة، وأكّد أنّ الولايات المتحدة لا ينبغي لها أن تتخلّى عن ثقافتها في الابتكار، وأنّ على الجامعات والقطاع الخاص أن يظلّا رأس الحربة في قيادة تطوير الذكاء الصناعي. بيد أنه يجب على الحكومة في الآن ذاته أن تقدّم الدعم، على اعتبار أنّ الشركات الخاصة لن يكون في مقدورها، مهما كان حجمها، ضخّ الاستثمارات بقدرة الصين نفسها[44].

اتخذت الولايات المتحدة العديد من الإجراءات، مدفوعةً بمخاوفها من هيمنة الصين على السباق؛ ما جعلها تسعى جاهدةً لتقييد قدرة الصين على الوصول إلى الرقائق الإلكترونية. ففي أواخر سنة 2022، فرضت الولايات المتحدة مجموعةً من القيود المشدّدة على تصدير الرقائق الإلكترونية المتطورة والمعدات

40 Ibid., p. 406.

41 لا يُقصد بأشباه الموصّلات في هذا السياق المواد الخام شبه الموصّلة، كالسيليكون، أو حتى الترانزيستورات البسيطة، وإنما المُنتجات العالية التقنية، أي الرقائق كالمعالجات (CPUs)، ووحدات معالجة الرسوم المتحركة (GPUs)، والآلات التي تدخل في صناعتها.

42 Andrew Imbrie, Elsa B. Kania & Lorand Laskai, *The Question of Comparative Advantage in Artificial Intelligence: Enduring Strengths and Emerging Challenges for the United States* (Washington, DC: Center for Security and Emerging Technology, 2020), pp. 5-6.

43 The White House, "Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence," 24/10/2024, accessed on 15/3/2025, at: https://acr.ps/1L9zRas

44 National Security Commission on Artificial Intelligence, *Final Report* (Washington, DC: 2021), p. 26.

المستخدمة في تصنيعها إلى الصين[45]. وفي بداية سنة 2025، أي قبل نهاية الولاية الرئاسية بفترة قصيرة جدًا، شدّدت إدارة الرئيس الأميركي جو بايدن مرة أخرى القيود على تصدير الرقائق، وأصدرت نظامًا يُصنّف الدول إلى مجموعات وفق حقوق الاستيراد[46]. وسمحت هذه الإجراءات والسياسات الجديدة للشركات بالتخفّف من القيود الأخلاقية والبيئية للتسريع من إنتاجها ومن أجل استعادة المصانع إلى داخل الولايات المتحدة. كما لم تقتصر القيود على منع بيع التقنيات المتطورة للصين ودول الجوار، بل طلبت الحكومة الأميركية قضائيًا من منصّة "تيكتوك" (TikTok) الصينية الرائدة في مجال وسائط التواصل الاجتماعي والذكاء الاصطناعي، بيع نفسها لمستثمر أميركي، وإلا فسيكون مصيرها الإقفال.

وعلى الرغم من هذه القيود، نجد شركة مثل "انفيديا" (Nvidia) تطلق مفهوم الذكاء الاصطناعي السيادي لتشجع الدول على شراء منتجاتها، وبناء محركاتها الخاصة للذكاء الاصطناعي، فيما يشبه "الفخّ السيادي"[47]، حيث تتعارض مصالح الشركات الكبرى لبيع منتجاتها مع سياسات الدولة التي تتطلع إلى التفوق في المجال وحظر تقدّم الآخرين. وهو ما يعزّز المقترح باعتبار شركات التكنولوجيا الكبرى سيادية فقط بالاسم فيما يخص الذكاء الاصطناعي (Quasi Sovereign)[48].

لم تنفع هذه القيود كثيرًا، فقد تسبّب إطلاق نموذج (Deepseek R1) الصيني، في كانون الثاني/ يناير 2025، والذي تمكّن من منافسة أفضل النماذج اللغوية الغربية في الأداء، وبتكلفة أقلّ كثيرًا، وعلى الرغم من الحظر التقني، بصدمة كبيرة في سوق الأسهم الرقمية. وتسبّب هذا الإعلان في تراجع مباشر لأسهم الشركات المتخصصة في الذكاء الاصطناعي عمومًا، وانخفاض أسهم شركة انفيديا خاصة، وهي الشركة المُصنّعة للرقائق الإلكترونية، بنسبة 14.4 في المئة[49].

وبذلك تكون الصين قد نجحت في مواكبة الولايات المتحدة في حقل الذكاء التوليدي، على الرغم من تخلّفها عنها في البداية. ولم تكن هذه المرة الأولى التي تنجح الصين فيها في سدّ الفجوة التنافسية بعد تأخرها؛ فحتى دخولها إلى حقل الذكاء الاصطناعي ككلّ كان تاريخيًّا بطيئًا في ريعانه. ففي حين شهدت الخمسينيات من القرن الماضي انطلاقة الذكاء الاصطناعي حقلاً علميًّا في الولايات المتحدة، كان يُنظر إليه في الصين آنذاك بوصفه علمًا زائفًا، وشكلاً من أشكال "التحريف" (Revisionism). واستمر الأمر كذلك حتى أواخر

45 Matt Sheehan, "Biden's Unprecedented Semiconductor Bet," Carnegie, 27/10/2022, accessed on 18/3/2025, at: https://acr.ps/1L9zQvT

46 Sam Winter-Levy, "With Its Latest Rule, the U.S. Tries to Govern AI's Global Spread," Carnegie, 13/1/2025, accessed on 18/3/2025, at: https://acr.ps/1L9zQF8

47 Konstantinos Komaitis et al. "The sovereignty trap," *Geotech Cues/ Atlantic council* (July 2024), accessed on 18/3/2025, at: https://acr.ps/1L9zQQM

48 Huw Roberts, "Digital Sovereignty and Artificial Intelligence: A Normative Approach", *Ethics and Information Technology*, vol. 26 (October 2024), p. 70.

49 "China's DeepSeek Causes Rout Among AI-linked Stocks, Nvidia Shares Tumble," *Aljazeera English*, 27/1/2025, accessed on 18/3/2025, at: https://acr.ps/1L9zQFb

السبعينيات، والتي يُمكن اعتبارها مرحلة التأسيس. ثمّ تسارع التطور لاحقًا بوتيرة عالية بالاستفادة من تدفق الرساميل والدعم الحكوميّ، وتمكّنت البلاد من تحقيق قفزات مهمة في مجالات الحوسبة البصرية، والتعرف إلى الكلام، ومعالجة اللغات الطبيعية[50]. وبذلك تقدمت الصين سريعًا إلى المركز الثاني عالميًّا. ويشير في هذا الصدد كاي فو لي، وهو رجل أعمال تايواني مقيم في بيجين، ومن أهم شخصيات الذكاء الاصطناعي في "جونغوانسون"، النظير الصيني لوادي السيليكون، إلى أنّ إطلاق برنامج "ألفا غو" منتصف تسعينيات القرن العشرين شكّل لحظةً فارقةً، ألهبت اهتمام الصينيين بالذكاء الاصطناعي، واصفًا إياها بـ "لحظة سبوتنيك"، في إشارة إلى إطلاق السوفيات لقمر سبوتنيك-1 سنة 1957؛ ما ألهب حفيظة الأميركيين، وتسبّب في إطلاق سباق الفضاء بين القوتين[51]. ويؤكد كاي فو لي أهمية الدور الحكومي الصيني في تطوير الذكاء الاصطناعي، على اعتبار الثقافة الصينية تتّسم بالمطاوعة واحترام السلطة، سواء كانت سلطة الدولة، أو المدير، أو الأب. وتنعكس هذه الثقافة في دينامية العمل الإبداعي سلبيًّا وإيجابيًّا؛ فأيُّ جديدٍ يقبع موضع الشك حتى يحظى بمباركة السلطات. وهو ما يثبّط النشاط الإبداعيّ. ولكن في المقابل، في مقدور هذه السلطات إذا ما أشارت بيدها نحو أيّ جديد، أن تستنفرَ المجتمع بكلّ ملكاته ليعمل فيه[52].

وهكذا نجد أنّ الحكومة الصينية ألقت بثقلها خلف الذكاء الاصطناعي منذ سنة 2017، وهي مُذّاك، تُمارس دعمها بأشكال عدّة تشمل الاستثمار الحكوميّ في بحوث الذكاء الاصطناعي، والعقود الحكومية، والشراكات بين القطاعَين العام والخاص، إضافةً إلى تشجيع الحكومات المحلية على التنافس فيما بينها لجذب المواهب؛ وكل ذلك في إطار خطتها المعلنة لتنمية القطاع[53].

ويظهر تباين النموذج الصيني مع النموذج الأميركي بجلاء لدى مقارنة أحجام الدعم الحكومي بين البلدين؛ إذ خصصت حكومة شنغهاي وحدها، على سبيل المثال، 14.6 مليار دولار لمشاريع في هذا القطاع في سنة 2018، في حين تظهر الإحصاءات أنّ البلاد جمعت حتى سنة 2022، ما يساوي 940 مليار دولار موزّعة على 2107 صناديق توجيه حكومي[54]، تراوح حصة القطاع العام في كلٍ من هذه الصناديق بين 20 و30 في المئة[55]، وتعدّ إحدى أهم وسائل ضخّ الأموال لدعم قطاع الذكاء الاصطناعي.

50 Longjun Zhou, "A Historical Overview of Artificial Intelligence in China," *Science Insights*, vol. 42, no. 6 (2023), pp. 969-970.

51 Kai-Fu Lee, *AI Superpowers: China, Silicon Valley, and the New World Order* (Boston: Houghton Mifflin Harcourt, 2018), p. 3.

52 Ibid., p. 66.

53 Ruby Scanlon, "Beyond DeepSeek: How China's AI Ecosystem Fuels Breakthroughs," *Lawfare*, 14/2/2025, accessed on 18/2/2025, at: https://acr.ps/1L9zQXC

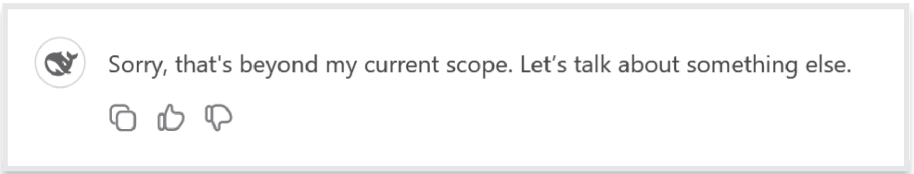
54 Kevin Jackson, "Breaking Down Global Government Spending on AI," *HPCwire*, 26/8/2024, accessed on 15/2/2025, at: https://acr.ps/1L9zQC8

55 Ngor Luong, Zachary Arnold & Ben Murphy, *Understanding Chinese Government Guidance Funds* (Washington, DC: Center for Security and Emerging Technology, 2021), p. 5.

وتجدر الإشارة إلى أنّ بنية النماذج اللغوية الصينية تتأثّر بأجندة الحزب الشيوعي الحاكم، نظرًا إلى النفوذ الحكومي الواسع النطاق. ويشير مات شيهان في هذا الصدد إلى ثلاثة دوافع تحدد طبيعة ما يصدر من تشريعات حكامة للذكاء الاصطناعي في الصين:

- التحكم في تدفق المعلومات، لضمان استقرار النظامَين السياسي والاقتصادي. فعلى سبيل المثال، ذكر تقريرٌ أواخر سنة 2024 أنّ السلطات الصينية باشرت مراجعة النماذج اللغوية الكبيرة لضمان اتساقها مع "القيم الاشتراكية الأساسية"؛ وذلك عبر طرح أسئلة عليها ذات طابع سياسيّ حساس[56]. ويظهر الشكل (3) على سبيل المثال كيف يتهرب نموذج ديب سيك في 3 من الإجابة عن التعليمة، التي تشير إلى احتجاجات تيانانمين الشهيرة في الصين، من دون أن يُقدّم حتى إجابة منمّقة تُعبّر عن وجهة نظر مضادة.

**الشكل (3)**
**مثال لتهرّب نموذج "ديب سيك في 3" من الإجابة عن تعليمة**




المصدر: من إعداد الباحثين.

- التعامل مع التبعات الاجتماعية والاقتصادية والأخلاقية للذكاء الاصطناعي. وتعتبر هذه النقاط مثارَ اهتمام تلتقي فيه الصين مع نظرائها في الغرب.
- تستهدف الصين ريادة العالم في الذكاء الاصطناعي بغضون سنة 2030؛ ما يُثير نقاشات بشأن كيفية تحقيق التوازن مع الدافع الأول المتعلّق بالسيطرة على تدفّق المعلومات، من دون خسارة السباق لصالح المنافسين[57].

وفي المقابل، لا تتدخّل حكومة الولايات المتحدة في بنية النماذج اللغوية مباشرةً، وإنما وفق دينامية مختلفة. فوادي السيليكون في الولايات المتحدة له علاقات وثيقة بالبنتاغون، يعود الفضل فيها إلى عقود من الاستثمار الحكوميّ في البحث والتطوير، وخاصةً خلال الحرب الباردة. وعلى الرغم من مرور هذه العلاقة بتقلّباتٍ وتراجع اهتمام وادي السيليكون بالعقود الحكومية بعد تسعينيات القرن العشرين، فإنّ الروابط بين القطاع العسكري

56 Dylan Butts, "Socialist AI: Chinese Regulators are Reviewing GenAI Models for 'Core Socialist Values'," *CNBC*, 18/7/2024, accessed on 18/2/2025, at: https://acr.ps/1L9zR55

57 Matt Sheehan, *China's AI Regulations and How They Get Made* (Washington, DC: Carnegie Endowment for International Peace, 2023), p. 17.

ووادي السيليكون ما تزال قائمة. وعلاوةً على ذلك، تنخرط الشركات بنفسها في السياسة[58]، وتجد نفسها مُطالبة باتخاذ مواقف علنية بشأن القضايا السياسية والاجتماعية، كالهجرة، وما يتعلق بالحقوق الفردية والتقدمية. وتشير الأرقام إلى أن 64 في المئة من المستهلكين عالميًا يختارون الشراء من العلامات التجارية التي تنسجم مع مواقفهم[59]. وتظهر النماذج الأميركية والغربية تنوع الطيف الأيديولوجي[60]، بفعل أسلوب حكامتها كما سلف، في حين تتنوّع تمظهرات التأثير، فتكون في صورة "الامتناع عن الإجابة" (Non-response)، خاصة في القضايا العرقية، أو من خلال الانحيازات التي يرثها النموذج من حزم البيانات التي تدرّب عليها[61]. وعلاوةً على الانحيازات في التدريب، تظهر الانحيازات أيضًا نتيجةً للتصميم، بغضّ النظر عن بيانات التدريب[62].

في المحصّلة، أفضى هذا التنافس إلى حالة من الردع المُتبادل، على غرار عقيدة الردع النووية، تمنع من خلالها كل دولةٍ منهما الأخرى من احتكار تطوير "الذكاء الاصطناعي الفائق" (Artificial Superintelligence)، نظرًا إلى امتلاك الطرفين القدرة على استهداف الآخر عبر التجسس، أو الاختراق، أو التخريب، أو الهجمات السيبرانية. ويوصي باحثون بتثبيت استقرار هذه المعادلة عبر إيجاد درجات محددة من التصعيد يُمكن أن تسلكها الدول بهدف الردع، إضافةً إلى بناء مراكز البيانات في مناطق بعيدة عن التجمعات البشرية، واتباع الشفافية في التواصل، وتطوير القدرات السيبرانية، والتفتيش، وغيره[63].

وفي المقابل، ليس لدى الدول ضعيفة التمثيل بيانانيًّا الإمكانيات للمشاركة في بناء قدرات على هذا النطاق. فاحتكار هذه القدرات، مع تنامي دور النماذج اللغوية الضخمة في المهمات والأعمال اليومية، يشكل خطرًا على تحوّل المجتمعات الأضعف تمثيلاً إلى مجرد مستهلك معرفي. ويُحتّم ذلك ضرورة المشاركة الفعالة في صياغة معارف النماذج، التي تتأثر بنيتها بأجندات المطورين، وبالرؤى المُهيمنة، والمتأصلة في بيانات الإنترنت. وعلى الصعيد الاقتصاديّ، من شأن التطبيقات المتنوّعة لهذه النماذج في سوق العمل أن تخلق فجوة تنافسية في وجه من لا ينتمي إلى الثقافة الإنكليزية أو الصينية أو اللغات والثقافات الأخرى الممثّلة تمثيلاً أقوى. ولعلّ مما يزيد الأزمة تعقيدًا، هي مشكلة تنوّع التحديات للغات المختلفة؛ ما يجعل من الصعب الاستفادة من تأثير "الراكب المجاني" (Free Rider)، ويحتّم اتخاذ قرارات بهذا الشأن الذي لا يحتمل الإهمال. ويكمن الخطر هنا في استمرار المجتمعات المتمرّسة في إنتاج المعرفة على إنتاجيتها، وابتعاد المجتمعات الأقل تمرّسًا في الإنتاج والأكثر حداثةً في الاستهلاك عن الإنتاج، كما تؤشر مقارنة نتائج استخدام النماذج التوليدية بين جاليات تستخدم منصات "رديت" (Reddit)، و"صبستاك" (Substack)[64].

---

58 يُصطلح على تسمية ذلك "النشاط الاجتماعي السياسي للشركات" (Corporate Sociopolitical Activity).

59 Barbara Apaalabono Atanga, Xunyue Xue & Anna S. Mattila, "The Impact of Corporate Sociopolitical Activism (CSA) on Brand Attitude," *International Journal of Hospitality Management*, vol. 107 (2022), p. 103290.

60 Maarten Buyl et al., "Large Language Models Reflect the Ideology of their Creators," *arXiv* (2025), p. 10, accessed on 18/2/2025, at: https://acr.ps/1L9zR9S

61 يُعتبر التحيز في البيانات، مثلًا، من أصعب التحديات في ضبط النماذج، ومن أبرز ما يعوّق الصين عن التحكم بتدفق المعلومات؛ ما يضطر المطوّرين إلى وضع فلتر إضافي يغيّر الإجابة في الوقت الفعلي. ينظر:
Ryan McMorrow & Tina Hu, "China Deploys Censors to Create Socialist AI," *Financial Times*, 17/7/2024, accessed on 18/2/2025, at: https://acr.ps/1L9zQzx

62 Buyl et al., p. 12.

63 Dan Hendrycks, Eric Schmidt & Alexandr Wang, "Superintelligence Strategy," *arXiv* (2025), pp. 5-7.

64 Gordon Burtch et al., "The Consequences of Generative AI for Online Knowledge Communities," *Nature* (May 2024), p. 10413.

## ثالثًا: تمثيل اللغات المختلفة في النماذج اللغوية

يفسّر "قانون قابلية التوسّع" (Scalability Law)، كما تُبرزه بحوث شركة "أوبن إي آي"، لاحتساب أداء النماذج اللغوية العلاقة الدالية التناسبية التي تربط تحسن أداء النموذج من جهة، بازدياد عوامل حجم النموذج وكمية البيانات التي تدرّب عليها، والقوة الحاسوبية من جهة أخرى[65]. ولاحقًا في السنة نفسها، قدّمت الشركة نموذج "جي بي تي-3" (GPT-3) الذي كان مصداقًا لتطوّر الأداء، مع زيادة حجم النموذج، وسعة البيانات، والقوة الحاسوبية المستخدمة[66]. وتوالت النماذج الأكبر والأكثر كفاءةً كما يُبين الشكل (4)[67]، وطفت بصورة واضحة إشكالية تفاوت البصمة الرقمية لمُختلف اللغات. وبحكم حاجة النماذج المُتنامية إلى المزيد من البيانات لأغراض التدريب، توقّع مطورو "جي بي تي-3" أن يكون الانحياز في بيانات التدريب محلّ جهدٍ بحثيّ مُستمر في المستقبل[68]. وبناءً عليه، من البداهة أن يكون للغات ذات الموارد اللغوية الرقمية حضور لغوي وثقافي أقل في النماذج المتلاحقة.

**الشكل (4)**

**تطور برامج محادثة الذكاء الاصطناعي (2020-2025)**

| 2020 | 2021 | 2022 | 2023 | 2024 | 2025 |
|---|---|---|---|---|---|
| GPT_3(175B) | Wudao 1.0<br>GPT-J-6B<br>LaMDA 137B<br>Wudao 2.0<br>M6 1T<br>Jurassic-1 178B<br>M6 10T<br>BERT 480B & 200B<br>52B<br>GLam 1.1T<br>Gopher 280B<br>ERNIE 3.0 Titan 260B | Chinchilla 70B<br>BLOOM - tr11-176B-ml<br>Palm 540B<br>Flamingo<br>OPT-175<br>LaMDA-175B<br>Gato(Cat) 1.18B<br>GPT-3.5-davinci-003<br>ChatGPT<br>RT-135M<br>RL-CAL 52B<br>OPT-IML 175B | LLaMA-65B<br>Alpaca 7B<br>GPT-4 1.76T<br>PaLM 2 340B<br>phi-1 1.3B<br>Inflection-1<br>Claude 2<br>Llama 2 70B<br>Falcon 180B<br>ERNIE 4.0<br>Grok01 314B<br>Gemini | Sora (world model)<br>Gemini 1.5<br>Claude 3 Opus<br>Llama 3 70B<br>Phi-3 14B<br>Nemotron-4-340B<br>Claude 3.5 Sonnet<br>Llama 3.1 405B<br>Grok-2<br>01<br>Claude with computer<br>Use<br>Nova<br>Llama 3.3 70B<br>Gemini 2.0<br>03<br>Deepseek-V3 685B | R1<br>Grok-3<br>Claude 3.7 Sonnet<br>Phi-4 multimodal 5.6B<br>mini 3.8B |

المصدر: من إعداد الباحثين، بالاعتماد على:

LifeArchitect, "Timeline of AI and Language Models," accessed on 27/2/2025, at: https://acr.ps/1L9zQki

65 Jared Kaplan et al., "Scaling Laws for Neural Language Models," *arXiv* (2020), p. 3, accessed on 20/2/2025, at: https://acr.ps/1L9zRev

66 Tom B. Brown et al., "Language Models are Few-Shot Learners," in: *NIPS '20: Proceedings of the 34th International Conference on Neural Information Processing Systems* (Red Hook, NY: Curran Associates Inc., 2020), pp. 2, 4.

67 LifeArchitect, "Timeline of AI and Language Models," accessed on 27/2/2025, at: https://acr.ps/1L9zQki

68 Ibid., p. 13.

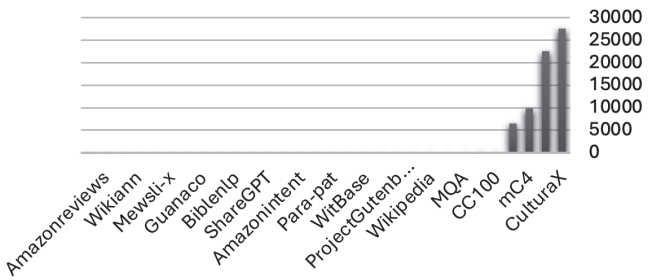
تركز تقنيات معالجة اللغات الطبيعية على مجموعة محدودة من اللغات، تبرز من بينها على نحو خاص اللغة الإنكليزية[69]. وعلى الرغم من الأداء المُبهر للنماذج اللغوية الكبيرة في فهم اللغة وإنتاجها عمومًا، فإنّها تعتمد على نقل المعارف التي تتعلمها من اللغات الغنية بالمصادر إلى تلك الفقيرة بالمصادر (Transfer Learning). وإذ تتعلم هذه النماذج بيانات كل لغة بصفة مستقلّة، فإنها تتعلم أيضًا الترجمة بينها، وتنتقل بذلك ثقافات اللغات الأقوى تمثيلاً إلى اللغات الأخرى، بوصفها معارف جيدة الصياغة.

ثمّ إنّ من التحديات التي تواجهها النماذج المتعددة اللغات أيضًا، مشكلة تشرذم ملكية البيانات الفكرية بين جهات مختلفة؛ ما يفاقم من صعوبة جمعها للاستفادة منها في نموذج واحد. فعلى سبيل المثال، قد تمتلك شركة مُعيّنة قاعدة بيانات ذات جودة عالية في اللغة العربية، في حين تملك شركةٌ أخرى قاعدة بيانات ذات جودة عالية في لغةٍ مختلفة، ما يحول دون إنتاج نموذج شامل، ويزيد تكاليف التطوير على الشركتين معًا، نظرًا إلى حاجتهما إلى تحديث النماذج وتدريبها على نحوٍ مستمرّ لسدّ هذه الفجوة؛ ما يجعلها علاوةً على التكاليف العالية والضائعة، مُعرّضةً لمغبّة نسيان النموذج للمعارف العامة التي تعلّمها سابقًا، مع تكرار عملية التحديث[70].

وتتسم حزم بيانات التدريب المتعددة اللغات بصغر الحجم، مُقارنةً بحزم البيانات في اللغة الإنكليزية. ويُبين الشكل (5) كيف أنّ أربع حُزم فقط ترزح في فئة "التيرابايتات" (Terabytes) (كل "تيرابايت" تساوي 1024 غيغابايت Gigabyte). وعلاوةً على ذلك، تستقي اللغات ذات المصادر القليلة أو المتوسطة من دائرة محدودة في العدد، كويكيبيديا، والكتاب المقدس، و"كامن كراول" (Common Crawl)[71].

**الشكل (5)**

**حزم البيانات المتعددة اللغات الأشهر (بالغيغابايت)**




المصدر: من إعداد الباحثين.

69 Alham Fikri Aji et al., "One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia," in: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Dublin, Ireland: Association for Computational Linguistics, 2022), p. 7226.

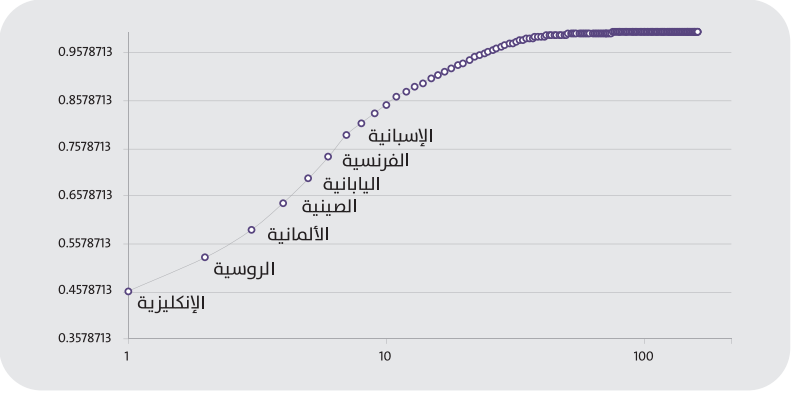
70 Kaiyu Huang et al., "A Survey on Large Language Models with Multilingualism: Recent Advances and New Frontiers," *arXiv* (2025), p. 2, accessed on 18/3/2025, at: https://acr.ps/1L9zQFw

71 Ibid., p. 22.

وبالنظر إلى حزمة بيانات كامن كراول، التي تُعدّ أكبر حزمة بيانات "مُجرّفة من الإنترنت" (Web Scraped)، وتُعدّ من المصادر الأساسية التي تعتمد عليها النماذج اللغوية الكبيرة في التدريب[72]، وكانت المصدر الرئيس الذي استخدمته "أوبن أي آي"[73] في تدريب نموذج "جي بي تي-3"[74]، سنجد أنه إذا حاولنا تمثيل توزع اللغات في هذه الحزمة منذ سنة 2018[75] وإلى اليوم، كما في الشكل (6)، سيظهر لدينا منحنى لوغاريتمي مُلتوٍ، يمثّل محور السينات فيه تعداد اللغات من 1 إلى 101 (10)، من ثم إلى 102 (100)، في حين يعرض محور الصادات التراكم المئوي لتمثيل اللغات. وكما هو متوقع، تتجمّعُ معظم اللغات في الجهة اليمنى، حيث تكون نسب المساهمة ضئيلة، في حين تظهر بضع لغات فقط ناحية اليسار وتساهم في معظم البيانات قبل أن تتضاءل هذه المساهمة، ويُصبح مسار المُنحنى مُسطّحًا. وفي الإصدار الأخير من كامن كراول (CC-MAIN-2025-08)، كانت اللغة الإنكليزية وحدها الأرجح[76] لـ 43.3 في المئة من محتوى صفحات الإنترنت المُجرّفة، وتليها الروسية بواقع 6 في المئة فقط[77].

**الشكل (6)**

**التوزيع التراكمي للغات (بالنسبة المئوية)**




المصدر: من إعداد الباحثين.

72 Stefan Baack, "A Critical Analysis of the Largest Source for Generative AI Training Data: Common Crawl," in: *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (New York: Association for Computing Machinery, 2024), p. 2199.

73 تذكر الورقة نفسها التي أصدرتها "أوبن أي آي" حول نموذج "جي بي تي-3"، أنّ اللغة الإنكليزية استحوذت على 93 في المئة من تعداد الكلمات في بيانات التدريب.

74 Brown et al., p. 6.

75 علمًا أن الجمع بدأ سنة 2008، ولكننا لم نستطع تمثيل كامل الإطار الزمني، بما في ذلك كامل ما جُمع خلال سنة 2018، لأسباب تتعلق بتوافر البيانات.

76 تقوم منهجية المكتبة البرمجية (Compact Language Detector 2, CLD2) باحتساب النسبة الاحتمالية مُرتّبةً لثلاث لغاتٍ يُرجح أن تكون لغة صفحة الويب المُجرّفة.

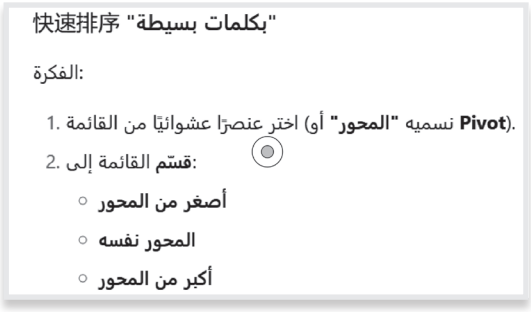
77 Common Crawl, "Statistics of Common Crawl Monthly Archives," *Common Crawl* (2025), accessed on 27/2/2025, at: https://acr.ps/1L9zQzY

ولا يتوقف الأمر على تعدد اللغات، بل يتجاوزه إلى تعدد اللهجات؛ إذ خلصت دراسة إلى أن أداء عدد من النماذج اللغوية (GPT-o1, GPT-4o, Claude-3.5-Sonnet) انخفض بمتوسط 10 في المئة عند تلقينها باستخدام لهجة "الإيبونكس" (African-American Vernacular English)، وهي لهجة الأميركيين من أصلٍ أفريقي، مُقارنةً بأدائها عند تلقينها "الإنكليزية القياسية" (Standard English)[78]. ويذكر الباحثون في الدراسة عدّة أسباب متفاوتة التأثير لانخفاض الأداء، يمكن أن تنسحب على اللغات الأخرى؛ إذ أدّت صعوبة فهم اللهجة ومعالجتها دورًا في تراجع الأداء بحسب نتائج الدراسة. كما وجدوا أن التطعيم التدريجي للإنكليزية القياسية بالقواعد التركيبية الصرفية للهجة الإيبونكس يؤثر في أداء النماذج بصفة ملحوظة. إضافةً إلى ذلك، تبيّن أنّ بعض الكلمات، التي لا تنفرد بها لهجة الإيبونكس، ارتبطت بتراجع أداء النموذج، بسبب أسلوب توظيفها في الجمل؛ وهو ما لا تحكمه قواعد معينة يسهل تلقينها للنموذج[79].

ومن الجوانب التي لا يُلتفت إليها عادةً أيضًا، التكلفة الكبيرة المرتبطة بمعالجة المدخلات بالنسبة إلى اللغات ضعيفة التمثيل[80]، خاصة ذات الأحرف غير اللاتينية، مثل العربية والفارسية. وهو ما قد يدفع النموذج إلى ارتكاب الأخطاء القواعدية، أو توليد كلماتٍ عشوائية، وتضمين كلمات أجنبية في الإجابات، كما يظهر في الشكل (7) على سبيل المثال؛ إذ نجد النموذج قد فشل في توليد الإجابة كاملة باللغة العربية، فأورد خطأً كلمة صينية محل كلمة أخرى عربية.

**الشكل (7)**

**مثال عن ارتكاب نموذج الذكاء الاصطناعي أخطاءً قواعدية أو توليده كلماتٍ عشوائية**




المصدر: من إعداد الباحثين.

يمتلك هذا النوع من اللغات تصريفات متنوعة للكلمة الواحدة. ونظرًا إلى أن النموذج لا يمتلك مقابلات في قائمة مفرداته لكل هذه التنويعات، يقوم بتجزيء الكلمات إلى وحدات أصغر (Tokenization) في محاولةٍ

78 Fangru Lin et al., "One Language, Many Gaps: Evaluating Dialect Fairness and Robustness of Large Language Models in Reasoning Tasks," *arXiv* (2025), p. 3, accessed on 27/2/2025, at: https://acr.ps/1L9zR3t

79 Ibid., p. 7.

80 Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (New Haven: Yale University Press, 2021).

لفهمها عبر فهم أجزائها. ويستعرض الجدول (1)[81] قائمة ببعض اللغات ومتوسط عدد الترميزات أو الأقسام (Tokens) التي تقابل الكلمة الواحدة لكل منها مُقارنةً بالإنكليزية[82]. وينعكس ارتفاع التقسيمات على معدل استخدام الذاكرة المطلوب، وحجم النصوص التي يُمكن تمريرُها إلى النموذج. فإذا كانت الكلمات في اللغة الأوكرانية تقسّم إلى ثلاثة أقسام، فإنها تحتاج 3*3؛ أي 9 أضعاف الطاقة الحاسوبية اللازمة للمعالجة[83].

**الجدول (1)**

**متوسط عدد الترميزات في بعض اللغات التي تقابل الكلمة الواحدة لكلٍّ منها مُقارنةً بالإنكليزية**

| اللغة | عدد التقسيمات |
|---|---|
| الأمهرية | 7.68 |
| الفرنسية | 1.6 |
| الألمانية | 1.58 |
| الهندية | 4.79 |
| العربية النجدية | 3.04 |
| الروسية | 2.49 |
| الأوكرانية | 3 |

المصدر: من إعداد الباحثين.

وقد اتجهت الكثير من الدول، على الرغم من الصعوبات اللوجستية والتقنية الآنفة الذكر، إلى بناء "نماذج لغوية ضخمة سيادية" (Sovereign LLMs) خاصة بها، باستخدام بنيتها التحتية والبيانات التي تمتلكها. ولا تقتصر جدوى بناء النماذج السيادية على حماية اللغة والثقافة، وإنما تجد الدول اليوم أنها تحتاجها لضمان الاستفادة من تطورات الذكاء الاصطناعي في احتياجاتها الأمنية، أو في المجالات التي تتطلب الوصول إلى بيانات ذات طابع حساس. فالشركات الكبرى، التي تزوّد المستخدم بالنماذج الضخمة، تحصل على بيانات المستخدم عندما يتفاعل مع خوادمها وواجهاتها البرمجية. كما أن امتلاك السيادة على النماذج اللغوية يقي الدول من مغبّات الاعتماد على الخارج[84] في حالات التوترات الجيوسياسية أو العقوبات الاقتصادية أو المشكلات التقنية وغيرها، مما قد يُعطّل الخدمات.

---

81 Aleksandar Petrov et al., "Language Model Tokenizers Introduce Unfairness between Languages," *arXiv* (2023), pp. 10-15, accessed on 27/2/2025, at: https://acr.ps/1L9zQnW

82 المُقسّم (Tokenizer) المُستخدم في الجدول هو (cl100k_base) المستخدم في "جي بي تي-4". فعلى سبيل المثال، بالنسبة إلى كلمة بسيطة مثل "لماذا"، فإن المرمز المستخدم في "جي بي تي-2" يُقسّمها إلى خمسة أجزاء، لكلّ حرفٍ جزء: Ibid., p. 2.

83 Artur Kiulian et al., "From English-Centric to Effective Bilingual: LLMs with Custom Tokenizers for Underrepresented Languages," *arXiv* (2024), p. 1, accessed on 27/2/2025, at: https://acr.ps/1L9zQRC

84 Muath Alduhishy, "Sovereign AI: What it is, and 6 Strategic Pillars for Achieving it," *World Economic Forum*, 25/4/2024, accessed on 27/2/2025, at: https://acr.ps/1L9zRdv

ويُعدّ نموذج "فالكون 40 بي" (Falcon 40B) الذي جرى إطلاقه في آذار/ مارس 2023 في الإمارات العربية المتحدة من التجارب العربية الرائدة في هذا السياق، ومن المساهمات المهمة عالميًّا نظرًا إلى إتاحة النسخة مفتوحة المصدر منه بعد إطلاقه بفترة قصيرة[85]. كذلك عربيًّا، يتميّز نموذج "علّام" السعودي بقدرته على توليد المحتوى باللغة العربية. ويتميّز نموذج "فنار" الذي أطلقته دولة قطر مؤخرًا، بقدرته على الفهم العميق للغة العربية بلهجاتها المختلفة ومصطلحاتها الثقافية.

أما عالميًا، فتوجد الكثير من الأمثلة البارزة، نذكر منها نموذج "جيغاشات" (GigaChat) الذي أطلقته روسيا سنة 2023، وجرى التسويق له بصفته "تشات جي بي تي" الروسي، والذي يختص بحسب الشركة المالكة بالقدرة على فهم اللغة الروسية وتوليدها أفضل من أيّ نموذج آخر آنذاك[86]. كذلك يُمكن الإشارة إلى نموذج "بروجيكت إندست" (Project Indust)، المخصص للغة الهندية (يشمل 37 لهجة من لهجاتها)، والذي نجح مطوروه في تصميمه بتكلفة 400 ألف دولار فقط، في حين تنفق ملايين الدولارات عادةً على تطوير نماذج مماثلة[87].

## رابعًا: السيناريوهات المستقبلية المحتملة لوضعية اللغات وإنتاج المعرفة العالمية

يُمكن أن تسلك النماذج اللغوية مسارات متعددة في نطاق السنوات المقبلة. وتحكم الكثير من "المجاهيل" (Unknowns) القدرة على التنبّؤ بشكل هذه المسارات. فمن جهة، تعتمد النماذج على التطور المستمر في طاقة المعالجة الحاسوبية. وقد صّرح في هذا الصدد جين-سون هوانغ، المدير التنفيذي لشركة "إنفيديا" أواخر السنة الماضية، أنّ الأداء الحاسوبي قد يتضاعف مرتين أو ثلاث مرات في السنوات العشر المقبلة[88]. وفي المقابل، كثيرًا ما تُثار تساؤلات بشأن قرب وصول تطوير المعالجات إلى حدودها أو تباطئها[89]. ومن جهةٍ أُخرى، يشكك بعضهم في استمرارية تطور أداء النماذج اللغوية الكبيرة باستخدام المقاربات المتبعة حاليًّا في التطوير. ويشير أحد أبرز الخبراء في هذا المجال، غاري ماركوس، إلى أنّ عوائد زيادة حجم النماذج والبيانات على أدائها بدأت تنخفض بالفعل، وأن النماذج اللغوية لن تصل إلى مستوى الذكاء الاصطناعي العام بمجرد ضخّ المزيد من الطاقة الحاسوبية والبيانات، محذرًا من أن هذه الصناعة لم تصل بعد إلى مستوى الربحية،

85 "الإمارات العربية المتحدة: معهد الابتكار التكنولوجي يشارك نموذج 'فالكون 40 بي' بشكل مفتوح المصدر لأغراض البحث والاستخدام التجاري"، معهد الابتكار التكنولوجي، 2023/5/25، شوهد في 2025/3/18، في: https://acr.ps/1L9zR7h

86 "Russia's Sberbank Releases ChatGPT Rival GigaChat," *Reuters*, 24/4/2023, accessed on 27/2/2025, at: https://acr.ps/1L9zQtx

87 "The Era of Sovereign LLMs and its Implications in the AI World: Insights from Cypher 2024," *Machine Hack Gen AI*, 29/11/2024, accessed on 27/2/2025, at: https://acr.ps/1L9zQM9

88 Tae Kim, "Nvidia CEO Jensen Huang Predicts 'Hyper Moore's Law' Pace for AI," *Barron's*, 7/11/2024, accessed on 21/2/2025, at: https://acr.ps/1L9zRgy

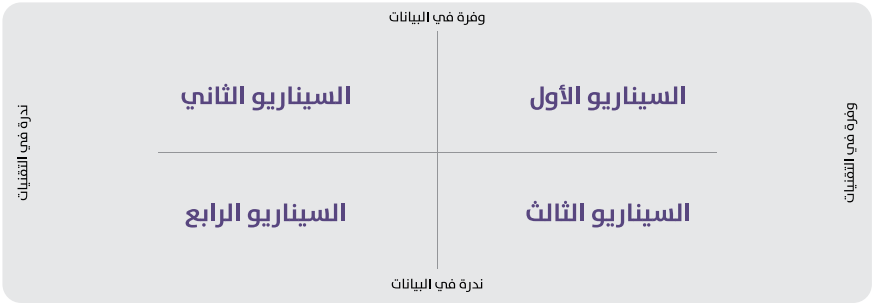
89 Thomas A. Campbell, "Moore's Law 2.0?" *Atlantic Council*, 5/9/2014, accessed on 21/2/2025, at: https://acr.ps/1L9zQV7

وقد لا تصل إليه قريبًا، بسبب التكاليف العالية جدًا في الطرائق المتبعة حاليًا لمواكبة السباق؛ ولذلك يتوقع أن ينسحب المستثمرون عند انفجار فقاعة الذكاء الاصطناعي، وإدراك الناس "لهذه الحقيقة"[90].

وتعدّ إشكالية محدودية بيانات التدريب، على ضخامتها، عاملاً آخر يحدّد مستقبل هذه النماذج. وقد خلصت دراسة لمؤسسة "إيبوك إي آي" (Epoch AI) إلى أن حجم ما يتوافر حاليًا من بيانات يساوي 300 مليار تقسيم (أي كلمة أو جزء من كلمة) مما ولّده البشر ويعتبر مُتاحًا للعامة. وبناءً عليه، يُقدّر الباحثون أن تستنزف النماذج هذه الكميات بين سنتي 2026 و2032، مع احتمال للتكرار المعرفي في البيانات المتبقية[91]. وتُطرح "البيانات المُصطنعة" (Synthetic Data)، بوصفها مصدرًا مُحتملاً للتدريب، وهي البيانات التي تولّدها النماذج اللغوية بنفسها وفق آلية مُنظّمة لتكون جاهزةً لتدريب النماذج. وبحسب دراسة أجراها هاو شين وآخرون، يوجد ترابطٌ إيجابي ذو دلالة إحصائية بين تنوع البيانات المُصطنعة وانخفاض أداء النماذج[92].

**الشكل (8)**

**سيناريوهات مُحتملة لمصير اللغات الضعيفة التمثيل في السنوات المقبلة**




المصدر: من إعداد الباحثين.

ومن المهم أيضًا التنبؤ بملامح الأطر التشريعية لحكامة النماذج اللغوية. وترتبط هذه الأخيرة في علاقة تأثير متبادلة بتطور أداء النماذج؛ إذ يحكم ذلك عدة عوامل، يشوبها اللايقين، كالتسابق بين الدول في مجال الذكاء الاصطناعي، الذي يستلزم مرونةً تشريعية من جانب، وتنامي دور النماذج في الحياة اليومية وماله من تبعات على الخصوصية وحقوق الملكية الفكرية وغير ذلك من جانب آخر.

90 Gary Marcus, "CONFIRMED: LLMs have Indeed Reached a Point of Diminishing Returns," *Marcus on AI*, 10/11/2024, accessed on 21/2/2025, at: https://acr.ps/1L9zR48

91 Pablo Villalobos et al., "Will We Run Out of Data? Limits of LLM Scaling Based on Human-Generated Data," *Epoch AI*, 6/6/2024, accessed on 21/2/2025, at: https://acr.ps/1L9zQwI

92 Hao Chen et al., "On the Diversity of Synthetic Data and its Impact on Training Large Language Models," *arXiv* (2024), p. 11, accessed on 27/2/2025, at: https://acr.ps/1L9zR9E

يستعرض الشكل (8) أربع سيناريوهات مُحتملة لمصير اللغات الضعيفة التمثيل في السنوات المقبلة. وتعرف هذه بمصفوفة 2x2، وتُعتبر التقنية الأكثر شيوعًا في رسم السيناريوهات المستقبلية[93]. وتعتمد التقنية على اختيار عاملين يشترط فيهما أن يكون تأثيرهما عاليًا، وتشوبهما درجة عالية من اللايقين[94]. ويفترض النموذج ثبات مجموعة من العوامل، هي استمرارية الجدوى الاقتصادية للنماذج اللغوية، وعدم حدوث مفاجآت نوعية على صعيد الأداء، كتطور هيكلية الشبكات العصبية، وغيره مما لا يُمكن التنبؤ به. وفي ضوء ذلك، فإننا نتحدث عن بضع سنوات مقبلة فقط.

## 1. السيناريو الأول: وفرة بيانات وتقنيات

أولاً، يتنامى في هذه الحالة ما هو مُتاح من اللغات القليلة الموارد عبر زيادة بصمتها الرقمية. ويكون ذلك بطرائق مُختلفة، كـ "تعبئة الجهود" (Crowdsourcing)، في حال لم يتوافر التمويل اللازم. ويُعد "مشروع كورباداِن" (The Crúbadán Project) من الأمثلة الأشهر على هذا النوع من المبادرات؛ إذ تقوم فكرته على تجريف الإنترنت والتعاون مع المتكلّمين الأصليين، لتقديم المعونة الاستشارية، وتزويد الباحثين بالنصوص اللازمة للتدريب لاستخدامها حين يتطلّب الأمر[95].

ثانيًا، يفترض هذا السيناريو تطوير التقنيات اللازمة ليكون تدريب النماذج على هذه البيانات ذا جدوى اقتصادية، كتصميم "مُرمّزاتٍ" (Tokenizers) تلائم خصائص اللغات الضعيفة المصادر؛ بحيث تُقسّم الكلمات بما يُناسب حالة كل لغة، ويُسهّل على النماذج تعلّم الأنماط فيها. ويفترض تطوير أدوات تنضيد وتوسيم تميّز نوعية البيانات في هذه اللغات، لتنقية ما نلقّنه للنماذج منها.

ويستتبع تحقق هذا السيناريو عدة مخرجات إيجابية على منتجي النماذج ومستهلكيها؛ إذ سيكون في متناول الشركات المُطوّرة للنماذج اللغوية الوصول إلى أسواق جديدة، ولن تقتصر حينها الخدمات المُقدّمة للمُجتمعات الضعيفة التمثيل على النماذج اللغوية فحسب، وإنما ستمتدّ لتشمل تطبيقات معالجة اللغات الطبيعية الأخرى، كالترجمة وغيرها. إضافةً إلى ذلك، ستعكس النماذج حينها الخصوصية الثقافية لهذه اللغات، وتُسهّل على المجتمعات غير الناطقة بها الوصول إليها.

## 2. السيناريو الثاني: وفرة في البيانات مع ندرة في التقنيات

يفترض هذا السيناريو تنمية بيانات اللغات الضعيفة التمثيل، مع بقاء التقنيات اللازمة لتدريب النماذج على وضعها. ولن يختلف الوضع في هذه الحالة بالنسبة إلى تمثيل اللغات في النماذج اللغوية

93 Alun Rhydderch, "Scenario Building: The 2x2 Matrix Technique," *Futuribles International* (June 2017), p. 16.

94 Ibid., p. 3.

95 Kevin P. Scannell, "The Crúbadán Project: Corpus Building for Under-Resourced Languages," *Cahiers du Cental*, vol. 5 (2007), pp. 4-5.

المعروفة؛ والبديل سيكون هو تدريب نماذج متخصصة تتكيّف ومتطلبات كل لغة[96]. ويُمكن الإشارة في هذا السياق مثلاً إلى نموذج "فنار" الذي أطلقته دولة قطر أواخر العام الفائت، كمثال على النماذج المتخصصة والضخمة[97]. ولأن القيام بهذا النوع من المشاريع يفرض تكاليف باهظة، فإنه من المتوقع في هذا السيناريو أن يقتصر وجود هذه النماذج على بعض الاقتصادات القوية. ويمكن أن تنتقل هذه التجارب إلى المربع الأول بسرعة عبر العمل على توفير المزيد من البيانات، وتطوير التقنيات المكتسبة في التجارب الأولى. ويمكن أيضًا لتضافر جهود فرق تعمل على اللغة والثقافة كفرق "جيس"، و"علام"، و"فنار"، وغيرها، أن تسرّع في عملية الانتقال إلى المربع الأول.

## 3. السيناريو الثالث: وفرة في التقنيات مع ندرة في البيانات

في حال توافرت التقنيات اللازمة لتدريب النماذج بتكاليف معقولة، لكن من دون وجود البيانات بوفرة، فسيكون هنالك تباين في أداء النماذج في اللغات المختلفة؛ ذلك أنّ الأداء سيبقى سيئًا بالنسبة إلى اللغات الضعيفة المصادر إجمالاً، لكنه سيتفاوت حينها بالنسبة إلى النماذج المعروفة، خاصة في اللغات التي تتحصل عليها حصريًّا لأسباب مختلفة. وقد ينجح مطورو نموذج معين في الحصول على بيانات ثرية للغة معينة، بحكم موقع جغرافي، أو من خلال الاستثمار في مبادرة خاصة بهذا الشأن، أو ربما من خلال اتفاقيات مع مؤسسات أكاديمية وبحثية تمتلك قواعد بيانات يُمكن الاستفادة منها.

ومن جهةٍ أخرى، سيسهل توافر هذه التقنيات تدريب نماذج صغيرة لا تتطلب قدرات معالجة كبيرة للقيام بمهمات بسيطة؛ ما يعني توسيع قاعدة المستخدمين ودمقرطة النماذج.

## 4. السيناريو الرابع: بقاء الوضع على ما هو عليه اليوم

في حال بقيت الحالة على ما هي عليه اليوم؛ أي ظلّت اللغات الضعيفة المصادر على فقرها، ولم تنجح التقنيات في تخفيض تكلفة تدريب النماذج بصفة مؤثرة، فإن الهوة الرقمية، مع ما يرتبط بها من غياب فرص تعليمية واقتصادية، ستتسع لمصلحة أصحاب اللغات المهيمنة، وستكون اللغات الضعيفة المصادر التي يعيشُ متكلّموها في دول ضعيفة اقتصاديًّا أو صغيرة، أكثر عرضة للتبعات السلبية، مُقارنةً بغيرها، مع ذهاب استثمارات الشركات المطورة للنماذج اللغوية تجاه تنمية حصتها في الأسواق الكبيرة، كالهند مثلاً.

وإضافةً إلى ذلك، سيؤسس عدم التوازن هذا لانحياز ثقافي لمصلحة اللغات المهيمنة، التي سيضطر المستخدمون إلى استخدامها للاستفادة على نحو كامل من خدمات النماذج اللغوية. ويُمكن أن

96 Tianyang Zhong et al., "Opportunities and Challenges of Large Language Models for Low-Resource Languages in Humanities Research," *arXiv* (2024), pp. 21-22, accessed on 27/2/2025, at: https://acr.ps/1L9zRbZ

97 "His Excellency the Prime Minister Unveils Fanar, a Breakthrough in Arabic AI," *Government Communication Office*, 11/12/2024, accessed on 22/2/2025, at: https://acr.ps/1L9zQU9

يكون هذا الانحياز بين اللغات وبين لهجات اللغة الواحدة نفسها[98]، كما سلف في نموذج لهجة "الإيبونكس" الإنكليزية.

## خامسًا: خلاصات وتوصيات

في ضوء رسمنا لخريطة الذكاء الاصطناعي العالمي، الموسومة بالسباق التنافسي المحموم بين عمالقة الذكاء الاصطناعي، لا سيما الولايات المتحدة والصين، وإبرازنا تطور النماذج اللغوية الضخمة السريع، ودخولها المتسارع في أتمتة عمليات إنتاج المعرفة وتداولها، ورسمنا عددًا من السيناريوهات المُستقبلية لمصائر الثقافات المحلية واللغات ومعارفها، نخلص إلى الخلاصات التالية، التي تضمّ عددًا من التوصيات بشأن سبل تجاوز فجوة البيانات بين اللغات، وتمثيل اللغات المختلفة في النماذج اللغوية، لا سيما منها الضعيفة التمثيل:

- تفتقر بعض اللغات إلى مجتمع ناطقين قادرٍ على تخديمها كما ينبغي للحفاظ عليها من الانقراض. وتشير منظمة اليونسكو، على سبيل المثال، إلى أنّ العالم يشهد انقراض لغة كل أسبوعين[99]. وإن كانت توجد بالفعل مبادرات لرقمنة مخزون اللغات الضعيفة المصادر، فإنّ غياب وجود رؤية حكومية تناسب خصوصية كل مجتمع، وتضع جدولًا زمنيًا للتنفيذ، يعرقل تنظيم تعبئة الموارد وتخصيصها على النحو الأمثل.
- تستفيد بعض المبادرات من تعبئة الجهود الجماعية لبناء حزم بيانات لغوية. ويُعتبر مشروع "موزيلا كومن فويس" (Mozilla Common Voice) من الأمثلة البارزة لهذا النوع من الجهود. وتقوم فكرة المشروع على توفير منصتين إلكترونيتين، تطلب الأولى من المستخدمين قراءة جمل مكتوبة، أو رفع جمل جديدة، أو مراجعة تسجيلات مُستخدمين آخرين. وتطلب المنصة الثانية من المستخدمين الإجابة عن أسئلة معينة، في حين يدوّن مستخدمون آخرون تسجيلات هذه الأجوبة. وتوفر المنصّة بياناتها مجانيًا، وتغطّي نحو 130 لغة. بيد أنّ مشكلة هذه المبادرة وغيرها من المبادرات الشبيهة تكمنُ، بحسب موقع برنامج الأمم المتحدة الإنمائي، في تشرذمها وافتقارها إلى التنظيم[100]. ويؤكد هذا أنّ تضافر الجهود التنفيذية مع الرؤى الحكومية ضروري حتى يتحقق لهذه المبادرات الشمول الجغرافي، ولكي تستفيد من القدرة الاستثنائية للحكومات على رفع مستوى المشاركة الشعبية اللازمة.

98 Regina Ta & Nicol Turner Lee, "How Language Gaps Constrain Generative AI Development," Brookings, 24/10/2023, accessed on 22/2/2025, at: https://acr.ps/1L9zQVb

99 Oleksii Sharavar, "Making AI Work for all: Bridging the Language Gap in AI by Focusing on Languages that are Underrepresented in the Digital World," *GSMA*, 17/1/2025, accessed on 23/2/2025, at: https://acr.ps/1L9zRgo

100 "Every Language Matters: Building a More Inclusive Digital Future," *UNDP*, 5/10/2024, accessed on 23/2/2025, accessed on 27/2/2025, at: https://acr.ps/1L9zQQb

- يرتبط تطوير اللغات الفقيرة المصادر بمستوى التنمية الاقتصادية والاجتماعية للناطقين بها، إضافةً إلى البنية التحتية اللازمة للاتصالات. فمن جهة، يحتاج المتكلمون الأصليون إلى امتلاك المقوّمات التي تجعلهم قادرين على المشاركة والتطوع في المشاريع من دون الانقطاع عن وظائفهم وأشغالهم؛ ومن جهة أخرى، يحتاج الباحثون إلى العمل مع من يمتلك الخبرة بالتنوعات اللغوية لإيجاد الترجمة الأمثل المقابلة لكل جملة[101].
- يُمكن تنمية بيانات اللغات ضعيفة المصادر بطريقة آلية لزيادة حجمها وتنويعها، وذلك باستخدام الترجمة العكسية (ترجمة النص إلى الإنكليزية مثلًا، أو أي لغة غنية في المصادر، ثم إعادة ترجمته إلى اللغة الأصلية لزيادة التنوع)، أو توليد البيانات المصطنعة، وغير ذلك. ويتعين على الباحثين الاستفادة من قدرة النماذج على تعلم الأنماط المشتركة بين اللغات، واكتشاف الطرائق التي يُمكنهم عبرها زيادة حصة اللغات الضعيفة الموارد في التدريب في المراحل اللاحقة على نحو تدريجيّ. وتستطيع تقنيات "التعلم مع صفر أمثلة" (Zero-Shot Learning)، وتقنيات "التعلم عبر أمثلة قليلة" (Few-Shot Learning)، أن تطور من أداء النماذج في اللغات الضعيفة المصادر. وأخيرًا، فإنه من المفيد تضمين تمثيل رياضي للّهجات المختلفة ضمن النموذج، تساعده على تحسين أدائه، أو تزويد النموذج بحزم بيانات شاملة ومتنوعة للهجات المختلفة[102].
- تحظى أدوات التقييم والمقايسة باهتمام متزايد[103]؛ لكن ما تزال اللغات الضعيفة المصادر تفتقر إلى كفايتها منها. وتساعد هذه الأدوات على تطوير أداء النماذج، وفهمه، وتحديد العيوب، لتشجيع المطورين على معالجتها. ويضع مطورو هذه الأدوات مهمات أو اختبارات للنماذج[104]، كاختبار قدرة النموذج على فهم الأمثال الشعبية بالنظر إلى حمولتها الثقافية[105].
- أظهر باحثون من شركة "أنثروبيك"، في دراسة أجروها سنة 2024، أن النماذج اللغوية يُمكنها أن تنتهج سلوكيات ضارّة، وحتى ولو بدَت آمنةً في مرحلة التدريب؛ وذلك عبر تعليم النماذج الاستجابة إلى مُدخلاتٍ معُينة، أو التقاط إشارات خاصّة، كاختلاف السياق. ويُمكن تعليم النموذج، على سبيل المثال، أن يكتب برنامجًا آمنًا في سنة 2023، ومن ثم يعدّل ويكتب برنامجًا قابلاً للاختراق لدى طلب الأمر نفسه منه في سنة 2024. وقد وجد الباحثون أن هذا السلوك

101 Joanito Agili Lopo & Radius Tanone, "Constructing and Expanding Low-Resource and Underrepresented Parallel Datasets for Indonesian Local Languages," *arXiv* (2024), p. 10, accessed on 27/2/2025, at: https://acr.ps/1L9zRfp

102 Zhong et al., pp. 21-22.

103 Israel Abebe Azime et al., "ProverbEval: Exploring LLM Evaluation Challenges for Low-resource Language Understanding," *arXiv* (2025), p. 1, accessed on 27/2/2025, at: https://acr.ps/1L9zQmm

104 Roberts Darģis et al., "Evaluating Open-Source LLMs in Low-Resource Languages: Insights from Latvian High School Exams," in: Mika Hämaläinen et al. (ed.), *Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities* (Miami: Association for Computational Linguistics, 2024), p. 289.

105 Azime et al., p. 9.

يستمرّ حتى بعد تطبيق "تقنيات السلامة"[106]. وتُبيّن هذه النتائج أهمية أن تستثمر المجتمعات اللغوية الضعيفة التمثيل في أبحاث سلامة الذكاء الاصطناعي لحماية نفسها من السلوكيات الضارة، إضافةً إلى تعزيزها لضرورة بناء نماذج سياسية لحماية نفسها من التطبيقات الحساسة، ذات الطابع الأمني، وغيرها من الجوانب الثقافية المؤمننة، كتلك المُرتبطة بالأحقية التاريخية في السيادة على أراضٍ مُعيّنة على سبيل المثال.

- من المهم عدمُ تشتيت الجهود والموارد بين الدول التي تجمعها تحديات مشتركة، وخاصةً الدول التي تجمعها اللغة نفسها. وفي هذا الصدد، تُعتبرُ دول مجلس التعاون لدول الخليج العربية، على سبيل المثال، بيئةً خصبةً بمشاريع الذكاء الاصطناعي. فبحسب استطلاع حديث لمجموعة بوسطن الاستشارية، فإن واحدة من بين كلّ أربع شركات استطلعت المجموعة آراء إداريها في الخليج العربي، تُخطط لاستثمار نحو 25 مليون دولار في الذكاء الاصطناعي[107]. ويُتوقّع أن تراوح مُساهمة الذكاء الاصطناعي في الناتج المحلي الإجماليّ لدول مجلس التعاون بين 8.2 و12.4 في المئة بحسب الدولة. وتبدي دول الخليج العربي اهتمامًا واضحًا بالذكاء الاصطناعي، كما تظهر استثماراتها الكبيرة، وخططها الاستراتيجية المُعلنة. بيد أنها، على الرغم من جهودها لبناء نماذج منفردة، ما تزال تفتقر إلى برنامج عملٍ مُشترك يُمكن أن يسمح بالاستفادة من البيانات الضخمة والبنية التحتية والرساميل البشرية الماهرة التي تمتلكها مجتمعة. وهو ما أخّر، إلى جوانب عوامل أُخرى، تقديم نموذج لغويّ ضخم، قادرٍ على فهم العربية وتوليدها بكفاءة ووعي ثقافي بما يواكب قدرات النماذج المتفوّقة، حتى تاريخ كتابة الدراسة.

## خاتمة

ينفرد الذكاء الاصطناعي التوليدي عن غيره من الاختراعات، بوظيفته المميزة بصفته حاملاً للمعرفة، وربما لا يقترب من قدرته الراديكالية على تغيير واقع التعلم وانتشاره على طول التاريخ الإنساني سوى اختراعات الكتابة، فالطابعة، فالإنترنت. ويعني ذلك أن هذه ليست المرة الأولى التي يواجه فيها البشر تغيرًا جذريًا في دينامية انتشار المعرفة، يفرض نفسه على الجميع من دون استئذان. لعل الفرق هو أن البشرية كانت تتبنّى هذه الاختراعات كلّ مرة في فترات مختلفة وبدرجات متفاوتة؛ فمنهم المتوجّس، ومنهم المتحمّس، ومنهم من يأخذ زمام الأمور، ومنهم المتلقّي السلبي. والأمر المختلف هنا هو أنّ سرعة تقدّم التكنولوجيا، فالكتابة، والطابعة، والإنترنت، لم تتغير على هذا النحو المتسارع الذي نشهده اليوم مع الذكاء الاصطناعي، ولم يكن تعقيدُها يزداد بالوتيرة نفسها.

106 Evan Hubinger et al., "Sleeper Agents: Training Deceptive LLMs that Persist Through Safety Training," arXiv (2024), pp. 2, 4, accessed on 27/2/2025, at: https://acr.ps/1L9zQSz

107 "One in Four GCC Companies Plan to Spend More than $25m on AI in 2025," *The Peninsula Qatar*, 20/3/2025, accessed on 23/3/2025, at: https://acr.ps/1L9zQic

ولذا، فإن الإطار الزمني للفعل صغير مُقارنةً بأي تغيرٍ فيما مضى. وقد ظهر ذلك جليًّا في التمهيد لمصفوفة السيناريوهات التي وضعناها؛ إذ يصعب التنبؤ بمسار الأمور حتى فترة طويلة، لأن ذلك رهنٌ بالكثير من المجاهيل التي يصعب الفصل فيها، خاصةً مع اكتساب النماذج اللغوية الكبيرة للطابع المؤمنن والسباقي، واجتذابها لرؤوس الأموال الكبيرة، والتشكيك في قدرتها على تحقيق أشواط طويلة في التطور، وإحداثها للصدمة نفسها التي جاءت بها النماذج الأولى. وبتعبير آخر، تراوح السيناريوهات في طيفٍ، حدّه الأول أبعدُ ممّا نحن عنده الآن بقليل، والآخر بعيدٌ جدًا؛ إذ يذهب بعضهم إلى أنّه سيتجاوز قدرات الإنسان، ويُنهي عصر الندرة، ويغيّر أسلوب عيشنا إلى الأبد.

تبقى النماذج اللغوية الضخمة، على الرغم من الإرباك الذي يكتنف مُستقبلها، ملزمة بالكثير من العمل بحكم ما تمتلكه في حاضرها فقط. فالهوة اليوم واسعة بما يكفي بين مجموعات المستخدمين، والنماذج اللغوية بقدراتها الحالية أصبحت جزءًا أساسيًّا من عمل الأفراد والمؤسسات المختلفة والفعاليات الاقتصادية، وتتجه إلى مزيد من الاستخدام. ولن يحول بذلك فشل توقعات المبشرين بالذكاء العام الاصطناعي دون أن تتبوّأ النماذج اللغوية دورًا رئيسًا بوصفها مصدرًا للمعرفة في أفضل الأحوال، وربما الدور المُهيمن، ومن ثمّ ستتحمل مجتمعات الدول غير الفاعلة تبعات السيناريوهات المختلفة على المستوى الاجتماعي المرتبط بارتدادات التحيّز ضد ثقافاتها، وعلى المستوى التنموي المُتعلّق بتفاوت فرص الوصول إلى الخدمات، والعجز عن التكيف كما ينبغي مع متطلبات السوق.

ويطرح مجموع هذه التحديات التي وردت في الورقة أسئلةً بشأن جدوى التفكير في التعاون الإقليمي العابر للحدود بين الأقاليم الثقافية، وعملية النشاط القانوني والاجتماعي، الذي يُمكن أن يحمل الشركات المطوّرة على مشاركة أعباء التحوّل، وحول السياسات العامة الموصى بها في حال تعثرت أو فشلت الجهود في تلبية الاحتياجات الحقيقية، لتخصيص الموارد حيث يكون مردودها أكبر.

# المراجع

## العربية

سعيد، إدوارد. **الاستشراق: المفاهيم الغربية للشرق**. ترجمة محمد العناني. القاهرة: رؤية للنشر والتوزيع، 2006 [1978].

هوميروس. **الأوديسة لشاعر الخلود هوميروس**. ترجمة دريني خشبة. القاهرة/ لندن: مؤسسة هنداوي، 2021.

الهيئة السعودية للبيانات والذكاء الاصطناعي. **التحيز في أنظمة الذكاء الاصطناعي: تحديات وحلول**. الرياض: الهيئة السعودية للبيانات والذكاء الاصطناعي، 2025. في: https://acr.ps/1L9zQU7

الهيئة السعودية للبيانات والذكاء الاصطناعي. **معجم البيانات والذكاء الاصطناعي**. الرياض: الهيئة السعودية للبيانات والذكاء الاصطناعي، 2022. في: https://acr.ps/1L9zQCd

## الأجنبية

Abebe Azime, Israel et al. "ProverbEval: Exploring LLM Evaluation Challenges for Low-resource Language Understanding." *arXiv* (2025). at: https://acr.ps/1L9zQmm

Alduhishy, Muath. "Sovereign AI: What it is, and 6 Strategic Pillars for Achieving it." *World Economic Forum*. 25/4/2024. at: https://acr.ps/1L9zRdv

AlShebli, Bedoor et al. "China and the U.S. Produce more Impactful AI Research when Collaborating Together." *Scientific Reports*. vol. 14, no. 28576 (2024).

Apaalabono Atanga, Barbara, Xunyue Xue & Anna S. Mattila. "The Impact of Corporate Sociopolitical Activism (CSA) on Brand Attitude." *International Journal of Hospitality Management*. vol. 107 (2022).

Bromham, Lindell et al. "Global Predictors of Language Endangerment and the Future of Linguistic Diversity." *Nature Ecology & Evolution*. vol. 6, no. 2 (February 2022).

Burtch, Gordon et al. "The Consequences of Generative AI for Online Knowledge Communities." *Nature* (May 2024).

Buyl, Maarten et al. "Large Language Models Reflect the Ideology of their Creators." *arXiv* (2025). at: https://acr.ps/1L9zR9S

Campbell, Thomas A. "Moore's Law 2.0?" *Atlantic Council*. 5/9/2014. at: https://acr.ps/1L9zQV7

Chen, Hao et al. "On the Diversity of Synthetic Data and its Impact on Training Large Language Models." *arXiv* (2024). at: https://acr.ps/1L9zR9E

Crawford, Kate. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press, 2021.

Darwiche, Adnan. "Human-level Intelligence or Animal-Like Abilities?" *Communications of the ACM*. vol. 61, no. 10 (September 2018).

Faraone, Christopher A. "Circe's Instructions to Odysseus. OD. 10.507–40. As an Early Sibylline Oracle." *The Journal of Hellenic Studies*. vol. 139 (November 2019).

Haenlein, Michael & Andreas Kaplan, "A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence." *California Management Review*. vol. 61, no. 4 (2019).

Heaven, Will Douglas. "OpenAI's New Language Generator GPT-3 is Shockingly Good—and Completely Mindless." *MIT Technology Review*. 20/7/2020. at: https://acr.ps/1L9zQSJ

Heaven, Will Douglas. "Why Meta's Latest Large Language Model Survived Only Three Days Online." *MIT Technology Review*. 18/11/2022. at: https://acr.ps/1L9zRkh

Hendrycks, Dan, Eric Schmidt & Alexandr Wang. "Superintelligence Strategy." *arXiv* (2025).

Huang, Kaiyu et al. "A Survey on Large Language Models with Multilingualism: Recent Advances and New Frontiers." *arXiv* (2025). at: https://acr.ps/1L9zQFw

Hubinger, Evan et al. "Sleeper Agents: Training Deceptive LLMs that Persist Through Safety Training." *arXiv* (2024). at: https://acr.ps/1L9zQSz

Imbrie, Andrew, Elsa B. Kania & Lorand Laskai. *The Question of Comparative Advantage in Artificial Intelligence: Enduring Strengths and Emerging Challenges for the United States*. Washington, DC: Center for Security and Emerging Technology, 2020.

Jackson, Kevin. "Breaking Down Global Government Spending on AI." *HPCwire*. 26/8/2024. at: https://acr.ps/1L9zQC8

Kaplan, Jared et al. "Scaling Laws for Neural Language Models." *arXiv* (2020). at: https://acr.ps/1L9zRev

Kiulian, Artur et al. "From English-Centric to Effective Bilingual: LLMs with Custom Tokenizers for Underrepresented Languages." *arXiv* (2024). at: https://acr.ps/1L9zQRC

Komaitis, Konstantinos et al. "The sovereignty trap." *Geotech Cues/Atlantic council* (July 2024). at: https://acr.ps/1L9zQQM

Kreps, Sarah. "The Global AI Race: Will US Innovation Lead or Lag?" Brookings. 6/12/2024. at: https://acr.ps/1L9zQII

Lee, Kai-Fu. *AI Superpowers: China, Silicon Valley, and the New World Order*. Boston: Houghton Mifflin Harcourt, 2018.

LifeArchitect. "Timeline of AI and Language Models." at: https://acr.ps/1L9zQki

Lin, Fangru et al. "One Language, Many Gaps: Evaluating Dialect Fairness and Robustness of Large Language Models in Reasoning Tasks." *arXiv* (2025). at: https://acr.ps/1L9zR3t

Lopo, Joanito Agili & Radius Tanone. "Constructing and Expanding Low-Resource and Underrepresented Parallel Datasets for Indonesian Local Languages." *arXiv* (2024). at: https://acr.ps/1L9zRfp

Luong, Ngor, Zachary Arnold & Ben Murphy. *Understanding Chinese Government Guidance Funds*. Washington, DC: Center for Security and Emerging Technology, 2021.

Mark Ryan et al. "An AI ethics 'David and Goliath': Value Conflicts between Large Tech Companies and their Employees." *AI & Society*. vol. 39 (March 2022).

Maslej, Nestor et al. *The AI Index 2024 Annual Report*. Stanford: AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, April 2024.

McCorduck, Pamela. *Machines Who Think*. 2nd ed. Natick, MA: AK Peters, Ltd., 2004.

Mika Hämaläinen et al. (ed.). *Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities*. Miami: Association for Computational Linguistics, 2024.

National Security Commission on Artificial Intelligence. *Final Report*. Washington, DC: 2021.

Newell, Allen & Herbert A. Simon, "The Logic Theory Machine: A Complex Information Processing System." *IRE Transactions on Information Theory*. vol. 2, no. 3 (September 1956).

*NIPS '20: Proceedings of the 34th International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc., 2020.

Petrov, Aleksandar et al. "Language Model Tokenizers Introduce Unfairness between Languages." *arXiv* (2023). at: https://acr.ps/1L9zQnW

*Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. New York: Association for Computing Machinery, 2024.

*Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Volume 1: Long Papers.* Dublin, Ireland: Association for Computational Linguistics, 2022.

Rhydderch, Alun. "Scenario Building: The 2x2 Matrix Technique." *Futuribles International* (June 2017).

Roberts, Huw. "Digital Sovereignty and Artificial Intelligence: A Normative Approach." *Ethics and Information Technology*. vol. 26 (October 2024).

Rumelhart, David E. Geoffrey E. Hinton & Ronald J. Williams. "Learning Representations by Back-Propagating Errors." *Nature*. vol. 323 (1986).

Scanlon, Ruby. "Beyond DeepSeek: How China's AI Ecosystem Fuels Breakthroughs." *Lawfare*. 14/2/2025. at: https://acr.ps/1L9zQXC

Scannell, Kevin P. "The Crúbadán Project: Corpus Building for Under-Resourced Languages." *Cahiers du Cental*. vol. 5 (2007).

Sharavar, Oleksii. "Making AI Work for all: Bridging the Language Gap in AI by Focusing on Languages that are Underrepresented in the Digital World." *GSMA*. 17/1/2025. at: https://acr.ps/1L9zRgo

Sheehan, Matt. "Biden's Unprecedented Semiconductor Bet." Carnegie. 27/10/2022. at: https://acr.ps/1L9zQvT

Sheehan, Matt. *China's AI Regulations and How They Get Made*. Washington, DC: Carnegie Endowment for International Peace, 2023.

Stanford HAI staff. "Global AI Power Rankings: Stanford HAI Tool Ranks 36 Countries in AI." *Stanford HAI*. 21/11/2024. at: https://acr.ps/1L9zQKl

Stryker, Cole & Eda Kavlakoglu, "What is Artificial Intelligence. AI?" *IBM*. 9/8/2024. at: https://acr.ps/1L9zQDu

Ta, Regina & Nicol Turner Lee. "How Language Gaps Constrain Generative AI Development." Brookings. 24/10/2023. at: https://acr.ps/1L9zQVb

The White House. "Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence." 24/10/2024. at: https://acr.ps/1L9zRas

Toosi, Amirhosein et al. "A Brief History of AI: How to Prevent Another Winter. A Critical Review." *PET Clinics*. vol. 16. no. 4 (October 2021).

Tsamados, Andreas et al. "The Ethics of Algorithms: Key Problems and Solutions." *AI & Society*. vol. 37 (2022).

Vaswani, Ashish et al. "Attention Is All You Need." *arXiv* (2023). at: https://acr.ps/1L9zRav

Winter-Levy, Sam. "With Its Latest Rule, the U.S. Tries to Govern AI's Global Spread." Carnegie. 13/1/2025. at: https://acr.ps/1L9zQF8

Zhong, Tianyang et al. "Opportunities and Challenges of Large Language Models for Low-Resource Languages in Humanities Research." *arXiv* (2024). at: https://acr.ps/1L9zRbZ

Zhou, Longjun. "A Historical Overview of Artificial Intelligence in China." *Science Insights*. vol. 42, no. 6 (2023).