

محاضرة 5: النماذج اللغوية الكبيرة وهندسة الأوامر — قراءة اجتماعية موسّعة

تقوم النماذج اللغوية الكبيرة على فكرة بسيطة في ظاهرها، عميقة في آثارها: عندما نعرض على الحاسوب كمًا هائلًا من النصوص، يصبح قادرًا على توقّع الكلمة التالية في الجملة بدقة متزايدة. غير أنّ هذه القدرة الإحصائية على ترتيب الكلمات لا تعني "معرفة" بالمعنى البشري، بل تعني مهارة في محاكاة أنماط اللغة التي رآها من قبل، وهذا فرق جوهري يجب أن ينتبّه منه الباحث الاجتماعي قبل أن يحمّل الأداة ما لا تحتمل من مسؤوليات معرفية أو أخلاقية. (Bender et al., 2021; Mitchell, 2019) هذه المحاضرة لا تستعرض الجانب التقني بقدر ما تشرح لماذا تُعد هذه النماذج مفيدة في عملنا اليومي كباحثين اجتماعيين، وكيف نُحسن توجيهها عبر صياغة أوامر واضحة تحوّل "القوة اللغوية" إلى مخرجات قابلة للعمل، وكيف نتحقق من النتائج تحقّقًا نقديًا يليق بمسؤولية القرار الاجتماعي، ونتناول في الطريق حدود هذه الأدوات ومزالقها الأخلاقية ومنهجية إدارتها بشفافية ومسؤولية (Bommasani et al., 2021; Russell & Norvig, 2020).

تتعلم هذه النماذج عبر مرحلتين متكاملتين. في الأولى، التي تُسمّى التعليم المسبق، تُدرّب الخوارزمية على كميات ضخمة من النصوص العامة بهدف التقاط البنية الإحصائية للغة: أنماط التراكيب، وروابط الجمل، وأساليب الربط بين الأفكار. هنا تتعلّم النموذج مهمة واحدة متكرّرة: **التنبؤ بالكلمة التالية** في كل سياق نصّي تراه (Brown et al., 2020). أمّا المرحلة الثانية فتُعنى بمواءمة السلوك مع توقعات البشر، وذلك عبر تدريب إضافي يتضمن أمثلة تعليمات حقيقية وتعقيبات بشرية حول جودة الاستجابة، بحيث يميل النموذج إلى طاعة الطلب، وتجنّب العبارات المؤذية، واعتماد نبرة متّزنة محترفة؛ ويُعرف هذا المسار بمواءمة التعليمات والتدريب بالتغذية الراجعة من البشر (Ouyang et al., 2022). عندما تكون التعليمات واضحة ومحدّدة، وتكون أكثر نفعًا كلّما وقّرنا لها سياقًا وعينات صغيرة لما نريد. لهذا السبب، يوصّف الاستخدام الرشيد لها بأنها "محرّر ومنظم" للنصوص والأفكار، لا "موسوعة نهائية" للحقائق (Bender et al., 2021).

تتبع قيمة هذه النماذج لعالم الاجتماع من قدرتها على تحويل الفكرة المبدئية المشتتة إلى منتج أولي منظم يمكن البناء عليه. عندما يشرع الطالب أو الباحث في مراجعة أدبيات موضوع واسع، يمكن للأداة أن تقترح مجموعة محاور رئيسية، وتشتق كلمات مفتاحية مناسبة لمسارات البحث اليدوي، وتنتج مسودّة فقرة تمهيدية توضّح ما هو معروف وما الذي لا يزال موضع سؤال. وفي مرحلة تصميم أدوات جمع البيانات، تستطيع أن تقترح مسودّات لدليل مقابلة نصف منظم تُراعي فيها حساسية الأسئلة، أو تولّد بنودًا أولية لمقياس "الاندماج الجامعي" موزّعة على

أبعاد معقولة، على أن تُراجع هذه المخرجات بعين بشرية خبيرة قبل اعتمادها. وفي التحليل النوعي الأولي، يمكنها أن تنظّم المقابلات الطويلة في جداول ترميز أولي تضمّ "أكوادًا" تفسيرية وثيمات مرشحة واقتباسات ممثّلة، كخطوة مساعدة تسبق التحليل النهائي. كما أنّها مجدية في الصياغة الأكاديمية: تنقيح النتائج، توحيد المصطلحات، وتكييف اللغة للجمهور الأكاديمي أو العام. (Salganik, 2017; Kvale & Brinkmann, 2009) غير أنّ هذه القوة التنظيمية لا تعني سلامة معرفية تلقائية؛ فالنموذج قد يُنتج معلومات تبدو صحيحة لكنها بلا سند، وقد يخلط بين مصادر أو يهلوس مراجع، كما أنّ لغته قد تعكس تحيّزات ثقافية وجندرية راسخة في النصوص التي دُرّب عليها. (Bender et al., 2021; Blodgett et al., 2020; Crawford, 2021) لذلك، يبقى تقييم البشر والتحقق المستقل من الحقائق شرطًا لا يُستغنى عنه قبل الاستشهاد الأكاديمي أو اتخاذ أي قرار يخصّ الناس.

تبدأ هندسة الأوامر من إدراك بسيط: النموذج لا يقرأ نواياك من بين السطور، بل يتبع الإشارات النبوية التي تقدّمها له في النص. عندما تكتب "اكتب عن البطالة"، فأنت تمنحه فضاءً واسعًا سيملؤه بجمل عامة. أمّا إذا حدّدت له الدور الذي تُريده أن يلعبه، كأن تقول "تخيّل أنك باحث اجتماعي يكتب فقرة تمهيدية لمقال أكاديمي"، فقد ضبطّ النبذة والمعايير. وإذا أوضحت المهمة بدقة، مثل "صياغة خمس أسئلة مقابلة حول تصوّرات شباب مدينة كذا للبطالة"، فقد قيّدت الناتج إلى شكل عملي. وعندما تضيف سياقًا واضحًا، كالفئة العمرية، والبيئة الحضرية أو الريفية، والزمن، والجمهور المقصود، فأنت تقلّص احتمال التعميمات الفارغة. وإذا سألت بلغة تراعي الحساسية الثقافية، وطلبت شكل الإخراج صراحةً—فقرة أكاديمية، أم جدولًا بعنوانين أعمدة محددة—فإنك تزيد من قابلية المخرج للاستخدام. وحين تُرفق مثالًا صغيرًا يمثل الشكل المطلوب، فأنت تُري النموذج النموذج الذي سيقّده، وقد أثبتت التجربة البحثية أن هذه العيّنات الصغيرة ترفع جودة النتائج بصورة ملحوظة لأنها تضبط النمط الذي يُحاكيه النموذج. (Brown et al., 2020; Mitchell, 2019) بهذه الروح، تتحول الكتابة إلى حوار بناء: صياغة أمر أول، قراءة نقدية للمخرج، إعادة كتابة الأمر بإضافة ما افنقر إليه من سياق أو قيود أو أمثلة، ثم تشغيل جديد وتوثيق الفروق. هذا التكرار الموجّه ليس مجرد حيلة تقنية، بل ممارسة منهجية تُشبه التحرير العلمي المتدرّج الذي نعرفه في الكتابة الأكاديمية. (Doshi-Velez & Kim, 2017; Lipton, 2018)

ومع أنّ إطار "الدور والمهمة والسياق والقيود وشكل الإخراج والأمثلة" عملي ويسير، إلّا أنّ المهام الكبيرة تتطلب غالبًا سلسلة من الأوامر بدل أمر واحد شامل. في مراجعة أدبيات معقّدة، قد تطلب أولاً تحديد المحاور الكبرى وتعريفها العاملة، ثم في أمر ثانٍ توليد قائمة بالكلمات المفتاحية وأسئلة البحث المفتوحة، ثم في أمر ثالث إنتاج فقرة تلخّص الثغرات النظرية والمنهجية، وفي أمر رابع اقتراح تصميم بيانات أولي. هذا التقسيم يحدّ من الأخطاء ويجعل كل خطوة قابلة للتقويم والإصلاح على حدة. وفي مسائل التحليل التي تتضمن استدلالًا، قد تطلب من

النموذج أن يكتب "خطوات التفكير" للوصول إلى النتيجة، مع تذكير مهم بأن "تفكير" النموذج **توسط لغوي** مفيد للتفكير وليس استدلالاً علمياً راسخاً؛ لذا يُستخدم هذا الأسلوب لإظهار الافتراضات وتنظيم الحجج، ثم يُراجع منطقيًا وبشريًا قبل الركون إليه. ويمكن، حين تتباين المخرجات، أن تنتج عدة مسودات ثم تجمع "خلاصة نقدية" تقارن بينها وتُبرز نقاط القوة والضعف، فتقلّ بذلك فرصة الارتهان لنص واحد قد يحوي ادّعاءات غير دقيقة (Wei et al., 2022; Wang et al., 2022).

يبقى السؤال العملي: **كيف نتحقق؟** يقترح النهج المسؤول مجموعة أسئلة ثابتة ترافق كل مخرج يولده النموذج. أولها سؤال الملاءمة: هل يخدم النص المطلوب فعلاً الغرض الذي حددها، أم أنه يدور حول الموضوع دون أن يقع فيه؟ وثانيها سؤال التغطية: هل تناول المحاور الأساسية التي اتفقنا عليها، أم ترك فراغات معرفية تجعل النص مضللاً بالسكوت؟ وثالثها سؤال الدقة الواقعية: هل يتضمن ادّعاءات يمكن التثبت منها بمصادر موثوقة؟ هنا تحديداً تبرز خطورة الاعتماد على "مراجع" مولدة؛ فالممارسة الرشيدة تُحتم علينا أن نذهب نحن لاحقاً إلى قواعد البيانات الأكاديمية أو التقارير الرسمية لنطابق أي إحالة قبل إدراجها. أما السؤال الرابع فهو سؤال الإجرائية: هل خرج النص بشكل **قابل للاستخدام فوراً**—فقرة سليمة في تقرير، جدول منضبط الأعمدة، مخطط بنود متماسك—أم أنه يتطلب إعادة صياغة شاقّة تتيح بقيمته الزمنية؟ (Doshi-Velez & Kim, 2017; Bender et al., 2021) بهذه الأسئلة الأربعة، يتحول التحقق إلى عادة منهجية، ويصبح "الإحساس الجيد بالنص" محكوماً بميزان واضح يمكن تدريسه وتقويمه.

لا يكتمل الحديث من دون الوقوف على الأخلاقيات والحوكمة. فالأصل أن لا تُدخل في الأداة بيانات تعريفية لأشخاص حقيقيين، وأن تُراعي النصوص المنتجة حساسية اللغة تجاه النوع الاجتماعي والهوية والثقافة، وأن تُراجع هذه الجوانب بشرياً قبل النشر خاصة عندما يتعلق الأمر بفئات هشّة. وعند استخدام النماذج في سياقات تتصل بالقرارات العامة، يصبح من الضروري توثيق كل شيء ببطاقات واضحة: **بطاقة بيانات** تصف مصدر البيانات وحدودها والجمهور المقصود، و**بطاقة نموذج** توضح الغرض الذي صُمّم من أجله، والبيانات التي تعلم منها، ومؤشرات أدائه العام وأدائه عبر الفئات، وحدود التعميم، وسيناريوهات الاستخدام غير المستحسن. هذه الوثائق ليست رفاهية إدارية؛ إنها **عقد اجتماعي مصغّر** يحدد حقوق الناس وواجباتهم ويؤسس للمساءلة عندما تسوء الأمور أو تظهر آثار غير مرغوبة (Gebru et al., 2021; Mitchell et al., 2019; Raji et al., 2020; Floridi et al., 2018).

ومن المهم أيضاً الحذر من الفخاخ الشائعة التي تتكرر في قاعات الدرس والممارسة على السواء. الفخ الأول هو **الغموض**: أمر عام ينتج نصوصاً عامة تبدو صحيحة لغوياً لكنها عاجزة عن خدمة غرض محدد. العلاج هنا هو بناء أمر غنيّ بالسياق والقيود وشكل الإخراج ومثال يُحتذى. الفخ الثاني هو **الاطمئنان الكاذب** إلى نصوص منسّقة كأنها حقيقة بديهية؛ والحلّ أن نفصل بين الجودة الأسلوبية وبين الدقة الواقعية، وأن لا ندخل ادّعاءً إلى تقرير أو إحالة إلى مرجع من غير تحقق مستقل. الفخ الثالث هو **تثبيت الانحياز** عبر قبول صياغات نمطية أو مصطلحات تُقضي فئات، وهو ما يستدعي مراجعة لغوية حسّاسة، وأحياناً إشراك المعنيين في المراجعة لضبط الأثر الرمزي للكلمات. والفخ الرابع هو **التضخيم غير المسؤول**، أي تحويل مسوِّدة مُساعدة إلى قرار سياساتي مباشر؛ والحل أن يبقى الإنسان في الحلقة، وأن تُقدّم المخرجات بوصفها اقتراحات أولية تحتاج دمجاً مع خبرة الميدان ونظرية الحقل قبل أن تُترجم إلى فعل. (Crawford, 2021; Raji et al., 2020)

ولكي تتجسّد هذه المبادئ في تجربة تعليمية، يمكن للمدرّس أن يهيئ تمريناً عملياً لا يقوم على التعداد بل على سرد مستمر: يُكلف الطلبة باختيار موضوع اجتماعي قريب—مثل بطالة الشباب أو الاندماج الجامعي—ثم يكتب كل منهم فقرة أمر مفصلة لاستدعاء “خريطة محاور أولية”، وبعد استلام المخرجات يكتب فقرة نقدية يجيب فيها عن أسئلة الملاءمة والتغطية والدقة الواقعية والإجرائية، ثم يعيد كتابة الأمر في فقرة ثانية أكثر إحصائياً تُبرز كيف تحسنت اللغة والسياق والقيود وشكل الإخراج. بهذا البناء السردية، يتعوّد الطلبة على أن هندسة الأوامر **تفكير** مكتوب أكثر منها تعويذة لغوية، وأن التحقق الأخلاقي والمعرفي جزء أصيل من الحرفة البحثية.

في الخلاصة، إنّ النماذج اللغوية الكبيرة أدوات قوية لتسريع التفكير وتنظيم المنتجات النصّية في العلوم الاجتماعية، لكنها أدوات تحتاج إلى حارس بشري يعرف حدودها ويحسن تسخيرها. عندما نكتب أوامراً مُحكمة غنية بالسياق، ونقسّم المهام الكبيرة إلى مراحل قابلة للتقويم، ونطلب أمثلة صغيرة تضبط النمط، ونمارس التحقق المنهجي من الحقائق والمراجع، ونوثق كل ذلك ببطاقات واضحة للبيانات والنموذج، تصبح هذه الأدوات **مُعِيناً** منهجياً يدفعنا نحو كتابة أدقّ وقرارات أكثر عدلاً، بدل أن تكون مصدرًا جديدًا للتضليل أو لتوسيع فجوات عدم المساواة. (Mitchell, 2019; Bommasani et al., 2021; Bender et al., 2021)

قائمة المراجع

1. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.

2. Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of “bias” in NLP. *Proceedings of the ACL*, 5454–5476.
3. Bommasani, R., Hudson, D. A., Adeli, E., et al. (2021). On the opportunities and risks of foundation models. *arXiv:2108.07258*.
4. Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. *NeurIPS*, 33, 1877–1901.
5. Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
6. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv:1702.08608*.
7. Floridi, L., Cows, J., Beltrametti, M., et al. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689–707.
8. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86–92.
9. Kvale, S., & Brinkmann, S. (2009). *InterViews: Learning the craft of qualitative research interviewing* (2nd ed.). Sage.
10. Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.
11. Lipton, Z. C. (2018). The mythos of model interpretability. *Queue*, 16(3), 31–57.
12. Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On faithfulness and factuality in abstractive summarization. *ACL*, 1906–1919.
13. Mitchell, M. (2019). *Artificial intelligence: A guide for thinking humans*. Farrar, Straus and Giroux.
14. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220–229).
15. Ouyang, L., Wu, J., Jiang, X., et al. (2022). Training language models to follow instructions with human feedback. *NeurIPS*, 35, 27730–27744.
16. Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *Proceedings of FAccT 2020*, 33–44.

17. Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
18. Salganik, M. J. (2017). *Bit by bit: Social research in the digital age*. Princeton University Press.
19. Suresh, H., & Gutttag, J. V. (2021). A framework for understanding sources of harm in machine learning. *EAAMO '21*.
20. Wang, X., Wei, J., Schuurmans, D., et al. (2022). Self-consistency improves chain-of-thought reasoning in language models. *arXiv:2203.11171*.
21. Wei, J., Wang, X., Schuurmans, D., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. *NeurIPS*, 35, 24824–24837.