

إشكاليّات اللغة العربية في عصر الذكاء الاصطناعيّ

حبيب عبد الرب سروري

Habib Abdulrab

<https://www.habibabdulrab.com>

Professeur Emérite, INSA de Rouen, Université de
Normandie, France



فهرس

- (١) تعريف الذكاء الاصطناعي
- (٢) تاريخ الذكاء الاصطناعي بكلمتين
- (٣) مجالات تفوق الذكاء الاصطناعي وتأخره على الذكاء الإنسانيّ حالياً
- (٤) الذكاء الاصطناعي «قيّد للماضي أم أفق للمستقبل»؟
- (٥) النواقص والعقبات الموروثة منذ دخول العربية العصر الرقمي
- (٦) التوليد الآلي للنصوص، ووضع العربية
- (٧) الترجمة الآلية، ووضع العربية
- (٨) آفاق ومقترحات لحضور لائق ومرموق للعربية

١) تعريف الذكاء الاصطناعي

- مجال في علوم الكمبيوتر هدفه محاكاة الكمبيوتر ملكات الدماغ الإنساني (في التعلّم، الاستيعاب، الاستنباط، البرهنة، حلّ الإشكاليات، واتخاذ القرارات...).
- تطبيقاته العملية:
- مختلف مجالات الحياة: الصحة، الإدارة، المواصلات، السياسة، التأليف، التعليم، أنظمة المعلومات...
- ما يهّمنا منها بشكل خاص هنا: معالجة اللغة الإنسانية: حوار، ترجمة...

(٢) تاريخ الذكاء الاصطناعي بكلمتين (١)

- ١٩٣٦، اختراع آلان تورنج لماكينة نظريّة (تجربة ذهنية) يمكن استخدامها للحلّ الآليّ لأي خوارزمية.
- التصميم العملي للكمبيوتر، ١٩٤٥: آلان تورنج، فان نيومان.
- تورنج نفسه مؤسس علم الذكاء الاصطناعي. راهن على امتلاك الكمبيوتر لذكاء الإنسان في نهاية القرن العشرين (امتحان تورنج).
- مؤتمر تأسيسي للعلم الجديد في ١٩٥٦: ماكارثي، مينسكي، شانون...

(٢) تاريخ الذكاء الاصطناعي بكلمتين (٢)

- تلتُهُ: عقود اللغات الذكيّة، المنظومات الخبيرة...
- شتاء الذكاء الاصطناعي (نهاية الثمانينات).
- انترنت وجبروت الكمبيوتر (التسعينات).
- ثمّ: تجليات طفرة تقنية «التعلّم العميق» ب «شبكة العصبونات الاصطناعية» ٢٠١٢، والانتصارات المتتالية للذكاء الاصطناعي بعد ذلك في مجالات عديدة: التعرف على الصور، لعبة الغو، الترجمة الآلية...

٣) مجالات تفوق الذكاء الاصطناعي وتأخره على الذكاء الإنساني حالياً

■ تفوقه:

الترجمة الآلية من ٢٠٠ لغة إلى ٢٠٠ لغة؛ تشخيص وعلاج الأمراض؛ هزيمة بطل العالم في أصعب لعبة عقلية: ألغو، ٢٠١٦...

■ تأخره:

تنظيم المهام: الطباخة، السفر...؛ الذكاء غير الموجه...

■ الذكاء ليس مجالاً واحداً نمطياً خطياً، والتفاعل بين الذكاء الاصطناعي والإنساني دياكتيكي، زجاجي، يتطور باستمرار (علاقتها ليست مانوية).

٤) سؤال المؤتمر: الذكاء الاصطناعي «قيّد للماضي أم أفق للمستقبل»؟

- الذكاء الاصطناعي أفق علم الكمبيوتر منذ ولادته (تكملة الطبيعة).
 - السؤال أعلاه يعادل: هل الكمبيوتر «قيّد للماضي»؟ بالطبع، لا!
 - ← عبارة «قيّد للماضي» لا محل لها من الإعراب! لكن هناك مخاطر تحتاج إلى ضبط **Régulation**:
 - يلزم قوينة شفافية ومجانية مدونات اللغة Corpus، ومآذجها LLM (التي يستند عليهما تعليم الآلة)، لتلافي الكثير من المخاطر.
- إذ كيف يمكن دراسة ردود الكمبيوتر وتحليل أخطائها دون معرفة
مكونات مكوّنات المدونة ونموذج اللغة؟

٥) أهم النواقص الموروثة اليوم منذ دخول العربية العصر الرقمي (١)

■ في مقال في ٢٠٠٧، بعنوان: «اللغة العربية في الزمن الرقمي: ٦ فجائع و٣ مقترحات» تناولتُ بعض النواقص التي يلزم استكمالها سريعا لانجاز البناء التحتي لدخول اللّغة العربية العصر الرقمي. ما زال بعضها غائبا أو ناقصا حتّى الآن. منها:

■ (١) Corpus

■ (٢) القارئ الضوئي للأحرف OCR

■ (٣) انسحاب اللغة العربية من كتابة التكنولوجيا والعلوم

٥) أهم النواقص الموروثة اليوم منذ دخول العربية العصر الرقمي (٢)

Corpus (١)

■ مدوَّنةُ أية لغة، (Corpus)، هي مجموعةٌ هائلةٌ (كانت تُعدُّ كلماتها بالمليارات عند كتابة المقال، بآلاف المليارات اليوم) من عيّنات النصوص المكتوبة أو المنطوقة، الآتية من قطاع متنوع عريض محايد من المصادر:

الكتب القديمة والحديثة، ويكيبيديا، نصوص المؤسسات الدولية، وما يرفده الإنترنت يوميا من صحف ومجلات، نقاشات، تقارير، مواقع إنترنت... تتلقفها وتؤرشفها مشاريع دولية مثل paraCrawl.eu...

٥) أهم النواقص الموروثة اليوم منذ دخول العربية العصر الرقمي (٣)

■ هي المختبر الذي تُستخلصُ منه الدراسات الاجتماعية والتاريخية واللغوية المتنوعة، وتتكئ عليه وسائل البحث المتقدمة مثل Ngram لغوغل (غائب عربياً!)...

■ وفي عصر الذكاء الاصطناعي صارت مدونة اللغة Corpus المادة الخام الرئيسة التي يتدرَّب عليها نموذج اللغة LLM، لتوليد النصوص آلياً للإجابة على أي سؤال. هي أيضاً مرجعيته للترجمة الآلية.

■ المفارقة الكبرى: العربية، لغة القواميس والمعاجم الأولى، ما زالت تفتقد إلى مدونةٍ تليق بها اليوم!

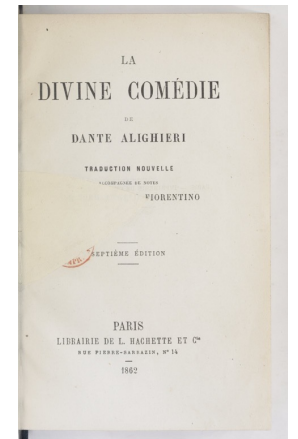
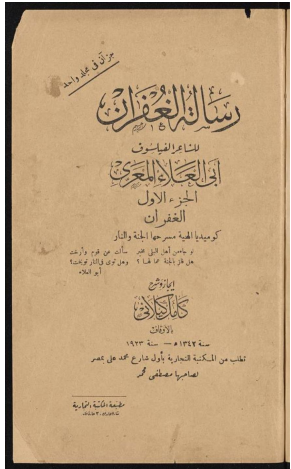
٥) أهم النواقص الموروثة اليوم منذ دخول العربية العصر الرقمي (٤)

٢) OCR

■ القارئ الضوئي للأحرف OCR يقوم بتحويل الكتب الورقية إلى نصوص رقمية. (في ٢٠٠٧، كانت اللغات الكبرى قد استكملت رقمه ملايين الكتب القديمة).

■ سمح ذلك حينها مثلا بتحويل كتاب «الكوميديا الإلهية» الورقي لدانتي إلى نص رقمي، في حين لا يمكن العثور على «رسالة الغفران» للمعري إلى اليوم إلا كصورة PDF!

■ التأخر في إنجاز رقمنة الكتب القديمة بالعربية معضلة قومية، قادت إلى ضعف موارد ذخائرها الرقمية، ومدونتها Corpus.



٥) أهم النواقص الموروثة اليوم منذ دخول العربية العصر الرقمي (٥)

٣) انسحاب اللغة العربية من كتابة التكنولوجيا والعلوم

- المواد العلمية تُدرّسُ باللغات الأجنبية.
- صار يُقال: لغة الضاد «لا تصلح للحدثاثة»!؟
- تفتقر إلى مفردات علمية وصيغ تعبيرية موحّدة.
- النصوص التوليدية والترجمة الآلية لا يمكنهما اختراع كلماتٍ وصيغٍ جُمَلٍ من العدم!

٥) أهم النواقص الموروثة اليوم منذ دخول العربية العصر الرقمي (٦)

٣) انسحاب اللغة العربية من كتابة التكنولوجيا والعلوم

- المفارقة: كانت لغة الخوارزمي وابن رشد لغة العلوم والفلسفة بامتياز.
- لها إمكانيات مرنة متميِّزة لاحتضان الجديد مثل:

Numérisation: رقمنة (انطلاقاً من صيغة «فعلنة» التي

تعكس التوليد والتحوّل: عقلنة، رهبنة، تيسنة...)،

Dématérialisation: استرقام، انطلاقاً من صيغة

«استفعل» التي تعني طلب الشيء واستدعاءه.

٦) التوليد الآلي للنصوص، ووضع العربية (١)

■ توليد النصوص الآلي يتكئ على راسيتين: مدونة اللغة التي ينبغي أن تكون ثرية متنوعة، ونموذج اللغة LLM الذي يتعلم من المدونة، عبر تقنية التعلم العميق بشبكات العصبونات الاصطناعية.

■ يسمح ذلك لمولد النصوص أن يجد أكثر الكلمات احتمالا لاستطراد نص ما أو مجموعة كلمات، ثم أكثر الكلمات احتمالا لمواصلة النص بعدها، وهكذا دواليك.

■ مثال: ميلانو إحدى أروع...

■ ← ... ميلانو إحدى أروع المدن...

٦) التوليد الآلي للنصوص، ووضع العربية (٢)

- عند توجيه الأسئلة المرتبطة بالثقافة العربية هناك فجوة هائلة، لأن مدونة لغة الضاد «ضعيفة الموارد والتموين» mal dot é.
- مثال: بعكس ردود ChatGPT المدهشة غالبا (والذي لا تخفي عليه خافية حول دانتى وكتابه)، عند استفساره عن رأي المعري بالمتنبي مثلا، يقول: «من الصعب تحديد الآراء الدقيقة للمعري بشأن المتنبي بدقة، نظرا لنقص المعلومات».
- وعن حوارات الجنة في رسالة الغفران، يرد: «لم يذكر المعري تفاصيل محدّدة عن الجنة»، أو «لا توجد إشارة إلى حوارات في رسالة الغفران مع شخصيات تاريخية».

٦) التوليد الآلي للنصوص، ووضع العربية (٣)

- يختلف الأمر مع أربيك جي بي تي (جيس) الذي تمّ رفضه بنصوص ثقافية عربية عن المعري ورسالة الغفران: ردّه جيّد حول رأي المعري بالمتنبي.
- لكن عند سؤال مثل: «كيف وجد ابن القارح جمال عبدالناصر في جنة رسالة الغفران؟».
- يردّ: «في رسالة الغفران ينطلق ابن القارح في رحلة رائعة إلى الآخرة يواجه فيها شخصيات أدبية مختلفة، وخلال رحلته يلاقي جمال عبدالناصر أحد الشخصيات البارزة التي كان لها تأثير على العالم العربي... يعكس إدراج جمال عبدالناصر في النص تقدير المعري للشخصيات السياسية الهامة وتأييده لها...!»

(٧) الترجمة الآلية اليوم، ووضع العربية (١)

- كانت الترجمة حتى أمد قريب تنطلق من القواعد النحويّة:
أي من القاعدة النحوية للجملة ج ١ في لغةٍ ما، إلى القاعدة النحوية المقابلة لها في لغةٍ ثانية، ثمّ بتّ ترجمة مفردات ج ١ في مواضعها في القاعدة المقابلة.
- منذ ٢٠١٧ اكتسحت تقنية المحوّل الآلي (**Transformer : Encoder, Decoder**) كل تقنيات الترجمة الآلية.
- تعتمد على نموذج اللغة الذي يتمّ تدريبه طويلا على مدوّنة اللغة، وعلى تقنية الإصغاء Attention للجملة.

(٧) الترجمة الآلية اليوم، ووضع العربية (٢)

- النتيجة: في اللغات ذات المدونات الغنية والنماذج المدربة: النتائج مذهلة، تفوق توقعات معايير التقييم التقليدية.
- ترجمة ١٠٠ لغة إلى ١٠٠ لغة، كما هو حال غوغل ترانسليت، لا يحتاج إلى $١٠٠ * ٩٩$ منظومة ترجمة، لكن إلى ٩٩ فقط: يكفي استخدام لغة وسيطة.
- كذلك حال مشروع فيسبوك nllb-200 لترجمة ٢٠٠ لغة إلى ٢٠٠ لغة!
- في العربية: النتائج ضعيفة نسبياً لمن يقارنها باللغات ذات التمويل العالي، وإن يمكن ملاحظة تطوّر سنويّ.

(٧) الترجمة الآلية اليوم، ووضع العربية (٣)

■ أمثلة: ترجمة: «فصل الربيع الجميل. فصل من الكتاب المبين»، غوغل:

a **chapter** from the clear book. The beautiful spring season

■ ترجمة « جوزي قال لي»، غوغل:

Ma muscade m'a dit : ٢٠٢٢ قبل

Josie m'a dit : ٢٠٢٢ في

Mon Mari m'a dit : اليوم

■ ماذا عن ترجمة: «فِتْنَةُ الفِتْنَةِ»؟

٨ آفاق ومقترحات لِحضورٍ لائقٍ ومرموقٍ للعربية (١)

■ إذا نريد الحفاظ على ثقافتنا العربية في عصر الذكاء الاصطناعي،
واللحاق بالعصر، فيلزمنا سريعا:

➡ ردم الهوة الشاسعة في مجال ترجمة العلوم وكتابتها، عبر
تطوير وإثراء القاموس العلميّ العربي، واستخدام اللغة العربية في
كتابة وتدريس العلوم والمعارف عامّة في المدارس والجامعات،
وتطوير مشاريع قومية لبوابات علمية نموذجية على الإنترنت في
مختلف الدروس العلمية والتكنولوجية.

٨ آفاق ومقترحات لِحضورٍ لائقٍ ومرموقٍ للعربية (٢)

↩ بناء LLM مُدربة على مدونة ثرية للغة العربية (وللغات المنطقة الحيّة الأخرى كالأمازيغية والسريانية والكردية) وما يضمن حفظ اللغات القديمة كالسّقطريّة والشحريّة والمهريّة وغيرهم.

ذلك يتضمّن ويستلزم:

← ضرورة إدخال معظم الكتب القديمة والحديثة، ونماذج الأحاديث الشفوية لجميع لغات ولهجات منطقتنا، كنصوص رقمية في المدونة، كما فعلت وتفعل شعوبُ اللغاتِ ثريةِ المواردِ الرقميةِ.

٨ آفاق ومقترحات لِحضورِ لائقٍ ومرموقٍ للعربية (٣)

➡ تأسيس مشاريع مرتبطة بخصوصيات لغة الضاد، مثل «منظومة الإعراب الآلي». سيساعد ذلك على تصحيح وإكمال تشكيل حروف النصوص، تحسين ترجمتها الآلية واستيعاب دلالاتها ومعانيها آليًا...

■ إذ لا يتوقّف الإعراب عند تحليل تركيب وتحريك وعلاقات الألفاظ، لكنه يمتدّ إلى الدلالة! (إعراب «هنا لندن» أمودجا).

■ قواعد السنسكريتية لِ بانيني تشمل ٤٠٠٠ قاعدة نحوية،

في كتابه Astadhyayi (٥ قرون قبل الميلاد)،

تمّ برمجتها آليًا في ال INRIA!



ملاحظة نهائية حول الذكاء الاصطناعي التوليدي (١)

■ استيعاب العالم وامتلاك المقدرة على التحليل والبرهنة والتخطيط هو الذكاء الحقيقي (أكثر من موهبة «البغاء الإحصائية» في إعادة صياغة وفذلكة المعارف، كما تفعل منظومات الذكاء الاصطناعي التوليدي).

■ سؤال: هل يمكن لهذه المنظومات أن تنطلق مثلا من معطيات أبيات لغز شعري قديم، لتشرح، باستدلالٍ آليٍّ وسيناريو تفصيلي، كيف يمكن أن يكون الرجل عَمًّا لِعَمَّتِهِ؟!

■ ماذا تردّ المنظومات التوليديّة على هذا السؤال؟

ملاحظة نهائية حول الذكاء الاصطناعي التوليدي (٢)

لي عمّة وأنا عمّها ولي خالة وأنا خالها
فأمّا التي أنا عمّ لها فإنّ أبي أمّه أمّها
أبوها أخي وأخوها أبي ولي خالة وكذا حكمها

خارطة الطريق لذكاء اصطناعيّ توليديّ جديرٌ باسمه هنا:
(١) تحويل الأبيات إلى برنامج كمبيوتر بلغة ذكية كـ
Prolog، تأخذ معطيات الأبيات كحقائق Fact، تضيف لها
قواعد منطقية Rules لشرح ما يعني: عمّ، خال...

ملاحظة نهائية حول الذكاء الاصطناعي التوليدي (٣)

(٢) تضيف لها قواعد أخرى لمعارف ضمنية مثل:

أ) إمكانية الطلاق؛

ب) للإنجاب عمرٌ محدد (ليست كل النساء مثل جدتنا سارة!);

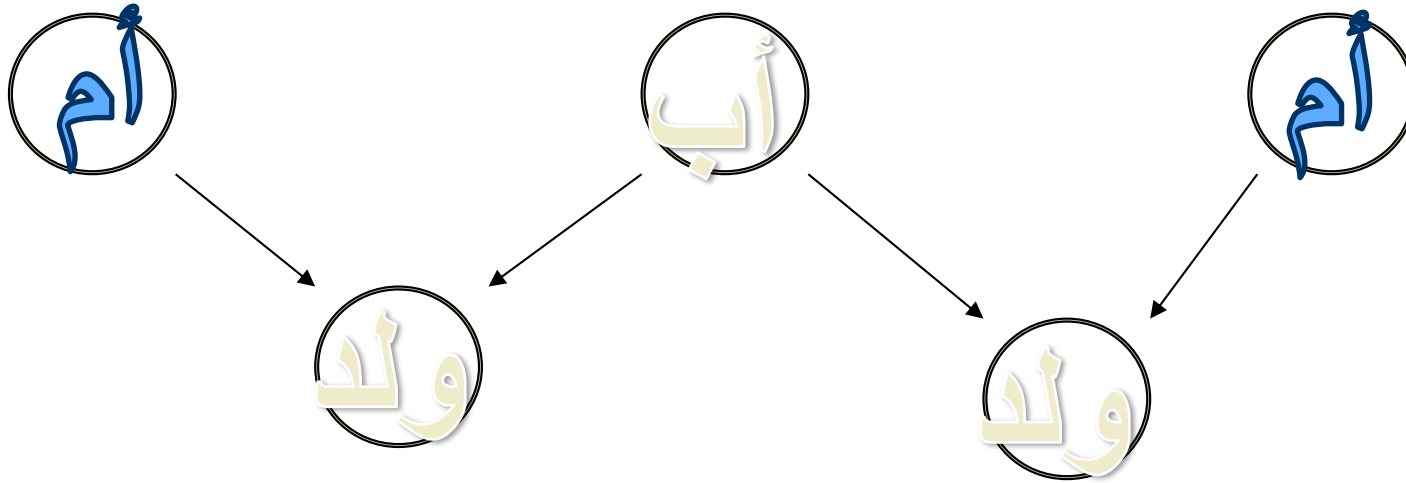
ج) ولا يأتي الإنجاب من معجزات (ليست كل النساء مثل أمنا مريم: «ابنة ابنها» كما قال دانتلي);

د) عدم اعتبار العلاقات المحرّمة كانتهاك المحارم...

(٣) تجيب على السؤال عبر الاستدلال المنطقي Logic inference للغة برنامج الكمبيوتر الذكية.

شفافية هذا البرنامج وكشفه للمستخدم ضرورية.

ملاحظة نهائية حول الذكاء الاصطناعي التوليدي (٤)



بعد حوالي 17 سنة، يتزوج كل ولد من أم أخيه (من الأب)
بعد طلاق أبيه منها!

ملاحظة نهائية حول الذكاء الاصطناعي التوليدي (٥)

