الحكم على صحة مخرجات النماذج اللغوية

**مقدمة:** مع الشهرة الكبيرة التي حظيت بها النماذج اللغوية الضخمة (مثل ChatGPT, GPT4)، وكثرة اعتماد المستخدمين على نتائجها؛ كان هناك حاجة **لتقييم صحة الحقائق** التي تقوم هذه النماذج اللغوية بإخراجها. الهدف من ذلك هو معرفة مدى صحة مخرجات النماذج اللغوية، وفي أي مواضيع تكون دقة المخرجات ضعيفة؛ وهذا من الممكن أن يساعد في اكتشاف الخلل أو التحيز في بيانات التدريب، وتحسين ذلك في المستقبل.

يمكن لاحقًا بناء برنامج لتنبيه المستخدمين عند اكتشاف معلومات غير صحيحة في مخرجات النماذج اللغوية.

**مثال:** عندما سألنا ChatGPT عن أسرع حيوان على اليابسة، كانت الإجابة: الجمل هو أسرع حيوان بسرعة 120 كم/ساعة.

إذا بحثنا في غوغل عن نفس السؤال، ستظهر أول إجابة من موقع **ناشيونال جيوغرافيك** كالتالي:

يعد الفهد أسرع حيوان على الأرض بسرعة 103 كم/ساعة.

هنا، كانت المعلومات التي أنتجها ChatGPT غير صحيحة.

تجهيز البيانات: لمحاولة استخلاص المعلومات من داخل ChatGPT و GPT4، قمنا بسؤال كل منهما أن يقوم بتأليف جُمل متنوعة تحتوي على كلمات مختلفة (مثل: حضارة، سرعة، إنشاء، جائزة..) بحيث تحتوي كل جملة على معلومات يمكن التحقق منها.

نريد الحكم على كل جملة من حيث كونها تحتوي على معلومات يمكن التحقق منها، وأيضا ما إذا كانت المعلومات صحيحة أم لا.

كذلك نريد ذكر رابط الموقع الذي ساعدنا في الحكم على صحة المعلومات، وإعادة كتابة الجملة بعد تصويب الأخطاء بها إن وجدت.

نعطي هنا أمثلة لكل نوع (1: معناها "نعم"، 0: معناها "لا")

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| الجملة | يمكن التحقق منها؟ | معلومات صحيحة | الجملة بعد تصويب الأخطاء | رابط المرجع للتأكد |
| ما هي أكبر دولة عربية من حيث المساحة؟ | **0** | - |  |  |
| الجمل أسرع حيوان بسرعة 120 كم/ساعة | 1 | **0** | الفهد أسرع حيوان بسرعة 103 كم/ساعة | https://ngalarabiya.com/article/4217654/ |
| برج خليفة أنشئ عام 2010 وهو أعلى مبنى في العالم بارتفاع 828 م | 1 | 1 |  | <https://ar.wikipedia.org/wiki/> برج\_خليفة |

التجهيز للعمل:

1. تم تجهيز جمل في ملفات إكسيل، يحتوي كل منها على 250 جملة. أمام كل جملة، هناك رقم مميز لها، ثم 4 أعمدة لإدخال بيانات التوسيم وهي: "يمكن التحقق منها؟"، "معلومات صحيحة"، "الجملة بعد تصويب الأخطاء"، و"رابط المرجع للتأكد" كما هو موضح.



2. المطلوب عدم حذف أو إضافة أي جملة، وأيضا عدم تغيير رقم الجملة أو نص الجملة. نريد فقط الكتابة في أعمدة التوسيم الأربعة.

3. يمكن فتح الملف ثم تكبير عمود "الجملة" لكي ترى الجملة كاملة. الرجاء قراءتها حتى النهاية والحكم عليها بعد قراءتها بدقة كبيرة.

سيتم **تقييم جودة التوسيم** لكل عنصر من عناصر التقييم، وذلك عن طريق منهجية علمية.

4. لا يجب أبدا استخدام ChatGPT أو GPT4 عند التأكد من صحة المعلومات. يمكننا استخراج كلمات مفتاحية من الجملة والبحث في غوغل، ثم ذكر رابط الموقع الذي تم استخدامه.

مثلا بالنسبة لهذه الجملة: "يعد الإسلام ثاني أكبر ديانة في العالم بعد المسيحية"

يمكننا الذهاب إلى غوغل والبحث عن الكلمة المفتاحية " الأديان في العالم"، فسنحصل على قائمة بالديانات وعدد معتنقيها. ومن هذه الصفحة (قائمة التعداد الديني - ويكيبيديا) يمكن تأكيد صحة الجملة: <https://ar.wikipedia.org/wiki/%D9%82%D8%A7%D8%A6%D9%85%D8%A9_%D8%A7%D9%84%D8%AA%D8%B9%D8%AF%D8%A7%D8%AF_%D8%A7%D9%84%D8%AF%D9%8A%D9%86%D9%8A>

ثم نقوم بكتابة رابط هذه الصفحة في عمود "رابط المرجع للتأكد".

5. عند وجود أكثر من معلومة داخل الجملة، يجب التأكد منها جميعا.

مثلا: إذا كانت الجملة هي: "برج خليفة أنشئ عام 2010 وهو أعلى مبنى في العالم بارتفاع 828 مترا"

هنا يجب التأكد من كل هذه المعلومات:

**أ.** هل تم إنشاء برج خليفة عام 2010؟ وأيضًا:

**ب.** هل برج خليفة أعلى مبنى في العالم؟ وأيضًا:

**ج.** هل ارتفاع برج خليفة 828 مترا؟

إذا كانت كل المعلومات صحيحة، نضع "1" في عمود "معلومات صحيحة"، وإلا نضع "0" عند وجود أي معلومة خاطئة.

6. عند البحث على موقع غوغل، يجب أن نبدأ بالصفحات العربية التي تظهر قبل الصفحات الإنجليزية، وأن نراعي أن تكون الصفحات التي نستخدمها في التحقق من المعلومات موثوقة (مثلا: ويكيبيديا العربية).

مثلا عند البحث على "طول البحر الأحمر"، نجد أن المعلومة في ويكيبيديا العربية مختلفة عن ويكيبيديا الإنجليزية. هنا، نأخذ في الاعتبار الصفحة العربية. من المحتمل أن تكون هناك فروق ثقافية بين الصفحات العربية والإنجليزية. لذلك، نفضل العربية.

7. في عمود "الجملة بعد تصويب الأخطاء"، يجب استخدام لغة فصيحة بدون أخطاء وأن تكون الجمل مكتملة.

مثلا للجملة التالية: "المملكة العربية السعودية هي أكبر دولة عربية من حيث المساحة. "

هذه الجملة غير صحيحة، ولذلك نضع "0" في عمود "معلومات صحيحة".

في عمود "الجملة بعد تصويب الأخطاء"، يمكن كتابة: "الجزائر هي أكبر دولة عربية من حيث المساحة" **أو** " المملكة العربية السعودية هي ثاني أكبر دولة عربية من حيث المساحة". ثم نضع رابط لصفحة مساحات الدول العربية مثلا.

مثال آخر، في الجملة "جواد الزريقي هو لاعب كرة قدم مغربي يلعب في مركز المدافع." بعد البحث على غوغل، سنجد أن هذه الجملة غير صحيحة. في عمود "الجملة بعد تصويب الأخطاء"، يمكن كتابة: "لا يوجد لاعب كرة قدم مغربي بهذا الاسم"

8. الجمل التي تحتوي على آراء، أو أن معلوماتها غير كافية للحكم على صحتها، يمكن وضع "0" في عمود "يمكن التحقق منها". أمثلة:

. الحضارة الفينيقية كان لها تأثير كبير على التجارة والفنون. (يمكن اعتبار أن هذا رأي من الصعب التحقق منه)

. قال الرئيس إن الدولة تريد رفع مستوى المواطن. (من هو الرئيس؟ ما هي الدولة؟ متى قال ذلك؟ هذا كلام عام يقوله كل المسئولين ولا يمكن التحقق منه، وغير مهم التحقق منه من الأساس)

. أشارت دراسة حديثة أن هناك نقص مياه في الوطن العربي. (ما هي هذه الدراسة الحديثة؟ أين ومتى نُشرت؟ وهكذا)

9. يمكن السماح ببعض الاختلاف في البيانات المتغيرة باستمرار، مثل عدد سكان دولة ما، أو ذكر البيانات التقديرية بشكل تقريبي.

مثلا: إذا كانت الجملة "الكرة الأرضية تبعد عن الشمس مسافة تقريبية قدرها 150 مليون كيلومتر".

موقع ويكيبيديا يقول إن الأرض تبعد عن الشمس بمسافة 149.6 مليون كيلومتر، وعلى غوغل، الإجابة هي 150.09 مليون كيلومتر.

هنا يمكن اعتبار أن كل هذه الإجابات صحيحة، ونضع "1" في عمود "معلومات صحيحة"

مثال آخر: الجملة "تبلغ مساحة مصر حوالي 1,010,408 كيلومتر مربع".

إذا سألنا موقع غوغل، ستكون الإجابة 1.002 مليون كم مربع. يمكننا القبول بهذه الفروق البسيطة، واعتبار أن الجملة صحيحة.

**في المقابل**، إذا كانت الجملة "توفي النبي محمد -صلى الله عليه وسلم- في العام العاشر من الهجرة"، فهنا يجب اعتبار الجملة خاطئة لأن عام الوفاة هو الحادي عشر من الهجرة.

10. من المهم تذكّر أن هذه النماذج الضخمة تم تدريبها على بيانات يرجع تاريخها إلى سبتمبر 2021. فعندما يقول النموذج مثلا "هذا العام" أو "حاليا/الآن/.."، هنا من الممكن أن نفترض أن التاريخ هو سبتمبر عام 2021.

مثلا: "يلعب كريم بنزيمة في نادي ريال مدريد هذا العام"

حسب تاريخ سبتمبر 2021، فإن هذه المعلومة صحيحة. يمكن كتابة ملاحظة أن هذه المعلومة صحيحة حتى تاريخ سبتمبر 2021 في عمود "ملاحظات" بعد أعمدة التوسيم. بشكل عام، يمكن كتابة أي تعليقات أو ملاحظات في العمود "ملاحظات" للاستفادة منها.

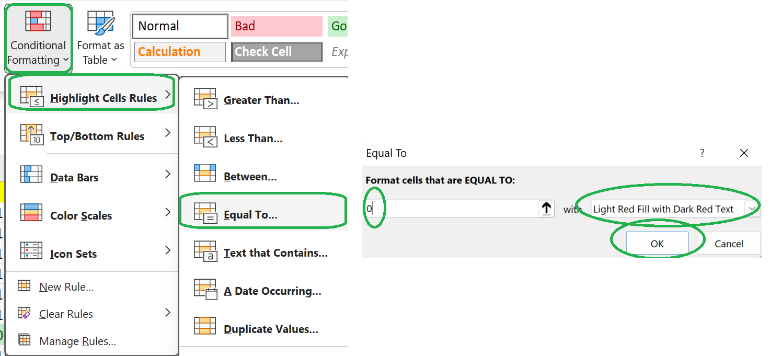
11. إذا وجدت أخطاء لغوية (إملائية أو نحوية أو منطقية) في الجملة، فيمكن تمييزها بوضع علامة "1" في عمود "خطأ لغوي" في نهاية الأعمدة. إذا كانت الجملة صحيحة، لا داعي لوضع "0" في العمود. هذه أمثلة من الجمل غير السليمة لغويا، أو غير المكتملة:

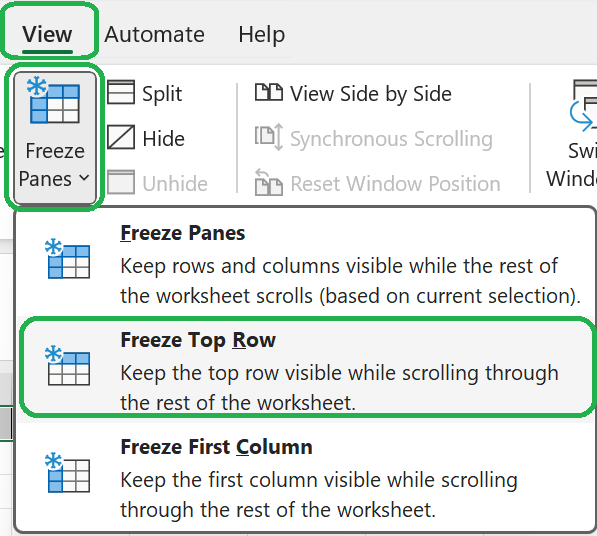
. هو ماري كوري هو أول شخص يفوز بجائزة نوبل في مجالين مختلفين. النفط مورد هام عن الإيرادات. (خطأ نحوي)

. قابلت صديقي الذي له 3 عيون. أصلي الجمعة يوم الخميس. الثعلب يرتدي ذيلا طويلا. (خطأ منطقي يعارض المعرفة العامة)

. هو الحيوان الوحيد الذي يمكنه البكاء بالدموع. (الجملة غير مكتملة الأركان. ما هو هذا الحيوان؟)

12. لمزيد من التوضيح أثناء التوسيم، يمكن تلوين الخلايا بألوان مختلفة بناء على القيم المكتوبة داخلها. يمكن استخدام خاصية "Conditional Formatting" داخل الإكسيل كما هو موضح. نبدأ باختيار عمود معين، مثلا "معلومات صحيحة"، ثم نختار من القائمة العلوية خاصية التنسيق المشروط "Conditional Formatting"، ثم نختار أن يتم تلوين الخلية باللون الأحمر إذا كانت تحتوي على القيمة "0" وذلك معناه أن الخلية تحتوي على معلومات غير صحيحة. وبالتالي تظهر المعلومات غير الصحيحة مظللة باللون الأحمر في العمود المختار بأكمله.



13. يمكن إظهار الصف الأول في الملف (الذي يحتوي على أسماء الأعمدة) وتثبيته بشكل دائم، وذلك عن طريق اختيار "تجميد الصف الأول" من قائمة "معاينة View" كما هو موضح.

14. لضمان عدم تغيير عمود رقم الجملة وعمود نص الجملة، يمكن اتباع الخطوات البسيطة في الفيديو القصير التالي:

<https://www.youtube.com/watch?v=VMaoPtVcPhk>