

تأثير الذكاء الاصطناعي على جودة اللسان العربي

أم كلثوم حكوم داود بن يحيى¹، محمد مدياني²

¹قسم الفقه وأصوله كلية الشريعة، جامعة الملك خالد (المملكة العربية السعودية)، obenyahia@kku.edu.sa

²كلية تقنية المعلومات، جامعة الإمارات العربية المتحدة (الإمارات العربية المتحدة)،

mohammed.mediani@uaeu.ac.ae

تاريخ الإرسال: 2025/05/24

تاريخ القبول: 2025/06/02

تاريخ النشر: جويلية 2026

ملخص:

يهدف هذا البحث إلى دراسة تأثير نموذج الذكاء الاصطناعي التوليدي ChatGPT على جودة اللسان العربي، وذلك من خلال تحليل النصوص التي ينتجها ChatGPT من حيث الدقة، الاتساق، والتنوع اللغوي، مع مقارنتها بالنصوص الطبيعية المكتوبة من قبل البشر.

كما يتناول البحث أبرز التحديات التي يواجهها النموذج عند معالجة اللغة العربية، مثل الأخطاء النحوية، عدم الدقة في التراكيب، والاعتماد على مصادر بيانات محدودة.

واقترضت طبيعة البحث الاعتماد على المنهج الوصفي التحليلي التجريبي، حيث سيتم تطبيق تجارب على عينة من النصوص المولدة بواسطة ChatGPT ومقارنتها مع نصوص طبيعية باستخدام أدوات تحليل لغوي متخصصة.

وكما سيتم تقييم النتائج من قبل مجموعة من خبراء اللغة العربية لقياس جودة النصوص، ومدى تطابقها مع القواعد النحوية والصرفية، بالإضافة إلى تنوع المفردات واتساق المعاني.

الكلمات المفتاحية: الذكاء الاصطناعي، النماذج اللغوية الكبرى، جودة اللغة العربية، الذكاء الاصطناعي التوليدي، معالجة اللغة الطبيعية.

Abstract:

This study aims to examine the impact of the generative AI model **ChatGPT** on the quality of the Arabic language. The research focuses on analyzing ChatGPT-generated texts in terms of accuracy, coherence, and lexical diversity, comparing them with naturally written human texts. It also explores the key challenges ChatGPT faces when processing Arabic, such as grammatical errors, structural inaccuracies, and reliance on limited data sources.

The study adopts a **descriptive, analytical, and experimental methodology**, where various experiments will be conducted on a sample of texts generated by ChatGPT. These texts will be compared with natural texts using specialized linguistic analysis tools. The results will be evaluated by Arabic language experts to measure the quality of generated texts, their compliance with grammatical and morphological rules, and their lexical diversity and semantic consistency.

Keywords: Artificial Intelligence, Large Language Models, Arabic Language Quality, Generative AI, Natural Language Processing.

Acknowledgement:

the authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Group Project under grant number (RGP/2/302/45)

مدخل (INTRODUCTION):

شهد العالم في السنوات الأخيرة تطوراً كبيراً في مجال الذكاء الاصطناعي، وبرز بشكل خاص الذكاء الاصطناعي التوليدي الذي أحدث نقلة نوعية في قدرة الأنظمة الحاسوبية على فهم اللغة الطبيعية وإنتاجها. يعتمد هذا النوع من الذكاء على النماذج اللغوية الكبرى، وهي شبكات عصبية عميقة يتم تدريبها باستخدام كميات هائلة من النصوص المأخوذة من مصادر متعددة مثل الكتب والمقالات والمحتويات الرقمية المتوفرة على الإنترنت، ولعل أبرز هذه النماذج ChatGPT و DALL-E من شركة OpenAI، بالإضافة إلى نموذج Gemini من Google*.

تتميز النماذج اللغوية الكبرى بقدرتها على إنتاج نصوص إبداعية، وتلخيص المعلومات، والإجابة عن الأسئلة بطريقة تحاكي الإنسان، وإن كانت هذه النماذج حققت نقلة نوعية في التعامل مع عدة لغات، إلا أنها تواجه تحديات كبيرة عند التعامل مع اللغة العربية، اللغة الغنية والمعقدة في نفس الوقت بسبب محدودية جودة -البيانات المستخدمة في تدريبها مقارنة باللغات الأخرى مثل الإنجليزية.

وعلى الرغم من الجهود المبذولة لتطوير نماذج تدعم اللغة العربية، مثل نموذج "عليم"، ونموذج "جيس"، إلا أن النصوص المنتجة باللغة العربية غالباً ما تكون أقل دقة وثراءً لغوياً مقارنة بالنصوص الطبيعية.

أهمية البحث:

تبرز أهمية هذا البحث في كونه محاولة لتسليط الضوء للوقوف على الأثر الذي يمكن أن تحدثه النماذج اللغوية الكبرى على اللغة العربية، وكذا الوقوف على التحديات التي تواجه العملية، ثم محاولة تقديم مقترحات لتحسين جودة النصوص المولدة وتقليل الفجوة بين النصوص الطبيعية والمولدة آلياً.

أسئلة البحث:

1. ما مدى تأثير النماذج اللغوية الكبرى على جودة اللسان العربي؟
2. ما أبرز التحديات التي تواجه تدريب النماذج اللغوية الكبرى على اللسان العربي؟
3. كيف يمكن تحسين جودة البيانات المستخدمة لتدريب النماذج اللغوية العربية؟

أهداف البحث:

1. تحليل تأثير النماذج اللغوية الكبرى على جودة النصوص العربية.
2. مناقشة التحديات المرتبطة بجمع وتدريب البيانات للنماذج العربية.
3. تقديم توصيات لتحسين جودة النماذج اللغوية بما يخدم تطوير اللغة العربية.

الدراسات السابقة:

تعددت الدراسات التي تناولت الذكاء التوليدي في اللغات الأجنبية، حيث ركزت معظمها على لغات مثل الإنجليزية والفرنسية، نظراً لتوافر بيانات ضخمة لهذه اللغات وسهولة معالجتها بواسطة تقنيات الذكاء الاصطناعي.

في المقابل، لا تزال الدراسات التي تتناول الذكاء التوليدي في اللغة العربية محدودة، نظراً لطبيعة اللغة العربية التي تتميز بتراكيب نحوية معقدة، ونظام صرفي غني، ووجود اختلافات بين الفصحى واللهجات العامية، كل هذه الفروقات تجعل من الصعب تعميم نتائج الدراسات الأجنبية على اللغة العربية، مما يستدعي إجراء أبحاث متخصصة تأخذ بعين الاعتبار خصوصية البنية اللغوية العربية وتحديات معالجتها بواسطة الذكاء الاصطناعي.

ومن الدراسات التي تقاطعت مع دراستنا من حيث تناول والخصوصية:

-تطبيقات الذكاء الاصطناعي وأثرها في تنمية الذات اللغوية الإبداعية لدى الطلاب الفائقين بالمرحلة الثانوية:

والدراسة بحث منشور في مجلة كلية التربية، المجلد 39، العدد 1، يناير 2023، ص، ص: (110)، (135)، أ.د/ عبد الرزاق مختار محمود، أ.د/ أحمد محمد علي رشوان، أ.د/ أحمد عبد الفتاح عبد الوهاب.

وهي من الدراسات الحديثة التي تناولت تأثير تقنيات الذكاء الاصطناعي على تطوير المهارات اللغوية الإبداعية لدى الطلاب المتميزين، حيث تم تطبيق الدراسة على عينة مكونة من 30 طالباً من الصف الثالث الثانوي الأزهرى، وقد سبق الدراسة اختبار قبلي، ثم بعد التطبيق تمت الإجابة على الاختبار البعدي لقياس الذات اللغوية الإبداعية، وأظهرت النتائج وجود فروق ذات دلالة إحصائية لصالح التطبيق البعدي، مما يشير إلى فعالية تطبيقات الذكاء الاصطناعي في تنمية هذه المهارات.

وقد أوصت الدراسة بضرورة دمج تطبيقات الذكاء الاصطناعي في العملية التعليمية، والاهتمام بتنمية المهارات اللغوية الإبداعية لدى الطلاب في مختلف المراحل الدراسية.

-تأثير الذكاء الاصطناعي على تعليم اللغة العربية:

وهي دراسة تقدم بها د. سيف الفطريانا ونُشرت في المؤتمر الدولي لطلبة الدراسات العليا في تدريس اللغة العربية وآدابها واللسانيات (ICON-POSTALL) عام 2023، وتناولت تأثير الذكاء الاصطناعي على تعليم اللغة العربية، مع التركيز على الفوائد والتحديات التي تواجه استخدام هذه التقنية في التعليم.

وناقشت الدراسة التأثيرات الإيجابية والسلبية للذكاء الاصطناعي باستخدام المنهج الوصفي التحليلي والذي ساعد في تقديم فهم شامل للظاهرة المدروسة، وتم دعم البحث بأمثلة على تطبيقات الذكاء الاصطناعي المستخدمة في تعليم اللغة العربية.

وتعتبر الدراسة أساساً جيداً لفهم العلاقة بين الذكاء الاصطناعي وتعليم اللغة العربية، ولكنها بحاجة إلى تعميق الجانب التجريبي وتوسيع التحليل ليشمل التطورات الحديثة في هذا المجال.

-تطبيقات الذكاء الاصطناعي ودورها في تنمية المحتوى الذكي من وجهة نظر أعضاء هيئة التدريس:

نُشرت الدراسة الباحثة نجوى محمد الشمري في مجلة العلوم التربوية والنفسية، المجلد 7، العدد 41، 2023م، وهدفت إلى الكشف عن دور تطبيقات الذكاء الاصطناعي في تطوير المحتوى التعليمي الذكي، وذلك من خلال تحليل آراء أعضاء هيئة التدريس حول فعالية هذه التقنيات في تحسين جودة التعليم وتلبية احتياجات المتعلمين.

اعتمدت الدراسة المنهج الوصفي التحليلي، حيث ركزت على استعراض كيفية استخدام الذكاء الاصطناعي في إنشاء محتوى تعليمي يتماشى مع متطلبات التعليم الحديث. كما ناقشت فوائد الذكاء الاصطناعي، مثل تحسين جودة المناهج، تعزيز التعليم التكيفي، وتوفير أدوات تعليمية مخصصة، إلى جانب التحديات التي تواجه تطبيقه، مثل التكلفة المرتفعة، تعقيد الخوارزميات، وضعف تفاعل الطالب مع المعلم.

- Can AI-generated materials help in Arabic teaching? A study of potential and pitfall:

والدراسة قدمها الباحثان: مي زكي، أحمد علي، للمؤتمر الدولي للذكاء الاصطناعي واللسانيات في الشارقة، 17 - 18 أكتوبر 2024، جامعة الشارقة الأمريكية ومركز الإمارات للبحوث والدراسات.

وناقشت الدراسة دمج تقنيات الذكاء الاصطناعي (AI) في تدريس اللغة العربية، خاصة في تطوير مواد تعليمية مخصصة ومتكيفة مع احتياجات المتعلمين، تم تقييم خمس أدوات ذكاء اصطناعي رئيسية: ChatGPT, Gemini, Co-Pilot, JAIS, Diffit وتمت دراستها وفقاً لقدرتها على إنشاء نصوص دقيقة لغوياً ومناسبة تربوياً لمتعلمي اللغة العربية في المستوى المتوسط.

وخلصت في الأخير إلى مجموعة من النتائج، وهي:

أولاً: وجود تحديات تواجه الذكاء الاصطناعي في العربية

- الإفراط في استخدام الجمل الاسمية.
- أخطاء متكررة في الأفعال والضمائر.
- تكرار غير منتظم للمفردات دون مراعاة السياق.
- تأثير اللغة الإنجليزية على النصوص، مما يؤدي إلى تعبيرات غير طبيعية وأخطاء نحوية.

ثانياً: إيجابيات الأدوات المدروسة

• توفير محتوى متنوع وأنواع مختلفة من الأسئلة.

• القدرة على توليد محتوى جذاب للمتعلمين.

- استخدام الذكاء الاصطناعي لتطوير مهارات اللغة العربية في التعلم:

نُشرت الدراسة من قبل الباحثين ديدموليانو، محمد زكي، وأرساد محمد علي ريدو في جامعة دار

السلام غونتور، ضمن مجلة *An-Nidzam: Jurnal Manajemen Pendidikan*

dan Studi Islam.

وناقشت دور الذكاء الاصطناعي في تحسين تعليم اللغة العربية، مع التركيز على تطوير المفردات، القواعد، والكفاءة في المحادثة، معتمدة المنهج التحليلي لدراسة تأثير أدوات الذكاء الاصطناعي مثل ChatGPT و JAIS و Gemini في دعم المتعلمين.

وأظهرت النتائج أن هذه التقنيات ساعدت في تحسين تجربة التعلم، لكنها واجهت تحديات مثل التكرار اللغوي، أخطاء القواعد، وتأثير الإنجليزية على الصياغة العربية، و رغم أهمية البحث، إلا أنه يفتقر إلى التحليل الكمي والمقارنة التجريبية مع أساليب التدريس التقليدية، وفي الأخير أوصت الدراسة بمزيد من الأبحاث التجريبية لضمان فعالية الذكاء الاصطناعي في تطوير المهارات اللغوية.

-واقع اللغة العربية في تطبيقات الذكاء الاصطناعي، دراسة وصفية تطبيقية:

وهو بحث منشور في مجلة "دراسات: العلوم الإنسانية والاجتماعية، العدد (25)، مجلد (2)، 2025م، للدكتور محمد إسماعيل العميرة، ص، ص: (459، 471)، وتناول العلاقة بين الذكاء الاصطناعي ودراسة اللغة العربية، وركز على كيفية تمثيل المعرفة الأساسية التي تدعم عملية الفهم.

ويسلط البحث معتمداً على المنهج الوصفي التطبيقي الضوء على التحديات التي تواجه اللغة العربية، وتأثير الذكاء الاصطناعي عليها من حيث تغير التراكيب اللغوية، انتشار اللغة المختلطة (العربية-الإنجليزية)، وتأثير الترجمة الآلية على الفصحى.

منهج البحث:

أُعتمد في البحث على المنهج الوصفي التحليلي باعتباره مناسباً لمعرفة مدى تطور الذكاء التوليدي في اللغة العربية، مع التركيز على خصائص اللغة العربية من حيث القواعد والتراكيب اللغوية وتأثيرها على أداء النماذج اللغوية.

وأُستعمل أيضاً المنهج التطبيق لدراسة التحديات التي تواجه النماذج اللغوية في فهم وإنتاج النصوص العربية بالتطبيق على أداة الذكاء الاصطناعي شات جي بي تي، وقياس جودة المخرجات من حيث الدقة، التنوع، والتماسك اللغوي.

المبحث الأول: الإطار النظري

أولاً: الذكاء الاصطناعي التوليدي:

هو نوع من أنواع الذكاء الاصطناعي الذي يستخدم تقنيات تعلم الآلة والشبكات العصبية العميقة لمحاكاة قدرة الإنسان في إنشاء بيانات جديدة أو محتوى أصيل ومبتكر، مثل النصوص، والصور، ومقاطع الفيديو، كما يمكن لنماذج الذكاء الاصطناعي التوليدي توليد مخرجات من نفس نوع المدخلات (مثل تحويل نص إلى نص)، أو من نوع مختلف (مثل تحويل نص إلى صورة أو مقطع فيديو)،⁽¹⁾ ومن أهم أسباب نجاح الذكاء الاصطناعي التوليدي:

o توفر كميات هائلة من البيانات عبر الإنترنت.

o التقدم في القدرات الحاسوبية.

ثانياً: النماذج اللغوية الكبرى (Large Language Models - LLMs)

هي أنظمة ذكاء اصطناعي متطورة تعتمد على الشبكات العصبية العميقة وتقنيات التعلم العميق (Deep Learning) لمعالجة اللغة الطبيعية (NLP).

ويتم تدريب هذه النماذج على كميات ضخمة من البيانات النصية لتوليد نصوص شبيهة بتلك التي يكتبها البشر، وفهم السياق، والإجابة على الأسئلة، وإنتاج محتوى جديد، ومن أمثلة هذه النماذج: **GPT-4** من (OpenAI، BERT)، من (Google)، نموذج لغوي عربي من جامعة محمد بن زايد للذكاء الاصطناعي (Jais)، و **LLaMA** من (Meta).

أبرز خصائص النماذج اللغوية الكبرى:

- تعتمد على بنية المحولات (Transformers)، كما هو الحال في نموذج GPT (Generative Pre-trained Transformer).
- تمتلك قدرة واسعة على فهم النصوص وتحليلها، مما يجعلها مفيدة في الترجمة، وتلخيص النصوص، وإنشاء المحتوى، وتحليل المشاعر.
- يتم تدريبها على كميات ضخمة من البيانات من مصادر مختلفة مثل الكتب، المقالات، والمواقع الإلكترونية.

وتشير بعض التقارير إلى أن النماذج اللغوية الكبيرة (LLMs) تم تطويرها في الأساس للغة الإنجليزية، مما يجعل أداءها في اللغات الأخرى أقل كفاءة نسبياً - بسبب نقص البيانات المتاحة لتدريبها على تلك اللغات. يوضح التقرير أن تقنيات الذكاء الاصطناعي التي تعتمد على هذه النماذج، مثل الدردشة الآلية (Chatbots) والإشراف على المحتوى وتحليل المشاعر، تعمل بشكل أفضل باللغة الإنجليزية مقارنة بأكثر من 7000 لغة أخرى في العالم. (2)

ومما يؤيد هذا الطرح أن موقع ستاتيسستا الذي يؤكد أن كمية البيانات المتاحة للإنجليزية تمثل حوالى 52 بالمائة من المادة اللغوية على الانترنت ويتوزع الباقي على جميع لغات العام الأخرى. وحصّة اللغة العربية من ذلك لا تتعدى 0.6 بالمائة. (3)

ثالثاً: جودة النصوص

ويقصد بها تحقيق النص لأهدافه بوضوح وفعالية، وتشمل معايير الجودة الأساسية:

1. تنوع المفردات: استخدام مجموعة واسعة من الكلمات والتعابير.
 2. الدقة: الالتزام بالقواعد النحوية والإملائية الصحيحة.
 3. الانسجام: ترابط الأفكار وسلسلة تدفق النص.
- كما تعتمد الجودة على مصداقية المعلومات وتأثيرها، خاصة في النصوص الإقناعية أو التفاعلية. النص الجيد يسهل فهمه، يجذب القارئ، ويوصل الفكرة بوضوح دون تعقيد، مما يجعله فعالاً في تحقيق غايته.

رابعاً: اللغة الطبيعية

وفقاً لمعجم البيانات والذكاء الاصطناعي الصادر عن الهيئة السعودية للبيانات والذكاء الاصطناعي (سدايا) يُعرّف مصطلح "اللغة الطبيعية" بأنها: "لغة بشرية يمكن التعبير عنها بنص أو كلام أو إشارة".⁽⁴⁾

تواجه النماذج اللغوية تحديات عند معالجة اللغة العربية مقارنة باللغات الأخرى، ورغم الجهود المبذولة في تطوير نماذج لغوية عربية متخصصة، إلا أن أبرز التحديات التي يمكن الوقوف عندها ما يلي:

1- تعقيد اللغة العربية:

فاللغة العربية لغة غنية ومتشعبة، ويميزها تعقيد النظام الصرفي والنحوي مقارنة باللغات الأخرى، ما يجعل مهمة فهمها فهماً دقيقاً صعباً على أدوات الذكاء الاصطناعي.

2- نقص البيانات:

يواجه الذكاء الاصطناعي تحديات كبيرة عند التعامل مع اللغات ذات الموارد المنخفضة، حيث إن كمية البيانات المتاحة لتدريب النماذج تكون قليلة ومتدنية الجودة، إذ يعتمد غالباً في عملية على تدريب النماذج اللغوية على مصدرين أساسيين، هما: موقع ويكيبيديا: وغالباً ما تعتمد على ترجمات من الإنجليزية. الأخبار: والتي تفتقر أحياناً إلى الدقة وتستخدم مزيجاً بين اللغة الدارجة والفصحى.

في بعض الحالات، يتم اللجوء إلى الترجمة الآلية لسد الفجوة، ولكن الأخطاء الناتجة عن الترجمة قد تزيد من سوء جودة النماذج بدلاً من تحسينها.

3- اختلاف اللهجات:

وجود اختلافات بين الفصحى واللهجات يزيد بلا شك من صعوبة تدريب النماذج على فهم جميع السياقات.

ورغم هذه التحديات التي تخص اللغة العربية دون غيرها من اللغات إلا أنه قد برزت جهود لمشاريع بحثية لتطوير نماذج لغوية عربية متخصصة مثل:

- علام (ALAM):

أعلنت الهيئة السعودية للبيانات والذكاء الاصطناعي (سدايا) عن تحقيق نموذج "علام" للمركز الأول عالمياً كأفضل نموذج لغوي توليدي باللغة العربية وفقاً للتقييم المعياري (Arabic MMLU)، مما

يعكس تقدمه في معالجة اللغة الطبيعية (NLP). جاء هذا الإعلان خلال القمة العالمية للذكاء الاصطناعي بمشاركة قيادات من سدايا، ويؤكد جهود المملكة في تعزيز اللغة العربية رقمياً، ضمن مستهدفات رؤية السعودية 2030.

وقد تم تطوير "عَلَام" في المركز الوطني للذكاء الاصطناعي بكوادر وطنية، بالتعاون مع 160 جهة حكومية، حيث تم تدريبه على 500 مليار وحدة نصية عربية، وطرحه بعدة إصدارات تتراوح بين 7 إلى 70 مليار متغير، كما أعلنت IBM عن توفيره على منصتها watsonx.ai، فيما أدرجته مايكروسوفت ضمن خدمات Azure، مما يعكس الثقة العالمية في قدراته على تحليل النصوص وتوليد المحتوى العربي بذكاء وكفاءة.⁽⁵⁾

- نموذج "أراجين ليدر بورد" (AraGen Leaderboard)

وهو إطار عمل مصمم لإعادة تعريف تقييم نماذج اللغة العربية الكبيرة (LLMs) أطلقتها شركة "إنسبشن" (Inception)، وهي شركة تابعة لمجموعة "جي 42" (G42) بالتعاون مع جامعة محمد بن زايد للذكاء الاصطناعي، ويُقدم النموذج مقياساً مبتكراً تم تطويره ليلبي احتياجات اللغة العربية، وهو "3 سي 3 اتش" (3C3H) والذي يُقيّم نماذج الذكاء الاصطناعي عبر ستة أبعاد رئيسية: الصواب، والاكتمال، والإيجاز، والفائدة، والصدق، وعدم الإضرار.⁽⁶⁾

- نموذج "جيس-13ب-شات" (Jais-13B-Chat)

وهو نموذج لغوي توليدي ثنائي اللغة (العربية والإنجليزية) يضم 13 مليار متغير، النموذج تم تطويره بالتعاون بين معهد الابتكار للذكاء الاصطناعي (Inception Institute of Artificial Intelligence) وجامعة محمد بن زايد للذكاء الاصطناعي (MBZUAI) وشركة سيربراسيسستمز (Cerebras Systems).

يعتمد النموذج على بنية المحولات (Transformers) ويستخدم تشفير المواقع ALiBi، مما يعزز قدرته على التعامل مع تسلسلات نصية طويلة وتحسين دقة السياق.

ومن أهم ما يميز النموذج أنه قادر على: معالجة المهام المعقدة، التعامل مع الفروق الدقيقة في اللغة العربية، حيث يُستخدم النموذج في تطبيقات مثل المساعدات الذكية، وخدمات العملاء، وتحليل النصوص، وتوليد المحتوى، مع التركيز على تقديم إجابات دقيقة وموثوقة باللغتين العربية والإنجليزية.⁽⁷⁾

المبحث الثاني: الإطار التطبيقي

أنشأنا مجموعة صغيرة من البيانات التي مصدرها شات جي بي تي، وجمعنا ما يقارب المليون من الأسئلة استخلصناها من البيانات العربية المتاحة.

من أجل ذلك فقد فحصنا كل البيانات المجانية للغة العربية والتي تمكنا من تنزيلها. وحجمها يزيد عن 141 مليون جملة أو بحدود 2.3 مليار كلمة. ثم من كل ذلك اقتصرنا على ما ينتهي منها بعلامة الاستفهام.

ومن هذه الأخيرة أخذنا عينة من التي تبتدئ منها بأداة استفهام مثل: اين، لماذا، كيف، متى، من، كم، الخ، ثم قدمنا الأسئلة تبعاً لبرنامج الذكاء الاصطناعي وقيدنا أجوبته، وفي التحليل النحوي المبسط اعتمدنا نظام ستانفورد⁸ لوسم الكلمات لتصنيف أقسام الكلام، لتتعرف على ما كثر تردده في بدايات الجمل. وفيما يلي نعرض بعض المقارنات بين النص التوليدي والنص الطبيعي:

- أولاً، وكما يظهر من الأرقام في الجدول أدناه، فإنه وعلى الرغم من أن الذكاء الاصطناعي يميل إلى تطويل الجمل (18 كلمة للنص الآلي في مقابل 16 للطبيعي) فإن نسبة التجديد فيه أقل من النص الطبيعي. بمعنى أن البرنامج له معجم من الكلمات المفضلات التي يميل إلى استخدامها كلما أمكن. ونتيجة لذلك نجد في النص الطبيعي كلمة جديدة في كل 29 كلمة في المتوسط. بينما لا نصادف كلمة جديدة في النص الآلي إلا بعد حول 34 كلمة.

النص الطبيعي	النص الآلي
متوسط عدد الكلمات في الجملة	16.38
التنوع المعجمي	33.89
	18.07
	27.93

- ثانياً إذا فحصنا التكرار في تتابع أقسام الكلام في النصين الطبيعي والآلي فإننا نجد أن برنامج الذكاء الاصطناعي كثيراً ما يبتدئ جملة بحرف جر. وفي الغالب يكون ذلك الحرف إما "في" أو "من". فالأولى لابتداء عبارات تتعلق بتاريخ ماضية. مثل "في الستينيات". والأخرى تتكرر بسبب استعمال عبارة "من المهم..." في بداية الجمل. ونعتقد أن هذا النوع من الجمل انما ينشأ أساساً من التأثير باللغة الإنجليزية، التي كثيراً ما يبتدأ فيها بحرف الجر، أو بعبارة "it is important" ...

- ثالثاً، يوازن النص الطبيعي في الابتداء بالأفعال والأسماء، بينما يميل النص الآلي إلى كثرة الابتداء بالأفعال. وغالباً ما يكون ذلك الفعل ماضياً مسنداً إلى ضمير الغائب المفرد. ولعل هذا الأخير بسبب غياب سياق يغطي أكثر من جملة. لان اسئلتنا مستقلة عن بعضها. وإذا غاب السياق كان المفرد الغائب هو الخيار الشائع.
- رابعاً، النص الطبيعي أكثر تنوعاً من حيث البنية النحوية. حيث إذا نسبنا عدد التراكيب المختلفة لخمس كلمات متتابعة إلى عدد الجمل فإن هذه النسبة تكون 94.72% في النص الطبيعي بينما تنخفض إلى 87.70% في النص الآلي.

الخاتمة:

يلعب الذكاء الاصطناعي التوليدي دوراً مهماً في تشكيل مستقبل اللغة العربية، لكنه يواجه تحديات تتعلق بجودة البيانات ودقتها، لكنه يمكن ومن خلال تحسين مصادر البيانات وتطوير تقنيات التدريب تعزيز جودة النماذج اللغوية العربية وجعلها أكثر فعالية في تمثيل اللغة.

التوصيات:

1. العمل على تطوير مجموعات بيانات لغوية عربية غنية لدعم تدريب النماذج بشكل أكثر كفاءة.
2. تحسين دقة الترجمة الآلية من وإلى اللغة العربية.
3. تطوير مساعدين افتراضيين ذوي قدرة أعلى على فهم اللغة العربية في السياقات المختلفة.
4. دعم التعليم الإلكتروني باللغة العربية من خلال محتوى ذكي أكثر دقة.
5. الاهتمام بالدراسات المقارنة بين تأثير الذكاء الاصطناعي على اللغة العربية مقارنة باللغات الأخرى، مما كان سيمنح القارئ فهماً أوسع لمدى تطور تقنيات الذكاء الاصطناعي في دعم اللغة العربية مقارنة بلغات مثل الإنجليزية أو الفرنسية.
6. محاولة فهم التحديات الثقافية واللغوية التي تواجه تعليم اللغة العربية بالذكاء الاصطناعي، مثل التعامل مع الفصحى واللهجات المختلفة، وتأثير ذلك على دقة النماذج اللغوية في الفهم والإنتاج اللغوي.

المراجع:

- 01- الذكاء الاصطناعي التوليدي، الهيئة السعودية للبيانات والذكاء الاصطناعي، نوفمبر 2023م، ص: 8.
- 02 - "Lost in Translation: Large Language Models in Non-English Content Analysis" نُشر في 22 "دُشِر في Center for Democracy & Technology – CDT)، مركز الديمقراطية والتكنولوجيا (2023 مايو
- 03-(<https://www.statista.com/statistics/262946/most-common-languages-on-the-internet/>)
- 04 - معجم البيانات والذكاء الاصطناعي (بالعربية والإنجليزية)، الهيئة السعودية للبيانات والذكاء الاصطناعي، 2022، ص. 88.
- 05 - <https://www.sba.sa/ar/nmwdhj-lawm-alafdl-almyana-dmn-fyth-byn-alnmadhj-altwlydyt-fy-alalm-ballght-alrbyt>.
- 06 - <https://inceptionai.ai/2024/12/04/rethinking-llm-evaluation-with-3c3h-aragen-benchmark-and-leaderboard/index.html>
- 07-<https://huggingface.co/inceptionai/jais-13b-chat>
- 08-<https://nlp.stanford.edu/projects/arabic.shtml>