مجلة الجامعة الإسلاميّة للغة العربيّة وآدابها

مجلة علميّة دوريّة محكمة

الجزء ٢

العدد :١٩
يناير - مارس ٢٠٢٦م

مجلة الجامعة الإسلاميّة للغة العربيّة وآدابها الجزء ٢


١٣٨
١٩٦


بسم الله الرحمن الرحيم

معلومات الإيداع

في مكتبة الملك فهد الوطنية

النسخة الورقية:

رقم الإيداع ١٤٤٣/٣٢٨٣ بتاريخ ١٤٤٣/٠٤/٠٢ هـ

ردمد: ١٦٥٨-٩٠٧٦

النسخة الإلكترونية:

رقم الإيداع ١٤٤٣/٣٢٨٤ بتاريخ ١٤٤٣/٠٤/٠٢ هـ

ردمد: ١٦٥٨-٩٠٨٤

الموقع الإلكتروني للمجلة

http: //journals. iu. edu. sa/ALS/index. html

ترسل البحوث باسم رئيس تحرير المجلة عبر المنصّة الإلكترونية

البحوث المنشورة في المجلّة تعبّر عن آراء الباحثين

ولا تعبر بالضرورة عن رأي المجلة

جميع حقوق الطّبع محفوظة للجامعة الإسلامية

هيئة التحرير

د. تركي بن صالح المعبدي
(رئيس هيئة التحرير)
أستاذ النحو والصرف المشارك بالجامعة الإسلامية

د. خليوي بن سامر العياضي
(مدير التحرير)
أستاذ تعليم اللغة العربية لغير الناطقين بها المشارك بالجامعة الإسلامية

أ.د. عبد الرزاق بن فراج الصاعدي
أستاذ أصول اللغة والمعاجم بالجامعة الإسلامية

أ.د. عبدالرحمن بن دخيل ربه المطرفي
أستاذ الأدب والنقد بالجامعة الإسلامية

أ.د. الزبير بن محمد أيوب
أستاذ أصول اللغة والمعاجم بالجامعة الإسلامية

د. مبارك بن شتيوي الحبيشي
أستاذ البلاغة المشارك بالجامعة الإسلامية

أ.د. محمد بن ظافر الحازمي
أستاذ اللسانيات بالجامعة الإسلامية

د. عبد المجيد بن عثمان اليتيمي
أستاذ أصول اللغة المشارك بالجامعة الإسلامية

أ.د. عبدالله بن عويقل السلمي
أستاذ النحو والصرف بجامعة الملك عبدالعزيز

أ.د. علي بن محمد الحمود
أستاذ الأدب والنقد بجامعة الإمام محمد بن سعود الإسلامية

أ.د. عبد الرحمن بن مصطفى السليمان
أستاذ اللّغات والآداب السامية والترجمة بجامعة لوفان – بلجيكا

أ.د. علاء محمد رأفت السيد
أستاذ النحو والصرف والعروض بجامعة القاهرة بمصر

أ.د. سعيد العوادي
أستاذ البلاغة وتحليل الخطاب بجامعة القاضي عياض بالمغرب

الهيئة الاستشارية

أ.د. محمد بن يعقوب التركستاني
أستاذ أصول اللغة بالجامعة الإسلامية

أ.د. محمد محمد أبو موسى
أستاذ ورئيس قسم البلاغة بكلية اللغة العربية جامعة الأزهر

أ.د. تركي بن سهو العتيبي
أستاذ النحو والصرف بجامعة الإمام محمد بن سعود الإسلامية

أ.د. عياد بن عيد الثبيتي
أستاذ النحو والصرف بجامعة أم القرى

أ.د. سالم بن سليمان الخماش
أستاذ اللغويات بجامعة الملك عبدالعزيز

أ.د. ناصر بن سعد الرشيد
أستاذ الأدب والنقد بجامعة الملك سعود

أ.د. منصور بن محمد الغامدي
أستاذ الصوتيات الحاسوبية بهيئة تقويم التعليم والتدريب

أ.د. عبد الفتاح محمد حبيب
أستاذ النحو والصرف والعروض بكلية الزقازيق بجامعة الأزهر

أ.د. حسن دربر
أستاذ الترجمة وعلم المصطلح بجامعة القاضي عياض بالمغرب

أ.د. عزالدين محمد المجدوب
أستاذ اللسانيات بالجامعة التونسية

أ.د. محمد بن صالح الشنقيطي
أستاذ النقد الأدبي الحديث بجامعة جدارة بالأردن

أ.د. صالح بن الهادي رمضان
أستاذ الأدب والنقد بجامعة تونس

قواعد النشر في المجلة (*)

- أن يكون البحث جديداً؛ لم يسبق نشره.
- أن يتّسم بالأصالة والجدّة والابتكار والإضافة للمعرفة.
- ألا يكون مستلاً من بحوثٍ سبق نشرها للباحث.
- أن تراعى فيه قواعد البحث العلميّ الأصيل، ومنهجيّته.
- أن يشتمل البحث على:

– عنوان البحث باللغة العربية وباللغة الإنجليزية.
– مستخلص للبحث لا يتجاوز (٢٥٠) كلمة؛ باللغتين العربيّة والإنجليزية.
– كلمات مفتاحيّة لا تتجاوز (٦) كلمات؛ باللغتين العربيّة والإنجليزية.
– مقدّمة.
– صلب البحث.
– خاتمة تتضمّن النّتائج والتّوصيات.
– ثبت المصادر والمراجع باللغة العربية.
– رومنة المصادر العربية بالحروف اللاتينية في قائمة مستقلة.

- في حال (نشر البحث ورقياً) يمنح الباحث نسخة مجانية واحدة من عدد المجلة الذي نُشر بحثُه فيه، و (١٠) مستلات من بحثه.
- في حال اعتماد نشر البحث تؤول حقوق نشره كافة للمجلة، ولها أن تعيد نشره ورقيّاً أو إلكترونيّاً، ويحقّ لها إدراجه في قواعد البيانات المحلّيّة والعالمية – بمقابل أو بدون مقابل – وذلك دون حاجة لإذن الباحث.
- لا يحقّ للباحث إعادة نشر بحثه المقبول للنّشر في المجلّة – في أي وعاء من أوعية النّشر – إلاّ بعد إذن كتابي من رئيس هيئة تحرير المجلة.
- نمط التوثيق المعتمد في المجلة هو نمط (شيكاغو).

(*) يرجع في تفصيل هذه القواعد العامة إلى الموقع الإلكتروني للمجلة: http: //journals. iu. edu. sa/ALS/index. html

محتويات العدد

م	البحث	الصفحة
١)	الأثرُ الْفَلْسَفِيُّ والْمَنْطِقِيُّ فِي شَرْحِ ابْنِ يَعِيشٍ عَلَى الْمُفَصَّلِ دراسة في تداخل الفكر الفلسفي والتحليل النحوي **د. عواض محمد عطيان القارحي**	٩
٢)	الاستراتيجية الحجاجية في الأبواب التأسيسية من كتاب الأصول لابن السراج **د. عائشة قاسم علي شماخي**	٦٣
٣)	وصف الحيوان في الشعر الجاهلي –دراسة بيئية **د. عادل بن علي الغامدي**	١١٣
٤)	أسرار التجاور في آيات الربا والإنفاق في الذكر الحكيم **د. علي بن محمد آل نومة القحطاني**	١٦٥
٥)	الهُويَّة في رواية ما بعد الكارثة –دراسةً نقديَّةً مقارنةً في فرانكشتاين في بغداد لأحمد سعداوي والطَّريق لكورماك مكارثي **د. محمد بن لافي اللويش**	٢٣٧

م	البحث	الصفحة
٦)	الاستلزامُ الحواريّ في خِطاب السِّيرة الذاتية السُّعودية دِراسة تداولية في سِير غازي القصيبي أنموذجًا **د. مريم معيبد عبيد الرشيد**	٢٨٣
٧)	نبرُ الكلمةِ وتدفّق الدلالةِ في الخطابِ دراسةٌ في الديناميّةِ التواصليّةِ "تنظير وإنجاز" **د. نجلاء علي مطري**	٣٣٥
٨)	خطاب الحبّ في كتاب (الزهرة) لأبي بكر الأصفهاني **د. منى بنت سعود آل تويم**	٣٧٧
٩)	تقييم أداء النماذج اللغوية الضخمة (LLMs) في كشف وتصنيف الأخطاء الكتابية لدى متعلمي اللغة العربية لغة ثانية **د. خليوي سامر خليوي العياضي**	٤٢٩
١٠)	مصفوفة مقترحة للمهارات اللازمة لمهارة التحدث لمتعلمي اللغة العربية الناطقين بلغات أخرى وفقًا للإطار الأوروبي المرجعي المشترك **د. علي بن محمد الغامدي**	٤٧٧

تقييم أداء النماذج اللغوية الضخمة (LLMs) في كشف وتصنيف الأخطاء الكتابية لدى متعلمي اللغة العربية لغة ثانية

Evaluating the Performance of Large Language Models (LLMs) in Detecting and Classifying Writing Errors among Arabic-as-a-Second-Language Learners

د. خليوي سامر خليوي العياضي

أستاذ تعليم اللغة العربية المشارك بمعهد تعليم اللغة العربية بالجامعة الإسلامية

البريد الإلكتروني: khlewe@gmail.com

استلام البحث A Research Receiving		اعتماد البحث A Research Approving
24/10/2025		06/12/2025
نشر البحث A Research Publication		
شوال ١٤٤٧هـ = April 2026		
DOI:10.36046/2356-000-019-019		

الملخص

هدفت هذه الدراسة إلى تقييم أداء النماذج اللغوية الضخمة في كشف وتصنيف الأخطاء الكتابية لدى متعلمي اللغة العربية لغة ثانية، ولتحقيق ذلك، استخدمت الدراسة المنهج الوصفي التحليلي، واختيرت أربعة نماذج من أجل تقييم أدائها وهي : Gpt-4o و Gemini 2.5 pro و DeepSeek v3.2 و ALLaM 34B ، وقد أسفرت نتائج الدراسة عن وجود قصور واضح في قدرة النماذج اللغوية الضخمة على كشف وتصنيف الأخطاء الكتابية بدقة كافية، حيث بلغ أعلى قيمة أداء محقق لدى نموذج Gemini 2.5 pro 0.5399 في مقياس F1-Score ، وهي نسبة لا يمكن الاعتماد عليها في سياق التعليم، وفي ضوء هذه النتائج. أوصت الدراسة بضرورة توخي الحذر عند توظيف هذه النماذج في ميدان التعليم بشكل عام وتعليم اللغة بشكل خاص، مع أهمية الإشراف المباشر من المعلم. كما أوصت بالعمل على بناء معايير تقييم دقيقة للنماذج اللغوية الضخمة في ميدان التعليم، وتحسين تصميم الموجّهات (Prompts) من خلال تجربة أساليب متقدمة مثل سلسلة الأفكار (CoT)؛ طلبا لرفع دقة النماذج وتقليل نسبة الهلوسة.

الكلمات المفتاحية: النماذج اللغوية الضخمة، تعليم اللغة العربية لغة ثانية، تحليل الأخطاء الكتابية، تقييم الأداء، هندسة الموجّهات.

Abstract

This study aimed to evaluate the performance of Large Language Models (LLMs) in identifying written errors made by non-native learners of Arabic. To achieve this, the study employed a descriptive-analytical methodology, selecting four models for performance evaluation: Gpt-4o, Gemini 2.5 pro, DeepSeek v3.2, and ALLaM 34B. The findings revealed significant shortcomings in the ability of LLMs to identify written errors with sufficient accuracy. The highest performance was achieved by the Gemini 2.5 pro model with an F1-Score of 0.5399, a score considered insufficient for full reliance in an educational context. Considering these findings, the study recommended exercising caution when employing these models in the field of education in general and language teaching in particular, emphasizing the importance of direct supervision by the teacher. It also recommends the development of rigorous evaluation criteria for LLMs in the educational domain and improving prompt design by experimenting with advanced techniques such as Chain-of-Thought (CoT) to enhance model accuracy and reduce the incidence of hallucination.

Keywords: Large Language Models (LLMs), Arabic as a Second Language (ASL), Written Error Analysis, Performance Evaluation, Prompt Engineering.

المقدمة:

يشهد العصر الحالي ثورة تقنية غير مسبوقة، تمثلت في الذكاء الاصطناعي؛ إذ تمكنت الآلة من قدرات بشرية متعددة، فظهرت نماذج لغوية تحاكي هذه القدرات، وكان أحد أبرز أنواعها، النماذج اللغوية الضخمة (Large Language Models - LLMs) مما أحدث نقلة نوعية في تفاعل الإنسان مع الآلة، التي أظهرت تمكنا في فهم اللغة وتوليدها وبرعت في مهام متقدمة مثل: التحليل المنطقي، والاستدلال، والترجمة، مما أثار اهتمامًا كبيرًا في الأوساط الأكاديمية[١].

فقام الباحثون والأكاديميون في مختلف الحقول العلمية باستثمار هذه الإمكانيات وتوظيفها في مجالاتهم. ففي ميدان الطب، أُستثمرت في تقديم الاستشارات الطبية[٢]، وفي البرمجة، أُختبرت في إنشاء شيفرات برمجية[٣]، وفي اللغويات أجريت دراسات متعددة مثل: توليد النصوص الأدبية والقصصية[٤]، واختصار النص وتلخيصه[٥]، وغيرها من الدراسات، وفي هذا كله إشارة إلى ضرورة استثمار منتجات الذكاء الاصطناعي وتطويعها في مختلف المجالات الممكنة، وأنه بات ضرورة ملحة لمواكبة التطور المعرفي والتقني.

(1) Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., and Xie, X. "A Survey on Evaluation of Large Language Models." ACM Transactions on Intelligent Systems and Technology 15, no. 3 (2024):2

(2) Biswas, Md. R., Mohsen, F., Shah, Z., and Zaghouani, W. "Potentials of ChatGPT for Annotating Vaccine-Related Tweets." 2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS) (2023)

(3) Al-khafaji, N. J., and B. K. Majeed. "Evaluating Large Language Models Using Arabic Prompts to Generate Python Codes." 2024 4th International Conference on Emerging Smart Technologies and Applications (eSmarTA) (2024)

(4) Alhussain, A., and A. M. Azmi. "Beyond Event-Centric Narratives: Advancing Arabic Story Generation with Large Language Models and Beam Search." Mathematics 12, no. 10 (2024): 1548.

(5) Einieh, Y., Almansour, A., and Jamal, A. "Fine-Tuning an AraT5 Transformer for Arabic Abstractive Summarization." 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN) (2022)

وفي هذا السياق، لم يكن ميدان تعليم اللغات بمعزل عن هذه التحولات، حيث برزت النماذج اللغوية الضخمة أداة أحدثت تغييرات جوهرية في الممارسات التعليمية[١]، ورأى فيها التربويون أداة لتحسين العملية التعليمية سواء كانت على مستوى المعلم أم المتعلم أم المنهج، وأجريت العديد من الدراسات في هذا الميدان، فمن حيث جهة المتعلم أثبتت دراسة نوعية[٢] أن منصات تعلم اللغة المتكاملة مع الذكاء الاصطناعي أظهرت تحسنًا ملحوظًا في الكفاءة اللغوية لدى متعلمي اللغة الناطقين بغيرها، خاصة في تنمية المفردات وفهم القواعد النحوية وتنمية مهارة المحادثة، وأجرت بني عرابة[٣] دراسة لقياس فاعلية تطبيقات الذكاء الاصطناعي في تنمية مهارة القراءة الإبداعية وبقاء أثر التعلم؛ مما يؤكد أن هذه الأدوات توفر تجربة تعليمية أكثر جاذبية وتفاعلية، وقد أشارت الدراسات إلى أن المتعلمين يقرّون بفعالية هذه الأدوات في توفير الوقت والإسهام في العملية التعليمية [٤].

ومن جهة المعلم تعد عملية التقويم والتغذية الراجعة الفورية أحد أهم أدواره التي تستهلك وقته وجهده، وتحد من قدرته على تقديم متابعة شخصية لكل متعلم، وهنا

(1) Dong, Bingyu, et al. "Large Language Models in Education: A Systematic Review." 2024 6th International Conference on Computer Science and Technologies in Education (CSTE) (2024); Hamaniuk, Vita A. "The Potential of Large Language Models in Language Education." Educational Dimension 5 (2021)

(2) Mulyanto, D., Zaki, M., Ridho, A., and Fata, K. "Artificial Intelligence Utilization for Arabic Language Skills Development in Arabic Language Learning." An-Nidzam: Jurnal Manajemen Pendidikan dan Studi Islam 11, no. 1 (2024)

(٣) بني عرابة، إخلاص بنت إبراهيم بن حمد، والكاف، فاطمة بنت محمد بن أحمد. "فاعلية بعض تطبيقات الذكاء الاصطناعي في تنمية مهارات القراءة الإبداعية وبقاء أثر التعلم لدى طالبات الصف السادس الأساسي." المجلة العربية للنشر العلمي، ع٧٨، (٢٠٢٥م)

(4) Al-Alami, Suhair E. "EFL Learners' Attitudes Towards Utilizing ChatGPT for Acquiring Writing Skills in Higher Education: A Case Study of Computing Students." Journal of Language Teaching & Research 15, no. 4 (2024): p.1029

يأتي دور أدوات الذكاء الاصطناعي لتسهم في حل هذا[١]، فيمكنها تقديم الدعم الفوري الذي يساعد المعلم في العملية التعليمية ومتابعة أداء المتعلم.

أما من ناحية دورها في المنهج التعليمي، فتسهم أدوات الذكاء الاصطناعي في عملية تطوير المناهج، إذ يمكنه توفير إطار عمل للمعلمين لتصميم الدروس، والمساعدة في توليد الأفكار والملخصات والخطوط العريضة للمقررات بشكل أسرع وأفضل، مما يوفر وقت وجهد المعلمين بشكل كبير[٢]، ويراعي مع ذلك حاجات وخصائص المتعلمين من خلال اقتراح الإستراتيجيات والتدريبات المناسبة لمستوياتهم[٣].

ومع هذا فإنه من الضروري التمييز بين قدرة أدوات الذكاء الاصطناعي على إنتاج نصوص صحيحة، وقدرتها على تقديم دور فاعل في العملية التعليمية؛ لذا يجب أن يسبق استثمارَ هذه الأدوات في العملية التعليمية، الوقوفُ على قدرتها ومعرفة الإستراتيجيات التي تجعل منها أداة فاعلة في العملية التعليمية بحق؛ وما ذاك إلا لأنها لم تبن في الأساس لتقوم بدور في العملية التعليمية، بل لتولد نصاً سليماً وفق السياق[٤]؛ لذلك يجب أن تنطلق البداية من تقويم أدائها في الغرض المراد تحقيقه قبل توظيفها في العملية التعليمية؛ لأنها عرضة لظاهرة "الهلوسة" وتوليد معلومات غير صحيحة[٥].

ويزداد تقييم أداء النماذج أهمية إذا كانت العملية تستهدف مهارات اللغة العربية؛

(1) BONNER, Euan; LEGE, Ryan; FRAZIER, Erin. Large Language Model-Based Artificial Intelligence in the Language Classroom: Practical Ideas for Teaching. Teaching English with Technology, (2023: p.23

(2) Ayodele, O. S., Aliu, J., Owoeye, F. O., Ajayi, E. A., and Sheidu, A. Y. "The Role of Artificial Intelligence in Curriculum Development and Management." Journal of Digital Innovations & Contemporary Research in Science, Engineering & Technology 11, no. 2 (2023): p.38

(3) Ejjami, R. "The Future of Learning: AI-Based Curriculum Development." International Journal for Multidisciplinary Research 6, no. 4 (2024): p.1

(4) Panagiotidis, P. "LLM-Based Chatbots in Language Learning." European Journal of Education 7, no. 1 (2024): p.107.

(5) Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., and Xie, X. "A Survey on Evaluation of Large Language Models.": p.15

لأن تمكن تلك الأدوات من اللغة العربية يعد أقل بكثير إذا ما قورن بتمكنها من اللغة الإنجليزية فهما وتوليدا؛ لأن اللغة العربية تعد من اللغات ذات الموارد المحدودة؛ لقلة البيانات العربية التي دُربت عليها(١)، وهذا يوضح أن ما أَثْبَتَت أدوات ذكاء الاصطناعي قدرتها فيه باللغة الإنجليزية لا يعد مسلمة في بقية اللغات.

وتجدر الإشارة إلى ضرورة استخدامها تحت إشراف المعلم الذي يُدرك الكيفية السليمة لتوظيفها، والحصول على أفضل نتيجة تنعكس بشكل إيجابي على الأهداف المرادة، فمن غير هذا قد لا يكون لأدوات الذكاء الاصطناعي دور بارز، وهذا ما أكدته الدراسات من أن استخدامها من قبل الطلاب لم يظهر أي تقدم يلاحظ في اكتساب مهارات اللغة على الرغم من رضا المتعلمين عن مخرجات تلك الأدوات(٢)، بل إن هناك مخاوف من أن الاعتماد المفرط عليها قد يعيق تطوير هذه المهارات(٣).

من هنا، تأتي هذه الدراسة لتسهم في هذا الباب لتقف على قدرات النماذج اللغوية الضخمة في ميدان تعليم اللغة العربية للناطقين بغيرها، متخذة من مهارة الكتابة حقلا لها؛ لمعرفة مدى قدرة النماذج اللغوية في كشف وتصنيف الأخطاء الكتابية لدى متعلمي اللغة العربية لغة ثانية.

وتبرز أهمية هذه الدراسة من خلال تقاطع مجالين: فمن المجال التربوي، يمكن أن تقدم نتائجها رؤى واضحة للمعلمين والمصممين التربويين حول واقع أدوات الذكاء الاصطناعي في قدرتها على كشف وتصنيف الأخطاء، وأيّ تلك الأدوات ذات جودة عالية في هذا الميدان، مما يسمح باتخاذ قرارات حول كيفية استخدامها أدوات مساعدة.

(1) Nehar, A., Bellaouar, S., Souffi, S., and Bouameur, M. "Dhati: A Fine-Tuned Large Language Model for Evaluating Subjectivity in Arabic Textual Data." *Proceedings of the 2023 5th International Conference on Pattern Analysis and Intelligent Systems (PAIS)* (2023): p.1

(2) Al-Alami, "EFL Learners' Attitudes Towards Utilizing ChatGPT for Acquiring Writing Skills in Higher Education" : p.1029.

(3) Dong, "Large Language Models in Education: A Systematic Review": p.131

أما من المجال التقني، فتسهم الدراسة في تقييم القدرات الحالية للنماذج اللغوية الضخمة في التعامل مع الأخطاء في كتابات متعلمي العربية الناطقين بغيرها، مما يسهم في تسليط الضوء على المجالات التي تحتاج فيها هذه النماذج إلى مزيد من التطوير.

مشكلة الدراسة:

يعد تحليل أخطاء المتعلمين وتقديم التغذية الراجعة لهم، أحد أهم المجالات التي تسهم في اكتسابهم لمهارات اللغة الكتابية، ومع أهميته إلا أنه يستنزف من المعلم وقتا وجهدا كبيرين؛ فينبغي منه متابعة شخصية ودقيقة لكل متعلم لتحديد أخطائه وتصويبها، وتقديم الشرح اللازم للمتعلم، في سبيل إكسابه المعرفة وتطوير مهاراته؛ ولهذا السبب يمكن استثمار النماذج اللغوية؛ لأنها تقدم حلا لهذا الإشكال(١).

وقد أُجريت بالفعل دراسات عديدة في ميدان تعليم اللغات تناولت توظيف أدوات الذكاء الاصطناعي، بدءا من وضع التصورات النظرية، والرؤى التربوية، والتطبيقات التعليمية، وانتهاءً بقياس فاعليتها في تنمية المهارات اللغوية المختلفة(٢). إلا

(1) Jaashan, H. M. S., and A. A. Alashabi. "Using AI Large Language Model (LLM-ChatGPT) to Mitigate Spelling Errors of EFL Learners." *Forum for Linguistic Studies* 7, no. 3 (2025)

(٢) عبد اللوي، محمد. "توظيف الذكاء الاصطناعي في تدريس اللغة العربية للناطقين بغيرها عن بعد." مجلة التطوير العلمي للدراسات والبحوث، ع١٧، (٢٠٢٤م)؛ عنتر، دينا. "دمج الذكاء الاصطناعي في تعليم اللغة العربية في التعليم العالي: دراسة حالة على ChatGPT 4 Education."مجلة الأرائك للعلوم والإنسانيات، ع٦، (٢٠٢٤م)؛ محسني، فاطمة. "دمج الذكاء الاصطناعي في تعليم اللغة العربية للناطقين بغيرها: إستراتيجيات فعالة لتطوير المهارات اللغوية." مجلة الباحث، (٢٠٢٥م)؛ الملحم، تركي عبد العزيز. "واقع استخدام تطبيقات الهواتف الذكية في تعليم اللغة العربية للناطقين بلغات أخرى في معهد تعليم اللغة العربية لغير الناطقين بها بالجامعة الإسلامية من وجهة نظر المعلمين." مجلة كلية التربية (أسيوط) ٣٧، ع٢، (٢٠٢١م)؛ الهذلول، علي بن هذلول علي. "أثر تطبيقات الذكاء الاصطناعي في تنمية مهارات الكتابة الأكاديمية لدى متعلمي اللغة العربية الناطقين بلغات أخرى." مجلة جامعة

أن هذه الجهود – رغم أهميتها – غالبًا ما أغفلت الخطوة السابقة والضرورية، المتمثلة في تحليل دقيق لكفاءة هذه النماذج وقدراتها في تمكنها من تلك المهارات التي يُستهدف تنميتها لدى المتعلم.

خاصة مع تجدد هذه النماذج على الساحة بشكل متسارع، وهذا التجدد حمل معه اختلاف بين هذه النماذج من حيث البيانات التي تدربت عليها ومعمارياتها وغير ذلك. هنا يبرز سؤال ملح يحتاج إلى إجابة وهو: أي هذه النماذج ذو الأداء الأفضل في ميدان تعليم اللغة لغير أهلها؟ والإجابة عنه تتطلب عملية تقويم لأدائها؛ للوقوف على نقاط القوة والضعف فيها في ضوء المهارات المطلوب تنميتها، قبل اعتمادها أو توظيفها في البيئات التعليمية.

ويجدر في هذا السياق الإشارة إلى أن جزءًا كبيرًا من الأبحاث التي قيّمت أداء النماذج اللغوية، أو عنيت ببناء معايير التقويم كانت تقف عند حدود مقدرة النموذج فقط. فعلى سبيل المثال هناك تقييم لمدى قدرة النموذج على التلخيص أو الترجمة، أو فهمه للنصوص المقروءة[1]، فليس الهدف من عملية التقويم هو التأكد من مستواها؛ لتطبيقها في ميدان حيوي، أي أن الباعث لعملية التقييم هو معرفة أداء النموذج بحد ذاته، أما هذه الدارسة فإنها تتجاوز اختبار قدرات النموذج في تحديد الأخطاء فقط، إلى تقويم مدى صلاحيتها بوصفها أداة تعليمية قابلة للدمج في ميدان تعليم اللغة العربية

الملك عبد العزيز – العلوم التربوية والنفسية، مج٤، ع٢، (٢٠٢٥م)

(1) G., Campesan, G., Farooq, M., Alhammadi, M., Launay, J., and Noune, B. "AlGhafa Evaluation Benchmark for Arabic Language Models." *Proceedings of the First Arabic Natural Language Processing Conference (ArabicNLP 2023)* (2023); AL-KHALIFA, Shahad; AL-KHALIFA, Hend. The Qiyas Benchmark: Measuring ChatGPT Mathematical and Language Understanding in Arabic. arXiv preprint arXiv:2407.00146, 2024; Lamsiyah, S., Zeinalipour, K., El Amrany, S., Brust, M., Maggini, M., Bouvry, P., and Schommer, C. "ArabicSense: A Benchmark for Evaluating Commonsense Reasoning in Arabic with Large Language Models." Proceedings of the 4th Workshop on Arabic Corpus Linguistics (WACL-4) (2025)

للناطقين بغيرها.

بناء على ما سبق تمثلت مشكلة الدراسة في غياب تقييم لأداء النماذج اللغوية الضخمة في كشف وتصنيف الأخطاء الكتابية لدى متعلمي اللغة العربية لغة ثانية، وجاءت هذه الدراسة لقياس مدى موثوقيتها وكفاءتها في كشف وتصنيف الأخطاء الكتابية لدى متعلمي اللغة العربية لغة ثانية، ومقارنة أداءاتها فيما بينها للوقوف على مستواها، ومعرفة أي هذه النماذج ذات كفاءة عالية في هذا الميدان

أسئلة الدراسة:

تسعى هذه الدراسة للإجابة عن الأسئلة الرئيسة التالية:

١. ما الأخطاء الكتابية التي يقع فيها متعلمو اللغة العربية الناطقون بغيرها؟
٢. ما مستوى كفاءة النماذج اللغوية الضخمة في كشف وتصنيف الأخطاء الكتابية في كتابات متعلمي اللغة العربية لغة ثانية؟
٣. ما الموجّه الذي يقدم أفضل أداء في نتائج النماذج اللغوية الضخمة في كشف وتصنيف الأخطاء الكتابية لدى متعلمي اللغة العربية لغة ثانية؟

أهداف الدراسة:

تهدف هذه الدراسة إلى تحقيق ما يلي:

١. قائمة بالأخطاء اللغوية التي يقع فيها متعلمو اللغة العربية الناطقون بغيرها.
٢. الوقوف على مستوى كفاءة النماذج اللغوية الضخمة في كشف وتصنيف الأخطاء الكتابية في كتابات متعلمي اللغة العربية لغة ثانية.
٣. التعرف على الموجّه الذي يقدم أفضل أداء في نتائج النماذج اللغوية الضخمة في كشف وتصنيف الأخطاء الكتابية لدى متعلمي اللغة العربية لغة ثانية

أهمية الدراسة:

تتمثل أهمية هذه الدراسة في النقاط التالية:

١. تُعد هذه الدراسة استجابة للحاجة المتزايدة إلى توظيف أدوات الذكاء الاصطناعي التي تنادي بها الدراسات في ميدان تعليم اللغات.
٢. تُسهم الدراسة في تزويد معلمي اللغة العربية للناطقين بغيرها بأفضل النماذج اللغوية الضخمة التي يمكن استخدامها في تعليم العربية بشكل عام وفي كشف وتصنيف الأخطاء الكتابية بشكل خاص.
٣. تُفيد هذه الدراسة مختصي الذكاء الاصطناعي في تقديم بيانات حول أداء النماذج اللغوية في سياق تعليمي، مما يساعدهم على التعرف على أوجه القوة والقصور فيها، وتوجيه جهود التطوير المستقبلي.
٤. يمكن أن تسهم نتائج الدراسة في تطوير أدوات تدقيق لغوي أكثر دقة وفعالية لمعالجة الأخطاء الكتابية في النصوص العربية.

مصطلحات الدراسة:

- النماذج اللغوية الضخمة :(LLMs) نماذج حاسوبية كبيرة تُدرب على كميات هائلة من البيانات النصية؛ لتكون قادرة على فهم اللغة البشرية وتوليدها.[١]
- كشف وتصنيف الأخطاء: تحديد المواطن التي تنحرف فيها استجابات المتعلم عن مقاييس الاستخدام اللغوي الصحيح[٢].
- التقييم: إعطاء قيمة لشيء ما وفقا لمستويات وضعت أو حددت سلفا[٣].

حدود الدراسة:

(1) Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., and Xie, X. "A Survey on Evaluation of Large Language Models.": p.3

(٢) طعيمة، رشدي أحمد. "تعليم العربية لغير الناطقين بها: مناهجه وأساليبه." المنظمة الإسلامية للتربية والعلوم والثقافة (الإيسيسكو)، (١٩٨٩م): ٥٤

(٣) بوبنديرة، عبد العزيز. "الامتحانات التقليدية ومشكلات التقييم." مجلة البحوث التربوية والتعليمية ١٣ (عدد خاص)، (٢٠٢٤م): ٤٤٤

الحد الموضوعي:

اقتصرت الدراسة على تقييم أداء أربعة نماذج في كشف وتصنيف الأخطاء الكتابية دون الشفوية لدى متعلمي اللغة العربية لغة ثانية في المستوى الرابع، وهذه النماذج هي: Gpt-4oو Gemini 2.5 pro و DeepSeek v3.2وALLaM 34B.

الحد الزماني:

طبقت الدراسة على النماذج اللغوية الضخمة المذكورة خلال الفصل الدراسي الأول من العام الجامعي ١٤٤٧هـ الموافق ٢٠٢٥م.

تقييم النماذج اللغوية وتعليم العربية:

تعد النماذج اللغوية الضخمة تطورًا كبيرًا على أنظمة تعلم اللغة بمساعدة الحاسوب، وقدمت قفزات كبيرة جدا في ميدان التعليم، حيث وفرت تفاعلًا حواريًا وتغذية راجعة فورية للمتعلمين تناسب مستواهم وقدراتهم [١].

وفي ميدان تعليم اللغات أجريت العديد من الدراسات التي أثبتت فعالية النماذج اللغوية الضخمة في مساعدة المتعلمين على اكتساب المفردات وتعزيز الاحتفاظ بها على المدى الطويل[٢]، ففي مهارة الكتابة تفوقت النماذج اللغوية الضخمة على أدوات التقييم الآلي التقليدية بتقديمها ملاحظات شاملة لا تقتصر على القواعد اللغوية فحسب، بل تشمل أيضًا المحتوى والتنظيم والأسلوب، مما يمنحها قبولًا وتفضيلًا أعلى لدى المتعلم[٣]بالإضافة إلى ما تقدمه من أدوار فاعلة في تخفيف عبء العمل على

(1) FINCHAM, Naiyi Xie; ALVAREZ, Aitor Arronte. Using large language models (llms) to facilitate l2 proficiency development through personalized feedback and scaffolding: An empirical study. In: Proceedings of the International CALL Research Conference. 2024: pp. 60,63

(2) Zhang, Z., and X. Huang. "The Impact of Chatbots Based on Large Language Models on Second Language Vocabulary Acquisition." *Heliyon* 10 (2024): p.1

(3) Zheng, R. "AWE and LLMs in L2 Writing Feedback: An Exploration of EFL Learners' Experiences." *Academic Journal of Humanities & Social Sciences* 8, no. 7 (2025): pp. 61,65

المعلمين؛ ليزيد من كفاءتهم في العملية التعليمية [١].

ويتبادر إلى الذهن من النظرة الأولى للنتائج التي توفرها النماذج اللغوية الضخمة، أنها قادرة على تقديم ما يحتاج إليه المستخدم من متعلم وغيره، إلا أن الأمر المنطقي يقتضي الاحتكام قبل إصدار الحكم، من خلال إجراء اختبار لقدراتها ومعرفة كفاءتها قبل توظيفها، خاصة في ميادين محورية مثل ميدان التعليم الذي له خصوصيته في جميع ما يتصل به لضمان سير العملية التعليمية على أفضل وجه، إلا أنه – على حد علم الباحث – لم تجر اختبارات في تقييم أداء النماذج اللغوية من منطلقات ميدان تعليم اللغات، وما أجري في هذا الميدان لا يعدو كونه نظرة عامة على قواعد العربية أو اختصار النصوص أو الاستدلال وفهم المقروء، ومن الملاحظ أن عددًا من هذه الدراسات أشارت إلى أن أداء النماذج اللغوية في اللغة العربية جاء دون المستوى بخلاف الإنجليزية[٢]، كما أجريت دراسة قامت باختبار أداء نماذج GPT-4 و GPT-4o و Gemini في الإجابة عن أسئلة اختبار القدرات العامة السعودي (GAT)، انتهت إلى أن مستويات تلك النماذج مازالت متوسطة فقد حصلت على درجات دقة تتراوح بين ٤٠٪ و ٦٠٪ ،وأشار الباحثون إلى أنها ليست كافية لتجعل من تلك النماذج أداة مساعدة للمتعلمين من أهل اللغة في فهم المهارات المتطلب معرفتها لتجاوز ذلك الاختبار[٣]، وهذا يطرح تساؤلًا عن قدرة تلك النماذج على مساعدة متعلمي اللغة العربية لغة ثانية في تنمية مهاراتهم اللغوية الحالية .

(1) Yavuz, F., Çelik, Ö., and G. Yavaş Çelik. "Utilizing Large Language Models for EFL Essay Grading: An Examination of Reliability and Validity in Rubric-Based Assessments." British Journal of Educational Technology 56 (2025): p. 163

(2) Koto, F., Li, H., Shatnawi, S., Doughman, J., Sadallah, A. B., Alraeesi, A., Almubarak, K., Alyafeai, Z., Sengupta, N., Shehata, S., Habash, N., Nakov, P., and T. Baldwin. "ArabicMMLU: Assessing Massive Multitask Language Understanding in Arabic." arXiv preprint arXiv:2402.12840 (2024).

(3) Alahmadi, M. D., Alharbi, M., Tayeb, A., and M. Alshangiti. "Evaluating Large Language Models' Proficiency in Answering Arabic GAT Exam Questions." Engineering, Technology & Applied Science Research 14, no. 6 (2024):

دور الموجّهات (Prompts) في تحسين أداء النماذج اللغوية الضخمة:

يعرف الموجّه بأنه مجموعة تعليمات تُقدّم إلى نموذج لغوي كبير (LLM)، تعمل على تهيئته وتخصيصه وتعزيز قدراته أو تنقيحها[١]، وتهدف صياغة الموجّهات إلى توجيه النماذج اللغوية الكبيرة (LLMs) للحصول على أفضل النتائج، وذلك عبر تصميم تعليمات محددة ودقيقة لا تقتصر على مجرد طرح سؤال، بل يتضمن تقديم سياق وإرشادات واضحة للنموذج، مما يتيح له إنتاج إجابات أكثر دقة، وتخصصًا، وملاءمة للمهمة المطلوبة [٢].

تكمن الأهمية الجوهرية لصياغة الموجّهات في كونها أداة فعالة لتحسين أداء النماذج اللغوية دون الحاجة إلى إعادة تدريبها أو تعديل معامِلاتها الداخلية[٣] فهي تُعد بمثابة جسر يربط بين ما يريده المستخدم وقدرات النموذج، مما يمكّن النماذج من التكيّف مع المهام المتنوعة[٤]، فبفضل هذا الأسلوب، يمكن للمستخدمين الحصول على نتائج موثوقة ومخصصة، والحد من الأخطاء والهلوسات التي قد تنتج عن الغموض في التعليمات[٥].

إن صياغة الموجّهات لها دور كبير في جودة مخرجات النماذج اللغوية الضخمة وقد عني بها الباحثون، وسعوا إلى اكتشاف أفضل أساليب تلك الموجّهات في ميادينهم

(1) White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., and D. C. Schmidt. "A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT." arXiv preprint (2023): p.2

(2) Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., and A. Chadha. "A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications." arXiv preprint (2024): p.1

(3) Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., and A. Chadha. "A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications.": p.1

(4) Bansal, P. "Prompt Engineering Importance and Applicability with Generative AI." Journal of Computer and Communications 12 (2024): p.15

(5) Zuccon, G., and B. Koopman. "Dr ChatGPT, Tell Me What I Want to Hear: How Prompt Knowledge Impacts Health Answer Correctness." arXiv preprint (2023): p.2

المختلفة، فقد قام Sahoo وزملاؤه (2024) بجمع تلك الأساليب من خلال مراجعة منهجية لهندسة تلك الموجّهات، وحصروها في 12 قسما واندرج تحتها 41 أسلوبا، تهدف جميعها إلى اكتشاف أفضل طريقة لتحقيق أفضل استجابة للنماذج اللغوية، وهو ما أثبتته الدراسات التي عادوا إليها. وفيما يلي أهم هذه الأساليب المتعلقة بأداء النماذج اللغوية[١]:

١. أسلوب الأمثلة:

في هذا الأسلوب يقوم المستخدم بتمرير أمثلة توضيحية لما يريده مع الموجّه المدخل للنموذج وقد تتعدد الأمثلة في الأمر لأكثر من مثال.

مثال: استخرج الأخطاء الإملائية في النص التالي وفق الأمثلة الموضحة:

[النـــــص]

الأخطاء الإملائية مثل:

أخطاء كتابة همزة القطع: ابراهيم

أخطاء كتابة التاء المربوطة: جميلت

.....

٢. أساليب المنطق والاستدلال:

تنوعت الأساليب تحت هذا القسم، والقاسم المشترك بينها هو تقديم طريقة للتفكير مع الموجّه؛ لكي تساعد النموذج في تنفيذ الأمور المعقدة خطوة بخطوة للحصول على أفضل نتيجة، وقد تصل طريقة التفكير إلى أمور تفصيلية مع إرفاق أمثلة مع التفكير لزيادة الدقة،

(1) Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., and A. Chadha. "A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications."

وغالبا ما تستخدم هذه الأساليب في العمليات التي يرى المستخدم فيها أنها تحتاج إلى تفكيكها إلى خطوات أصغر.

مثال: استخرج الأخطاء الكتابية في النص التالي واستخدم الخطوات التالية في كشف وتصنيف الأخطاء.

الطريقة: استخرج أولا الأخطاء الإملائية ثم النحوية ثم الدلالية.

[النـــص]

٣. أساليب تقليل الهلوسة Hallucination

إحدى الأساليب التي ناقشها الباحثون ما عني بمشكلة الهلوسة التي قد تصيب النماذج، فتقدم ردوداً غير صحيحة أو غير منطقية، وسعوا من خلال تلك الأساليب إلى ضمان الدقة في استجابات النماذج، وأحد هذه الأساليب التي لا تدخل بالضرورة في هندسة الموجّهات ما يعرف بالتوليد المعزز بالاسترجاع Retrieval Augmented Generation (RAG)، وهو تزويد النموذج بمعلومات خارج معرفته، ويطلب منه أن يجيب من خلال تلك المستندات المقدمة؛ لكي لا يأتي بمعلومات غير صحيحة لتجنب الهلوسة، إلا أن هناك أساليب تفيد في جانب الهلوسة ذكرها الباحثون، منها سلسلة التحقق، وفكرتها أن يتحقق النموذج من الإجابة التي قدمها ثم يؤكدها أو يعدلها، مثالها:

الموجّه: أين الحال في الجلمة التالية:

رأيت زيدا باكياً

الإجابة الحال: زيداً

التحقيق: هل كلمة (زيد) تدل على هيئة حال الشيء؟

الإجابة: لا.

الإجابة النهائية: الحال: باكيا.

٤. أساليب التحسين والكفاءة.

وهنا يطلب من النموذج بعد الإجابة عن الموجّه أن يقوم بانتقاد إجابته وفحصها لتقديم إجابة أفضل من المقدمة سلفا، وتقارب الأسلوب السابق في المجمل.

٥. أساليب ما وراء المعرفة والتأمل الذاتي:

تحفز هذه الأساليب النموذج على التفكير بشكل تجريدي حول الأمر المقدم قبل تقديم إجابة من خلال النظر بالمبادئ والمفاهيم المذكورة في الأمر واستعراضها ثم يقدم الإجابة في ضوء تلك الأسس، وهو أشبه بمراجعة ما ذكر حول تلك المفاهيم لتقديم إجابة مؤكدة، وأحد هذه الأساليب أسلوب "أخذ خطوة إلى الوراء" Take a Step Back، كأن يتضمن التوجيه مسائل لحساب قيم رياضية مبنية على نظريات وخوارزميات متعددة؛ فيطلب من النموذج في الأول أن يتأكد من كل نظرية على حدة ويراجعها ثم يكمل التوجيه وينجزه.

تجدر الإشارة إلى أن جميع ما ذكر من أساليب ليس بالضرورة أن تكون مناسبة للمهام اللغوية، بالإضافة إلى أن ما ذكروه يراعي اللغات ذات الموارد الضخمة بعكس اللغة العربية التي تعتبر من اللغات ذات الموارد المحدودة، إلا أن هناك من الباحثين العرب من قدم أساليب تفيد في تحسين أداء النماذج اللغوية، عبر تقديم موجهات تستثمر قدرات النماذج اللغوية في اللغات ذات الموارد الضخمة، ومن تلك الأساليب:

١. تقديم الأمر باللغة الإنجليزية:

لاحظ الباحثون أنهم إذا قدموا الموجّه باللغة الإنجليزية في مهام تخص اللغة العربية، فإن أداء النموذج يكون أفضل من تقديم الموجّه باللغة العربية؛ وذلك بسبب تمكن النموذج من اللغة الإنجليزية؛ لأن كمية البيانات الإنجليزية التي تدرب عليها أكبر بكثير من اللغة العربية(١)، مثاله:

بدلا من كتابة الأمر: لخص النص التالي:

[النص باللغة العربية]

يصبح الأمر: Summarize the following text

[النص باللغة العربية]

٢. نقل المعرفة (Knowledge Transfer)

لا يعتبر نقل المعرفة في الأصل شيئاً يمكن توظيفه من خلال كتابة الموجّهات، ولكنه يحسّن نتائج اللغة العربية ،ويقصد به أن قدرة النموذج ومعرفته لنظام اللغات ذات الموارد الضخمة يمكن أن ينعكس على النتائج في اللغات ذات الموارد المحدودة، فعلى سبيل مقدرته وفهمه للقواعد اللغوية في اللغة الإنجليزية ونمطها يكون له أثر في معالجة اللغات الأخرى(٢)؛ لأن هناك قواعد عالمية تتماثل فيها جميع اللغات، من ذلك أن اللغات تحوي على أفعال وأسماء وصفات،

(1) Khondaker, M. T. I., Naeem, N., Khan, F. L., Elmadany, A. A., and M. Abdul-Mageed. "Benchmarking LLaMA-3 on Arabic Language Generation Tasks." *Proceedings of the Second Arabic Natural Language Processing Conference* (2024): p.285

(2) Alshammari, W., and S. Alhumoud. "TAQS: An Arabic Question Similarity System Using Transfer Learning of BERT with BiLSTM." *IEEE Access* 10 (2022)

فقدرته على فهم الصفة في اللغة الإنجليزية يمكن أن ينعكس على قدرته على استخراج الصفات في العربية.

ومن الجدير في هذا السياق أن يُشار إلى أن كتابة الموجّهات تتشكل من عدة مكونات قد تزيد أو تنقص بحسب غرض المستخدم، وهي تفيد في سياق تحسين أداء ومخرجات النماذج اللغوية الضخمة، وهناك عدة أشكال لمكونات الموجّهات ومن تلك المكونات ما يلي(١):

١. الدور :هذا المكون يحدد "شخصية" أو هوية النموذج التي يجب أن يتقمصها، مما يساعد النموذج ويوجهه لأخذ هذا الدور، مثاله" :أنت خبير في اللغة العربية".
٢. التوجيه: يعد الأمر الرئيسي أو السؤال الأساسي الذي يحدد المهمة المطلوبة من النموذج، مثاله" :استخرج الأخطاء الكتابية في النص التالي".
٣. المعلومات الإضافية: يوفر هذا المكون سياقًا إضافيًا أو تفاصيل تساعد النموذج على فهم المهمة بشكل أفضل. كإضافة قيود أو سياق للمهمة أو توضيحات، مثاله" لا تكتف بتحديد الأخطاء فقط بل صححها أيضا" أو "النص المقدم أحد نصوص متعلمي اللغة العربية لغة ثانية"
٤. تنسيق المخرجات :يحدد هذا المكون الهيكل أو الشكل الذي يجب أن تكون عليه إجابة النموذج، مثاله: " قدم إجابتك على شكل جدول مكون من ثلاثة أعمدة: الخطأ، والصواب، والشرح.
٥. الأمثلة :تقديم مثال ونموذج للأخطاء أو شكل المخرج مما يزيل أي غموض في التعليمات السابقة، ومثال:

(1) Zheng, Y., Li, T., Huang, H., Zeng, T., Lu, J., Chu, C., Huang, Y., Jiang, Z., Xiong, Q., Ge, Y., and M. Li. "Are All Prompt Components Value-Neutral? Understanding the Heterogeneous Adversarial Robustness of Dissected Prompt in Large Language Models." *arXiv preprint* (2025)

"هذا مثال على المخرجات المطلوبة:

الخطأ	الصواب	الشرح
رأيت رجلان	رأيت رجلين	كلمة 'رجلين' مفعول به منصوب، وعلامة نصب المثنى هي الياء وليس الألف

منهجية الدراسة وإجراءاتها

في هذا القسم عرض للإجراءات المنهجية في تقييم أداء النماذج اللغوية الضخمة (LLMs)في كشف وتصنيف الأخطاء الكتابية لدى متعلمي اللغة العربية لغة ثانية. وقد اتبعت الدراسة المنهج الوصفي التحليلي؛ لتصف واقع قدرة النماذج اللغوية في كشف وتصنيف الأخطاء الكتابية لدى متعلمي اللغة العربية لغة ثانية، وتحلل تلك النتائج للوقوف على نقاط القوة والضعف في أداء النماذج؛ بغية الوصول إلى الحكم على مدى إمكانية استثمارها في ميدان تعليم العربية للناطقين بغيرها.

مجتمع الدراسة:

تمثل مجتمع الدراسة في النماذج اللغوية الضخمة (LLMs) المتاحة للاستخدام العام والبحثي في الفترة التي أجريت فيها الدراسة.

عينة الدراسة:

اختار الباحث عينة قصدية مكونة من أربعة نماذج لغوية ضخمة، وهي:

١. Gpt-4o
٢. Gemini 2.5 pro
٣. DeepSeek v3.2
٤. ALLaM 34B

واختيرت النماذج الثلاثة الأولى؛ لأنها من أكثر النماذج زيارة من قبل

المستخدمين، كما أشارت إلى ذلك المواقع التي عُنيت بالتسويق والتجارة، ومن تلك المواقع، موقع onelittleweb الذي أشار إلى أن أكثر الزيارات السنوية لعام ٢٠٢٥م حصلت عليها المواقع التالية[١]:

جدول (١) عدد الزيارات السنوية لمواقع النماذج اللغوية الضخمة

موقع النموذج اللغوي الضخم	عدد الزيارات السنوية
ChatGPT	46.59 مليار
DeepSeek	2.74 مليار
Gemini	1.66 مليار

أما النموذج الرابع ALLaM 34B، فهو النموذج المخصص للغة العربية التي أشرفت عليه سدايا، ويعد النموذج الأقوى في أسرته، وأضيف للقائمة؛ لأنه مخصص للغة العربية، وليس كبقية النماذج المذكورة متعددة اللغات.

أداة الدراسة: تمثلت أداة الدراسة في بناء بطاقة تقييم لأداء النماذج اللغوية الضخمة في التعرف على الأخطاء الكتابية التي قد يقع فيها متعلمو اللغة العربية الناطقون بغيرها.

مصادر الأداة: لبناء قائمة الأخطاء الكتابية استعرض الباحث عددا من الأبحاث التي هدفت إلى تحليل أخطاء متعلمي العربية الناطقين بغيرها لجمع قائمة شاملة لتلك الأخطاء، وتلك الدراسات هي:

(1) Sarkar, S. (2025, August 8). The AI 'Big Bang' study 2025: Best AI chatbots and what 55.88 billion visits reveal. OnelittleWeb. https://onelittleweb.com/data-studies/best-ai-chatbots/

دراسة الملحم[١]، ودراسة الهذلول[٢]، ودراسة محسني [٣]ودراسة عبد اللوي[٤]، ودراسة عنتر[٥]، بالإضافة إلى عدد من الأخطاء التي اقترحها المقيمون لنصوص متعلمي اللغة العربية مما لم تذكر في الدراسات السابقة.

تحكيم الأداة: قدمت الأداة إلى (٤) محكمين من المختصين في تعليم العربية لغير الناطقين بها، لإبداء ما يرونه لازما من إضافة أو تعديل أو حذف؛ لاعتمادها في عملية التقييم.

وجاءت الصورة النهائية لبطاقة التقييم التي حوت جميع الأخطاء التي قد يقع فيها متعلمو اللغة العربية على النحو التالي:

(1) Alfaifi, Abdullah, and Eric Atwell. "Potential Uses of the Arabic Learner Corpus." Paper presented at the Leeds Language, Linguistics and Translation PGR Conference 2013, Leeds, UK (2013)

(٢) أمين، مروة مصطفى السيد. "أخطاء الكتابة بين متعلمي العربية الناطقين بها والناطقين بغيرها في المرحلة الجامعية الأولى: كلية الألسن جامعة عين شمس نموذجًا." فيلولوجي: سلسلة في الدراسات الأدبية واللغوية، ع٦٥، (٢٠١٦م).

(3) Hilal, Mahmud, and Ahmet İsmailoğlu. "Arapça Öğretimi Gören Öğrencilerin Yaygın Dilbilimsel Hataları ve Çözümler Üzerine Uygulamalı Alan Çalışması: Kırıkkale Üniversitesi K.K.Ü.FEF. Arapça M.T.B Örneği." Kırıkkale Üniversitesi Sosyal Bilimler Dergisi 10, no. 2 (July 2020)

(٤) الأفيوني، بشار مصطفى، وشالكينسكايا، أرينا. "تحليل الأخطاء اللغوية في الإنتاج الكتابي لدى متعلمي اللغة العربية الروس في جامعة موسكو الحكومية." مجلة الدراسات اللغوية والأدبية، س١٥، ع٢، (٢٠٢٤م)

(5) Farraj, K. K. "TAHLIL AL-AKHTHA'INDA MUTA'ALLIMI AL-LUGHAH AL-'ARABIYAH LI AL-NAATHIQIINA BI GHAIRIHA." Arabiyat: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban 2, no. 1 (2015)

جدول (١) الصورة النهائية لبطاقة تقييم أداء النماذج في التعرف على الأخطاء الكتابية لمتعلمي العربية الناطقين بغيرها

مجال الخطأ	العدد	نوع الخطأ	كلمات أصاب فيها النموذج	كلمات أخطأ فيها النموذج
القواعد الكتابية		عدم مراعاة المسافة بين الكلمات		
		وضع مسافة بين حروف الكلمة		
الإملاء		الخطأ في همزة الوصل والقطع		
		الخطأ في الهمزة المتوسطة والمتطرفة		
		الخطأ في التاء المفتوحة والمربوطة (ت،ة)		
		الخطأ في الهاء والتاء المتطرفتين (هـ، ة)		
		الخطأ في الألف اللينة (ا، ى)		
		الخطأ بين النون والتنوين (ن ، ـًـٍـٌ)		
		تقصير الصوائت الطويلة (و،ي،ا ← ـُـِـَ)		
		تطويل الصوائت القصيرة (ـُـِـَ ← و،ي،ا)		
		الخطأ في كتابة اللام الشمسية والقمرية		
		الخطأ في كتابة الحروف التي تكتب ولا تنطق		
		الخطأ في كتابة الحروف التي تنطق ولا تكتب		
		حذف نقاط الحروف أو زيادتها (الإعجام)		
		استبدال حرف أو أكثر من الكلمة		
		زيادة حرف أو أكثر		
		نقص حرف أو أكثر		
		الخطأ في ترتيب الحروف داخل الكلمة		
علامات الترقيم		علامة ترقيم خاطئة		
		علامة ترقيم زائدة		
		علامة ترقيم مفقودة		
الصرف		الخطأ في اختيار بنية الكلمة المناسبة		

النحو		الخطأ في الإعراب		
		الخطأ في التعريف والتنكير		
		الخطأ في الجنس (التذكير والتأنيث)		
		أخطاء إسناد الفعل للفاعل		
		الخطأ في العدد (الإفراد والتثنية والجمع)		
		الخطأ في ترتيب الكلمات في الجملة		
		كلمة زائدة		
		كلمة ناقصة		
		استبدال كلمة مكان أخرى		
الدلالة		خطأ في كتابة المفردة المعجمية الصحيحة		
		الخطأ في اختيار الكلمة المناسب المناسبة		
		الخطأ في اختيار التعبير المناسب		
		انعدام الدلالة في التركيب		

يظهر الجدول (١) الشكل النهائي لبطاقة التقييم التي طبقت على النماذج اللغوية الضخمة، لمعرفة قدرتها في التعرف على الأخطاء الكتابية، وهذه البطاقة تقيس أداء أحد النماذج في طريقة واحدة، وبناء على ذلك فقد بلغ عدد البطاقات (١٢٠) بطاقة، لكل نص يقدم للنموذج ثلاث بطاقات، كل بطاقة منها تمثل أحد الموجّهات التي وضحت في أثناء الخطوات الإجرائية، مضروبة في عدد النصوص التي بلغت (١٠) نصوص، مجموع كلماتها (١٨٤٠) كلمة.

النماذج اللغوية المختبرة:

حددت أربعة نماذج، ثلاثة من أشهر النماذج اللغوية التوليدية، وأكثرها استخداما، وواحد من النماذج التي بنيت للغة العربية وهي:

DeepSeek v3.2: نموذج تكون من 671 مليار معامِل، وتدرب على ما يقارب من 14.8 تريليون مفردة عالية الجودة، مستهدفا التركيز على اللغة الإنجليزية والصينية ثم

تأتي بقية اللغات بعد ذلك [١]

Gemini 2.5 pro: نموذج صدر عن شركة قوقل، ولم تعلن حين الكشف عنه عن عدد المعامِلات أو المفردات التي تدرب عليها النموذج، إلا أن الاختبارات التي أعلنت عن أداء هذا النموذج تفوقت على نماذج لغوية ضخمة أخرى في عدد من المهام[٢] مثل: deepseek و Grok و GPT

GPT-4o: نموذج صدر عن شركة Open AI، ولم يعلن شيء بشكل رسمي عن عدد المعامِلات ولا المفردات التي تدرب عليها، ولكن هناك ورقة صدرت عن باحثين من شركة مايكروسوفت أشارت إلى أن عدد معامِلات النموذج بلغ حوالي ٢٠٠ مليار معامِل[٣].

ALLam: نموذج "علّام" الذي أصدرته SDAIA "الهيئة السعودية للبيانات والذكاء الاصطناعي" بحجم معامِلات ٣٤ مليار، وبلغ العدد الإجمالي للمفردات التي تدرب عليها ٥.٢ ترليون مفردة، تكوّن في جزء منه من مزيج بين اللغة الإنجليزية والعربية[٤].

خطوات تنفيذ التقييم:

الخطوة الأولى:

أخذت نصوص عشوائية من كتابات متعلمي اللغة العربية لغة ثانية في المستوى

(1) DeepSeek-AI. "DeepSeek-V3 Technical Report (arXiv:2412.19437v2)." *DeepSeek-AI* (2025)

(2) Google DeepMind. "Gemini Model Thinking Updates: Advanced Coding." Google Blog, March 2025. Retrieved October 19, 2025, from https://blog.google/technology/google-deepmind/gemini-model-thinking-updates-march-2025/#advanced-coding.

(3) Ben Abacha, A., W.-w. Yim, Y. Fu, Z. Sun, M. Yetisgen, F. Xia, and T. Lin. "MEDEC: A Benchmark for Medical Error Detection and Correction in Clinical Notes (arXiv:2412.19260v1)." *arXiv preprint* (2024).

(4) BARI, M. Saiful, et al. Allam: Large language models for arabic and english. arXiv preprint arXiv:2407.15390, 2024.

الرابع، بلغ عدد مفرداتها (١٨٤٠) كلمة وقد اعتمد الباحث في جمع هذه النصوص على المدونة اللغوية لمتعلمي اللغة العربية[١].

الخطوة الثانية:

قدمت نصوص متعلمي اللغة العربية للناطقين بغيرها لمختصين في تعليم العربية، بلغ عددهم (١٠) مختصين، وطلب منهم تقييم كتابات متعلمي اللغة واستخراج الأخطاء الكتابية باستخدام بطاقة التقييم التي بنيت في ضوء الأخطاء الكتابية لدى متعلمي العربية الناطقين بغيرها.

الخطوة الثالثة:

عرضت نصوص متعلمي العربية الناطقين بغيرها على النماذج المحددة سلفا؛ للتعرف على الأخطاء الكتابية، وتنوعت الموجّهات المدخلة لتحقيق أغراض الدراسة، ثم قورنت نتائج النماذج اللغوية الضخمة في كشف وتصنيف الأخطاء مع الأخطاء المحددة سلفا من قبل المختصين في المرحلة الثانية؛ لمعرفة إن كانت النماذج استطاعت أن تتعرف على تلك الأخطاء.

الخطوة الرابعة:

بعد رصد ما أخطأت فيه النماذج وأصابت من الأخطاء، طبق عدد من المقاييس؛ للوقوف على النسبة التي حققها كل نموذج في القدرة على التعرف على الأخطاء، وتمثلت هذه المقاييس فيما يلي: مقياس الدقة Accuracy ومقياس الاستدعاء Recall ومقياس الدقة المحددة Precision ومقياس النسبة الموزونة F1-Score

الموجّهات المدخلة:

لم تهدف الدراسة إلى تقييم نتائج تلك النماذج فقط، ولكن أيضا سعت للوصول إلى استكشاف أَمْثَل طريقة لتوظيفها في تعرف أخطاء متعلمي اللغة؛ لذلك لم تكتف

(1) Alfaifi, Abdullah, and Eric Atwell. "Potential Uses of the Arabic Learner Corpus."

بموجه واحد (prompt) بل جعل هناك ثلاث صور لتلك الموجّهات؛ لمحاولة التوصل إلى أفضل أداء لها في التعرف على الأخطاء، وجاءت تلك الأوامر على النحو التالي:

الموجّه الأول:

حدد الأخطاء الكتابية في النص التالي:

[النص]

الأمر الثاني:

أنت خبير لغوي في اللغة العربية متخصص في تحليل الأخطاء الكتابية والنحوية والصرفية والدلالية والإملائية. مهمتك هي تحليل النص التالي واستخراج الأخطاء الكتابية.

التعليمات:

تعرف على الأخطاء في جميع النص.

حلل الأخطاء وفقًا للقائمة التفصيلية أدناه. لكل خطأ:

اذكر **الكلمة أو الجزء الخاطئ**. وعدد كل نوع من الأخطاء:

قائمة أنواع الأخطاء:

المجال: أخطاء القواعد الكتابية:

عدم مراعاة المسافة بين الكلمات، مثل: ويعلم – وعدمحمد، والصحيح: ويعلم – وعد محمد.

وضع مسافة بين حروف الكلمة، مثل: مح مد –كا تب والصحيح محمد – كاتب.

المجال: الأخطاء الإملائية:

.....

النص المراد تحليله:

[النص]

الأمر الثالث:

You are an Arabic language expert specializing in analyzing orthographic, syntactical, morphological, semantic, and spelling errors. Your task is to analyze the following text and identify all writing-related errors.

Instructions:

1. **Review the entire text** carefully and identify any writing issues.
2. Analyze each error according to the detailed list below.

For each error, provide:

- The **incorrect word or phrase**.
- The **type of error**, based on the categories listed.
- A **count** of each error type.

Error Categories:

Domain: Orthographic (Writing) Errors

1. Lack of proper spacing between words
   Example:
   Incorrect ⟶ وعدمحمد –ويعلم
   Correct ⟶ وعد محمد –و يعلم
2. Unnecessary spaces within a single word
   Example:
   Incorrect ⟶ كا تب – مح مد
   Correct ⟶ كاتب –محمد

Text to be analyzed:
[Text]

مقياس التقييم:

للحكم على درجة تمكن النماذج من قدرة التعرف على الأخطاء الكتابية في نصوص متعلمي اللغة العربية لغة ثانية، حددت المقاييس التالية: (Accuracy,

Precision, Recall, F1-Score)

ويهدف كل مقياس منها لغرض محدد للحكم على مخرجات النماذج اللغوية الضخمة[١]:

:Accuracy

يستخدم مقياس الدقة العام Accuracy لتقييم الأداء الكلي للنموذج، من خلال حساب نسبة جميع الإجابات الصحيحة من إجمالي النص. والإجابات الصحيحة هنا يقصد بها ما تعرف بالفعل أنه خطأ، وما تجاهله؛ لأنه صواب. فلم يخطئ بوضعه في قائمة الأخطاء وهو صحيح، فمثلًا، إذا تعرف النموذج على ٦٠ كلمة أنها تحوي على خطأ من أصل ١٠٠ كلمة، ومن هذه الـ 60 كلمة، 10 كلمات لا يوجد بها خطأ فإن حساب دقته يكون كالتالي:

$$\text{Accuracy} = ((60 - 10) + ((100 - 60) - 10)) \div 100 = (50 + 30) \div 100 = 80\%$$

وشرحها كالتالي:

عدد الإجابات الصحيحة = 60 – 10

الكلمات الصحيحة التي تجاهلها لعدم وجود خطأ بها = (60 - 100) – 10 الدقة = عدد الإجابات الصحيحة + الكلمات الصحيحة التي تجاهلها لعدم وجود خطأ بها ÷ 100 = 80%

:Precision

(1) Fadel, A. S., O. A. Abulnaja, and M. E. Saleh. "Multi-Task Learning Model with Data Augmentation for Arabic Aspect-Based Sentiment Analysis." *CMC-Computers, Materials & Continua* 75, no. 2 (2023): p. 4434; Karajeh, O., M. N. Al-Kabi, and E. A. Fox. "Fusing AraBERT and Graph Neural Networks for Enhanced Arabic Text Classification." *2023 24th International Arab Conference on Information Technology (ACIT)*. IEEE (2023): p.5; Van Huyssteen, G. B., E. R. Eiselen, and M. J. Puttkammer. "Re-Evaluating Evaluation Metrics for Spelling Checker Evaluations." In *Proceedings of First Workshop on International Proofing Tools and Language Technologies*, (2004)

أما الدقة المحددة (Precision) فتركّز على قياس مدى صحة النموذج في تحديد الأخطاء. فمثلا لو حدّد النموذج 30 كلمة على أنها تحتوي على أخطاء كتابية، وتبين أن 25 منها فقط كانت أخطاء كتابية بالفعل وباقي الكلمات لم يكن فيها أخطاء، فالدقة المحددة تُحسب كالتالي:

$$\text{Precision} = 25 \div 30 = 83\%$$

:Recall

الاستدعاء (Recall) يقيس مدى قدرة النموذج على اكتشاف كل حالات الخطأ في النص. فمثلًا، إذا كان هناك ٤٠ خطأ كتابيا فعليًا في النص، واستطاع النموذج اكتشاف ٢٥ منها فقط، فإن الاستدعاء يكون:

$$\text{Recall} = 25 \div 40 = 62.5\%$$

:F1-Score

أما مقياس F1 (F1-Score) فيُستخدم للجمع بين الدقة المحددة والاستدعاء في رقم واحد متوازن، خصوصًا عند عدم توازن أنواع الأخطاء. فلو كانت الدقة ٨٣٪ والاستدعاء ٦٢,٥٪، فإن قيمة F1-Scoreهي كالتالي:

$$F1 = 2 \times (0.83 \times 0.625) \div (0.83 + 0.625) \approx 71.4\%$$

ويُعد هذا المقياس الأنسب لتقييم النماذج التي تتعامل مع أخطاء متباينة في التوزيع والتكرار.

نتائج الدراسة ومناقشتها

للإجابة عن سؤال الدراسة الأول الذي نصه " **ما الأخطاء الكتابية التي يقع فيها متعلمو اللغة العربية الناطقون بغيرها** "، فقد جمعت قائمة بالأخطاء الكتابية لدى متعلمي اللغة العربية من الدراسات والأدبيات السابقة، وبعد عرضها على مختصين جاءت الصورة النهائية للأخطاء الكتابية على النحو التالي:

جدول (٢) الصورة النهائية لقائمة الأخطاء الكتابية لدى متعلمي اللغة العربية لغة ثانية

مجال الخطأ	عدد أنواع الأخطاء
القواعد الكتابية	٢
الإملاء	١٦
علامات الترقيم	٣
الصرف	١
النحو	٩
الدلالة	٤
المجموع	٣٥

يعرض جدول (٢) التصنيف النهائي لأنواع الأخطاء الكتابية التي وقع فيها متعلمو اللغة العربية من الناطقين بغيرها، وقد بلغ مجموع أنواع الأخطاء المرصودة (35) نوعًا، موزعة على ستة مجالات رئيسة، وقد سبق تفصيل هذه الأنواع في جدول (١).

وتصدّر مجال الإملاء القائمة بـ 16نوعًا من الأخطاء، ما نسبته 45.7٪ من إجمالي الأنواع، وشمل أنواعًا كثيرة من الأخطاء، من ذلك الأخطاء المتعلقة بكتابة الهمزات، والأخطاء في كتابة التاء المربوطة والمفتوحة وغيرها من الأخطاء التي أشارت إليه الدراسات. تلاه مجال النحو الذي بلغت فيه الأخطاء (٩) أنواع ما نسبته (25.7٪)، فقد ذكرت الدراسات أخطاء في المستوى النحوي متنوعة وقع فيها متعلمو العربية بدءا من الإعراب وضبط الكلمات بناء على الموقع مرورا بالأخطاء المتعلقة بالجنس والعدد

والزمن وإسناد الفعل وغيرها من الأخطاء.

أما أخطاء الدلالة فقد جاءت في المرتبة الثالثة حيث بلغت (٤) أنواع ما نسبته (11.4٪)، وتتمحور حول أخطاء المعنى، مثل أخطاء اختيار الكلمة المناسبة، أو أخطاء انعدام الدلالة. ثم جاءت أخطاء علامات الترقيم التي بلغت (٣) أنواع ما نسبته (8.6٪)، وتليها أخطاء القواعد الكتابية، وتمثلت في نوعين ما نسبته (5.7٪)، أشار إليها المختصون في نقطة يكثر فيها أخطاء المتعلمين وهي تجاهل أو زيادة فراغ بين كلمة وأخرى أو بين حروف الكلمة الواحدة، وورد نوع واحد في مجال الأخطاء الصرفية ما نسبته (2.9٪).

وللإجابة عن سؤال الدراسة الثاني الذي نصه "**ما مستوى كفاءة النماذج اللغوية الضخمة في كشف وتصنيف الأخطاء الكتابية في كتابات متعلمي اللغة العربية لغة ثانية** "، جاءت النتائج بناء على أفضل أداء للنماذج على النحو التالي:

جدول (٣) نتائج تقييم أداء النماذج اللغوية الضخمة في تعرف أخطاء متعلمي اللغة

النموذج	Accuracy	precision	Recall	F1-Score
Gemini 2.5 Pro	0.8027	0.7802	0.4127	0.5399
GPT-4o	0.8016	0.8682	0.3449	0.4937
DeepSeek v 3.2	0.7516	0.5880	0.3817	0.4629
Allam 34B	0.6119	0.1008	0.0484	0.0654

يعرض الجدول (٣) مقارنة بين أداء أربعة نماذج لغوية ضخمة في التعرف على الأخطاء الكتابية في نصوص متعلمي اللغة العربية لغة ثانية، وذلك من خلال إجراء أربعة مقاييس أساسية على نتائج تلك النماذج وهذه المقاييس هي: الدقة

(Accuracy)، والاسترجاع (Recall) ، والدقة المحددة (Precision) ، ومعدل التوازن (F1-Score)، وقد أظهرت النتائج أن نموذج Gemini 2.5 Pro قدّم أفضل أداء نسبي بين النماذج الأربعة، حيث حقق أعلى معدل دقة (Accuracy) بلغ 0.80، وأداء متوازن نسبيًا في باقي المقاييس، إذ بلغت الدقة المحددة 0.78، بينما كان الاسترجاع منخفضًا نسبيًا عند 0.41، مما انعكس على معدل التوازن (F1) الذي سجّل 0.54 تقريبًا.

تلاه نموذج GPT-4o الذي سجّل أداءً قريبًا من Gemini 2.5 Pro في الدقة (0.80)مع تفوقه في الدقة المحددة (Precision) التي بلغت 0.87، إلا أن ضعف الاسترجاع (0.34) أثّر سلبًا على معدل التوازن الذي بلغ 0.49، مما يشير إلى أن النموذج دقيق في تحديد نوع الخطأ إلا أنه ليس ذا قدرة عالية على تعرف جميع الأخطاء في النص. أما نموذج DeepSeek v 3.2، فقد جاء في المرتبة الثالثة بدقة إجمالية 0.75، مع توازن نسبي في الدقة المحددة (0.59) والاسترجاع (0.38)، مما أدى إلى تحقيق معدل F1قدره 0.46. أما نموذج Allam فقد كان أضعف أداء بين النماذج المختبرة؛ إذ لم تتجاوز دقته 0.61، مع تدنٍ شديد في الدقة المحددة (0.10) والاسترجاع (0.05)، مما نتج عنه أدنى معدل F1 وهو 0.065 فقط.

وهذا يشير إلى ما لحجم النموذج وضخامة معامِلاته من دور كبير في رفع كفاءته، فإن نموذج علام بلغت معامِلاته ٣٤ مليار، أما بقية النماذج فقد تجاوزت هذا بكثير، فنموذج DeepSeek v 3.2بلغت معامِلاته 671 مليار معامِل، وبلغ نموذج GPT-4o ٢٠٠ مليار معامِل. أيضا هناك دور كبير لضخامة البيانات التي دربت عليها تلك النماذج، فإن نموذج Allam "علام" بلغ عدد المفردات التي دُرب عليها ٥.٢ ترليون مفردة، بينما نموذج DeepSeek v 3.2درب على 14.8 تريليون مفردة.

وفي نهاية المطاف فجميع أداءات النماذج لم ترق للمستوى المطلوب الذي يمكن الاعتماد عليه، فجميع قيم F1-Score تعد قيماً لا يمكن الركون إليها في أداء مهمة

كشف وتصنيف أخطاء متعلمي اللغة العربية في النصوص، ويظل الأمر رهين إشراف معلم اللغة العربية؛ لاستثمار هذه النماذج في تنمية مهارات متعلمي اللغة، وهذا يتوافق مع ما ذكرته دراسة Al-Alami [1]. ويحيل الباحث هذا التدني في القيم إلى كثرة الأخطاء في النصوص من جهة، ومن جهة أخرى التباس نوع الخطأ على النماذج في عدد منها فعلى سبيل المثال لوحظ أن النماذج تحيل الخطأ في حذف النقاط (الإعجام) إلى استبدال حرف في مثل كلمة (الرياص)، وهذا يعد خطأ إعجام وليس استبدال كما في (بعد) وأراد المتعلم كلمة (بعض).

أيضا وُجد أن النماذج يلتبس عليها إسناد الخطأ المتعلق بالكلمة؛ لتردد هذا الخطأ في مجال الصرف والنحو والدلالة، فجميع هذه المجالات ورد في أنواعها مفردة كلمة. أيضا من الأمور التي تربك النماذج الأخطاء المركبة في كلمة واحدة، فمثلا كلمة (اسيارت) حوت خطأين الأول في اللام الشمسية والثاني في كتابة التاء المربوطة تاء مفتوحة. كما لوحظ ظهور الهلوسة على هذه النماذج فتأتي بنتائج غير صحيحة، ولكن فقط تحاكي الموجّه المدخل لها، فقدمت النماذج أنواع أخطاء جديدة غير صحيحة وغير موجودة، وأشارت إلى كلمات صحيحة بأنها تضمنت أخطاء وأسندتها إلى أنواع لا تنتمي إلى الخطأ المذكور.

وتجدر الإشارة هنا إلى أن هذه النماذج لم تبن في الأصل للتعرف وتحليل الأخطاء، إنما هي نماذج عامة تركز على إنتاج النصوص من خلال الرد على ما يطلب منها، وهذا يجب أن يؤخذ بعين الاعتبار؛ لأن بنية هذه النماذج التقنية وتدريبها وبياناتها لا تركز على هذه المهمة.

وللإجابة عن سؤال الدراسة الثالث الذي نصه "**ما الموجّه الذي يقدم أفضل أداء لنتائج النماذج اللغوية الضخمة في كشف وتصنيف الأخطاء الكتابية لدى**

(1) Al-Alami, "EFL Learners' Attitudes Towards Utilizing ChatGPT for Acquiring Writing Skills in Higher Education"

متعلمي اللغة العربية لغة ثانية؟ " وبعد تنفيذ جميع صيغ الموجّهات، جاءت النتائج النحو التالي:

جدول (٤) نتائج النماذج اللغوية بناء على صيغة الموجّه في كشف وتصنيف الأخطاء الكتابية

النموذج	الموجّه	Accuracy	precision	Recall	F1-Score
GPT-4o	أمر ١	0.8016	0.8900	0.3365	0.4889
GPT-4o	أمر ٢	0.8016	0.8683	0.3450	0.4938
GPT-4o	أمر ٣	0.8149	0.8814	0.3801	0.5294
Gemini Pro 2.5	أمر ١	0.7869	0.7784	0.3646	0.4970
Gemini Pro 2.5	أمر ٢	0.7869	0.7692	0.3690	0.4992
Gemini Pro 2.5	أمر ٣	0.8027	0.7802	0.4128	0.5399
DeepSeek v 3.2	أمر ١	0.7500	0.6780	0.3165	0.4305
DeepSeek v 3.2	أمر ٢	0.7516	0.5881	0.3818	0.4630
DeepSeek v 3.2	أمر ٣	0.7335	0.5997	0.3081	0.4066
Allam 34B	أمر ١	0.5982	0.1117	0.0317	0.0483
Allam 34B	أمر ٢	0.6119	0.1008	0.0484	0.0654
Allam 34B	أمر ٣	0.5778	0.0778	0.0317	0.0447

يعرض الجدول (٤) تأثير اختلاف صيغة الموجّه في أداء النماذج اللغوية الضخمة في مهمة كشف وتصنيف الأخطاء الكتابية. ويتضح من النتائج أن هناك تأثيرا واضحا في هذا الباب كما أشارت الدراسات السابقة؛ إذ أظهرت بعض الصيغ تحسناً في التوازن بين مقاييس الأداء، ولا سيما في قيم الـ F1-Score التي تعكس مدى المستوى الدقيق للنموذج في التعرف على الأخطاء.

وقد أظهر نموذج GPT-4o عند استخدام صيغة الموجّه الثالث أفضل أداء إجمالي بين جميع التركيبات؛ إذ حقق أعلى دقة (Accuracy = 0.81) وأعلى دقة محددة (Precision = 0.88)، مع معدل توازن (F1 = 0.53)، مما يشير إلى أن هذه الصيغة كانت الأنسب في توجيه النموذج للتعرف على الأخطاء على نحو متزن نسبيًا. أما نموذج Gemini Pro 2.5 فقد سجل أداءً جيدًا في جميع صيغ موجهاته الثلاثة، مع تفوق ملاحظ لصيغة الموجّه الثالث التي حققت أفضل قيمة لـF1 بلغت (0.5399) مقارنة ببقية الموجّهات، وهو ما يدل على أن هذا الموجّه ساعد النموذج على تحقيق توازن أفضل بين الدقة المحددة (0.7802) والاسترجاع (0.4128). أما نموذج DeepSeek v3.2فقد سجل أداءً متقاربًا في صيغ موجهاته الثلاثة، مع تفوق ملاحظ لصيغة الموجّه الثاني التي حققت أفضل قيمة لـ F1 بلغت (0.46) مقارنة ببقية الموجّهات، وهو ما يدل على أن هذا الموجّه ساعد النموذج على تحقيق توازن أفضل بين الدقة المحددة (0.59) والاسترجاع .(0.38)

أما نموذج Allam 34B فقد أظهر أداءً ضعيفًا في جميع صيغ موجهاته الثلاثة، مع تفوق طفيف لصيغة الموجّه الثاني التي حققت أفضل قيمة لـF1 بلغت (0.0654) مقارنة ببقية الموجّهات، مما يشير إلى أن هذا الموجّه كان الأنسب نسبيًا لهذا النموذج رغم محدودية قدرته العامة على التعرّف على الأخطاء؛ إذ بلغت دقته المحددة (0.1008) واسترجاعه (0.0484).

وهذا يشير كما ذكرت الدراسات إلى أن صيغة الموجّه تمثل عاملاً حاسمًا في

تحسين أداء النماذج اللغوية الضخمة؛ إذ يمكن لتعديل بسيط في طريقة صياغة الموجّه أن يُحدث فرقا ملموسًا في النتائج. كما تكشف النتائج أن النماذج الأضخم تستجيب بشكل أوضح لتغيّر صيغ الموجّهات أكثر من النماذج الصغيرة وهو في هذا السياق نموذج ALLam الذي بلغت معامِلاته 34 مليار.

وتجدر الإشارة إلى أن الأرقام في الجدول (٤) تشير إلى أن أداء النماذج في صيغة الموجّه الأول كان الأكثر دقة في تحديد نوع الخطأ فعلا، ولكن لم يكن قادرا على تعرف جميع الأخطاء في النص، وهنا اتضح ما لبقية صيغ الموجّهات من حث للنماذج لاستظهار جميع ما في النص من أخطاء كما توضحه قيم Recallالاسترجاع في الجدول.

ومع ذلك، فإن قيم F1-Score المتحققة عمومًا لا تزال متواضعة، ولا ترقى إلى مستوى الاعتماد الكامل على النماذج في مهمة كشف وتصنيف الأخطاء الكتابية دون إشراف بشري. وهذا يؤكد ما توصلت إليه نتائج السؤال السابق من أن هذه النماذج لم تُصمَّم أساسًا لتحليل الأخطاء، بل لإنتاج اللغة ومعالجتها بصورة عامة، مما يجعل دقة المخرجات مرهونة في الوقت الحالي بجودة الموجّه وتحديد المهمة بدقة ووضوح، واستخدام أساليب أخرى، وطرائق في التعامل مع النماذج لرفع كفاءتها.

النتائج والتوصيات:

ألقى الذكاء الاصطناعي بظلاله على العديد من المجالات، وأحد هذه المجالات الحيوية مجال التعليم، وقد قام المختصون مذ ظهوره بتوظيفه في شتى مجالات التعليم؛ لما يوفره من مميزات تقدم حلولا وتيسيرا في كثير من زوايا العملية التعليمية. إلا أن هذا التوظيف يجب أن يسبقه وقوف على القدرات الفعلية لتلك الأدوات، وأنها قادرة على تقديم نتائج دقيقة جدا في الميدان الذي تُوظف فيه، وبدون هذا الأمر قد يكون في توظيفها أثر عكسي على المتعلمين؛ لذلك سعت الدراسة إلى الوقوف على قدرتها في التعرف على أخطاء المتعلمين الناطقين بلغات أخرى، وأسفرت النتائج إلى أن النماذج اللغوية الضخمة لا يمكن الاعتماد عليها في هذا الميدان وأقصى نسبة دقة حققتها بلغت53%، وهذا ليس بكافٍ.

وتجدر الإشارة في هذا السياق إلى أنه لوحظ أن قدرة النماذج على تصحيح النص أفضل من قدرتها على كشف وتصنيف الأخطاء ، وكأنّ الأمر فيه أمر عربي قحٌ قادر على استخدام اللغة وتصحيح الخطأ، ولكن لو سئل عن الاستثناء المفرغ في النص لما اهتدى إليه، وهنا تقترح الدراسة النظر إلى استثماره بطرائق أخرى يوصي بدراستها، منها أن تستخدم أساليب تقوم على سلسلة الأفكار CoT من خلال تقسيم النص إلى جمل أقصر، ثم تصحيحها ومطابقتها من النص الأصلي، ثم يطلب من النموذج أن يستخرج الفرق ليحدد نوع الخطأ في النص، وربما دُمج معها أحد أساليب الاستدلال للتأكد من الإجابة وخفض هلوسة النماذج.

نتائج الدراسة:

توصلت الدراسة إلى عدد من النتائج المهمة في إطار تقييم أداء النماذج اللغوية الضخمة في كشف وتصنيف الأخطاء الكتابية لدى متعلمي اللغة العربية لغة ثانية، ومن أبرز هذه النتائج:

١. قائمة بالأخطاء الكتابية لدى متعلمي اللغة العربية لغة ثانية، اشتملت على ٣٥ نوعًا من الأخطاء موزعة على المجالات التالية: قواعد الكتابة، والإملاء، وعلامات الترقيم، والصرف، والنحو، والدلالة.

٢. أظهرت نتائج التقييم تفوق نموذج Gemini Pro 2.5 حيث بلغت قيمة F1-Score 0.5399 في التعرف على الأخطاء الكتابية، تلاه نموذج Gpt-4o بقيمة بلغت 0.4937ثم DeepSeek v3.2الذي حقق 0.4629، تلا ذلك بمستوى منخفض بشكل كبير نموذج ALLaM 34Bفقد حصل على 0.0654 في قيمة F1-Score.

٣. بينت الدراسة أن حجم النموذج وضخامة معامِلاته، وحجم البيانات التي يُدرب عليها، تؤدي دورا بارزا في قدرتها على إدراك ما هو مطلوب منها تنفيذه.

٤. أظهرت الدراسة أن صياغة الموجّهات (Prompts) تؤدي دورًا كبيرًا في تحسين نتائج النماذج، حيث حقق الموجّه التفصيلي سواء بالإنجليزية أم العربية أعلى أداء، موزعا بين النماذج.

التوصيات:

في ضوء النتائج التي توصلت إليها الدراسة فإنها توصي بالنقاط التالية:

١. ضرورة بناء معايير تحليل دقيقة وشاملة لتقييم أداء النماذج اللغوية الضخمة في المجالات اللغوية المختلفة، لا سيما في مهام اللغة العربية المرتبطة بتعليمها.

٢. ضرورة توظيف النماذج اللغوية في ميدان التعليم، ولكن بحذر مع ضرورة إشراف المعلم على العملية التعليمية؛ لتحقيق أقصى فائدة من النماذج اللغوية.

٣. تحسين تصميم الموجّهات (Prompts) المقدمة للنماذج، وجعلها أكثر وضوحًا وتفصيلًا؛ لتفادي الإجابات غير الدقيقة والهلوسة قدر الإمكان التي قد تنتج من النماذج.

٤. ضرورة اعتماد بيانات أصيلة ناتجة عن متعلمين حقيقيين عند اختبار النماذج؛ لضمان تمثيل كافٍ للواقع اللغوي والتحديات الكتابية الفعلية التي يواجهها متعلمو العربية.

٥. تشجيع البحث المستقبلي على مقارنة أداء النماذج اللغوية في سياقات متنوعة، مع تجربة المزيد من أساليب صياغة الموجّهات التي قد تحسن أداء النماذج بشكل أكبر.

٦. تدريب معلمي اللغة العربية على استخدام النماذج اللغوية الضخمة في التعليم، ويقصد بالتدريب هنا أن تقدم لهم أمثل الأساليب وأفضلها؛ لاستثمار هذه الأدوات وتحقيق الاستفادة القصوى منها في مجال التعليم.

المصادر والمراجع:

أولا– المراجع العربية:

الأفيوني، بشار مصطفى، وشالكينسكايا، أرينا. "تحليل الأخطاء اللغوية في الإنتاج الكتابي لدى متعلمي اللغة العربية الروس في جامعة موسكو الحكومية." مجلة الدراسات اللغوية والأدبية، س١٥، ع٢، (٢٠٢٤م): ٦٧–٩٠.

أمين، مروة مصطفى السيد. "أخطاء الكتابة بين متعلمي العربية الناطقين بها والناطقين بغيرها في المرحلة الجامعية الأولى: كلية الألسن جامعة عين شمس نموذجًا." فيلولوجي: سلسلة في الدراسات الأدبية واللغوية، ع٦٥، (٢٠١٦م): ٣٥–٦٩.

بني عرابة، إخلاص بنت إبراهيم بن حمد، والكاف، فاطمة بنت محمد بن أحمد. "فاعلية بعض تطبيقات الذكاء الاصطناعي في تنمية مهارات القراءة الإبداعية وبقاء أثر التعلم لدى طالبات الصف السادس الأساسي." المجلة العربية للنشر العلمي، ع٧٨، (٢٠٢٥م): ١١٢–١٣١.

بوبنديرة ، عبدالعزيز. "الامتحانات التقليدية ومشكلات التقييم." مجلة البحوث التربوية والتعليمية ١٣ (عدد خاص)، (٢٠٢٤م): ٤٤١–٤٥٨.

طعيمة، رشدي أحمد. "تعليم العربية لغير الناطقين بها: مناهجه وأساليبه." المنظمة الإسلامية للتربية والعلوم والثقافة (الإيسيسكو)، (١٩٨٩م)

عبد اللوي، محمد. "توظيف الذكاء الاصطناعي في تدريس اللغة العربية للناطقين بغيرها عن بعد." مجلة التطوير العلمي للدراسات والبحوث، ع١٧، (٢٠٢٤م): ٤٠٢–٤١٧.

عنتر، دينا. "دمج الذكاء الاصطناعي في تعليم اللغة العربية في التعليم العالي: دراسة حالة على ChatGPT 4 Education." مجلة الأرائك للعلوم والإنسانيات، ع٦، (٢٠٢٤م): ٤٦٧–٤٩١.

محسني، فاطمة. "دمج الذكاء الاصطناعي في تعليم اللغة العربية للناطقين بغيرها: إستراتيجيات فعالة لتطوير المهارات اللغوية." مجلة الباحث، (٢٠٢٥م): ١–١٣.

الملحم، تركي عبد العزيز. "واقع استخدام تطبيقات الهواتف الذكية في تعليم اللغة العربية للناطقين بلغات أخرى في معهد تعليم اللغة العربية لغير الناطقين بها بالجامعة الإسلامية من وجهة نظر المعلمين." مجلة كلية التربية (أسيوط) ٣٧، ع٢، (٢٠٢١م): ٣٩–١٠٨.

الهذلول، علي بن هذلول علي. "أثر تطبيقات الذكاء الاصطناعي في تنمية مهارات الكتابة الأكاديمية لدى متعلمي اللغة العربية الناطقين بلغات أخرى." مجلة جامعة الملك عبد العزيز – العلوم التربوية والنفسية، مج٤، ع٢، (٢٠٢٥م): ٥٠–٦٩.

ثانيا– المراجع الأجنبية:

Al-Alami, Suhair E. "EFL Learners' Attitudes Towards Utilizing ChatGPT for Acquiring Writing Skills in Higher Education: A Case Study of Computing Students." *Journal of Language Teaching & Research* 15, iss. 4 (2024): 1029.

Al-khafaji, N. J., and B. K. Majeed. "Evaluating Large Language Models Using Arabic Prompts to Generate Python Codes." *2024 4th International Conference on Emerging Smart Technologies and Applications (eSmarTA)* (2024.

Al-Khalifa, Shahad; Al-Khalifa, Hend. The Qiyas Benchmark: Measuring ChatGPT Mathematical and Language Understanding in Arabic. arXiv preprint arXiv:2407.00146, 2024.

Alahmadi, M. D., Alharbi, M., Tayeb, A., and M. Al-Shanqiti. "Evaluating Large Language Models' Proficiency in Answering Arabic GAT Exam Questions." *Engineering, Technology & Applied Science Research* 14, iss. 6 (2024): 17774–17780.

Alfaifi, Abdullah, and Eric Atwell. "Potential Uses of the Arabic Learner Corpus." Paper presented at the *Leeds Language, Linguistics and Translation PGR Conference 2013*, Leeds, UK (2013).

Al-Husain, A., and A. M. Azmi. "Beyond Event-Centric Narratives: Advancing Arabic Story Generation with Large Language Models and Beam Search." *Mathematics* 12, no. 10 (2024): 1548.

Almazrouei, E., Cojocaru, R., Baldo, M., Malartic, Q., Alobeidli, H., Mazzotta, D., Penedo, G., Campesan, G., Farooq, M., Alhammadi, M., Launay, J., and Noune, B. "AlGhafa Evaluation Benchmark for Arabic Language Models." *Proceedings of the First Arabic Natural Language Processing Conference (ArabicNLP 2023)* (2023): 244–275.

Al-Shammari, W., and S. Alhumoud. "TAQS: An Arabic Question Similarity System Using Transfer Learning of BERT with BiLSTM." *IEEE Access* 10 (2022): 91509–91522.

Ayodele, O. S., Aliu, J., Owoeye, F. O., Ajayi, E. A., and Sheidu, A. Y. "The Role of Artificial Intelligence in Curriculum Development and Management." *Journal of Digital Innovations & Contemporary Research in Science, Engineering & Technology* 11, no. 2 (2023): 37–46.

Bansal, P. "Prompt Engineering Importance and Applicability with Generative AI." *Journal of Computer and Communications* 12 (2024): 14–23.

Bari, M. Saiful, et al. Allam: Large language models for Arabic and English. *arXiv preprint arXiv:2407.15390*, 2024.

Ben Abacha, A., W.-w. Yim, Y. Fu, Z. Sun, M. Yetisgen, F. Xia, and T. Lin. "MEDEC: A Benchmark for Medical Error Detection and Correction in Clinical Notes (arXiv:2412.19260v1)." *arXiv preprint* (2024).

Biswas, Md. R., Mohsen, F., Shah, Z., and Zaghouani, W. "Potentials of ChatGPT for Annotating Vaccine-Related Tweets." *2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS)* (2023): 1-6.

Bonner, Euan; Lege, Ryan; Frazier, Erin. Large Language Model-Based Artificial Intelligence in the Language Classroom: Practical Ideas for Teaching. Teaching English with Technology, (2023), 23.1: 23-41.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., and Xie, X. "A Survey on Evaluation of Large Language Models." *ACM Transactions on Intelligent Systems and Technology* 15, no. 3 (2024): 1–45.

DeepSeek-AI. "DeepSeek-V3 Technical Report (arXiv:2412.19437v2)." *DeepSeek-AI* (2025)

Dong, Bingyu, et al. "Large Language Models in Education: A Systematic Review." *2024 6th International Conference on*

Computer Science and Technologies in Education (CSTE) (2024): 131–134.

Einieh, Y., Almansour, A., and Jamal, A. "Fine-Tuning an AraT5 Transformer for Arabic Abstractive Summarization." *2022 14th International Conference on Computational Intelligence and Communication Networks (CICN)* (2022): 194–197.

Ejjami, R. "The Future of Learning: AI-Based Curriculum Development." *International Journal for Multidisciplinary Research* 6, no. 4 (2024): 1–31.

Fadel, A. S., O. A. Abulnaja, and M. E. Saleh. "Multi-Task Learning Model with Data Augmentation for Arabic Aspect-Based Sentiment Analysis." *CMC-Computers, Materials & Continua* 75, no. 2 (2023): 4419–4442.

Farraj, K. K. "Tahlil Al-Akhtha 'inda Muta'allimi Al-Lughah Al-'Arabiyah Li Al-Naathiqiina Bi Ghairiha." *Arabiyat: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban* 2, no. 1 (2015): 112–129.

Fincham, Naiyi Xie; Alvarez, Aitor Arronte. Using large language models (llms) to facilitate l2 proficiency development through personalized feedback and scaffolding: An empirical study. In: Proceedings of the International CALL Research Conference. 2024. p. 59-64.

Hamaniuk, Vita A. "The Potential of Large Language Models in Language Education." *Educational Dimension* 5 (2021): 208–210.

Hilal, Mahmud, and Ahmet İsmailoğlu. "Arapça Öğretimi Gören Öğrencilerin Yaygın Dilbilimsel Hataları ve Çözümler Üzerine Uygulamalı Alan Çalışması: Kırıkkale Üniversitesi K.K.Ü.FEF. Arapça M.T.B Örneği." *Kırıkkale Üniversitesi Sosyal Bilimler Dergisi* 10, no. 2 (July 2020): 731–742.

Jaashan, H. M. S., and A. A. Alashabi. "Using AI Large Language Model (LLM-ChatGPT) to Mitigate Spelling Errors of EFL Learners." *Forum for Linguistic Studies* 7, no. 3 (2025): 328–339.

Karajeh, O., M. N. Al-Kabi, and E. A. Fox. "Fusing AraBERT and Graph Neural Networks for Enhanced Arabic Text Classification." *2023 24th International Arab Conference on Information Technology (ACIT)*. IEEE (2023).

Khondaker, M. T. I., Naeem, N., Khan, F. L., Elmadany, A. A., and M. Abdul-Mageed. "Benchmarking LLaMA-3 on Arabic Language Generation Tasks." *Proceedings of the Second Arabic Natural Language Processing Conference* (2024): 283–297.

Koto, F., Li, H., Shatnawi, S., Doughman, J., Sadallah, A. B., Alraeesi, A., Almubarak, K., Alyafeai, Z., Sengupta, N., Shehata, S., Habash, N., Nakov, P., and T. Baldwin. "ArabicMMLU: Assessing Massive Multitask Language Understanding in Arabic." *arXiv preprint* arXiv:2402.12840 (2024).

Lamsiyah, S., Zeinalipour, K., El Amrany, S., Brust, M., Maggini, M., Bouvry, P., and Schommer, C. "ArabicSense: A Benchmark for Evaluating Commonsense Reasoning in Arabic with Large Language Models." *Proceedings of the 4th Workshop on Arabic Corpus Linguistics (WACL-4)* (2025): 1–11.

Mohseni, Fatima. "Integrating Artificial Intelligence into Teaching Arabic as a Foreign Language: Effective Strategies for Developing Linguistic Skills." (in Arabic). Al-Bāḥith Journal, (2025): 1–13.

Mulyanto, D., Zaki, M., Ridho, A., and Fata, K. "Artificial Intelligence Utilization for Arabic Language Skills Development in Arabic Language Learning." *An-Nidzam: Jurnal Manajemen Pendidikan dan Studi Islam* 11, no. 1 (2024): 18–32.

Nehar, A., Bellaouar, S., Souffi, S., and Bouameur, M. "Dhati: A Fine-Tuned Large Language Model for Evaluating Subjectivity in Arabic Textual Data." *Proceedings of the 2023 5th International Conference on Pattern Analysis and Intelligent Systems (PAIS)* (2023): p. 1-7

Panagiotidis, P. "LLM-Based Chatbots in Language Learning." *European Journal of Education* 7, no. 1 (2024): 102–123.

Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., and A. Chadha. "A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications." *arXiv preprint* (2024).

Van Huyssteen, G. B., E. R. Eiselen, and M. J. Puttkammer. "Re-Evaluating Evaluation Metrics for Spelling Checker Evaluations." In *Proceedings of First Workshop on International Proofing Tools and Language Technologies*, (2004) :91–99

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., and D. C. Schmidt. "A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT." *arXiv preprint* (2023).

Yavuz, F., Çelik, Ö., and G. Yavaş Çelik. "Utilizing Large Language Models for EFL Essay Grading: An Examination of Reliability and Validity in Rubric-Based Assessments." *British Journal of Educational Technology* 56 (2025): 150–166.

Zhang, Z., and X. Huang. "The Impact of Chatbots Based on Large Language Models on Second Language Vocabulary Acquisition." *Heliyon* 10 (2024): e25370.

Zheng, R. "AWE and LLMs in L2 Writing Feedback: An Exploration of EFL Learners' Experiences." *Academic Journal of Humanities & Social Sciences* 8, no. 7 (2025): 60–69.

Zheng, Y., Li, T., Huang, H., Zeng, T., Lu, J., Chu, C., Huang, Y., Jiang, Z., Xiong, Q., Ge, Y., and M. Li. "Are All Prompt Components Value-Neutral? Understanding the Heterogeneous Adversarial Robustness of Dissected Prompt in Large Language Models." *arXiv preprint* (2025).

Zuccon, G., and B. Koopman. "Dr ChatGPT, Tell Me What I Want to Hear: How Prompt Knowledge Impacts Health Answer Correctness." *arXiv preprint* (2023).

ثالثا– الموقع الالكتروني:

Google DeepMind. "Gemini Model Thinking Updates: Advanced Coding." Google Blog, March 2025. Retrieved October 19, 2025, from https://blog.google/technology/google-deepmind/gemini-model-thinking-updates-march-2025/#advanced-coding.

Bibliography

Abdelloui, Muhammad. “Employing Artificial Intelligence in Teaching Arabic as a Foreign Language Remotely.” (in Arabic). Journal of Scientific Development for Studies and Research, iss. 17 (2024): 402–417.

Afyouni, Bashar Mustafa, & Shalkinskaya, Arina. “Linguistic Error Analysis in the Written Production of Russian Learners of Arabic at Moscow State University.” (in Arabic). Journal of Linguistic and Literary Studies, vol. 15, iss. 2 (2024): 67–90.

al-Hadhlūl, ʻAlī ibn Hadhlūl ʻAlī. “The Impact of Artificial Intelligence Applications on Developing Academic Writing Skills among Arabic Learners Who Speak Other Languages.” (in Arabic). King Abdulaziz University Journal – Educational and Psychological Sciences, vol. 4, iss. 2 (2025): 50–69.

al-Mulḥim, Turkī ʻAbd-al-ʻAzīz. “The Reality of Using Smartphone Applications in Teaching Arabic to Speakers of Other Languages at the Institute for Teaching Arabic to Non-Native Speakers at the Islamic University, from the Teachers’ Perspective.” (in Arabic). Assiut University, Faculty of Education Journal, vol. 37, iss. 2 (2021): 39–108.

Amin, Marwa Mustafa Al-Sayed. “Writing Errors among Native and Non-Native Learners of Arabic in the First University Stage: The Faculty of Al-Alsun, Ain Shams University as a case study.” (in Arabic). Philology: Series in Literary and Linguistic Studies, iss. 65 (2016): 35–69.

ʻAntar, Dīnā. “Integrating Artificial Intelligence into Teaching Arabic in Higher Education: A Case Study on ChatGPT-4 Education.” Al-Araek Journal for Sciences and Humanities, iss. 6 (2024): 467–491.

Banī ʻUrābah, Ikhlāṣ bint Ibrāhīm ibn Ḥamad and al-Kāf, Fāṭimah bint Muḥammad ibn Aḥmad. “The Effectiveness of Selected Artificial Intelligence Applications in Developing Creative Reading Skills and Retention of Learning Among Sixth-Grade Female Students.” (in Arabic). Arab Journal for Scientific Publishing, isss. 78 (2025): 112–131.

Boubandirah, Abdelaziz. "Traditional Examinations and Assessment Problems." (in Arabic). Journal of Educational and Pedagogical Research, vol. 13 (Special Issue), (2024): 441–458.

Tu'aimah, Rushdi Ahmad. "Teaching Arabic to Non-Native Speakers: Its Curricula and Methods". (in Arabic). Islamic Educational, Scientific and Cultural Organization (ISESCO), 1989.


الجامعة الإسلامية بالمدينة المنورة
١٣٨
١٩٦

THE ISLAMIC UNIVERSITY JOURNAL OF ARABIC LANGUAGE AND LITERATURE PART 2

THE ISLAMIC UNIVERSITY JOURNAL OF ARABIC LANGUAGE AND LITERATURE

Part 2

Issue: 19
January - March 2026