تقييم أداء نماذج الذكاء الاصطناعي في وزن الأفعال العربية: دراسة وصفية مقارنة بين (ChatGPT) و(Gemini)

Evaluating the Performance of Artificial Intelligence Models in Identifying Arabic Verb Patterns: A Comparative Descriptive Study of ChatGPT and Gemini

10.35781/1637-000-173-007

د. محمد عدلي محمد عودة*

د. طارق إبراهيم محمود الزيادات **

*جامعة زايد، الإمارات العربية المتحدة

**جامعة الإسراء، الأردن

الملخص

هدف هذا البحث إلى تقييم كفاية نماذج الذكاء الاصطناعي (ChatGPT, Gemini) وفي معالجة الميزان الصرفي للأفعال العربية بوصفه أحد الأسس الرئيسة في تحليل البنية اللغوية. واعتمد البحث المنهج الوصفي التحليلي المقارن مدعوماً بالتحليل الكمي لقياس دقة الأداء، وذلك من خلال عينة مكونة من (16) فعلًا تمثل الأوزان الثلاثية والرباعية والخماسية والسداسية في حالتي الصحة والاعتلال، إضافة إلى ظاهرتي الإعلال والإدغام. وأظهرت النتائج أن نسبة الخطأ بلغت (43.75%) في نموذج (ChatGPT) مقابل (6.25%) في نموذج(Gemini) ، في حين سجلت نسبة الدقة (56.25%) و(93.75%) على التوالي، مما يعكس تفوقًا واضحًا لصالح (Gemini) في أداء مهمة تحديد الأوزان الصرفية. كما بينت النتائج تحسن أداء النموذجين في الأوزان القياسية، مقابل ضعف نسبي في الأفعال المعتلة وصيغ الأمر. وخلصت الدراسة إلى تفوق (Gemini) نسبيًا وإمكانية توظيفه أداةً مساندة في تعليم الصرف، مع التأكيد على ضرورة المراجعة اللغوية المتخصصة لمخرجات نماذج الذكاء الاصطناعي، والتوصية بتطوير نماذج عربية مدرّبة على بيانات صرفية أعمق للحد من الأخطاء والهلوسة اللغوية، بما يعزز دقة تطبيقاتها في تعليم اللغة العربية ومعالجة الظواهر الصرفية المعقدة، ويفتح آفاقًا بحثية مستقبلية في الدراسات التطبيقية للصرف العربي والذكاء الاصطناعي.

الكلمات المفتاحية :الذكاء الاصطناعي، ChatGPT، Gemini، الميزان الصرفي، الأفعال العربية، المعالجة الآلية، تعليم اللغة العربية.

ISSN : 2410-1818

Evaluating the Performance of Artificial Intelligence Models in Identifying Arabic Verb Patterns: A Comparative Descriptive Study of ChatGPT and Gemini

Dr. Mohammad Adli Mohammad Odeh

Dr. Tareq Ibrahim Mahmoud Al-Zyadat

Abstract

This study aimed to evaluate the effectiveness of artificial intelligence models (ChatGPT and Gemini) in processing the Arabic morphological scale (al-mīzān al-ṣarfī) in verbs, as a key component of linguistic structure analysis. The study adopted a descriptive-analytical comparative approach supported by quantitative analysis to measure performance accuracy. The sample consisted of 16 verbs representing triliteral, quadriliteral, quinquiliteral, and sextuple patterns in both regular and irregular forms, in addition to morphological phenomena such as vowel change (iʿlāl) and assimilation (idghām).

The results indicated that the error rate was (43.75%) for ChatGPT compared to (6.25%)for Gemini, while the accuracy rates were (56.25%)and (93.75%) respectively, demonstrating a clear superiority of Gemini in morphological pattern identification. Both models showed better performance in standard patterns, whereas weaknesses appeared in irregular verbs and imperative forms. The study concluded that Gemini can be considered a more effective supportive tool in teaching Arabic morphology, while emphasizing the need for linguistic review of AI outputs and the development of Arabic-specific trained models to reduce errors and hallucinations and improve computational linguistic applications.

Keywords: Artificial Intelligence, Arabic Morphology, Verb Patterns, ChatGPT, Gemini, Arabic Language Teaching

.

ISSN : 2410-1818

المقدمة:

شهد العالم في العقود الأخيرة تطورًا ملحوظًا في تقنيات الذكاء الاصطناعي، وأصبح لها دور متزايد في مجالات متعددة، من التعليم إلى الصناعة والصحة. وقد أسهم هذا التطور في ظهور أدوات رقمية قادرة على معالجة اللغات وتحليلها بدرجات متفاوتة، الأمر الذي أتاح إمكانية توظيفها في تعليم اللغة العربية وتحليل بنيتها اللغوية.

وفي مجال تعليم اللغة العربية، بدأت تطبيقات الذكاء الاصطناعي تسهم في دعم العملية التعليمية، من خلال تقديم خدمات متنوعة، مثل: تحليل الكلمات، وتحديد الأوزان الصرفية، وتوضيح البنى الصرفية المختلفة. ويعد الميزان الصرفي من أهم أدوات التحليل في اللغة العربية، لما له من دور في الكشف عن بنية الكلمة وضبط أوزانها.

أهمية البحث:

تكمن أهمية هذا البحث في إسهامه في إثراء الدراسات المتعلقة بالذكاء الاصطناعي ومعالجة اللغة العربية، من خلال تقديم دراسة تطبيقية تقيّم قدرة نماذج الذكاء الاصطناعي على تحديد الأوزان الصرفية للكلمات العربية. ويساعد هذا البحث المعلمين والباحثين وصنّاع النماذج على فهم مستوى أداء هذه الأدوات في التعامل مع الميزان الصرفي، بوصفه أحد أهم أدوات التحليل الصرفي في اللغة العربية، مما يسهم في توظيفها بصورة أكثر وعيًا في تعليم الصرف العربي.

مشكلة البحث:

تتمثّل مشكلة هذا البحث في التساؤل عن مدى قدرة نماذج الذكاء الاصطناعي الحديثة على التعامل مع الميزان الصرفي في اللغة العربية بدقة وموثوقية، ولا سيما في تحديد الأوزان الصرفية للأفعال العربية. فعلى الرغم من التطور الكبير الذي شهدته هذه النماذج في معالجة اللغات الطبيعية، إلا أن تطبيقها في المجال الصرفي العربي ما يزال يثير إشكالات تتعلق بالدقة والاتساق في تحليل البنية الصرفية للكلمة. ومن هنا تتحدد مشكلة البحث في تقييم مدى كفاية نماذج الذكاء الاصطناعي في تحديد الأوزان الصرفية للأفعال العربية، والكشف عن درجة موثوقية مخرجاتها مقارنة بالتحليل الصرفي الصحيح.

أهداف البحث:

1 -دراسة قدرة نموذجي الذكاء الاصطناعي(ChatGPT) و (Gemini) على تحديد الأوزان الصرفية للأفعال العربية.

2 -تقييم دقة هذين النموذجين في التعامل مع الميزان الصرفي في اللغة العربية.

ISSN : 2410-1818

أسئلة البحث:

1 -ما مدى قدرة نموذجي الذكاء الاصطناعي (ChatGPT) و (Gemini)على تحديد الأوزان الصرفية للأفعال العربية؟

2 -ما مدى دقة هذين النموذجين في التعامل مع الميزان الصرفي للأفعال العربية؟

3- ما نقاط القوة والضعف في أداء هذين النموذجين في تحديد الأوزان الصرفية؟

فرضيات البحث:

1- توجد فروق ذات دلالة في مستوى دقة تحديد الأوزان الصرفية للأفعال العربية بين نموذجي الذكاء الاصطناعي (ChatGPT) و (Gemini) ضمن العينة المدروسة.
2- تختلف دقة أداء نموذجي الذكاء الاصطناعي باختلاف نوع الفعل من حيث البنية الصرفية (ثلاثي، رباعي، خماسي، سداسي).
3- تتباين قدرة النموذجين في التعامل مع صيغ الأفعال المختلفة (الماضي، المضارع، الأمر)، مع توقع انخفاض الدقة في صيغ الأمر للأفعال المعتلة.
4- يمكن توظيف نماذج الذكاء الاصطناعي بوصفها أدوات مساندة في تعليم الميزان الصرفي، شريطة إخضاع مخرجاتها للمراجعة اللغوية المتخصصة.

منهج البحث وإجراءاته:

اعتمد هذا البحث على المنهج الوصفي التحليلي القائم على المقارنة؛ إذ هدف إلى وصف أداء نموذجي الذكاء الاصطناعي (ChatGPT) و(Gemini) في تحديد الأوزان الصرفية للأفعال العربية وتقييمهما، من خلال مقارنة مخرجاتهما بالتحليل الصرفي المعياري المستند لرأي الخبراء والمصادر اللغوية المعتبرة. وتندرج هذه الدراسة ضمن الدراسات الاستكشافية، التي تهدف إلى رصد ملامح أداء النماذج اللغوية في سياق صرفي محدد، دون السعي إلى تعميم النتائج على الحالات جميعها، وهو ما يفسر الاكتفاء بعينة مركزة وممثلة.

إجراءات البحث: لضمان موضوعية المقارنة ودقة النتائج، مرّت الدراسة بالخطوات الإجرائية الآتية:

- **بناء عينة الدراسة:** تَمّ حصر (16) فعلاً عربياً اختيرت بأسلوب العينة الاستكشافية القصدية، لتغطي أربعة مستويات صرفية (الثلاثي، الرباعي، الخماسي، والسداسي). وروعي في الاختيار التنوع بين الأفعال الصحيحة، والمعتلة (أجوف، ناقص)، والمضعفة، وصيغ الأزمنة (ماضي، مضارع، أمر)؛ لضمان اختبار قدرة النماذج في مواجهة التحديات الصرفية المعقدة كالإعلال والإدغام والحذف.
- **لغة الأوامر:** اعتمدت الدراسة استراتيجية الأسئلة الموحدة والمباشرة؛ حيث وُجّه سؤال نمطي واحد لكل فعل بصيغة: "ما الوزن الصرفي لكلمة (الفعل)؟"، باللغة العربية حصراً، ودون تقديم أي سياقات

ISSN : 2410-1818

لغوية أو شروحات إضافية؛ لقياس القدرة الذاتية المجردة للنموذج على التحليل الصرفي، وللحد من أثر المتغيرات الخارجية على دقة الإجابة.

- **نموذجا الاختبار وزمنه:** نُفِذت الاختبارات خلال شهر أبريل (April 2026)، باستعمال النسخ المحدثة المتاحة وقت الدراسة: نموذج (GPT-5.3-mini) عبر الواجهة الرسمية لشركة (OpenAI)، ونموذج (Gemini 3 Pro) عبر واجهة (Google API)، مع تثبيت إعدادات الاستعمال الافتراضية؛ لضمان حيادية النتائج.
- **المعالجة والتحليل:** وظفت الدراسة أدوات التحليل الكمي من خلال رصد إجمالي الاستجابات الصحيحة والخاطئة وحساب النسب المئوية للدقة، متبوعة بـ تحليل نوعي لتفسير طبيعة الأخطاء الصادرة عن النماذج وتصنيفها، وصولاً إلى ربط هذه النتائج بالخصائص البنيوية للغة العربية وصياغة التوصيات النهائية.

وقد اقتضت طبيعة البحث أن يُقسَّم إلى أربعة مباحث، تسبقها مقدمة، وتتلوها خاتمة تتضمن أهم النتائج والتوصيات، وذلك على النحو الآتي:

المبحث الأول: الأفعال الثلاثية

المبحث الثاني: الأفعال الرباعية

المبحث الثالث: الأفعال الخماسية

المبحث الرابع: الأفعال السداسية

الدراسات السابقة:

1- بحث بعنوان: "استثمار الذكاء الاصطناعي في تعليم اللغة العربية للناطقين بغيرها: الواقع والتحديات"، أعدّته سعاد بسناسي، جامعة وهران 01 – الجزائر، ونُشر في مجلة "لغة – كلام"، المجلد 12، العدد 01، 2026م.

هدفت الدراسة إلى استعراض واقع استعمال الذكاء الاصطناعي في تعليم اللغة العربية للناطقين بغيرها، وتسليط الضوء على التحديات المرتبطة بتوظيف هذه التقنية، بما في ذلك صعوبات المعالجة اللغوية للنصوص العربية، وتحليل النطق، وتصحيح الأخطاء الإملائية والنحوية، إضافة إلى عرض التطبيقات والبرامج الداعمة لمهارات الكتابة والاستماع والتحدث. وأظهرت النتائج أن الذكاء الاصطناعي يسهم في تسهيل تعلم اللغة العربية وتحسين مهارات التواصل، إلا أنه ما يزال يواجه تحديات تتعلق بالدقة اللغوية وفهم السياق.

ISSN : 2410-1818

2- بحث بعنوان: "استخدام الذكاء الاصطناعي في تعليم اللغة العربية بين الإمكانيات والمشكلات"، أعدّه محمد أمين الحق، الجامعة الإسلامية العالمية شيتاغونغ – بنغلاديش، ونُشر في المجلة العربية للنشر العلمي، الإصدار الثامن، العدد (79)، (2025م).

هدف البحث إلى استعراض إمكانيات الذكاء الاصطناعي في تعليم اللغة العربية ومناقشة التحديات المصاحبة له. وقد بيّنت النتائج أن هذه التقنيات تسهم في تسهيل تعلم القواعد، وتشكيل النصوص، والترجمة، والتلخيص، وتسريع عملية التعلم. في المقابل، أشارت الدراسة إلى وجود تحديات مثل ضعف تنمية المهارات اللغوية الطبيعية لدى المتعلمين، وضعف التفاعل التربوي بين المعلم والطالب، وصعوبة تقديم تقييم دقيق للفروق الفردية.

3- بحث بعنوان: "تطبيقات الذكاء الاصطناعي التربوية وفاعليتها في تعليم اللغة العربية: مراجعة منهجية للأدبيات"، للباحثة مريم حمدان العنزي، مجلة العلوم التربوية، العدد(3)، الجزء (2)، (2022) م.

هدفت الدراسة إلى استكشاف فاعلية تقنيات الذكاء الاصطناعي في تعلم اللغة العربية وتنمية مهاراتها، مع تحديد أبرز التطبيقات التربوية والتحديات المرتبطة بإدماجها، مثل قضايا الخصوصية ونقص تدريب المعلمين. واعتمدت الدراسة على تحليل (50) دراسة سابقة، وأظهرت النتائج أن التطبيقات الذكية تسهم في تحسين أداء الطلبة وزيادة دافعية التعلم، مع التأكيد على ضرورة وجود معايير تقييم دقيقة وتعاون بين المؤسسات الأكاديمية والقطاع التقني.

يتضح من استعراض الدراسات السابقة أنها اهتمت بعرض إمكانات الذكاء الاصطناعي في تعليم اللغة العربية والتحديات المرتبطة بتوظيفه، أو قدّمت مراجعات نظرية عامة دون الدخول في اختبار تطبيقي دقيق لقدرات هذه الأدوات في الجانب الصرفي. أما هذا البحث فينحو إلى التركيز على الجانب التطبيقي المقارن، إذ يعمد إلى تقييم قدرة نموذجي الذكاء الاصطناعي (ChatGPT) و (Gemini) على تحديد الأوزان الصرفية للأفعال العربية، من خلال اختبارات عملية شملت الأفعال الثلاثية والرباعية والخماسية والسداسية، ومقارنة نتائجها بالتحليل الصرفي الصحيح. وبذلك يقدّم البحث بيانات موضوعية دقيقة حول مستوى أداء هذين النموذجين في الميزان الصرفي للأفعال، والكشف عن مدى دقتهما في معالجة الظواهر الصرفية مثل الإعلال والإدغام وصيغ الأمر، وهو ما يمنح البحث طابعًا تحليليًا تقييميًا يتجاوز الطرح النظري الذي ساد في الدراسات السابقة.

ISSN : 2410-1818

المبحث الأول: الأفعال الثلاثية

يُعَدّ تمييز الأوزان الثلاثية حجر الزاوية في التحليل الصرفي الآلي، نظراً لكونها تمثل الأصول المجردة التي تُبنى عليها معظم اشتقاقات اللغة. ويهدف هذا المطلب إلى قياس مدى دقة نموذجي (ChatGPT)و (Gemini) في التعامل مع البنية الصرفية الأساسية، ولا سيما في الحالات التي تتطلب تمييزاً دقيقاً بين الفعل الصحيح والمعتل (الأجوف) والمضعّف، كما يظهر في الجدول الآتي:

جدول (1): أسئلة وزن الأفعال الثلاثية

نموذج الذكاء الاصطناعي	ما الوزن الصرفي لكلمة (كَتَبَ)؟	ما الوزن الصرفي لكلمة (يَقُولُ)؟	ما الوزن الصرفي لكلمة (اسْعَ)؟	ما الوزن الصرفي لكلمة (مدّ)؟	عدد الأخطاء
ChatGPT	فَعَلَ	يَفُولُ	افْعَ	فعَّ	2
Gemini	فَعَلَ	يَفْعُلُ	افْعَ	فَعَلَ	0

يُظهر الجدول مقارنةً بين أداء نموذجي الذكاء الاصطناعي (ChatGPT) و (Gemini) في تحديد الأوزان الصرفية لأربعة أفعال ثلاثية، حيث تبيّن أن النموذجين قد أصابا في تحديد وزن الفعل (كَتَبَ) على (فَعَلَ)، وكذلك في تحديد وزن فعل الأمر (اسْعَ) على (افْعَ)، مع مراعاة أن هذه الصيغة مبنية على حذف حرف العلة.

غير أنّ موضع الاختلاف برز في معالجة الفعل المضارع (يقول)، إذ أخطأ (ChatGPT) عندما وزنه على (يَفُولُ)، وهو وزن غير صحيح صرفيًا؛ لأن الفعل (قال) فعل أجوف، وأصله الثلاثي على وزن (فَعَلَ). وقد أشار عبد القاهر الجرجاني إلى وزن (فال) باعتبار الصورة الصوتية للكلمة بعد الإعلال (الأستراباذي، 1975، ص18)، وهو توجيه يراعي البنية الملفوظة لا الحروف الأصلية. وعليه فإن الوزن الصرفي المعياري للفعل المضارع هو (يَفْعُلُ)، وهو ما أصاب فيه(Gemini) ، إذ إن حرف العلة في الفعل الأجوف لا يُحذف في هذه الصيغة، وإنما يبقى على صورة الواو.

وظهر تفاوت آخر في تحليل الفعل (مَدَّ)، حيث قدّم (ChatGPT) وزنًا غير منضبط (فعَّ)، وهو تمثيل لا يعكس البنية الصرفية الكاملة، في حين أصاب (Gemini) عندما أعاده إلى أصله الثلاثي (فَعَلَ) بوصفه فعلًا مضعّفًا أُدغمت عينه في لامه.

ويتضح أن نموذج (Gemini) تفوّق في هذه العيّنة من خلال خلوّ إجاباته من الأخطاء، في مقابل وقوع (ChatGPT) في خطأين صرفيين، مما يشير إلى ضعف في معالجة الفعل الأجوف والفعل المضعّف.

ISSN : 2410-1818

المبحث الثاني: الأفعال الرباعية

يُعَدّ تحليل الأفعال الرباعية والمزيدة من المستويات المتقدمة في علم الصرف العربي، نظرًا لما تتسم به من تنوّع في الأبنية وتعدّد في الصيغ المرتبطة بالزمن والدلالة. وتنبع أهمية هذا الجانب من كونه يختبر قدرة النماذج اللغوية على التعامل مع التحويلات الصرفية الأكثر تعقيدًا، مثل صيغ الأمر والمضارع والماضي في الأفعال المزيدة، فضلًا عن التمييز بين الأوزان القياسية المختلفة. ومن هذا المنطلق، تسعى هذه العيّنة إلى تقييم أداء نموذجَي الذكاء الاصطناعي (ChatGPT) و (Gemini) في تحديد الأوزان الصرفية لمجموعة مختارة من الأفعال الرباعية المجردة والمزيدة، بهدف الوقوف على مدى دقتهما في معالجة هذه الأبنية، والكشف عن أوجه الاتفاق والاختلاف في تحليلهما الصرفي.

جدول (2): أسئلة وزن الأفعال الرباعية

نموذج الذكاء الاصطناعي	ما الوزن الصرفي لكلمة (دحرج)؟	ما الوزن الصرفي لكلمة (يوسوس)؟	ما الوزن الصرفي لكلمة (أهدِ)؟	ما الوزن الصرفي لكلمة (اطْمأنَّ)؟	عدد الأخطاء
ChatGPT	فَعْلَلَ	يُفَعْلِلُ	أَفْعِلْ	افْعَلَّ	2
Gemini	فَعْلَلَ	يُفَعْلِلُ	أَفْعِ	افْعَلَلَّ	0

يُظهر الجدول مستوى الأداء لدى نموذجي الذكاء الاصطناعي (ChatGPT) و (Gemini) في تحليل الأوزان الصرفية للأفعال الرباعية المجردة والمزيدة، حيث اتفق النموذجان على التحليل الصحيح للفعلين (دحرج) على وزن (فَعْلَلَ)، و(يوسوس) على وزن (يُفَعْلِلُ)، مما يدل على قدرة جيدة في التعامل مع الأبنية الرباعية المجردة.

إلا أن موضع الاختلاف ظهر في تحليل الفعل (أهدِ)، حيث أخطأ (ChatGPT) عندما وزنه على (أفْعِلْ)، وهو وزن يخصّ الأمر من (أهدى) المعتل، في حين أصاب (Gemini) في تحديده على وزن (أَفْعِ)، وهو الوزن الصحيح؛ لأن فعل الأمر معتل الآخر يُبنى على حذف حرف العلة (حسن، 2010، ص81). وظهر موضع اختلاف آخر في تحليل الفعل (اطمأنّ)، حيث أخطأ (ChatGPT) في تحديد وزنه على (افعلّ)، وهو وزن لا يستوعب البنية الصرفية الكاملة للفعل، في حين أصاب (Gemini) في وزنه (افعللّ)، وهو الصواب؛ لأن الفعل (اطمأنّ) فعل رباعي مزيد، وتقتضي بنيته الصرفية هذا الوزن الذي يحفظ جميع أصوله وحروفه (العكبري، 1995، 218/2).

وبناءً على ذلك، يُسجَّل تفوّق (Gemini) في هذه العينة بخلوّ إجاباته من الأخطاء، مقابل وقوع (ChatGPT) في خطأين، مما يشير إلى أن (ChatGPT) يواجه صعوبة في وزن الفعل الناقص والفعل المزيد. وهو ما جرى في وزن الأفعال الثلاثية.

ISSN : 2410-1818

المبحث الثالث: الأفعال الخماسية

يتطلّب تحليل الأفعال المزيدة دقّة في تحديد الأوزان، نظرًا لتعدّد صيغها وتداخل بنيتها، وما يطرأ عليها من ظواهر صرفية كالحذف والإعلال والإدغام. وتبرز أهمية هذا التحليل في تقييم مدى قدرة نماذج الذكاء الاصطناعي على فهم البنية الصرفية العربية. ويهدف هذا الجدول إلى قياس دقّة نموذجي (ChatGPT)و (Gemini) في تحديد أوزان هذه الأفعال، والكشف عن مدى كفايتهما في التعامل مع هذه الظواهر.

جدول (3): أسئلة وزن الأفعال الخماسية

نموذج الذكاء الاصطناعي	ما الوزن الصرفي لكلمة (انطلق)؟	ما الوزن الصرفي لكلمة (يجتمع)؟	ما الوزن الصرفي لكلمة (ارْتقِ)؟	ما الوزن الصرفي لكلمة (اشْتَدّ)؟	عدد الأخطاء
ChatGPT	انْفَعَلَ	يَفْتَعِلُ	افْتَعِلْ	افْتَعَلّ	2
Gemini	انْفَعَلَ	يَفْتَعِلُ	افْتَعِ	افْتَعَلَّ	1

يُظهر الجدول تقاربًا في أداء نموذجي الذكاء الاصطناعي (ChatGPT) و (Gemini) في تحليل الأوزان الصرفية للأفعال الخماسية، إذ اتفق النموذجان على الوزن الصحيح للفعلين (انطلق) على وزن (انفعل)، و(يجتمع) على وزن (يفتعل)، مما يدلّ على قدرة جيّدة في التعامل مع الأفعال الخماسية في صيغتي الماضي والمضارع.

غير أنّ موضع التباين برز في تحليل الفعل (ارْتقِ)، حيث قدّم (ChatGPT) وزنه على (افْتَعِلْ)، في حين قدّم (Gemini) وزنه على (افْتَعِ)، وهو الصواب؛ لأن الفعل أمر من (ارتقى) على وزن (افْتَعَلَ)، وقد حُذفت لامه لاعتلالها، فصار على هذه الصورة بعد الإعلال (الأزهري، د.ت، ص20).

وأخطأ النموذجان في تحليل الفعل (اشْتَدَّ)، إذ لم يُراعيا أثر الإدغام، والصواب أن وزنه (افْتَعَلَ)، مع إدغام العين في اللام (د + د)، فصارت صورته الصوتية (اشْتَدَّ).

وبناءً على ذلك، يتبيّن أنّ كلا النموذجين لم يحقّقا تطابقًا كاملًا في هذه العيّنة، إذ وقع كلٌّ منهما في أخطاء تتعلّق بصيغ الأمر وظاهرة الإدغام، مما يشير إلى حاجة هذه النماذج إلى مزيد من الدقّة في معالجة التحوّلات الصرفية الدقيقة، ولا سيما في الأفعال المزيدة والمعتلّة. ومع ذلك، فقد اقتصر خطأ (Gemini)على موضع واحد، في حين وقع (ChatGPT) في موضعين، الأمر الذي يدلّ على تفوّق (Gemini)نسبيًا في هذا الاختبار.

ISSN : 2410-1818

المبحث الرابع: الأفعال السداسية

يُعدّ باب الأفعال السداسية من أهم أبواب الصرف العربي وأكثرها تعقيدًا، إذ يتكوّن غالبًا من ستة أحرف على وزن (استفعل)، وما يطرأ عليه من تغييرات صرفية عند تصريفه في الأزمنة المختلفة، مثل: الماضي والمضارع وصيغة الأمر. وتزداد أهمية هذا الباب في كونه يكشف قدرة المتعلم أو النماذج اللغوية على التعامل مع الظواهر الصرفية الدقيقة، كالإعلال والحذف والتحويل بين الصيغ. ومن هذا المنطلق، يهدف هذا المطلب إلى تحليل أداء نموذجي الذكاء الاصطناعي (ChatGPT) و (Gemini) في تحديد الأوزان الصرفية لمجموعة من الأفعال السداسية، وبيان مدى دقتهما في التعامل مع هذا النوع من الأبنية.

جدول (4): أسئلة وزن الأفعال السداسية

نموذج الذكاء الاصطناعي	ما الوزن الصرفي لكلمة (استخرج)؟	ما الوزن الصرفي لكلمة (يستعمل)؟	ما الوزن الصرفي لكلمة (استعن)؟	ما الوزن الصرفي لكلمة (استعدّ)؟	عدد الأخطاء
ChatGPT	اسْتَفْعَلَ	يَسْتَفْعِلُ	استفعِل	اسْتَفْعَلَ	1
Gemini	اسْتَفْعَلَ	يَسْتَفْعِلُ	اسْتَفِلْ	اسْتَفْعَلَ	0

يُبرز هذا الجدول مستوى أداء نموذجي الذكاء الاصطناعي(ChatGPT) و ((Gemini) في تحليل الأوزان الصرفية للأفعال السداسية على وزن (استفعل)، حيث اتفق النموذجان على الوزن الصحيح للفعلين (استخرج) و(استعدّ) على وزن (اسْتَفْعَلَ)، وكذلك الفعل المضارع (يستعمل) على وزن (يَسْتَفْعِلُ)، مما يدل على قدرة جيدة في التعامل مع الأبنية الصرفية القياسية في صيغتي الماضي والمضارع.

إلا أن موضع الاختلاف ظهر في تحليل فعل الأمر (استعِنْ) ، إذ وزنه(ChatGPT) على (استفعِل)، في حين قدّم (Gemini) الوزن الصحيح (اسْتَفِلْ)، وفق القاعدة الصرفية التي تقضي بحذف حرف العلة في عند صياغة الأمر من الفعل المعتل. ويعكس هذا الموضع دقة في معالجة صيغ الأمر في الأفعال السداسية المعتلة لدى(Gemini) ، مقابل وجود خطأ لدى (ChatGPT) في هذا الجانب.

وبناءً على ذلك، تشير النتائج إلى تفوّق (Gemini) في هذه العينة من حيث خلوّ إجاباته من الأخطاء، مقابل وقوع (ChatGPT) في خطأ واحد، مما قد يعكس حاجة الأخير، في حدود هذه العينة، إلى مزيد من الدقة في التعامل مع التحولات الصرفية المرتبطة بصيغة الأمر في الأفعال السداسية.

ويمكن تفسير الأخطاء التي ظهرت في أداء نماذج الذكاء الاصطناعي (ChatGPT) و (Gemini)في هذا البحث في ضوء ما أشار إليه(Habash & Rambow, 2005, p. 547) ، إذ إن تعدد التحليلات الصرفية المحتملة للكلمة الواحدة يُعدّ تحديًا مباشرًا لأنظمة المعالجة الآلية، وهو ما

ISSN : 2410-1818

يفسر التباين في تحديد بعض الأوزان الصرفية، ولا سيما في الحالات التي تتضمن الإعلال أو الإدغام أو صيغ الأمر في الأفعال المعتلة.

ويُعزى هذا التفاوت في الأداء أحيانًا إلى طبيعة التفاعل مع هذه النماذج؛ إذ يتأثر أداء النماذج اللغوية في معالجة اللغة العربية بلغة الأوامر (Prompts) المستخدمة، حيث وُجد أن الأوامر المكتوبة باللغة الإنجليزية تؤدي إلى نتائج أفضل في المهام العربية مقارنة بالأوامر المكتوبة بالعربية نفسها (Khondaker et al., 2023, p. 4).

ورغم أن الدراسة الحالية استخدمت أوامر عربية موحدة لكلا النموذجين، إلا أن تفوق (Gemini)قد يشير إلى قدرة أعلى على معالجة الأوامر العربية المباشرة وتجاوز فجوة "لغة المطالبة" التي قد يعاني منها.(ChatGPT)

ويُضاف إلى ذلك أن هذا التعقيد يرتبط بطبيعة النظام الصرفي في العربية بوصفه نظامًا صرفيًا غير خطي، حيث لا تُشتق الكلمات بطريقة خطية قائمة على إضافة اللواصق فحسب، بل من خلال تفاعل الجذر مع الأوزان الصرفية ضمن بنية داخلية معقدة، مما يزيد من صعوبة التحليل الآلي وتحديد البنية الصرفية الدقيقة للكلمة(Naudé, 2003, p. 16) ، بالإضافة إلى أن اللغة العربية تُعدّ من أكثر اللغات تعقيدًا صرفيًا على مستوى العالم، ويعود ذلك إلى نظام الجذر والوزن الذي يحكم بنيتها الصرفية.(Alayba, 2025, p. 4)

إن الأدوات والتقنيات الحالية في معالجة اللغات الطبيعية (NLP) التي طُوِّرت للغات أخرى قد لا تنتقل صلاحيتها مباشرةً إلى اللغة العربية، الأمر الذي يستدعي تطوير مقاربات مصممة خصيصًا للتعامل مع خصائصها الصرفية .(Alasmari, 2025, p. 2) ويظهر أن التفوق الملحوظ لنموذج (Gemini)يعود إلى قدرته على إدراك التفسير الوظيفي للبنية الصرفية، وتجاوز حدود الشكل السطحي (Surface Form) الذي قد يضلل بعض النماذج الأخرى (Adel, Alhafni, & Habash, 2026, pp. 7–8).

إن الأدوات والتقنيات الحالية في معالجة اللغات الطبيعية (NLP) التي طُوِّرت للغات أخرى، قد لا تنتقل صلاحيتها مباشرةً إلى اللغة العربية، الأمر الذي يستدعي تطوير مقاربات مصممة خصيصًا للتعامل مع خصائصها الصرفية (هلال وحلاوة، 2025، ص278-279).

ويظهر أن التفوق الملحوظ لنموذج (Gemini) يعود إلى قدرته على إدراك التفسير الوظيفي للبنية الصرفية، وتجاوز حدود الشكل السطحي (Surface Form) الذي قد يضلل بعض النماذج الأخرى

وفي هذا الإطار، قد يُفضي الاستخدام الواسع لتقنيات الذكاء الاصطناعي في المجال التعليمي إلى إغفال الأبعاد الاجتماعية التفاعلية والجوانب الأخلاقية التي تُشكّل عنصرًا أساسيًا في العملية التعليمية. إذ إن التركيز المفرط على الأتمتة والمعالجة الكمية قد يؤدي إلى تقليص دور التفاعل الإنساني

ISSN : 2410-1818

والعلاقات المباشرة بين أطراف العملية التعليمية، مما ينعكس سلبًا على تنمية مهارات التواصل الاجتماعي وتعزيز التفكير النقدي، وقد يسهم في زيادة مظاهر العزلة الاجتماعية. كما تبرز في هذا السياق إشكاليات أخلاقية متعددة، من أبرزها قضايا الخصوصية وتكافؤ فرص الوصول إلى التعليم في ظل التوسع في استخدام هذه التقنيات. وإلى جانب ذلك، فإن الاعتماد المتزايد على الذكاء الاصطناعي يؤدي إلى إعادة تشكيل أدوار المعلمين، حيث تبقى القدرات الإنسانية مثل الدعم العاطفي والتوجيه الفردي من المهام التي تعجز النماذج الآلية عن تقديمها بالفاعلية ذاتها التي يوفرها المعلم (العنزي، 2024، ص477).

ولا يمكن للذكاء الاصطناعي أن يؤدي وظيفته في مجال التعليم دون توافر بنية تحتية ملائمة تدعمه، تشمل توفر إنترنت عالي السرعة وواسع التغطية وبتكلفة معقولة. وعلى الرغم من أن هذه المتطلبات متوفرة في عدد من الدول ذات الإمكانات الاقتصادية الكبيرة، فإن العديد من الدول النامية ما تزال بعيدة عن تحقيق هذه الشروط. كما يعتمد نجاح وفعالية استخدام الذكاء الاصطناعي في التعليم على توافر المعدات الرقمية، وتدريب الكوادر الفنية المتخصصة، إضافة إلى ضرورة تأمين وحماية البيانات الضخمة التي يتم التعامل معها (شاهين، 2023، ص147).

ومع ذلك، ينبغي الحذر من ظاهرة الهلوسة (Hallucination) -وهي توليد النموذج لمخرجات أو أوزان غير صحيحة أو غير موجودة في الاستعمال اللغوي أو القاعدة الصرفية- والتي قد تظهر في بعض حالات التوليد غير المنضبط (Adel, Alhafni, & Habash, 2026, p. 8).

ISSN : 2410-1818

الخاتمة

سعى هذا البحث إلى تقييم أداء نموذجي الذكاء الاصطناعي (ChatGPT) و (Gemini) في تحليل الأوزان الصرفية للأفعال العربية، من خلال تطبيق مجموعة من الأسئلة التي شملت الأفعال الثلاثية والرباعية والخماسية والسداسية، بما تضمنته من ظواهر صرفية متنوعة مثل الإعلال والإدغام وصيغ الأمر والمضارع. وقد اعتمد البحث على المقارنة التحليلية بين النموذجين بهدف الكشف عن مدى دقتهما في معالجة البنية الصرفية العربية، وتحديد مواطن القوة والقصور في كل منهما.

وقد أظهرت المقارنة بين إجابات النموذجين تفاوتًا في مستوى الدقة، وفيما يأتي بيان ذلك:

جدول (5): مقارنة إجمالي الأخطاء في تحليل الأوزان الصرفية

الموضوع	عدد الأسئلة	عدد أخطاء (ChatGPT)	موطن الخطأ	نسبة الدقة	نسبة الخطأ	عدد أخطاء (Gemini)	موطن الخطأ	نسبة الدقة	نسبة الخطأ
الأفعال الثلاثية	4	2	يقول (يفول) مدّ (فعّ)	٪50	٪50	-	-	٪100	٪0
الأفعال الرباعية	4	2	أهدِ (أفعل) اطمأن (افعلّ)	٪50	٪50	-	-	٪100	٪0
الأفعال الخماسية	4	2	ارتقِ (افتعل) (اشتدّ افتعلّ)	٪50	٪50	1	اشتدّ(افتعلّ)	93.75%	٪ 6.25
الأفعال السداسية	4	1	استعِن (استفعل)	٪75	%25	-	-	٪100	٪0
المجموع	16	7	-	٪56.25	٪43.75	1	-	%93.75	**6.25%**

وبناءً على التحليل الكمي للأخطاء، فقد بلغت نسبة الخطأ في نموذج (ChatGPT) (43.75%)مقابل (6.25٪) في نموذج(Gemini) ، في حين سجلت الدقة (56.25٪) و(93.75٪) على التوالي، مما يشير إلى تفوق واضح لصالح نموذج (Gemini) في أداء مهمة تحديد الأوزان الصرفية للأفعال العربية. ويُلاحظ أن أخطاء (ChatGPT) توزعت على معظم أنماط الأفعال محلّ الاختبار، لا

ISSN : 2410-1818

سيما الأفعال الثلاثية والرباعية والخماسية، الأمر الذي يعكس تذبذبًا في قدرته على ضبط البنية الصرفية مقارنة بـ (Gemini)الذي اقتصر خطؤه على موضع محدود واحد فقط، ما يدل على درجة أعلى من الاستقرار والدقة في المعالجة الصرفية. وبالتالي، فإن هذه النتائج تُبرز تفاوتًا ملحوظًا بين النموذجين في التعامل مع الظواهر الصرفية العربية، لصالح (Gemini) في هذه العينة محل الدراسة.

النتائج

بناءً على النتائج الواردة في المقارنة بين نموذجي الذكاء الاصطناعي (ChatGPT) و(Gemini)، توصل البحث إلى النتائج الآتية:

- أظهر نموذج (Gemini) مستوى أعلى من الدقة العامة مقارنةً بنموذج (ChatGPT) في تحديد الأوزان الصرفية ضمن نطاق العينة، مما يعكس تفاوتًا في كفاية النموذجين في معالجة البنية الصرفية العربية .
- سجّل نموذج (ChatGPT) عددًا أكبر من الأخطاء، تمركزت بصورة رئيسة في الأفعال الثلاثية والرباعية والخماسية، مع انخفاض نسبي في الأخطاء في الأفعال السداسية .
- أظهر النموذجان قدرة جيدة على التعامل مع الأوزان الصرفية القياسية، ولا سيما في الأفعال الصحيحة، وهو ما يشير إلى استجابة أفضل للأنماط المنتظمة في البنية الصرفية .
- برزت صعوبات لدى النموذجين في معالجة الظواهر الصرفية الدقيقة، مثل: الإعلال والإدغام، إضافة إلى صيغ الأمر في الأفعال المعتلة، مع ملاحظة تفوق نسبي لنموذج (Gemini) في هذا الجانب .
- يحتاج نموذج (ChatGPT) إلى مزيد من التحسين في معالجة بعض الأنماط الصرفية غير المنتظمة، وهو ما يفسر التباين في مستوى أدائه مقارنةً بالنموذج الآخر ضمن العينة المدروسة .
- لم يحقق نموذج (ChatGPT) درجة كافية من الدقة تسمح بالاعتماد عليه بصورة مستقلة في معالجة الميزان الصرفي، مقابل أداء أفضل لنموذج(Gemini) ، الأمر الذي يدعم إمكانية توظيفه تعليميًا ضمن إشراف متخصص.

ISSN : 2410-1818

التوصيات:

- التأنّي في توظيف نماذج الذكاء الاصطناعي في تعليم الصرف العربي، وحصر استعمالها في أنشطة تطبيقية داعمة للفهم والتحليل .
- عدم الاعتماد الكلي على مخرجات هذه النماذج دون إخضاعها لمراجعة لغوية علمية متخصصة .
- تدريب الطلبة على تحليل الأخطاء الصرفية الصادرة عن نماذج الذكاء الاصطناعي، وتنمية مهارات النقد اللغوي لديهم .
- تعزيز دور المعلم في تقييم مخرجات الذكاء الاصطناعي وتوجيه استعمالها توجيهًا علميًا سليمًا .
- إجراء دراسات مستقبلية مقارنة تشمل نماذج أحدث، مع توسيع نطاق التحليل اللغوي ليشمل مستويات صرفية أعمق .
- تطوير تطبيقات ذكاء اصطناعي متخصصة في اللغة العربية تراعي الدقة الصرفية، ولا سيما قضايا الإعلال والإدغام .
- بناء اختبارات معيارية في الميزان الصرفي تُعتمد مرجعًا لقياس أداء نماذج الذكاء الاصطناعي وتقويمها .
- إجراء دراسات لاحقة تعتمد على عينات أوسع وأكثر تنوعًا؛ للتحقق من النتائج وتعزيز إمكانية تعميمها.

ISSN : 2410-1818

قائمة المصادر والمراجع

المراجع العربية:

- الأزهري، خالد. (د.ت). "شرح الأزهرية". المطبعة الكبرى ببولاق، القاهرة.
- الأستراباذي، رضي الدين. (1975). "شرح شافية ابن الحاجب (حاشيته)". تحقيق: محمد محيي الدين عبد الحميد وآخرون. دار الكتب العلمية، بيروت.
- حسن، عباس. (2010). "النحو الوافي". دار المعارف، القاهرة.
- شاهين، هالة عبد المؤمن محمد. (2023). "الذكاء الاصطناعي وتحويل التعليم من التلقين إلى تطبيق أدوات تضمن استدامة التعليم". المجلة العربية للتربية النوعية، المنظمة العربية للتربية والعلوم والآداب، مج (7)، ع (26)، 219-239.
- العكبري، أبو البقاء. (1995). "اللباب في علل البناء والإعراب". تحقيق: عبد الإله النبهان. دار الفكر، دمشق.
- هلال، منتصر عثمان، وحلاوة، حمادة فوزي. (2025). "أثر استخدام تقنيات معالجة اللغة الطبيعية (NLP)المعتمدة على الذكاء الاصطناعي لتحسين تجربة المتعلم والاتجاه نحوها". مجلة الذكاء الاصطناعي وأمن المعلومات، مج (3)، ع (7)، 278-279.

المراجع الأجنبية

- Alasmari, A. (2025). "A scoping review of Arabic natural language processing for mental health". Healthcare. MDPI (Switzerland), 13(9), 963.

https://doi.org/10.3390/healthcare13090963

- Alayba, A. M. (2025). "Arabic natural language processing (NLP): A comprehensive review of challenges, techniques, and emerging trends". Computers. MDPI (Switzerland), 14(11), 497.

https://doi.org/10.3390/computers14110497

- Habash, N., & Rambow, O. (2005). "Arabic tokenization, part-of-speech tagging and morphological disambiguation in one fell swoop". Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL). Association for Computational Linguistics, 547.

https://doi.org/10.3115/1219840.1219911

ISSN : 2410-1818

- Khondaker, M. T. I., Waheed, A., Nagoudi, E. B., & Abdul-Mageed, M. (2023). "GPTAraEval: A comprehensive evaluation of ChatGPT on Arabic NLP". Proceedings of EMNLP 2023. Association for Computational Linguistics, 220–247. https://doi.org/10.18653/v1/2023.emnlp-main.16
- Naudé, J. A. (2003). "The consonantal root in Semitic languages". Journal of Northwest Semitic Languages, 29(2), 16.

ISSN : 2410-1818

مجلـــة الأندلس للعلوم الإنسانيـة والاجتماعيـة
مجلـــة دوليـــة شهـريـــة علميـــة محكمـــة
الترقيـم الدولي الإلكتروني: ISSN:2410- 521X
الترقيـم الدولـي الورقـي :ISSN:2410- 1818
البريد الإلكتروني : journal@andalusuniv.net

المجلـة مفهرسـة فـي المواقـع الآتيـة :

العام	2021	2022	2023	2024	2025
معامل أرسيف	0.2692	0.1954	0.3759	0.3068	0.5978
معامل التأثير العربي	1.60	1.73	1.25	1.55	1.59