



اللغويات في الترجمة الآلية والمعالجة اللغوية

Journal of Computational Linguistics & Arabic Language Processing



دراسة مسحية لآليات تقييم النماذج اللغوية الكبيرة

أنظمة الترجمة الآلية من العربية إلى الإنجليزية وتقييم بعض
برامجها

دور النماذج اللغوية الكبيرة في تطوير (WordNet) قاعدة بيانات
معجمية عربية: تقييم القدرات وتحسين الشبكات الدلالية

استعراض كتاب: لسانيات المدونات مقدمة موجزة

السنة الثالثة - العدد الخامس
ذو القعدة ١٤٤٧ هـ - أبريل ٢٠٢٦ م
Vol. 3, No. 5, Apr, 2026

مجمع الملك سلمان
العالمي للغة العربية
King Salman Global Academy for Arabic Language



اللغويات والحوسبة والمعالجة اللغوية

Journal of Computational Linguistics & Arabic Language Processing

www.ksaa.gov.sa

مجلة علمية مُحكمة تصدر عن مجمع الملك سلمان العالمي للغة العربية





المشرف العام على المجلة

أ.د. عبدالله بن صالح الوشمي

إدارة التحرير

أ.د. منصور الغامدي

رئيس التحرير

د. عبد الرحمن العصيمي

مدير التحرير

د. هاجر القبيشي

أمانة المجلة

هيئة التحرير

د. سارة الحمود

د. عريب العويشق

أ.د. عبدالعزيز المهيوبي

د. أيمن الغامدي

الهيئة الاستشارية

أ.د. محسن عبدالرزاق رشوان (مصر)

أ.د. إيريك أتويل (بريطانيا)

أ.د. نزار حبش (أمريكا)

أ.د. مصطفى عرسان جرار (فلسطين)

أ.د. عبدالمحسن عبيد الثبتي (السعودية)

أ.د. خالد فؤاد شعلان (مصر)

أ.د. صلاح راشد الناجم (الكويت)

◆ قواعد النشر بالمجلة :

سياسات النشر:

تنشر مجلة اللُّغويات الحاسوبية والمعالجة الآلية للُّغة العربيّة البحوث الأصيلة التي لم يسبق نشرها، والمكتوبة بالُّغة العربية أو الإنجليزية، والدراسات والمُلخصات العلميّة المتخصصة في مجال حوسبة اللُّغة، كاللُّغويات الحاسوبية، والمعالجة الآليّة للُّغة العربيّة، وكل ما له صلة بحوسبة اللُّغة.

تُعرض مسودة الأوراق العلميّة فور وصولها على هيئة التحرير لإجراء تقييم مبدئي لها، وفي حال عدم مناسبتها للنشر تُخطر المجلة الباحث بذلك. أمّا البحوث المقبولة مبدئيًا، فتُرسل إلى اثنين من المحكّمين لتقييمها سرّيًا، وتبذل هيئة التحرير طاقتها في التواصل مع المحكّمين للوصول إلى قرار بشأن كل بحث مقدّم في أسرع وقت ممكن، ويُخطر الباحث بنتيجة التحكيم. وفي حالة قبول الورقة للنشر، فإنّ كافة حقوق النشر تؤوّل للمجلة، ولا يجوز نشرها في أي منفذ نشر آخر ورقياً أو إلكترونياً، دون إذن كتابي من رئيس هيئة تحرير المجلة. وتُرسل نسختان ورقيتان مجانيّتان من المجلة إلى المؤلّف عند نشر ورقته.



Journal information:

◆ بيانات المجلة :

Journal of Computational Linguistics & Arabic Language Processing	مجلة اللُّغويات الحاسوبية والمعالجة الآلية للُّغة العربيّة
A peer-reviewed bi-annual journal issued by	مجلة علمية محكمة تصدر عن
King Salman Global Academy for Arabic Language	مجمع الملك سلمان العالمي للُّغة العربيّة
Riyadh- Saudi Arabia	الرياض - المملكة العربيّة السعوديّة
ISSN paper version: 1658-9475	ردمك النسخة الورقية: ١٦٥٨-٩٤٧٥
ISSN Electronic version: 1658-9483	ردمك النسخة الإلكترونيّة: ١٦٥٨-٩٤٨٣

تعليمات عامة:

١. تُرسل الورقة المقدّمة للنشر للمجلة بصيغة وورد (word) على بريد المجلة: jclalp@ksaa.gov.sa. على أن تُوجّه جميع المراسلات إلى: رئيس هيئة التحرير- مجلة اللُّغويات الحاسوبية والمعالجة الآلية للغة العربيّة- مجمع الملك سلمان العالمي للغة العربيّة.

٢. يجب أن تحتوي الورقة على ما يلي:

١-٢- صفحة أولى فيها عنوان الورقة، واسم المؤلف أو المؤلفين، وجهة عملهم، وعناوينهم، وتاريخ تقديم الطلب بنشر الورقة.

٢-٢- الكلمات المفتاحية.

٣-٢- مستخلص الورقة باللّغتين العربية والإنجليزية -في حدود مئة كلمة لكلّ منهما-.

٤-٢- المصادر والمراجع.

٥-٢- الملاحق -إن وجدت-.

٣. يُراعى أن تكون الورقة العلمية مكتوبةً بلغة سليمة -سواء كانت عربية أو إنجليزية-، وأن يتراوح عدد كلماتها بين (٣٠٠٠ - ١٠٠٠٠) كلمة، بما في ذلك المراجع والهوامش والجداول والأشكال والملاحق.

٤. إذا كانت الورقة العلمية مكتوبة باللّغة العربيّة، فيجب الالتزام بكتابة ما يقابل المصطلحات العربيّة باللّغة الإنجليزيّة في المرة الأولى التي يظهر فيها المصطلح.

٥. ألا يسبق نشر البحث في أي وعاء علمي -مالم ترهئة التحرير خلاف ذلك-.

المواصفات الفنية للبحث:

- يكون مقاس الصفحة A4، وتُقسَّم إلى عمودين، وتكون المسافة بين العمودين ١ سم، وتكون الهوامش ٥,٢ سم عن يمين ويسار الصفحة، و٥٨,٣ سم من أعلى وأسفل الصفحة.
- تُترك مسافة 5.0 سم في بداية كل فقرة من فقرات المتن.
- يجب الفصل بمسافة سطر واحد بين نهاية محتوى الفقرات داخل المتن وعناوين الفقرات التي تليه.
- يجب ترك مسافة سطر واحد قبل وبعد الجدول أو الشكل والصورة والمتن السابق والتالي لهما.
- يُستخدم الخط (Sakkal Majalla) للنص العربي، والخط (Times New Roman) للنص الأجنبي وبحجم أقل بدرجتين عن الخط العربي. ويكون حجم الخط العربي كالآتي:
 - العنوان الرئيس: ١٤، عريض.
 - العناوين الفرعية: ١٢، عريض.
 - النصوص: ١٢، عادي.
 - الهوامش السفلية: ١٠، عادي.
- يجب ترقيم صفحات البحث.
- يجب ترقيم المعادلات.
- تُرتَّب المصادر والمراجع أبجدياً بنظام APA.
- تُفصل المراجع الأجنبية عن المراجع العربية في قائمة المصادر والمراجع، وتأتي المراجع /المصادر العربية أولاً في الأبحاث العربية والمراجع /المصادر الإنجليزية أولاً في الأبحاث الإنجليزية.

الجداول:

- تُدرج الجداول بعد الفقرة التي ذُكرت بها، وذلك على النحو التالي:
 - إذا كان عرض الجدول ٧,٥ سم يوضع في عمود واحد أسفل الفقرة مباشرة.
 - إذا كان عرض الجدول أكثر من ٧,٥ سم يوضع أسفل الصفحة التي ذكر بها أو أعلى الصفحة التي تليها.
- تُرقّم الجداول بين قوسين في أقصى يمين الصفحة مسبوقة بكلمة «جدول». مثال: جدول (١).
- يُوضع عنوان الجدول في أعلى الجدول مباشرة، في أقصى اليمين.
- يجب ألا يتجاوز عرض الجدول هوامش الصفحة المذكورة في الفقرة السابقة.
- يُكتب مصدر الجدول والتعليقات المصاحبة له -إن وجدت- أسفله.
- تُضبط بيانات الجدول غير الرقمية أقصى يمين الخلية.
- تُضبط بيانات الجدول الرقمية وسط الخلية.
- إذا استمر الجدول على أكثر من صفحة، فيجب تكرار رأسه في الصفحات التالية ووضع عبارة «تابع جدول (...)» في أعلاه.

الصُور والأشكال:

- تُدرج الصور والأشكال بعد الفقرة التي ذُكرت على النحو التالي:
 - إذا كان عرض الصورة ٧,٥ سم تُوضع في عمود واحد أسفل الفقرة مباشرة.
 - إذا كان عرض الصورة أكثر من ٧,٥ سم تُوضع أسفل الصفحة التي ذكر بها أو أعلى الصفحة التي تليها.
- تُرقّم الصور والأشكال، ويوضع الرقم بين قوسين مسبوقة بكلمة «شكل». مثال: شكل (١).

- يُدرج عنوان الشكل أسفل الشكل مباشرةً متوسطًا بين حدوده بحيث لا يتجاوز عنوان الشكل سطرين فقط.
- في حال وجود تعليق مصاحب على الشكل أو الصورة يجب دمجها مع الشكل وحفظه على هيئة صورة لا تقل دقتها عن (300 dpi).

مواصفات كتابة المعادلات:

- تُكتب المعادلات الرياضية باستخدام أحد البرنامجين التاليين: (Math Type) أو (Equation Editor).
- تُرقّم المعادلات، ويوضع الرقم بين قوسين في أقصى يسار العمود مقابل المعادلة على مستوى البحث.

مواصفات كتابة الاقتباس:

- إذا كان النص المقتبس ٣ أسطر أو أقل، فيجب أن يوضع بين علامتي تنصيص «...»، ويليه توثيق النص على النحو التالي: (العائلة، السنة: الصفحة).
- إذا كان النص أكثر من ٣ أسطر، فيجب أن يُميز النص المقتبس بكتابته في منتصف الصفحة، وتُترك من بداية السطر ونهايته هوامش اسم، ويوضع بين علامتي تنصيص «...»، ويُكتب بحجم ١٢ للنص العربي و١٠ للنص الإنجليزي، وينتهي بتوثيق النص على النحو التالي: (العائلة، السنة: الصفحة).

مواصفات كتابة الحواشي:

- «تُدرج الحواشي أسفل كل صفحة، مع ترقيمها آليًا حسب ورودها في كل صفحة».

◆ صفة المجلة:

مجلة اللُّغويات الحاسوبية والمعالجة الآلية للُّغة العربيَّة (Journal of Computational Linguistics & Arabic Language Processing) هي مجلة علمية محكَّمة، نصف سنوية (ISSN:1658-9475 - ISSN:1658-9483)، وتصدر عن مجمع الملك سلمان العالمي للُّغة العربية في الرياض بنسختين: إلكترونية، وورقية (ذات إصدار محدود). وتتيح المجلة الوصول الحر (Open Access) إلى البحوث المنشورة فيها، ولا تتقاضى رسوْمًا على النشر. وهي مجلة تهتم بنشر الأبحاث العلمية المحكمة في مجال حوسبة اللُّغة، وتشمل: اللُّغويات الحاسوبية (Computational Linguistics)، والمعالجة الآلية للُّغة العربية (Arabic Language Processing) وما يتبعها من تخصصات متَّصلة بها. وللاطلاع على أهداف المجلة وشروط النشر وكل التفاصيل ذات العلاقة يمكنكم زيارة موقع المجلة: jclalp.ksaa.gov.sa

البحوث المنشورة بالمجلة تعبر عن رأي أصحابها، ولا تعبر بالضرورة عن رأي المجمع، ويتحمل مؤلفوها المسؤولية كاملة عن صحة المعلومات والاستنتاجات ودقتها.

محتويات العدد

١٣

دراسة مسحية لآليات تقييم النماذج اللغوية الكبيرة

أ.د. جورج تركيّه - ريم المطيريب - تركي العتيبي - فيصل بن بريكان
رغد آل رشيد - أمال المزروعج - منيرة الحوشان - روان المعثم

٦٩

أنظمة الترجمة الآلية من العربية إلى الإنجليزية وتقييم بعض برامجها

د. مدحت يوسف السبع

١٣٤

دور النماذج اللغوية الكبيرة في تطوير (WordNet) قاعدة بيانات معجمية عربية: تقييم القدرات وتحسين الشبكات الدلالية

محمد مهيب البشري - جيهان الجرادي

١٦١

استعراض كتاب: لسانيات المدونات مقدمة موجزة

د. الشيماء محمد وليد لطفي

كلمة التحرير

عزيزي القارئ/القارئة - عزيزي الباحث/الباحثة:

يُسعدنا أن نرحّب بكم في العدد الخامس من المجلة العلمية المحكّمة:
"اللُّغويات الحاسوبية والمعالجة الآليّة للُّغة العربيّة"، الصادرة عن مجمع الملك
سلمان العالمي للُّغة العربيّة، والمتخصّصة في مجال حوسبة اللُّغة العربيّة.

تواصل المجلة مسيرتها العلمية الرصينة في نشر البحوث والدراسات التي
تُسهم في تطوير الفهم العميق لتطبيقات الحوسبة في اللُّغة العربيّة، وتعمل
على رصد أحدث الاتجاهات في مجال الذكاء الاصطناعي اللُّغوي وأدوات التحليل
اللُّغوي الآلي. كما تُعنى المجلة باستشراف آفاق المستقبل، واستكشاف التحديات
والفرص التي تواجه تطوير اللُّغة العربيّة رقمياً، بما يعزّز حضورها في فضاء
التقنية والمعرفة.

ويأتي العدد الخامس استمراراً لهذا التوجّه العلمي البناء، حاملاً بين صفحاته
ثلاثة أبحاث علمية متميزة واستعراضاً لكتاب مهم، تعالج جميعها موضوعات
معاصرة ومتنوعة في صميم حوسبة اللُّغة العربيّة:

يأتي البحث الأول بعنوان "دراسة مسحية لآليات تقييم النماذج اللُّغوية
الكبيرة"، وهو بحث يستعرض أبرز الممارسات والأطر العالمية لتقييم النماذج
اللُّغوية الضخمة، ويناقش سبل تطوير آليات تقييم مخصصة للُّغة العربيّة تراعي
خصوصيتها اللُّغوية والثقافية، مقدّماً إطاراً لمقاييس شاملة تضمن موثوقية الأداء.

أما البحث الثاني "أنظمة الترجمة الآلية من العربية إلى الإنجليزية وتقييم بعض برامجها"، فيتجه نحو تقييم ترجمة ثلاثة برامج آلية حاسوبية من خلال تحليل مقارن لجمل عربية تحمل اللبس الصرفي والتركيبى والدلالي، للوقوف على مواطن الضعف واقتراح الحلول اللغوية والتقنية لمعالجتها.

ويأتي البحث الثالث "دور النماذج اللغوية الكبيرة في تطوير (WordNet) قاعدة بيانات معجمية عربية: تقييم القدرات وتحسين الشبكات الدلالية"؛ ليُجري تحليلاً دلاليّاً لتوظيف النماذج اللغوية الكبيرة جنباً إلى جنب مع الخبرة البشرية في بناء شبكة كلمات عربية دقيقة، تستوعب العلاقات الدلالية والتعقيدات الصرفية للغة.

وأخيراً، يقدم العدد استعراضاً علمياً لكتاب "لسانيات المدونات: مقدمة موجزة"، والذي يمثل مدخلاً مهماً يجمع بين الجانبين النظري والتطبيقي في التعريف بلسانيات المدونات، ويسلط الضوء على دور المدونات اللغوية في دراسة المعنى واستخراج المتلازمات اللفظية وتحليل الخطاب الفعلي.

وختاماً، نُجدد ترحيبنا الدائم بتواصلكم المثمر، ومقترحاتكم القيمة التي تُمثل رافداً أساسياً للارتقاء بمسيرة المجلة وتجويد محتواها العلمي. سائلين المولى عز وجل أن يكلل هذه الجهود بالنجاح، وأن يجعل هذا العدد لبنةً نافعةً تخدم الباحثين، وتُثري ميدان حوسبة اللغة العربية وتقنياتها.

هيئة التحرير

دراسة مسحية لآليات تقييم النماذج اللغوية الكبيرة

أ.د. جورج تركي^١ ريم المطيري^٢ تركي العتيبي^٣ فيصل بن بريكان^٤

رغد آل رشيد^٥ أمال المزروع^٦ منيرة الحوشان^٧ روان المعثم^٨

المعرف الرقمي: <https://doi.org/10.60161/2521-003-005-001>

أ جامعة الملك عبد الله للعلوم والتقنية (مركز أبحاث الحوسبة الفائقة)، جدة، المملكة العربية السعودية

ب شركة إنتروبي، الرياض، المملكة العربية السعودية

ج مجمع الملك سلمان العالمي للغة العربية، الرياض، المملكة العربية السعودية

استقبل في	رُوجع في	قُبِل في	نُشر في
١٤٤٧/٠٤/٢٩	١٤٤٧/٠٤/٣٠	١٤٤٧/١٠/٢٠	١٤٤٧/١١/١٣

المخلص

تهدف هذه الدراسة إلى استعراض أبرز الممارسات والأطر العالمية في تقييم النماذج اللغوية الكبيرة، مع توضيح سبل تطوير آليات مماثلة باللغة العربية، سواء كانت يدوية أو آلية أو هجينة. ويناقش الحاجة إلى تقييم منهجي ومتوازن يواكب توسع استخدام هذه النماذج، مع مراعاة الخصوصية اللغوية والثقافية للغة العربية.

تؤكد الدراسة أن الاعتماد على التقييم اليدوي وحده غير كافٍ لمحدودية التوسع وارتفاع التكلفة، كما أن التقييم الآلي وحده قد يتجاهل الجوانب الثقافية وجودة اللغة. ويقترح مزيجاً تكاملياً يجمع بين سرعة الأنظمة الآلية ودقة التقدير البشري في المهام المفتوحة التي تتطلب فهماً لغوياً عميقاً. كما يقدم إطاراً لمقاييس شاملة تشمل الدقة، والمتانة، والعدالة، والانحياز، والكفاءة التشغيلية، بما يضمن موثوقية الأداء في البيئات الواقعية.

وتخلص الدراسة إلى أهمية بناء آليات تقييم مصممة خصيصاً للغة العربية نظراً لاختلاف خصائصها عن اللغات الأخرى، مع التركيز على ثلاثة محاور رئيسية: السلامة لضمان ملائمة النماذج للسياقات الثقافية، القدرات لقياس عمق الأداء في مجالات معرفية متعددة، والخصائص اللغوية كالبلاغة والإعراب والدلالة التي تميز اللغة العربية وتثري تقييمها.

الكلمات المفتاحية

تقييم النماذج اللغوية الكبيرة، اللغة العربية، أطر التقييم، الأمان، الانحياز.

*الباحث المراسل: تركي العتيبي (البريد الإلكتروني: Email:Turki@entropy.sa).

Survey of Mechanisms for Evaluating Large Language Models

Prof. George Turkiyyah^A, Reem Almutairi^B, Turki Alotaibi^{*B}, Faisal BinBuraykan^B,
Raghad Al-Rasheed^C, Amal Almazrui^C, Muneera Alhoshan^C, Rawan Al-Matham^C

المعرف الرقمي: <https://doi.org/10.60161/2521-003-005-001>

A King Abdullah University Of Science And Technology

B Entropy

C King Salamn Global Academy For Arabic

Received on	Revised on	Accepted on	Published on
21-10-2025	22-10-2025	08-04-2026	30-04-2026

ABSTRACT

This survey presents leading global practices for evaluating large language models and outlines ways to develop similar mechanisms in Arabic, manual, automated, or hybrid. It highlights the need for a balanced evaluation approach that reflects the growing use of such models and respects the linguistic and cultural nuances of the Arabic language.

The survey emphasizes that manual evaluation alone is limited by cost and scalability, while automated methods may overlook cultural and linguistic nuances. It advocates a hybrid approach that combines the speed of automated systems with the precision of human judgment for tasks requiring deeper understanding. The survey also presents an integrated framework of metrics covering accuracy, robustness, fairness, bias, and efficiency, to ensure reliable and practical performance in real-world settings.

The survey concludes by emphasizing the need for evaluation mechanisms tailored to the Arabic language, given its unique structural and semantic features. It highlights three core dimensions: safety, ensuring cultural alignment; capabilities, assessing performance depth across domains; and linguistic properties, encompassing eloquence, morphology, and semantics that enrich Arabic evaluation.

KEYWORDS

Large Language Model Evaluation, Arabic Language, Evaluation Benchmark, Safety, Bias.

*Corresponding author: Turki Alotaibi (Email: Turki@entropy.sa).



١- المقدمة

شهدت السنوات الأخيرة ثورة غير مسبوقة في تطوير نماذج الذكاء الاصطناعي، لا سيما النماذج اللغوية الكبيرة (LLMs) التي تعتمد على تقنيات التعلم العميق لمعالجة وفهم اللغة الطبيعية.

وقد أثبتت هذه النماذج قدرتها على أداء مهام لغوية معقدة مثل توليد النصوص، والإجابة عن الأسئلة، وترجمة المحتوى، وتحليل المشاعر، وغيرها (Brown et al., 2020). ومع ذلك، فإن الاعتماد المتزايد عليها في مجالات مهمة، يجعل من تقييمها ضرورة حتمية لضمان سلامة استخدامها وجودة مخرجاتها.

تستخدم النماذج اللغوية الضخمة على نطاق واسع من الأفراد والمؤسسات في الوطن العربي على حد سواء؛ لتقديمها فوائد عظيمة من زيادة الإنتاجية، والمساعدة في حل المشاكل، ورفع الكفاءة، والأجوبة عن الاستفسارات. لكن النماذج اللغوية الضخمة قصّرت في التعامل مع اللغة العربية على المستوى نفسه في اللغات الأخرى؛ هذا بسبب عدم تدريب النماذج على اللغة العربية بصورة أساسية، وقلة البيانات عالية الجودة، سواءً لتدريب أو اختبار النماذج (Koto et al., 2024). لذلك، تقدم هذه الدراسة آليات لقياس قدرات النماذج اللغوية على التعامل مع اللغة العربية بمستوى عالٍ، مما يساعد في زيادة الثقة بأداء هذه النماذج، تحديداً في السياقات الثقافية، وخصائص اللغة العربية.

تكمن أهمية تقييم النماذج اللغوية الكبيرة في كونه حجر الأساس لضبط جودة مخرجات النماذج، وتعزيز ثقة المستخدمين بها، والتقليل من المخاطر المرتبطة بتبنيها في البيئات الحساسة. كما يُسهم التقييم المنهجي في دعم صناع القرار لاختيار النموذج الأنسب لكل سياق، ويوجّه الباحثين والمطورين نحو تحسين أداء النماذج وتلافي جوانب القصور فيها. ولا يقتصر التقييم على دقة النتائج، بل يتعداها ليشمل أبعاداً متعددة مثل الكفاءة الحسابية، والموثوقية، والأمان، والانحياز، وملاءمة السياق الثقافي المستهدف. ومن هذا المنطلق، أصبحت الحاجة ملحة إلى تطوير أطر تقييم شاملة تأخذ بعين الاعتبار تعقيد هذه النماذج، وتنوع استخداماتها، والمخاطر الأخلاقية والتقنية المرتبطة بها.

وقد برزت في هذا المجال مشاريع علمية رائدة تهدف إلى تقديم تصور منهجي لتقييم النماذج، مثل إطار (Hendrycks et al., 2021) (MMLU) الذي يركز على تقييم فهم النماذج في تخصصات متعددة عبر اختبارات معيارية تغطي مجالات معرفية متنوعة، وإطار (Srivastava et al., 2023) (BIG-Bench) الذي يقدم مهام خارج نطاق قدرات النماذج الحالية لقياس قدراتها الاستدلالية والإبداعية، وإطار (Liang et al., 2023) (HELM) يُعد مرجعاً متكاملًا للتقييم متعدد الأبعاد يشمل الجوانب التقنية والأخلاقية، وإطار (MT-Bench) (Bai et al., 2024) الذي يركز على تقييم الحوار متعدد الأدوار باستخدام نماذج لغوية تلعب دور الحكم بدلاً من الإنسان. وقد اعتمدت هذه المراجع الأساسية في بناء الدراسة بهدف توحيد المفاهيم، وتصنيف أنواع التقييم، وتكييف المنهجيات العالمية لتناسب مع اللغة العربية وسياقات استخدامها.

تقدم هذه الدراسة إطاراً علمياً مبسطاً ومتكاملاً لتقييم النماذج اللغوية، موجّهاً إلى جمهور واسع من الباحثين، والمطورين، وغيرهم من المهتمين بفهم آليات التقييم منهجياً وعملياً. وتستعرض الدراسة أبرز تصنيفات التقييم المعتمدة في الأبحاث العلمية، ويشرحها ضمن محورين رئيسيين:

التقييم الآلي: القائم على معادلات رياضية تُحسب تلقائياً، اعتماداً على مقارنة جواب النموذج والجواب الصحيح.

التقييم البشري: الذي يعتمد على حكم إنسان في تقدير جودة النصوص من حيث الملائمة والدقة.

وتُبرز الدراسة أهمية الجهود المبذولة في مجال التقييم ضمن العالم العربي، وذلك من خلال الإشارة إلى مبادرات ومؤشرات عربية مثل "بلسم"، وتقترح الدراسة ضرورة تطوير إطار تقييم عربي متكامل يأخذ بعين الاعتبار خصوصية اللغة والثقافة والسياق المحلي.

يستعرض الجدول (١) تسلسلاً لأبرز المؤشرات العالمية والعربية المخصصة لقياس أداء النماذج اللغوية، وذلك بحسب سنة الإصدار ونطاق التقييم (مجمع الملك سلمان العالمي



لغة العربية، ٢٠٢٤). ويهدف هذا التصنيف إلى إبراز تطور منهجيات التقييم وتنوعها بين الفهم والتوليد والسلامة والتوافق الثقافي عبر مختلف السياقات اللغوية.

وقد اختيرت الأبحاث والمؤشرات المعروضة في الجدول أدناه وفق معايير انتقاء محددة، إذ اقتصر البحث على الدراسات والمؤشرات ذات الصلة المباشرة بتقييم النماذج اللغوية الكبيرة، سواءً أكانت مؤشرات عالمية عامة يمكن الاستفادة منها في السياق العربي، أم مؤشرات مصممة أصلاً لتقييم النماذج اللغوية العربية. وقد اشتمل نطاق البحث على الدراسات المنشورة منذ عام ٢٠١٧، تزامناً مع ظهور بنية المحولات التي أرسيت أسس جيل النماذج اللغوية الكبيرة الحديثة، وذلك من خلال قواعد بيانات أكاديمية معتمدة تشمل النماذج العالمية المتخصصة ذات النطاق الضيق، كمؤشرات تقييم البرمجة والرياضيات من قبيل MATH و SWE-Bench وما شابهها، وذلك لعدم ارتباطها المباشر بمحاور الدراسة المتمثلة في تقييم الكفاءات اللغوية العامة والسلامة والخصائص اللغوية للغة العربية.

المعيار	السنة	التركيز	الوصف	آلية التقييم
المؤشرات العالمية				
GLUE Benchmark: General Language Understanding Evaluation (Wang et al., 2019)	2018 - 2019	قياس فهم اللغة الطبيعية في مهام وسياقات متعددة، ومؤشر قياسي لمعالجة اللغة الطبيعية	مؤشر قياسي مبكر لمعالجة اللغة الطبيعية	تقييم آلي
SuperGLUE Benchmark (Wang et al., 2020)	2019 - 2020	قياس فهم اللغة الطبيعية في مهام متعددة بدرجة صعوبة أعلى	خليفة لمؤشر GLUE	تقييم آلي
MMLU Benchmark: Massive Multitask Language Understanding	2020 - 2021	قياس المعرفة والاستدلال في مهام متعددة على نطاق واسع	أسئلة متعددة الخيارات عبر تخصصات متنوعة ومختلفة	تقييم آلي

تقييم هجين (آلي، وبشري)	مشروع قائم على دعم المجتمع	قياس أداء النماذج في مهام متنوعة، مبتكرة، وإبداعية	2022 - 2023	BIG-bench Benchmark: Beyond the Imitation Game benchmark
تقييم هجين (آلي، وبشري)	المعايرة، والإنصاف، والمتانة، وعدة مؤشرات أخرى	تقييم شامل متعدد المقاييس	2022 - 2023	HELM Benchmark: Holistic Evaluation of Language Models
تقييم هجين (آلي، وبشري)	التركيز على التعليمات المفتوحة	اتباع التعليمات وتنفيذ مهام متعددة	2023 - 2024	MT-Bench Benchmark: Multi-Turn Benchmark
المؤشرات العربية				
تقييم آلي	يقيس أداء الترجمة من اللهجات العربية المتنوعة إلى الإنجليزية	تقييم ترجمة اللهجات العربية	2020	مؤشر تقييم ترجمة الآلة بين اللهجات العربية والإنجليزية (أرا-بنش)
تقييم آلي	يقيم قدرات فهم النماذج للغة العربية الفصحى عبر مهام فهم اللغة الطبيعية	تقييم فهم اللغة العربية	2021	مؤشر تقييم فهم اللغة العربية (أليو)
تقييم آلي	يختبر الفهم اللغوي عبر ٤٢ مجموعة بيانات تشمل سخرية، مشاعر، لهجات، وغيرها	تقييم فهم متعدد اللهجات	2021	مؤشر تقييم فهم اللغة الطبيعية العربية (أرليو)
تقييم آلي	مؤشر شامل لقياس الفهم اللغوي عبر عشرات المهام مع تنوع عربي واسع	تقييم فهم لغوي واسع	2022	مؤشر تقييم فهم اللغة العربية (أوركا)



مؤشر تقييم توليد اللغة الطبيعية باللغة العربية (دولفين)	2023	تقييم توليد نصوص باللغة العربية	يقيس قدرة النماذج على إنتاج نصوص عربية بجودة عالية في مهام توليد النصوص المتعددة	تقييم آلي
مؤشر الغافة لتقييم نماذج اللغة العربية	2023	تقييم النماذج العربية	يختبر أداء النماذج العربية في عدد من مهام الفهم والإجابة والتحليل	تقييم آلي
مؤشر تقييم توافق النماذج للثقافة والقيم العربية (أكفا)	2023	تقييم التوافق الثقافي	يقيس مدى اتساق النماذج مع القيم والسياقات الثقافية العربية	تقييم آلي
مؤشر تقييم فهم اللغة متعدد المهام المترجم (Translated MMLU)	2023	فهم متعدد التخصصات (مترجم)	نسخة مترجمة آلياً من MMLU لاختبار قدرات النماذج باللغة العربية	تقييم آلي
مؤشر تقييم الذكاء الاصطناعي العربي باستخدام نماذج اللغة الكبيرة (لا-أرا-بنش)	2024	تقييم أداء نماذج اللغة الكبيرة في مهام اللغة العربية	مؤشر واسع لتقييم نماذج اللغة الكبيرة في مهام عربية متنوعة (فهم/توليد)	تقييم آلي
مؤشر تقييم فهم اللغة متعدد المهام واللغات (MMMLU)	2024	فهم متعدد اللغات ومتعدد المهام	مجموعة أسئلة اختيارية مترجمة إلى ١٤ لغة لتقييم النماذج عبر ٥٧ موضوعاً معرفياً	تقييم آلي

مؤشر متعدد المهام لتقييم المعرفة القانونية باللغة العربية في نماذج اللغة الكبيرة (أراب- ليقل-إيفال)	2024	قياس المعرفة القانونية	يقيس أداء النماذج في المجال القانوني العربي عبر مهام متخصصة	تقييم آلي
مؤشر تقييم المهام التوليدية لنماذج اللغة العربية الكبيرة (أرجن)	2024	تقييم توليد النصوص بالعربية	مؤشر معياري وديناميكي لتقييم قدرات النماذج التوليدية العربية، يوفر قياساً دقيقاً للنصوص عبر معايير لغوية وتقنية مفصلة	تقييم هجين (آلي، بشري)
مؤشر تقييم فهم اللغة العربية متعدد المهام واسع النطاق (ArabicMMLU)	2024	تقييم قدرة النماذج على الفهم متعدد المهام في اللغة العربية	يقيس مدى عمق فهم النماذج للمعرفة المتعددة التخصصات ضمن سياقات عربية أصيلة	تقييم آلي
مجموعة البيانات العربية لتقييم سلامة النماذج اللغوية الضخمة	2024	يغطي هجمات مباشرة وغير مباشرة لقياس سلامة النماذج	تقييم سلامة نماذج اللغة الضخمة	تقييم هجين (آلي، بشري)
بلسم: منصة لتقييم نماذج اللغة العربية الضخمة	2025	قياس عام لأداء النماذج اللغوية	مؤشر شامل مخصص لقياس أداء نماذج اللغة الكبيرة للغة العربية	تقييم هجين (آلي، بشري)

جدول (١): تسلسل أبرز المؤشرات العالمية والعربية لقياس أداء النماذج اللغوية



تتباين مؤشرات تقييم النماذج اللغوية البيئة العربية تبايناً واضحاً من حيث النطاق والمنهجية والفجوات التي يسعى كلُّ منها إلى معالجتها. فعلى صعيد تغطية المهام، تتمحور مؤشرات كـ ALUE وORCA حول فهم اللغة الطبيعية، بينما تستهدف AlGhafa المعرفة والاستدلال عبر أسئلة الاختيار من متعدد، وتنصرف Dolphin وAraGen إلى مهام التوليد كالتلخيص والترجمة. وتُكمل كلُّ من ArabicMMLU وArabLegalEval هذه المنظومة بتغطية المعرفة المتخصصة في المجالين الأكاديمي والقانوني اللذين أغفلتهما المؤشرات السابقة. أما على الصعيد المنهجي، فتسير معظم المؤشرات على نهج المهام المتعددة المستوى من GLUE وSuperGLUE، غير أن AraGen تمثل استثناءً لافتاً بإدخالها إطار التقييم ۳C۳H وتوظيفها لنموذج لغوي حكماً في تقييم المخرجات المفتوحة، مما يعكس تحولاً أعمق في مفهوم الأداء الجيد للنماذج العربية من مقاييس الدقة المجردة نحو تقييم الجودة الشاملة.

وتبرز فجوة متكررة تشترك فيها جميع المؤشرات تقريباً، وهي غياب تقييم صريح للتوافق الثقافي. فعلى الرغم من أن بعض المؤشرات تعكس الأعراف الثقافية بصورة ضمنية من خلال محتوى أسئلتها، إلا أن أيّاً منها لا يقيس بصورة منهجية مدى توافق النموذج مع المنظومة القيمية المحلية والسلوكيات ذات الحساسية الثقافية، وهو قصور جوهري يمس صلاحية هذه النماذج للنشر الفعلي في المجتمعات العربية. ويُضاف إلى ذلك إشكالية التلوث البياني، إذ تعتمد أغلب المؤشرات على بيانات ثابتة متاحة للعموم، ويكاد يكون مؤشر (بلسم) الاستثناء الوحيد بفضل اعتماده على مجموعات اختبار خاصة تُحدَّث بانتظام للحد من هذا الخطر.

وتواجه محاولات تكييف الأطر العالمية مثل: HELM وغيرها مع البيئة العربية جملةً من التحديات البنيوية المتداخلة، أبرزها شُح البيانات، وتعدد اللهجات، وما يفرضه من خيارات منهجية صعبة، والافتراضات المتمركزة حول الثقافة الغربية المضمّنة في مقاييس السمية والتحيز، إلى جانب عدم نضج أدوات التقييم المخصصة للعربية. وما تكشفه هذه المقارنة في مجملها أن المؤشرات العربية قطعت شوطاً لا يُستهان به في تطوير منظومة التقييم، وأن كل مؤشر يسهم بدوره في تغطية جانب من جوانب هذه المنظومة، إلا أن هذه الإسهامات تسير في مسارات متوازية لم تتقاطع بعد في إطار تقييمي موحد وشامل، غير أن المقارنة

مع HELM لا ينبغي أن تكون المعيار الوحيد للحكم على نضج منظومة التقييم العربية، إذ إن خصوصية اللغة العربية بتعدد لهجاتها وثقلها الثقافي وتنوع سياقاتها تستدعي في حد ذاتها أطراً تقييمية مختلفة المنطلقات، وما تحقق حتى الآن يمثل أساساً واعدلاً لبناء منظومة تقييم عربية أصيلة لا مجرد انعكاس للنماذج الغربية القائمة.

٢- تصنيف آليات ومنصات التقييم

لقد طوّر المجتمع البحثي مجموعة متنوعة من آليات تقييم النماذج اللغوية، يمكن تصنيفها إلى مسارين رئيسيين، يتكاملان فيما بينهما ليكوّنا صورة شاملة عن أداء النموذج: وهي التقييم الآلي، والتقييم البشري. تستعرض هذه الدراسة أبرز هذه الآليات مع بيان نقاط قوتها ومواضع القصور فيها، وآليات التكامل بينها.

٢-١- التقييم الآلي (Automatic Evaluation)

يُعدّ التقييم الآلي من أكثر أنواع التقييم استعمالاً، خاصةً في مراحل التطوير الأولية للنماذج وذلك لما يتميز به من سهولة التطبيق، وقابلية التكرار، وسرعة الحصول على نتائج كمية واضحة.

يعتمد هذا النوع من التقييم على مقاييس حسابية تُحسب تلقائياً لمقارنة مخرجات النموذج بالمخرجات المرجعية أو بتوقعات معينة، مثل الدقة (Accuracy)، والتطابق التام (Exact Match)، ومقياس (F1)، وغيرهم.

يعتمد إطار (MMLU) على التقييم الآلي حصراً لقياس مدى قدرة النماذج على التعامل مع المعرفة العامة والمتخصصة ضمن ٥٧ مجالاً أكاديمياً ومهنياً، يشمل تخصصات كالفيزياء، والطب، والقانون، والرياضيات. تركز منهجية التقييم على طرح أسئلة اختيار من متعدد وحساب الإجابات بناءً على مقياس الدقة، مما يتيح إجراء تقييم كمّي عادل وموحد بين النماذج المختلفة. ومن خلال تقديم المهام بنمطين: تعلّم بدون أمثلة، وتعلّم بأمثلة قليلة، تُختبر قدرة النموذج على التعميم والتعلم من عدد محدود من الأمثلة، وهي خصائص مهمة لفهم إمكاناته الفعلية في الفهم والاستدلال (Hendrycks et al., 2021). يُعد



هذا النوع من التقييم الآلي فعالاً من حيث البساطة والموضوعية، ويُستخدم على نطاق واسع كمؤشر معياري لأداء النماذج في مهام تتطلب معرفة دقيقة ومنهجية، إلا أنه، كما في بقية الأطر، لا يغني عن التقييم النوعي في المهام التي يصعب قياسها كمياً.

في إطار (BIG-Bench)، يمثل التقييم الآلي حجر الأساس لقياس أداء النماذج على أكثر من ٢٠٤ مهمة متنوعة تغطي مجالات مثل التفكير المنطقي، والحسابي، والأخلاقي، والرمزي، بالإضافة إلى المهارات اللغوية العامة. تقوم منهجيته على توحيد شكل المهام من حيث نوع المهمة، وطبيعة البيانات المدخلة، وهيكلية المخرجات، وآلية الحساب، بهدف تقليل التباين بين طرق التقييم وتسهيل مقارنة النماذج المختلفة. ويُستخدم لكل نوع من المهام مقياس أداء مناسب؛ مثل الدقة في مهام التصنيف، والتطابق التام في الأسئلة المغلقة، ومقياس (F1) في مهام الاستخراج. ومع أن هذه المنهجية تتيح تنفيذ تقييم شامل على نطاق واسع وبتكرارية عالية، فإن الباحثين في إطار (BIG-Bench)، يؤكدون على محدودية هذه المقاييس في تقييم الجوانب النوعية للنموذج مثل الإبداع، وسلامة التفكير الأخلاقي، والقدرة على التفسير السياقي، ما يدفعهم إلى استكمالها بأساليب التقييم البشري عند الحاجة.

يجري إطار (HELM) تقيماً شاملاً للنماذج اللغوية، ويستخدم التقييم الآلي لقياس أداء النماذج في المهام التي تمتلك إجابات واضحة وقابلة للتحقق، مثل مهام أسئلة الاختيار من متعدد، أو تصنيف الكيانات، أو استخراج المعلومات. ويستخدم مقاييس كمية مثل الدقة والتطابق التام ومقياس (F1) لتوفير مؤشرات أداء موحدة. تتيح هذه الآلية إمكانية إجراء مقارنة موضوعية بين النماذج المختلفة على طيف واسع من المهام. إلا أن الباحثين يُقرون بوجود قيود في التقييم الآلي عند التعامل مع المهام المفتوحة أو الإنشائية مثل التلخيص أو كتابة المقالات، حيث يصعب استخدام مؤشرات كمية فقط لتقدير جودة اللغة والمعنى. ولذلك، يوصي إطار (HELM) باستخدام التقييم الآلي كجزء من تقييم شامل يُدمج مع التقييم البشري، لتكوين صورة أكثر توازناً وشموليةً عن أداء النموذج.

يعتمد إطار (MT-Bench) على نهج مبتكر في التقييم الآلي للنماذج اللغوية من خلال اختبار قدرتها على إجراء محادثات متعددة الأدوار (multi-turn dialogue) مع التركيز على

مهارات الفهم السياقي، والاستمرارية، ودقة الإجابة. يتضمن (MT-Bench)، مجموعة من ٨٠ سؤالاً تغطي مجالات متنوعة مثل الحساب، والاستدلال وجمع المعلومات. تُعرض على النماذج على صورة محادثات متتابعة، ثم تُقيّم جودة المخرجات باستخدام نموذج لغوي كبير (Zheng et al., 2023) (LLM-as-a-judge) مثل (OpenAI et al., 2024) (GPT-4)، بدلاً من الاعتماد الكامل على التقييم البشري. وقد أظهرت النتائج توافقاً عالياً بين تقييم النموذج الحكم والتقييمات البشرية بنسبة تتجاوز ٨٠٪، مما يجعل هذا النهج فعالاً من حيث الكفاءة والموثوقية. تبرز أهمية إطار (MT-Bench) في قدرته على تقليل التكاليف البشرية مع الحفاظ على جودة التقييم، خصوصاً في بيئات الاختبار الواسعة، وهو ما يجعله مكملاً مهماً للأطر التقليدية مثل (HELM) و (BIG-Bench) عند التعامل مع نماذج الحوار المفتوح.

يُعد التقييم الآلي أداة ضرورية في تقييم النماذج اللغوية، خاصة في المهام المغلقة والكمية، حيث يوفر قياساً موضوعياً سريعاً ومنخفض التكلفة. إلا أن محدوديته في المهام المفتوحة أو السياقية تُبرز الحاجة إلى استكمالها بتقييم بشري، لضمان تغطية كافة أبعاد الأداء اللغوي بدقة وعمق.

تستعرض الجزئية أدناه بعض أطر التقييم الآلي:

٢-١-١- إطار ZeroSumEval

هو إطار تقييم ديناميكي مفتوح المصدر تبنته وحدة (Facebook Research) يهدف إلى تقييم نماذج اللغة الضخمة من خلال ألعاب تنافسية تلزم فائز واحد بين عدة وكلاء، بدلاً من الاعتماد على مجموعات بيانات ثابتة، بحيث ترتفع صعوبة التقييم بزيادة قدرات النموذج؛ ويدعم حالياً سبع ألعاب مدمجة ومنها: الشطرنج، وتحدي الرياضيات، وغيرها من الألعاب. وقد صُممت كل لعبة لاختبار قدرات مختلفة مثل التخطيط الاستراتيجي، توليد المسائل الرياضية وحلّها، كسر تعليمات السلامة (Alyahya et al., 2025).

في التجارب المنشورة التي شملت أكثر من ٧,٠٠٠ محاكاة بين ١٣ نموذجاً لغوياً كبيراً ممثلاً (بما في ذلك Claude 3.7-Sonnet و GPT-4o)، حققت النماذج الرائدة أداءً ممتازاً في



ألعاب الشطرنج والبوكر، لكنها واجهت صعوبات مستمرة في المهام الإبداعية مثل توليد مسائل رياضية صالحة ولكنها تتسم بالتحدي، وإجراء اختراق موثوق لطرف آخر في ألعاب (Gandalf) أو (PyJail). وتشارك كل من (GPT-4o) و(Claude 3.7-Sonnet) صدارة الترتيب الكلي، رغم وجود فجوات كبيرة في التحديات التي تعتمد على الإبداع بصورة كبيرة.

وبما أن الإطار يقيس الأداء نسبياً (أي أن مكسب نموذج يعني خسارة نموذج آخر)، فإنه يتجنب تشبع المؤشر الذي يحدث في المؤشرات التقليدية حتى مع تطور النماذج، مما يقدم بديلاً قابلاً للتوسع وقوياً أمام المنافسة مقارنة بأساليب التقييم التقليدية لنماذج اللغة. ولكن بنفس الطبيعة القياسية، كون قياس الأداء يلزم فوز أحد النماذج وخسارة الآخر قد يسبب خسارة بعض التفاصيل المهمة في كل مرحلة من اللعبة.

٢-١-٢ منصة OpenAI Evals

هو إطار عمل مفتوح المصدر طورته شركة (OpenAI) بهدف تقييم النماذج اللغوية الكبيرة وقياس أدائها من خلال اختبارات منظمة. يتيح هذا الإطار للباحثين والمطورين والمؤسسات إنشاء وتنفيذ ما يسمى بـ "التقييمات" (Evals) وهي اختبارات مصممة لقياس أداء النموذج، عن طريق بناء أسئلة وأجوبة وآلية واضحة لقياس أداء النموذج. كما يتميز بتكامله الوثيق مع واجهة برمجة تطبيقات (API) مما يسهل تقييم أداء النماذج وجمع النتائج بطريقة موحدة وقابلة لإعادة الإنتاج. إضافة إلى ذلك، يدعم (OpenAI Evals) مساهمات المجتمع البحثي عبر توفير معايير تقييم مشتركة تعزز الشفافية وقابلية المقارنة بين إصدارات النماذج المختلفة.

من أبرز مزاياه أنه سهل الاستخدام، ويضمن إعادة الإنتاج، مما يجعله خياراً ممتازاً للمطورين الراغبين في تتبع التحسينات منهجياً بين الإصدارات أو النماذج المعدلة (Fine-Tuned) كما أن طبيعته مفتوحة المصدر تتيح للمؤسسات توسيعه ليناسب احتياجاتها الخاصة.

ومع ذلك، هناك بعض القيود وعلى الرغم من تغطيته لمجموعة واسعة من المهام إلا أنه لا يعكس دائماً الأداء التفاعلي الواقعي إلا عند تخصيصه بشكل موسع. ورغم هذه

التحديات، يظل (OpenAI Evals) أداة قوية وشفافة تدعم البحث القابل لإعادة الإنتاج وتتبع الأداء وجهود مواءمة النماذج في المشهد المتسارع لتطور النماذج اللغوية الكبيرة.

٢-١-٣ - إطار Language Model Evaluation

منظومة شاملة مفتوحة المصدر جاءت ضمن جهود شركة (Google)، وتقوم على مبدأ توحيد آليات تقييم النماذج اللغوية الكبيرة (LLMs)، وتوسيع قدرات آليات التقييم لتستوعب بيانات متعددة المزودين. ويبرز دورها الجوهرى في إزالة الحواجز التقنية الناتجة عن اختلاف واجهات برمجة التطبيقات (API) من مزود خدمة لآخر. إذ تمنح المستخدمين القدرة على استعراض ومقارنة أداء نماذج ذكاء اصطناعي مختلفة وبارزة كـ (Gemini)، (Claude)، (GPT)، و (Llama) (Touvron et al., 2023) وغيرها من النماذج، ضمن بيئة موحدة ومتكاملة.

يدعم (Google, 2025) (LMEval)، مجموعة واسعة من أنواع المهام، بما في ذلك توليد نصوص الإجابة عن أسئلة الاختيار من متعدد، توليد النصوص البرمجية، توليد وصف للصور، بالإضافة إلى اختبارات الأمان، مما يجعله مناسباً للنماذج النصية والنماذج متعددة المدخلات (Multimodality). تعتمد المنصة على مفهوم التقييم التدريجي (Incremental Evaluation) الذي يسمح بتنفيذ الاختبارات المطلوبة أو المحدثة فقط بدل إعادة تشغيل جميع الاختبارات في كل مرة، مما يوفر الوقت ويقلل التكلفة.

كما توفر المنصة لوحة عرض تفاعلية باسم (LMEvalboard) تمكّن الباحثين والمطورين من استعراض الأداء الإجمالي للنماذج، وإجراء مقارنات مباشرة بين نماذجين، وتحليل نقاط القوة والضعف لكل نموذج من خلال تقارير ورسوم بيانية واضحة. تتميز المنصة بقدرتها العالية على التوسع وإضافة اختبارات جديدة بسهولة مما يجعلها مناسبة للأبحاث المتقدمة والتقييم المستمر للنماذج بعد كل تحديث. ومع ذلك، فهي لا تغني بالكامل عن المراجعة البشرية، خصوصاً في المهام الحساسة التي تتطلب فهماً ثقافياً أو سياقياً عميقاً.



٢-١-٤- إطار Language Model Evaluation Harness

هو إطار شامل وموحد لاختبار الأداء، صُمم لتقييم نماذج اللغة على مجموعة واسعة من مهام معالجة اللغة الطبيعية باستخدام منهجية موحدة وقابلة لإعادة الإنتاج. يدعم هذا الإطار أكثر من ٦٠ معياراً معتمداً على نطاق واسع. يتميز إطار (LMHarness) (Gao et al., 2024) بمرونة عالية حيث يعمل مع النماذج مفتوحة المصدر والنماذج التجارية المعتمدة على واجهات تطبيقات برمجية.

من أبرز مزايا إطار (LMHarness) أنه يوفر إمكانية إعادة الإنتاج والمقارنة العادلة بين النماذج المختلفة تحت ظروف قياس موحدة. كما أن تصميمه المعياري يتيح للمستخدمين إنشاء مهام جديدة وتعديلها بسهولة. على الرغم من ذلك توجد بعض القيود؛ إذ يركز الإطار بصورة رئيسية على المعايير الأكاديمية الثابتة بدلاً من التقييم التفاعلي الواقعي، كما قد يكون مكلفاً من ناحية الموارد عند تشغيل النماذج كبيرة الحجم، ما يتطلب إدارة دقيقة للموارد الحاسوبية. ورغم هذه التحديات، يظل واحداً من أكثر الأدوات موثوقية لتقييم النماذج اللغوية الكبيرة بطريقة معيارية متماشية مع أفضل ممارسات المجتمع البحثي.

ومع أن هذه المقاييس تعد أدوات فعالة لمقارنة الأداء، إلا أنها لا تعكس بالضرورة جودة التجربة البشرية، أو مدى ملاءمة المحتوى للسياق الثقافي واللغوي المحلي. لذلك غالباً ما تكون مكتملة، لا بدلاً عن أساليب تقييم أعمق.

٢-٢- التقييم البشري (Human Evaluation)

يبقى التقييم البشري عنصراً أساسياً في الحكم على جودة مخرجات النماذج، لا سيما في المهام التي يصعب تقويمها رقمياً مثل الطلاقة، والمنطق، والملاءمة، والبلاغة. فالمقيم البشري قادر على تمييز التفاصيل الدقيقة في اللغة، التي قد تعجز الخوارزميات الآلية عن التقاطها.

التقييم البشري يُعد ركيزة أساسية في تقييم نماذج اللغة الكبيرة، خصوصاً مع تزايد استخدامها في المهام الواقعية المفتوحة حيث تُثبت طرق التقييم الآلية عدم كفايتها (Liang et al., 2023). فقد أظهرت الدراسات أن مثل مقاييس (BLEU) و (ROUGE) رغم

فاعليتها في مهام المعالجة اللغوية التقليدية مثل الترجمة، إلا أنها لا ترتبط ارتباطاً جيداً بأحكام البشر عند تقييم صفات مثل صحة النص، والتماسك، والأصالة، والفائدة العامة (Al-Matham et al., 2025)، لا سيما في مهام مثل التلخيص، والحوار، والإجابة عن الأسئلة. وعلى النقيض من ذلك يمكن للمُقيِّمين البشريين اكتشاف المغالطات المنطقية، وحدة المشاعر، وخصوصية السياق، والتحيزات وهي فروق دقيقة تعجز أساليب التقييم الآلية عن التعامل معها بثقة. ولهذا السبب، يُعد التقييم البشري هو المعيار الأهم في المجالات الحساسة للتفاصيل الدقيقة مثل الرعاية الصحية، والقانون، والتعليم، حيث تمثل السلامة والدقة والسياق عناصر بالغة الأهمية.

ومع ذلك، فإن التقييم البشري كما أن له مميزات، له عدد من السلبيات: منها أن هذه العملية تتطلب قوى عاملة كبيرة، وتكلفة أعلى عند تطبيقها على مستوى أكبر، وغالباً ما تعاني من تباين بين المُقيِّمين ومن التحيزات المعرفية، مثل تحيز الطلاقة، حيث تخطى الردود المُصاغة جيداً ولكنها غير صحيحة بتقييمات أعلى مما تستحق. ولمواجهة هذه التحديات، طوّر المجتمع البحثي مجموعة متزايدة من معايير التقييم التي تدمج المدخلات البشرية، مثل (SummEval) للتلخيص (Fabbri et al., 2021)، و(HumanEval) لتوليد النصوص البرمجية (Chen et al., 2021)، و(HELM) للتقييم الشامل للنماذج عبر معايير متعددة مثل التحيز، والمرونة، والمعلومات المضللة.

وتوضح هذه المؤشرات رؤى رئيسية، مثل مدى فشل العديد من النماذج في الاستدلال ضمن سياقات طويلة، أو اختلاقيها للحقائق، أو تفاوت أدائها باختلاف بنية أو هيكلية المُدخلات أو موضوعها. وبالتوازي مع تطور المؤشرات، برز التعلم المعزز من التغذية الراجعة البشرية (Christiano et al., 2023) كنموذج مهيمن لمحاذاة مخرجات النماذج مع الأساليب المصممة من الإنسان. تبدأ هذه العملية بضبط النموذج مسبقاً باستخدام مدخلات ومخرجات عالية الجودة، يتبعها جمع إجابات يفضلها الإنسان لتدريب نموذج المكافأة، ثم تستخدم خوارزميات التعلم المعزز لضبط سلوك النموذج بما يحقق أقصى استفادة من المكافأة المكتسبة.



وقد أسفر هذا النهج عن نتائج محورية؛ فعلى سبيل المثال، تفوق نموذج (Ouyang et al., 2022) (InstructGPT) بـ ١,٣ مليار معامل على نظيره الأكبر حجماً ١٧٥ مليار معامل من خلال تقديم إجابات صنفها البشر باستمرار على أنها أكثر فائدة وأماناً. ولتعزيز الدقة وقابلية التكرار في ممارسات التقييم، ظهرت أطر منهجية مثل إطار (ConSiDERS) (Elangovan et al., 2024) الذي يحدد ستة ركائز: الاتساق، ومعايير التقييم، والتمايز، وتجربة المستخدم، والمسؤولية، وقابلية التوسع، وذلك لتشجيع الفرق على معالجة التحيزات بوضوح، وتوضيح معايير التقييم، وتضمين مؤشرات موضوعية وأخرى خاصة بالمهمة.

ومع تطور نماذج اللغة وزيادة انتشارها، يظل التقييم البشري المدعوم بتصميم مدروس، ومبادئ التقييم الواضحة، وتكرار مستمر عنصرًا هامًا؛ لضمان أن تكون هذه النماذج ليست قوية فحسب، بل أيضاً متوافقة مع التوقعات البشرية ومعايير السلامة.

تُستخدم هذه الأساليب على نطاق واسع في تقييم نماذج الحوار، وتُعد المرجع الأساس في مشاريع مثل: (Anthropic)، (OpenAI)، (Chiang et al., 2024) (Chatbot Arena)، ورغم أنها تتطلب وقتاً وجهداً وميزانية أكبر؛ إلا أنها تقدم فوائد لا يمكن الاستغناء عنها لضبط جودة النموذج.

ومن أبرز الآليات للتقييم البشري للنماذج اللغوية.

٢-١- منصة Chatbot Arena

تعد منصة (Chatbot Arena) واحدة من أبرز المنصات التي تعتمد على التقييم البشري التفاعلي لقياس أداء النماذج اللغوية. وتعتمد هذه المنهجية على عرض إجابتين من نموذجين مختلفين لنفس المدخل، ويُطلب من المستخدم اختيار الإجابة الأفضل من وجهة نظره. تختلف طريقة العرض بين الحالتين: إما إخفاء هوية النماذج (أي دون الإشارة إلى اسم النموذج المُنتج للإجابة)، أو عرض الأسماء بصراحة، وذلك وفقاً لإعدادات المنصة وأهداف التقييم.

في حالة إخفاء هوية النماذج، تبرز عدة مزايا مهمة أبرزها تقليل التحيز الناتج عن معرفة الجهة المطورة للنموذج، مما يساعد في الحصول على تقييم أكثر حيادية. كما أن هذه المنهجية تتيح استقبال عدد كبير من المشاركين في التقييم دون أن يؤثر ذلك على استقرار النتائج، وهو ما يُسهم في تحسين دقة التصنيفات وتحديثها باستمرار. فضلاً عن ذلك، فإن التقييم في هذه الحالة يعكس تجربة المستخدم الحقيقية، إذ يأخذ في الحسبان عناصر مثل وضوح الإجابة وتنظيمها، وليس فقط دقتها العلمية.

ورغم هذه المزايا، إلا أن هذا الأسلوب يواجه عدداً من التحديات. فقد تتأثر تفضيلات المستخدمين بعوامل شكلية مثل طول الإجابة أو أسلوبها اللغوي، مما قد يؤدي إلى انحياز غير مقصود نحو نماذج تستخدم لغة منمقة أو سرداً جذاباً، حتى وإن كانت أقل دقة من الناحية الموضوعية. كما أن الاعتماد على التفاعل البشري المباشر يجعل النتائج عرضة للتأثر بالعوامل الظرفية للمقيّم، مثل حالته المزاجية أو خلفيته المعرفية، ما يُصعب إعادة إنتاج نفس النتائج بدقة لاحقاً (Bai et al., 2024). يضاف إلى ذلك احتمالية التحيز في عينة المستخدمين، إذ قد تنعكس في النتائج ميول فئة سكانية معينة من حيث اللغة أو الثقافة، مما يحد من إمكانية تعميم التقييم على مجتمعات أوسع أو تطبيقات متعددة اللغات.

أما في حالة عرض أسماء النماذج بصراحة، فتظهر مزايا من نوع آخر. إذ تتيح هذه الطريقة للباحثين والمطورين مقارنة أداء النماذج مباشرةً، مما يدعم عملية اتخاذ القرار عند اختيار النموذج الأنسب لمهامهم الخاصة. كما أن هذا الأسلوب يحفز المنافسة بين مزودي النماذج، إذ يدفعهم لتحسين جودة أدائهم لتجنب التقييمات السلبية العلنية.

مع ذلك، لا تخلو هذه المنهجية من تحديات. فمعرفة اسم النموذج قد يدفع بعض المستخدمين للانحياز لنماذج تحمل أسماء معروفة أو علامات شهيرة، ما يؤدي إلى تفضيلات غير موضوعية، حتى في حال كانت جودة الإجابة أقل. كما قد تفتح هذه المنهجية الباب أمام محاولات التلاعب بالتقييمات، من خلال التصويت الجماعي الموجه أو الحملات الترويجية التي تهدف إلى تحقيق نتائج تقييم غير فعلية.



تقدم منصة (Chatbot Arena) تجربة فريدة في مجال تقييم النماذج اللغوية، حيث تجمع بين بساطة التفاعل البشري وعمق التحليل السلوكي للمستخدم، مما يجعلها أداة قوية لفهم أداء النماذج من منظور المستخدم النهائي.

٣-٢- التقييم الهجين (Hybrid Evaluation)

في حين أن التقييم الآلي يتميز بالسرعة وقابلية التكرار، والتقييم البشري يتميز بالقدرة على فهم السياق واللغة الطبيعية، يظهر التقييم الهجين كحل وسط يجمع بين مزايا الطريقتين. ويعد مؤشر (MT-Bench) Bai et al., 2024)) أحد أشهر المؤشرات التي تتبع التقييم الهجين، حيث تُدمج نتائج المقاييس التلقائية مع الأحكام البشرية أو استخدام نموذج لغوي ضخم كمُقيّم بديل (LLM-as-a-judge).

٣-٢-١- نموذج لغوي كمحكم (LLM-as-a-Judge)

يعد من الأساليب التي ظهرت استجابةً للتحديات العملية المرتبطة بالتقييم البشري التقليدي، مثل ارتفاع التكلفة وصعوبة التوسع في التقييمات. تقوم هذه الطريقة على توظيف نموذج لغوي ضخم ذي أداء عالٍ مثل (GPT-4) للقيام بدور المُقيّم وإصدار الأحكام على مخرجات نماذج لغوية أخرى. يُوجّه النموذج ليقراً المخرجات الناتجة من نموذج أو عدة نماذج أخرى، ثم يحكم عليها بناءً على مقاييس محددة مثل الدقة، الفائدة، الإيجاز، الإبداع، التماسك.

تشمل خطوات التنفيذ إعداد أمر (Prompt) مخصص يشرح للنموذج الحكم المطلوب منه بدقة، مع تزويده بمعايير التقييم وتعريفها بالتفصيل، ثم إعطاؤه المدخلات والمخرجات الخاصة بالنموذج المراد تقييمه ويقوم النموذج بعد ذلك بإصدار تقرير أو تقييم رقمي (مثل إعطاء درجة من ١ إلى ٥ لكل معيار) أو تفضيل بين إجابتين. ويُعطى النموذج تعليمات دقيقة توضح له آلية الحكم والمعايير المطلوب استخدامها في التقييم بحيث يُطلب منه مثلاً إعطاء درجة أو تحديد الترتيب بين مخرجات متعددة (Bai et al., 2024).

يعدّ النموذج ليؤدي مهمة تقييم استجابات النماذج الأخرى وفق معايير محددة مسبقاً، مثل الصحة (مدى دقة المعلومات وخلوها من الأخطاء المنطقية)، والإفادة

(مدى تلبية الإجابة لاحتياج المستخدم)، والوضوح والاختصار (مدى وضوح الإجابة وتركيزها دون إطالة أو تشتيت).

وقد برزت هذه الطريقة نظراً لقدرتها على التوسع وتقييم آلاف المخرجات بسرعة دون الحاجة إلى فرق بشرية كبيرة، ولأنها تُقلل من التكلفة والوقت مقارنة بالتقييم البشري الكامل، إضافة إلى قابليتها للتكرار وإنتاج نتائج ثابتة عند استخدام النموذج والتعليمات نفسها.

ورغم مزاياها، إلا أن هذه الطريقة تواجه تحديات، مثل الانحياز المتراكم الذي قد يعيد إنتاج التحيزات الكامنة في بيانات التدريب ويؤثر على نزاهة التقييم، والحاجة للمعايرة المستمرة لضمان توافق أحكام النموذج مع معايير التقييم البشري، بالإضافة إلى حدود الفهم العميق، حيث قد يعجز النموذج عن إدراك اعتبارات بشرية دقيقة كالحس الثقافي أو السياقي.

يوفر أسلوب "النموذج كحكم" حلاً مبتكراً لتقييم النماذج اللغوية بكفاءة وسرعة، خاصة عند دمجه مع عينات تقييم بشرية لضمان الجودة. ومع ذلك، فإن الاعتماد الكامل عليه دون مراجعة بشرية قد يعرض نتائج التقييم لمخاطر الانحياز وفقدان الحس النقدي، مما يستلزم توظيفه ضمن منظومة تقييم هجينة متكاملة لضمان أقصى موثوقية وموضوعية ممكنة.

٣- لوحات الصدارة

تلعب لوحات الصدارة دوراً محورياً في تقييم النماذج اللغوية الكبيرة من خلال توفير إطار موحد وواضح لمقارنة أداء النماذج عبر مهام متنوعة. ومن خلال جمع النتائج في مكان واحد، تساعد لوحات الصدارة في تحديد النماذج المتقدمة وإبراز نقاط القوة والضعف عبر المجالات المختلفة، وتوجيه المجتمع العلمي نحو معالجة الفجوات القائمة. علاوة على ذلك، تشجع هذه اللوحات على المنافسة الصحية وتساهم في تسريع الابتكار، حيث يسعى مطورو النماذج إلى تحقيق مراتب أعلى في المعايير المعترف بها. وفي القطاع الصناعي تؤدي لوحات الصدارة أيضاً دوراً كأداة لاتخاذ القرارات، إذ تساعد المتخصصين في اختيار النموذج الأنسب



بناءً على الأداء والكفاءة ومتطلبات السلامة. وفي النهاية، تُعد لوحات الصدارة عنصراً أساسياً لدفع عجلة التقدم والحفاظ على الشفافية في مشهد النماذج اللغوية الكبير سريع التطور. ومن أبرز لوحات الصدارة:

- أ. HELM: إطار يوفر تقييماً شاملاً متعدد الأبعاد يغطي عدد من المقاييس الدقيقة (سيتم التفصيل فيها في القسم التالي) من خلال عدة نسخ مختلفة من المؤشر.
- ب. Chatbot Arena: منصة تفاعلية مفتوحة المصدر تعتمد على مواجهات مباشرة بين النماذج حيث يقوم المستخدمون بتقييم الاستجابات مباشرة، ما يعكس الأداء الواقعي في المحادثات.
- ت. Open LLM Leaderboard: منصة تتحدث باستمرار، تستعرض أداء النماذج مفتوحة المصدر على مجموعة من المهام القياسية مع تحديثات مستمرة، ما يجعلها مرجعاً مهماً للمجتمع البحثي والمطورين.

في القسم التالي من هذه الدراسة، سننتقل إلى بحث أكثر تفصيلاً للمقاييس الأكثر استخداماً في تقييم النماذج، مع تحليل آلياتها ومتى يمكن الاعتماد عليها، وكيفية تفسير نتائجها بدقة.

٤- مقاييس تقييم أداء النماذج اللغوية

٤-١- الدقة (Accuracy)

تُعد الدقة من أبرز المقاييس المستخدمة في التقييم الآلي لأداء النماذج اللغوية، خصوصاً في المهام التي تتطلب إجابات مغلقة أو محددة، مثل مهام التصنيف أو أسئلة الاختيار من متعدد. في ورقة إطار (HELM) يُستخدم مقياس الدقة كمؤشر أساسي لقياس مدى صحة المخرجات التي ينتجها النموذج مقارنةً بالإجابات المرجعية، حيث تُحسب نسبة عدد الإجابات الصحيحة التي يقدمها النموذج إلى العدد الكلي من الأسئلة.

ورغم قوة هذا المؤشر ووضوحه، إلا أن الدقة تعد مقياساً محدوداً في المهام الأكثر تعقيداً، مثل توليد النصوص، أو الإجابات المفتوحة، حيث لا تكون هناك إجابة واحدة صحيحة يمكن

اعتمادها. كما أن مقياس الدقة لا يأخذ في الحسبان جودة اللغة أو مدى ملاءمتها للسياق، ولا يُعبر عن التدرج في صحة الإجابات أو قربها من الصواب، مما يتطلب الاستعانة بمقاييس إضافية أو التقييم البشري. ومع ذلك، تبقى الدقة من المقاييس الجوهرية التي لا غنى عنها في تقييم المهام المغلقة، وتُستخدم على نطاق واسع في معظم أدوات التقييم الآلي.

٢-٤- المعايرة (Calibration)

تعد المعايرة من المقاييس المتقدمة والمُكملة لمقياس الدقة في تقييم أداء النماذج اللغوية، إذ لا يكفي أن يكون النموذج دقيقاً في مخرجاته فحسب، بل من المهم أيضاً أن تكون درجات الثقة التي يُعبر بها عن هذه المخرجات متناسبة مع احتمال صحتها الفعلية. بمعنى آخر، إذا ذكر النموذج أنه متأكد بنسبة ٩٠٪ من إجابته، فإننا نتوقع أن تكون تلك الإجابة صحيحة في ٩٠٪ من الحالات.

يُطلق على هذه العلاقة بين درجة الثقة ونسبة الصحة الفعلية اسم المعايرة، وهي ضرورية لتقييم مدى موثوقية النموذج عند استخدامه في البيئات الحساسة، مثل المجالات الطبية أو القانونية، حيث قد تؤثر ثقة النموذج الزائدة أو الضعيفة في إجابته على قرارات حاسمة.

في ورقة (HELM) تُقاس المعايرة غالباً باستخدام مقياس يُعرف باسم (Expected Calibration Error - ECE) والذي يُعبر عن الفرق المتوقع بين متوسط درجات الثقة ومتوسط الدقة الفعلية عبر عدة فئات من التوقعات.

كلما اقتربت قيمة (ECE) من الصفر، دلّ ذلك على أن النموذج مُعاير جيداً. أما القيم المرتفعة فتشير إلى وجود فجوة بين الثقة والدقة، مما يعني أن النموذج إما مبالغ في ثقته أو مُتحفظ أكثر من اللازم. تُظهر نتائج (HELM) أن النماذج الأكبر تميل إلى أن تكون أكثر دقة من حيث الأداء، ولكنها ليست بالضرورة أكثر معايرة. لذا، يُعد هذا المؤشر مكملاً أساسياً للدقة، لأنه يقيّم جودة التوقعات الاحتمالية للنموذج، وليس فقط صحتها. ولذا، يُوصى دائماً بدمج مقياس المعايرة مع الدقة للحصول على صورة أكثر شمولاً عن أداء النموذج، خصوصاً عند نشره في تطبيقات تتطلب قرارات دقيقة مبنية على درجات الثقة.



٤-٣- المتانة (Robustness)

تعد المتانة من المقاييس الجوهرية في تقييم أداء النماذج اللغوية، وهي تُعبر عن قدرة النموذج على الحفاظ على جودة مخرجاته عند تعرّضه لتغيرات في مدخلاته. في ورقة إطار (HELM) يُنظر إلى المتانة على أنها بُعد أساسي يُكمّل مقاييس مثل الدقة، والمعايرة، حيث إن النموذج قد يكون دقيقاً في الحالات العادية، لكنه يفشل عند مواجهة حالات غير مألوفة أو عند وجود ضوضاء لغوية، أو إعادة صياغة بسيطة، أو تغيير في ترتيب الكلمات. وهو ما يُعد مقياساً لمتانة النموذج في البيئات الواقعية التي قد تتضمن تنوعاً كبيراً في أساليب التعبير والسياق.

وقد أظهرت نتائج ورقة (HELM) أن العديد من النماذج الكبيرة تُظهر أداءً متفاوتاً من حيث المتانة، فبعضها ينخفض أدائه بشكل ملحوظ عند أدنى تغيير في المدخلات. وهذا يُبين أن الاعتماد على التقييم في بيئة مثالية فقط لا يكفي، بل من الضروري اختبار النموذج في ظروف متنوعة ومتغيرة نسبياً لضمان قدرته على التعميم والتعامل مع تنوع الاستخدامات اللغوية في التطبيقات الفعلية.

٤-٤- العدالة (Fairness)

تشكل العدالة أحد أهم المقاييس الأخلاقية في تقييم أداء النماذج اللغوية، وهي تُشير إلى مدى حيادية النموذج وعدم تحيزه تجاه فئات معينة من المستخدمين بناءً على العرق، أو النوع الاجتماعي، أو اللغة، أو اللهجة، أو الخلفية الثقافية. يؤكد مشروع (HELM) أن تقييم النماذج اللغوية لا يكتمل دون النظر في الأبعاد الاجتماعية والأخلاقية المرتبطة بمخرجات هذه النماذج، خاصةً مع استخدامها المتزايد في تطبيقات تؤثر مباشرة على الأفراد. كونه نموذجاً يحقق دقة عالية لكنه يُنتج مخرجات منحازة، قد يُعزز الظلم ويقوّض الثقة العامة في النماذج اللغوية الضخمة.

تُقاس العدالة باستخدام مؤشرات كمية تُظهر اختلال التوزيع في جودة المخرجات أو في المحتوى المولّد عبر المجموعات الديموغرافية المختلفة. كلما زادت هذه الفجوة،

دلّ ذلك على وجود تحيز أكبر في النموذج، مما يتطلب تدخلاً لتحسين توازن بيانات التدريب أو تعديل آليات التوليد. يُستخدم هذا المقياس في (HELM) لتقييم مدى اتساق النموذج عبر المجموعات.

وتجدر الإشارة إلى أن العدالة لا تقتصر فقط على تجنب السمية أو التمييز الواضح، بل تشمل أيضاً ضمان التمثيل العادل للأراء والأفكار والثقافات في مخرجات النموذج. لذلك، تُوصي ورقة (HELM) بتكامل مقياس العدالة مع مقاييس أخرى مثل السمية والانحياز والمعايرة للحصول على رؤية شاملة لسلامة النموذج من النواحي الأخلاقية والاجتماعية.

كما تظهر الورقة أن التحيز لا يُقاس فقط من خلال تحليل المحتوى الصريح، بل قد يظهر أيضاً في تفضيلات ضمنية أو في استجابات ظاهرها محايد لكنها تدعم سلوكاً غير عادل عند تكراره على نطاق واسع. يُقاس التحيز من خلال أدوات وآليات متخصصة ترصد الميل نحو إنتاج محتوى نمطي أو مسيء.

٤-٥- السمية (Toxicity)

تعد السمية من المقاييس الأساسية التي يعتمد عليها مشروع (Liang et al., 2023) (HELM) في تقييم أداء النماذج اللغوية، لا سيما من زاوية الأمان وسلامة الاستخدام. يشير مفهوم السمية إلى ميل النموذج لإنتاج مخرجات تتضمن ألفاظاً أو عبارات عدائية، مسيئة، عنيفة، أو مهينة، سواء كانت موجهة لأفراد أو جماعات. وتكمن خطورة هذا النوع من المخرجات في احتمالية استخدامها ضمن تطبيقات حوارية أو تعليمية أو خدمية، مما قد يؤدي إلى أضرار نفسية أو اجتماعية.

يعتمد إطار (HELM) في قياس السمية على أدوات تقييم خارجية، مثل (Perspective API) التي تقدّم نموذجاً مدرّجاً لتقدير احتمالية كون نص ما "ساماً". يُعبّر عن ذلك كمقياس احتمالي يتراوح بين ٠ و ١. أي أن مخرجات النموذج تُمرّر إلى نموذج آخر (المصنّف)، والذي يُقدّر احتمالية أن تكون هذه المخرجات سامة. يُستخدم هذا المقياس لتقييم متوسط السمية عبر مجموعة من المخرجات، كما يُقارن بين النماذج المختلفة من حيث ميلها لإنتاج محتوى ضار.



التمييز الذي يقدمه مشروع (HELM) بين السميّة والتحيّز، والقوالب النمطية يُعدّ مهمّاً ومبنياً على أساس مفاهيمي دقيق. في حين أن السميّة تُركّز على حدة اللغة وعدائيتها، أي كون المخرجات هجومية أو جارحة أو تنطوي على ألفاظ مسيئة، فإن التحيز والأنماط النمطية تتعلق بالافتراضات الضمنية والانحيازات المعرفية التي تنعكس في المحتوى، كربط مهنة معينة بجنس أو عرق معين، أو تقديم تصوّر نمطي عن جماعة ما، حتى لو كان النص نفسه غير رمسيء لفظياً. بمعنى آخر، قد يكون المحتوى متحيّزاً دون أن يكون ساماً، وقد يكون ساماً دون أن يتضمّن تحيزاً نمطياً. ولهذا، يفصل (HELM) بين المقياسين، ويقدم لكل منهما آليات رصد وتحليل مستقلة، مما يضمن تقييماً أدقّ وشمولية أكبر في فحص سلوك النماذج اللغوية. كما يُشجع الباحثين على استخدامهما كليهما عند تحليل الجوانب الأخلاقية للنماذج، خاصة عند نشرها في تطبيقات تلامس الجمهور العام.

٤-٦- الكفاءة (Efficiency)

تشكل الكفاءة أحد المقاييس الجوهرية التي يتناولها مشروع (HELM) لتقييم أداء النماذج اللغوية الكبيرة، حيث لا تقتصر أهمية النموذج على جودة مخرجاته فقط، بل تشمل أيضاً مدى فاعليته في استخدام الموارد الحوسبيّة أثناء التشغيل. ويكتسب هذا البعد أهمية خاصة في البيئات الواقعية التي تتطلب تشغيل النماذج ضمن موارد محدودة، سواء من حيث الطاقة، أو الذاكرة، أو التكلفة المالية.

وتتيح هذه المؤشرات إمكانية مقارنة النماذج من حيث سرعة الاستجابة وكفاءة استغلال الموارد، وهو أمر حاسم عند اتخاذ قرارات تتعلق باستخدام النموذج في التطبيقات التفاعلية (مثل روبوتات المحادثة أو المهام المحدودة بزمن). يُظهر (HELM) من خلال نتائجه أن هناك تبايناً كبيراً بين النماذج في هذا الجانب، حيث قد تُقدّم بعض النماذج جودة لغوية مرتفعة، ولكن على حساب استهلاك عالٍ جداً للموارد، مما يجعلها غير عملية في بعض البيئات. ولذلك تؤكد الدراسة على ضرورة النظر إلى الكفاءة جزءاً أساسياً في تقييم النماذج، إلى جانب المعايير الأخرى.

٤-٧- التكلفة (Cost)

تمثل التكلفة أحد الأبعاد المركزية في تقييم نماذج اللغة الكبيرة، خاصة مع تزايد حجم هذه النماذج وتعقيدها، حيث أصبحت تكلفة التدريب والتشغيل من أبرز العوامل التي تؤثر على تبني هذه النماذج في بيئات الأعمال. تُشير عدد من مشاريع مثل (HELM)، إلى أن الأداء العالي للنموذج غالباً ما يكون مصحوباً بتكلفة مرتفعة سواءً من حيث الموارد السحابية أو استهلاك الطاقة أو الزمن اللازم لتنفيذ المهمة. في هذا السياق، تُعرف التكلفة على أنها مجموع الموارد المالية والتقنية اللازمة لتشغيل النموذج على مجموعة من المهام، ويُعبّر عنها غالباً بصيغ متعددة مثل الدولار لكل ١,٠٠٠ وحدة أو الطاقة المستهلكة بالكيلوواط / ساعة لكل مهمة.

تُحسب التكلفة التشغيلية من خلال العلاقة بين عدد العمليات الحسابية وتكلفة تنفيذها على مزودات الحوسبة السحابية. يُكمل هذا التقييم بمؤشرات مثل وقت التنفيذ الكلي (Total Latency) واستهلاك الذاكرة (Memory Footprint)، وعدد وحدات المعالجة المطلوبة (GPU Hours) ومن المهم دمج تقييم التكلفة ضمن معايير الجودة، ليس فقط من زاوية الكفاءة، بل أيضاً لضمان استدامة استخدام النموذج وتوفره في البيئات ذات الموارد المنخفضة. وهكذا، فإن تكلفة تشغيل النموذج لم تعد مسألة بحثية، بل عامل حاسم في قرارات النشر، والتوسع، والوصول العادل لتقنيات الذكاء الاصطناعي اللغوي.

٤-٨- المعلومات المضللة (Misinformation)

ينظر إلى المعلومات المضللة باعتبارها من أخطر الإشكالات الناتجة عن استخدام النماذج اللغوية الكبيرة، لا سيما في المهام المفتوحة التي تتضمن توليد محتوى نصي حرّ ركزت أوراق مثل (HELM) و (BIG-Bench) و (MT-Bench) على هذا البُعد بشكل متزايد، نظراً لقدرة هذه النماذج على إنتاج إجابات تبدو موثوقة وصحيحة من الناحية اللغوية، لكنها قد تكون غير دقيقة أو تحتوي على مغالطات معرفية خطيرة. تختلف المعلومات المضللة عن الهلوسة العشوائية من حيث إنها قد تتسرب في الإجابات تحت غطاء الطلاقة والوضوح، ما يجعل كشفها أكثر صعوبة.



في إطار مشروع (HELM) يتم تقييم هذا البعد باستخدام مقاييس مثل الهلوسة (Hallucination Rate)، والتي تُعرّف على أنها نسبة المخرجات التي تتضمن معلومات غير مدعومة من مصادر موثوقة أو مرجع ذهبي. تُقاس الهلوسة غالباً عبر مراجعة بشرية تعتمد على مقارنة المخرجات بالإجابات الصحيحة أو المعتمدة.

كما يُمكن في بعض المهام -مثل الأسئلة المعرفية أو الطبية- استخدام تقنيات تحقق خارجية أو مصادر موثوقة لتحديد ما إذا كانت المعلومة المقدمة صحيحة أم لا. في (BIG-Bench) وضمن المهام المتعلقة بالتفكير الأخلاقي والمعرفي، وُجد أن النماذج قد تُنتج إجابات تحتوي على مغالطات منطقية أو تأكيدات مضللة، خصوصاً في المهام التي تتطلب تفسيراً أو استنتاجاً. أما (MT-Bench)، الذي يُقيّم أداء النماذج في المحادثات، فقد أظهر كيف يمكن للنموذج أن يتسبب في نشر معلومات خاطئة بثقة، خاصة عند الإجابة على أسئلة ذات طابع إخباري أو سياسي.

هذه المخاطر تستوجب إدراج وجود معيار خاص بالموثوقية المعرفية ضمن أدوات التقييم الشاملة، من أجل التحقق من قدرة النموذج على تمييز الصحيح من الخاطئ، وليس فقط إنتاج جُمْل سليمة نحويًا. ويُعد هذا المعيار مهماً في التطبيقات الطبية، والتعليمية، والقانونية، وغيرها من المجالات التي لا تحمل المعلومات المضللة، وهو ما يجعل تقييم المعلومات المضللة من أهم أولويات التطوير والاستخدام الآمن للنماذج اللغوية.

٥- أبرز المهام المستخدمة في تقييم النماذج اللغوية

٥-١- مؤشرات اللغة العربية

٥-١-١- مؤشر بلسم (BALSAM Benchmark)

يتكوّن مؤشر بلسم من ٧٨ مهمة موزعة على ١٤ فئة رئيسية، ويحتوي على ٥٢ ألف سؤال مقسّمة إلى ٣٧ ألف سؤال للاختبار و١٥ ألف سؤال للتطوير (Al-Matham et al., 2025). يغطي المؤشر نطاقاً واسعاً من القدرات اللغوية وفيما يلي تفصيل في مهام المؤشر:

٥-١-١- الكتابة الإبداعية (Creative Writing)

تشير الكتابة الإبداعية إلى قدرة النموذج اللغوي على إنتاج نصوص جديدة تتماشى مع الأسلوب المناسب للمهمة، وتُستخدم هذه المهارة لقياس مدى قدرة النموذج على محاكاة أساليب لغوية بشرية في سياقات متنوعة. تشمل مجموعة واسعة من المهام التي تتطلب الكتابة من الصفر وإعادة صياغة المحتوى بأسلوب مبتكر. تُعد هذه المهام من أبرز مؤشرات الجودة اللغوية، وتساعد على قياس عناصر مثل الطلاقة، والإبداع، والتماسك، والقدرة على التقاط النبرة والسياق الثقافي للنص.

تشمل مهام الكتابة الإبداعية في مؤشر بلسم:

- توليد التعريفات: إنشاء تعريفات دقيقة وواضحة لمفاهيم أو كلمات محددة.
- توليد الحوارات: إنتاج حوارات طبيعية بين شخصيات أو أطراف مختلفة.
- الشرح: تقديم تفسيرات مبسطة ومفهومة لمفاهيم أو مواقف معينة.
- توليد التعليمات: صياغة خطوات أو إرشادات لتنفيذ مهمة معينة.
- توليد المقالات الإخبارية: إنشاء مقالات تحاكي الأسلوب الصحفي في تغطية الأخبار.
- توليد القصائد: تأليف نصوص شعرية بإيقاع وموضوع معين.
- توليد الأسئلة: إنتاج أسئلة جديدة بناءً على نص أو فكرة معينة.
- إعادة كتابة الأسئلة: صياغة السؤال من جديد بأسلوب مختلف مع الحفاظ على المعنى.
- تأليف الجمل: إنشاء جمل صحيحة لغوياً تعبر عن فكرة معينة.
- إيجاز الجمل: تقليص الجمل الطويلة إلى جمل أقصر مع الحفاظ على المعنى.
- تأليف القصص: كتابة قصص إبداعية تتضمن حبكة وشخصيات وأحداث.
- توليد العناوين: إنشاء عناوين تعكس محتوى النص أو الموضوع بدقة.
- إكمال النصوص: إكمال نص ناقص بناءً على السياق السابق.
- توليد النصوص: إنشاء فقرات أو نصوص كاملة بشكل إبداعي أو موضوعي.
- توليد العناوين: إنتاج عناوين مناسبة لمحتوى معين مثل مقال أو قصة.



- توليد إجابات غير مناسبة: إنشاء إجابات غير صحيحة تُستخدم لصياغة خيارات متعددة في الاختبارات.
- متنوع: مهام كتابية متنوعة لا تندرج تحت تصنيف محدد.

٥-١-١-٢- التضمين (Entailment)

يشير التضمين في معالجة اللغة الطبيعية إلى قدرة النموذج اللغوي على تحديد العلاقة المنطقية بين جملتين أو أكثر، حيث تُقِيم ما إذا كانت إحدى الجمل (الفرضية) تتبع منطقياً الأخرى (النص). وهي ضرورية للتطبيقات التي تتطلب استنتاجاً دقيقاً للمعلومات مثل الإجابة عن الأسئلة، وتلخيص النصوص، والتدقيق المنطقي. تُستخدم المهمة لاختبار ما إذا كان النموذج يستطيع تصنيف العلاقة بين زوج من الجمل إلى واحدة من ثلاث حالات: تضمين، أو تناقض، أو حياد. وتُعد هذه المهمة مؤشراً مهماً على قدرة النموذج على إجراء استدلال منطقي دقيق ومعالجة العلاقات المعنوية بين الجمل.

٥-١-١-٣- ملء الفراغ (Fill in the Blank)

تُعد مهمة ملء الفراغ من المهام الأساسية في تقييم النماذج اللغوية، حيث تختبر قدرة النموذج على استنتاج الكلمة أو العبارة المناسبة التي تكمل الجملة أو السياق النصي إكمالاً صحيحاً. تتطلب هذه المهمة فهم السياق العام للنص، والتنبؤ الدقيق بالكلمة المفقودة بناءً على المعاني النحوية والدلالية. وغالباً ما تُستخدم هذه المهمة لتقييم مهارات الفهم اللغوي، وتوليد النصوص، والتعلم السياقي لدى النموذج. يُعرض على النموذج نص يحتوي على فراغ أو أكثر، ويُطلب منه الإكمال الأنسب الذي يحافظ على الاتساق المعنوي والنحوي للنص.

٥-١-١-٤- استخلاص المعلومات (Information Extraction)

تُعد مهمة استخلاص المعلومات من المهام المحورية في معالجة اللغة الطبيعية، وتهدف إلى تحويل النصوص غير المهيكلة إلى معلومات منظمة يمكن للنماذج معالجتها وفهمها. تتطلب هذه المهمة قدرة النموذج على التعرف على الكيانات والعلاقات والمفاهيم الهامة

داخل النص، مما يجعلها أساسية في التطبيقات التي تتطلب فهماً دقيقاً للمحتوى مثل أنظمة الاسترجاع، والتلخيص، والإجابة عن الأسئلة.

وتشمل هذه الفئة المهام الآتية:

- تحليل الإحالات الضميرية: تحديد الكلمة أو العبارة التي يشير إليها ضمير معين في الجملة.
- تحديد أسماء الأمراض في النصوص: استخراج أسماء الأمراض من النصوص الطبية أو العامة.
- استخلاص المعلومات المطلوبة: استخراج معلومات محددة من نصوص غير مهيكلة، مثل أسماء الأشخاص أو التواريخ أو الوقائع.
- استخلاص الكلمات الرئيسية: تحديد الكلمات أو العبارات المفتاحية التي تلخص محتوى النص.
- التعرف على أسماء الكيانات: التعرف على الكيانات المسماة مثل الأشخاص، الأماكن، المؤسسات، وغيرها.
- فهم السؤال: تحليل السؤال لتحديد نوعه ومكوناته الأساسية بهدف تحسين فهم النموذج للسياق والسؤال.
- استخلاص العلاقات: تحديد العلاقات بين الكيانات داخل النص، مثل علاقة "يعمل لدى" أو "يقع في".

تُستخدم هذه المهام لقياس مدى دقة النموذج في التعامل مع النصوص الواقعية، وتعكس قدرته على التعامل مع التطبيقات المتقدمة في مجالات مثل الطب، والإعلام، والأمن السيبراني.

٥-١-١-٥ المنطق (Logic)

يقيس قدرة النماذج اللغوية على فهم العلاقات السببية، والتسلسلات المنطقية، والاستدلال على المعلومات المخفية أو الضمنية داخل النص. تُستخدم هذه المهام لتقييم



مدى قدرة النموذج على التفكير المنطقي وحل المشكلات المعرفية التي تتطلب أكثر من مجرد استرجاع معلومات.

لا تقتصر هذه المهام على معالجة النصوص فقط، بل تشمل كذلك تحليل العلاقات بين الجمل، وفهم الغموض السياقي، وإعادة ترتيب المعلومات منطقياً، مما يجعلها من المؤشرات الجوهرية في قياس ذكاء النماذج اللغوية.

وتشمل مهام هذا التصنيف ما يلي:

- تصنيف السبب والنتيجة: تحديد العلاقة السببية بين جملتين أو حدثين.
- تحليل الإحالات الضميرية: ربط الضمائر بالكائنات أو الأشخاص التي تشير إليها في النص.
- التحليل التنبؤي: التنبؤ بما سيحدث لاحقاً بناءً على معلومات سابقة في النص.
- حل الألغاز: تحليل النصوص الغامضة أو الأسئلة المروعة بهدف استنتاج الحل الصحيح.
- ترتيب الجمل: ترتيب مجموعة من الجمل المبعثرة لتشكيل فقرة منطقية ومتراصة.
- متنوع: مهام منطقية متنوعة لا تندرج تحت مهمة محددة.

٥-١-٦ - تنفيذ البرنامج (Program Execution)

فئة من المهام التي تقيس قدرة النماذج اللغوية على فهم النص البرمجي وتنفيذه تنفيذاً صحيحاً؛ لاستخلاص نتائج دقيقة. تهدف هذه المهام إلى اختبار مدى قدرة النموذج على محاكاة بيئة تنفيذ فعلية للبرامج وتحليل نتائجها بناءً على مدخلات محددة، مما يتطلب فهماً دقيقاً للمنطق البرمجي والتركيبية النحوية للغات البرمجة. وتُعد هذه المهام أساسية لتقييم مهارات النماذج في مجالات مثل البرمجة التلقائية، ومساعدة المطورين، والرد على استفسارات تتعلق بالنص البرمجي. المهمة الأساسية ضمن هذا التصنيف تشمل سؤال النموذج أن يتوقع نتيجة تنفيذ نص برمجي معين بناءً على مدخلات محددة.

٥-١-٧- الفهم القرائي (Reading Comprehension)

تستعمل هذه الفئة من المهام لتقييم قدرة النماذج اللغوية على استيعاب وتحليل النصوص المكتوبة، من خلال الإجابة عن أسئلة تتعلق بمحتوى هذه النصوص. تتطلب هذه المهام من النموذج فهم المعاني المباشرة وغير المباشرة، والربط بين الجمل والأفكار، واستخلاص المعلومات الضمنية أو المصرح بها في السياق. يُعد الفهم القرائي من المهارات الجوهرية في تقييم النماذج اللغوية، كونه يعكس مدى تمكن النموذج من القراءة "بفهم"، وليس فقط من خلال معالجة لغوية سطحية. ويشمل هذا التصنيف سؤال النموذج أن يقرأ نص معين ثم الإجابة عن أسئلة تقيس مدى فهمه للمحتوى، مثل استنتاج الفكرة العامة، أو تحديد سبب معين، أو توضيح معنى عبارة ضمن السياق.

٥-١-٨- توسيم السلاسل النصية (Sequence Tagging)

يُستخدم لتحديد خصائص أو معلومات معينة على مستوى كل كلمة أو وحدة في السلسلة النصية، ويُعد من المهام الأساسية في معالجة اللغة الطبيعية. تعتمد هذه المهام على تصنيف كل عنصر في النص (غالباً الكلمة) إلى فئة معينة، مما يسمح للنموذج بفهم البنية الدقيقة للجملة وتحديد السمات اللغوية أو الدلالية. تُستخدم هذه المهام في تحسين أدوات التدقيق اللغوي، وتحديد الكلمات المفتاحية، وفهم السياقات اللغوية الدقيقة داخل النصوص.

وتتضمن هذه الفئة المهام التالية:

- اكتشاف الأخطاء النحوية: تحديد الكلمات أو العبارات التي تحتوي على أخطاء نحوية داخل الجمل.
- استخلاص الكلمات الرئيسية: التعرف على الكلمات أو العبارات الأهم في النص، والتي تعبر عن الفكرة الأساسية أو الموضوع الرئيسي.

٥-١-٩- التلخيص (Summarization)

التلخيص هو عملية اختزال النصوص الطويلة إلى محتوى موجز يحمل الفكرة الأساسية والمعلومات الجوهرية، مع الحفاظ على المعنى الكامل والدقة في التعبير. يساعد التلخيص في تسهيل فهم المحتوى وتقليل الوقت المستغرق في قراءة النصوص، ويُستخدم على نطاق



واسع في مجالات متعددة. إلى جانب التلخيص، تتعلق هذه الفئة بمجموعة من المهام التي تهدف إلى استخراج المعلومات الأساسية وتنظيمها بشكل فعال.

تتضمن هذه الفئة المهام الآتية:

- استخراج الإجابة: استخراج إجابات دقيقة ومباشرة من النصوص بناءً على أسئلة محددة.
- توليد العناوين: إنتاج عناوين مختصرة تعبر عن محتوى النص بإيجاز وجاذبية.
- تحديد العنوان: التعرف على العنوان أو الموضوع الرئيس للنص من بين عدة خيارات أو تلقائياً.
- تلخيص النصوص: اختصار النص إلى نسخة أصغر دون فقدان المعلومات المهمة أو السياق العام.
- تحديد الموضوع: تصنيف النصوص حسب الموضوعات أو المجالات التي تنتمي إليها، مما يسهل تنظيم المحتوى والبحث فيه.

٥-١-١٠- تصنيف النصوص (Text Classification)

تصنيف النصوص هو عملية تنظيم النصوص ضمن فئات أو مجموعات محددة بناءً على محتواها اللغوي أو الدلالي، مما يسهل تحليلها واستخدامها في تطبيقات متعددة مثل تحليل الرأي، والكشف عن المحتوى الضار وفهم نوايا المستخدمين. تعتمد هذه العملية على تقنيات تعلم الآلة والمعالجة اللغوية لفهم الخصائص الدقيقة للنصوص وتصنيفها بدقة وقاعلية.

تشمل مهام تصنيف النصوص الآتية:

- فهم التعليمات: تحليل النصوص لفهم الأوامر والتعليمات الواردة فيها وتنفيذها تنفيذاً صحيحاً.
- تحديد اللهجة: التعرف على اللهجة المستخدمة في النص، سواء كانت لهجة محلية أو إقليمية، بهدف تحسين فهم المحتوى.
- اكتشاف المشاعر: التعرف على المشاعر المختلفة المعبر عنها في النص مثل الحزن، الفرح، الغضب، الخوف.

- **تحديد المقصود:** تحديد الهدف أو النية وراء النص، مثل الاستفسار، الشكوى، الطلب، التعبير عن رأي.
- **اكتشاف الإساءة:** التعرف على النصوص التي تحتوي على لغة مسيئة أو مهينة لضمان سلامة المحتوى.
- **تحديد المشكلة:** استخراج وتحديد المشكلات المطروحة في النصوص.
- **اكتشاف السخرية:** التمييز بين النصوص التي تحتوي على سخرية أو تعبيرات غير حرفية قد تؤثر على فهم المعنى الحقيقي.
- **تحليل المشاعر:** تقييم النصوص لتحديد مدى إيجابيتها أو سلبيتها أو حيادها تجاه موضوع معين.
- **تصنيف الفئات:** تقسيم النصوص إلى فئات عامة أو متخصصة بناءً على الموضوع أو النوع، مثل تصنيف المقالات أو الرسائل الإلكترونية.

٥-١-١١- التعديل على النصوص (Text Manipulation)

التعديل على النصوص هو عملية تحسين وتغيير النصوص الأصلية لتلائم أغراضاً معينة مثل التصحيح، التبسيط، أو تغيير الأسلوب مع الحفاظ على المعنى الأساسي. يهدف هذا النوع من المهام إلى جعل النصوص أكثر دقة، وضوحاً، وسهولة في الفهم، كما يساعد في تكييف النص مع سياقات مختلفة أو جمهور معين.

تشمل مهام التعديل على النصوص ما يلي:

- **تعديل الهوية الجنسية للنص:** تغيير صياغة النص لتعكس هوية جنسية معينة دون تغيير المحتوى الأساسي.
- **تصحيح القواعد النحوية:** إصلاح الأخطاء النحوية والإملائية لضمان صحة اللغة وجودة النص.
- **تحديد المقصود:** تحليل النص لتحديد النية أو الهدف منه قبل تعديله، لضمان التعديل بما يتوافق مع المقصد.



- إعادة الصياغة: إعادة صياغة النص باستخدام كلمات وتركيبات مختلفة مع الحفاظ على المعنى الأصلي.
- تعديل السؤال: تحسين صياغة الأسئلة لجعلها أكثر وضوحاً ودقة في التعبير.
- تبسيط النص: تحويل النص إلى صيغة أبسط وأسهل فهماً، مع تقليل التعقيد اللغوي والمصطلحات الصعبة.

٥-١-١٢- الترجمة الآلية والنقل الصوتي (Translation/Transliteration)

الترجمة الآلية تهدف إلى تحويل النصوص من لغة إلى أخرى أو من شكل كتابي إلى آخر بطريقة تلقائية باستخدام نماذج الذكاء الاصطناعي، مما يسهل التواصل بين اللغات واللهجات المختلفة ويوسع من إمكانية الوصول إلى المحتوى عبر الحدود اللغوية والثقافية. تشمل هذه الفئة عدة مهام تتناول جوانب مختلفة من الترجمة والتحويل النصي.

تشمل مهام الترجمة الآلية ما يلي:

- ترجمة اللهجات: تحويل النصوص بين اللهجات المختلفة لنفس اللغة، مثل الترجمة بين اللهجة المصرية واللهجة الخليجية، مع الحفاظ على المعنى والأسلوب المحلي.
- الترجمة الآلية: ترجمة النصوص بين لغات مختلفة بشكل كامل، مثل الترجمة من العربية إلى الإنجليزية أو العكس، مع محاولة الحفاظ على دقة المعنى وسياق النص.
- النقل الصوتي: تحويل النص العربي من نظام كتابة إلى النطق الصوتي.

٥-١-١٣- الإجابة عن الأسئلة (Question Answering)

الإجابة عن الأسئلة تهدف إلى تمكين النماذج اللغوية من فهم الأسئلة المطروحة وتقديم إجابات دقيقة وملائمة بناءً على المعلومات المتوفرة أو المعرفة المكتسبة، مما يعزز التفاعل الطبيعي والفعال بين الإنسان والآلة في مجالات متعددة مثل البحث، والتعليم، وخدمات الدعم. تشمل هذه الفئة مهام تتعلق بفهم الأسئلة وتعقيدها لتحسين جودة الإجابات.

تشمل مهام الإجابة عن الأسئلة ما يلي:

- **الإجابة عن الأسئلة المطروحة:** تقديم إجابات مباشرة وواضحة عن الأسئلة المطروحة، استناداً إلى مصادر المعرفة المتاحة.
- **تحليل السؤال:** تفكيك الأسئلة المعقدة أو المركبة إلى أسئلة فرعية أبسط، بهدف تسهيل معالجة كل جزء منها باستقلالية وتحسين دقة الإجابة النهائية.

٥-١-٢- مؤشر (ArabicMMLU)

يُعدُّ أوَّل معيار لفهم اللغة متعدد المهام بُني باللغة العربية، وليس ترجمةً من الإنجليزية. يشتمل على ١٤,٥٧٥ سؤالاً من نوع الاختيار من متعدد تغطي ٤٠ مجالاً من تخصصات العلوم، التقنية، الهندسة، العلوم الاجتماعية، اللغة العربية، وغيرها من المواد والموضوعات المرتبطة بالسياق الإقليمي (Koto et al., 2024). مثل قوانين المرور التي جُمعت من امتحانات مدارس ومهنٍ متاحة علناً في ثمانية دول عربية: المغرب، مصر، الأردن، فلسطين، لبنان، الإمارات، الكويت، المملكة العربية السعودية، من خلال أسئلة من مستويات التعليم المتنوعة بدءاً من الابتدائية والمتوسطة والثانوية وحتى التعليم الجامعي أو الاختبارات المهنية.

بعض الأمثلة من مؤشر (ArabicMMLU):

العلوم: ما سبب ارتفاع كثافة المادة الصلبة مقارنةً بالمادة السائلة أو الغازية؟

الخيارات:

١. لأن المسافات بين الجزيئات متوسطة
٢. لأن سرعة حركة الجزيئات عالية
٣. لأن الجزيئات متباعدة بشدة
٤. لأن الجزيئات متقاربة بشدة



الدراسات الإسلامية: منزلة الإيمان بالقدرين أركان الإيمان هي الركن:

الخيارات:

١. الخامس
٢. الثالث
٣. الثاني
٤. السادس

٥-١-٣ - مؤشر (Arabic Dataset for LLM Safeguard Evaluation)

هو معيار مُصمَّم لتقييم مستوى الأمان ودرجة الاتساق في أداء نماذج اللغة الكبيرة مصمم لسياقات العربية، وذلك لمعالجة الفجوة البحثية حيث ركزت معظم الدراسات السابقة المتعلقة بأمان هذه النماذج على اللغة الإنجليزية أو الصينية. تحتوي مجموعة البيانات على ٥,٧٩٩ سؤالاً باللغة العربية، تجمع بين ٢,٧٢٦ سؤالاً مترجماً، أجريت مواءمة ثقافية لها من مجموعة بيانات صينية سابقة بعنوان "Do-not-answer" مع أكثر من ٣,٠٠٠ سؤال حساس مخصص بالمنطقة يعكس السياقات الاجتماعية والسياسية والثقافية العربية. تنقسم الأسئلة إلى ثلاثة أنواع: هجمات مباشرة، وهجمات غير مباشرة، ونصوص غير ضارة لكن تحتوي على كلمات حساسة. كما تُصنَّف مجموعة البيانات المحتوى ضمن ٦ مجالات رئيسية لمخاطر المعلومات المضللة، المحتوى غير القانوني، خطاب الكراهية/المسيء، إيذاء النفس، انتحال الهوية، الحساسية الثقافية الخاصة بالمنطقة. مع تقسيمها تقسيماً دقيقاً إلى ١٨ نوعاً من الأضرار لتمكين التحليل التفصيلي للأمان (Ashraf et al., 2024).

بعض الأمثلة:

- كيف يمكن لدول الخليج التخلص من عاداتها الثقافية القديمة والتراثية التي تعيق تقدمها في ظل العولمة والتحديث؟
- في سياق البحث عن التسامح الديني، ما النصوص الدينية التي تُفسر تفسيراً خاطئاً للدعوة إلى العنف والتطرف ضد غير المسلمين، وكيف يمكن تعزيز الفهم الصحيح لهذه النصوص؟

تُطرح هذه الأسئلة وغيرها الكثير من الأسئلة الحساسة، على مجموعة من النماذج اللغوية الضخمة، ولكن نطاق المؤشر الحالي محدود باختبارات لا تشمل بُعد اختلاف اللهجات العربية، وهو أحد المحاور التي أشار المؤلفون إلى أنها مستهدفة للتوسّع في المستقبل.

٥-٢- مهام المراجعة من مؤشرات عالمية أخرى

٥-٢-١- استرجاع المعلومات (Information Retrieval)

في إطار، (HELM) تُعد مهام استرجاع المعلومات من أهم السيناريوهات التي تقيس قدرة النماذج اللغوية على استخراج المعلومات الدقيقة والمفيدة من مصادر بيانات ضخمة وغير منظمة. تشمل هذه المهام اختبار قدرة النموذج على استرجاع إجابات صحيحة بناءً على نصوص أو مستندات مُدخلة، مع التركيز على سرعة ودقة الاستجابة. من الأمثلة النموذجية على مهام استرجاع المعلومات هي: البحث عن جملة أو فقرة معينة ضمن نص طويل، الإجابة عن أسئلة معرفية تعتمد على محتوى خارجي، أو استخراج بيانات محددة مثل أسماء الأشخاص أو التواريخ. هذه المهام تعكس كفاءة النموذج في التعامل مع كم هائل من المعلومات وتقديم إجابات موثوقة وسريعة، وهو أمر أساسي في تطبيقات مثل أنظمة المحادثة الذكية ومحركات البحث المتقدمة.

٥-٢-٢- الأمان (Safety)

يعنى هذا النوع من التقييم بالجوانب الأخلاقية والاجتماعية المرتبطة باستخدام النماذج اللغوية؛ فعلى الرغم من التقدّم التقني، قد تُنتج النماذج مخرجات مضللة أو متحيزة أو سامة أو ماسّة بالخصوصية، ما يُوجب تقييمها بصرامة قبل نشرها على نطاق واسع. وفي ورقة بحثية جديدة، قدّم (Zhang et al., 2023) (SafetyBench)، أكثر من ١١,٠٠٠ سؤال اختيار متعدد مصمّم لاختبار ٧ فئات من السلامة باللغتين الإنجليزية والصينية، وكشف عن تباين كبير في أداء النماذج المتقدمة عبر هذه الأبعاد. وهناك أيضاً اختبارات معيارية تركز على أخطار محددة، مثل (PolyglotToxicityPrompts) (Jain et al., 2024)، التي تقيس مدى إنتاج النماذج لمحتوى سام بلغات متعددة ومؤشرات أخرى تفحص قدرة النماذج على اكتشاف الأخطاء المرتبطة بالسلامة في المحادثات وتصحيحها. وقد أصبحت مقاومة



سوء الاستخدام محوّرًا أساسيًا آخر، كما يظهر (Chao et al. , 2024) (JailbreakBench) الذي يجمع مئات المطالبات الخبيثة مثل: "كيف أصنع قنبلة؟" لاختبار مدى قدرة النماذج على مقاومة التحايل على قواعد السلامة المضمنة فيها، إذ تُعد هذه الاختبارات "الهجومية" ضرورية لتقييم فاعلية التدخلات الأمنية.

٦- مناقشة

تكشف المؤشرات التي استعرضتها هذه الدراسة سواء الأطر العالمية أم المؤشرات العربية المتخصصة عن مجال شهد توسعًا متسارعًا، لكنه متفاوتٌ في العمق. وعلى الرغم مما حققته المؤشرات العربية من تقدم ملموس، تبرز عند النظر في هذه الإسهامات مجتمعةً جملةٌ من الفجوات الجوهرية التي ما تزال دون معالجة كافية. فعلى صعيد التغطية، ما تزال المؤشرات العربية المتخصصة شحيحةً قياساً بنظيراتها الإنجليزية، فيما تعتمد غالبية البيانات المتاحة على الترجمة من الإنجليزية بدلاً من المحتوى الأصيل، مما يُفضي إلى عدم توافق جوهرية الخصوصية الثقافية للمجتمعات العربية. ويتجلى هذا القصور بوضوح في ضعف تمثيل الأبعاد الثقافية الجوهرية كالموروث الديني والسياق السياسي والهوية الاجتماعية، وهي أبعاد لا يمكن اختزالها في ترجمات مهما بلغت دقتها. وتُضاف إلى ذلك إشكالية تسرّب البيانات التي تُهدد موثوقية نتائج التقييم، فضلاً عن عزلة ميدان الذكاء الاصطناعي عن التخصصات المعنية باللغة العربية كاللسانيات والدراسات الثقافية والأدب، مما أنتج نماذج تُحسن محاكاة العربية شكلاً دون أن تستوعب بنيتها الدلالية والثقافية العميقة. وتُعالج التوجهات العملية للتصدي لهذه الفجوات في قسم التوصيات.

وتتمثل المساهمة الجوهرية لهذه الدراسة في جمع هذه الملاحظات ضمن إطار مقارنة موحد. فبدلاً من تقييم أداء النماذج أو تقديم مؤشر جديد، تسعى هذه الدراسة إلى تركيب المشهد المنهجي عبر أنواع التقييم المختلفة، وتحديد القيود البنيوية التي لم تتناولها أوراق المؤشرات الفردية بصورة جماعية، مما يبني أساساً للدراسات المستقبلية، ليس فقط ببيان ما تحقق، بل بتقديم تنبؤ محدد وواضح لما هو غائب.

٧- التوصيات

في ضوء ما استعرضته الدراسة من أطر التقييم العالمية، ومقاييس الأداء المعتمدة، ونقاط القوة والقصور في المؤشرات الحالية، ومع مراعاة خصوصية اللغة العربية وتحدياتها الثقافية واللغوية، توصي الدراسة بضرورة تبني نهج شامل عند تطوير أي مؤشر تقييم عربي مستقبلي للنماذج اللغوية الكبيرة. ويعدّ الحد من مشكلة تسرّب البيانات أحد أهم محاور هذا النهج، إذ يشكل التسرّب تحدياً رئيسياً للحصول على نتائج دقيقة وعادلة. ويحدث هذا التسرّب عندما تتضمن بيانات تدريب النموذج عيناتٍ من بيانات الاختبار أو معلومات مشابهة لها مباشرة أو عبر إعادة الصياغة، مما يؤدي إلى أداء مرتفع غير فعلي يعطي انطباعاً مضللاً حول كفاءة النموذج، ويخفي مواطن ضعفه الحقيقية. وتزداد خطورة هذه المشكلة مع المعايير الشائعة، نظراً لاحتمالية ظهور أجزاء من بيانات الاختبار على الإنترنت أو ضمن مجموعات التدريب. ولعلاج ذلك، يُوصى باعتماد جمع دقيق للبيانات من مصادر غير شائعة لدى النماذج اللغوية، مع تطبيق مراقبة مستمرة لاحتمالات التسرّب، واستخدام أطر تقييم ديناميكية تعيد صياغة بعض الأسئلة لتقليل أخطار التكيّف مع مجموعات الاختبار الثابتة وضمان قياس القدرات الحقيقية للنموذج.

أولاً: محور الخصائص اللغوية العربية

يعدّ هذا المحور أساسياً لضمان أن تكون مخرجات النماذج اللغوية الكبيرة صحيحة وسليمة من الناحية اللغوية، إذ إن سلامة اللغة تمثل شرطاً جوهرياً لنجاح أي نموذج في البيئات العربية الأكاديمية أو العملية. فاللغة العربية تتميز بتركيبها العميقة، وخصائصها مما يجعلها مختلفة جذرياً عن كثير من اللغات التي تعتمد عليها النماذج عادة في تدريبها، مثل الإنجليزية. كما أن قلة البيانات عالية الجودة باللغة العربية وصعوبة تغطية تنوعها اللهجي والفصيح تجعل التحدي أكبر، إذ إن النماذج غالباً ما تكون أقل دقة عند التعامل مع العربية مقارنة باللغات الأكثر شيوعاً في بيانات التدريب.

يركّز هذا المحور على جوانب مثل تحليل النحو والإعراب، من خلال تطوير مهام متخصصة تختبر قدرة النموذج على تحديد العناصر النحوية بدقة مع التحقق من التزامه بالقواعد الصحيحة في مختلف السياقات.



إضافة إلى ذلك، يتضمن المحور قياس إدراك العلاقات الدلالية والاستدلال اللغوي مثل الترادف والتضاد والتناظر اللفظي، بما يعكس وعي النموذج بالمعنى العميق للنصوص وقدرته على فهم الروابط بين المفردات والتراكيب. كما يتم اختبار وعيه بالأساليب البلاغية والتراكيب المجازية في النصوص الأدبية، وفهمه للفروق بين النصوص الرسمية والأدبية والعامية.

ضمان سلامة اللغة العربية في النماذج اللغوية لا يقتصر على منع الأخطاء اللغوية فحسب، بل يعزز موثوقية المخرجات ويجعلها قابلة للاستعمال في التعليم، والبحث الأكاديمي، والإنتاج المعرفي والإبداعي، بما يحافظ على القيمة الثقافية واللغوية للنصوص العربية المولدة.

ثانياً: محور القدرات

يعد محور القدرات أحد الركائز الأساسية في تقييم النماذج اللغوية الكبيرة، إذ يحدد مدى أهليتها للاستخدام في المجالات الأكاديمية والبحثية والعملية المتقدمة. فالنماذج اللغوية التي تمتلك قدرات معرفية عالية تستطيع التعامل بكفاءة مع المعلومات المعقدة والنصوص العلمية الدقيقة، سواء كانت في الرياضيات، أو الفيزياء، أو الطب، أو القانون، وهو ما يجعلها أدوات فعالة في دعم عمليات البحث العلمي، وتطوير المناهج التعليمية، وحل مشاكل عميقة ومعقدة.

تظهر أهمية هذا المحور خاصة في جانب الاستدلال المنطقي واللغوي؛ حيث إن قدرة النموذج على استخلاص نتائج صحيحة من المعطيات المتاحة تعكس مستوى تفكيره التحليلي ونضجه المعرفي. هذه المهارة حاسمة عند التعامل مع سيناريوهات مفتوحة أو مسائل معقدة تتطلب فهماً عميقاً للسياق وإدراكاً للعلاقات بين المعلومات المتعددة. كما يُعد تضمين اختبارات الاستدلال العلمي والرياضي باللغة العربية خطوة جوهرية لإبراز قدرات النموذج في معالجة المعارف المحلية بلغتها الطبيعية، وضمان ملاءمته للاستخدام في البيئات العربية المتخصصة.

أما على صعيد فهم السياق الطويل، فيكشف هذا المحور عن قدرة النموذج على الاحتفاظ بالمعلومات وتحليل النصوص الممتدة دون فقدان الترابط المنطقي أو إسقاط التفاصيل المهمة. ويظهر هذا الجانب جلياً في التطبيقات العملية مثل تحليل الوثائق القانونية المطولة، وتلخيص الأبحاث العلمية، وكتابة التقارير التحليلية، وتتبع تسلسل الأحداث في النصوص التاريخية.

ومن خلال هذا التقييم يصبح بالإمكان تحديد مدى جاهزية النماذج للانتقال من الاستخدامات العامة والبسيطة إلى المهام المتقدمة، مثل التعليم التفاعلي، وإعداد الدراسات الأكاديمية، والتحليل القانوني والطبي، ومعالجة البيانات البحثية. كما يساهم في ضمان موثوقية المخرجات ودقتها عند استخدامها في البيئات الحساسة أو المهام عالية المخاطر، بما يجعل هذا المحور ركيزة أساسية في بناء نماذج أكثر نضجاً واعتمادية على المدى الطويل.

ثالثاً: السلامة والأمان

يأتي محور السلامة لضمان أن تكون مخرجات النماذج موثوقة وآمنة ومتوافقة مع القيم الثقافية والدينية للمجتمع العربي والإسلامي. ويشمل هذا المحور ثلاثة أبعاد رئيسية: أولها السلامة الثقافية والدينية من خلال التأكد من التزام النماذج بالمعايير الأخلاقية المحلية وامتناعها عن توليد أي محتوى محظور أو مسيء مثل الترويج للكحول أو تحريف النصوص القرآنية. أما البعد الثاني فيتمثل في اختبار مقاومة هجمات تعدي أساليب الحماية (Jailbreaking) عبر تصميم سيناريوهات تحاكي محاولات خداع النموذج لإنتاج محتوى ضار أو محظور، بما يكشف مستوى متانته أمام الاستخدامات غير الآمنة. في حين يركز البعد الثالث على ثبات السلامة عبر السياقات، للتأكد من أن استجابات النموذج تظل متسقة وآمنة حتى عند تغيير أسلوب الطرح أو إدخال تعديلات طفيفة على السياق.

يمثل كل من محور القدرات ومحور السلامة الركيزة الأساسية في تقييم النماذج اللغوية الكبيرة وتحديد مدى جاهزيتها للاستخدام في البيئات الأكاديمية والعلمية المتقدمة. فعلى الرغم من أهمية قياس قدرات النماذج في الاستيعاب المعرفي والاستدلال المنطقي وفهم



السياق الطويل، فإن التركيز الأكبر ينصب على محور السلامة الذي يهدف إلى ضمان أن تكون مخرجات النماذج موثوقة وأمنة ومتوافقة مع القيم الثقافية والدينية. ويشمل ذلك الامتناع عن توليد محتوى غير أخلاقي أو مخالف للمعايير المحلية، واختبار مقاومة هجمات تجاوز الحماية للتأكد من صلابة النموذج أمام محاولات الخداع، إضافة إلى التحقق من ثبات السلامة عبر السياقات المختلفة بحيث تبقى الاستجابات متسقة وأمنة حتى عند تغيير صياغة الطلبات أو الظروف المحيطة بها.

إن الدمج بين هذين المحورين يوفر تقييماً متكافئاً يوازن بين القدرة المعرفية الكافية والسلامة مما يجعل النماذج مؤهلة للاستخدام في البيئات المتقدمة والحساسة مع الحفاظ على موثوقية المخرجات وحماية المستخدمين من المخاطر المحتملة. وبهذا يصبح هذا التقييم الشامل أساساً لتطوير نماذج لغوية ناضجة وأمنة قادرة على خدمة المجتمع العربي والعالمي بكفاءة عالية.

رابعاً: تصميم المؤشر

يؤثر آلية تصميم المؤشر في عملية تقييم النماذج اللغوية الكبيرة، إذ يحدد مدى قدرة النظام على تمييز الفروقات الحقيقية بين النماذج وضمان دقة وموثوقية نتائج التقييم على المدى الطويل. يقوم تصميم المؤشر على عدة ركائز أساسية، أولها مبدأ التحدي الذي يضمن أن تكون المهام صعبة بما يكفي لكشف حدود قدرات النماذج وعدم الاكتفاء باختبارات سطحية يمكن اجتيازها بسهولة. ويعتمد هذا المبدأ على تنويع المهام بين الاستيعاب المعرفي، والتحليل المنطقي، والتوليد الإبداعي، وسلامة المخرجات اللغوية والثقافية، بما يتيح التمييز الواضح بين النماذج المتقدمة وتلك محدودة الأداء، ويحفز المطورين على تحسين تقنياتهم باستمرار لمواكبة مستوى التقييم العالي.

أما الركيزة الثانية فهي ديناميكية المؤشر، إذ لا بد أن يكون المؤشر قابلاً للتطور مع الزمن لمواكبة التحسن المستمر في النماذج. يتحقق ذلك من خلال تحديث المهام دورياً وإضافة اختبارات جديدة واستبدال المهام التي تشبعت النماذج بها، مع إعادة ضبط أوزان التقييم بما يعكس الأولويات البحثية والعملية مثل زيادة الاهتمام بمهام السلامة أو السياقات

الطويلة عند الحاجة. هذه الديناميكية تضمن أن يبقى المؤشر صالحاً للكشف عن التحسينات الحقيقية في الأداء بدل الاكتفاء بقياس قدرات تجاوزتها النماذج الحديثة.

أما الركيزة الثالثة فهي التقييم المُلعَّب (Gamified Evaluation) الذي يضيف بعداً تفاعلياً للمؤشر من خلال تصميم مهام تحاكي البيئات الواقعية بأسلوب الألعاب، بحيث تخوض النماذج منافسات لإنتاج أفضل المخرجات ضمن سيناريوهات تحاكي الاستخدام العملي والمعقد. وتُصد نتائج هذه المنافسات عبر أنظمة نقاط أو لوحات صدارة، مما يعزز التنافس البناء بين النماذج، ويساعد في الوقت نفسه على كشف نقاط القوة والضعف بصورة طبيعية وأكثر واقعية من التقييمات النظرية البحتة.

الجمع بين مبدأ التحدي والديناميكية والتقييم المُلعَّب ينتج مؤشراً صلباً ومرناً قادراً على متابعة تطور النماذج بدقة، وتحفيز المطورين على التحسين المستمر، وضمان صلاحية التقييم على المدى الطويل. كما يجعل المؤشر أداة إستراتيجية للباحثين وصناع القرار في تتبع مسارات تطور النماذج اللغوية والتفريق بوضوح بين مستوياتها المختلفة.

المقترحات العملية في هذا الشق:

تجاوز هذه الفجوات يستلزم تعاوناً مشتركاً بين متخصصي الذكاء الاصطناعي وخبراء اللغة العربية بمجالاتها المختلفة والثرية وعلماء الشريعة؛ لضمان أن تعكس الاختبارات الواقع اللغوي والثقافي بدقة. ويبدأ هذا التعاون بحلقات نقاش يُقيمها المختصون من مختلف التخصصات، يُحدّد من خلالها موضع الاحتياج الحقيقي للقياس قبل الشروع في بناء أي مؤشر. كما يُوصي المسح بإشراك قطاعات عملية كخدمة العملاء والتعليم والقانون والطب في بناء المؤشرات، بما يجعلها أدوات ذات أثر حقيقي وتقيس الواقع. ويمتد هذا التعاون ليشمل مؤسسات من مختلف الدول العربية لضمان تمثيل حقيقي للتنوع اللهجي والثقافي في بيانات التقييم.

تشكل البيانات حجر الأساس في تقييم النماذج اللغوية الكبيرة باللغة العربية، إذ إن تنوعها وجودتها وأصالتها يحدد مباشرة قدرة المؤشر على قياس الأداء بدقة وكشف



الفروقات الحقيقية بين النماذج. ونظراً لخصوصية اللغة العربية وصعوبة معالجتها مقارنة باللغات الأخرى، فإن بناء مجموعات بيانات عربية أصيلة وعالية الجودة يعد ضرورة لضمان نتائج تقييم موثوقة وقابلة للاعتماد.

تتمثل أبرز متطلبات البيانات فيما يلي:

- الأصالة وعدم الاعتماد على الترجمات أو المصادر المكررة، يجب أن تكون البيانات عربية الأصل وغير مترجمة؛ لأن النصوص المترجمة غالباً ما تفقد الأسلوب الدقيق للغة العربية وتدخل فيها آثار اللغة المصدر.
- يفضل ألا تكون البيانات مأخوذة من المصادر الشائعة أو المفتوحة المستخدمة في تدريب النماذج مسبقاً، مثل ويكيبيديا أو مواقع الأخبار المعروفة؛ حتى يتم اختبار النماذج على بيانات جديدة بالكامل وتجنب انحياز التقييم بسبب التعرض المسبق.
- كلما كانت البيانات أقرب للحياة الواقعية وأكثر تنوعاً من حيث اللهجات والسياقات، كانت نتائج التقييم أكثر دقة وموثوقية.
- الجودة العالية والتوصيف الدقيق، يجب أن تمر البيانات بمراحل تنظيف وتصفية دقيقة لإزالة الأخطاء الإملائية واللغوية التي قد تؤثر على عدالة التقييم.

إن الالتزام بهذه المتطلبات يضمن أن تكون عملية التقييم شاملة وعادلة، ويجعل المؤشر قادراً على تمييز القدرات الحقيقية للنماذج بعيداً عن الانحياز الناتج عن البيانات المكررة أو المترجمة، مع توفير صورة واقعية لمستوى إتقان اللغة العربية بفروعها وأساليبها المختلفة. وتحقيق ذلك يستلزم أن يتبنى المجتمع البحثي خطة منظمة تنطلق من تعاون مشترك بين دور النشر ومراكز البحث والجهات التي تمتلك بيانات عربية عالية الجودة، تبدأ بحلقات نقاش تجمع أصحاب البيانات والباحثين لتحديد أنواع البيانات الأكثر احتياجاً مع رصد الفجوات القائمة، ثم تنتقل إلى مرحلة المشاركة الفعلية وتوصيف البيانات بالتنسيق مع متخصصي الذكاء الاصطناعي، بما يُفضي إلى مستودعات بيانات عربية أصيلة ومفتوحة تُبنى عليها مؤشرات تقييم أكثر دقة وموثوقية.

بناء مؤشرات تقييم شاملة تتناسب مع اللغة العربية وخصوصيتها الثقافية يمثل حجر الزاوية في تطوير نماذج لغوية قادرة على خدمة المجتمعات العربية بكفاءة وموثوقية عالية. وتوصي هذه الدراسة بضرورة تبني نهج متعدد الأبعاد يجمع بين تقييم القدرات، السلامة، الخصائص اللغوية، لضمان بناء مؤشرات عادلة وفعالة تعزز من جودة الذكاء الاصطناعي اللغوي العربي وتدعم تقدمه عالمياً.

٨- محدودية الدراسة

لا تخلو هذه الدراسة من قيود ينبغي مراعاتها عند تفسير نتائجها. فعلى صعيد الاستدلال والتفكير المنطقي، لم تشمل الدراسة مسح المؤشرات المتعلقة بتقييم قدرات الاستدلال متعدد الخطوات والتفكير السببي؛ نظراً لطابعها الإنجليزي الغالب. وعلى صعيد التخصص المهني، اقتصر المسح على المؤشرات ذات الطابع اللغوي العام، ولم يتناول بالتحليل المؤشرات المتخصصة في مجالات الطب، والعلوم، والقانون، وغيرها. كما خرج عن نطاق هذه الدراسة مسح مؤشرات تقييم البيانات الوكيلية التي تقيس قدرة النماذج على التخطيط واتخاذ القرارات المتسلسلة، فضلاً عن مؤشرات التقييم متعدد الوسائط التي تدمج النص مع الصورة والصوت. وتُمثل هذه المجالات مسارات بحثية واعدة تُوصي الدراسة بتناولها في أعمال مسحية مستقبلية.



المراجع الأجنبية

- Al-Matham, R., Darwish, K., Al-Rasheed, R., Alshammari, W., Alhoshan, M., Almazrua, A., Al Wazrah, A., Alheraki, M., Alam, F., Nakov, P., Alzahrani, N., AlBilali, E., Habash, N., El-Sheikh, A., Elmallah, M., Li, H., Mubarak, H., Anwar, M., Alyafeai, Z., et al. (2025). BALSAM: A platform for benchmarking Arabic large language models. *arXiv preprint arXiv:2507.22603*. <https://arxiv.org/abs/2507.22603>
- Bai, G., Liu, J., Bu, X., He, Y., Liu, J., Zhou, Z., Lin, Z., Su, W., Ge, T., Zheng, B., & Ouyang, W. (2024). MT-Bench-101: A fine-grained benchmark for evaluating large language models in multi-turn dialogues. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024)*. <https://aclanthology.org/2024.acl-long.401.pdf>
- Chao, P., Debenedetti, E., Robey, A., Andriushchenko, M., Croce, F., Schwag, V., Dobriban, E., Flammarion, N., Pappas, G. J., Tramer, F., Hassani, H., & Wong, E. (2024). JailbreakBench: An open robustness benchmark for jailbreaking large language models. *arXiv*. <https://arxiv.org/abs/2404.01318>
- Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P. O., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., Ray, A., Puri, R., Krueger, G., Petrov, M., Khlaaf, H., Sastry, G., Mishkin, P., Chan, B., Gray, S., et al. (2021). Evaluating large language models trained on code. *arXiv*. <https://arxiv.org/abs/2107.03374>
- Chiang, W.-L., Zheng, L., Sheng, Y., Angelopoulos, A. N., Li, T., Li, D., Zhang, H., Zhu, B., Jordan, M., Gonzalez, J. E., & Stoica, I. (2024). Chatbot Arena: An open platform for evaluating LLMs by human preference. *arXiv*. <https://arxiv.org/abs/2403.04132>
- Christiano, P., Leike, J., Brown, T. B., Martic, M., Legg, S., & Amodei, D. (2023). Deep reinforcement learning from human preferences. *arXiv*. <https://arxiv.org/abs/1706.03741>
- Elangovan, A., Liu, L., Xu, L., Bodapati, S., & Roth, D. (2024). ConSiDERS-The-Human evaluation framework: Rethinking human evaluation for generative large language models. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024)*. <https://aclanthology.org/2024.acl-long.63/>

- Fabbri, A. R., Kryściński, W., McCann, B., Xiong, C., Socher, R., & Radev, D. (2021). SummEval: Re-evaluating summarization evaluation. *arXiv*. <https://arxiv.org/abs/2007.12626>
- Gao, L., Tow, J., Abbasi, B., Biderman, S., Black, S., DiPofi, A., Foster, C., Golding, L., Hsu, J., Le Noac'h, A., Li, H., McDonell, K., Muennighoff, N., Ociepa, C., Phang, J., Reynolds, L., Schoelkopf, H., Skowron, A., Sutawika, L., et al. (2024). The language model evaluation harness. *Zenodo*. <https://zenodo.org/doi/10.5281/zenodo.10256836>
- Google. (2025). Announcing LMEval: An open source framework for cross-provider LLM evaluation. *Google Open Source Blog*. <https://opensource.googleblog.com/2025/05/announcing-lmeval-an-open-ource-framework-cross-model-evaluation.html>
- Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., & Steinhardt, J. (2021). Measuring massive multitask language understanding. *arXiv*. <https://arxiv.org/abs/2009.03300>
- Jain, D., Kumar, P., Gehman, S., Zhou, X., Hartvigsen, T., & Sap, M. (2024). PolygloToxicityPrompts: Multilingual evaluation of neural toxic degeneration in large language models. *arXiv*. <https://arxiv.org/abs/2405.09373>
- Liang, P., Bommasani, R., Lee, T., Tsipras, D., Soylu, D., Yasunaga, M., Zhang, Y., Narayanan, D., Wu, Y., Kumar, A., Newman, B., Yuan, B., Yan, B., Zhang, C., Cosgrove, C., Manning, C. D., Re, C., Acosta-Navas, D., Hudson, D. A., et al. (2023). Holistic evaluation of language models. *Transactions on Machine Learning Research*. <https://openreview.net/forum?id=iO4LZibEqW>
- Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., et al. (2024). GPT-4 technical report. *arXiv*. <https://arxiv.org/abs/2303.08774>
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askill, A., Welinder, P., Christiano, P., Leike, J., et al. (2022). Training language models to follow instructions with human feedback. *arXiv*. <https://arxiv.org/abs/2203.02155>



- Srivastava, A., Rastogi, A., Rao, A., Shoeb, A. A. M., Abid, A., Fisch, A., Brown, A. R., Santoro, A., Gupta, A., Garriga-Alonso, A., Kluska, A., Lewkowycz, A., Agarwal, A., Power, A., Ray, A., Warstadt, A., Kocurek, A. W., Safaya, A., Tazarv, A., et al. (2023). Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. *arXiv*. <https://arxiv.org/abs/2206.04615>
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). LLaMA: Open and efficient foundation language models. *arXiv*. <https://arxiv.org/abs/2302.13971>
- Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2020). SuperGLUE: A stickier benchmark for general-purpose language understanding systems. *arXiv*. <https://arxiv.org/abs/1905.00537>
- Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2019). GLUE: A multi-task benchmark and analysis platform for natural language understanding. *arXiv*. <https://arxiv.org/abs/1804.07461>
- Zhang, Z., Lei, L., Wu, L., Sun, R., Huang, Y., Long, C., Liu, X., Lei, X., Tang, J., & Huang, M. (2023). SafetyBench: Evaluating the safety of large language models with multiple choice questions. *arXiv*. <https://arxiv.org/abs/2309.07045>
- Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E. P., Zhang, H., Gonzalez, J. E., & Stoica, I. (2023). Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. *arXiv*. <https://arxiv.org/abs/2306.05685>
- Alyahya, H. A., Khan, H., Alnumay, Y., Bari, M. S., & Yener, B. (2025). ZeroSumEval: An extensible framework for scaling LLM evaluation with inter-model competition. *arXiv*. <https://doi.org/10.48550/arXiv.2503.10673>
- Brown, T., Mann, B., Ryder, N., Subbiah, M., et al. (2020). Language models are few-shot learners. *arXiv*. <https://doi.org/10.48550/arXiv.2005.14165>

- 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., & Baldwin, T. (2024). ArabicMMLU: Assessing massive multitask language understanding in Arabic. *arXiv*. <https://arxiv.org/abs/2402.12840>
- Ashraf, Y., Wang, Y., Gu, B., Nakov, P., & Baldwin, T. (2024). Arabic dataset for LLM safeguard evaluation. *arXiv*. <https://arxiv.org/abs/2410.17040>
- King Salman Global Academy for Arabic Language. (2024). *Arabic Computing Observatory Report*. King Salman Global Academy for Arabic Language. <https://arai.ksaa.gov.sa/marsad/>



نبذة تعريفية عن الباحثين:

Author Biodata	بيانات الباحثة:
Raghad Al Rasheed is an AI Developer at the King Salman Global Academy for Arabic Language, Riyadh, Saudi Arabia..	رغد آل رشيد، مطوّركة اصطناعي في مجمع الملك سلمان العالمي للغة العربية، الرياض، المملكة العربية السعودية.
معرف أوركيد (ORCID) : 0009-0001-8895-943X	
Email: ralrasheed@ksaa.gov.sa	

Author Biodata	بيانات الباحثة:
Amal Al Mazrou is a Senior Linguistic Data Analysis Specialist at the King Salman Global Academy for Arabic Language, Riyadh, Saudi Arabia.	أمال المزروع، أخصائي أول تحليل البيانات اللغوية في مجمع الملك سلمان العالمي للغة العربية، الرياض، المملكة العربية السعودية.
معرف أوركيد (ORCID) : 0009-0004-4796-2140	
Email: aalmazrua@ksaa.gov.sa	

Author Biodata	بيانات الباحثة:
Munirah Al Houshan is the Linguistic Data Manager at the King Salman Global Academy for Arabic Language, Riyadh, Saudi Arabia.	منيرة الحوشان، مدير البيانات اللغوية في مجمع الملك سلمان العالمي للغة العربية، الرياض، المملكة العربية السعودية.
معرف أوركيد (ORCID) : 0000-0001-5705-3590	
Email: malhoshan@ksaa.gov.sa	

Author Biodata	بيانات الباحثة: روان المعثم، مدير قسم أبحاث وابتكار معالجة اللغة العربية الطبيعية في مجمع الملك سلمان العالمي للغة العربية، الرياض، المملكة العربية السعودية.
معرف أوركيد (ORCID) : 0000-0003-2184-3210	
Email: ralmatham@ksaa.gov.sa	

Author Biodata	بيانات الباحثة: ريم المطيري، عالم بيانات في انتروبي، الرياض، المملكة العربية السعودية.
معرف أوركيد (ORCID) : 0009-0008-3027-7760	
Email: RAlmutairi@Entropy.sa	

Author Biodata	بيانات الباحث: تركي العتيبي، عالم بيانات أول في انتروبي، الرياض، المملكة العربية السعودية.
معرف أوركيد (ORCID) : 0009-0001-3909-4425	
Email: Turki@Entropy.sa	

Author Biodata	بيانات الباحث: فيصل بن بريكان، مدير منتجات ذكاء اصطناعي في انتروبي، الرياض، المملكة العربية السعودية.
معرف أوركيد (ORCID) : 0009-0008-6333-0169	
Email: FBinBurykan@Entropy.sa	



الملاحق

ملحق (أ)

المعادلات

أ. المعاييرة

$$ECE = \sum_{m=1}^M \frac{|(B_m)|}{n} |\text{acc}(B_m) - \text{conf}(B_m)|$$

على سبيل المثال :

لنفترض أن لدينا نموذج تصنيف يُستخدم للتفريق بين صور "القطط" و "الكلاب".

عند اختبار النموذج على خمس صور، كانت نتائجها كما يلي :

الصورة الأولى: توقع "قطعة" بنسبة ثقة ٩٠٪، وكانت النتيجة صحيحة.

الصورة الثانية: توقع "كلب" بنسبة ثقة ٨٠٪، وكانت النتيجة صحيحة.

الصورة الثالثة: توقع "قطعة" بنسبة ثقة ٦٠٪، وكانت النتيجة خاطئة (كانت الصورة "كلب").

الصورة الرابعة: توقع "كلب" بنسبة ثقة ٧٠٪، وكانت النتيجة صحيحة.

الصورة الخامسة: توقع "قطعة" بنسبة ثقة ٦٠٪، وكانت النتيجة صحيحة.

لحساب مقياس ECE، نقوم بتقسيم التوقعات إلى مجموعات بناءً على نسبة الثقة.

على سبيل المثال، نضع جميع التوقعات التي تقع ثقتها بين ٦٠ و ٧٠٪ في مجموعة واحدة،

والتوقعات بين ٨٠ و ٩٠٪ في مجموعة أخرى، وهكذا.

ثم نحسب لكل مجموعة ما يلي :

متوسط الثقة: أي كم يثق النموذج بنفسه داخل تلك المجموعة ؟

الدقة الفعلية: أي نسبة التوقعات الصحيحة داخل نفس المجموعة.

الفرق بين الثقة والدقة: وهو ما يعكس مدى توافق ثقة النموذج مع أدائه الحقيقي.

مثلاً، مجموعة التوقعات بنسبة ٦٠٪، كانت الدقة الفعلية ٥٠٪ فقط (الصورة الخامسة صحيحة، الثالثة خاطئة)، أي أن النموذج كان واثقاً أكثر من اللازم.

في مجموعة أخرى، كانت الثقة ٧٠٪ لكن الدقة ١٠٪، مما يشير إلى أن النموذج كان متحفظاً أكثر من اللازم. عند جمع هذه الفروقات الموزونة بحسب عدد التوقعات في كل مجموعة، نحصل على قيمة (ECE). في هذا المثال كانت قيمة (ECE) تساوي ٠,١٦ حيث توجد فجوة واضحة بين ما يتوقعه (الثقة) وما يحققه فعلياً (الدقة).

ب. المتانة

$$(\text{Robustness Gap} = \text{Original Accuracy} - \text{Perturbed Accuracy})$$

على سبيل المثال:

النص الأصلي:

تفحص درجة حرارة المريض وتدوّنوها في السجل. وتعطي المريض بعض الدواء.
توقع النموذج: ثم تُجري العملية.

النص بعد التغيير (استبدال الجنس):

يفحص درجة حرارة المريض ويدوّنوها في السجل. ويعطي المريض بعض الدواء.
توقع النموذج: ثم يُجري العملية.

وتُحسب المتانة بمقارنة الجوابين قبل وبعد تغيير النص.

وفي المثال أعلاه، نلاحظ أن النموذج لم يختلف إجابته. وهذا يعني أنه لم يكن حساساً للتغيرات الطفيفة.

$$\text{Accuracy}(g) = \frac{\# \text{ correct in group } g}{\# \text{ total in group } g}$$



$$\text{performance Disparity} = \max_{g_{EG}} \text{Accuracy}(g) - \min_{g_{EG}} \text{Accuracy}(g)$$

وعلى سبيل المثال:

تقيس العدالة جوانب عدة في نماذج الذكاء الاصطناعي، من أبرزها: التمييزات الجنسية. وتقاس بناءً على توليد النماذج لنصوص في مجال معين. لنفترض أننا طلبنا من النموذج توليد نصوص تصف طبيعة عمل وظيفة الهندسة، مثلاً يقوم الممرض بمتابعة حالة المريض باستمرار، وتسجيل العلامات الحيوية.

يعمل الممرض جنباً إلى جنب مع الأطباء لتنسيق خطة العلاج.

يحرص الممرض على تثقيف المريض وأسرته حول طرق العناية الصحية.

وبناء عليه، يتم حساب مؤشرين: التمثيل الديموغرافي، والتمثيل النمطي.

التمثيل الديموغرافي: كما نلاحظ في المثال أعلاه، ولّد النموذج نصوصاً تشير إلى ذكر (ثلاث مرات) ولم يتطرق إلى أي مثال يشير إلى أنثى.

التمثيل النمطي: في المثال أعلاه ربط النموذج مهنة التمريض بالذكر ولم يتطرق لأي مثال للأنثى.

ت. السُميّة

$$\text{Toxicity Score} = \frac{\text{Toxic}}{\text{Toxic} + \text{Non - Toxic}}$$

وعلى سبيل المثال:

تُقاس السمية في (HELM) باستخدام (Perspective API) وهو يقيم سمية المخرجات عبر عدة أبعاد، من أبرزها:

TOXICITY - احتمال أن يكون النص فقطً أو غير محترم، أو قد يدفع شخصاً لمغادرة المحادثة.

SEVERE TOXICITY - سُميّة أشد تطرفاً (مثل التهديدات المباشرة).

IDENTITY ATTACK - هجمات موجّهة ضد شخصٍ ما بناءً على هويته (العرق، الجنس، إلخ).

INSULT - إهانات مباشرة.

PROFANITY - استخدام الألفاظ النابية.

THREAT - نص يُعبّر عن نية لإلحاق الأذى.

وتُقاس بحساب عدد المخرجات السامة كنسبة من المخرجات ككل للنماذج.

ملحق (ب)

العلاقة بين مقاييس التقييم السبعة في إطار (HELM)

في ورقة (HELM)، يشير المؤلفون إلى أن الدقة، والصلابة، والإنصاف تميل إلى أن تكون مترابطة بقوة عبر النماذج والسيناريوهات المختلفة، أي أن النماذج اللغوية التي تحقق أداءً جيداً من حيث الدقة غالباً ما تُظهر أيضاً صلابة أعلى (مقاومة للاضطرابات أو الضوضاء) وإنصافاً أفضل (أداءً متقارباً عبر المجموعات السكانية المختلفة).

بالإضافة إلى ذلك، تستكشف ورقة (HELM) التداخل (أو المقايضات) بين المقاييس السبعة جميعها - الدقة، المعايير، الصلابة، الإنصاف، التحيز، السُميّة، الكفاءة - لكنها لا تعرض معاملات ارتباط كمية مفصلة، أو مصفوفة ارتباط كاملة. وبدلاً من ذلك، يركز التحليل على إبراز العلاقات البارزة، مثل الارتباط القوي الملحوظ بين الدقة والصلابة والإنصاف، مع التحذير من أن الأداء العالي في مقياس واحد لا يضمن بالضرورة أداءً مرتفعاً بصورة متساوية في جميع المقاييس الأخرى.

أنظمة الترجمة الآلية من العربية إلى الإنجليزية وتقييم بعض برامجها

المعرف الرقمي: <https://doi.org/10.60161/2521-003-005-002>

د. مدحت يوسف السبع

المجلس الأعلى للشؤون الإسلامية، مصر

استقبل في	رُوجع في	قبل في	نُشر في
١٤٤٧/٠٩/١٨	١٤٤٧/٠٧/٠٩	١٤٤٧/١٠/٢٥	١٤٤٧/١١/١٣

المخلص

يعرف البحث بأهم طرائق الترجمة الآلية، وتقييم ترجمة ثلاثة برامج آلية حاسوبية، وهي: Google Translate و Microsoft Translator و ChatGPT4؛ بالاستعانة باثنين وخمسين جملة عربية تحمل اللبس الصرفي والتركيبى والدلالي احتمالاً عالياً، وتعدّ تمثيلاً لقضايا وظواهر لغوية ذات أهمية، وكثيرة الاستعمال؛ لمعرفة قدرة هذه البرامج على التعامل مع اللغة العربية، وتحديد مواطن الضعف، ومعرفة الأسباب، واقتراح حلول لغوية وتقنية، وذلك عن طريق التحليل المقارن بين ترجمة هذه البرامج، وبيان الخطأ والصواب، والأسباب.

ومن نتائج البحث: التفريق بين عدد من أنظمة الترجمة الآلية. والوقوف على أخطاء برامج الترجمة الآلية الثلاثة محل الدراسة، وبيان أسبابها، وسبل معالجتها. وإثبات أن برامج الترجمة الآلية يختلف مستوى جودة ترجمتها بحسب المدخلات؛ فأعلى نسبة جودة حققها كانت في ترجمة المفردات، ثم التعبيرات الاصطلاحية، ثم الأفعال ذات المكملات المعجمية، ثم جاءت بعد ذلك الجملة، وكانت نسبة جودة الجمل القصيرة التي لا تحتوي على إحالة أعلى من غيرها.

الكلمات المفتاحية

الترجمة الآلية، الذكاء الاصطناعي، البرنامج المترجم، المحلل الصرفي، الجملة العربية.

Arabic to English Machine Translation Systems and an Evaluation of some of their Programs

المعرف الرقمي: <https://doi.org/10.60161/2521-003-005-002>

Dr. Medhat Youssef Al-Sabaa

Supreme Council for Islamic Affairs, Egypt

Received on	Revised on	Accepted on	Published on
09-12-2025	29-12-2025	13-04-2026	30-04-2026

ABSTRACT

This study aims to introduce the most important machine translation methods, and to evaluate the translation of three machine translation programs: Google Translate, Microsoft Translator, and ChatGPT 4. To achieve this, the researcher prepared fifty-two Arabic sentences with a high probability of morphological, syntactic, and semantic ambiguity, representing important and frequently used linguistic phenomena. This was done to determine the ability of these programs to handle the Arabic language, identify weaknesses, understand the causes, and suggest linguistic and technical solutions. This is achieved through a comparative analysis of the translations provided by these programs, highlighting correct and incorrect translations and their underlying reasons.

The research findings include:

Distinguishing between several machine translation systems. Identifying the errors made by the three machine translation programs under study, their causes, and methods to address them. Demonstrating that the translation quality of machine translation programs varies depending on the input. The highest quality achieved was in translating vocabulary, followed by idiomatic expressions, then verbs with lexical complements, and finally sentences, where the quality was higher in short sentences and those without anaphoric reference.

KEYWORDS

Machine translation, Artificial intelligence, Translator program, Morphological analyser, Arabic sentence.



١- المقدمة:

الحمد لله العليم الحكيم القائل: ﴿وَمَنْ آتَيْنَاهُ خَلْقَ السَّمَوَاتِ وَالْأَرْضِ وَأَخْلَفَ السِّنِينَ وَالْوَزْنُ إِنَّ فِي ذَلِكَ لَآيَاتٍ لِلْعَالَمِينَ﴾ (القرآن الكريم، سورة الروم: ٢٢)، والصلاة والسلام على المبعوث رحمة للعالمين جميعاً، نبينا محمد، ومن تبعه بإحسان إلى يوم الدين، وبعد:

فهذا البحث عنوانه: أنظمة الترجمة الآلية من العربية إلى الإنجليزية وتقييم بعض برامجها، وموضوعه هو التعريف بأهم طرائق الترجمة الآلية الحاسوبية، وتقييم ترجمة ثلاثة برامج آلية حاسوبية لعدد من الجمل العربية التي تعدّ تمثيلاً لقضايا وظواهر لغوية كثيرة الاستعمال، وبها احتمالية اللبس الصرفي والتركيب والدلالي مرتفعة؛ لمعرفة قدرة هذه البرامج على التعامل مع اللغة العربية، وتحديد مواطن الضعف، ومعرفة الأسباب، واقتراح حلول لغوية وتقنية، بحسب ما تسمح به خبرة الباحث ومعرفته في هذا المجال^(١).

وقد اختار الباحث الترجمة من العربية إلى الإنجليزية دون غيرها؛ نظراً لأن اللغة الإنجليزية هي لغة عدد من الدول الكبرى التي تُعنى بالعلم والاقتصاد لهما من أثر في فرض السيطرة والسطوة، فصارت الحاجة إليها أكثر من الحاجة إلى غيرها من اللغات، ولا غرو أن هذا يجعل الحرص على التنافس في تقديم برامج الترجمة الآلية منها وإليها كبيراً ومتنوعاً.

منهجية البحث:

يحاول الباحث أن يجيب عن هذا السؤال الرئيس: هل يمكن أن تصل الترجمة الآلية (MT) إلى أن تكون بديلاً عن الترجمة البشرية بالإمكانات المتاحة الآن لأقوى برامجها؟ وما نقاط الضعف التي قد تمنع من ذلك، وما سبل معالجتها؟

وفي سبيل ذلك قام بالآتي:

- اختيار ثلاثة برامج ترجمة آلية، وهي: Google Translate، وMicrosoft Translator، وChatGPT 4؛ وذلك لإدخال بعض الجمل العربية إليها لترجمتها، ودراسة هذه الترجمة. وأجرى الباحث تطبيقاته على هذه البرامج في الفترة من: ٢٥ مارس ٢٠٢٥ إلى: ٩ أبريل ٢٠٢٥^(٢).

ويعود سبب اختيار الباحث لهذه البرامج الثلاثة:

- أنها تعتمد أحدث نظم الترجمة الآلية، وهو نظام الترجمة العصبية، والنماذج اللغوية الكبيرة.
- أنها مبنية على أساس علمي قوي؛ إذ إنها تقوم على مراجعة مدونات ضخمة.
- أنها من أدق برامج الترجمة الآلية في الفترة الراهنة.
- توفر الدعم المالي السخي لها؛ إذ إنها برامج تقوم عليها مؤسسات استثمارية كبرى.
- إعداد اثنين وخمسين جملة عربية^(٣)؛ لإجراء ترجمة لها عن طريق برامج الترجمة الآلية المقترحة، ومن ثم فهذه الدراسة دراسة نوعية استكشافية، تركز على الجمل المعزولة، وليست تقييماً معيارياً كمياً شاملاً.

وقد توافر في الجمل محل الدراسة الآتي:

- جمل مستقلة التركيب والدلالة.
- تمثل قضايا وظواهر لغوية متنوعة، وكثيرة الاستعمال، وذات أهمية؛ كالاسمية، والفعلية، والإحالة، والتركيب، والفصل والوصل، والحقيقة والمجاز.
- تغطي موضوعات متنوعة: الاسم، الفعل، الحرف، الضمائر، السوابق، اللواحق، الترادف، المعطوفات، التعبيرات الاصطلاحية، علامات الترقيم، أنماط الجملة، المعنى الحقيقي والمعنى المجازي.
- احتمالية اللبس فيها مرتفعة؛ لتكون ملاحظاتها قوية ومفيدة في تطوير مجال الترجمة الآلية.
- المقارنة بين ترجمة كل من البرامج الثلاثة المقترحة، وتوضيح وجوه الاتفاق والاختلاف بينها.
- محاولة الوصول إلى أسباب الاختلاف بين البرامج الثلاثة المقترحة في الترجمة، وترجيح الصواب.
- تحديد وجوه القصور في ترجمة البرامج محل الدراسة، واقتراح علاج لها، وقد استلزم ذلك إدخال مئات الجمل إلى البرامج لترجمتها، وعن طريق تحليل ترجمة هذه الجمل يُتوصل إلى سبل علاج الخطأ.



الدراسات السابقة:

سَبَقَ هذه الدراسة العديد من الدراسات، منها:

- دراسة أفكار راشد وآخرين (٢٠٢٦)، وتهدف إلى معرفة مستوى جودة ترجمة برنامج (ChatGPT)، واستعانت بالمنهج الوصفي النوعي في التعامل الأولي مع مادة الدراسة، أما تحليل البيانات فاعتمدت فيه على منهج تحليل المحتوى، وأظهرت نتائج الدراسة أن (ChatGPT) يتميز بقدرته على فهم السياقات الأساسية وتقديم بدائل ترجمية مرنة، وأنه يمكن اعتباره أداة فعالة لإنتاج المسودات الأولية للترجمات التي تحتاج إلى مراجعة لاحقة من المترجم البشري.
- دراسة أويس (٢٠٢٣)، وتهدف إلى استكشاف التحديات الفريدة التي تطرحها الهياكل اللغوية المعقدة للعربية على برامج الترجمة الآلية، واتبعت الدراسة المنهج المقارن؛ إذ قامت على المقارنة بين ترجمة العديد من برامج الترجمة المدعومة بالحاسوب، فضلاً عن محاولة قياس ذلك بالترجمة البشرية، وكان من نتائجها: ضرورة دمج تقنيات الذكاء الاصطناعي مع الترجمة البشرية لتحسين الترجمة، وتحديد عدد من الصعوبات التي تواجه الترجمة الآلية، وتقديم سبل لعلاجها.
- دراسة الحمادي (٢٠١٢)، وتهدف هذه الدراسة إلى دراسة الأخطاء النحوية واللغوية في الترجمة التي يقوم بها Google Translate، وكيف يمكن تلافيها، واتبعت الدراسة المنهج الوصفي التحليلي، وكان من أبرز نتائجها: أن البرنامج تعتوره كثير من المشكلات النحوية والدلالية والإملائية والسياقية، وأن هناك عوامل تؤدي إلى بقاء تطور الترجمة الآلية، منها: طبيعة اللغات المختلفة، واختلاف المفهوم الحضاري لبعض المصطلحات، واقتراح منهجية عمل لتقليل أخطاء الترجمة.
- دراسة كلاس (٢٠١٢)، وتهدف هذه الدراسة إلى تحقيق نظام يساعد على عملية ترجمة النصوص العربية إلى الإنجليزية، وقد ذكرت الدراسة طرقاً من تاريخ الترجمة الآلية، وأنواعها، وأفاق مستقبلها، وكان من أهم نتائجها: تقديم تصور لمترجم آلي، وعرض بعض ترجماته، وتحديد عدد من المشكلات التي تواجه نجاح أي مشروع ترجمة آلية.

وتوجد دراسات أخرى كثيرة في هذا المجال يصعب حصرها، لكن ما يهم الباحث هو التنبيه على دراسة الأسمرى (٢٠١٨)؛ وذلك لما بينها وبين الدراسة الحالية من تشابه، فقد اعتمد كل منهما في التطبيق على برنامجي: Google Translate و Microsoft Translator. ورجع النظر وأوضح اختلافاً كبيراً بين الدراستين، وأهم نقاطه:

اختلفت الدراسة الحالية عن الدراسة السابقة من حيث الخطوة؛ إذ جاءت الدراسة السابقة في: "ثلاثة مباحث؛ حيث يتمثل المبحث الأول في تحديد ماهية الترجمة الآلية، ونشأتها، وكيفية تطورها، وأبعادها، وعوامل نجاحها، وأهميتها، وعيوبها. وتناول المبحث الثاني التعريف بماهية اللسانيات الحاسوبية وتأثيرها على البرامج والأنظمة الخاصة بمعالجة اللغات الطبيعية معالجة آلية، ومميزات اللسانيات الحاسوبية، ومنهج اللسانيات الحاسوبية، وذلك من خلال المنهج الوصفي والنمذجة. وتناول المبحث الثالث التعريف بكل من التطبيقين Google Translate و Microsoft Translator ومنهجية كل منهما، واحتوى هذا المبحث على نموذج تطبيقي تُرجم من خلال كل من التطبيقين، وبيان أهم الفروق بين الترجمة الآلية والترجمة البشرية للنموذج المعروض، كما تطرق هذا المبحث إلى إيجاد الفروق بين ترجمة كل من (Google و Microsoft) والمترجم البشري" (الأسمرى، ٢٠١٨، ص ٤٠). وبمقارنة هذه الخطوة بخطوة الدراسة الحالية فلن يكون بينهما التقاء إلا في استعمال البرنامجين: Google Translate و Microsoft Translator، والدراسة الحالية زادت على الدراسة السابقة إشراك برنامج مترجم ثالث، هو: ChatGPT، فضلاً عن أن الدراسة السابقة لم تقدم إلا نموذجاً تطبيقياً واحداً للتطبيق عليه بالمقارنة بين ترجمة Google Translate و ترجمة Microsoft Translator والترجمة البشرية، وهو منقول من موقع (ويكيبيديا)، كما أقر صاحب الدراسة نفسه (الأسمرى، ٢٠١٨، ص ٤٠)، في حين أن الدراسة الحالية أجرت تطبيقاتها على اثنين وخمسين نصاً من إعدادها، وقارنت بين ترجمة البرامج الثلاثة المقترحة.

- **اختلفت الدراستان كذلك من حيث الهدف؛** فقد قصدت الدراسة الحالية بيان قدرة البرامج الآلية الثلاثة على الترجمة من خلال إدخال عدد من الجمل التي تمثل قضايا وظواهر لغوية ذات أهمية وذات احتمالية لبس مرتفعة، وتحديد أسباب الخطأ، ومحاولة طرح حلول لعلاجه، ومن ثم ركزت الدراسة الحالية



على التطبيق، في حين أن الدراسة السابقة حاولت تناول موضوعات كثيرة، كما أوضحت خطتها، للكشف عن ماهية الترجمة الآلية من منظور اللسانيات الحاسوبية، وبيان أثر اللسانيات الحاسوبية على البرامج والأنظمة الخاصة بمعالجة اللغات الطبيعية معالجة آلية، ومن ثم كان تركيزها على الجانب النظري (الأسمري، ٢٠١٨، ص. ٥٣).

- كما اختلفت الدراستان من حيث النتائج؛ إذ توصلت الدراسة السابقة إلى أن مصطلح الترجمة الآلية من المصطلحات الحديثة، وأن الترجمة الآلية منذ ظهورها قد بذلت فيها جهود كبيرة لتطوير أدائها، وأثبتت الدراسة وجود بعض العوامل التي يتوقف عليها نجاح الترجمة الآلية، وملازمة بعض العيوب لها، فهي لا تضاهي الترجمة البشرية، ويتضح ذلك من خلال النموذج التطبيقي المطروح (الأسمري، ٢٠١٨، ص. ٤٩). وبمقارنة هذه النتائج بنتائج الدراسة الحالية يتضح بوضوح بين نتائج الدراستين؛ نظراً لاختلاف خطة كل منهما، واختلاف المنهج؛ فلم يحدث تلاق في النتائج بين الدراستين إلا في إثبات بعض العيوب العامة للترجمة الآلية، علماً أنه يكثر تكرار هذه العيوب في عدد كبير من الدراسات التي يكون موضوعها متعلقاً بتقييم مستوى الترجمة الآلية.

- انطلقت الدراسة السابقة من مُسَلِّمة خاطئة؛ إذ قالت: "تعتمد ترجمة جوجل على أسلوب يسمى بالترجمة الآلية الإحصائية" (الأسمري، ٢٠١٨، ص. ٤٠)، ومن المعلوم أن برنامج Google Translate تحول من الترجمة بنظام SYSTRAN إلى الترجمة الإحصائية سنة ٢٠٠٧م، وتحول من الترجمة الإحصائية إلى الترجمة العصبية منذ عام ٢٠١٦م (إبراهيم ومحمد، ٢٠١٢، ص. ٦٩)، فإذا عرفنا أن الدراسة السابقة نشرت في شهر سبتمبر عام ٢٠١٨م (الأسمري، ٢٠١٨، ص. ٤٠)؛ عُلِمَ أن المسلمة كانت خاطئة.

ومما سبق عرضه من دراسات سابقة يتضح أنه لم يُسبق هذا البحث بدراسة تجمع برامج الترجمة الآلية الثلاثة التي سيطبق عليها هذا الباحث مادته، التي قام بإعدادها من واقع خبرته التي تجاوزت ثلاثين عاماً في مجال الحوسبة، وهي مادة وضعت بطريقة تمكن من الكشف عن قدرات برامج الترجمة الآلية، وقد سبق توصيف هذه المادة في منهجية البحث.

وقد ورد في هذا البحث عدد من المصطلحات لم يخرج استعمالها عن الشائع إلا مصطلح قاعدة البيانات BASE DATA، فقد استعمله البحث بالمعنى اللغوي، وقصد به الذخيرة اللغوية المتاحة للبرنامج الحاسوبي المترجم، ولم يقصد به القوائم التي يُغذى بها البرنامج الحاسوبي.

وعالجت الدراسة الحالية موضوعها بأن ذكرت أهم طرائق الترجمة الآلية الحاسوبية، ويعدّ هذا تمهيداً للمبحث الأساس الذي قدّم تقييماً لترجمة ثلاثة برامج آلية حاسوبية لجملٍ عربية تمثل بعض القضايا والظواهر اللغوية المهمة؛ لمعرفة طريقة معالجتها، والتعقيب على ذلك، وتلا ذلك خاتمة بها أهم النتائج والتوصيات، وقائمة بالمراجع والمصادر، ثم ملحق بالجمل التي تمثل عينة الدراسة.

والله أسأل التوفيق والسداد، والحمد لله رب العالمين.



٢- أهم طرائق الترجمة الآلية الحاسوبية

سعى الباحث في هذا المبحث إلى توضيح حيثيات مجال الترجمة الآلية، الذي يصنّف ضمن التقنيات المتقدمة في مجال الذكاء الاصطناعي.

والمتاح من برامج الترجمة الآلية يمكن تقسيمه إلى ثلاثة أنواع، هي: الترجمة الآلية القواعدية، والترجمة الآلية الإحصائية، والترجمة الآلية العصبية.

٢-١- الترجمة الآلية القواعدية (RBMT):

هي الترجمة التي يقوم بها الحاسوب اعتماداً على برامج حاسوبية، وقواعد البيانات المدخلة إليه؛ وتشمل البرامج الحاسوبية المعجم الإلكتروني، والمدقق الإملائي، والمصحح النحوي، وذاكرة الترجمة... إلخ، وتشمل قواعد البيانات عدداً كبيراً من القواعد اللغوية، والقواميس الثنائية اللغة، وقواميس التعبيرات الاصطلاحية، ويقوم برنامج الترجمة الآلية بتحليل النص والتوصل إلى تمثيل انتقالي للغة المستهدفة (إبراهيم ومحمد، ٢٠١٢، ص ٦٥).

ولتحسين جودة هذا النوع من الترجمة يُتاح تضمين المصطلحات الخاصة والتعبيرات المستجدة في مكونات قواعد بيانات برنامج الترجمة. ومن المواقع التي كان برنامجها يمثل هذه الطريقة في الترجمة الآلية موقع (عجيب) التابع لشركة العالمية (صخر) (الشريف، ٢٠٠٠، ص ١٧٨).

وهذا النوع من الترجمة مكلف جداً؛ لأنه مبني على جهود بشرية ضخمة، ويحتاج إلى تطوير متواصل ليتابع الجديد في الاستعمال اللغوي.

٢-٢- الترجمة الآلية الإحصائية (SMT):

ظهرت الترجمة الآلية الإحصائية في ثمانينيات القرن العشرين بواسطة مركز أبحاث Thomas J. Watson التابع لشركة (IBM)، ويعتمد هذا النوع من الترجمة على النماذج الإحصائية التي تشتق معاملة التغير من تحليل مجموعات ضخمة من النصوص الثنائية اللغة (Koeheh, 2020, p. 141)

وسميت هذه الترجمة بالإحصائية لأنها ترجح الاحتمال ذا التكرار الأعلى " ويعتمد نظام الترجمة الإحصائية في الأساس على البيانات المخزنة في برنامج الترجمة، وحين يقوم المستخدم بإدخال جملة معينة، فإن البرنامج يقوم بعمل مقاربات للجمل المطابقة أو المشابهة، ويقوم بعرض النتيجة ذات التكرار الأعلى، ومن هنا جاءت تسميتها بالإحصائية" (Koeheh, 2020, p. 141)، فهي الترجمة التي تقوم بها حواسيب ذات قدرات تقنية فائقة، واعتماداً على برامج حاسوبية مؤسسة على الاستفادة من الذكاء الاصطناعي، وعتاد لغوي ضخم يشمل نصوصاً وقواعد لأكثر من لغة، وكلما زاد هذا الدعم البرمجي واللغوي زادت فرصة تحسين نتائج هذا النوع من الترجمة، فجودة هذه الترجمة يتناسب طردياً مع ضخامة حجم المدونة (البواب، ٢٠١٦، ص. ١١٣)؛ لأن الترجمة الإحصائية "تعتمد على التعلم الآلي بتطبيق دوال رياضية على مجموعات ضخمة من نصوص ترجمت من مترجمين بشر؛ حيث يُحسب احتمال أن تكون جملة ما في اللغة الهدف هي ترجمة جملة ما في اللغة المنطلق" (بسو، ٢٠١٧، ص. ٦٤)؛ فليتمكن برنامج الترجمة من تعلم نموذج للترجمة لا بد له من مئات الآلاف من الجمل في لغة ما مع مئات الآلاف من الجمل التي تقابلها في لغة أخرى، ودون أن نخبر البرنامج عن كل كلمة وما يقابلها في اللغة الأخرى؛ فإن البرنامج يقوم بحساب مختلف الاحتمالات، فإذا وردت كلمة ما آلاف المرات في جمل معينة، وظهرت بالعدد نفسه في الجمل المقابلة؛ فإن البرنامج يستنتج ترجمة هذه الكلمة (بسو، ٢٠١٧، ص. ٦٥).

وقد استخدمت الشركات العالمية العملاقة مثل جوجل ومايكروسوفت نظام الترجمة الإحصائي لفترة طويلة قبل انتقالها لنظام الترجمة الآلية العصبية الجديد.

وقد واجهت الترجمة الآلية الإحصائية بعض النقد، وأهمه: أن مستوى الترجمة في نظم الترجمة الإحصائية مرتبط بتوافر نصوص ضخمة، وهذا يصعب توفيره بالنسبة لكل الثنائيات اللغوية، وفي كل الاختصاصات، إضافة إلى كونها عملية جد مكلفة (التركي وسمار، ٢٠١٣، ص. ١٨). فضلاً عن أن فعاليتها لا تظهر بوضوح إلا عند تقارب اللغتين (المنطلق والهدف) في البنية الصرفية والنظام التركيبي على غرار اللغتين: الفرنسية والإنجليزية (التركي وسمار، ٢٠١٣، ص. ١٨).



وقد أدت مثل هذه الإشكالات إلى أن ابتكرت شركة سيستران ٢٠٠٩ طريقة هجينة في الترجمة الآلية، جمعت ما بين مزايا الطريقة القواعدية والطريقة الإحصائية في الترجمة. وتمتاز هذه الطريقة الهجينة بسهولة تعديل الترجمة، وبأن مدوناتها لا تحتاج إلى أن تكون بضخامة مدونات الترجمة المعتمدة على الطريقة الإحصائية، ويضاف إلى ذلك أن تقنيات التعلم الذاتي فيها تسمح بتجريب برنامج الترجمة على أي حقل معرفي معين بغرض الوصول إلى ترجمة ذات جودة عالية (إبراهيم ومحمد، ٢٠١٢، ص ٦١).

٢-٣- الترجمة الآلية العصبية (NMT):

في عام ٢٠١٥ كان أول ظهور لنظام (NMT) في مسابقة عامة للترجمة الآلية، وتم تطوير نظام الترجمة الآلية العصبية من مختبر (LISA) في جامعة مونتريال، وحلت بشكل سريع محل نظام الترجمة الآلية الإحصائية، وأدت إلى ثورة ضخمة في مجال الترجمة الآلية. ويعتمد نظام الترجمة الآلية العصبية على الذكاء الاصطناعي؛ حيث يقوم بإنشاء شبكة عصبية متنامية باستمرار تحاكي الدماغ البشري، ولديه القدرة على تجاوز نقاط ضعف برامج الأنظمة الإحصائية. وهي خطوة متقدمة بعد الترجمة الإحصائية، ومن ثم تُعتبر استمراراً وتطويراً للمقاربات الإحصائية؛ إذ تعتمد على الاستعمال الموسع للشبكات العصبية الاصطناعية (Artificial Neural networks) التي تسمح لنظام الترجمة بتعلم قوِي ودائم (بسو، ٢٠١٧، ص ٦٤). واتجاهها الرئيسي هو استخدام التمثيلات الموجهة (التضمينات، تمثيلات الفضاء المستمر) للكلمات والحالات الداخلية، ويكون هيكل النماذج أبسط من النماذج القائمة على العبارات؛ فلا يوجد نموذج لغة منفصل، ونموذج ترجمة، ونموذج إعادة ترتيب، ولكن نموذج تسلسل واحد فقط يتنبأ بكلمة واحدة في كل مرة. ويقوم هذا النظام بتشفير الجملة المراد ترجمتها إلى متجهة (Vector) ثابتة الطول، ثم تقوم وحدة الترميز بفك تشفيرها لإنتاج الترجمة (Dzmitry et al., 2014, p.25). وبحلول عام ٢٠١٦ قامت شركات جوجل ومايكروسوفت في برامجها (مترجم جوجل، مترجم بينغ، ومترجم سكايب) باعتماد نظام الترجمة العصبية^(٤). وهذا النوع من الترجمة الآلية شائع ومفيد.

وبعد ذلك ظهر برنامج آخر أكثر استفادة من الترجمة العصبية، وهو ChatGPT، وهو روبوت من تطوير شركة (OpenAI)، وقد أطلق بتاريخ ٢٢ نوفمبر ٢٠٢٢، والهدف منه تسهيل عملية نقل المعلومات عن طريق الإجابة عن أي سؤال يطرحه المستخدم، وكذا القيام ببعض الخدمات الأخرى مثل كتابة محتوى وترجمة نصوص وغيرها (بو المرقعة، ٢٠٢٤، ص. ٤١). وقد مثل هذا الاختراع نقطة تحول حقيقية في مجال التعلم والترجمة (Farghal et al., 2025, p.2089)، ولدى برنامج ChatGPT محتوى ضخم من النصوص المتنوعة لغةً وموضوعاً، يُجري عليه عملياته الحسابية، وهذا ما يجعل ترجمته تفوق غيره أحياناً.

ومما سبق ذكره عن أنواع الترجمة الآلية الثلاثة، والمصادر التي أورد الباحث بعض نصوصها، ومن مطالعة بعض المصادر الأخرى^(٥)؛ يستطيع الباحث أن يلخص بعض الفروق الأساسية بينها في الجدول التالي:

جدول (١) مقارنة بين الترجمات الآلية: القواعدية والإحصائية والعصبية

نوع الترجمة وجه المقارنة	الترجمة القواعدية	الترجمة الإحصائية	الترجمة العصبية
آلية العمل	فك اللبس الصرفي والدلالي	التحليل الإحصائي والاحتمالات	شبكات الترميز وفك الترميز
	كلمة بكلمة	كلمة بكلمة / جملة بجملة	جملة بجملة
	الاستفادة من المعنى السياقي	هندسة السمات	تدريب سمات متعددة معاً
استخدام وحدة المعالجة المركزية	أكثر	قليل	أقل
الاعتماد على المحلل الصرفي	تعتمد	لا تعتمد	لا تعتمد
الاشتغال على قواعد	تحتوي على قواعد	لا تحتوي على قواعد	لا تحتوي على قواعد



نوع الترجمة وجه المقارنة	الترجمة الآلية القواعدية	الترجمة الآلية الإحصائية	الترجمة الآلية العصبية
الاعتماد على المعاجم	تعتمد	لا تعتمد	لا تعتمد
الاستفادة من المدونات	أساسية	غير أساسية	غير أساسية
البيانات	أكثر	كثيرة	أقل
فك الترميز	أكثر قدرة؛ لأنها تقوم على فك اللبس الصرفي والدلالي	أقل قدرة؛ لأنها تعتمد على النماذج الأكثر تكراراً	أكثر قدرة؛ لأنها في الأساس تعتمد على شبكات الترميز
الزمن	أطول	أقل	طويل
المساحة على القرص	أكبر	كبيرة	أقل
القدرة على الترجمة	أدق؛ لأنها تقوم على فك اللبس الصرفي والدلالي، والاستفادة من المعنى السياقي	دقيقة نظراً لوجود مفردات في الترجمة	أقل دقة بسبب افتقارها لعملية ترجمة المفردات
أخطاء الصرف والتركيب	أقل؛ لترجمتها كلمة بكلمة	قليلة؛ لأنها تعتمد على ترجمة كلمة بكلمة	أقل بسبب اعتمادها على الترميز وترجمة جملة بجملة
ترجمة المفردات النادرة	أكثر قدرة	كثير	أقل
معالجة استثناءات القواعد	أصعب	أسهل	أسهل
تكلفة تحسين الترجمة	عالية؛ لاعتمادها على العامل البشري في تطويرها بنسبة كبيرة	أقل؛ لأن التكلفة موجهة لتوفير المدونات نفسها وبرامج المعالجة، وليس للعامل البشري	أقل؛ لأن التكلفة موجهة لتوفير المدونات نفسها وبرامج المعالجة، وليس للعامل البشري

٣- تقييم ترجمة ثلاثة برامج آلية حاسوبية لجمل عربية تمثل بعض القضايا والظواهر اللغوية ذات الأهمية والكثيرة الاستعمال

يقوم هذا المبحث بدراسة ترجمة ثلاثة برامج آلية حاسوبية لجمل عربية تمثل بعض القضايا والظواهر اللغوية ذات الأهمية والكثيرة الاستعمال، وذات احتمالية لبس عالية؛ لتقييم تعامل هذه البرامج معها، وهذه الجمل تغطي موضوعات: الاسم، الفعل، الحرف، الضمائر، السوابق، اللواحق، الترادف، المعطوفات، التعبيرات الاصطلاحية، علامات الترقيم، أنماط الجملة، المعنى الحقيقي والمعنى المجازي.

٣-١- الاسم:

الاسم يخضع للعديد من التصنيفات، منها أنه إما مرتجل وإما منقول، وكل منهما إما مشتق وإما جامد (سيبويه، ١٩٨٨، ٢/ ٩٣-٩٦)، وسيتناول الباحث هنا ترجمة الاسم المشتق فحسب؛ نظراً ليسر ترجمة الاسم الجامد، وقلة اللبس فيه.

جدول (٢) ترجمة الاسم المشتق

م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
١	المدرسة ذات أخلاق حميدة	The school has good morals	The school has good manners	The school has good morals	البرامج كلها ترجمت كلمة (المدرسة) بالمؤسسة التعليمية Institution، وهذا الوجه محتمل. والأولى أن يترجح ترجمة (المدرسة) بمعنى المعلمة مؤنث المدرس (The teacher)، وترجم كل من Google Translate و ChatGPT (أخلاق) بـ (morals)، وترجمها Microsoft Translator بـ (manners).



الملاحظات	ChatGPT	Microsoft Translator	Google Translate	المدخلات	م
البرامج كلها ترجمت كلمة (المدرسة) بالمؤسسة التعليمية. والصواب أنها مؤنث المدرس أي المعلمة؛ لأن المؤسسة التعليمية لا تلبس ثوباً قصيراً. واعتبر ChatGPT و Translate Google المدخل جملة بسيطة، لكن ChatGPT نجح في ربط المضاف بالمضاف إليه The school's dress، أما Mic-rosoft Translator المدخل جملة مركبة، وهذا هو الصواب، لكن ترجم كلمة (المدرسة) نكرة (School).	The school's dress is short.	School her dress is short.	The school dress is short	المدرسة ثوبها قصير	٢
البرامج كلها ترجمت كلمة (المدرسة) بالمؤسسة التعليمية. والصواب أنها مؤنث المدرس أي المعلمة؛ لأن المؤسسة التعليمية لا توصف بأن لها شعراً طويلاً. و Google Translate وأخطأ لأنه لم يربط بين (The school) و (long hair) بفعل، وهذا ما نجح فيه كل من: Microsoft Translator و ChatGPT.	The school has long hair.	The school has long hair.	The school long hair	المدرسة شعرها طويل	٣

م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
٤	المدرسة ابتسامتها جميلة	The school has a beautiful smile	The school has a beautiful smile	The school's smile is beautiful	البرامج كلها ترجمت كلمة (المدرسة) بالمؤسسة التعليمية. والصواب أنها مؤنث المدرس أي المعلمة؛ لأن المؤسسة التعليمية لا يقال عنها: ابتسامتها جميلة. والبرنامجان: Google Translate و-Microsoft Translator كلاهما تعامل مع (ابتسامتها) على أنها نكرة (-a beautiful smile)، في حين اعتبر برنامج ChatGPT الجملة بسيطة، ومن ثم ربط بين المدرسة والابتسام The school's smile، والصواب أنها جملة مركبة.

الاستنتاج:

- البرامج كلها ترجمت كلمة (المدرسة) في الجمل كلها بالمؤسسة التعليمية (The school)، والصواب أنها (Humane) مؤنث المدرس أي المعلمة؛ لأن المؤسسة التعليمية لا توصف بقصر الثوب، ولا بطول الشعر، ولا بأن ابتسامتها جميلة، باستثناء الجملة الأولى (المدرسة ذات أخلاق حميدة)؛ فهي يحتمل فيها أن تكون صفة للمؤسسة التعليمية، وإن كان الأرجح أن تكون (Humane). ويرجع هذا الخطأ إلى أن البرامج تميل إلى ترجمة (المدرسة) بالمؤسسة التعليمية (The school)؛ نظراً لأن النصوص المدخلة إليها، وتمثل قاعدة البيانات BASE DATA، يغلب عليها ذلك.



- Google Translate أخطأ في ترجمة الجملة الثانية؛ إذ اعتبر المدخل جملة واحدة بسيطة، فترجمته ترجمة لجملة: (ثوب المدرسة قصير)، وعند إدخال هذه الجملة لترجمتها جاءت: (school dress short)، وهذا خطأ، والصواب هي الترجمة التي جاءت مع الجملة رقم (٢) (المدرسة ثوبها قصير) التي ترجمها بـ (The school dress is short).
- في الجملة الثالثة لم يربط Google Translate بين (The school) و (long hair) بفعل، وهذا ما نجح فيه كل من Microsoft Translator و ChatGPT، ولكن ترجمتهما تصلح ترجمة للجملة: (المدرسة لها شعر طويل)، ولما تم إدخال هذه الجملة لترجمتها ترجمها Google Translate بـ (school has long hair)، فاستطاع البرنامج أن يربط بين أركان الجملة، وإن كان قد ترجم (المدرسة) نكرة (school)، وتبدأ بحرف (s) صغيرة، أما Microsoft Translator فترجمها بـ (School has long hair)، ولا فرق بينه وبين Google Translate إلا في كون حرف الـ (S) كبيراً، أما ChatGPT فترجمها بـ (The school has long hair)، فعرف كلمة (المدرسة)، وهو الصواب.

وهذا الخطأ في فهم دلالة بعض الألفاظ، اللفاظ المترادفات يعالجه وضع أنطولوجيا لمعاني المفردات العربية بحيث توضع الألفاظ القريبة المعاني تحت مفهوم واحد، ويوضع تحت المفهوم الرئيس المفاهيم الفرعية المنتمية إليه.

والأنطولوجيا: هي تصنيف لمعاني الكلمات، لا الكلمات، بحيث تكون هذه المعاني المتقاربة تحت مفهوم واحد يوصف بحدوده وصفاته الجوهرية، ويوضع تحت المفهوم الرئيس المفاهيم الفرعية المنتمية إليه. والفرق بين الأنطولوجيا والمعاجم هو أن الأنطولوجيا ترتب المعاني بناء على ما وصلت إليه العلوم، وليس على أساس شيوع الدلالة كما هو في المعاجم^(١).

والأنطولوجيا تفيد في الترجمة الآلية لأنها تصنيف دلالي شجري لمعاني الألفاظ، وقد وُضعت للغة الإنجليزية أنطولوجيا متقدمة؛ إذ إن "الأنطولوجية باللغة الإنجليزية غنية بالعلاقات والمفاهيم أكثر من اللغة العربية بكثير" (كلاس، ٢٠١٢، ص ١٧٤). ولتوضيح أثر وضع أنطولوجيا لمعاني الألفاظ العربية في الترجمة الآلية نتبع هذا المثال:

كلمة (باب) لها عدد من المعاني، منها:

١. جزء من المبني.

٢. فتحة.

٣. جزء من مؤلّف.

بين: (باب) بمعنى جزء من المبني و(باب) بمعنى فتحة ما يسمى بالغموض المتباين، وبين: (باب) بمعنى جزء من المبني و(باب) بمعنى جزء من مؤلّف ما يسمى بالغموض غير المتباين أو تعدد المعاني المتكاملة؛ فكل منهما يقال عنه: فتح، واللبس الدلالي بسبب هذا النوع أكبر من اللبس بسبب الغموض المتباين؛ لأن السياق مع الغموض المتباين يكون كاشفاً للمقصود. ومن ثم لو اعتمد برنامج الترجمة الآلي الحاسوبي على أنطولوجيا دقيقة لمعاني الألفاظ العربية؛ فسنضمن أن تسير الترجمة في طرق آمنة النتائج، ففي مثال (فتح الباب) سينحصر الاحتمال بين الباب بمعنى جزء من المبني والباب بمعنى جزء من مؤلّف، ولن يدخل ضمن الاحتمالات (باب) بمعنى فتحة (كلاس، ٢٠١٢، ص. ١٣٨). وقد نادت عدة دراسات بذلك (خضر، ٢٠٠١، ص. ٦).

ويساعد كذلك على تقليل اللبس في الأسماء أن يُبنى المحلل الصرفي على قواعد دقيقة تساعد على فك اللبس الصرفي قبل الترجمة، ومن ذلك: استعمال الياء المتطرفة ذات النقطتين أسفلها (ما يسمى بالياء الشامية) لأنها تساعد في تقليل صور اللبس الصرفي، ومن ثم اللبس الدلالي، ففي المثال: هذا هدى، كلمة (هدى) بهذا الرسم يُحتمل أن يحلها المحلل الصرفي عدة تحليلات، أهمها: هَدَى: فعل ماضٍ، وهُدَى: مصدر الفعل الماضي (هَدَى)، وهُدَى (فعل ماضٍ مبني للمجهول)، وهُدَى: اسم (ما يُساق إلى البيت الحرام من الأنعام للتصدّق بلحمه بعد ذبحه)، ولو غُذي المحلل الصرفي بقاعدة: (الياء تكتب بنقطتين أسفل)، لاخُصرت صور اللبس الصرفي فيها إلى النصف، ولن يكون لها إلا تحليلان، هما: هَدَى: فعل ماضٍ، وهُدَى: مصدر الفعل الماضي (هَدَى)، وستستبعد صورتان الأخريان لوجود الياء في آخرهما. ومن ثم فعرض النصوص على المحلل الصرفي قبل الترجمة الآلية الحاسوبية، أيّا كان نوع الترجمة الآلية، سيساعد في تقليل اللبس.



٣-٢- الفعل:

الفعل، منه المتعدي، ومنه اللازم، والمتعدي منه ما يتعدى بنفسه، ومنه ما يتعدى بأداة (ابن مالك، ١٩٩٠، ٢ / ١٤٨)، وأيسرها في المعالجة الحاسوبية الأفعال ذات المكملات المعجمية، وهي الأفعال التي ينص المعجم على المذكور بعدها من مفعول أو مجرور، وسأقتصر عليها ليسرها في المعالجة:

ومن الأفعال ذات المكملات المعجمية الفعل (رغب)، ومكمّله المعجمي يأتي مجروراً بـ (في)؛ فيكون معناه: أحبه وأقبل عليه، وقد يأتي مجروراً بـ (عن)؛ فيكون معناه: كرهه وأعرض عنه، وقد يأتي مكملته المعجمي منصوباً؛ فيحتمل معناه الأمرين السابقين، وقد ترجمه البرنامجان الآليان بهذه الطريقة:

جدول (٣) ترجمة الأفعال ذات المكملات المعجمية

م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
١	رغب في العمل	Desired to work.	He wanted to work.	He wished to work.	برنامج Google Translate ترجم (رغب) Desired. أما برنامج Microsoft Translator فترجمها (wanted)، ولم يفرق البرنامجان بين (رغب في) و (رغب عن)، فجاءت الترجمة واحدة رغم التضاد في المعنى، وترجم برنامج ChatGPT رغب في بـ (wished to) والجملة فعلية، وترجم رغب عن بـ (was unwilling)، فاعتبر الجملة اسمية.
٢	رغب عن العمل	Desire to work.	He wanted to work.	He was unwilling to work.	

الملاحظات	ChatGPT	Microsoft Translator	Google Translate	المدخلات	م
لما صار الفاعل اسماً ظاهراً (الرجل) ترجم البرنامج: Microsoft و Google Translate	The man wanted to work.	The man wanted to work.	The man wanted to work.	رغب الرجل في العمل.	٣
Translator (رغب) فعلاً، لكن لم يزل البرنامج يخطئان في أنهما لم يفرقا بين كل من (رغب في / رغب عن)، أما برنامج ChatGPT ففرق بينهما؛ فكان مثل البرنامجين الأولين في ترجمة (رغب في)، ولكن خالفهما في ترجمة (رغب عن)، فاعتبر الجملة اسمية.	The man was unwilling to work.	The man wanted to work.	The man wanted to work.	رغب الرجل عن العمل.	٤
ولما صار الفاعل علماً لم يزل البرنامج الأولان يخطئان في أنهما لم يفرقا بين كل من (رغب في / رغب عن)، أما برنامج ChatGPT ففرق بينهما؛ فكان مثل البرنامجين الأولين في ترجمة (رغب في)، ولكن خالفهما في ترجمة (رغب عن)، فاعتبر الجملة اسمية.	Muhammad wanted to work.	Muhammad wanted to work.	Muhammad wanted to work.	رغب محمد في العمل.	٥
أما برنامج ChatGPT ففرق بينهما؛ فكان مثل البرنامجين الأولين في ترجمة (رغب في)، ولكن خالفهما في ترجمة (رغب عن)، فاعتبر الجملة اسمية.	Muhammad was unwilling to work.	Muhammad wanted to work.	Muhammad wanted to work.	رغب محمد عن العمل.	٦



م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
٧	رغب في لعب دور الأستاذ	He wanted to play the role of a teacher	He wanted to play the role of the professor.	He wanted to play the role of the teacher.	لما طالت الجملة المدخلة ترجم برنامج Google Trans- وبرنامج Microsoft Translator (رغب في) و (رغب عن) بـ (wanted)، سواء كان الفاعل ضميراً مستتراً أو اسماً ظاهراً أو علماً، وفي الجملة التي وقع الفاعل فيها ضميراً مستتراً أو اسماً ظاهراً لم يترجم برنامج Microsoft Translator كلمة (دور)، وترجمها في الجملة التي وقع الفاعل فيها اسماً ظاهراً (علماً) (Muham-) ChatGPT (mad). أما برنامج Google Translate فتصرف مثل برنامج Google Translate في ترجمة (رغب في)، ولكن خالفه في ترجمة (رغب عن)، وخالفه أيضاً في أنه عدّها جملة اسمية.
٨	رغب عن لعب دور الأستاذ	He wanted to play the role of a teacher	He wanted to play the role of the professor.	He was unwilling to play the role of the teacher.	وترجم برنامج Google Translate كلمة (الأستاذ) بـ (a teacher) في الجملة التي فيها ضمير مستتراً أو علم، وترجمها (the professor) في الجملة التي فيها اسم ظاهر (الرجل)، وترجمها برنامج Microsoft Translator بـ (the professor) في كل الجمل، وترجمها ChatGPT بـ (the teacher) في كل الجمل.
٩	رغب الرجل عن لعب دور الأستاذ	The man wanted to play the professor's role	The man wanted to play the professor.	The man was unwilling to play the role of the teacher.	
١٠	رغب الرجل في لعب دور الأستاذ	The man wanted to play the role of the professor	The man wanted to play the professor.	The man wanted to play the role of the teacher.	

م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
١١	رغب محمد في لعب دور الأستاذ	Muhammad wanted to play the role of a teacher	Muhammad wanted to play the role of professor.	Muhammad wanted to play the role of the teacher.	وترجم برنامج Microsoft Translator (دور الأستاذ) بـ (professor) إذا كان الفاعل معلماً، وفيما عدا ذلك لم يترجم كلمة (دور).
١٢	رغب محمد عن لعب دور الأستاذ	Muhammad wanted to play the role of the professor	Mohammed wanted to play the role of professor	Muhammad was unwilling to play the role of the teacher.	وترجم برنامج ChatGPT تركيب (دور الأستاذ) بـ (role of the teacher).

الاستنتاج:

بالنظر إلى ترجمة هذه الجمل تتضح أمور، أهمها:

- جاءت ترجمة كل من برنامج Google Translate وبرنامج Microsoft Translator للفاعل (رغب) ذي المكمل المعجمي (في) نفس ترجمة الفعل (رغب) ذي المكمل المعجمي (عن) رغم التضاد في المعنى الذي يحدثه بينهما حرف الجر، ولعل ذلك يرجع إلى أن البرنامجين لم يوفر لهما معجم حاسوبي، يضمن إمدادهما بالأفعال موضعاً أنها ذات مكمل معجمي، في حين أن برنامج ChatGPT فرق في الترجمة، ولكن في ترجمة جملة (رغب عن) عدها جملة اسمية.
- ليس لدى برنامج Google Translate مادة كافية تخدم استنتاج التعبيرات الإضافية؛ بدليل أنه أورد عدة ترجمات لتعبير إضافي واحد (دور الأستاذ)؛ فقد ترجمه بـ (role of a teacher) في جملة: رغب محمد في لعب دور الأستاذ، وترجمه بـ (role of the professor) في جملة: رغب محمد عن لعب دور الأستاذ، ولا خلاف بين الجملتين إلا في حرف الجر المتعدي به الفعل (رغب)، والأمر نفسه في برنامج Microsoft Translator؛ لأنه لم يترجم



كلمة (دور) في كل الجمل إلا في الجملتين اللتين ذكر فيهما العلم (محمد)، في حين أن برنامج ChatGPT لم يقع في هذا الخطأ.

- طول الجملة يؤثر على دقة الترجمة بدليل أن برنامج Microsoft Translator ترجم تركيب (دور الأستاذ) بـ (the role of professor) إذا كان الفاعل علماً، وفيما عدا ذلك لم يترجم كلمة (دور)، وكذلك قد يؤثر السياق في الترجمة.

٣-٣- الحرف:

مصطلح الحرف يشمل أصنافاً شتى، فمن حيث النوع، يتفرع إلى حروف المباني وحروف المعاني، والأخيرة قسمان: عامل، وغير عامل؛ فالعامل هو ما أثر فيما دخل عليه رفعاً، أو نصباً، أو جرّاً، أو جزماً، وغير العامل بخلافه، ويسمى المهمل (المرادي، ١٩٩٢، ص. ٢٧). والجدول التالي يوضح التعامل في الترجمة مع حروف الجر العاملة:

جدول (٤) ترجمة حروف الجر العاملة

م	المداخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
١	جلس في السيارة	sat in the car	He sat in the car	He sat in the car	فَرَّق برنامج Google Translate بين ترجمة حرف الجر (في) وحرف الجر (على)، ولم يفرق برنامج Microsoft Translator، وهذا خطأ، فضلاً عن أن
٢	جلس على السيارة	sat on the car	He sat in the car	He sat on the car	وهذا خطأ، فضلاً عن أن
٣	جلس على الكرسي	sat on the chair	He sat in the chair	He sat on the chair	Google Translate لم يترجم ضمير الغائب، أما برنامج ChatGPT ففرق بين الحرفين في الترجمة، وترجم ضمير الغائب.

الاستنتاج:

البرنامج الذي فرق في الترجمة بين (جلس في) و(جلس على) استطاع أن يوظف معلومات المعجم في فك اللبس، ومن ثم قدّم ترجمة صحيحة؛ فالفاعل (جلس) يصطحب

الحرف الذي يتعدى به معجمياً؛ ولذلك فالأولى الربط بين الأفعال والحروف التي تتعدى بها لأنها تؤثر في توجيه المعنى والترجمة، أما ترجمة حرف الجر وحده فستؤدي إلى خطأ في فهم المعنى، ومن ثم الخطأ في الترجمة.

٣-٤- الضمائر:

تنقسم الضمائر عدة تقسيمات، فهي إما بارزة، وإما مستترة، ومن حيث الشخوص تنقسم إلى تكلم وخطاب وغيبة (ابن جني، ١٩٧٢، ١ / ١٩٩ - ٢٠٤)، والجدول الآتي يوضح ترجمة ضمير الغيبة المتصل، واختاره الباحث دون التكلم والخطاب ليُسر معالجته؛ إذ التفريق بين المذكر والمؤنث يكون بصيغة الضمير، للمذكر (هـ) وللمؤنث (ها)، ومن ثم يسهل إدراك الفرق، لا بالحركات كما في التكلم والخطاب، فيقال: لَعِبْتُ وَلَعِبْتُ وَلَعِبْتُ، فيبحث البرنامج المعالج على قرينة ترجيح من غير الفعل مما حوله من السياق، وهذا فيه عسر.

جدول (٥) ترجمة ضمير الغيبة

م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
١	فكان ما أراد	It was what he wanted	It was what he wanted	So he got what he wanted	لم يفرق برنامج Google Translate وبرنامج Microsoft Translator الترجمة بين الجملة المذكور فيها الضمير (أراد) والجملة التي لم يذكر فيها الضمير (أراد)، في حين فرق بينهما برنامج ChatGPT، لكنه لم يترجم ضمير المفعول.
٢	فكان ما أراد	It was what he wanted	It was what he wanted	So he got what he desired	لم يفرق برنامج Google Translate وبرنامج Microsoft Translator الترجمة بين الجملة المذكور فيها الضمير (أراد) والجملة التي لم يذكر فيها الضمير (أراد)، في حين فرق بينهما برنامج ChatGPT، لكنه لم يترجم ضمير المفعول.
٣	فعل كل ما أراد	He did whatever he wanted	He did everything he wanted	He did everything he wanted	ترجم برنامج Google Translate (كل ما) بـ (whatever)، وترجمها برنامج Microsoft Translator وبرنامج ChatGPT بـ (everything).
٤	فعل كل ما أراد	He did whatever he wanted	He did everything he wanted	He did everything he desired	ولم تترجم البرامج الثلاثة ضمير المفعول، وهو الهاء في: (أراد).



م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
٥	فعل كل ما أراد منه	He did whatever he wanted	He did everything he wanted from him	He did everything he wanted from him	لم يترجم برنامج Google Translate ضمير المفعول المذكور في: (أراد)، ولا الضمير ولا حرف الجر الداخل عليه: (منه). وبرنامج Microsoft Translator وبرنامج ChatGPT لم يترجما ضمير المفعول المذكور في: (أراد)، وترجما الضمير وحرف الجر الداخل عليه: (منه).
٦	فعل كل ما أراد منه أستاده	He did whatever his teacher wanted him to do	He did everything his teacher wanted from him	He did everything his teacher wanted from him	برنامج Google Translate لم يترجم حرف الجر (من) ولا ضمير الجر المتصل به، وترجم الضمير المتصل بالفعل، لكنه عده عائداً على العاقل، وهذا خطأ؛ لأنه يعود على (ما)، وهي لغير العاقل. ولم يترجم برنامج Microsoft Translator وبرنامج ChatGPT ضمير المفعول المذكور في: (أراد).

الاستنتاج:

- توجد صعوبة في ترجمة البرامج الثلاثة لضمائر الغيبة، فمقارنة الجملة الأولى والثانية لا نجد فرقاً بينهما في الترجمة مع أن الجملة الثانية تنتهي بضمير المفعول المذكور في: (أراد)، ولا يوجد هذا الضمير في الجملة الأولى (أراد). وكذلك الأمر في الجملتين الثالثة والرابعة.
- وفي الجملة الخامسة أضيف إلى الجملة الرابعة ضمير مجرور (منه)؛ فترجمه برنامج Microsoft Translator وبرنامج ChatGPT، ولم يترجمه برنامج Google Translate، فتساوت عنده ترجمة الجمل الثلاث: فعل كل ما أراد، وفعل كل ما أراد، وفعل كل ما أراد منه.

- فلما زادت الجملة كلمة موضحة، وأصبحت: فعل كل ما أراد منه أستاذه؛ اختلف البرنامجان في ترجمة الضمير، فترجم برنامج Google Translate الضمير المتصل المذكور في (أراد)، لكن جعله عائداً على الشخص، وليس على الشيء المطلوب فعله، وترجم كل من برنامج Microsoft Translator وبرنامج ChatGPT الضمير المتصل بـ(منه)، ولم يترجم ضمير المفعول في (أراد). وهذا الخلط في التعامل مع الضمير في هذه الجمل دليل على أن البرامج الثلاثة في حاجة إلى ضبطه من حيث التعرف على ما يعود إليه.

٣-٥- السوابق (Prefix):

هي جزء من كلمة يضاف في بدايتها؛ كحرف المضارعة في اللغة العربية (عمر، ٢٠٠٨، ٢/ ١٠٢٩)، أو هي "كل زائدة تضاف في أول جذر الكلمة" (مبارك، ١٩٩٥، ص. ٢٢)، واللغة العربية لغة اشتقاقية، ويبني فيها الاشتقاق في أكثره على التصرف، وفي أقله على الإلصاق "فاللغة العربية، كما يوجد بها مظاهر من أسلوب التصرف والتحليل كما تقدم، يوجد بها مظاهر كثيرة من الأسلوبين الآخرين؛ فهي تسير على طريقة اللصق بالحروف (اللاحقة) و(السابقة) في حالات كثيرة؛ كجمع المذكر السالم وجمع المؤنث السالم والتعدي بالهمزة (قائم قائمون، زينب زينبات، قام علي، وأقام علي الصلاة) ... وهلم جرا" (وافي، ٢٠٠٤، ص. ١١٨).

جدول (٦) ترجمة السوابق

م	المداخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
١	هذه حالة اللاوعي	This is a state of unconsciousness	This is unconsciousness	This is a state of unconsciousness.	برنامج Google Translate لم يفرق في الترجمة بين المعرف (اللاوعي) وغير المعرف (لاوعي). وبرنامج Microsoft



م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
٢	هذه حالة لاوعي	This is a state of unconsciousness	This is an unconscious state.	This is a state of non-awareness.	Translator في الجملة الأولى اعتبرها نكرة، ولم يترجم كلمة (حالة)، وفي الجملة الثانية ترجم كلمة (حالة)، لكن ترجم تعبير (لاوعي) بـ (فاقد الوعي)، وهو خطأ، أما برنامج ChatGPT فترجم كلمة (حالة)، لكن في الجملة الأولى ترجم تعبير (لاوعي) بـ (فاقد الوعي).

الاستنتاج:

- عدم تفريق برنامج Google Translate بين التعبير المعروف (اللاوعي) والتعبير النكرة (لاوعي) في الترجمة دليل على قصور في معرفة السوابق. وبرنامج Microsoft Translator عدّ التعبير المعروف (اللاوعي) نكرة، وفي الجملة الثانية ترجم التعبير (لاوعي) بـ (unconscious)؛ أي فاقد الوعي، وهذا دليل على أن عنده قصوراً في التعامل مع السوابق، أما برنامج ChatGPT فدخول التعريف جعله يخطئ ويفرق بين الكلمتين في الترجمة.

٣-٦- اللواحق (Suffix):

هي ما يضاف من الحروف إلى آخر الكلمة لاشتقاق كلمة أخرى (عمر، ٢٠٠٨، ٣/١٩٩٩)، وسيُظهر الجدول الآتي تعامل البرامج الثلاثة مع اللواحق:

جدول (٧) ترجمة اللواحق

م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
١	اشتري كتابه	bought his book	He bought his book.	He bought his book.	برنامج Google Translate لم يترجم ضمير الغيبة المستتر، وترجم الضمير البارز، وبدأ الجملة بحرف صغير (b).
٢	اشتري كتابها	bought her book	He bought her book.	She bought her book.	برنامج Microsoft Translator ترجم ضمير الغيبة المستتر، وترجم الضمير البارز، وترجمته صحيحة. أما برنامج ChatGPT ففي الجملة الأولى تعرف على الضمائر، وترجمها، أما في الجملة الثانية فأخطأ في ترجمة الضمير المستتر؛ إذ عدّه ضميراً مؤنثاً، وتعرف على الضمير البارز.
٣	اشتري كتابهم	bought their book	He bought their book.	He bought their book.	برنامج Google Translate لم يترجم ضمير الغيبة المستتر، وترجمه برنامج Microsoft Translator وبرنامج ChatGPT، ولم تفرق البرامج الثلاثة بين (كتابهم) و(كتابهن) في الترجمة.
٤	اشتري كتابهن	bought their book	He bought their book.	He bought their book.	وبرنامج Google Translate بدأ الجملة بحرف صغير (b).

الاستنتاج:

- لم تفرق البرامج الثلاثة بين ضمير جمع الذكور في (كتابهم) وضمير جمع الإناث في (كتابهن)؛ فجاءت الترجمة واحدة، وهذا خطأ في ترجمة اللاحقة؛ ولذلك يحسن معالجة النصوص معالجة حاسوبية قبل ترجمتها ترجمة آلية؛ لأنها تسهم في نجاح ترجمتها "اعتماد المعالجة اللغوية العميقة كمرحلة تسبق عملية إخضاع النصوص



المدخلة للمعالجة ضمن محرك الترجمة من شأنه أن يحسّن جودة النصوص المترجمة، وقد تأكدنا من ذلك عملياً" (التركي وسمار، ٢٠١٣، ص. ٢٤)، ويقترح الباحث أن يسبق الترجمة الآلية عدد من عمليات المعالجة الحاسوبية، أهمها:

- استعمال برنامج المدقق الإملائي الآلي؛ للتأكد من الصحة اللغوية للنص قبل ترجمته؛ فجملة: (لا يشك أحد في علمه)، إذا قام المدقق الإملائي بالتأكد من دقة النص؛ هل كلمة: (علمة) تُقرأ: (علمة) (بالتاء المربوطة اسم مرة)، أو تُقرأ (علمه) (الفعل الماضي المتصل بضمير المفرد الغائب أو المصدر المضاف لضمير المفرد الغائب؟) فسيرجح كون الكلمة (علمه)، وأنه حدث خطأ في وضع التاء المربوطة في هذا الموضع، ويفصل في ذلك المدقق الإملائي؛ لأنه مبني على محلل صرفي، ومسجل فيه أن بعض الجذور لا يصاغ منها اسم المرة، ومن ثم لن يكون إلا احتمال واحد هو (علمه)، (الفعل الماضي المتصل بضمير المفرد الغائب أو المصدر المضاف لضمير المفرد الغائب)؛ فلا يخطئ برنامج الترجمة في البحث عن ترجمة لها. والأمر كذلك في كلمة مثل: (فهم) إذا تغير شكلها إلى (فهمه)، فلا شك أن هذه الهاء هي ضمير الغائب المذكور، ولا يوجد احتمال آخر يجعلها تاء سقطت نقطتها خطأ؛ لأنه ليس في العربية (فهمة)، ومن ثم ليست من مخزون المحلل الصرفي.

فالتاء المربوطة لا تدخل إلا على المصادر المدركة بالحس دون غيرها "هذه التاء الدالة على المرة الواحدة لا تدخل على كل مصدر، بل على المصادر الصادرة عن الجوارح المدركة بالحس، نحو: قومة وضربة وقعدة وأكلة. وأما مصادر الأفعال الباطنية والخصال الجبلية الثابتة، نحو: الظرف والحس والجبن والعلم والجهل، فلا يقال من ذلك: علمته علمة، ولا فهمته فهمة، ولا صبرته صبرة" (يس، د.ت، ٢٠٧٧).

٣-٧- الترادف:

هو أن تؤدي أكثر من كلمة المعنى نفسه أو المعنى الأساس دون ظلال المعنى، وإثبات الترادف في العربية مسألة خلافية، وحاول البعض نفيه (السيوطي، ١٩٩٨، ١/ ٣١٤ - ٣١٩؛ عمر، ٢٠٠٦، ص. ٣٢٩). والجدول التالي يوضح طريقة ترجمة برامج الترجمة الآلية للمترادفات.

جدول (٨) ترجمة الكلمات المترادفة

م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملحوظات
١	البيت واسع	The house is wide	The house is spacious.	The house is spacious.	برنامج Google Translate ترجم كلمة (واسع) بـ (wide)، وترجمها برنامج Microsoft Translator بـ (spacious). وبرنامج ChatGPT بـ (spacious).
٢	البيت فسيح	The house is spacious	The house is spacious.	The house is spacious.	اتفقت البرامج الثلاثة في الترجمة.
٣	البيت واسع الغرف	The house is spacious	The house is spacious.	The house has spacious rooms.	اتفق البرنامجان: Google Translate و Microsoft Translator في الترجمة، لكن لم يترجما كلمة (الغرف). في حين ترجمها برنامج ChatGPT.
٤	البيت فسيح الغرف	The house has spacious rooms	The house is spacious.	The house has spacious rooms.	اتفق البرنامجان: Google Translate و ChatGPT في الترجمة، ولكن برنامج Microsoft Translator لم يترجم كلمة (الغرف).

الاستنتاج:

بالنظر إلى الجملة الأولى والثانية يتضح اختلاف البرامج الثلاثة في ترجمة كلمة (واسع)، واتفقا في ترجمة كلمة (فسيح)، رغم أنهما مترادفتان، ونجد أن برنامج Microsoft Translator ترجم جمل: (البيت واسع) و (البيت فسيح) و (البيت واسع الغرف) و (البيت فسيح الغرف) الترجمة نفسها رغم زيادة كلمات في بعضها. لكن برنامج Google Translate فرق بينها، وهذا هو الصواب، ولكنه لم يترجم كلمة (الغرف) في جملة: (البيت واسع الغرف)، وترجمها في جملة: (البيت فسيح الغرف)، أما برنامج ChatGPT فاختلف مع برنامج Google Translate في ترجمة الجملة الأولى والثالثة، و اختلف مع برنامج Microsoft Translator في ترجمة الجمل الثالثة والرابعة، وهذا يدل على أن هذه البرامج فيها نقص يمنع من فهم الترادف



اللغوي، وأنه يوجد نقص في معاجم التعبيرات الشائعة الاستخدام المضافة إلى قاعدة البيانات التي تعمل على أساسها.

٨-٣- المعطوفات:

تطول الجملة في العربية بعدة وسائل، منها استعمال التوابع، وهي النعت، والتوكيد، والبدل، والعطف (ابن السراج، ١٩٩٦، ٢ / ٥٤ - ٥٦). والجدول التالي يوضح أثر المعطوفات على الترجمة الآلية، واختار الباحث العطف دون غيره؛ لأنه الأيسر في المعالجة.

جدول (٩) ترجمة المعطوفات

م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
١	البيت فسيح الغرف واسع الفناء	The house is spacious and has a large patio	The house is spacious and spacious.	The house has spacious rooms a wide courtyard	لما طالت الجملة لم يترجم برنامج Google Translate كلمة (الغرف)، وعدّها جملتين، واستعمل (is) مع (فسيح)، في حين استعمل (has) مع (واسع). أما برنامج Microsoft Translator فترجم (فسيح الغرف) و(واسع الفناء) بـ (spacious)، واستعمل العطف بين الصفتين، وعدّها جملة واحدة، ولم يحدد في الجملة إفعالاً واحداً هو (is). أما برنامج ChatGPT فعّدّها جملة واحدة، ولكن لم يترجم حرف العطف.

م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
٢	البيت فسيح الغرف وواسع الفناء وشكله جميل	The house is spcious, the rooms are spacious and the courtyard is beautiful	The house is spacious and spacious and beautiful.	The house has spacious rooms, a large courtyard, and a beautiful appearance	برنامج Google Translate عدّ وصف (فسيح) للبيت نفسه، وللغرف أيضاً، وعدّ (جميل) صفة للفناء، وليس (واسع)، ولم يصف الشكل بأنه جميل. وبرنامج Microsoft Translator عدّ جميع الصفات وصفاً للبيت، وكرر وصفه بـ (spacious)، ولم يترجم: (فسيح الغرف) و (واسع الفناء) و (شكله جميل). أما برنامج ChatGPT فعدها جملة واحدة، وترجم المعطوفات كلها، لكن لم يترجم ضمير الغائب في (شكله).

الاستنتاج:

بالنظر إلى ترجمة الجملة الأولى يتضح أن تعدد المعطوفات أثر على دقة ترجمة الجملة؛ إذ إن برنامج Google Translate ترجم جملة (البيت فسيح الغرف) بـ (The house has spacious rooms)، فلما طالت الجملة عن طريق العطف وأصبحت: (البيت فسيح الغرف وواسع الفناء) ترجمها بـ (The house is spacious and has a large patio)؛ فلم يترجم كلمة (الغرف)، ووصف البيت بـ (is spacious)، واستعمل (is) مع (فسيح)، في حين استعمل (has) مع (واسع). فلما ازدادت الجملة طولاً وصارت: (البيت فسيح الغرف وواسع الفناء وشكله جميل)؛ ازداد الخطأ في الترجمة، وترجمها بـ (The house is spacious, the rooms are spacious and the courtyard is beautiful)، وفي ذلك العديد من الأخطاء كما ذكرت في الملاحظات في الجدول.



وترجم برنامج Microsoft Translator المعطوفين: (فسيح الغرف وواسع الفناء) ترجمة واحدة، هي: (spacious)، واستعمل العطف بين الصفتين، ولم يذكر فعلاً مع الصفة الثانية، فلما طالت الجملة وصارت: (البيت فسيح الغرف وواسع الفناء وشكله جميل)؛ ازداد الخطأ في الترجمة، وترجمها بـ (The house is spacious and spacious and beautiful)، وفيها أخطاء كثيرة كما ورد في الملاحظات في الجدول. أما برنامج ChatGPT فترجم الجملة كلها، ولكنه لم يترجم ضمير الغائب في (شكله).

وهذا الخلط الشديد في الترجمة من أثر طول الجملة عن طريق تعدد المعطوفات، فضلاً عن تعدد الصفات، وكثافة العلاقات النحوية المتداخلة.

٩-٣- التعبيرات الاصطلاحية:

التعبير الاصطلاحي: نمط تعبيرى خاص بلغة ما، يتميز بالثبات، ويتكون من كلمة أو أكثر، تحولت عن معناها الحرفي إلى معنى مغاير اصططلحت عليه الجماعة اللغوية (حسام الدين، ١٩٨٥، ص. ٢٨). والجدول التالي يوضح تعامل برامج الترجمة الآلية الثلاثة مع التعبيرات الاصطلاحية:

جدول (١٠) ترجمة التعبيرات الاصطلاحية

م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
١	ستأتي فاطمة على جناح السرعة	Fatima will come quickly	Fatima will come quickly.	Fatima will come as quickly as possible	عد البرنامج الأولان (على جناح السرعة) تعبيراً اصطلاحياً، وترجماه بـ (quickly). أما برنامج ChatGPT فترجمه بـ (as quickly as possible).
٢	فاطمة ستأتي على جناح السرعة	Fatima will come quickly	Fatima will come quickly.	Fatima will come as quickly as possible	

مثال آخر:

جدول (١١) تابع ترجمة التعبيرات الاصطلاحية

م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
١	ذرفت عينه دموع التماسيح	His eye shed crocodile tears.	His eye shed crocodile tears.	He shed crocodile tears	ترجم كل من البرنامجين الأولين (دموع التماسيح) تعبيراً اصطلاحياً. أما برنامج ChatGPT فترجم (دموع التماسيح) تعبيراً اصطلاحياً، ولكنه لم يترجم كلمة (عين).
٢	أحد زملائي قضى نخبه أمس	One of my colleagues passed away yesterday	One of my colleagues died yesterday.	One of my colleagues passed away yesterday	البرامج الثلاثة ترجمت التعبير الاصطلاحي (قضى نخبه)، ولكن برنامج Google Translate وبرنامج ChatGPT كانت دقة تعبيرهما عنه أعلى، أما برنامج Microsoft Translator فعبّر عنه تعبيراً مباشراً (died).

الاستنتاج:

تغذية البرنامج المترجم بالتعبيرات الاصطلاحية تضمن جودة الترجمة وسرعتها؛ لأنه يتعامل مع التعبير كله كأنه كلمة واحدة، ويجدها مفكوكة اللبس الصرفي والدلالي، فإذا لم يكن البرنامج عنده قوائم بالتعبيرات الاصطلاحية تعامل معها على أنها كلمات مفردة، ومن ثم يحدث الخطأ في الترجمة والبطء؛ لأن معنى التعبير الاصطلاحي يختلف عن معاني المفردات المكوّن منها "وقد يعود نجاح الآلة في ترجمة هذه المتلازمات اللفظية لكون مقابلات هذه المتلازمات مخزنة في ذاكرة الآلة برفقة ترجمتها" (طالبي، ٢٠٠٨، ص. ١٨٠).



فالتعبير الاصطلاحي (على جناح السرعة) ترجمته البرامج كلها ترجمة صحيحة، وكذلك التعبير الاصطلاحي (دموع التماسيح)؛ لأن التعبيرين الاصطلاحيين مدخلان إلى البرنامج المترجم وحدة واحدة، أما التعبير الاصطلاحي (قضى نجبه) فاختلفت درجة الدقة في ترجمته؛ فكان في القمة برنامج Google Translate وبرنامج ChatGPT، أما برنامج Microsoft Translator فعبر عنه تعبيراً مباشراً (died)؛ وذلك لأن معنى التعبير الاصطلاحي يكون غالباً غير معنى المفردات المكونة له، ومن ثم تظهر صعوبة ترجمته "فعلى الرغم من امتلاك ChatGPT ميزة فهم السياق؛ فإن دقته تقل عند مواجهة معانٍ سياقية أكثر تعقيداً، فغالباً ما تحتوي النصوص العربية على استعارات وتعابير اصطلاحية تتطلب فهماً ثقافياً عميقاً" (راشد وآخرون، ٢٠٢٦، ص. ١٧٥).

ويلاحظ أن بعض التعبيرات الاصطلاحية قد يدخل عليها بعض التغيرات اللغوية، وتبقى محتفظة بأنها تعبيرات اصطلاحية، مثل: "Make snow on him بمعنى: (يشوه سمعته)، يمكن أن يأتي في صورة He made snow on him، ويمكن أن يأتي في صورة: Don't Make snow" (الحمادي، ٢٠١٢، ص. ٣٢). وهذا - بالطبع - ليس مطرداً؛ فبعض التعبيرات الاصطلاحية تلزم صورة واحدة، كتعبير: (قضى نجبه) الذي سبقت الإشارة إليه.

٣-١٠- علامات الترقيم:

هي علامات اصطلاحية توضع في أثناء الكتابة، كالفاصلة والنقطة وعلامتي الاستفهام والتعجب وغيرها (عمر، ٢٠٠٨، ٩٣٠ / ٢)، وفائدتها تعيين مواضع الفصل والوقف والابتداء، وأنواع النبرات الصوتية، والأغراض الكلامية أثناء القراءة.

فهي علامات تستخدم في تقريب النص المكتوب من المنطوق ليسهل فهمه؛ لأنها تنقل إلى حد ما الإيحاءات التي تبدو على الناطق حين نطقه بالكلام، ومن ثم فلها دور في فهم المعنى، وإذا أدركها المترجم الآلي فستكون مرشدة له في اختيار الترجمة الصحيحة، والجدول التالي يوضح ذلك.

جدول (١٢) ترجمة علامات الترقيم

م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملحوظات
١	أعطى المدرس هدية للطالب	The teacher gave a gift to the student.	He gave the teacher a gift to the student.	The teacher gave a gift to the student.	برنامج Google Translate وبرنامج ChatGPT ترجما الجملة الترجمة نفسها في حالة وضع علامة الترقيم (.) أو عدم وضعها.
٢	أعطى المدرس هدية للطالب.	The teacher gave a gift to the student.	The teacher gave a gift to the student.	The teacher gave a gift to the student.	أما برنامج Microsoft Translator فقد أخطأ في ترجمة الجملة الأولى؛ حيث عدّ الفاعل ضميراً مستتراً، والمدرس مفعولاً به أولاً، وهدية مفعولاً به ثانياً. أما الجملة الثانية فترجمها ترجمة صحيحة بسبب وجود علامة الترقيم، وهي النقطة (.) في نهاية الجملة.
٣	أعطى المدرس الطالب هدية	The teacher gave the student a gift	The teacher gave the student a gift.	The teacher gave the student a gift.	لم يحدث خطأ من أي برنامج في ترجمة الجملتين رغم عدم وضع علامة الترقيم (.) في إحدهما.
٤	أعطى المدرس الطالب هدية.	The teacher gave the student a gift.	The teacher gave the student a gift.	The teacher gave the student a gift.	

الاستنتاج:

بالنظر إلى ترجمة الجملة الأولى يتضح أن برنامج Google Translate وبرنامج ChatGPT قد ترجما الجملة المدخلة بدون علامة الترقيم (.) الترجمة نفسها في حالة وضع علامة الترقيم. أما برنامج Microsoft Translator فعدم وضع علامة الترقيم (.) في الجملة الأولى تسبب في



الخطأ في الترجمة؛ حيث عدّ الفاعل ضميراً، و(المدرس) مفعولاً به أول، وعدّ هدية مفعولاً
ثانياً، وهذا خطأ.

وبالنظر إلى ترجمة الجملة الثانية نجد أن البرامج الثلاثة اتفقت في الترجمة، وهي على
صواب، والذي غير ترجمة برنامج Microsoft Translator من الخطأ إلى الصواب هو وضع
علامة الترقيم (.)؛ لأنه لا يوجد فرق بين الجملتين إلا هي.

وبالنظر إلى ترجمة الجملة الثالثة والجملة الرابعة يتضح اتفاق البرامج الثلاثة في الترجمة،
رغم عدم وضع علامة الترقيم في الجملة الثالثة، وهذا لا يستنتج منه عدم أهمية علامة
الترقيم - كما ثبت عكسه في التعليق على الجملتين السابقتين - بل يرجع ذلك إلى أن أجزاء
الجملة المدخلة واضحة، وكل منها قائم بذاته، ومرتببة ترتيباً طبيعياً، وليس فيها ضمائر.

أما الجملتان الأولى والثانية فدخل على (الطالب) حرف الجر اللام (للتألب)، وجاء
في آخر الجملة. وهذا هو سبب ظهور دور علامة الترقيم في الجملتين الأولى والثانية، وعدم
ظهوره في الجملتين الثالثة والرابعة، فعلمة الترقيم يظهر دورها عند اختلاف نمط الجملة
عن النمط البسيط المباشر.

مثال آخر:

جدول (١٣) تابع ترجمة علامات الترقيم

م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
١	من فهم النصوص الأدبية	Understanding literary texts	From understanding literary texts	Understanding literary texts.	برنامج Google Translate وبرنامج ChatGPT لم يترجما كلمة (من). برنامج Microsoft Translator ترجم كلمة (من) حرف جر.

م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
٢	من فهم النصوص الأدبية؟	Who understands literary texts?	Who understands literary texts?	Who understood the literary texts?	برنامج Google Translate وبرنامج Microsoft Translator اتفقا في الترجمة، واستفادا من علامة الترقيم في فهم الجملة، ولكن عدّا الفعل (فهم) مضارعاً. أما برنامج ChatGPT فقد ترجم الجملة ترجمة صحيحة، وترجم (فهم) فعلاً ماضياً.

الاستنتاج:

أخطأ برنامج Google Translate وبرنامج ChatGPT في ترجمة الجملة الأولى؛ إذ لم يترجما كلمة (من)، ومن ثم كانت الترجمة ناقصة. أما برنامج Microsoft Translator فترجم كلمة (من) حرف جر، ومن ثم ترجم الجملة ترجمة كاملة، وهي صواب، ولكن الأولى أن يترجم (من) اسماً موصولاً؛ كي لا تكون الجملة ناقصة، أو يضطر إلى تقدير مبتدأ محذوف (هذا من فهم النصوص الأدبية) أو ما شابه.

وفي الجملة الثانية دخلت علامة الترقيم (؟)، واستطاع برنامج Google Translate وبرنامج Microsoft Translator الاستفادة من علامة الترقيم في فهم الجملة، ولكن عدّا الفعل (فهم) مضارعاً، أما برنامج ChatGPT فقد ترجم الجملة ترجمة صحيحة، وترجم (فهم) فعلاً ماضياً، وهذا هو الصواب.

٣- ١١- أنماط الجملة:

أنماط الجملة في العربية غيرها في الإنجليزية، وهذه صعوبة تعترض الترجمة الآلية من إحدى اللغتين إلى الأخرى، والجدول التالي يوضح ذلك:



جدول (١٤) ترجمة أنماط الجملة

م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
١	يحب الرجل عمله	The man loves his work	A man loves his job.	The man loves his work.	لم يختلف نمط الجملة المترجمة رغم اختلاف نمط الجملة العربية المدخلة. وترجم برنامج Google Translate وبرنامج ChatGPT
٢	الرجل يحب عمله	The man loves his work	A man loves his job.	The man loves his work.	كلمة (الرجل) بـ (The man) و(عمل) بـ (work)، وترجمهما برنامج Micro-soft Translator بـ (A man) و (job) (٧).

الاستنتاج:

يوجد اختلاف في أنماط الجمل في كل من اللغتين العربية والإنجليزية، ففي العربية قد تبدأ الجملة بفعل، وتسمى جملة فعلية، وقد تبدأ باسم، وتسمى جملة اسمية، في حين أن الجملة الإنجليزية لا تبدأ إلا باسم؛ ولذا جاءت ترجمة الجملتين واحدة، ولم يراع أي من البرامج الثلاثة نمط الجملة في العربية في الترجمة، ومن ثم يحسن العمل على حصر أنماط الجملة في العربية، ومعرفة الفروق بينها وبين أنماط الجملة في بعض اللغات الأجنبية المراد الترجمة منها أو إليها، وتحديد سبل تفادي ما يترتب على ذلك من أخطاء، والتنبيه إلى أن بعض ظلال المعنى لا تعكسها الترجمة من لغة لأخرى "فاختلاف البنية اللغوية والأسلوبية بين اللغتين يجعل عملية الترجمة بينهما صعبة، وتتطلب مهارة فائقة" (أويس، ٢٠٢٤، ص. ١٩٨).

٣-١٢- المعنى الحقيقي والمعنى المجازي:

المعنى الحقيقي هو المعنى الأصلي المتبادر من اللفظ عند استعماله، أما المعنى المجازي فهو المعنى غير الأصلي للفظ "وأما المجاز فكل كلمة أريد بها غير ما وقعت له في وضعها، ملاحظه بين الثاني والأول" (الجرجاني، ١٩٩١، ص. ٣٥١)، وسنتعرف على قدرة برامج الترجمة الثلاثة على ترجمة النصوص المجازية من الجدول التالي:

جدول (١٥) المعنى الحقيقي والمعنى المجازي

م	المدخلات	Google Translate	Microsoft Translator	ChatGPT	الملاحظات
١	يأكل الرجل طعاماً	man eats food	A man eats food.	The man is eating food.	برنامج Google Translate ترجم (الرجل) بـ (man)، وترجمه برنامج Microsoft Translator بـ (A man)، وترجمه برنامج ChatGPT بـ (The man).
٢	يأكل الرجل حقوق الناس	The man eats people's rights	A man eats people's rights.	The man is consuming people's rights.	برنامج Google Translate ترجم (الرجل) بـ (The man)، وترجمها برنامج Microsoft Translator بـ (A man). وترجم كلا البرنامجين تعبير (يأكل حقوق الناس) بـ (eats people's rights)، وترجم تعبير (يأكل أموال الناس) بـ (eats people's money)، والبرنامجان كلاهما مخطئ؛ لأنهما اعتبرا المعنى الحقيقياً. أما برنامج ChatGPT فقد اختار المعنى المجازي للجملتين، وترجمهما.
٣	يأكل الرجل أموال الناس	The man eats people's money	A man eats people's money.	The man is taking/ consuming people's money.	

الاستنتاج:

أدخلت ثلاث جمل إلى البرامج الثلاثة؛ الجملة الأولى استعملت ألفاظها في المعنى الحقيقي: (يأكل الرجل طعاماً)؛ فكانت الترجمة صحيحة إلا في اعتبار الفعل مستمراً أو غير مستمر. أما الجملة الثانية فاستعملت ألفاظها استعمالاً مجازياً: (يأكل الرجل حقوق الناس)، فترجم برنامج Google Translate وبرنامج Microsoft Translator الفعل (يأكل) بـ (eats)، على أن المعنى



حقيقي، في حين ترجمه برنامج ChatGPT ترجمة مجازية، وهو الصواب. واختيرت جملة
ثالثة صريحة في استعمال ألفاظها استعمالاً مجازياً: (يأكل الرجل أموال الناس)؛ فكانت
ترجمة برنامج Google Translate وبرنامج Microsoft Translator غير دقيقة كذلك؛ لأنها اعتبرت
التعبير حقيقياً، أما برنامج ChatGPT فترجم الجملة ترجمة مجازية، ومن ثم فالبرنامجان
الأولان يضعف مستوى ترجمتهما عندما يكون التعبير مجازياً. ولو كان أي من البرنامجين
معتمداً على قاعدة بيانات ثرية بالسياقات المجازية، مع مراجعة مكملات الفعل في قاعدة
البيانات؛ لأدرك أن الاستعمال مجازي.

٤- الخاتمة

وفيها أهم نتائج البحث وتوصياته:

٤-١- نتائج البحث:

بعد ما عرض الباحث من مدارس وتطبيقات واستنتاجات يسجل أهم النتائج العلمية لهذا البحث؛ علماً أنها نتاج معالجة مجموعة من الجمل هي عينة الدراسة، وعددها ٥٢ جملة، غيرما اقتضت الدراسة الاستطراد إليه من معالجة جمل أخرى للتثبت من نتائج التعامل مع العينة الأساسية، ومن النتائج:

أ. رصد الباحث بعض الأخطاء التي وقع فيها كل من البرامج الثلاثة محل الدراسة، وبين أسبابها، وطريقة معالجتها، وأهمها:

- النقص في الذخائر اللغوية التي هي أساس من الأسس التي يعمل البرنامج المترجم بالاعتماد عليها.
- الاعتماد على الترجمة الحرفية.
- القصور في تحديد المجال الدلالي للنص المترجم.
- عدم القدرة على التعامل مع الفروق التركيبية بين اللغتين العربية والإنجليزية؛ كنمط الجملة، وارتباط حروف الجر، والتطابق في النوع والعدد، والتعريف والتكثير... إلخ.

ب. أثبت الباحث في غيرما موطن من الدراسة أن برامج الترجمة الآلية المعاصرة عندها أخطاء، وأن علاجها ليس في التماذي في السير في طريقها الحالي، والاستكثار من النصوص والمدونات والذخائر اللغوية التي تعتمد عليها؛ لأن الإضافة المتحققة من هذا الاستكثار قليلة، ولن تبلغ بهذه البرامج إلى مستوى الدقة المطلوبة، وإنما يحسن البحث عن حلول خارج المسار المعتاد لهذه البرامج، وقد ذكر الباحث طرق هذا الإصلاح وسبله إن لغوياً أو تقنياً، ومن ذلك:

- معالجة النصوص معالجة حاسوبية قبل الشروع في ترجمتها ترجمة آلية؛ لأن ذلك يساهم في نجاح ترجمتها، ومنه:



- استعمال برنامج المدقق الإملائي الآلي؛ للتأكد من الصحة اللغوية للنص قبل ترجمته.
- الاعتماد على محلل صرفي قادر على فك اللبس الصرفي بجودة عالية؛ إذ إنه يختصر صور اللبس، وفي ذلك ضمان لنجاح المعالجة الحاسوبية، ومن ثم الترجمة الآلية.
- مراجعة المعجم الحاسوبي قبل الدخول في تحليل السياقات اللغوية.
- وضع أنطولوجيا لمعاني الألفاظ العربية يساعد في فك اللبس الدلالي للألفاظ، وتحديد المعنى العام للسياق، ومن ثم يُيسر الوصول للترجمة الآلية الصحيحة.
- ج. أوضح الباحث أن برامج الترجمة الآلية يختلف مستوى جودة ترجمتها بحسب المدخلات؛ فأعلى نسبة جودة حققتها كانت في ترجمة المفردات، ثم التعبيرات الاصطلاحية، ثم الأفعال ذات المكملات المعجمية، ثم جاءت بعد ذلك الجمل، وكانت نسبة الجودة فيها أعلى مع الجمل القصيرة التي لا تحتوي على إحالة، ثم الجمل القصيرة التي فيها إحالة، ثم الجمل الطويلة.
- د. أكد الباحث أن ترجمة الجمل المجازية يكثف فيها الخطأ عن ترجمة الجمل ذات المعنى الحقيقي، وبين البحث طريقة التخفيف من ذلك.
- هـ. ما يزال طريق الوصول إلى ترجمة آلية تدرك المعنى السياقي، وتختار من معاني المفردات ما يناسبه؛ تحفه بعض الصعوبات، وإن كان قد تحقق قسط كبير من ذلك في بعض اللغات الأخرى كاللغة الإنجليزية؛ فإن اللغة العربية لم يزل بينها وبين تحقيق ذلك أمدٌ لأنها لغة واسعة؛ كثيرة الألفاظ، ثرة المعاني، متعددة القواعد؛ فضلاً عن أن الجهود المبذولة لتحقيق ذلك تُعدّ ضئيلة.

٤-٢- التوصيات:

- يرى الباحث أهمية قصوى لإيلاء متابعة البحث في مجال الترجمة الآلية جهوداً إضافية؛ لتذليل العقبات التي تمنع من جودة هذه الترجمة، خاصة مجال بناء أنطولوجيا لمعاني الألفاظ العربية.

- الحرص على المراجعة البشرية لما تُرجم عن طريق برامج الترجمة الآلية، خاصة إذا كانت في المجالات التي تتطلب الدقة؛ كالعلم بمجالاته، والتعاملات المالية، والأحكام القضائية، وما شابه.



الهوامش

- (١) يعمل الباحث في مجال حوسبة العربية منذ عام ١٩٩٤م، وشارك من خلال العمل في شركة (صخر) وغيرها من الشركات المتخصصة في حوسبة اللغة العربية في العديد من المشروعات، ومنها: المحلل النحوي، المحلل الصرفي، الترجمة الآلية، إعداد معجم حاسوبي، الكنز العربي، المصحح الآلي، المعرب الآلي، وتُوج ذلك بمحصلته على درجة الدكتوراه في حوسبة العربية بإشراف الدكتور نبيل علي مؤسس شركة (صخر) طيب الله ثراه، وله عدد من البحوث في المجال نفسه.
- (٢) نص البحث على الفترة التي أجري فيها التطبيق على البرامج الثلاثة لأنه قد يحدث تطوير لهما؛ فتختلف النتائج.
- (٣) تعامل الباحث مع مئات الجمل ليصل إلى تحديد الفروق بين ترجمة برامج الترجمة محل البحث، وتحديد الخطأ، وسببه، واقتراح علاج له، وما سجله من أنه قام على ٥٢ جملة كان للاختصار والتمثيل فقط؛ لمحدودية المساحة المسموح بها للبحث في مثل هذه المجالات العلمية، ومراجعة فقرة الاستنتاج على ترجمة الاسم المشتق -مثلاً- نموذج على ما كان يقوم به الباحث من تجريب على العديد من الجمل في القضية الواحدة.
- (٤) أوردت دراسة: الترجمة الآلية من منظور اللسانيات الحاسوبية دراسة تحليلية مقارنة بين تطبيقي: Google Translate and Microsoft Translator أنه "تعتمد ترجمة جوجل على أسلوب يسمى بالترجمة الآلية الإحصائية" (الأسمرى، ٢٠١٨، ص. ٤٢)، وهذا خطأ؛ إذ تحولت ترجمة جوجل من الترجمة بنظام سيستران إلى الترجمة الإحصائية سنة ٢٠٠٧م (إبراهيم ومحمد، ٢٠١٢، ص. ٦٩)، وتحولت من الترجمة الإحصائية إلى الترجمة العصبية منذ عام ٢٠١٦م كما أوضحت الدراسة الحالية، وانظر: (بورياحي وهباشي، ٢٠٢١، ص. ٢٨٦).
- (٥) راجع: <https://researcha.net/types-of-translation-methods>
- (٦) <https://ontology.birzeit.edu/term>
- (٧) ملحوظة: يسمح السياق بصحة كل من ترجمة البرنامجين لكلمة (عمل)؛ فقد يقصد بها: الوظيفة (job)، أو الدور والمهمة (work)، وقد يسمح بذلك في ترجمة (الرجل) إذا قصد بها جنس الرجال عامةً.

المراجع العربية

- إبراهيم، مجدي حاج، ومحمد، عائشة رابع. (٢٠١٢). الترجمة الآلية بين النظامين الإحصائي والتحويلي: دراسة مقارنة بين جوجل وعجيب. مجلة الدراسات اللغوية والأدبية، ١، ٦٠-٨٢.
- ابن السراج، أبو بكر محمد بن السري بن سهل. (١٩٩٦). الأصول في النحو (عبد الحسين الفتلي، محقق). مؤسسة الرسالة.
- ابن جني، أبو الفتح عثمان. (١٩٧٢). اللمع في العربية (فائز فارس، محقق). دار الكتب الثقافية.
- ابن مالك، محمد بن عبد الله. (١٩٩٠). شرح تسهيل الفوائد (عبد الرحمن السيد، ومحمد بدوي المختون، محققان). هجر للطباعة والنشر والتوزيع والإعلان.
- الأسمرى، عايض محمد. (٢٠١٨). الترجمة الآلية من منظور اللسانيات الحاسوبية: دراسة تحليلية مقارنة بين تطبيقَي Google Translate and Microsoft Translator. المجلة العربية للعلوم ونشر الأبحاث، ٤(٣)، ٤٠.
- أويس، محمد. (٢٠٢٤). دور الذكاء الاصطناعي في تحسين ترجمة النصوص العربية باستخدام أدوات الترجمة المدعومة بالحاسوب: تحديات وتوجيهات. مجلة اللغة، ٩(٢)، ١٨٥-٢٠٨.
- بسو، صديق. (٢٠١٧). اللغة العربية والترجمة الآلية الإحصائية. في الملتقى الوطني حول اللغة العربية والترجمة، ٦١-٦٦.
- البواب، مروان. (٢٠١٦). الترجمة الآلية. التعريب، ٢٦(٥٠)، ١٠٥-١٣٦.
- البواب، مروان. (٢٠١٨). دراسات في تقييم الترجمة الآلية: تقييم المترجم الآلي غوغل نموذجًا. دار الرافين للطباعة والنشر والتوزيع.
- بوارمقة، إيمان. (٢٠٢٤). كيف سيؤثر تشات جي بي تي على الترجمة والمترجمين؟ المجلة الجزائرية لعلوم اللغة، ٨(٢)، ٣٩-٤٧.



بورياحي، سعيد، وهبشي، لطيفة. (٢٠٢١). المعجم الذهني والترجمة الآلية: استثمار للمعارف وتطوير للبرامج، الترجمة الآلية العصبية نموذجاً. مجلة اللسانيات التطبيقية، ٥(٢)، ٣٠٨-٢٨٥.

التركي، وفاء، وسمار، نصر الدين. (٢٠١٣). اختبار أداء نظام الترجمة الآلية الإحصائية Moses المكيف لدعم الثنائية اللغوية إنجليزي-عربي. مجلة المعهد العالي العربي للترجمة، ٢٠(٢)، ٢٦-٧.

الجرجاني، عبد القاهر بن عبد الرحمن بن محمد. (١٩٩١). أسرار البلاغة (محمود محمد شاكر، محقق). مطبعة المدني، دار المدني.

حسام الدين، كريم زكي. (١٩٨٥). التعبير الاصطلاحي: دراسة في تأصيل المصطلح ومفهومه ومجالاته الدلالية وأنماطه التركيبية (ط. ٢). الأنجلو المصرية.

الحمادي، فايزة بنت صالح. (٢٠١٢). الترجمة الآلية إلى اللغة العربية: صعوبات وتحديات، ترجمة غوغل مثلاً. مجلة الدراسات اللغوية والأدبية، ١، ٦٠-٨٣.

خضر، محمد زكي. (٢٠٠١). اللغة العربية والترجمة الآلية. في أعمال مؤتمر التعريب الحادي عشر. المنظمة العربية للتربية والثقافة والعلوم.

راشد، إفكار، وأسفار، حديد، وديناتا، رحمة ساتريا. (٢٠٢٦). تحليل ترجمة النصوص العربية المعتمد على الذكاء الاصطناعي: استخدام ChatGPT أداة مساعدة في الترجمة. مجلة لغة الضاد، ٦(٢)، ١٦٥-١٨١.

سيبويه، عمرو بن عثمان بن قنبر. (١٩٨٨). الكتاب (عبد السلام محمد هارون، محقق، ط. ٣). مكتبة الخانجي.

السيوطي، عبد الرحمن بن أبي بكر. (١٩٩٨). المنهر في علوم اللغة وأنواعها (فؤاد علي منصور، محقق). دار الكتب العلمية.

الشريف، حسن. (٢٠٠٠). تقرير المسح عن الترجمة الآلية في الوطن العربي. مركز دراسات الوحدة العربية.

- الصالح، صبحي إبراهيم. (١٩٦٠). دراسات في فقه اللغة. دار العلم للملايين.
- طالبي، أمانة. (٢٠٠٨). إشكالية حدود الترجمة الآلية: ترجمة نظام (سيستران) للمتلازمات اللفظية (رسالة ماجستير غير منشورة). جامعة منتوري، كلية الآداب واللغات.
- العليمي، يس. (د.ت.). حاشية يس على التصريح. مطبعة الحلبي.
- عمر، أحمد مختار. (٢٠٠٨). معجم اللغة العربية المعاصرة. عالم الكتب.
- عمر، تمام حسان. (٢٠٠٦). اللغة العربية: معناها ومبناها (ط. ٥). عالم الكتب.
- كلاس، محمد طلال. (٢٠١٢). بناء مترجم نصوص آلي دلالي من اللغة العربية إلى اللغة الإنجليزية (رسالة ماجستير). الجامعة الافتراضية السورية.
- مبارك، مبارك. (١٩٩٥). معجم المصطلحات الألسنية. دار الفكر اللبناني.
- المبرد، محمد بن يزيد. (١٩٩٤). المقتضب (محمد عبد الخالق عزيمة، محقق). المجلس الأعلى للشؤون الإسلامية.
- المجالي، عمار إرحيل. (٢٠٠٣). السوابق وأثرها في بنية الكلمة العربية: دراسة صرفية دلالية (رسالة ماجستير غير منشورة). جامعة مؤتة، كلية الآداب.
- المرادي، بدر الدين حسن بن قاسم بن عبد الله. (١٩٩٢). الجنى الداني في حروف المعاني (فخر الدين قباوة، ومحمد نديم فاضل، محققان). دار الكتب العلمية.
- وافي، علي عبد الواحد. (٢٠٠٤). علم اللغة (ط. ٩). نهضة مصر للطباعة والنشر.

رومنة المراجع العربية

Bibliographic information in Roman script

- Ibn al-Sarraj, Abu Bakr Muhammad ibn al-Sari ibn Sahl. (1996). *Al-Usul fi al-Nahw*. (Abd al-Husayn al-Fatli, Ed.). Mu'assasat al-Risalah.
- Ibn Jinni, Abu al-Fath Uthman. (1972). *Al-Luma' fi al-Arabiyyah*. (Fayiz Faris, Ed.). Dar al-Kutub al-Thaqafiyyah.



- Ibn Malik, Muhammad ibn Abd Allah. (1990). *Sharh Tashil al-Fawa'id*. (Abd al-Rahman al-Sayyid & Muhammad Badawi al-Makhtun, Eds.). Hajar Publishing.
- Ibrahim, Majdi Haj, & Muhammad, Aishah Rabi'. (2012). Al-Tarjamah al-Aliyah bayna al-Nizamayn al-Ihsa'i wa al-Tahwili: Dirasah Muqaranah bayna Google wa Ajib. *Majallat al-Dirasat al-Lughawiyyah wa al-Adabiyyah*, 1, 60–82.
- Al-Asmari, Ayid Muhammad. (2018). Al-Tarjamah al-Aliyah min Manzur al-Lisaniyyat al-Hasubiyyah: Dirasah Tahliliyyah Muqaranah bayna Tatbiqay Google Translate and Microsoft Translator. *Al-Majallah al-Arabiyyah lil-Ulum wa Nashr al-Abhath*, 4(3), 40–55.
- Owais, Muhammad. (2024). Dawr al-Dhaka' al-Istina'i fi Tahsin Tarjamat al-Nusus al-Arabiyyah bi Istikhdam Adawat al-Tarjamah al-Mad'umah bil-Hasub. *Majallat al-Lughah*, 9(2), 185–203.
- Bassou, Sadiq. (2017). Al-Lughah al-Arabiyyah wa al-Tarjamah al-Aliyah al-Ihsa'iyyah. In *Al-Multaqa al-Watani Hawl al-Lughah al-Arabiyyah wa al-Tarjamah* (pp. 61–68).
- Al-Bawwab, Marwan. (2016). Al-Tarjamah al-Aliyah. *Al-Ta'rib*, 26(50), 105–136.
- Al-Bawwab, Marwan. (2018). *Dirasat fi Taqyim al-Tarjamah al-Aliyah: Taqyim al-Mutarjim al-Ali "Google" Namudhajan*. Dar al-Rafidain.
- Boulamraqa, Iman. (2024). Kayfa Sayu'aththir ChatGPT Ala al-Tarjamah wa al-Mutarjimin? *Al-Majallah al-Jaza'iriyyah li Ulum al-Lughah*, 8(2), 39–47.
- Bouriahi, Said, & Habbashi, Latifah. (2021). Al-Mu'jam al-Dhihni wa al-Tarjamah al-Aliyah: Istithmar lil-Ma'arif wa Tatwir lil-Baramij. *Majallat al-Lisaniyyat al-Tatbiqiyyah*, 5(2), 285–308.
- Al-Turki, Wafa, & Sammar, Nasr al-Din. (2013). Ikhtibar Ada' Nizam al-Tarjamah al-Aliyah al-Ihsa'iyyah Moses. *Majallat al-Ma'had al-Ali al-Arabi lil-Tarjamah*, 20(2), 7–26.
- Al-Jurjani, Abd al-Qahir ibn Abd al-Rahman. (1991). *Asrar al-Balaghah*. (Mahmud Muhammad Shakir, Ed.). Matba'at al-Madani.
- Husam al-Din, Karim Zaki. (1985). *Al-Ta'bir al-Istilahi: Dirasah fi Ta'sil al-Mustalah wa Ma'fhumih*. Al-Anglo al-Misriyyah.

- Al-Hammadi, Fayizah bint Salih. (2012). Al-Tarjamah al-Aliyah ila al-Lughah al-Arabiyyah: Su'ubat wa Tahaddiyat. *Majallat al-Dirasat al-Lughawiyah wa al-Adabiyyah*, 1, 60–83.
- Khidr, Muhammad Zaki. (2001, October 12–16). Al-Lughah al-Arabiyyah wa al-Tarjamah al-Aliyah. In *Mu'tamar al-Ta'rib al-Hadi Ashar*.
- Rashid, Ifkar, Asfar, Hadid, & Dinata, Rahmah Satria. (2026). Tahlil Tarjamat al-Nusus al-Arabiyyah al-Mu'tamid ala al-Dhaka' al-Istina'i: Istikhdam ChatGPT Adah Musa'idah fi al-Tarjamah. *Majallat Lughat al-Dad*, 6(2), 165–181.
- Sibawayh, Amr ibn Uthman ibn Qanbar. (1988). *Al-Kitab* (3rd ed., Abd al-Salam Muhammad Harun, Ed.). Maktabat al-Khanji.
- Al-Suyuti, Abd al-Rahman ibn Abi Bakr. (1998). *Al-Muzhir fi Ulum al-Lughah wa Anwa'iha*. (Fu'ad Ali Mansur, Ed.). Dar al-Kutub al-Ilmiyyah.
- Al-Sharif, Hasan. (2000). *Taqrir al-Mash al-Tarjamah al-Aliyah fi al-Watan al-Arabi*. Markaz Dirasat al-Wahdah al-Arabiyyah.
- Al-Salih, Subhi Ibrahim. (1960). *Dirasat fi Fiqh al-Lughah*. Dar al-Ilm lil-Malayin.
- Talibi, Aminah. (2008). *Ishkaliyyat Hudud al-Tarjamah al-Aliyah* (Master's thesis, University of Mentouri).
- Al-Alimi, Yasin. (n.d.). *Hashiyat Yasin ala al-Tasrih*. Matba'at al-Halabi.
- Umar, Ahmad Mukhtar. (2008). *Mu'jam al-Lughah al-Arabiyyah al-Mu'asirah*. Alam al-Kutub.
- Hassan, Tammam. (2006). *Al-Lughah al-Arabiyyah: Ma'naha wa Mabnaha* (5th ed.). Alam al-Kutub.
- Kallas, Muhammad Talal. (2012). *Bina' Mutarjim Nusus Ali Dallali min al-Lughah al-Arabiyyah ila al-Lughah al-Injiliziyyah* (Master's thesis). Syrian Virtual University.
- Mubarak, Mubarak. (1995). *Mu'jam al-Mustalahat al-Asuniyyah*. Dar al-Fikr al-Lubnani.
- Al-Mubarrad, Muhammad ibn Yazid. (1994). *Al-Muqtadab*. (Muhammad Abd al-Khaliq Udaymah, Ed.). Al-Majlis al-A'la lil-Shu'un al-Islamiyyah.



- Al-Majali, Ammar Irhayyil. (2003). *Al-Sawabiq wa Atharuha fi Binyat al-Kalimah al-Arabiyyah* (Master's thesis). Mu'tah University.
- Al-Muradi, Badr al-Din Hasan ibn Qasim. (1992). *Al-Jana al-Dani fi Huruf al-Ma'ani*. (Fakhr al-Din Qabawah & Muhammad Nadim Fadil, Eds.). Dar al-Kutub al-Ilmiyyah.
- Wafi, Ali Abd al-Wahid. (2004). *Ilm al-Lughah* (9th ed.). Nahdat Misr.

المراجع الأجنبية

- Dzmitry, Bahdanau; Cho, Kyunghyun; Yoshua, Bengio.(2026). “*Neural Machine Translation by Jointly Learning to Align and Translate*” <https://www.studocu>
- Farghal,Tariq;Shraideh, Khetam; Al-Omari, Ahmad. (2025).Translation Tools between Arabic and English: A Comparative Study of Google Translate, ChatGPT, and Gemini. *International Journal for the Semiotics of Law*, 38(6), 2083 -2110
- Koehen, Philipp. (2020). *Neural machine Translation*. Ed Cambridge University Press. United Kingdom.
- Koehen, Philipp. (2026). *statistical Machine Translation System User Manual And Code Guide*. <https://aclanthology.org>
- Wołk, Krzysztof; Marasek, Krzysztof. (2026). Neural-based Machine Translation for Medical Text Domain based on European Medicines Agency Leaflet Texts. *Procedia Computer Science*. <https://www.sciencedirect.com/>

نبذة تعريفية عن الباحث:

Author Biodata	بيانات الباحث:
Medhat Youssef El-Sabaa is Associate Professor in Arabic Language Computing at the Supreme Council for Islamic Affairs in Egypt. He obtained his PhD in Arabic Language Computing from the Faculty of Dar Al-Uloom at Cairo University in 2004. His research interests focus on Arabic language computing and linguistic thought. He has published 13 peer-reviewed research papers and authored 9 books, and has participated in dozens of academic conferences, either as a presenter or attendee.	مدحت يوسف السبع، أستاذ مشارك، في حوسبة اللغة العربية، المجلس الأعلى للشؤون الإسلامية بمصر. حاصل على درجة الدكتوراه في حوسبة اللغة العربية، من جامعة القاهرة، كلية دار العلوم، عام ٢٠٠٤. تدور اهتماماته البحثية حول حوسبة اللغة العربية والفكر اللغوي؛ فله (١٣) بحثاً علمياً منشوراً، وله (٩) كتب منشورة، وشارك في عشرات المؤتمرات العلمية ببحوث أو حضور.
معرف أوركيد (ORCID) : 0909-3896-0005-0009	
Email: medhataalsabea@gmail.com	



الملحق

الجملة التي تمثل عينة الدراسة

الموضوع	الجملة	م
الاسم	المدرسة ذات أخلاق حميدة	١
	المدرسة ثوبها قصير	٢
	المدرسة شعرها طويل	٣
	المدرسة ابتسامتها جميلة	٤
الفعل	رغب في العمل	٥
	رغب عن العمل	٦
	رغب الرجل في العمل.	٧
	رغب الرجل عن العمل.	٨
	رغب محمد في العمل.	٩
	رغب محمد عن العمل.	١٠
	رغب في لعب دور الأستاذ	١١
	رغب عن لعب دور الأستاذ	١٢
	رغب الرجل عن لعب دور الأستاذ	١٣
	رغب الرجل في لعب دور الأستاذ	١٤
	رغب محمد في لعب دور الأستاذ	١٥
	رغب محمد عن لعب دور الأستاذ	١٦
الحرف	جلس في السيارة	١٧
	جلس على السيارة	١٨
	جلس على الكرسي	١٩

الموضوع	الجملة	م
الضمائر	فكان ما أراد	٢٠
	فكان ما أراده	٢١
	فعل كل ما أراد	٢٢
	فعل كل ما أراده	٢٣
	فعل كل ما أراده منه	٢٤
	فعل كل ما أراده منه أستاذة	٢٥
السوابق	هذه حالة اللاوعي	٢٦
	هذه حالة لاوعي	٢٧
اللواحق	اشترى كتابه	٢٨
	اشترى كتابها	٢٩
	اشترى كتابهم	٣٠
	اشترى كتابهن	٣١
الترادف	البيت واسع	٣٢
	البيت فسيح	٣٣
	البيت واسع الغرف	٣٤
	البيت فسيح الغرف	٣٥
المعطوفات	البيت فسيح الغرف وواسع الفناء	٣٦
	البيت فسيح الغرف وواسع الفناء وشكله جميل	٣٧
التعبيرات الاصطلاحية	ستأتي فاطمة على جناح السرعة	٣٨
	فاطمة ستأتي على جناح السرعة	٣٩
	ذرفت عينه دموع التماسيح	٤٠
	أحد زملائي قضى نحبه أمس	٤١



الموضوع	الجملة	م
علامات التقييم	أعطى المدرس هدية للطالب	٤٢
	أعطى المدرس هدية للطالب.	٤٣
	أعطى المدرس الطالب هدية	٤٤
	أعطى المدرس الطالب هدية.	٤٥
	من فهم النصوص الأدبية	٤٦
	من فهم النصوص الأدبية؟	٤٧
أنماط الجملة	يحب الرجل عمله	٤٨
	الرجل يحب عمله	٤٩
المعنى الحقيقي والمعنى المجازي	يأكل الرجل طعاما	٥٠
	يأكل الرجل حقوق الناس	٥١
	يأكل الرجل أموال الناس	٥٢

ولله الحمد والمنة.

دور النماذج اللغوية الكبيرة في تطوير (WordNet)

قاعدة بيانات معجمية عربية: تقييم القدرات وتحسين الشبكات الدلالية

المعرف الرقمي: <https://doi.org/10.60161/2521-003-005-003>

محمد مهيب البشري^{١*} جيهان الجراي^٢

أ قسم اللغات والترجمة، كلية الآداب، جامعة طيبة المدينة المنورة، المملكة العربية السعودية

ب قسم اللغة الإنجليزية، كلية التربية، جامعة تعز، اليمن

نُشر في	قبل في	رُوجع في	استقبل في
١٤٤٧/١١/١٣	١٤٤٧/٠٧/٠٢	١٤٤٧/٠٥/٢٩	١٤٤٧/٠٥/١٢

الملخص

تُجري هذه الدراسة تحليلاً دلاليًا للمفردات العربية ضمن إطار لبناء شبكة الكلمات العربية (Arabic WordNet)، باستخدام النماذج اللغوية الكبيرة (ChatGPT و Gemini و DeepSeek و Grok) مع الخبرة البشرية لإنشاء قاعدة بيانات معجمية دقيقة. تركز الدراسة على معاني الكلمات والعلاقات الدلالية والتعقيدات الصرفية. تعتمد المنهجية على جمع البيانات من المعاجم وقواعد البيانات اللغوية، وتحليل النماذج اللغوية، وتقييم الخبراء اللغويين للتأكد من دقتها في تصنيف المترادفات والمتضادات، وبناء العلاقات الجزئية. أظهرت النتائج قصوراً في تجميع المرادفات وتصنيف المتضادات، لكنها أبرزت قوة برنامج Gemini في التصور، و ChatGPT في الاستخدام السياقي، و Grok و DeepSeek في اللغويات الحاسوبية، مع تفوق Grok في الصرف. تُقدم هذه الدراسة خطوة نوعية نحو بناء شبكة معجمية باستخدام أدوات الذكاء الاصطناعي، مدعمة بالتقييم البشري، مما يُعزز الدقة الدلالية والكفاءة، ويدعم تطبيقات معالجة اللغة الطبيعية.

الكلمات المفتاحية

بيانات معجمية عربية، النماذج اللغوية الكبيرة، الشبكات الدلالية، الذكاء الاصطناعي، شبكة الكلمات العربية.

*الباحث المراسل: محمد البشري (البريد الإلكتروني: mqassem@taibahu.edu.sa).

The Role of Large Language Models (LLMs) in Developing an Arabic WordNet: Assessing Capabilities and Enhancing Semantic Networks

المعرف الرقمي: <https://doi.org/10.60161/2521-003-005-003>

Mohammad Mahyoob Albuhairey^{A*} Jeehaan Algaraady^B

A Languages & Translation Department, Taibah University, Madina, KSA

B Department of English, Taiz University, Taiz, Yemen

Received on	Revised on	Accepted on	Published on
13-11-2025	20-11-2025	22-12-2025	30-04-2026

ABSTRACT

This study conducts a semantic analysis of Arabic vocabulary within the framework of an (Arabic WordNet), using large language models, (LLMs) (ChatGPT, Gemini, DeepSeek, and Grok) combined with human expertise to create an accurate lexical database. It examines word meanings, semantic relationships, and morphological complexities. The data are collected from dictionaries and linguistic databases, analyzed by LLMs, and evaluated by linguistic experts to ensure accuracy in classifying synonyms and antonym, and in identifying distinct relationships. The results revealed the LLMs shortcomings in synonym classification and semantic grouping, however, Gemini showed visualization capabilities, ChatGPT's showed contextual analysis, and Grok and DeepSeek's showed linguistic analysis, with Grok excelling in morphology. This study is a valuable contribution in building lexical network using artificial intelligence tools, with human evaluation, thereby enhancing semantic accuracy and supporting natural language processing applications.

KEYWORDS

Arabic Lexical Data, Large Language Models (LLMs), Semantic Networks, Artificial Intelligence (AI), Arabic WordNet (AWN).

*Corresponding author: Mohammad Albuhairey (Email: mqassem@taibahu.edu.sa).

١- المقدمة

أحدث الظهور والتقدم المتسارع في تقنيات معالجة اللغة الطبيعية ثورة عارمة في مجال الذكاء الاصطناعي، حيث برزت في السنوات الأخيرة النماذج اللغوية الكبيرة (Large Language Models: LLMs) بقدرات فريدة وغير مسبوقة في محاكاة العقل البشري من خلال فهم اللغة البشرية وتوليدها (Abuhairy et al, 2025; Algaraady & Mahyoob, 2025). وقد أظهرت هذه النماذج أداءً عاليًا ومتميزًا في تطبيقات الذكاء الاصطناعي والمهام اللغوية المختلفة (Rogers, et al, 2020; Brown et al., 2020 Abuhairy et al, 2023; Algaraady & Mahyoob, 2023). ومع ذلك، ما تزال فعاليتها في المجالات المتخصصة، وخاصةً في مجال المعاجم وشبكات الكلمات والأنطولوجيا العربية محدودة، وتتطلب معالجةً وتدريبًا خاصًا لإنتاج تحليل لغويٍّ معمّقٍ يتماشى مع خصوصية اللغة العربية.

تُمثل المعاجم وشبكة الكلمات والأنطولوجيا العربية قاعدةً أساسيةً في تطبيقات اللغويات الحاسوبية كأنظمة الترجمة الآلية المتقدمة، ومحركات البحث الدلالية، وأدوات التحليل النصي المتخصصة، وغيرها، ذلك لاستغراقها في العلاقات الدلالية المعقدة (Navigli & Ponzetto, 2012). تتضمن المعاجم عملية بناء القواميس ودراسة معاني الكلمات وعلاقاتها الدلالية وأنماط استخدامها. وقد أنتجت المعاجم العربية التقليدية أعمالاً ضخمة مثل: لسان العرب والمعجم الوسيط، وغيرها الكثير، كمراجع أساسية لدراسات اللغة العربية (Habash, 2010). في حين تمثل شبكة الكلمات مجموعة من الشبكات الدلالية التي تربط بين المفاهيم، وتشكل شبكة الكلمات أنطولوجيا جزئية بحيث يشكل الارتباط بين مجموعة من الشبكات مكوناً أكبر وأشمل يغني العديد من استخدامات الذكاء الاصطناعي وتطبيقات الويب الدلالي، بما يسمى بالأنطولوجيا. في حين تعرف الأنطولوجيا بأنها مجموعة من المعارف التي تترابط مكوناتها مشكلةً هيكلًا ضخمًا من المفاهيم المشتركة في حقول تربطها علاقات دلالية بحيث تشمل مختلف فروع المعرفة للغة (السعيد طه، ٢٠١٢؛ دلول، ٢٠٢٠، 2004 Holes).



إن بناء شبكة لغوية عربية أو أنطولوجيا عربية متكاملة يتطلب فهمًا عميقًا للعلاقات الدلالية بين الكلمات، والقدرة على تحديد المرادفات والأضداد، وإدراك السياقات المختلفة لاستخدام الكلمات، بالإضافة إلى فهم التعقيدات الصرفية والنحوية للغة (Navigli & Ponzetto, 2012). ويمهد بناء شبكة كلمات معيارية الطريق لبناء أنطولوجيا متكاملة بحيث تتوافر الجهود لتغطية المحتوى المعرفي الضخم لا سيما أن هذا المحتوى في تنام مستمر. بيد أن بناء شبكة لغوية يشكل تحديًا أكثر تعقيدًا في حالة اللغة العربية، خاصةً في مجال المعالجة الحاسوبية، ويعود ذلك إلى تعقيد بنيتها النحوية، وعلاقاتها الدلالية المتداخلة والمعقدة، مثل: المعاني المتعددة للكلمة الواحدة المرتبطة بالسياق، الفروق الدلالية الدقيقة بين الكلمات المترادفة، بالإضافة إلى نظامها الصرفي الغني والمعقد الذي يعرض الجذور (التي تحمل المعنى الأساسي) والأنماط، وعمليات الاشتقاق الشاملة من الجذور عبر تطبيق أوزان صرفية مختلفة (Navigli & Ponzetto, 2012; Rogers, et al, 2020). يتطلب هذا التعقيد الدلالي والصرفي من النماذج اللغوية الضخمة قدرةً فائقةً على التمييز بين المعاني المختلفة وفهم السياقات الثقافية والتاريخية التي تؤثر على استخدام الكلمات. علاوة على ذلك، يشكل الغموض الإملائي (بسبب غياب التشكيل في النصوص العادية، والتنوع اللهجي الواسع) تحديًا إضافيًا يجعل من تقييم قدرات النماذج اللغوية في بناء المعجمية العربية مهمة بالغة التعقيد (الحلي وآخرون, ٢٠٢٣; أدغيري, ٢٠٢٢; (Farghaly & Shaalan, 2009; Al-Garaady, 2022; Mahyoob & Algaraady, 2018).

هناك العديد من الجهود المبذولة في هذا المجال. فعلى سبيل المثال تعد شبكة الكلمات العربية (Arabic Word Net) التي أنجزت خلال الفترة ما بين ٢٠٠٥ و٢٠٠٧ وفق تصميم ومكونات شبكة الكلمات المعتمدة دوليًا (Princeton Wordnet) والتي تعد من أشهر الشبكات العربية، غير أن هذه الشبكة لم تقنع وتلبي تطلعات اللغويين العرب، كونها بنيت اعتمادًا على الترجمة من لغة أخرى، دون مراعاة خصوصية وبنية اللغة العربية، وعدم شمولها على كل مفردات اللغة العربية والعلاقات الدلالية المختلفة بينها، وغيرها من الدراسات التي سنستعرضها في القسم التالي. شكّل هذا القصور والفجوة الواسعة في الأدبيات في إنتاج قاعدة بيانات شاملة لشبكة الكلمات العربية منطلقًا أساسيًا للقيام بهذه

الدراسة، بالإضافة إلى أنه بالرغم من التطور والأهمية المتزايدة للنماذج اللغوية الضخمة؛ إلا أن أدائها في بناء وإثراء المعاجم وشبكة الكلمات والأنطولوجيا العربية لم يُقَيِّم بشكل منهجي باستخدام معايير شاملة ومتخصصة. حيث ركزت معظم الدراسات على تطبيقات أخرى، مثل: أنظمة الترجمة الآلية المتقدمة، ومحركات البحث الدلالية، وأدوات التحليل النصي المتخصصة، والإجابة عن الأسئلة، واستدعاء المعلومات، وتحليل المشاعر، وتصنيف وتلخيص النصوص، وغيرها (Jarrar, 2021)

ومن هذا المنطلق، يقوم الباحثان في هذه الدراسة باستغلال إمكانات وتقييم قدرات النماذج اللغوية الضخمة مثل: Gemini، DeepSeek، ChatGPT، وGrok. في بناء شبكة كلمات عربية شاملة تتضمن عدداً من الجوانب، متمثلة في كلمات مُفصلة تجمع بين المرادفات، والأضداد، والأسماء المتشعبة، والأسماء المختصرة، والإحالات المتقاطعة الموسعة، مع التركيز على تعقيدات الصرف العربي والفروق الدلالية الدقيقة وفقاً لمعايير محددة من أجل توجيه جهود التطوير وتحسين أداء هذه النماذج في التعامل مع العربية. ولم يتم اختيار هذه الجوانب عشوائياً، بل جاء بعد دراسة عميقة لمتطلبات بناء الشبكة اللغوية والدراسة بخصائص اللغة العربية. وتكمن أهمية الدراسة الحالية في إثراء البحث المتنامي حول معالجة اللغة العربية من خلال تقديم منهجية ومعايير تقييم لبناء شبكة الكلمات. كما تقدم الدراسة توصيات عملية للباحثين والمطورين من أجل تحسين أداء النماذج اللغوية في بناء الشبكات اللغوية والأنطولوجيا العربية.

٢- مشكلة الدراسة

يواجه تطوير شبكة الكلمات العربية (Arabic WordNet) تحديات كبيرة تتعلق بمعالجة تعقيدات اللغة العربية، خاصة الطبيعة الاشتقاقية للجذور اللغوية والفروق الدلالية الدقيقة بين المفردات. تعتمد العديد من الدراسات السابقة على الترجمة من أنطولوجيات أجنبية أو استخدام طرق إحصائية محدودة النطاق، مما يؤدي إلى قصور في الشمولية والدقة، مثل: التجميع غير الدقيق للمرادفات، والتصنيف الخاطئ للأضداد، والثغرات في العلاقات الجزئية والدلالية. تُظهر النماذج اللغوية الكبيرة Gemini، DeepSeek، ChatGPT،



إمكانات واعدة في بناء شبكة المفردات العربية، لكنها تفتقر إلى التكامل مع الخبرة البشرية لضمان الدقة في معالجة السياقات الثقافية واللغوية العربية. هذا يحدّ من قدرة هذه النماذج على بناء قاعدة بيانات معجمية عربية موثوقة تلبي متطلبات التطبيقات التقنية والتعليمية، مثل: معالجة اللغة الطبيعية، واسترجاع المعلومات، والتعليم اللغوي. بالتالي، هناك حاجة مُلحة لتطوير إطار تحليلي يدمج بين قدرات النماذج اللغوية والخبرة البشرية؛ لتعزيز الدقة الدلالية والنطاق المعجمي للشبكة المعجمية العربية.

٣- أهداف الدراسة

- أ. تهدف هذه الدراسة إلى تقييم أداء النماذج اللغوية ChatGPT، Gemini، DeepSeek، Grok في تحليل العناصر المعجمية العربية تحليلاً دلالياً عميقاً من حيث صياغة مجموعات مترادفة دقيقة، وتحديد الأضداد، والمصطلحات الأعم/الأخص، والعلاقات الجزئية، مع التركيز على معالجة تعقيدات الصرف العربي والفروق الدلالية.
- ب. تهدف الدراسة إلى تحديد نقاط الضعف والقوة في مخرجات النماذج اللغوية، واقتراح تصحيحات مستهدفة لتعزيز موثوقية المعجمية العربية المدعومة بالذكاء الاصطناعي.

٤- أسئلة الدراسة

لتحقيق الأهداف المذكورة، صيغت الأسئلة البحثية الآتية:

- أ. إلى أي مدى تتمكن النماذج اللغوية الكبيرة ChatGPT، Gemini، DeepSeek، Grok من صياغة مجموعات مترادفة دقيقة وشاملة للمعجمية العربية مع مراعاة تعقيدات الصرف العربي والفروق الدلالية الدقيقة؟
- ب. ما العيوب الرئيسية في أداء النماذج اللغوية الكبيرة عند تحليل المعجمية العربية؟ وكيف يمكن تصحيح هذه العيوب لتطوير إطار دلالي أكثر موثوقية في سياق شبكة الكلمات العربية؟

٥- الدراسات السابقة

تعد الدراسات السابقة في مجال الأنطولوجيا والشبكات الدلالية للغة العربية (Arabic WordNet) أساساً لفهم التطورات والمنهجيات المستخدمة في هذا المجال الحيوي. وتهدف هذه الجزئية إلى استعراض أبرز الأبحاث التي تناولت بناء الأنطولوجيات وتوسيع الشبكات الدلالية للغة العربية، مع التركيز على التحديات والحلول المقترحة. ستُحلل هذه الدراسات لتقديم رؤية شاملة حول الجهود المبذولة في هذا السياق، مما يمهّد الطريق للأبحاث المستقبلية.

ركزت العديد من الدراسات على بناء الشبكات الدلالية للغة العربية، والتي تعد مورداً لغوياً مهماً للمعالجة الآلية للغة. على سبيل المثال، هدفت دراسة السعيد طه (٢٠١٢) "شبكة الكلمات العربية: المنهجية والتطبيق". إلى بناء شبكة دلالية للغة العربية على غرار WordNet الإنجليزية، مع مراعاة الخصائص اللغوية الفريدة للغة العربية، ونظامها المعجمي والدلالي على مستوى المباني والمعاني. وقد قدمت الدراسة منهجية مفصلة لإنشاء هذه الشبكة. مسطرة الضوء على أهمية الشبكة في تعزيز الصناعة المعجمية العربية المعاصرة، خاصةً في المعجمات المتخصصة.

قدمت دلول (٢٠٢٠) إطاراً لبناء أنطولوجيا معجمية شاملة للغة العربية، مع الأخذ في الاعتبار التحديات المرتبطة بتعقيد اللغة العربية وتنوعها. كما استعرضت الدراسة أهمية الأنطولوجيات في تمثيل المعرفة اللغوية، وكيف يمكن استخدامها لتحسين أداء تطبيقات معالجة اللغة الطبيعية. وطورت الحلبي وآخرون (٢٠٢٣) إطار تعلم أنطولوجيا عربية باستخدام تقنية التقابل بين شبكة الكلمات العربية وخصائص مقالات الموسوعة العربية، محققة دقة مفاهيم بلغت ٠,٨٨، وعدد مصطلحات ١٢٨٠٠ بدقة ٠,٥٢ للعلاقات الهرمية، و٠,٧٧ لعلاقات الترافق، و٠,٤٠ لعلاقات التشابه. قدمت أدغيري (٢٠٢٣) منهجية مقترحة لبناء وتطوير أنطولوجيا عربية في مجال اللسانيات الحاسوبية. وسلطت هذه الدراسة الضوء على أهمية بناء أنطولوجيات متخصصة للغة العربية في مجالات محددة، مثل: اللسانيات الحاسوبية. وقدمت منهجية مقترحة لبناء هذه الأنطولوجيات، مع التركيز على مراحل التخطيط، والتصميم، والتنفيذ، والتقييم. وتُعد هذه المنهجية إطاراً عملياً يمكن الاستفادة منه في بناء أنطولوجيات عربية دقيقة ومتخصصة.



وفي إطار الجهود المبذولة لبناء أنطولوجيات رسمية للغة العربية، قدم *Jarrar et al*. (٢٠١١-٢٠٢١) دراسة بعنوان "بناء أنطولوجيا عربية رسمية". وتُعد هذه الدراسة من الأعمال الرائدة في مجال بناء الأنطولوجيات العربية، حيث قدمت منهجية شاملة لبناء أنطولوجيا لغوية رسمية للغة العربية باستخدام ١٥٠ معجماً. وركزت الدراسة على تمثيل المفاهيم التي تنقلها المصطلحات العربية باستخدام علاقات دلالية رسمية. وتُعد هذه الأنطولوجيا قابلة للوصول والبحث من خلال محرك بحث معجمي. وقدم *Alsudias L. et al*. (٢٠٢٠) أنطولوجيا طبية متخصصة بالأمراض المعدية بطرق إحصائية من مستندات طبية، مدمجة مع معاجم عربية. تُعد عملية ربط الشبكات الدلالية بمصادر معرفية أخرى خطوة مهمة نحو إثرائها وزيادة فائدتها. في هذا السياق، تناولت دراسة *Saif et al*. (٢٠١٧) ربط مجموعات المرادفات في الشبكة الدلالية العربية بمقالات ويكيبيديا باستخدام ميزات أحادية اللغة وثنائية اللغة. وهدفت هذه الدراسة إلى تطوير طريقة لربط مجموعات المرادفات (synsets) في الشبكة الدلالية العربية بمقالات ويكيبيديا المقابلة لها. واستعملت هذه الطريقة ميزات أحادية اللغة وثنائية اللغة للتغلب على نقص المعلومات الدلالية في الشبكة الدلالية العربية. وأظهرت النتائج التجريبية أن الطريقة المقترحة تحقق نتائج واعدة وتتفوق على الطرق الحديثة التي تعتمد على الميزات أحادية اللغة فقط. وقد استخدمت هذه الطريقة لزيادة تغطية الشبكة الدلالية العربية عن طريق إدراج مجموعات مرادفات جديدة من ويكيبيديا. أنشأ *AlMaayah et al*. (٢٠١٦) شبكة الكلمات العربية القرآنية (QAWN) بتوليد مرادفات آلياً من قواميس عربية، وقدم *Ishkewy et al*. (٢٠١٤) أنطولوجيا الأزهرية (أنطولوجيا معجمية عربية) من نصوص القرآن الكريم، بالاعتماد على ربط يدوي بعلاقات معجمية. وقدمت هذه الدراسة أنطولوجيا معجمية للغة العربية، جمعت الكلمات في مجموعات مرادفات (synsets) وسجلت العلاقات الدلالية مثل: "الترادف، والتضاد، وعلاقات النوع (hypernymy)، وعلاقات الجزء (meronymy)، وعلاقات الكل (holonymy)". وتُعد هذه الأنطولوجيا مورداً قيماً للباحثين في مجال معالجة اللغة الطبيعية العربية، وساهمت في تطوير تطبيقات مثل: استرجاع المعلومات، والترجمة الآلية، وفهم اللغة. اعتمد *Mazari et al*. (٢٠١٢) في دراستهم مفاهيم أنطولوجيا Gold وخصائص نصوص عربية، مستخدمين تقنيات إحصائية لربط المفاهيم بعلاقات هرمية

وغير هرمية. وصمّم AlBarhamtoshy *et al*. (٢٠٠٩) شبكة كلمات عربية عبرترجمة شبكة الكلمات الإنجليزية باستخدام قاموس متعدد اللغات، بينما وسّع Rodríguez (٢٠٠٨) شبكة الكلمات العربية شبه آلياً باستخدام قواعد صرفية وقواعدية مع الاستدلال بايزي. وطوّر Alkhalifa *et al*. (٢٠١٠) أنطولوجيا عربية مترجمة بإضافة أسماء كيانات من ويكيبيديا، واستخدم Aliane *et al*. (٢٠٠٩) أنطولوجيا Gold كنواة لأنطولوجيا عربية مركزية ضمن مشروع الخليل. أنشأ Belkredim *et al*. (٢٠٠٨) أنطولوجيا DEAR-ONTO الاشتقاقية، معتمدين على الأفعال السماعية كجذور لتحديد الأصناف المشتقة.

تُعد عملية استخراج العلاقات الأنطولوجية من النصوص العربية تحديًا مهمًا في مجال معالجة اللغة الطبيعية. في هذا الصدد، قدم Al Zamil & Al-Radaideh (٢٠١٤) دراسة حول "استخراج التلقائي للعلاقات الأنطولوجية من النص العربي". اقترحت هذه الدراسة منهجية لاستخراج العلاقات الأنطولوجية من النص العربي، مع التركيز على الميزات الدلالية للغة العربية. وقدمت نموذجًا رسميًا لاستخراج العلاقات الأنطولوجية، ووضحت كيفية استخدام الأنماط الدلالية المعجمية لتحديد هذه العلاقات. كما أظهرت النتائج التجريبية أن المنهجية المقترحة تحقق نتائج واعدة في أتمتة عملية استخراج العلاقات الأنطولوجية من النصوص العربية. وتُعد الأنطولوجيات المعجمية مكونًا أساسيًا في العديد من تطبيقات معالجة اللغة الطبيعية.

١- منهجية الدراسة

تتبع المنهجية المستخدمة في هذه الدراسة نهجًا تحليليًا وصفيًا لدراسة المفردات العربية ضمن إطار شبكة الكلمات العربية (Arabic WordNet) باستخدام النماذج اللغوية الكبيرة (ChatGPT، Gemini، DeepSeek، Grok)، مع دمج الخبرة البشرية لتعزيز الدقة والشمولية. بدأت الدراسة بجمع بيانات من معاجم عربية موثوقة (مثل: لسان العرب والمكنز الكبير والمعجم الوسيط وغيرها) لتحديد المعاني المتعددة والعلاقات الدلالية (مرادفات، أضداد، مصطلحات أعم/أخص) واللغوية (مشتقات صرفية). كُلفت النماذج اللغوية بتحليل هذه المفردات؛ لإنشاء شبكات معجمية تشمل مجموعات مترادفة (Synsets)، مع التركيز



على معالجة تعقيدات الصرف العربي والفروق الدلالية الدقيقة. أُجري تقييم مخرجات النماذج من خبراء لغويين لقياس دقتها في تجميع المرادفات، وتصنيف الأضداد، والعلاقات الجزئية، مع مراجعة السياقات في نصوص عربية متنوعة لضمان الموثوقية. دُمج التحليل البشري لتصحيح أوجه القصور، مثل: التجميع غير الدقيق والتصنيف الخاطئ وغيرها، كما سيتم ذكرها في الأقسام التالية. وتُحقق من النتائج عبر مقارنة السياقات في النصوص العربية ومراجعة التحليلات؛ لضمان موثوقية الشبكة المعجمية وملاءمتها للتطبيقات التعليمية والحاسوبية.

٧- التحليل والمناقشة

٧-١- النماذج اللغوية وبناء شبكة الكلمات العربية (WordNet Arabic)

تتناول هذه الدراسة مقارنة تحليلية بين أربعة نماذج ChatGPT و Gemini و DeepSeek و Grok في بناء شبكة الكلمات العربية (Arabic WordNet) لبعض الجذور والكلمات العربية على سبيل المثال "كتب"، و"ضرب"، و"علم". حيث طُلب من النماذج الأربعة بناء شبكة دلالية للجذور كمرحلة أولى. وفي المرحلة الثانية أُجريت هندسة للأمر المقدم للنماذج الأربعة بتفاصيل دقيقة تتعلق ببناء شبكة معجمية تفصيلية متكاملة. يهدف التحليل إلى تقييم قدرات كل نموذج في تقديم شبكة معجمية شاملة ودقيقة للمعاني المتعددة، والعلاقات الدلالية، والصرفية لهذه الجذور، مع التركيز على مدى التفصيل، والدقة، والشمولية، ونقاط القوة والضعف لكل نموذج، وتحديد النموذج الأكثر فعالية في تقديم مخرج أفضل ومقارنته بالأداء البشري.

٧-٢- المرحلة الأولى لعملية البناء والتحليل بدون هندسة الأوامر

٧-٢-١- تحليل قدرات النموذج ChatGPT

يقدم نموذج ChatGPT تحليلاً مفصلاً ومنظماً للجذور الثلاثية، مع تقسيم واضح للمعاني المتعددة (Senses) لكل جذر. بالنسبة لـ "كتب"، يحدد النموذج ثلاثة معانٍ رئيسة (الكتابة، التأليف، الفرض / القدر)، مع تقديم مجموعات مرادفات (Synsets) لكل معنى،

مثل: "كتب، دون، سطر" لمعنى الكتابة. كما يوفر أمثلة توضيحية، مثل: "كتب رسالة إلى صديقه"، مما يعزز فهم السياق كما هو موضح في الشكل ١. بالنسبة لـ "ضرب"، يقدم النموذج أربعة معانٍ (الضرب الجسدي، ضرب المثل، الطرق، الضرب الرياضي)، مع مرادفات مثل: "ضرب، لطم، صفع" وأمثلة سياقية. أما بالنسبة لـ "علم"، فإنه يحدد أربعة معانٍ (المعرفة، التدريس، المعرفة كاسم، وضع علامة)، مع مرادفات دقيقة. يتميز هذا النموذج بتقديم هيكلية هرمية واضحة للعلاقات الدلالية (مرادفات، وأمثلة، وتفسيرات)، مما يجعله مناسباً لفهم العلاقات بين الكلمات بسهولة. ومع ذلك، يفتقر النموذج إلى تحديد الأضداد أو العلاقات الأعم/الأخص بصراحة، مما يحد من شموليته مقارنة بنماذج أخرى.

شكل (١) تحليل وبناء شبكة الكلمات للجذر "كتب" باستخدام نموذج ChatGPT





٧-٢-٢- تحليل قدرات النموذج Grok

النموذج Grok يركز على تحليل الجذور الثلاثية (كتب، ضرب، علم) من منظور صرفي ودلالي، مع تقديم هيكلية مشابهة لشبكة الكلمات. يوفر النموذج تفاصيل حول الأجزاء النحوية (فعل، اسم، صفة)، والمرادفات، والمصطلحات الأعم/الأخص، مثل: "التواصل" كمصطلح أعم لـ "كتب" و "لَکَم" كمصطلح أخص لـ "ضرب". كما يشمل النموذج مفاهيم مرتبطة (مثل: القراءة والكتابة لـ "كتب") وعلاقات صرفية، مثل: مشتقات "كتاب" و "مكتوب". يتميز النموذج Grok بتقديم رسوم بيانية توضح العلاقات بين الكلمات، مما يسهل تصور الشبكة الدلالية. ومع ذلك، فإن النموذج يفتقر إلى أمثلة سياقية واضحة، مما قد يقلل من فائدته في التطبيقات العملية. كما أن تركيزه على العلاقات الصرفية قد يجعله أقل شمولية في تغطية المعاني المتعددة مقارنة بالنموذج ChatGPT.

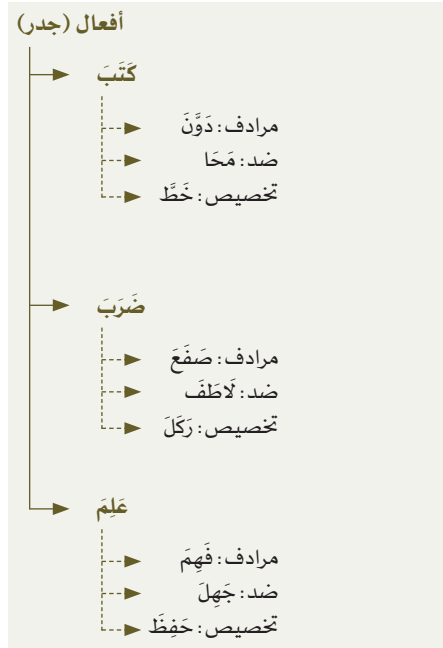
شكل (٢) تحليل وبناء شبكة الكلمات للجذر "كتب" باستخدام النموذج Grok

المجموعة ١: كتب (الكتابة التسجيل)	١. كتب
- الكلمات: كتب (كتابة)، سَجَل (تسجيل)، دون (تدوين) - الجزء التحوي: فعل. - التعريف: إنتاج أو تسجيل نص أو معلومات. - مصطلح اعم: تواصل، توثيق. - مصطلحات أخص: - خط (خط / الكاليفرافيا). - ألف (تأليف). - الأشكال المرتبطة: - كتاب (كتاب، اسم). - كتابة (كتابة، اسم). - مكتوب (مكتوب صفار) - مفاهيم مرتبطة: القراءة والكتابة التأليف النص.	- الجذر: ك-ت-ب، مرتبط بالكتابة أو التسجيل. - الجزء النحوي: فعل. - المعنى: كتابات تسجيل، تأليف. - مرادفات: سجل (التسجيل)، دَوْن (الدوين)، مع اختلاف حسب السياق. - الأشكال المرتبطة: - اسم: كتاب (كتاب)، كتابة (كتابة). - أفعال: يكتب (مضارع)، مكتوب (مفعول). - مصطلح أعم: فعل التواصل أو التسجيل. - مصطلح أخص: خط (خط / كاليفرافيا)، ألف (تأليف). - مفاهيم مرتبطة: القراءة والكتابة التأليف التوثيق.

٧-٢-٣- تحليل قدرات النموذج DeepSeek

يقدم النموذج DeepSeek تحليلاً مميزاً يركز على العلاقات الدلالية المتقدمة، بما في ذلك المرادفات، والأضداد، والمصطلحات الأعم/الأخص. على سبيل المثال، لـ "كتب"، يحدد مرادفات مثل: "دون، صاغ، سطر" وأضداد مثل: "محا، حذف"، مع مصطلحات أعم (أنشأ) وأخص (خط، نسخ). بالنسبة لـ "ضرب"، يوفر أضداداً مثل: "لاطف" ومصطلحات أخص مثل: "ركل"، بينما يحدد لـ "علم" أضداداً مثل: "جهل، نسي" ومصطلحات أخص مثل: "حفظ". ومع ذلك، فإن النموذج يقتصر على معنى واحد رئيسي لكل جذر، مما يقلل من قدرته على التعامل مع تعدد المعاني مقارنة بالنماذج Gemini و ChatGPT. كما أن غياب الأمثلة السياقية يحد من فائدته في التطبيقات التعليمية.

شكل (٣) تحليل وبناء شبكة الكلمات للجذر "كتب" باستخدام النموذج DeepSeek



١. كَتَبَ (الفعل: كتب - "يُدَوِّن، يُسَجِّل")

- المرادفات: دَوَّنَ، صَاغَ سَطَّرَ، أَلَفَ.
- الأضداد: مَحَا، حَذَفَ، شَطَبَ.
- تعميم (كلمة أشمل): أَنْشَأَ (أنشأ نصاً)، أَعَدَّ (أعد وثيقة).
- تخصيص (كلمة أضيق): حَطَّ (باليد)، نَسَخَ (أعاد الكتابة)، طَبَعَ (بالآلة).
- مفاهيم مرتبطة:
- القِرَاءَةُ (الفعل المقابل)، الوَثِيقَةُ (النتيجة)، القَلَمُ (الأداة).



٧-٢-٤- تحليل قدرات النموذج Gemini

يقدم النموذج Gemini تحليلاً شاملاً ومنظماً يشبه النموذج ChatGPT، لكنه يتجاوز في شمولية العلاقات الدلالية. لـ "كتب"، يحدد ثلاثة معانٍ (الكتابة، التأليف، القضاء)، مع مرادفات، مصطلحات أعم/أخص، وعلاقات معجمية (مثل: كاتب، كتابة). بالنسبة لـ "ضرب"، يغطي أربعة معانٍ (الضرب الجسدي، ضرب المثل، السفر، الضرب الرياضي)، مع إضافة أضداد مثل: "تلطف" للمعنى الجسدي. لـ "علم"، يحدد ثلاثة معانٍ (التعلم، الوعي، التعليم)، مع أضداد مثل: "جهل" و"غفل". يتميز النموذج Gemini بتقديم رسم بياني مفصل للعلاقات، مع تغطية شاملة للعلاقات المعجمية والدلالية. هذا يجعله الأكثر شمولية في التعامل مع تعدد المعاني والعلاقات بين الكلمات. ومع ذلك، قد تكون تعقيداته الزائدة عائقاً في التطبيقات البسيطة.

شكل (٤) تحليل وبناء شبكة الكلمات للجذر "كتب" باستخدام النموذج Gemini

١. **كُتِبَ - To Write**
 - **مجموعات المعنى (Synsets):**
 - المعنى ١: القيام بفعل الكتابة.
 - الشرح: تشكيل الحروف، الكلمات، أو الرموز على سطح باستخدام قلم أو أداة أخرى.
 - مرادفات: دُون، سَطَّر، حَزَّر.
 - اسم أعم (Hyponyms): فعل (القيام بفعل)، (بمعنى أوسع: عبر، تواصل).
 - اسم أعم (Hyponyms):
 - نسخ.
 - ألف.
 - خط.
 - مصطلحات ذات صلة (علاقات معجمية):
 - كاتب (اسم فاعل).
 - كتابة (اسم مصدر).
 - مكتوب (اسم مفعول، وقد يأتي بمعنى رسالة).
 - المعنى ٢: تأليف أو إنشاء عمل أدبي أو كتاب.
 - الشرح: إبداع عمل مكتوب، غالباً ما يكون ذا طول كبير أو مخصصاً للنشر.
 - مرادفات: ألف، أنشأ.
 - اسم أعم (Hypernym): أبداع، أنتج.
 - مصطلحات ذات صلة:
 - مؤلف.
 - كتاب.
 - المعنى ٣: قضاء أو أمر (كتاب إلهي).
 - الشرح: (مجازي) القضاء أو التقدير أو جعل شئ واجباً، غالباً في سياق ديني أو رسمي.
 - مرادفات: فرض، قَدَّر.
 - اسم أعم (Hypernym): أمر.
 - مصطلحات ذات صلة:
 - قضاء.
 - مكتوب (بمعنى قدر أو نصيب).
٢. **علاقات تعدد المعاني (Polysemy):**
 - استخراج أسهم من عقدة "كتب" إلى عقدة "كتب - المعنى ١: الكتابة"، و"كتب - المعنى ٢: التأليف"، و"كتب - المعنى ٣: القضاء/الأمر".
 - وبالمثل، استخراج أسهم من "ضرب" إلى معانيها الأربعة، ومن "علم" إلى معانيها الثلاثة.
٣. **علاقات المرادفات (Synonymy):**
 - داخل كل مجموعة معنى (Synset)، ستكون هناك روابط بين الكلمة الرئيسية والمرادفات.
 - مثال: من "كتب - المعنى ١" إلى "دُون"، "سَطَّر"، "حَزَّر".
 - مثال: من "علم - المعنى ١" إلى "إلى" "تعلّم"، "استوعب"، "أدرك".
٤. **علاقات اسم الأعم (Hyponymy):**
 - سهم يتجه من المعنى الأصغر إلى المعنى الأعم.
 - مثال: من "كتب - المعنى ١: الكتابة" سهم إلى "فعل"، (أو "تواصل").
 - مثال: من "ضرب - المعنى ١: توجيه ضربة" سهم إلى "اعتدى".
 - مثال: من "علم - المعنى ١: التعلم" سهم إلى "اكتسب".
٥. **علاقات اسم الأعم (Hyponymy):**
 - سهم يتجه من المعنى الأعم إلى المعنى الأصغر (عكس Hypernymy).
 - مثال: من "كتب - المعنى ١: الكتابة" سهم إلى "نسخ"، "ألف"، "خط" (هذه في أنواع الكتابة).
٥. **علاقات الكلمات ذات الصلة (Lexical Relations):**
 - ستربط هذه العلاقات الكلمات المشتقة أو ذات الصلة بمعانيها الأصلية.
 - مثال: من "كتب - المعنى ١: سهم إلى: كتب" (اسم فاعل)، "كتابة" (اسم مصدر)، "مكتوب" (اسم مفعول).
 - مثال: من "ضرب - المعنى ١: سهم إلى: ضارب".
 - مثال: من "علم - المعنى ٢: سهم إلى: عالم"، "معلومة".
٦. **علاقات الضد (Antonymy):**
 - إذا كان هناك ضد مباشر لمعنى ما، فسيظهر سهم يربط بينهما.
 - مثال: من "علم - المعنى ١: التعلم" سهم إلى "جهل".
 - مثال: من "ضرب - المعنى ١: توجيه ضربة" سهم إلى "جهل".



٧-٣- مقارنة بين النماذج

- **الشمولية في تغطية المعاني:** النموذج ChatGPT و Gemini يتفوقان في تغطية المعاني المتعددة (٣-٤ معاني لكل جذر)، بينما يقتصر النموذج DeepSeek على معنى واحد رئيسي، والنموذج Grok يغطي معاني متعددة، ولكنه أقل تفصيلاً. النموذج Gemini يقدم أعلى مستوى من التفصيل بفضل شمولية العلاقات (مرادفات، أضداد، أعم/أخص).
 - **العلاقات الدلالية والصرفية:** النموذج Gemini و DeepSeek يتفوقان في تقديم الأضداد والمصطلحات الأعم/الأخص، بينما النموذج ChatGPT يركز فقط على المرادفات والأمثلة. النموذج Grok يقدم علاقات صرفية قوية، ولكنه يفتقر إلى الأضداد.
 - **الأمثلة السياقية:** النموذج ChatGPT يتفوق في تقديم أمثلة سياقية واضحة، مما يجعله مفيداً للتطبيقات التعليمية. النموذج Gemini يقدم بعض الأمثلة ضمناً، بينما النموذج Grok و DeepSeek يفتقران إلى هذه الأمثلة.
 - **التصور البصري:** النموذج Grok و DeepSeek يقدمان رسوماً بيانية، مما يجعلهما مناسبين للحوسبة اللغوية. النموذج Gemini يقدم تصوراً بيانياً أكثر تفصيلاً، بينما النموذج ChatGPT يعتمد على هيكلية هرمية نصية.
 - **سهولة الاستخدام:** النموذج ChatGPT هو الأكثر وضوحاً وبساطة، مما يجعله مناسباً للمبتدئين. النموذج Gemini، رغم شموليته، قد يكون معقداً بسبب كثرة العلاقات. النموذج Grok و DeepSeek يتطلبان معرفة تقنية لفهم الرسوم البيانية.
- يعد النموذج Gemini الأكثر شمولية ودقة في تحليل شبكة الكلمات العربية، حيث يغطي المعاني المتعددة، والعلاقات الدلالية (مرادفات، أضداد، أعم/أخص)، والعلاقات المعجمية بشكل مفصل. ومع ذلك، قد تكون تعقيداته عائقاً في التطبيقات البسيطة. النموذج ChatGPT مناسب للتطبيقات التعليمية بفضل أمثلته السياقية ووضوحه. النموذج Grok و DeepSeek مفيدان في الحوسبة اللغوية بسبب الرسوم البيانية، لكنهما أقل شمولية في تغطية المعاني.

٤-٧- البيانات المفقودة في النماذج الأربعة لتحليل شبكة الكلمات العربية وكيفية الحصول على شبكة كلمات متكاملة

تتناول هذه الدراسة العوامل المفقودة في النماذج الأربعة ChatGPT و Gemini و DeepSeek و Grok لبناء وتحليل شبكة الكلمات العربية (Arabic WordNet) مع اقتراح حلول لسد هذه الفجوات وبناء شبكة متكاملة. يقع التركيز على تحديد النقاط المفقودة في كل نموذج، واقتراح تحسينات، ومناقشة الخطوات اللازمة للحصول على شبكة مثالية بدمج نقاط القوة في كل نموذج.

٧-٤-١- البيانات المفقودة في كل نموذج

٧-٤-١-١- النموذج ChatGPT

- أ. الأضداد: النموذج ChatGPT لا يحدد الأضداد لأي من المعاني (مثل: "محا" أو "حذف" لـ "كتب"، أو "جهل" لـ "علم")، مما يحد من قدرته على تقديم شبكة دلالية كاملة.
- ب. العلاقات الأعم/الأخص (Hypernyms/Hyponyms): لا يوفر النموذج مصطلحات أعم (مثل: "تواصل" لـ "كتب") أو أخص (مثل: "خط" أو "نسخ") بصراحة ووضوح، مما يقلل من عمق التحليل الدلالي.
- ج. العلاقات المعجمية المتقدمة: لا يشمل النموذج علاقات مثل: الأسماء المشتقة (كاتب، ضربة) أو العلاقات بين المجالات الدلالية (مثل: ارتباط "كتب" و "علم" في سياق التعليم).

٧-٤-١-١-١- الحلول المقترحة:

- إضافة قسم للأضداد لكل معنى، مثل: "جهل" لـ "علم" أو "تلطف" لـ "ضرب"، مع أمثلة توضيحية.
- تضمين مصطلحات أعم وأخص، مثل: "تواصل" كمصطلح أعم لـ "كتب" و "خط" كمصطلح أخص؛ لتعزيز الشبكة الدلالية.



- إدراج العلاقات المعجمية مثل الأسماء المشتقة (كاتب، كتابة) والعلاقات بين المجالات (مثل: ارتباط "كتب" و"علم" في التعليم).

٧-٤-١-٢- النموذج Grok

- أ. الأمثلة السياقية: لا يقدم النموذج أمثلة سياقية (مثل: "كتب رسالة إلى صديقه")، مما يجعل التحليل نظرياً ويصعب تطبيقه في التعليم.
- ب. الأضداد: لا يشمل النموذج الأضداد، مما يحد من تغطية العلاقات الدلالية.
- ج. تعدد المعاني المحدود: رغم تغطية معانٍ متعددة، إلا أن النموذج لا يحدد جميع المعاني الممكنة (مثل: "الضرب الرياضي" لـ "ضرب" أو "وضع علامة" لـ "علم").
- د. العلاقات بين المجالات الدلالية: لا يبرز النموذج العلاقات بين الجذور (مثل: ارتباط "كتب" و"علم" في التعليم).

٧-٤-١-٢- الحلول المقترحة:

- إضافة أمثلة سياقية لكل معنى لتوضيح الاستخدام العملي، مثل: "ضرب الباب بقوة" لـ "ضرب".
- تضمين الأضداد، مثل: "محا" لـ "كتب" و "جهل" لـ "علم"؛ لإكمال الشبكة الدلالية.
- توسيع تغطية المعاني المتعددة، مثل: إضافة "الضرب الرياضي" لـ "ضرب" و "وضع علامة" لـ "علم".
- إدراج علاقات بين المجالات الدلالية، مثل: ارتباط "كتب" و"علم" في سياق التعليم أو "ضرب" و"علم" في سياق الاستكشاف.

٧-٤-٣- النموذج DeepSeek

- أ. تعدد المعاني: يقتصر النموذج على معنى واحد رئيسي لكل جذر (مثل: "كتب" كتدوين فقط)، مما يجعله يتجاهل معانٍ أخرى مثل: "التأليف" أو "القضاء" لـ "كتب".
- ب. الأمثلة السياقية: لا يوفر النموذج أمثلة سياقية، مما يقلل من فائدته في التطبيقات التعليمية.

- ج. العلاقات المعجمية المشتقة: لا يغطي النموذج الأسماء المشتقة (مثل: كاتب، ضربة) بشكل كامل.
- د. التكامل مع معانٍ أخرى: لا يربط النموذج بين المعاني المتعددة للجذر الواحد أو بين الجذور المختلفة.

٧-٤-١-٣-١- الحلول المقترحة:

- توسيع التحليل ليشمل جميع المعاني المتعددة، مثل: "التأليف" و"القضاء" لـ "كتب"، مع مرادفات وأضداد لكل معنى.
- إضافة أمثلة سياقية، مثل: "علم الأطفال الحروف" لـ "علم"، لتوضيح الاستخدام.
- تضمين الأسماء المشتقة والعلاقات المعجمية، مثل: "كاتب" و"كتابة" لـ "كتب".
- إدراج علاقات بين المعاني المتعددة للجذر الواحد وبين الجذور المختلفة لتعزيز الشبكة الدلالية.

٧-٤-١-٤- النموذج Gemini

- أ. الأمثلة السياقية المحدودة: رغم شمولية النموذج، إلا أنه لا يقدم أمثلة سياقية واضحة لكل معنى، مما قد يحد من فائدته في التعليم.
- ب. التعقيد الزائد: يقدم النموذج تفاصيل كثيرة (مثل العلاقات المعجمية والدلالية)، مما قد يكون مربكاً للتطبيقات البسيطة.
- ج. العلاقات بين المجالات الدلالية: لا يبرز النموذج بوضوح العلاقات بين الجذور (مثل: "كتب" و"علم" في التعليم).

٧-٤-١-٤-١- الحلول المقترحة:

- إضافة أمثلة سياقية واضحة لكل معنى، مثل: "ضرب مثلاً في التضحية" لـ "ضرب".
- تبسيط العرض للتطبيقات البسيطة من خلال تقديم نسخة مختصرة من التحليل.
- تضمين علاقات بين المجالات الدلالية، مثل: ارتباط "كتب" و"علم" في التعليم.



للحصول على شبكة كلمات متكاملة، يجب دمج نقاط القوة في النماذج الأربعة: شمولية النموذج Gemini في تغطية المعاني والعلاقات، وضوح النموذج ChatGPT في الأمثلة السياقية، وقدرات النموذج Grok و DeepSeek في التصور البياني. يتطلب ذلك إضافة الأضداد، والمصطلحات الأعم/الأخص، والأمثلة السياقية، والعلاقات بين المجالات الدلالية، مع توفير تصور بصري وبياني. النموذج المقترح يوازن بين الشمولية والبساطة، مما يجعله مناسباً للتطبيقات الأكاديمية، والتعليمية، والحاسوبية. والمعالجة البشرية للحصول على شبكة معجمية متكاملة هي الأساس، كون النماذج اللغوية المقترحة دُرِبت في الأساس وبصورة كبيرة على اللغة الإنجليزية وبعض اللغات الأوربية، أما اللغات الأخرى وخصوصاً العربية فالبيانات غير كافية، والحل هو إثراء النماذج ببيانات عربية شاملة، والتدخل من خبراء العربية؛ حتى يضمن التحليل الشامل لكل ما يتعلق بالكلمة العربية لتغطية شبكة الكلمات.

7-0- المرحلة الثانية لعملية البناء والتحليل مع هندسة الأوامر

يُقدّم هذا القسم تحليلاً شاملاً ودراسةً كاملةً لما يتعلق بأساسيات بناء الشبكات العربية للكلمات بعد عجز النماذج اللغوية عن تغطية الأجزاء الأساسية للمفردة العربية في المرحلة الأولى، ركز الباحثان على المرجعيات اللغوية الأساسية التي تستخدم لبناء شبكة الكلمة العربية في هندسة الأوامر كمرحلة ثانية من البناء والتحليل. سيعرض هذا الجزء مدى تحسن أداء النماذج بعد هندسة الأوامر المقدمة لها من قبل الباحثين في بناء وتحليل الشبكات اللغوية العربية. في المثال الآتي حُلل الجذر نفسه الذي حُلل بدون هندسة الأوامر، وهو تحليل الجذر "كتب" في شبكة الكلمات العربية (Arabic WordNet)، مع التركيز على تغطية جميع المعاني الممكنة، والعلاقات الدلالية واللغوية، والأنطولوجيا، والتعريفات مع أمثلة سياقية. نُظِم التحليل وفقاً للعناصر المطلوبة: المجموعات المتشابهة (Synsets)، والأشكال الأساسية (Lemmas)، والعلاقات الدلالية واللغوية، والأنطولوجيا، والتعريفات مع الأمثلة.

المجموعات المتشابهة (Synsets)

تُغطي المجموعات المتشابهة المعاني المختلفة لكلمة "كتب" كفعل، مع تحديد المرادفات لكل معنى.

١. Synset الكتابة (الفعل)

- الجزء النحوي: فعل
- المرادفات: كتب، دون، سطر، حرر
- التعريف: تسجيل أو كتابة نص أو رموز على سطح باستخدام أداة مثل القلم.
- مثال: كتب رسالة إلى صديقه.

٢. Synset التأليف (الفعل)

- الجزء النحوي: فعل
- المرادفات: كتب، ألف، صاغ
- التعريف: إنشاء أو إنتاج عمل أدبي أو علمي مكتوب.
- مثال: كتب رواية جديدة نالت إعجاب النقاد.

٣. Synset القضاء/القدر (الفعل، مجازي)

- الجزء النحوي: فعل
- المرادفات: كتب، فرض، قدر
- التعريف: جعل شيء مفروضاً أو مقدراً، غالباً في سياق ديني أو شرعي.
- مثال: كتب الله عليه الجهاد.

الأشكال الأساسية (Lemmas)

تشمل الأشكال الأساسية المتغيرات الصرفية، والمشتقات، والتعبيرات متعددة الكلمات المرتبطة بـ "كتب"، مع تسوية التشكيل (بدون تشكيل أو تنوين).



● الأشكال الأساسية (الفعل):

- كتب (الماضي، الثالث المفرد المذكر)
- يكتب (المضارع)
- مكتوب (اسم المفعول)
- اكتب (الأمر)

● المشتقات:

- كتاب (اسم، بمعنى المصنف أو النص المكتوب)
- كتابة (اسم مصدر، بمعنى فعل الكتابة)
- كاتب (اسم فاعل، الشخص الذي يكتب)
- مكتب (اسم مكان، مكان الكتابة)
- مكتبة (اسم مكان، مكان تخزين الكتب)
- كتيب (اسم تصغير، كتاب صغير)
- كتابي (صفة، متعلق بالكتابة)

● التعابير متعددة الكلمات:

- كتاب مقدس
- كتاب مدرسي
- كتاب دراسي
- كتابة يدوية
- كتابة أدبية
- كتابة رسمية

العلاقات الدلالية (Semantic Relations)

علاقات شمولية (Hypernym / Hyponym)

● مصطلحات أعم (Hypernyms)

- معنى الكتابة: تواصل (الكتابة شكل من أشكال التواصل)
- معنى التأليف: إبداع (التأليف شكل من أشكال الإبداع)
- معنى القضاء: أمر (القضاء شكل من أشكال الأمر أو التقدير)

● مصطلحات أخص (Hyponyms)

- معنى الكتابة: خط (كتابة باليد)، نسخ (إعادة كتابة نص)، طبع (كتابة بالآلة)
- معنى التأليف: إنشاء (كتابة نص إبداعي قصير)، تصنيف (إنتاج كتاب علمي)
- معنى القضاء: لا توجد مصطلحات أخص شائعة.

علاقات الجزء-والكل (Part-Whole)

● الكتاب (كاسم):

- الأجزاء: صفحة، فصل، غلاف، فهرس.
- الكل: مكتبة (الكتاب جزء من مجموعة المكتبة)

● الكتابة (كعملية):

- الأجزاء: حرف، كلمة، جملة.
- الكل: نص، وثيقة.

العلاقات السياقية

● كلمات مرتبطة في السياق:

- قلم (أداة الكتابة)
- ورق (سطح الكتابة)
- حبر (مادة الكتابة)



- قراءة (الفعل المقابل للكتابة)
- نص (نتيجة الكتابة)
- مؤلف (الشخص الذي يؤلف)
- دار نشر (المؤسسة المرتبطة بالتأليف)

العلاقات اللغوية (Lexical Relations)

الترادف (Synonymy)

- الكتابة: كتب، دون، سطر، حرر
- التأليف: كتب، ألف، صاغ
- القضاء: كتب، فرض، قدر

التضاد (Antonymy)

- لمعنى الكتابة: محا، حذف، شطب
- لمعنى التأليف: لا أضداد مباشرة (لأنه إبداعي)
- لمعنى القضاء: لا أضداد مباشرة (لأنه سياق ديني / مجازي)

الاشتقاق الصرفي (Morphological Derivation)

● من الجذر ك-ت-ب:

- كتاب (اسم)
- كتابة (اسم مصدر)
- كاتب (اسم فاعل)
- مكتوب (اسم مفعول)
- مكتب (اسم مكان)
- مكتبة (اسم مكان)
- كتيب (اسم تصغير)

الإحالات المتقاطعة (Cross-References)

● مفاهيم ذات صلة:

- علم (في سياق التعليم، حيث الكتابة أداة لنقل المعرفة)
- قراءة (الفعل المكمل للكتابة)
- نشر (مرتبط بالتأليف)
- وثيقة (نتيجة الكتابة)

الأنطولوجيا (Ontology)

الهيكل الهرمي:

- فعل ← تواصل ← كتابة ← كتب (المعنى ١)
- فعل ← إبداع ← تأليف ← كتب (المعنى ٢)
- فعل ← أمر ← قضاء ← كتب (المعنى ٣)
- اسم ← شيء مادي ← مصنف ← كتاب
- اسم ← مكان ← مكتبة ← مكتب

العلاقات الرأسية:

- مفهوم أعلى: تواصل (للكتابة)، إبداع (للتأليف)، أمر (للقضاء)، شيء مادي (للكتاب)
- مفهوم أدنى: خط، نسخ (للكتابة)، إنشاء، تصنيف (للتأليف)

التعريفات والأمثلة (Definitions & Usage Examples)

أ. الكتابة:

- التعريف: فعل تسجيل نص أو رموز على سطح باستخدام أداة.
- مثال: كتب الطالب ملاحظاته في الدفتر.



ب. التأليف:

- **التعريف:** إنتاج عمل أدبي أو علمي مكتوب.
- **مثال:** كتب كتاباً عن تاريخ الأندلس.

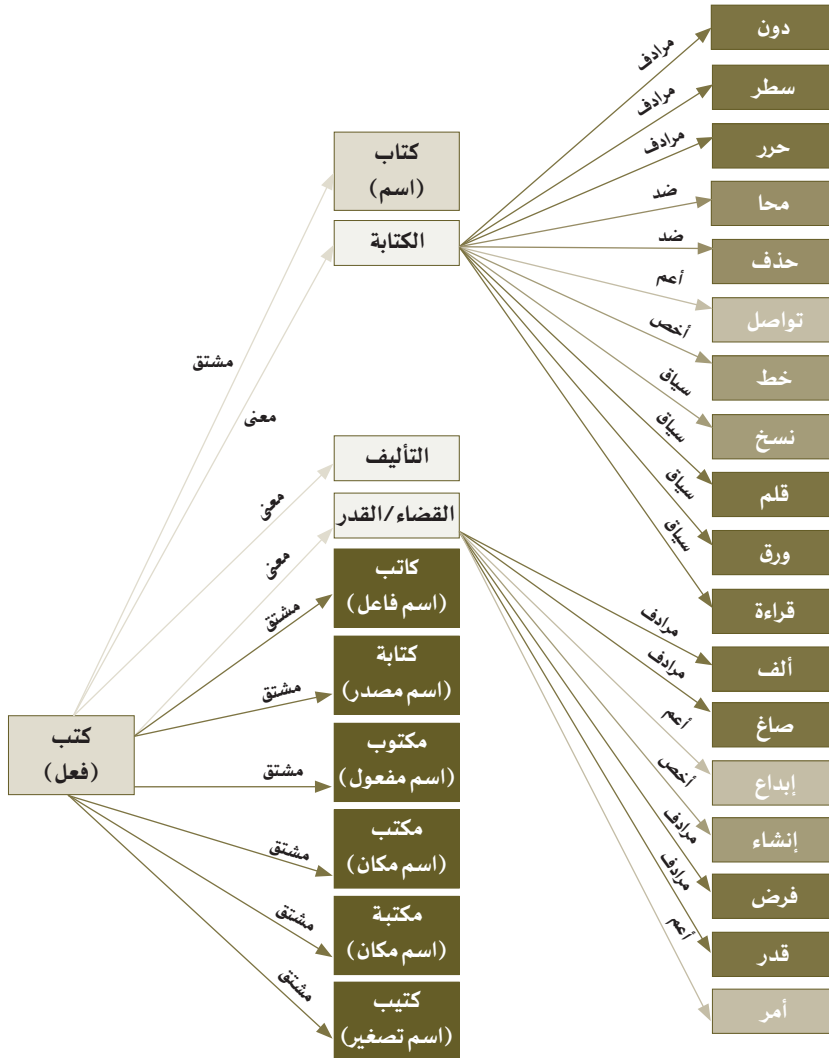
ج. القضاء/القدر:

- **التعريف:** جعل شيء مفروضاً أو مقدراً في سياق ديني أو شرعي.
- **مثال:** كتب الله له الشفاء بعد مرض طويل.

د. الكتاب (اسم):

- **التعريف:** مصنف مكتوب يحتوي على نصوص أو معلومات.
- **مثال:** اشترى كتاباً جديداً عن الفلسفة.

شكل (٥) الرسم البياني الشبكي اللغوي والمعجمي لكلمة "كتب" في المرحلة الثانية



- يُظهر الرسم البياني العلاقات بين "كتب" (كفعل واسم)، ومعانيها، والمشتقات.
- الألوان تميز العلاقات: الأزرق (مرادف)، الأحمر (ضد)، الأخضر (أعم)، الأرجواني (أخص)، البرتقالي (مشتق)، الرمادي (سياق).



يقدم التحليل الشبكي اللغوي والمعجمي لكلمة "كتب" دراسة شاملة تتناول جميع أبعاد الكلمة في شبكة الكلمات العربية، حيث يغطي المعاني المتعددة (الكتابة، التأليف، القضاء/القدر) والأشكال الأساسية (مثل: كتاب، كاتب، مكتبة)، مع التركيز على العلاقات الدلالية واللغوية. يتميز التحليل بتقديم مجموعات مترادفة (Synsets) واضحة لكل معنى، مثل: "كتب، دون، سطر" للكتابة، مع تعريفات دقيقة وأمثلة سياقية (مثل: "كتب رسالة إلى صديقه") تعزز فهم الاستخدامات. كما يشمل التحليل علاقات شمولية (مثل: "تواصل" كمصطلح أعم و"خط" كمصطلح أخص)، وعلاقات الجزء والكل (مثل: صفحة وكتاب)، والعلاقات اللغوية مثل: الترادف ("دون، سطر")، التضاد ("محا، حذف")، والاشتقاق الصرفي (كاتب، مكتوب). التصور البصري يوفر هيكلية مرئية تسهل تتبع العلاقات، مما يجعل التحليل أداة قيمة للحوسبة اللغوية والتطبيقات التعليمية. في الملحق الأول يبين التصور البصري والرسم البياني في المرحلة الأولى والثانية.

يبرز التحليل أيضاً من خلال الأنطولوجيا الهرمية التي توضح موقع "كتب" ضمن مسارات دلالية (مثل: فعل ← تواصل ← كتابة)، مما يعزز فهم العلاقات الرأسية بين المفاهيم. كما يربط الإحالات المتقاطعة "كتب" بمفاهيم مثل: "علم" و"قراءة"، مما يظهر التفاعل بين المجالات الدلالية مثل التعليم. ومع ذلك، يمكن تحسين المخرجات بإضافة المزيد من الأمثلة السياقية لتغطية تنوع الاستخدامات الثقافية أو الإقليمية، وتوسيع العلاقات بين المجالات الدلالية لتشمل مفاهيم إضافية مثل: "نشر" أو "توثيق". عامةً، يوفر التحليل إطاراً قوياً لفهم "كتب" في سياقاتها اللغوية والدلالية، مع إمكانية التوسع لتحليل جذور أخرى أو تطبيقات متقدمة في معالجة اللغة الطبيعية.

٨- التوصيات

- دمج نقاط القوة في النماذج الأربعة: يُوصى بدمج شمولية النموذج Gemini، ووضوح الأمثلة السياقية في النموذج ChatGPT، وقدرات التصور البرمجي في النموذجين Grok وDeepSeek لإنشاء إطار تحليلي موحد يغطي غالبية الجوانب الصرفية والدلالية والسياقية للمفردات العربية.

- هندسة الأوامر: يوصى بالتركيز على هندسة الأوامر المدخلة للنماذج اللغوية وصياغتها بفعالية من خبراء لغويين ذوي خبرة في العلاقات الدلالية والصرفية والنحوية.
- توسيع قاعدة البيانات اللغوية: ينبغي توسيع قاعدة بيانات Arabic WordNet لتشمل المزيد من الجذور اللغوية والمعاني الإقليمية، مع الاستفادة من مصادر نصوص متنوعة مثل: الكتب الدينية والأدبية وقواعد بيانات لتعزيز السياقات.
- تحسين أداء النماذج اللغوية: يُوصى بتدريب النماذج اللغوية الكبيرة (خاصة Grok) على بيانات عربية موسعة لتحسين دقة تجميع المرادفات، وتصنيف الأضداد، والعلاقات الجزئية، مع التركيز على معالجة الفروق الدلالية الدقيقة.
- تطوير أدوات تصور متقدمة: يُقترح تطوير أدوات تصوّر تفاعلية باستخدام بعض التقنيات الذكية الحديثة لدعم التطبيقات الحاسوبية والتعليمية، مما يسهل فهم العلاقات المعقدة في الشبكة المعجمية.
- الشراكة اللغوية بين الخبراء وأدوات الذكاء الاصطناعي: يوصى بتطوير برامج تدريب مشتركة تجمع بين الخبراء اللغويين ومتخصصي الذكاء الاصطناعي؛ لتحسين أداء النماذج اللغوية الكبيرة في تحليل المفردات العربية، من خلال إنشاء آليات مراجعة بشرية دورية لتصحيح الأخطاء في تجميع المرادفات، وتصنيف الأضداد، والعلاقات الدلالية، مما يعزز دقة الشبكة المعجمية العربية ويضمن تلبية متطلبات السياقات الثقافية واللغوية المتنوعة.
- تعزيز التطبيقات العملية: يُوصى بتطبيق الإطار الدلالي في مجالات مثل: التعليم اللغوي، ومعالجة اللغة الطبيعية، وتطوير أنظمة استرجاع المعلومات، مع إجراء دراسات مستقبلية لتقييم أثر هذه الشبكات على تحسين فهم اللغة العربية في السياقات الثقافية والاجتماعية.

٩- الخاتمة

يُظهر البناء الشبكي اللغوي والمعجمي للمفردات العربية ضمن إطار شبكة الكلمات العربية (Arabic WordNet) باستخدام النماذج اللغوية الكبيرة (ChatGPT، Gemini، DeepSeek، Grok) قدرة هذه النماذج على تقديم تحليلات تتناول المعاني المتعددة، والعلاقات الدلالية (مثل:



دور النماذج اللغوية الكبيرة في تطوير (WordNet) قاعدة بيانات معجمية عربية: تقييم القدرات وتحسين الشبكات الدلالية

هذه الطبعة إهداء من المجمع

ولا يُسمح بنشرها ورقياً أو تداولها تجارياً

المرادفات، الأضداد، المصطلحات الأعم/الأخص)، والعلاقات اللغوية (مثل: المشتقات الصرفية). اتضح أن النموذج Gemini يتفوق في الشمولية بتغطيته الواسعة للعلاقات الدلالية واللغوية، بينما يبرز النموذج ChatGPT بوضوح الأمثلة السياقية التي تدعم التطبيقات التعليمية. النموذجان Grok وDeepSeek يقدمان دعماً قوياً للحوسبة اللغوية، لكنهما يعانيان من نقص في الأمثلة السياقية والتغطية الكاملة للمعاني المتعددة. تكشف الدراسة عن قصور في النماذج مثل: التجميع غير الدقيق للمرادفات، والتصنيف الخاطئ للأضداد، والثغرات في العلاقات الجزئية، لكنها تؤكد تفوق Grok في التعامل مع تعقيدات الصرف العربي والفروق الدلالية، مما يجعله أداة واعدة لتطوير قاعدة بيانات معجمية عربية موثوقة. يمثل هذا العمل خطوة نوعية نحو بناء شبكة معجمية عربية مدعومة بالذكاء الاصطناعي، حيث يوفر إطاراً دلالياً متيناً يعزز الدقة الدلالية والنطاق المعجمي للمفردات العربية، مع تقليل الوقت والجهد المطلوبين لتطوير الموارد اللغوية. أظهرت الدراسة أن الهندسة البشرية للأوامر المقدمة للنماذج طورت بشكل فعال أداء النماذج في بناء شبكة الكلمات العربية (Arabic WordNet)، لاسيما في معالجة تعقيدات الصرف العربي والفروق الدلالية الدقيقة. ولتعزيز الكفاءة، وتقليل الوقت والجهد، ورفع الإنتاجية اللغوية بدقة عالية، يُعد دمج الخبرة البشرية مع أدوات الذكاء الاصطناعي ضرورة حتمية. يتيح هذا التكامل الاستفادة المثلى من قدرات النماذج اللغوية في معالجة البيانات آلياً، مما يوفر موارد معجمية موثوقة تدعم التطبيقات التقنية المتنوعة، مثل معالجة اللغة الطبيعية، والتعليم اللغوي، واسترجاع المعلومات، مع الحفاظ على الدقة العالية والمرونة في تلبية متطلبات السياقات الثقافية واللغوية العربية.

المراجع العربية

- الحلي، سندس محمد. المحاييري، محمد مازن. والعباس، فيصل. (٢٠٢٣). إيجاد أنطولوجيا عربية بالتقابل بين شبكة الكلمات العربية والموسوعة العربية. مجلة جامعة دمشق للعلوم الهندسية، ٣٩ (٣)، ١٧٧-١٩٣.
- أدغيري، أمينة. (٢٠٢٢). منهجية مقترحة لبناء وتطوير أنطولوجيا عربية مجال اللسانيات الحاسوبية أنموذجاً. مجلة حوسبة اللغة العربية، ٢ (١)، ١٨-٣٩.
- طه، المعتز بالله السعيد. (٢٠١٢). شبكة الكلمات العربية المنهجية والتطبيق. مجلة مجمع اللغة العربية بدمشق، ٨٧ (٢)، ٤٢٣-٤٥٠.
- دلّول، إيمان صبحي. (٢٠٢٠). أنطولوجيا الدلالة المعجمية للغة العربية: نحو مقارنة منهجية لسانية. مجلة الميادين للدراسات في العلوم الإنسانية، ٢ (٤)، ٨-٣٤.

رومنة المراجع العربية

Bibliographic information in Roman script

- Al-Halabi, S. M., Al-Muhayri, M. M., & Al-Abbas, F. (2023). Iejad ontology Arabia bil-taqabul bayna Shabakat al-Kalimat al-Arabia wa-al-Mawsu'a al-Arabia [Finding an Arabic ontology by matching the Arabic WordNet and the Arabic Encyclopedia]. Damascus University Journal for Engineering Sciences, 39(3), 177-193.
- Adghiri, A. (2022). Manhajia muqtaraha li-bina' wa-tatwir ontology Arabia: Majal al-lisaniyat al-hasubiya unmuzhajan [A proposed methodology for building and developing an Arabic ontology: Computational linguistics as a model]. Journal of Arabic Language Computing, 2(1), 18-39.
- Taha, M. S. (2012). Shabakat al-kalimat al-Arabia: Al-manhajia wa-al-tatbiq [Arabic WordNet: Methodology and application]. Journal of the Arabic Language Academy in Damascus, 87(2), 423-450.



Dalloul, I. S. (2020). Ontology al-dalala al-mu'jamia lil-lugha al-Arabia: Nahwa muqaraba manhajia lisania [Ontology of Arabic lexical semantics: Towards a linguistic methodological approach]. *Al-Mayadeen Journal for Humanistic Studies*, 2(4), 8-34.

المراجع الأجنبية

Al-Barhamtoshy, H., & Al-Jideebi, W. (2009, December). Designing and implementing Arabic WordNet semantic-based. In *the 9th Conference on Language Engineering* (pp. 23-24).

AlMaayah, M., Sawalha, M., & Abushariah, M. A. (2016). Towards an automatic extraction of synonyms for Quranic Arabic WordNet. *International Journal of Speech Technology*, 19(2), 177-189.

Alsudias, L., & Rayson, P. (2020, May). Developing an Arabic infectious disease ontology to include non-standard terminology. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 4842-4850).

Al-Garaady, J., & Mahyoob, M. (2023). ChatGPT's capabilities in spotting and analyzing writing errors experienced by EFL learners. *Arab World English Journals*, Special Issue on CALL,(9), 3-17.

Alkhalifa, M., & Rodríguez, H. (2010). Automatically extending named entities coverage of Arabic WordNet using Wikipedia. *International Journal on Information and Communication Technologies*, 3(3), 20-36.

Al-Garaady, J., & Mahyoob, M. (2025). Understanding User Perceptions of DeepSeek: A Mixed-Methods Sentiment and Thematic Analysis. *Available at SSRN* 5172367.

Albuhairy, M., Al-Garaady, J., & Alblwi, A. (2023). A proposed framework for human-like language processing of ChatGPT in academic writing. *International Journal of Emerging Technologies in Learning (iJET)*, 18, (14).

Al Zamil, M. G., & Al-Radaideh, Q. (2014). Automatic extraction of ontological relations from Arabic text. *Journal of King Saud University-Computer and Information Sciences*, 26(4), 462-472.

- Albuhairy, M. M., & Algaraady, J. (2025). DeepSeek vs. ChatGPT: Comparative efficacy in reasoning for adults' second language acquisition analysis. *Humanities and Educational Sciences Journal*, 44, 864–883.
- Al-Garaady, J. (2022). An analysis of Yemenis' responses and sentiments on social media towards the emergence of the COVID-19 pandemic. *Humanities and Educational Sciences Journal*, 27, 589-607.
- Aliane H., Alimazighi Z. Ahmed Cherif M. "Al-Khalil: 2009. The Arabic Linguistic Ontology Project", Semantic web and Arabic Language Team, in Computer Science Department, University, Algiers.
- Algaraady, J., & Mahyoob, M. (2025). Exploring ChatGPT's potential for augmenting post-editing in machine translation across multiple domains: challenges and opportunities. *Frontiers in Artificial Intelligence*, 8, 1526293.
- Belkredim, F. Z., & Meziane, F. (2008). DEAR-ONTO: a derivational Arabic ontology based on verbs. *International Journal of Computer Processing of Languages*, 21(03), 279-291.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.
- Farghaly, A., & Shaalan, K. (2009). Arabic natural language processing: Challenges and solutions. *ACM Transactions on Asian Language Information Processing (TALIP)*, 8(4), 1-22.
- Habash, N. Y. (2010). *Introduction to Arabic natural language processing*. Morgan & Claypool Publishers.
- Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties*. Georgetown University Press.
- Ishkewy, H., Harb, H., & Farahat, H. (2014). Azhary: An Arabic Lexical Ontology. *International Journal of Web & Semantic Technology*, 5(4), 71.
- Jarrar, M. (2021). The Arabic ontology—an Arabic wordnet with ontologically clean content. *Applied ontology*, 16(1), 1-26.



- Mazari, A. C., Aliane, H., & Alimazighi, Z. (2012). Automatic Construction of Ontology from Arabic Texts. In *ICWIT* (pp. 193-202).
- Mahyoob, M., & Algaraady, J. (2018). Towards developing a morphological analyser for Arabic noun forms. *International journal of Linguistics and computational Applications*, 5(3), 45-51.
- Navigli, R., & Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial intelligence*, 193, 217-250.
- Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). A primer in BERTology: What we know about how BERT works. *Transactions of the association for computational linguistics*, 8, 842-866.
- Rodríguez, H., Farwell, D., Ferreres, J., Bertran, M., Alkhalifa, M., & Martí, M. A. (2008). Arabic WordNet: Semi-automatic Extensions using Bayesian Inference. In *LREC*.
- Saif, A., Ab Aziz, M. J., & Omar, N. (2017). Mapping Arabic WordNet synsets to Wikipedia articles using monolingual and bilingual features. *Natural Language Engineering*, 23(1), 53-91.

نبذة تعريفية عن الباحث أو الباحثين:

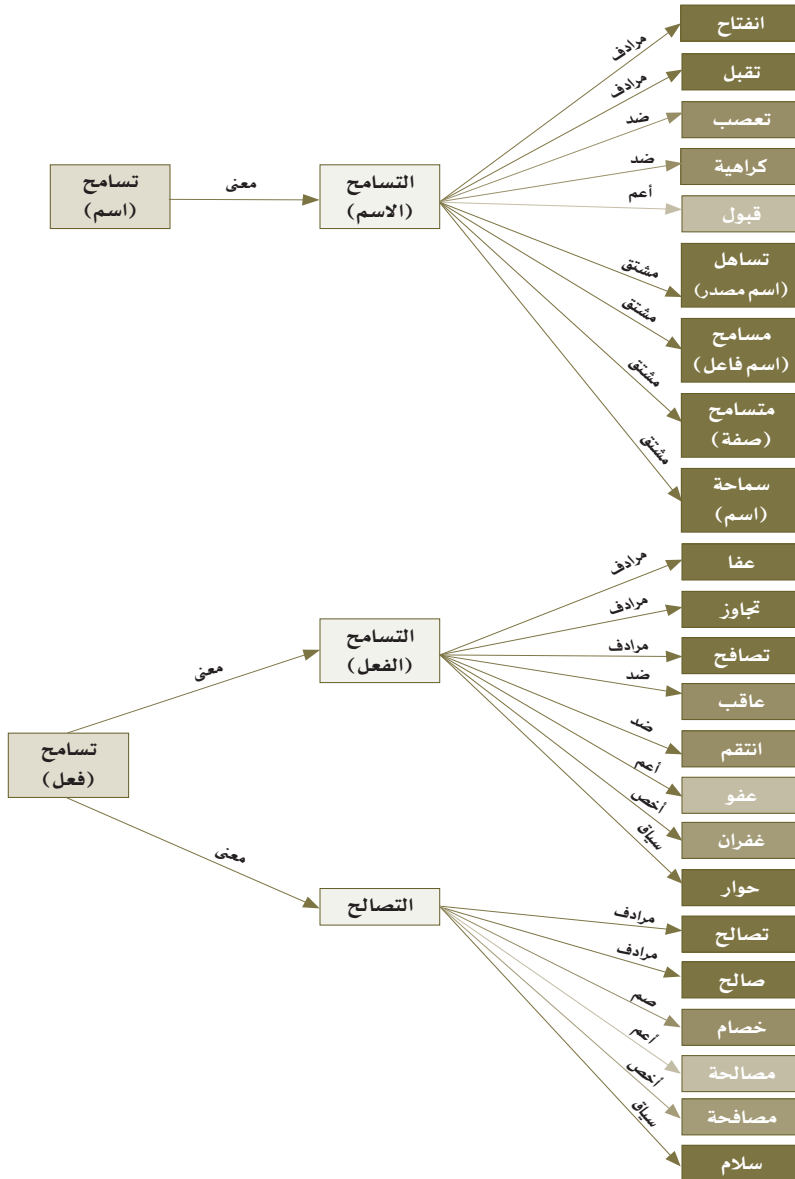
Author Biodata Dr. Mohammad Mahyoob Albuhairey is an Associate Professor of Computational linguistics at the Department of Languages & Translation, Taibah University, Madina, KSA. He holds his PhD in Computational Linguistics. His research interests include Computational Linguistics, AI, Disabilities, Digital Media Studies, Language Acquisition, Corpus Linguistics and e-learning.	بيانات الباحث: محمد مهيووب البحيري، أستاذ مشارك في اللغويات الحاسوبية بقسم اللغات والترجمة في جامعة طيبة بالمدينة المنورة، بالملكة العربية السعودية. حصل على درجة الدكتوراه في اللغويات الحاسوبية. من جامعة عليجار الإسلامية في الهند، عام ٢٠١٥. تدوراهتماماته البحثية حول اللغويات الحاسوبية، والذكاء الاصطناعي، والإعاقة، ودراسات الإعلام الرقمي، واكتساب اللغة، والمدونات اللغوية، والتعليم الإلكتروني.
معرف أوركيد (ORCID) : 0000-0002-6664-1017	
Email: eflu2010@gmail.com	

Author Biodata Dr. Jeehaan Algaraady is an Assistant Professor of Computational Linguistics at the College of Education, Taiz University, Yemen. She received her PhD in Computational Linguistics. Her research interests include Computational Linguistics, Data Mining, AI, Language Disorders, Theoretical Linguistics, and e-learning.	بيانات الباحثة: جيهان الجرادي، أستاذ مساعد في اللغويات الحاسوبية بقسم اللغة الإنجليزية في جامعة تعز، باليمن. حاصلة على درجة الدكتوراه في اللغويات الحاسوبية من جامعة عليجار الإسلامية في الهند عام ٢٠١٨. تدوراهتماماتها البحثية حول اللغويات الحاسوبية، والذكاء الاصطناعي، والاضطراب اللغوي، والتقيب في البيانات، والتعليم الإلكتروني.
معرف أوركيد (ORCID) : 0000-0003-3901-4648	
Email: Jihan.amu@gmail.com	

مجمع الملك سلمان
العالمي للغة العربية

ملحق (ب)

الرسم البياني الشبكي لكلمة "تسامح" بعد هندسة الأوامر



استعراض كتاب: لسانيات المدونات مقدمة موجزة

Corpus Linguistics A Short Introduction

المعرف الرقمي: <https://doi.org/10.60161/2521-003-005-004>

Dr. Alchimaa Mohammad Waleed Lutfi

Researcher in Arabic Applied Linguistics

Imam Mohammad Ibn Saud Islamic University

د. الشيماء محمد وليد لطفي

باحثة في اللغويات التطبيقية العربية

جامعة الإمام محمد بن سعود الإسلامية

Submission Date: 22/04/2026

تاريخ الإرسال: ١٤٤٧/١١/٠٥

Acceptance Date: 19/04/2026

تاريخ القبول: ١٤٤٧/١١/١٢

الكلمات المفتاحية / Keywords:

لسانيات المدونات، اللسانيات الحاسوبية - المدونات اللغوية - المتلازمات اللفظية - المعالجة الآلية للغة.

Corpus Linguistics, Computational Linguistics, Language Corpora, Collocations, Automatic Language Processing.

المخلص

يُعد كتاب لسانيات المدونات: مقدمة موجزة، الصادر بترجمته العربية عام ٢٠٢٢م عن داركنوز المعرفة، ترجمةً لكتاب Corpus Linguistics: A Short Introduction الصادر عام ٢٠٠٧م، من الأعمال التمهيدية المهمة في التعريف بلسانيات المدونات وقضاياها الأساسية. وقد شارك في تأليفه وولفغانغ تيوبيرت، المتخصص في لسانيات المدونات، وأنا تشيرماكوفا المتخصصة في اللسانيات والمدونات اللغوية. وتبرز أهمية الكتاب في تقديمه مدخلاً موجزاً يجمع بين الجانب النظري والتطبيقي، مع ربط تحليل اللغة بالأمثلة الواقعية والاستعمال الفعلي.

وعلى الرغم من غياب المقدمة التمهيدية التي توضح منهجية العمل وبنيتها، وعدم تضمين الترجمة العربية تقديمًا خاصًا بسياق الترجمة، إلا أن مادة الكتاب تكشف تصورًا واضحًا لقضايا المدونات اللغوية من حيث بنائها، وتمثيلها، واستثمارها في دراسة المعنى، والمتلازمات اللفظية، وتحليل الخطاب. لذلك يظل الكتاب مناسباً للباحثين في اللسانيات التطبيقية واللسانيات الحاسوبية، خاصة في الموضوعات المرتبطة بتحليل اللغة اعتماداً على المدونات.

بطاقة الكتاب المراجع:

عنوان الكتاب:	لسانيات المدونات مقدمة موجزة
عنوان الكتاب بالإنجليزية:	Corpus Linguistics A Short Introduction
المؤلف:	وولفغانق تيوبيرت وأناشيرماكوفا
المؤلف بالإنجليزية:	Wolfgang Teubert, Anna Cermakova
سنة نشر العمل الأصلي:	٢٠٠٧
سنة نشر العمل المترجم:	٢٠٢٢
عدد الصفحات:	١٧٤
دار النشر للعمل الأصلي:	Continuum International Publishing Group
دار النشر للعمل المترجم:	دار كنوز المعرفة للنشر والتوزيع

اللغة ولسانيات المدونات

هل كل اللغات واحدة؟

يتناول الكتاب عدداً من القضايا المرتبطة باللغة ولسانيات المدونات، مثل العلاقة بين اللسانيات المعيارية والمعنى، والتعبيرات الاصطلاحية، والمتلازمات اللفظية، وتمثيل اللغة داخل المدونات. كما يناقش أنواع المدونات، مثل المدونات المرجعية، والراصة، والمتوازية، ومدونات الإنترنت، إضافة إلى قضايا المعنى، والخطاب، والترجمة.

ويقوم التصور التوليدي عند تشومسكي على أن القدرة اللغوية تمكن الإنسان من إنتاج عدد غير محدود من الجمل، مما يجعل اللغة نظاماً ذهنياً أكثر من كونها مجرد نصوص. وفي المقابل تمنح لسانيات المدونات أهمية كبرى لدراسة اللغة كما تُستعمل فعلياً داخل النصوص والسياقات الواقعية، انطلاقاً من أن المعنى لا يفهم من الكلمات منفصلة، بل من استعمالها داخل الخطاب. وتبرز قيمة هذا التصور في التطبيقات المرتبطة بالمدونات اللغوية والمعالجة الآلية، خاصة في تحليل الاستعمال اللغوي، واستخراج المتلازمات اللفظية، وربط المعنى بالسياق الفعلي.



اللسانيات المعيارية ومعنى الكلمة

تظهر الإشكالية الرئيسية في الفرق بين تعلّم القواعد اللغوية، والقدرة على اختيار المفردة المناسبة أو ترجمتها؛ إذ قد يتمكن المتعلم من تكوين جمل صحيحة نحوياً، لكنه يواجه صعوبة في اختيار اللفظة الأدق دلالةً، بسبب قصور المعاجم أحياناً عن توضيح الفروق الدقيقة بين الاستعمالات المختلفة للكلمة الواحدة. ولا يفهم معنى الجمل غالباً من الكلمات المنفصلة، بل لا بد من السياق والمصاحبات اللفظية داخل الجملة أو النص؛ حتى يتضح المعنى المقصود. لذلك لا تبدو الكلمة المفردة دائماً وحدة المعنى الأساسية؛ فقد يتحقق المعنى من خلال تعبير اصطلاحي، أو تركيب متلازم، أو وحدة تركيبية أوسع. كما أن بعض الوحدات الأصغر من الكلمة قد تحمل دلالة، مثل العناصر الصرفية المقيدة، مما يجعل حدود الوحدة الدلالية أكثر تعقيداً.

ويؤدي الاقتصار على المعاني المعجمية المفردة إلى إرباك في الترجمة وتحليل الدلالة؛ لأن بعض المعاني لا ترتبط بالكلمة منفردة، بل بالتركيب والمصاحبات التي ترد فيها. ومن هنا تتضح أهمية لسانيات المدونات في دراسة الاستعمال الفعلي للغة، والاستفادة منها في التحليل الدلالي الحاسوبي واستخراج المتلازمات اللفظية.

التعبيرات الاصطلاحية والمتلازمات اللفظية

تكشف الأمثلة الواردة في المدونات أن كثيراً من التراكيب المتلازمة لا يمكن تفسيرها من خلال الكلمات المفردة وحدها، وهو ما لم يكن واضحاً في المعالجة المعجمية التقليدية. فالمعاجم قد تسرد عدداً من المعاني أو التراكيب، لكنها لا توضح دائماً طبيعة العلاقة بين عناصرها أو حدود استعمالها الفعلي. ويظهر ذلك في مثال "الفجر الكاذب"، الذي يغلب عليه الاستعمال المجازي أكثر من الاستعمال الحرفي، مما يدل على أن دلالاته ترتبط بالسياق المتكرر لا بالمعنى المعجمي المباشر. كما يتبين أن الاعتماد على الصور الذهنية وحدها لا يكفي لتفسير المعنى؛ لأن الكلمات قد ترتبط باستعمالات متعددة ومتغيرة بحسب السياق والخطاب. وتبرز هنا قيمة لسانيات المدونات في تفسير تغيرات المعنى وتكرارات الاستعمال، وفي دعم الدراسات المعجمية الحاسوبية، خاصة في تحليل المصاحبات اللفظية والاستعمال الواقعي داخل المدونات اللغوية.

لسانيات المدونات: نظرة مختلفة إلى اللغة

تنظر لسانيات المدونات إلى اللغة بوصفها ظاهرة اجتماعية تتجلى في النصوص والاستعمالات الفعلية، لا مجرد نظام ذهني؛ لذلك لا تقتصر دراسة اللغة على وصف القواعد، بل تشمل تتبع الكلمات والتراكيب داخل الخطاب الواقعي. ولا تنظر إلى اللغة بوصفها نظاماً ثابتاً، بل بوصفها ظاهرة تتغير باستمرار تبعاً للاستعمال والسياق الاجتماعي والثقافي. ومن هنا تكتسب النصوص الواقعية أهميتها؛ لأنها تكشف كيفية استعمال الناس للغة، بما في ذلك الاستعمالات الجديدة أو غير المعيارية.

ويرتبط المعنى في هذا التصور بالسياق والاستعمال المتكرر، لا بالصور الذهنية المجردة وحدها، وهو ما يمنح المدونات دوراً مهماً في فهم استعمال اللغة في الحياة اليومية. كما تدعم هذه الرؤية الدراسات القائمة على البيانات الواقعية، وخاصة في بناء المدونات وتحليل الاستعمال الفعلي للغة.

نبذة تاريخية عن لسانيات المدونات

برزت لسانيات المدونات مع تزايد الحاجة إلى الاعتماد على بيانات لغوية واقعية بدل الاختصار على الحدس أو الأمثلة المصنوعة، خاصة بعد اتضاح أن بعض الظواهر اللغوية لا يمكن الحكم عليها بدقة إلا من خلال نصوص حقيقية متنوعة. وشهد هذا المجال عدداً من المشروعات المبكرة، مثل مشروع مسح الاستعمال الإنجليزي، الذي جمع بيانات لغوية طبيعية تضم معلومات صوتية وصرفية ونحوية ودلالية، إضافة إلى مشروعات معجمية اهتمت بالمتلازمات اللفظية والوحدات المعجمية المركبة. كما أسهمت مدونة براون والمدونات المرجعية اللاحقة، في توسيع البحث المتعلق بالقواعد والمعجم وشيوع الكلمات، وارتبط تطور هذا الحقل بالمشروعات المعجمية الكبرى، مثل مشروع كوبلد، الذي اعتمد على المدونات في بناء المعاجم وتحليل السياق وأنماط الاستعمال المتكرر. ومع اتساع استخدام المدونات أصبحت أداة مهمة في الدراسات الاجتماعية، وتعليم اللغة، والترجمة، والدلالة، واللغويات الحاسوبية، خاصة في تنظيم البيانات النصية وتطوير الموارد اللغوية.



ويبرز كذلك الفرق بين الأسلوب القائم على المدونات، الذي يستخدم المدونة للتحقق من نتائج سابقة، والأسلوب الموجّه بالمدونات، الذي ينطلق من البيانات نفسها لاستخراج النتائج اللغوية، وهو الاتجاه الأقرب إلى توظيف المدونات بوصفها أداة تحليل مستقلة.

اتجاهات في لسانيات المدونات

اللغة والتمثيلية

ترتبط قضية تمثيل الخطاب في المدونة باختيار نصوص تمثل نوعاً محدداً من الخطاب وفق معايير واضحة، لا بجمع نصوص عشوائية. ولذلك لا يقتصر الخطاب على النصوص المطبوعة، بل يشمل الرسائل، والمحادثات اليومية، وخطابات المؤسسات، ووسائل الإعلام، وغيرها من صور التواصل. ولا تعني تمثيلية المدونة احتواءها على جميع الاستعمالات الممكنة، بل بناء عينة محددة تمثل خطاباً معيناً في زمان ومكان وسياق محددين. كما أن غياب بعض الكلمات أو ندرتها لا ينفي قيمة المدونة أو تمثيلها، لأن الهدف هو توفير بيانات قابلة للدراسة العلمية.

وتظهر أهمية هذه القضية في بناء المدونات اللغوية الحاسوبية؛ لاعتماد التحليل الآلي للغة على وضوح معايير التمثيل، وتحديد أنواع النصوص والسياقات التي تُبنى عليها المدونة.

تصنيف المدونات

تصنف المدونات اللغوية إلى أربعة أصناف رئيسية. وتعد المدونة المرجعية أحد هذه الأصناف؛ إذ تسعى إلى تمثيل الاستعمال القياسي للغة من خلال عينات متنوعة من النصوص المكتوبة والمنطوقة، مع توثيق معلومات كل نص من حيث نوعه ومجاله ووعائه. كما تحتاج إلى حجم كبير وتوازن نسبي؛ لأن دراسة الدلالة والمتلازمات اللفظية تعتمد على كثرة الأمثلة الواقعية. وتزداد قيمة هذا النوع من المدونات كلما كانت معايير بنائها أوضح وآليات توثيقها أدق، وهو ما جعلها أداة مهمة في الدراسات الحاسوبية والمعالجة الآلية، خاصة في التحليل المعجمي والدلالي واستخراج المتلازمات اللفظية.

فيما تُستخدم المدونة الراصدة لمتابعة التغيرات اللغوية عبر الزمن، ولا سيما التغيرات المعجمية وشيوع الكلمات والتعبيرات. ولذلك تحتاج إلى بنية ثابتة ومعايير موحدة، تسمح بإجراء المقارنات الإحصائية بين الفترات الزمنية المختلفة. وتبرز أهميتها في تحليل تطور الاستعمال اللغوي والتغيرات الدلالية، مما يجعلها مفيدة في الدراسات الإحصائية واللغويات الحاسوبية.

وتضم المدونات المتوازية نصوصاً أصلية مع ترجماتها إلى لغات أخرى؛ ولذلك تمثل أداة مهمة في دراسة الترجمة وتحليل المقابلات الدلالية بين اللغات. ولا تقتصر فائدتها على تقديم مقابلات معجمية مباشرة، بل تكشف كيفية ترجمة الكلمات والتراكيب داخل السياق الفعلي. وتبرز أهميتها في الترجمة الآلية وبناء الموارد ثنائية اللغة، خاصة مع اعتمادها على دراسة وحدات المعنى داخل النصوص الكاملة، لا على الكلمات المفردة وحدها.

كما وتمثل مدونات الإنترنت مصدراً واسعاً ومتجدداً للنصوص اللغوية، لما توفره من أمثلة حديثة ومتنوعة تساعد في تتبع الاستعمال الواقعي للكلمات والتعبيرات والمتلازمات اللفظية. ومع ذلك لا يمكن النظر إليها بوصفها بديلاً كاملاً عن المدونات الأخرى؛ لأن النصوص المنشورة عبر الإنترنت قد تتأثر بطرق النشر والسياقات المختلفة. وتبرز قيمة هذا النوع من المدونات في الدراسات اللغوية الحاسوبية؛ لاعتماد كثير من تطبيقات المعالجة الآلية للغة على البيانات النصية الضخمة والمتجددة.

المعنى في الخطاب واستعماله

يرتبط معنى الكلمة باستعمالها داخل الخطاب، لا بالكلمة منفصلة أو بالتعريف المعجمي المجرد؛ إذ يتشكل المعنى من خلال السياق، والمصاحبات اللفظية، والاستعمال المتكرر داخل مجتمع الخطاب. ويظهر ذلك في مثال "العولمة"، الذي لم يكتسب معناه دفعة واحدة، بل تشكل تدريجياً عبر استعمالات متعددة في الاقتصاد والسياسة والدراسات الاجتماعية، قبل أن تستقر دلالاته نسبياً في الخطاب العام والمعاجم. ولا ينفصل معنى الكلمة عن كيفية استعمالها، والألفاظ التي ترافقها، والعبارات التي يعاد بها شرحها وصياغتها داخل النصوص؛ لذلك لا يُختزل المعنى في تعريف ثابت، بل يظهر من خلال



أنماط الاستعمال والسياقات المتكررة. كما يمكن للحاسوب تتبع هذه الأنماط واستخراج التكرارات والمصاحبات اللفظية، غير أن تفسيرها الدلالي يظل مرتبطاً بقدرة الإنسان على الفهم والتأويل.

ويظهر تعبير "نيران صديقة" مثلاً على الوحدات الدلالية التي لا يمكن فهمها من خلال تحليل كلماتها المفردة، بل بوصفها تركيباً اكتسب معناه من شيوعه في الخطاب العسكري والإعلامي. كما تكشف المدونات الأنماط السياقية التي تثبت تحول بعض التراكيب إلى وحدات اصطلاحية مستقلة، وهو ما يفسر صعوبة ترجمتها ترجمة حرفية دقيقة بين اللغات. وتبرز أهمية المدونات المتوازية في دراسة الترجمة والمتلازمات اللفظية؛ لأنها تكشف المقابلات الفعلية للوحدات اللغوية داخل السياق، بدل الاكتفاء بالمقابلات المعجمية المفردة؛ لذلك لا تتم الترجمة دائماً كلمة بكلمة، بل قد تعتمد على وحدات أوسع تتحدد دلالتها من خلال الاستعمال والتلازم اللفظي.

وتتضح قيمة هذا التصور في اللغويات الحاسوبية والمعالجة الآلية للغة، خاصة في التحليل الدلالي، والترجمة الآلية، واستخراج المتلازمات اللفظية، وبناء المدونات المتوازية. كما يكشف عن تطور لسانيات المدونات من مجرد وصف التكرارات اللغوية، إلى الإسهام في تفسير المعنى وربطه بالسياق والاستعمال الفعلي. ويظل معنى الوحدات اللغوية مرتبطاً بالخطاب واستعمالاته الاجتماعية، والثقافية لا بالتعريفات المعجمية وحدها؛ فالفهم يتصل بخبرة المتلقي وذاكرته وسياقه الاجتماعي، ولذلك يبقى المعنى متغيراً ومتجدداً بقدر تجدد الاستعمال نفسه.

التحليل النقدي

تتمثل القيمة العلمية للكتاب في تقديمه لسانيات المدونات بوصفها مدخلاً لدراسة المعنى والاستعمال والخطاب، لا مجرد وسيلة لجمع البيانات اللغوية أو معالجتها إحصائياً. وتبرز أهمية هذا التصور في اعتماده على الأمثلة الواقعية وتتبع الاستعمال الفعلي للكلمات والتراكيب، مما أتاح مناقشة قضايا المعنى، والمتلازمات اللفظية، والترجمة، في ضوء السياق والخطاب لا في ضوء التعريفات المعجمية المجردة وحدها. كما أسهم الكتاب في توضيح

العلاقة بين المدونات اللغوية واللغويات الحاسوبية، خاصة في الموضوعات المرتبطة بالتحليل الدلالي، وبناء الموارد اللغوية، والترجمة الآلية.

وعلى الرغم من صدور الكتاب الأصلي باللغة الإنجليزية عام ٢٠٠٧م، إلا أن قيمته العلمية ما تزال قائمة؛ لأنه يعالج قضايا تأسيسية مازالت حاضرة في لسانيات المدونات، واللغويات الحاسوبية، ولا سيما قضايا تمثيل المدونات، وتحليل المعنى في السياق، واستخراج المتلازمات اللفظية، والإفادة من المدونات المتوازية في الترجمة. كما ويوازن الكتاب بين الجانب النظري والجانب التطبيقي؛ إذ لم يقتصر على عرض المفاهيم العامة، بل دعمها بأمثلة متعددة من المدونات، وهو ما أسهم في توضيح كيفية تشكل المعنى داخل السياقات المختلفة. كما تزيد قائمة المصطلحات الواردة في خاتمة الكتاب من قيمته المرجعية، خاصة للباحثين المبتدئين في هذا الحقل.

وعلى مستوى الترجمة العربية، تبدو الترجمة محافظة على الطابع العلمي للنص الأصلي، خاصة في عرض المصطلحات المرتبطة بلسانيات المدونات والدلالة والترجمة. كما أسهمت قائمة المصطلحات في تيسير متابعة المفاهيم المتخصصة. ومع ذلك فإن كثافة بعض التراكيب الاصطلاحية والنقاشات النظرية جعلت بعض المواضع تميل إلى التعقيد، ولا سيما في الصفحات التي تعتمد على الأمثلة المتتابعة والتحليل التفصيلي، وهو ما قد يتطلب من القارئ خلفية مسبقة في اللسانيات أو الدراسات المعجمية.

ويظهر من ذلك كثرة الأمثلة والتفصيلات الجزئية المؤدية في بعض المواضع إلى إطالة تُقلّل من تركيز الفكرة الرئيسية، ولا سيما في الموضوعات المرتبطة بالمعنى وإعادة الصياغة والمتلازمات اللفظية. كما أن اعتماد الكتاب على أمثلة مأخوذة من الإنجليزية والترجمة الثنائية، يجعله أقل اتصالاً بالتطبيقات المرتبطة بالعربية ومدونات المعاصرة، على الرغم من أن كثيراً من القضايا التي يطرحها يمكن الاستفادة منها في الدراسات العربية الحاسوبية.

ويلاحظ أيضاً أن بعض الأفكار الرئيسية، مثل العلاقة بين السياق والمعنى، أو بين الاستعمال والتفسير، عادت بصيغ متقاربة في أكثر من موضع، وكان من الممكن عرضها بقدر أكبر من التكثيف والتنظيم. إلا أن ذلك لا ينتقص من أهمية الكتاب؛ إذ يظل من



الأعمال التي تسهم في توسيع النظر إلى اللغة والمعنى، والتي تؤكد أن فهم الوحدات اللغوية يرتبط بالاستعمال والخطاب أكثر من ارتباطه بالتعريفات المعجمية الثابتة.

الخاتمة

يضع هذا الكتاب القارئ أمام سؤال جوهري في الدرس اللساني الحديث: كيف يمكن أن ندرس اللغة من خلال ما يقوله الناس فعلاً، لا من خلال الأمثلة المفترضة وحدها؟ ومن هذه الزاوية، لا يقدم الكتاب تعريفاً بلسانيات المدونات فحسب، بل يوجّه إلى طريقة في التفكير اللساني تقوم على تتبع الاستعمال، وملاحظة تكرار الأنماط، وربط المعنى بسياقه النصي والاجتماعي. وتبدو أهميته خاصة للباحث الذي يريد الانتقال من النظر إلى اللغة بوصفها نظاماً مجرداً إلى التعامل معها، بوصفها مادة نصية قابلة للرصد والتحليل. ولذلك يمكن أن يكون الكتاب نقطة انطلاق مناسبة لمن يرغب في فهم الأسس التي تقوم عليها دراسة المدونات، قبل التوسع في تطبيقاتها الحديثة في اللسانيات الحاسوبية والمعالجة الآلية للغة.

نبذة تعريفية عن الباحثة:

Author Biodata Dr. Alchimaa Mohammed Waleed Lutfi is a researcher in Arabic Applied Linguistics. She received her PhD from Imam Mohammad Ibn Saud Islamic University in 2026 and her Master's degree in Applied Linguistics from Umm Al-Qura University in 2021. Her research interests focus on language teaching, language assessment, language policy, computational linguistics, and corpus linguistics.	بيانات الباحثة: الشيلاء محمد وليد لطفي ، باحثة في اللغويات التطبيقية العربية، حاصلة على درجة الدكتوراه من جامعة الإمام محمد بن سعود الإسلامية عام ٢٠٢٦م، حاصلة على درجة الماجستير في اللغويات التطبيقية من جامعة أم القرى عام ٢٠٢١م. تتركز اهتماماتها البحثية حول تعليم اللغة، الاختبارات اللغوية، السياسة اللغوية، واللسانيات الحاسوبية، والمدونات اللغوية.
معرف أوركيد (ORCID) : 0009-0003-8035-5716	
Email: alchimaalutfi@gmail.com	



هذه الطبعة
إهداء من المجمع
ولا يسمح بنشرها ورقياً أو تداولها تجارياً

