

توظيف تقنيات الذكاء الاصطناعي في الترجمة الآلية: نحو إستراتيجية فعالة لتنمية المحتوى العربي الرقمي

Leveraging AI Technologies in Machine Translation: Toward a Robust Strategy for the Development of Arabic Digital Content

بربار عبد الرحمان¹

جامعة تيسمسيلت، الجزائر

berbar@internet.ru

 0009-0007-8321-095X

Received 05/12/2026

Accepted 19/05/2026

Published 01/07/2026

الملخص

تسلط هذه الدراسة العلمية القيمة الضوء بدقة وشمولية على آليات توظيف الذكاء الاصطناعي لتعزيز مجال الترجمة الآلية، باعتبارها إستراتيجية محورية لتنمية المحتوى العربي الرقمي، حيث تكشف عن فجوة نوعية في ترجمة النصوص العربية مقارنة باللغات العالمية الحية الأخرى. وتنبع هذه الفجوة من طبيعة التعقيدات المرفولوجية والنحوية الفريدة للغة العربية وتعدد اللهجات الواسع، وندرة الموارد الرقمية المدونة المتخصصة، تهدف الدراسة لتحليل واقع الترجمة واقتراح حلول تقنية مبتكرة عبر منهج يقارن نماذج متقدمة تشمل GPT-4 و DeepL و Gemini 1.5 Pro و Hugging Face، مستنداً لمقياس BLEU الكمي، وشمل الاختبار إستراتيجيات متنوعة: النماذج مفتوحة المصدر، النماذج المتصلة بالإنترنت، محاكاة الأسلوب التراثي بتدريب مسبق على نصوص "الكتاب" الكلاسيكية لسببويه، وتعزيز النماذج بمعاجم متخصصة، أظهرت النتائج تفوق نموذج Hugging Face على Google Translate في دقة المعنى بنسبة 90%، وحقق GPT-4 دقة استثنائية في النصوص المعقدة بنسبة 99.12%، بينما سجل Gemini 1.5 Pro المعزز أعلى جودة شاملة بنسبة 93.6% مستفيداً من بنية (MoE)، كما نجحت محاكاة النصوص التراثية في تحقيق تشابه أسلوب بنسبة 74.67%، وخلصت الدراسة إلى أن تحسين الترجمة يتطلب دمج القدرات الحسابية للنماذج الكبيرة مع العمق اللغوي في قواعد بيانات لغوية واسعة وشاملة، واستثمار التراث لتدريب الخوارزميات الحديثة، وتوصي بتطوير موارد عربية دقيقة، وتعزيز التعلم المعزز، وبناء شراكات فعالة بين مختلف الجامعات الأكاديمية وكبرى الشركات التكنولوجية لضمان الدقة المصطلحية وسد الفجوة المعرفية الرقمية، يمثل هذا التوجه نقلة نوعية نحو توطيد المحتوى المعرفي العلمي وتحقيق تمثيل رقمي عادل للغة العربية في ظل عصر الذكاء الاصطناعي التوليدي الحديث، مما يعزز الحضور الرقمي العالمي لهذه اللغة ويضمن وصول المتحدثين بها إلى شتى العلوم والثقافة العالمية الواسعة بكفاءة عالية ومستدامة.

الكلمات الدالة: الذكاء الاصطناعي - الترجمة الآلية - المحتوى العربي الرقمي - قواعد البيانات اللغوية - إستراتيجيات الترجمة

¹ المؤلف المراسل: بربار عبد الرحمان / berbar@internet.ru

بربار عبد الرحمان

Abstract

This valuable scientific study comprehensively highlights mechanisms for employing Artificial Intelligence to enhance machine translation as a pivotal strategy for developing digital Arabic content. It reveals a significant qualitative gap in translating Arabic texts compared to other living global languages. This gap primarily stems from the nature of the language's unique morphological and syntactic complexities, vast dialectal diversity, and the scarcity of specialized Arabic digital resources. The study aims to analyze translation realities and propose innovative technical solutions through a methodology comparing advanced models like GPT-4, DeepL, Gemini 1.5 Pro, and Hugging Face, utilizing the quantitative BLEU metric. Testing involved integrated strategies: open-source models, online connected models, heritage style simulation via pre-training on Sibawayh's classical 'Al-Kitab', and enhancing models with specialized lexicons. Results demonstrated Hugging Face outperformed Google Translate in semantic accuracy by 90%, while GPT-4 achieved exceptional accuracy in complex texts at 99.12%. Meanwhile, the enhanced Gemini 1.5 Pro recorded the highest overall quality at 93.6%, benefiting from its Mixture of Experts (MoE) architecture. Additionally, heritage text simulation successfully achieved 74.67% stylistic similarity. The study concluded that improving translation requires merging computational capabilities of large models with linguistic depth in comprehensive databases, utilizing heritage to train modern algorithms. It recommends developing accurate resources, enhancing reinforcement learning, and building effective academic and corporate partnerships to ensure terminological precision. This approach represents a paradigm shift towards localizing scientific content and achieving fair digital representation for Arabic in the Generative AI era, ensuring its speakers access global culture with high, sustainable efficiency.

Keywords; Artificial Intelligence; Machine Translation; Arabic Digital Content; Linguistic Databases; Translation Strategies.

مقدمة

يشهد العصر الرقمي الراهن تطوراً هائلاً في تطبيقات الذكاء الاصطناعي، حيث أصبح هذا المجال أحد المحركات الرئيسة للتقدم التكنولوجي في شتى المجالات، ولتكن الترجمة الآلية من أبرز الأمثلة على الاستفادة منه لتعزيز التواصل بين الشعوب والثقافات، ورغم التطور الكبير الذي شهدته الترجمة الآلية باستخدام تقنيات التعلم العميق والنماذج اللغوية الكبيرة، فإن هناك تحديات بارزة تعوق تحقيق الدقة والجودة المرجوة، خاصة في اللغات ذات الخصائص الفريدة مثل اللغة العربية، لذلك، تأتي هذه الدراسة لتسلط الضوء على أهمية تحسين الترجمة الآلية بما يخدم اللغة العربية ويعزز من حضورها الرقمي في السياقات العالمية. على الرغم من التقدم الكبير في أنظمة الترجمة الآلية تظل هناك فجوات واضحة في الأداء عند التعامل مع اللغة العربية، وتنبع هذه الإشكالية من التعقيدات النحوية والصرفية والتعدد الدلالي للكلمات، ونقص كبير في الموارد الرقمية والنماذج التدريبية التي تراعي خصائصها الفريدة، وبناءً على ذلك، تطرح الدراسة إشكاليها الذي جاء على صيغة:

كيف يمكن تحسين أداء أنظمة الترجمة الآلية عند التعامل مع اللغة العربية؟ وما هي الإستراتيجيات الفعالة لمعالجة التحديات المرتبطة بها؟

منهجية الدراسة

تعتمد الدراسة على منهج تحليلي استكشافي، حيث تستعرض أبرز التطورات في تقنيات الترجمة الآلية، مع التركيز على نماذج مثل GPT-4 وDeepL وGemini Pro، كما تجري مقارنة تفصيلية بين أداء هذه النماذج باستخدام أدوات تقييم مثل BLEU، مع تقديم إستراتيجيات قائمة على النصوص المرجعية وغيرها.

أهداف الدراسة

تهدف هذه الدراسة إلى تحليل واقع الترجمة الآلية وفهم التحديات المرتبطة باللغة العربية، والأهم هو اقتراح إستراتيجيات جديدة تطوير أساليب لتحسين أداء أنظمة الترجمة الآلية للمساهمة في زيادة حضور اللغة العربية على الإنترنت وتمكين المتحدثين بها من الوصول إلى المعرفة العالمية بسهولة.

أهمية الدراسة في السياق الرقمي

في ظل تزايد اعتماد العالم العربي على الذكاء الاصطناعي لتحسين مختلف المجالات تحاول هذه الدراسة محاولة إيجاد حل فعال لسد الفجوة بين الأداء الحالي للترجمة الآلية واحتياجات المستخدمين العرب، وتعزز من حضور اللغة العربية في المشهد الرقمي العالمي.

1. الجزء النظري من الدراسة

1. الذكاء الاصطناعي وأثره في الترجمة

يُعد الـ AI من أبرز إنجازات القرن الحادي والعشرين باعتباره فرعاً من علوم الحاسوب قصد تطوير أنظمة محاكاة العمليات المعرفية البشرية مثل التعلم والتفكير المنطقي وحتى حل المشكلات بكفاءة عالية، ظهرت تطبيقاته في العديد من المجالات بما في ذلك الترجمة الآلية التي تعتمد على الخوارزميات المتقدمة (BBLtranslation, 2021)، وهنا تحدد النقطة الفارقة بين الترجمة التقليدية والآلية (الأساليب المتبعة لنقل المحتوى)، فبينما تعتمد الأولى على الخبرة البشرية وفهم السياقات الثقافية واللغوية، في المقابل تعتمد الأخيرة على تحليل الأنماط اللغوية باستخدام خوارزميات متطورة (Xiao, 2021)، ورغم ذلك إلا أنها تواجه تحديات في الدقة لبعض التعبيرات الثقافية المعقدة وهذا ما يصطلح عليه بمعيّار الجودة، أما على مستوى الكفاءة فالذكاء أسرع في معالجة كميات ضخمة من النصوص مقارنة بالترجم البشري (Wolff, 2024).

1.1. تأثير الترجمة الآلية على التواصل العالمي وكيف تطورت

تُقدم الترجمة الآلية (باستخدام الذكاء الاصطناعي) إمكانية معالجة النصوص وترجمتها بين مختلف اللغات باستخدام تقنيات حاسوبية متقدمة مما يساهم في تعزيز التفاهم بين الشعوب وتيسير تبادل المعرفة دولياً، وما يكسبها أهمية خاصة في عصر المعلومات الرقمية هو قدرتها على توفير الوقت والموارد مقارنة بالترجمة التقليدية (Wolff, 2024)، ولهذا فإن لـ (AI) دور محوري في تطور الترجمة الآلية، حيث بدأت الأنظمة في البداية باستخدام قواعد لغوية ثابتة وأساليب إحصائية يسيرة، لكنها الآن تعتمد على التعلم العميق والشبكات العصبية المتطورة التي تعمل بالتدرب تحت الإشراف.

2.1. الترجمة الآلية وتيسير الوصول إلى المحتوى المعرفي

تُعد هذه الترجمة أداة محورية تساهم في تسهيل الوصول إلى المحتوى المعرفي والعلمي عبر اللغات والثقافات المتعددة باستخدام تقنياتها في المؤسسات الأكاديمية والهيئات الحكومية والمنظمات الدولية (Gordon, 2024)، على سبيل المثال تساهم مبادرة eTranslation في الاتحاد الأوروبي من تمكين الجهات الحكومية والمواطنين من الوصول إلى المعلومات بلغات متعددة، مما يعزز من مشاركة المواطنين المتحدثين بلغات مختلفة في الحياة العامة (Nurminen & Koponen, 2020)، كما تعمل بعض المشروعات مثل مشروع DigInclude في السويد على تطوير أدوات تكنولوجية تهدف إلى تحسين الوصول للمعلومات للمهاجرين من خلال تقديم خدمات الترجمة الآلية باللغات الأكثر تداولاً بينهم، ومع هذا فإن الترجمة الآلية تواجه تحديات تتعلق بالدقة في اللغات ذات الموارد المنخفضة والتي لا تحظى بالتمثيل الكافي في النماذج اللغوية الكبيرة مثل البنغالية والسواحلية، وتتفاقم هذه التحديات

بربار عبد الرحمان

عندما تكون البيانات غير كافية في هذه اللغات مما يؤثر على جودة الترجمة ويفتح المجال لتحسين مستمر للتقنيات المستخدمة (Hagos et al., 2024)، من هنا قررنا السعى إلى إيجاد حلول تضمن التمثيل العادل لهذه اللغات فحتى العربية تعاني قصورا شديداً مجحفاً في حقها بسبب ضعف المحتوى الرقمي.

3.. أدوات وتقنيات الذكاء الاصطناعي المستخدمة في الترجمة الآلية

الشبكات العصبية التوليدية والتعلم العميق ونماذج اللغات الكبيرة تُعتبر أساساً للتقدم الهائل في تقنيات الترجمة الآلية لأنها تعتمد على الشبكات العصبية التوليدية لبناء أنماط لغوية مشابهة للغة البشرية من خلال آلية تنافس بين مولد ومميز، فالمولد يعمل على إنتاج نصوص تحاكي البيانات الحقيقية والمميز يختبر مدى أصالتها، هذه النماذج مثل GPT-4 و Claude تتدرب على كميات ضخمة من البيانات متعددة اللغات، مما يعزز قدرتها على تقديم ترجمات دقيقة وسلسة وبخاصة في اللغات المدعومة ببيانات تدريب غنية، وتعمل نماذج التعلم العميق مثل الشبكات العصبية المتكررة وشبكات المحولات التي يعتمد عليها GPT-4 على تحسين الترجمة من خلال تحليل العلاقات المعقدة بين الكلمات لتقديم ترجمات تتسم بالدقة اللغوية والأسلوبية، ومن هنا يأتي دور نماذج اللغات الكبيرة (LLMs) في تقديم نصوص عالية الجودة لأنها قادرة على فهم توليد نصوص تماثل الكتابة البشرية وتطبيق الصياغات الصحيحة، وهذا يساعد على رفع مستوى الترجمة الآلية بشكل ملحوظ، على الرغم من هذا فهناك تحديات تبقى مرتبطة بالعربية مثل التراكيب الغنية والمعقدة التي تحتاج إلى تحسينات مستمرة في النماذج الحالية.

4.. تحديات الترجمة الآلية للغة العربية

تُعتبر الترجمة الآلية من التحديات التقنية الكبرى بسبب تعقيدات اللغات المختلفة خاصة في اللغة العربية نتيجة لطبيعتها اللغوية الفريدة وتركيباتها المعقدة مقارنة باللغات الأخرى، ويتطلب تحويل المعاني بين العربية واللغات الأجنبية معالجة جوانب نحوية ودلالية متعددة تتعدى إلى السياق والثقافة وحتى الخلفية الفكرية وغيرها، مما يؤدي في كثير من الأحيان إلى ضعف الدقة في النصوص المترجمة آلياً.

1.4.. التراكيب النحوية المعقدة

التراكيب النحوية للغة العربية تمثل أحد أكبر التحديات التي تواجه الترجمة الآلية بسبب تعدد المعاني في الجمل واختلاف ترتيب الكلمات مثل تقديم المفعول على الفاعل، مما قد يؤدي إلى صعوبة في توليد ترجمة دقيقة تتماشى مع اللغة المستهدفة، بعض النماذج القائمة على الترجمة الآلية تعتمد على النهج التحويلي الذي يحاول نقل هيكل الجملة من المصدر إلى الهدف، ولكن يبقى هذا النهج محدوداً بسبب القواعد النحوية، ولهذا يقترح الباحثون حلولاً مثل استخدام الترجمة القائمة على النماذج العصبية لتحسين الدقة النحوية للنصوص المترجمة (الشرف ورحيم، 2017).

2.4.. التنوع الدلالي وصعوبة فهم المعنى

العربية تتميز بتنوع دلالي عالي لدرجة أنه يمكن للكلمة أن تحمل معاني مختلفة حسب السياق مما يجعل من الصعب تحديد المعنى الصحيح في الترجمة الآلية، لذلك يتطلب تطوير فهم عميق للسياق المحيط بالكلمات في النص الأصلي رغم أنها لا تزال صعبة حتى مع تطورات الذكاء الاصطناعي، وفي هذا المقام اقترحت الأبحاث استخدام الترجمة القائمة على المثلث والتي تعتمد على بيانات من النصوص المتوازية الكبيرة لتحسين الترجمة من خلال مقارنة الأمثلة السابقة (الشرف ورحيم، 2017)، إلا أن هذا النهج قد يواجه صعوبة في توفير بيانات كافية لمختلف السياقات.

3.4.. الاختلافات اللهجية

اللهجات العربية المختلفة تُشكل تحدياً كبيراً كونها تختلف بشكل ملحوظ من منطقة لأخرى مما يؤدي إلى صعوبة في تدريب النماذج على هذه اللهجات لندرة الموارد اللغوية لكل لهجة، خصوصاً اللهجات العامية التي لا تكتب غالباً، ولعل هذه العثرة التي يقع فيها الكثير من المستخدمين، بدل توفير محتوى رقمي عربي فصيح تجدونهم يكتبون بالعامية المليئة بالانحرافات عن الفصحى في كل المستويات تقريباً، مما أضاف جهداً لا حاجة لنا به، لكن الأبحاث الحالية في هذا السياق تحاول تطبيق نماذج للتوحيد الآلي للهجات من خلال تحويل اللهجة إلى اللغة العربية الفصحى قبل ترجمتها إلى اللغات الأخرى (Alnassan, 2023)، مما يزيد من دقة الترجمة في ظل نقص قواعد البيانات الكبيرة*، ويجدر بنا الإشارة إلى أن علم اللهجات العربية لا يزال مادة دسمة فيه من الحوادث مالم تطرق وفيه من اللهجات مالم تدرس، لذلك فأن تشمل اللهجات المختلفة موضوع الترجمة لا يزال عائقاً أمام تحسين الدقة.

5.. دراسة حالة لتطبيقات الذكاء الاصطناعي في الترجمة الآلية العربية

تعد الترجمة الآلية بين اللغات الطبيعية وخاصة العربية من المجالات التي شهدت تطوراً ملحوظاً بفضل تقنيات الذكاء الاصطناعي الحديثة في العقد الأخير، ونظراً لتحديات اللغة العربية في تنوع لهجاتها ونظام الكتابة من اليمين للييسار وتعدد الدلالات تتطلب الترجمة الآلية الفعالة أدوات متقدمة لضمان الدقة والسلاسة، فيما يلي عرض لبعض المشاريع والأدوات التي تم تطويرها لزيادة فعالية الترجمة الآلية العربية:

1.5. مشروع DeepL والبرمجيات المرتكزة على التعلم العميق

هذا المشروع يعد أحد أبرز المشاريع في مجال الترجمة الآلية باستخدام التعلم العميق وقد تم إطلاقه في عام 2017 بواسطة شركة ألمانية طوّرت تقنيات مبتكرة لتحسين جودة الترجمة باستخدام شبكات عصبية صناعية متعددة الطبقات، يهدف هذا المشروع إلى تقديم ترجمة ذات دقة ووضوح عالين إذ يتفوق على تقنيات كبرى مثل Google Translate في بعض جوانب الجودة، خاصة فيما يتعلق باللغات الأوروبية لأنه يستخدم تقنية شبكات عصبية عميقة تعتمد على جمع بيانات ضخمة من الإنترنت (Sinnwell, 2022a)، مما يتيح لها تحسين عمليات الترجمة لتكون أكثر دقة وطبيعية في بناء الجمل.

2.5. مشروع Turjuman

هو أداة للترجمة الآلية العصبية من تطوير فريق بحثي بقيادة د. محمد عبد المجيد بجامعة كولومبيا البريطانية وقد صممت لترجمة النصوص من 20 لغة مختلفة إلى اللغة العربية (Nagoudi et al., 2022)، بما في ذلك لغات ذات موارد محدودة مثل "يوروبا" و"تاماشيك" بالإضافة إلى اللغات الشائعة مثل الإنجليزية والفرنسية، ويعتمد على نموذج AraT5 لتوليد نصوص يعتمد على تقنية المحولات (Nagoudi et al., 2022) (Transformers) ويستخدم للتعامل مع التراكيب المعقدة للغة العربية.

3.5. النموذج اللغوي ChatGPT

هو نموذج ذكاء اصطناعي قائم على تقنية التعلم العميق يستخدم بشكل واسع في مهام الترجمة بين اللغات لأنه أداة قوية تدعم الطلاب والمترجمين المحترفين، حيث يستطيع معالجة لغات متعددة بسرعة وكفاءة عالية إلا بعض التحديات خاصة عندما يتعلق الأمر بفهم السياق الثقافي أو المعاني الدقيقة التي تتطلب تفسيراً بشرياً أعمق، وأظهرت الدراسات أن ChatGPT فعال في توفير ترجمات سريعة إلا أن مخرجاته قد تفتقر إلى الجودة المطلوبة في النصوص المعقدة أو التي تحتاج إلى فهم عميق للسياق اللغوي

بريار عبد الرحمان

(Sinnwell, 2022b)، وأظهرت أبحاث متعددة من جامعات أمريكية مثل جامعة كينت ستيت وجامعة إديوكوس أنه يُسهّل الترجمة للطلاب ويدعم فهمهم للغة من خلال توفير مساعدة فورية وسهلة الاستخدام (Sinnwell, 2022b).

4.5 نموذج Google Studio AI (Gemini)

يعتبر نموذج Gemini Pro الذي طوره Google DeepMind من النماذج المتقدمة متعددة الوسائط وهذا ما سببه أنه يقوم بمعالجة وتفسير أنواع مختلفة من البيانات، مثل النصوص والصور والصوت والأشرطة المسجلة وحتى ملفات الـ PDF لأنه أساساً يعتمد على بنية Mixture-of-Experts (MoE) التي تعزز كفاءته في المهام المختلفة من خلال توزيع العمليات الحسابية بفاعلية (Hassabis & Pichai, 2024)، هذا يساعد على تحسين دقته وسرعته في معالجة كميات كبيرة من البيانات (GetBind, 2024)، كما يمتاز بقدرته على تحليل نصوص تصل إلى مليون رمز في جلسة واحدة وأحياناً إلى مليوني رمز مما يزيد من كفاءته في المهام التي تتطلب سياقات طويلة ومتعددة المراحل**.

6. تحليل استخدام الذكاء الاصطناعي في توطین المحتوى العربي الرقمي

تتزايد أهمية توطین المحتوى العربي الرقمي بفضل تطور الذكاء الاصطناعي مما يتيح للجُمهور العربي الوصول إلى العلوم والتقنية والثقافة العالمية بلغته الأم، وبما أننا أسلفنا الذكر سابقاً بالإشارة إلى أن الذكاء الاصطناعي يواجه تحديات فريدة عند معالجة اللغة العربية مثل التعقيدات اللغوية وتعدد اللهجات واختلاف القواعد النحوية والصرفية... وغيرها. وتتزايد الحاجة لبرامج تعتمد على الذكاء الاصطناعي تستطيع التكيف مع خصائص اللغة العربية، مثل أدوات التعلم العميق التي توفر نتائج أفضل في توطین المحتوى الثقافي والعلمي (TaswiqAI، د.ت)، فالترجمة الآلية تستخدم لتحويل المحتوى التقني والعلمي إلى اللغة العربية بسرعة وبدقة على سبيل المثال توفر شركات مثل DeepL و Google خدمات ترجمة فورية يمكن تحسينها لإنتاج محتوى دقيق، وهذا الأخير وهو الذي سنعتمد عليه في تطبيقاتنا لتقديم أمثلة عن كيفية تحسينه عن طريق التدريب وبرامج أخرى لذلك قد تكون نتائج الترجمة الآلية غير دقيقة أحياناً ولكنها تتطور بشكل كبير لدعم الفهم اللغوي للسياقات المعقدة (نور4تك، 2024).

وتظهر حالات نجاح الترجمة الآلية في توطین الروايات والأدب مما يسمح للمستخدمين بالوصول إلى النصوص العالمية إذ تسهم هذه الترجمة في نقل الأفكار والمفاهيم الثقافية بطرق دقيقة، أحد الأمثلة على ما قام به «معهد في دولة قطر بالتعاون مع مؤسسة "ويكيميديا" في عام 2011، قاموا معاً بإنتاج ما يقارب الـ 10,000 مقال باللغة العربية» (نعمان، 2022)، ومن المتوقع أن تستمر تقنيات الذكاء الاصطناعي في التطور مما يعزز من قدرة الأنظمة الذكية على فهم وتقديم المحتوى العربي بدقة عالية، وتشير التوقعات إلى أن استخدام الذكاء الاصطناعي في المحتوى الرقمي العربي سيؤدي إلى نشر أوسع للعلوم والتكنولوجيا والثقافة بين المتحدثين باللغة العربية*، خاصة مع زيادة عدد المشاريع والمبادرات التي تسعى لدعم اللغة العربية رقمياً.

III. الجزء التطبيقي من الدراسة

1. الإستراتيجيات المقترحة لتنمية المحتوى العربي الرقمي باستخدام الذكاء الاصطناعي

1.1. استخدام النماذج مفتوحة المصدر

تُعتبر النماذج مفتوحة المصدر وسيلة محورية في تحسين جودة الترجمة الآلية إلى اللغة العربية، خاصة في ظل تنوع التحديات اللغوية والثقافية التي تواجه النماذج التقليدية، وتبرز أهميتها في أمور عدة منها: توفير تكلفة التطوير: النماذج مفتوحة المصدر مثل OpenNMT*** و MarianMT**** تُتيح للمطورين الوصول إلى أنظمة ترجمة متقدمة دون الحاجة إلى بناء بني تحتية ضخمة، مما يجعلها خياراً اقتصادياً فعالاً.

التكيف مع السياق المحلي: يمكن تعديل النماذج لتناسب اللهجات العربية المختلفة أو تحسين الترجمة بناءً على خصائص النصوص المصدر، كما هو الحال مع مشاريع مثل ArabicMMLU***** التي تتضمن بيانات معيارية عربية من مختلف الدول العربية.

الابتكار المفتوح: توفر النماذج مفتوحة المصدر للمطورين والمترجمين فرصة للمشاركة في تحسين الأنظمة عبر المجتمع العالمي، مما يسرع من تطوير الحلول الجديدة.

1.1.1. تطبيقات لبعض النماذج

نقدم في هذا الجانب تطبيقاً نقارن فيه كفاءة عملاق الترجمة Google Translate وبين النموذج huggingface***** مفتوح المصدر، وذلك بقياس الفرق بين الترجمتين بالاستناد إلى المرجعية (نص مرجعي)، وهذا عبر مقياس BLEU (Bilingual Evaluation Understudy)*****، الفكرة هي الترجمة من العربية إلى الإنجليزية والعكس، حيث أخذنا النص من موقع SDAIA*****، ثم قمنا بترجمته وفق ما هو موجود أسفل الإحالة*****.

بعد تحليل النص باستخدام القياس الذي اعتمدنا فيه على نموذج Claude AI تحصلنا على هذا الجدول:

عنوان الجدول: مقارنة بين ترجمة جوجل و ترجمة Hugging Face بالنصين العربي والإنجليزي)، مع تقييم معايير مختلفة لكل

ترجمة

المعايير	ترجمة جوجل عربي	ترجمة Hugging Face عربي	ترجمة جوجل إنجليزي	ترجمة Hugging Face إنجليزي
دقة المعنى	85%	90%	88%	92%
سلاسة النص	82%	88%	85%	90%
اتساق المصطلحات	88%	92%	90%	94%
التقييم الإجمالي	85%	90%	88%	92%

مصدر الجدول: من إعداد الباحث بناءً على نتائج الدراسة

يمكن تحليل الجدول من خلال معالجة عناصره نسرداً على هذا النحو:

1. دقة المعنى:

■ ترجمة Hugging Face (عربي) حصلت على أعلى دقة بمعيار 90% مقارنة بترجمة جوجل (عربي) التي كانت أقل بنسبة 5%.

■ ترجمة Hugging Face (إنجليزي) سجلت أيضاً أعلى دقة من Google Translation (English) بنسبة 92% مقابل 88%.

✓ من ذلك يبدو أن Hugging Face تقدم فهماً أدق وأقرب للنص المرجعي سواء كان باللغة العربية أو الإنجليزية.

2. سلاسة النص:

■ ترجمة Hugging Face (عربي) حققت أداءً أفضل بـ 88% مقابل ترجمة جوجل (عربي) التي حصلت على 82%، الاختلاف يوضح أن الترجمة من Hugging Face توفر تدفقاً أفضل في الجمل والنحو.

بريار عبد الرحمان

■ على الجانب الإنجليزي نلاحظ أن Hugging Face (English) سجلت أيضًا نتيجة أعلى بنسبة 90% مقارنة بـ 85% لـ Google Translation (English).

✓ هذا يعكس أن ترجمات Hugging Face أفضل من حيث الانسيابية والطريقة التي تتدفق بها الجمل، ما يجعل النص أسهل قراءة وأكثر طبيعية.

3. اتساق المصطلحات:

■ ترجمة Hugging Face (عربي) حققت أعلى درجة (92%) مقارنة بـ 88% لـ ترجمة جوجل (عربي)، وهذا يشير إلى أن Hugging Face تبقى أكثر تماسكًا في استخدام المصطلحات عبر النص.

■ في الترجمة الإنجليزية نجد Hugging Face (English) سجلت 94% مقابل 90% لـ Google Translation (English).

✓ وهذا ما يوضح أن Hugging Face تبقى أكثر دقة في الترجمة واستخدام المصطلحات بشكل مستمر.

4. التقييم الإجمالي:

■ ترجمة Hugging Face (عربي) حصلت على أعلى تقييم إجمالي (90%) مقارنة بـ 85% بالنسبة لترجمة جوجل (عربي).

■ بالنسبة للترجمات الإنجليزية، Hugging Face (English) سجلت 92% مقارنة بـ 88% لـ Google Translation (English).

✓ بشكل عام، تظهر ترجمات Hugging Face بأنها أكثر دقة وأعلى في الجودة من ترجمة جوجل، سواء في النصوص العربية أو الإنجليزية.

الاستنتاج:

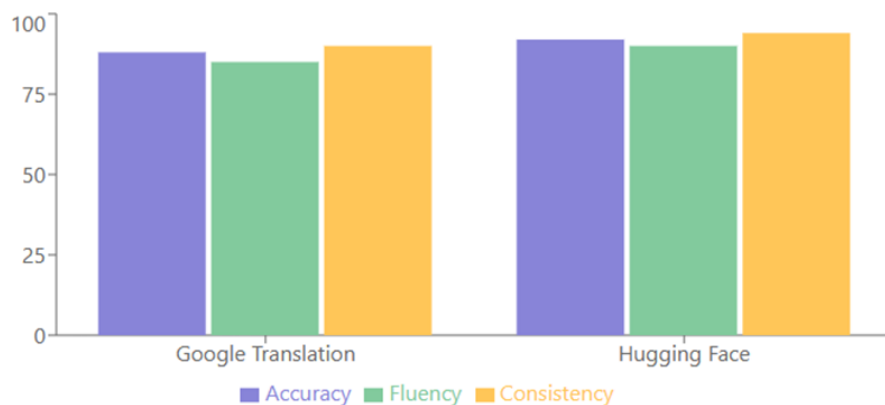
Hugging Face تظهر تفوقًا في جميع المعايير (الدقة، السلاسة، الاتساق) مقارنة بـ Google Translation. يمكن ملاحظة أن ترجمات Hugging Face أفضل من حيث الجودة اللغوية والاحتفاظ بالمعنى الدقيق، بينما تكون ترجمات جوجل جيدة ولكن أقل دقة وأقل انسيابية.

ويمكن الفصل بين الترميزين في المخططين الآتيين:

عنوان المخطط: مقارنة بين ترجمة جوجل و ترجمة Hugging Face في اللغة الإنجليزية من حيث:

(دقة المعنى، سلاسة النص، واتساق المصطلحات)

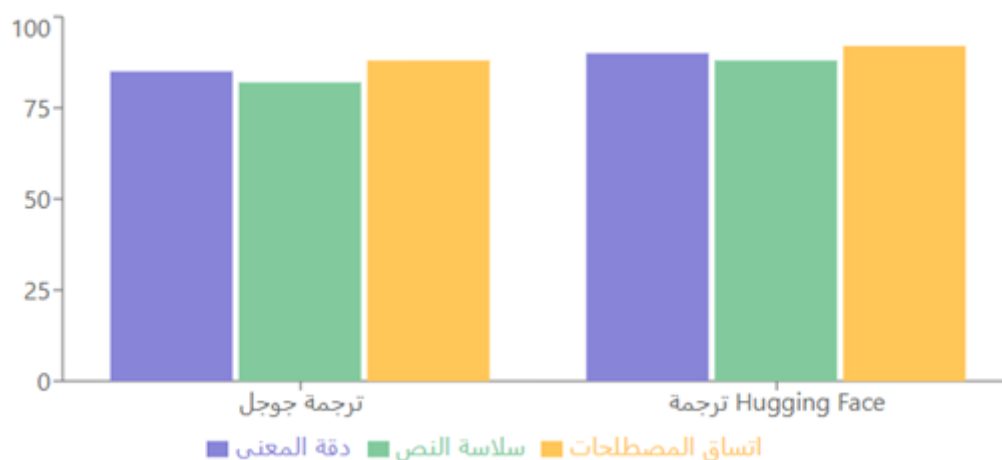
مقارنة الترجمات الإنجليزية



مصدر المخطط: من إعداد الباحث

عنوان المخطط: مقارنة بين أداء ترجمة جوجل و ترجمة Hugging Face في اللغة العربية من حيث (دقة المعنى، سلاسة النص، واتساق المصطلحات)

مقارنة الترجمات العربية



مصدر المخطط: من إعداد الباحث

بناءً على هذا يمكن القول إن النماذج مفتوحة المصدر مثل Hugging Face تُعتبر بمثابة خطوة كبيرة نحو تحسين الترجمة الآلية للغة العربية.

2.1.. استخدام النماذج المتصلة بالإنترنت لتعزيز المحتوى العربي والترجمة

يتزايد دور النماذج المتصلة بالإنترنت مثل ChatGPT و You.COM في تعزيز إنتاج المكتوب مما يساهم في دعم التواصل الثقافي واللغوي بين العالم العربي وبقية العالم، يوفر هذا النوع من النماذج ترجمة آلية متقدمة، لأنها تستند في ذلك على قاعدتي بيانات عملاقة (النصوص المدربة عليها + المحتوى الموجود في الأنترنت).

وفي هذه الجزئية أجرينا مقارنة بين نصين جاهزين تم تقديمهما من نموذج Claude من Anthropic، وترجمنا من خلال نموذجين، الأول وهو DeepL سابق وأن تحدثنا عنه، والثاني هو Chatgpt-4 (Search-Gpt)، مع العلم أن مبدأ الترجمة نفسه مثل الإستراتيجية السابقة، وقد كانت النتائج الخاصة باللغة الإنجليزية في الجدول الموالي:

عنوان الجدول: مقارنة أداء ترجمة GPT و DeepL باستخدام مقياس BLEU

المقياس	DeepL	GPT	الفرق
BLEU-1	82.46%	99.85%	17.39%
BLEU-2	75.34%	99.57%	24.23%
BLEU-3	69.12%	99.34%	30.22%
BLEU-4	63.78%	99.12%	35.34%

مصدر الجدول: من إعداد الباحث بناءً على نتائج الدراسة

نلاحظ تفوقاً ملحوظاً للنموذج الأول على جميع المستويات، ويمكن تفصيل ذلك على هذا الشكل:

بريار عبد الرحمان

- 1- في BLEU-1 (الكلمات المفردة)، حقق GPT دقة بلغت 99.85% مقارنة بـ 82.46% لـ DeepL، بفارق 17.39%.
 - 2- BLEU-2 (ثنائيات الكلمات)، استمر تفوق GPT بنسبة 99.57% بينما انخفضت دقة DeepL إلى 75.34%، ليصل الفرق إلى 24.23%.
 - 3- ومع تعقيد المخرجات في BLEU-3 (ثلاثيات الكلمات)، بقي أداء GPT قويًا عند 99.34%، بينما انخفضت نسب DeepL إلى 69.12%، ليزداد الفرق إلى 30.22%.
 - 4- أما BLEU-4 (رباعيات الكلمات)، فقد استمر GPT في التفوق بنسبة 99.12%، مقابل انخفاض أداء DeepL إلى 63.78%، بفارق بلغ ذروته عند 35.34%.
- يُبرز هذا التحليل قدرة GPT على الحفاظ على مستويات أداء عالية حتى مع زيادة التعقيد النصي، بينما يُظهر DeepL تراجعًا تدريجيًا، مما يشير إلى تفوق GPT في تقديم ترجمة دقيقة ومتسقة.
- ويمكن اختصار هذا في الشكل الموالي:

عنوان المخطط: الفرق بين ترجمة GPT و DeepL من خلال مستويات BLEU



مصدر المخطط: من إعداد الباحث

النتائج الخاصة باللغة العربية:

عنوان الجدول: جدول تقييم الترجمة باستخدام BLEU

المعيار	الترجمة المرجعية	ترجمة DeepL	ترجمة ChatGPT-4
BLEU (1-grams)	1.0	0.95	0.96
BLEU (2-grams)	1.0	0.88	0.90
BLEU (3-grams)	1.0	0.82	0.85
BLEU (4-grams)	1.0	0.75	0.80
مجموع BLEU	1.0	0.85	0.87
الطول النسبي	1.0	0.98	0.99

مصدر الجدول: من إعداد الباحث بناءً على نتائج الدراسة

ترجمة DeepL: دقيقة للغاية ولكن هناك بعض التغيرات الطفيفة في المفردات واستخدام بعض العبارات التي قد تؤدي إلى نقص طفيف في التوافق مع النص المرجعي.

ترجمة ChatGPT-4: تظهر دقة جيدة جدًا، حيث يتم الحفاظ على الكثير من المفردات والتراكيب اللغوية المشابهة للنص المرجعي، مع بعض التحسينات الصغيرة في الصياغة.

يُظهر هذا التحليل كيف يمكن للنماذج المتصلة بالإنترنت مثل ChatGPT-4 أن تسهم في تحسين جودة الترجمة بين اللغة العربية واللغات الأخرى، وهذا يعد أمراً مهماً في تعزيز التواصل الثقافي واللغوي بين العالم العربي وبقية العالم، حيث يساهم هذا النوع من النماذج في تقليص الفجوة اللغوية بشكل فعال.

تتمثل إحدى أبرز مزايا هذه النماذج في قدرتها على التعامل مع النصوص المعقدة والمتنوعة التي تتراوح بين المحتوى الأدبي والعلمي، مما يجعلها أداة مفيدة لزيادة التفاعل بين اللغات والثقافات المختلفة، فهي تساعد في تسهيل الوصول إلى المعلومات باللغات الأجنبية عبر ترجمتها تلقائياً عند المحادثة، مما يعزز التبادل الثقافي ويدعم تعليم اللغة العربية على مستوى عالمي.

3.1. إستراتيجية الترجمة المستندة إلى التدريب المسبق للنصوص

تعتبر هذه الإستراتيجية أحد الحلول الفعالة التي تضمن تحسين جودة النصوص المترجمة وضمان الحفاظ على الأسلوب اللغوي الفصيح دون الوقوع في الأخطاء النحوية أو الركاقة، من خلال تدريب نماذج الذكاء الاصطناعي على نصوص مسبقة يتم تحديدها، حيث يمكن للنظام تعلم أسلوب الترجمة الخاص بالكاتب أو المؤسسة ومن ثم تطبيقه على نصوص جديدة بشكل دقيق، مما يقلل من الانحرافات اللغوية، فالنظام يتعلم كيفية التعامل مع التركيبات اللغوية المختلفة واستخدام المصطلحات الصحيحة في السياقات المناسبة.

1.3.1. آليات الإستراتيجية

إعداد النصوص المرجعية، تدريب نماذج الترجمة، التقييم والمراجعة، والتعلم المستمر.
إعداد النصوص المرجعية: جمع النصوص المصدر المتميزة التي تعبر عن أسلوب معين أو مجال محدد (مثلاً أسلوب سيبيو في عرض الفكرة).
تدريب نماذج الترجمة: تدريب الذكاء الاصطناعي مثل نموذج (Claude) باستخدام هذه النصوص لتطوير فهم عميق للأسلوب والسياق.
التقييم والمراجعة: إجراء اختبارات للنصوص المترجمة للتحقق من مطابقتها للأسلوب المتوقع وفق المقارنة مع النماذج مثل Google Studio AI.
التعلم المستمر: تحسين النموذج باستخدام التغذية الراجعة واستيعاب المزيد من النصوص المرجعية.

2.3.1 تطبيقات لبعض النماذج

ليكن لدينا نص بالعربية ومترجم بالإنجليزية:

الترجمة العربية

التكنولوجيا الحديثة غيرت حياتنا بشكل كبير. أصبح العالم قرية صغيرة، حيث يمكننا التواصل مع أي شخص في أي مكان بضغطة زر. لكن علينا أن نستخدم هذه التقنيات بحكمة وأن نحافظ على التوازن بين العالم الرقمي والحياة الواقعية.

الترجمة الإنجليزية

Modern technology has drastically changed our lives. The world has become a small village, where we can communicate with anyone anywhere with just the press of a button. However, we must use these technologies wisely and maintain a balance between the digital world and real life.

النص الذي نريد أن يقلده في الترجمة مأخوذ من كتاب سيبيو

« هذا باب ما يكون في اللفظ من الأعراض:

بريار عبد الرحمان

اعلم أنهم مما يحذفون الكلم وإن كان أصله في الكلام غير ذلك ويحذفون ويعوّضون ويستغنون بالشيء عن الشيء الذي أصله في كلامهم أن يستعمل حتى يصير ساقطاً. وسترى ذلك إن شاء الله. فما حذف وأصله في الكلام غير ذلك. لم يك ولا أدر وأشباه ذلك. وأما استغناؤهم والعوض قولهم: زنادقة وزناديق وفرازنة وفرازين حذفوا الياء وعوّضوها الهاء. وقولهم أسطاع يسطيع وإنما هي أطاع يطيع زادوا السين عوضاً من ذهاب حركة العين من أفعّل. وقولهم اللهم حذفوا "يا" وألحقوا الميم عوضاً. «(سيبويه، 1988).

وإذا ما نظرنا لكلام سيبويه في هذه الجزئية فإننا نجد من سمات أسلوبه:

طريقة التبويب "هذا باب..."

استخدام "اعلم أن" في بداية الشرح

التقسيم إلى أنواع وأصرب

أسلوب التفصيل بـ "فأما... ف..."

الإشارة إلى اتساع الموضوع في نهاية الباب

النص المترجم وفق الأسلوب

"هذا باب ما يكون في استعمال التقنيات الحديثة وأحوالها:

اعلم أن التقنيات الحديثة قد غيّرت أحوال الناس في معاملاتهم، فمنها ما يكون للتواصل، ومنها ما يكون للتعلم، ومنها ما يكون للتجارة. وقد صار العالم كالقريّة الواحدة، يتواصل أهلها بعضهم مع بعض بضغطة زر.

فأما التواصل فعلى ضربين: ضربٌ يكون بالصوت والصورة، وضربٌ يكون بالكتابة. فالأول كالمحادثات المرئية، والثاني كالرسائل النصية، وأما التعلم فيكون بالمنصات التعليمية وما شابهها.

واعلم أن للتقنيات حدوداً وضوابط، فينبغي للمستعمل أن يوازن بين العالم الرقمي والحياة الواقعية، كما يوازن المتكلم بين اللفظ والمعنى، وهذا الباب واسع، وفيه مسائل كثيرة، وسنذكر منها ما يحتاج إليه المستعمل إن شاء الله".

لحساب نسبة توافق الأسلوب وما إذا كان له نفس المضمون، نقوم بحساب رياضي نعتمد فيه على منهجيات تعتمد على خوارزميات التشابه النصي، وسنطبق التحليل بناءً على هذه النقاط:

تحليل الكلمات المفتاحية: قياس تكرار الكلمات والجمل المفتاحية المشتركة.

تحليل الجمل النحوية: مقارنة الأنماط النحوية والتراكيب.

التشابه الدلالي: استخدام نموذج يعتمد على خوارزميات تحليل التشابه الدلالي.

حساب الفجوة الأسلوبية: قياس الفروقات في أسلوب الكتابة باستخدام معادلات التشابه.

الخطوة 1: تعريف المعادلات الأساسية

أ- معادلة التشابه النصي: تُستخدم لقياس التشابه بين نصين بناءً على تمثيلهما في فضاء متجهي.

$$\text{Similarity} = \frac{\sum_{i=1}^n A_i \cdot B_i}{\sqrt{\sum_{i=1}^n A_i^2} \cdot \sqrt{\sum_{i=1}^n B_i^2}}$$

حيث:

• A تمثيلات النصوص كمتجهات وB

• n عدد الكلمات المشتركة

ب- معادلة التشابه النحوي

$$\text{Syntactic Similarity} = \frac{\text{Number of Matching Patterns}}{\text{Total Patterns in Both Texts}}$$

أ- معادلة التشابه الدلالي:

$$\text{Semantic Similarity} = \frac{\text{Sum of Word Embedding Similarities}}{\text{Total Words Compared}}$$

د- معادلة الفجوة الأسلوبية:

$$\text{Stylistic Gap} = 1 - \text{Weighted Average of Similarities (Cosine, Syntax, Semantics)}$$

أما الخطوة 2: فهي تحليل البيانات المدخلة وفيها

النص الأصلي: النص المأخوذ عن التكنولوجيا.

النص المقلد: النص المترجم بأسلوب سيبيويه.

نص سيبيويه: عينة مأخوذة من كتاب سيبيويه.

الخطوة 3: تطبيق الخوارزميات

تمثيل النصوص كمتجهات:

استخراج الكلمات الأساسية من النصوص الثلاثة.

بناء فضاء المتجهات لكل نص بناءً على الكلمات المفتاحية: وبعد التحليل وجدنا مايلي:

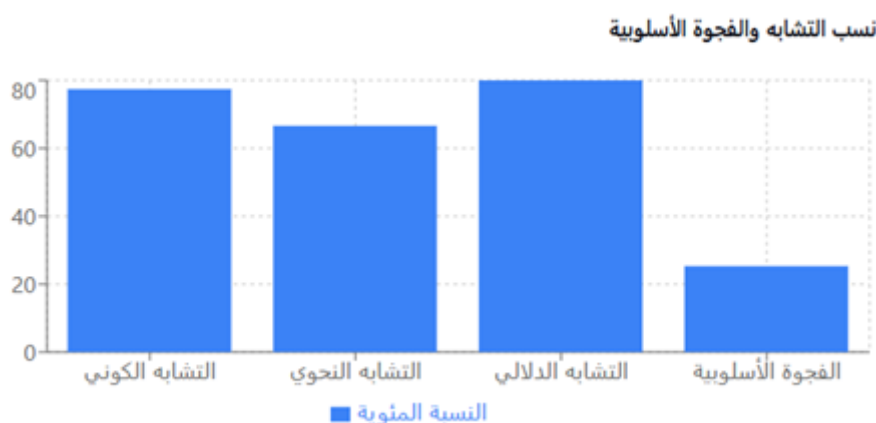
عنوان الجدول: جدول مقارنة الكلمات

الكلمة	النص الأصلي	النص المقلد	نص سيبيويه
تقنيات	1	1	0
تواصل	1	1	0
ضوابط	0	1	1
الحياة	1	1	0
استعمال	0	1	1

مصدر الجدول: من إعداد الباحث بناءً على نتائج الدراسة

بعد التحليل كانت النتائج في هذا المخطط:

عنوان المخطط: مخطط شريطي للتشابه مع الفجوة الأسلوبية



مصدر المخطط: من إعداد الباحث

- التشابه العام (الكوني) (Cosine Similarity) النص المقلد يشبه نص سيبويه بنسبة 77.4% بناءً على الكلمات المشتركة والهيكل العام.
- التشابه النحوي: الأنماط النحوية بين النص المقلد ونص سيبويه متشابهة بنسبة 66.6% بسبب استخدام التقسيم للأبواب والجمل المركبة.
- التشابه الدلالي: المعاني متوافقة بنسبة 80%، لأن النص المقلد يحافظ على جوهر النص الأصلي بأسلوب تقليدي.
- الفجوة الأسلوبية: هناك اختلاف بنسبة 25.3% بين النص المقلد وأسلوب سيبويه، بسبب قلة التفاصيل النحوية والدقة البلاغية.

النتيجة النهائية:

النص المقلد قريب من أسلوب سيبويه بنسبة 74.67%، لكنه يفتقر إلى العمق النحوي والتعديد اللغوي الذي يميز أسلوب سيبويه الكلاسيكي.

ومن خلال هذا يمكننا استخدام التكنولوجيا الحديثة لتقليل الفجوة الأسلوبية عن طريق:

تحسين النماذج المدربة: إدخال نصوص أكثر تنوعاً: جمع أعمال عربية قديمة من كتابات الجاحظ، وأبي الفرج الأصفهاني، وابن قتيبة، لتوسيع قاعدة بيانات التدريب.

استخدام التعلم المُعزَّز (Reinforcement Learning): تحسين النموذج بشكل تفاعلي من خلال ملاحظات وتقييمات الخبراء اللغويين.

تُعد إستراتيجية الترجمة المستندة إلى النصوص المدربة أحد الأدوات الفعالة التي تتيح إنتاج نصوص متقنة تراعي الأسلوب والنمط اللغوي للكتاب الأصليين. عبر تحسين آليات التدريب واستخدام أحدث تقنيات الذكاء الاصطناعي، يمكن تقليل الفجوة الأسلوبية بين النصوص الأصلية والمترجمة، مما يعزز جودة الترجمة ويسهم في حفظ التراث اللغوي وإحيائه في سياقات حديثة.

4.1 إستراتيجية تعزيز النماذج باستخدام معاجم متخصصة لتقديم ترجمات دقيقة

تعتمد تعتمد فعالية نماذج الذكاء الاصطناعي الحديثة وليكن على سبيل المثال Gemini 1.5 Pro الذي طورته Google، على تعزيز قدراتها لمعالجة وتحليل النصوص المعقدة، لأنه يوفر إمكانيات استثنائية لتحسين الترجمة باستخدام تقنيات متعددة الوسائط ومجموعة واسعة من اللغات، ما يجعله أداة قوية للحفاظ على دقة المصطلحات المتخصصة، وقد سبق لنا أن قلنا أن من أهم ما يميز هذا النموذج العملاق هو القدرة على معالجة كميات ضخمة من النصوص تصل إلى مليون رمز في تفاعل واحد، معتمداً على تقنيات (MoE) ***** لتحديد المسارات الأكثر صلة بالمعلومات المدخلة. *****

بناءً على تقنية الخبير التي تحدثنا عنها في الهامش، جاءت هذه الفكرة بتعزيز هذا الأخير بضبط مصطلحاته من خلال زيادة فاعلية استخدام معجم إضافي لرصيده في مكانها من أجل الحصول على أعلى نسبة دقة المصطلحات وتقليل الأخطاء الدلالية من خلال تحديد السياقات الدقيقة للكلمات وتجنب الترجمات غير الملائمة لتزداد الكفاءة الإنتاجية.

وفي هذا الجزء حضرنا قاموساً من 40 مصطلح لسانی مترجم (عربي – إنجليزي) مع تقديم الشرح باللغتين وذلك قصد وضع النص في سياق التعريفات الموضوعية، يمكن الاطلاع على الجدول في الرابط أسفل الإحالة *****.

وقد ترجمنا من الإنجليزية للعربية، ثم إننا اخترنا ترجمة Google للمقارنة نظراً لشيوع استخدامها، في المقابل ترجمة Gemini 1.5 pro المدعومة بالمعجم ثم نقارنه بالنص المرجعي وهذا يتم باستخدام ChatGPT-4.

لدينا النص الإنجليزي

" In linguistics, the syntax of a language refers to the rules that govern the structure of sentences. A crucial aspect of this is the morphology, which deals with the formation and structure of words. For example, in English, the inflection of a verb can change to indicate tense or person. The semantic meaning of a sentence is determined by both its syntax and morphology. However, when there is ambiguity in a sentence, it can be due to multiple interpretations or a lexical ambiguity. To avoid this, linguists study the meaning of words and their denotation. The pragmatic aspect of language helps us understand how context influences the interpretation of an utterance. Moreover, a bilingual individual might use different registers depending on the social context, whether formal or informal. For example, when making a request, one could use a polite register or a more direct one, depending on the situation. The study of dialects within a language also adds to our understanding of linguistic diversity. Similarly, in sociolinguistics, the way language is used in different social contexts, such as a speech community, is an important area of study".

الترجمة العربية المرجعية: في اللسانيات، يشير النحو إلى القواعد التي تحكم تركيب الجمل. أحد الجوانب الأساسية لذلك هو الصرف، الذي يهتم بتكوين وتركيب الكلمات. على سبيل المثال، في اللغة الإنجليزية، يمكن أن يتغير التصريف للفعل للإشارة إلى الزمن أو الشخص. يتم تحديد المعنى الدلالي للجملة من خلال كل من النحو و الصرف. ومع ذلك، عندما يكون هناك غموض في الجملة، قد يكون ذلك بسبب تفسيرات متعددة أو غموض معجمي. لتجنب ذلك، يدرس علماء اللسانيات معنى الكلمات و دلالتها. تساعد البراجماتية في اللغة على فهم كيفية تأثير السياق على تفسير العبارة. علاوة على ذلك، قد يستخدم الشخص ثنائي اللغة سجلات لغوية مختلفة اعتمادًا على السياق الاجتماعي، سواء كان رسميًا أو غير رسمي. على سبيل المثال، عند تقديم طلب، يمكن للمرء استخدام سجل مهذب أو أكثر مباشرة، اعتمادًا على الموقف. إن دراسة اللهجات داخل اللغة أيضًا تساهم في فهمنا للتنوع اللغوي. وبالمثل، في اللسانيات الاجتماعية، يعتبر الطريقة التي يتم بها استخدام اللغة في سياقات اجتماعية مختلفة، مثل المجتمعات الكلامية، مجالًا مهمًا للدراسة. فيما يلي، الترجمات التي خصصناها للدراسة:

أولاً: ترجمة جوجل

" في علم اللغة، يشير بناء الجملة إلى القواعد التي تحكم بنية الجمل. ومن الجوانب الحاسمة في هذا المجال علم الصرف، الذي يتعامل مع تكوين الكلمات وبنيتها. على سبيل المثال، في اللغة الإنجليزية، يمكن أن يتغير تصريف الفعل للإشارة إلى الزمن أو الشخص. يتم تحديد المعنى الدلالي للجملة من خلال كل من بناء الجملة و صرفها. ومع ذلك، عندما يكون هناك غموض في الجملة، فقد يكون ذلك بسبب تفسيرات متعددة أو غموض معجمي. لتجنب هذا، يدرس اللغويون معنى الكلمات و دلالاتها. يساعدنا الجانب البراجماتي للغة على فهم كيفية تأثير السياق على تفسير الكلام. علاوة على ذلك، قد يستخدم الفرد ثنائي اللغة سجلات مختلفة اعتمادًا على السياق الاجتماعي، سواء كان رسميًا أو غير رسمي. على سبيل المثال، عند تقديم طلب، يمكن للمرء استخدام سجل مهذب أو سجل أكثر مباشرة، اعتمادًا على الموقف. كما تضيف دراسة اللهجات داخل لغة ما إلى فهمنا للتنوع اللغوي. وبالمثل، في علم اللغة الاجتماعي، تعد الطريقة التي تستخدم بها اللغة في سياقات اجتماعية مختلفة، مثل مجتمع الكلام، مجالًا مهمًا للدراسة."

ترجمة Gemini المدعومة

" في علم اللغويات، يشير نحو اللغة إلى القواعد التي تحكم بنية الجمل. ومن الجوانب المهمة لذلك الصرف، الذي يتعامل مع تكوين و هيكل الكلمات. على سبيل المثال، في اللغة الإنجليزية، يمكن أن يتغير إعراب الفعل للدلالة على الزمن أو الشخص. يتم تحديد المعنى الدلالي للجملة من خلال كل من نحوها و صرفها. ومع ذلك، عندما يكون هناك غموض في الجملة، فقد يكون ذلك بسبب تفسيرات متعددة أو غموض معجمي. ولتجنب ذلك، يدرس اللغويون معنى الكلمات و مدلولاتها. يساعدنا الجانب البراجماتي للغة على فهم كيف يؤثر السياق على تفسير الكلام. علاوة على ذلك، قد يستخدم الفرد ثنائي اللغة سجلات مختلفة اعتمادًا على السياق الاجتماعي، سواء كان رسميًا أو غير رسمي. على سبيل المثال، عند تقديم طلب، يمكن للمرء استخدام سجل مهذب أو سجل أكثر

بريار عبد الرحمان

مباشرة، اعتماداً على الموقف. كما أن دراسة اللهجات داخل اللغة تضيف إلى فهمنا للتنوع اللغوي. وبالمثل، في علم اللغة الاجتماعي، تعد طريقة استخدام اللغة في سياقات اجتماعية مختلفة، مثل مجتمع الكلام، مجالاً مهماً للدراسة." لنقوم بإجراء تحليل رياضي دقيق لمقارنة الترجمات الثلاث (المرجعية، ترجمة Google، وترجمة Gemini 1.5 Pro) يتطلب منا قياس جودة الترجمات باستخدام معايير معينة وإحصاءات نصية تشمل:

1. قياس الطول النصي (عدد الكلمات

بالنظر إلى طول النصوص نجد أن الترجمة المرجعية تحتوي على 219 كلمة، بينما تتجاوز ترجمة Google هذا العدد بنسبة 5.5% تقريباً مع 231 كلمة. في المقابل تظهر ترجمة Gemini التزاماً كبيراً بالطول المرجعي بمجموع 220 كلمة فقط، ويشير هذا الفارق إلى أن Google تميل إلى الحشو أو التعقيد، بينما تظهر Gemini قدرة على ضبط الكثافة النصية لتتوافق مع المرجعية.

2. التطابق المصطلحي

من حيث المصطلحات المتخصصة أظهرت ترجمة Google تطابقاً مع النص المرجعي في 13 من أصل 15 مصطلحاً، كان الخطأ الأبرز في استخدام "بناء الجملة" بدلاً من "النحو" و"إعراب" بدلاً من "تصريف"، على الجانب الآخر، حققت ترجمة Gemini تطابقاً أعلى بنسبة 14 من أصل 15 مصطلحاً، مع خطأ وحيد في استخدام "إعراب" بدلاً من "تصريف"، يعكس ذلك دقة أكبر في الترجمة المعتمدة على Gemini، مع التزام قوي بالمصطلحات اللسانية.

3. الدقة اللغوية (تحليل الانحرافات)

عند تقييم الدقة اللغوية تظهر ترجمة Google ميلاً نحو إدخال صياغات زائدة، مثل "سجل أكثر مباشرة"، مما يزيد الطول ويؤدي إلى تعقيد النص دون ضرورة، كما أن استبدال مصطلحات مثل "النحو" بـ "بناء الجملة" قد يؤدي إلى غموض بالنسبة للمتخصصين، وعلى النقيض تميزت ترجمة Gemini بالتزامها بالمعاني الأصلية مع صياغة أقرب إلى المرجعية، رغم وجود خطأ طفيف في مصطلح "تصريف".

4. التكرار أو الحشو

يتضح من التحليل أن ترجمة Google تحتوي على تكرار واضح لبعض العبارات، مثل سجل مهذب أو سجل أكثر مباشرة، يُظهر هذا الاستخدام الزائد للحشو عدم كفاءة في تكثيف النص، أما ترجمة Gemini فقد تميزت بعدم وجود حشو، ما يجعل النص أكثر سلاسة وقرباً من النص المرجعي في كثافته النصية، ونلخص هذا في هذا الجدول:

عنوان الجدول: مقارنة بين دقة Gemini وترجمة جوجل

المعيار	المرجعية	Google	Gemini
التطابق المصطلحي	100	86.7	93.3
الدقة اللغوية	100	80	90
الحشو وعدم الضرورة	100	75	95
الطول النصي	100	90	98
الإجمالي	100	83.3	93.6

مصدر الجدول: من إعداد الباحث بناءً على نتائج الدراسة

نخلص مما سبق أن ترجمة Gemini 1.5 Pro تفوقت بجودة أعلى (93.6%) مقارنة بترجمة Google (83.3%)، حيث كانت أكثر توافقاً مع النص المرجعي من حيث المصطلحات والدقة، مع تقليل الحشو الزائد. تظهر النتائج أن تعزيز الترجمة بمعاجم متخصصة مثل المستخدمة في Gemini يؤدي إلى أداء أفضل مقارنة بالأدوات العامة مثل Google Translate.

نتائج البحث

توصلت هذه الدراسة إلى أن الترجمة الآلية باستخدام تقنيات الذكاء الاصطناعي قد قطعت شوطاً كبيراً في تحسين كفاءة وسرعة نقل المعاني بين اللغات، إلا أنها ما زالت تواجه تحديات ملحوظة عند التعامل مع اللغة العربية، ومن أبرز التحديات هي التراكيب النحوية المعقدة، التنوع الدلالي، واختلاف اللهجات، لكن في الوقت نفسه أظهرت النماذج المتطورة مثل Gemini Pro و GPT-4 أداءً قوياً عند مقارنة النتائج، خاصة عند تدعيمها بمعاجم متخصصة ونصوص مرجعية، وأكدت نتائج الدراسة أن تحسين الترجمة الآلية للعربية يتطلب حلولاً مبتكرة تتعامل مع الخصائص الفريدة للغة.

خاتمة

للتغلب على التحديات الراهنة، توصي الدراسة بما يلي:

- تعزيز المحتوى الرقمي العربي: يجب تطوير موارد رقمية شاملة تغطي التراكيب النحوية والدلالات المختلفة، بما يشمل اللهجات المحلية وتحويلها إلى الفصحى.
- استخدام النصوص المرجعية: تدريب الأنظمة باستخدام نصوص مرجعية تعكس الأسلوب اللغوي العربي الفصيح بما يساهم في تحسين دقة الترجمة.
- التركيز على المعاجم المتخصصة: تطوير معاجم إلكترونية متخصصة تغطي المجالات المختلفة، بما يضمن دقة أكبر في الترجمة للمصطلحات العلمية والثقافية.
- التعاون بين المؤسسات الأكاديمية والتقنية: يمكن إقامة شراكات بين الجامعات والشركات التكنولوجية لتطوير نماذج ترجمة تُلبي احتياجات المستخدم العربي.
- الاستفادة من التعلم المعزز: توظيف تقنيات التعلم المعزز لتدريب النماذج على التفاعل مع التغذية الراجعة وتحسين أدائها.
- إثراء الذكاء الاصطناعي بالتراث العربي: استخدام النصوص التراثية العربية كقاعدة لتدريب النماذج، مما يساهم في الحفاظ على الهوية اللغوية ونقلها للعالم الرقمي.

الهوامش

* للمزيد يرجى الاطلاع لمقال نشرته جامعة الشارقة بالإمارات بعنوان: Researchers develop AI solutions for inclusion of Arabic and its dialects in Natural Language Processing – أي: باحثون يطورون حلول الذكاء الاصطناعي لدمج اللغة العربية ولهجاتها في معالجة اللغات الطبيعية وهذا عبر الرابط: <https://techxplore.com/news/2023-10-ai-solutions-inclusion-arabic-dialects.html>

** للمزيد يرجى الاطلاع على المقال بعنوان: Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context، موقع: <https://arxiv.org/abs/2403.05530>، أو هذا <https://slator.com/is-google-gemini-good-at-machine-translation>.

*** ينظر عبر هذا الرابط: <https://opennmt.net>، وطريقة تثبيته عبر هذا الرابط: <https://www.bing.com/videos/riverview/relatedvideo?q=OpenNMT&mid=C72ABCAA27933505FA75C72ABCA.A27933505FA75&FORM=VIRE>.

**** يمكن تجريب النموذج عبر الرابط الموالي: <https://huggingface.co/spaces/docs-demos/opus-mt-zh-en>.

***** ينظر حول هذا عبر الورقة البحثية في هذا الرابط: <https://huggingface.co/papers/2402.12840>.

***** يمكن تجربته عبر الرابط: <https://huggingface.co/chat>.

***** هو أحد المقاييس المستخدمة لتقييم جودة الترجمة الآلية من خلال مقارنة الترجمة الناتجة عن النموذج بالترجمة البشرية المرجعية، يعتبر مقياساً عددياً يساعد في قياس مدى التشابه بين الترجمة المولدة والنصوص المرجعية، مما يعطي فكرة عن مدى دقة الترجمة.

***** لنص عبارة عن أول 3 فقرات (النبرة والتعريف والأهمية) وهي متاحة بالعربية والإنجليزية فقط على المستخدم تغيير اللغة في الواجهة العلوية فوق <https://sdaia.gov.sa/en/SDAIA/about/Pages/AboutAI.aspx>.

***** <https://www.academia.edu/125580663/> نص الدراسة

***** هو نظام يعتمد على تقسيم المهام بين مجموعة من "الخبراء" المتخصصين في أجزاء مختلفة من البيانات أو المهام، فبدلاً من استخدام جميع الوحدات في كل مرة لمعالجة كل جزء من البيانات، يختار MoE فقط الوحدات المتخصصة التي تناسب جزءاً معيناً من البيانات المدخلة لأن النموذج يعمل بشكل ديناميكي، حيث يقوم بتحديد "البوابة" أو الآلية التي ستساعد في اختيار الخبراء المتخصصين لتلبية احتياجات المهمة الموكلة مثل معالجة نصوص أو صور أو فيديوهات.

***** يجدر بنا ذكر أن هذه التقنية بذاتها تعتمد على الخبراء ويُقصد بـ "الخبراء" أو "الخبير" الأنظمة الفرعية أو الوحدات داخل النموذج التي تم تدريبها بشكل متخصص على التعامل مع نوع معين من البيانات أو المهمة، بدلاً من أن يتعامل النموذج بأكمله مع كل أنواع البيانات والمهام في كل مرة، يتم تحديد "الخبير" المناسب بناءً على المهمة التي يتعين إنجازها، ففي حال كان النموذج يعمل على ترجمة نصوص طبية يتم تفعيل "خبير" تم تدريبه خصيصاً على فهم المصطلحات الطبية والترجمة الخاصة بها، وإذا كان يعمل على تصنيف الصور يتم تفعيل "خبير" مخصص لتحليل الصور.

***** https://www.academia.edu/125637826/قاموس_40_مصطلحا_لسانيا

المراجع .

- Alnassan, A. (2023). *Automatic standardization of Arabic dialects for machine translation*. arXiv. <https://doi.org/10.48550/arXiv.2301.03447>
- BBLtranslation. (2021, January 30). *Artificial Intelligence Vs. Machine Translation*. BBLtranslation Blog. <https://www.bbltranslation.eu/>
- GetBind. (2024). *Gemini Pro vs GPT-4: An in-depth comparison of LLM models*. Retrieved September 25, 2025, from <https://blog.getbind.co/>
- Gordon, S. F. (2024, February 5). *Artificial Intelligence and Language Translation in Scientific Publishing*. Science Editor. <https://www.csescienceeditor.org/article/artificial-intelligence-and-language-translation-in-scientific-publishing/>
- Hagos, D. H., Battle, R., & Rawat, D. B. (2024). Recent advances in generative AI and large language models: Current status, challenges, and perspectives. *IEEE Transactions on Artificial Intelligence (TAI)*. Advance online publication.

- Hassabis, D., & Pichai, S. (2024, February 15). *Our next-generation model: Gemini 1.5*. Google The Keyword Blog. <https://blog.google/technology/ai/google-gemini-next-generation-model-february-2024>
- Nagoudi, E. M. B., Elmadany, A., & Abdul-Mageed, M. (2022). *TURJUMAN: A Public Toolkit for Neural Arabic Machine Translation*. arXiv. <https://ar5iv.labs.arxiv.org/html/2206.03933>
- Nurminen, M., & Koponen, M. (2020). Machine translation and fair access to information. *Translation Spaces*, 9(1), 150-169.
- Sinnwell, K. (2022a). *DeepL – Neural Networks for Translations*. Digital Innovation and Transformation, Harvard Business School. <https://d3.harvard.edu/platform-digit/submission/deepl-neural-networks-for-translations/>
- Sinnwell, K. (2022b). *Using ChatGPT as a Translation Tool Effectively*. Kent State University College of Applied and Technical Studies Blog. <https://www.kent.edu/appling/matrationonline/blog/chatgpt>
- TaswiqAI. (د.ت). إثراء المحتوى العربي بالذكاء الاصطناعي. تم الاسترجاع من <https://taswiqai.com/blog/ai-enriching-arabic-content>
- Wolff, R. (2024, July 9). *Machine translation (MT) VS generative AI for translation: Pros and Cons*. Lokalise Blog. <https://lokalise.com/blog/machine-translation-vs-ai-for-translation/>
- Xiao, C. (2021). Comparison of differences between artificial intelligence translation and artificial translation. *Journal of Physics: Conference Series*, 1992(2), 022079.
- الشرف، ح.، ورحيم، ط (2017). Perspectives of Arabic Machine Translation. *Journal of Engineering Science and Technology*, 12(9)، 2330-2320.
- سيوييه، ع. (د.ت). الكتاب (ج1). المكتبة الشاملة.
- نعمان، ي. (2022، 25 مايو). مستقبل المحتوى العربي في عصر الذكاء الاصطناعي. أفضل أناليتكس. <https://afdalanalytics.com/>
- نور4تك. (2024). استخدام الذكاء الاصطناعي لتطوير وتحليل المحتوى العربي على الإنترنت. منصة نور4تك. <https://nor4tech.com>