استخدام الذكاء الاصطناعي لتقييم جودة الترجمة: دراسة حالة تشات جي بي تي وترجمات النصوص القانونية *

تأليف: فاطمة أ. الغامدي وهند العتيبي
كلية علوم اللغات، جامعة الملك سعود، الرياض، المملكة العربية السعودية

ترجمة: رامي بوودن
المجمع الجزائري للغة العربية

ملخص المقال

أضحى توظيف الذكاء الاصطناعي في سياق تقييم جودة الترجمة مجالا بحثيا واعدا، إذ يسعى إلى استكشاف ما تختزنه هذه التكنولوجيا التحويلية من إمكانات في ميدان دراسات الترجمة عموما وما يترتب عن اعتمادها من انعكاسات على منظومة تكوين المترجمين وتأهيلهم. وتروم هذه الدراسة تقصي مدى التقارب بين تقييم الذكاء الاصطناعي لترجمات الطلبة القانونية وتقييم الأساتذة، مع تشخيص ما يكتنف استثمار الذكاء الاصطناعي في تقييم مخرجات الترجمة القانونية من مزايا ونقائص. وقد جمعت عشر ترجمات من امتحان منتصف الفصل من الإنجليزية إلى العربية بإغفال هويات أصحابها في إطار مقرر جامعي في الترجمة القانونية بإحدى الجامعات السعودية، ثم أعيد تقييم هذه الترجمات بتشات جي بي تي 4o،(ChatGPT-4o)، وذلك بعد توجيهه لرصد أخطاء الترجمة ومنح الدرجات استنادا إلى نفس سلم التقييم الذي اعتمده الأساتذة. كما أجريت مقارنة تحليلية دقيقة، فقرة بفقرة، بين تقييم الأداة

---

* **العنوان الأصلي للمقال:**

Alghamdi, F.A.; Alotaibi, H. Using AI in translation quality assessment: A case study of ChatGPT and legal translation texts. Electronics 2025, 14, 3893.
https://doi.org/10.3390/electronics14193893

الذكية والتقييم البشري، وصنفت فيها الأخطاء حسب أنماطها، وقيست درجة الاتساق بالمعالجة الإحصائية للدرجات للتحقق من وجود أي فروق مهمة. وقد أفضت النتائج إلى تسجيل مستوى مرتفع من التوافق بين تقييم تشات جي بي تي 4o وتقييم الأساتذة، في حين كشفت اختبارات العينات المزدوجة (t-test) عن غياب فروق ذات دلالة إحصائية بين درجات الطرفين ($p > 0.05$)، بينما اتسم الارتجاع الذي قدمه تشات جي بي تي 4o بالوضوح والدقة، مشتملا تعليلات للأخطاء وتصحيحات مقترحة. وعلى الرغم من أن هذه المعطيات تشجع على إدماج أدوات الذكاء الاصطناعي إدماجا فاعلا في تقييم جودة الترجمة ضمن سياقات تكوين المترجمين، فإن هنالك حاجة ماسة لتوظيف استراتيجي يوازن بين الأتمتة والحكم البشري المتبصر، وبالتالي، يمكن للذكاء الاصطناعي، متى أحسن تصميمه وتدريبه والإشراف عليه، أن يسهم إسهاما كبيرا في دعم بيداغوجيا الترجمة المعاصرة.

الكلمات الدالة: الذكاء الاصطناعي التوليدي، الترجمة الآلية، النماذج اللغوية الكبيرة، الذكاء الاصطناعي في التعليم، التعليم العالي، الترجمة القانونية.

1. مقدمة

لقد كان أداء النماذج اللغوية الكبيرة (Large Language Models - LLMs)[i] وتكنولوجيات الذكاء الاصطناعي (Artificial Intelligence - AI)[ii] أثناء تأدية مهام الترجمة لافتا في الآونة الأخيرة، خاصة في سياق توليد النصوص المترجمة [1،2،3]، إلا أن البحوث التي تناولت قدراتها على تقييم الترجمات ما زالت تتسم بالمحدودية، وتحديدا في سياقات تكوين المترجمين وتأهيلهم، ويبدو أن هذا المسار البحثي قد انبثق استجابة لجملة من الإكراهات العملية التي يواجهها الأساتذة، إذ يطلب منهم تقييم أعمال أعداد كبيرة من الطلبة، مع الحرص في الوقت ذاته على ضمان الإنصاف والاتساق والدقة اللغوية والملاءمة الثقافية والسلامة الوظيفية للنصوص المترجمة [4]. وتشير بعض الدراسات [5،6] إلى وجود درجة عالية من الارتباط بين تقييمات

الأدوات المعتمدة على الذكاء الاصطناعي وحكم الخبير البشري، وهو ما يفتح المجال أمام توظيف هذه الأدوات وجعلها وسائل موثوقة أثناء تقييم الترجمات، والغاية من ذلك تخفيف العبء الواقع على عاتق الأساتذة والحرص على تنسيق عمليات التقييم. كما أبرز باحثون آخرون (مثل [7]) ما يمكن أن تقدمه أدوات الذكاء الاصطناعي في سياق تنمية استقلالية المتعلم وصقل تفكيره النقدي في تعليم اللغات. وعلى الرغم من هذه الإسهامات فإن هنالك فجوة بحثية بينة تتصل باستخدام أدوات الذكاء الاصطناعي في تقييم الترجمة فيما نشر من دراسات علمية؛ فأولا، انصرفت معظم الدراسات إلى لغات مهيمنة عالميا مثل الإنجليزية والإسبانية والماندرين، مع إغفال لغات أقل حضورا في الموارد الرقمية كالعربية؛ وثانيا، لم تعن أي من هذه الدراسات بمهام الترجمة القانونية رغم ما يكتنف هذا المجال من خصوصيات وتحديات نوعية [5،6]؛ وثالثا، ركزت معظم البحوث على الأخطاء النحوية على حساب الاتساق القانوني والسياق الثقافي [8،9].

وبذلك تسعى هذه الدراسة إلى سد هذه الفجوة البحثية باستكشاف إمكانات توظيف أدوات الذكاء الاصطناعي في تقييم جودة الترجمة (Translation Quality Assessment - TQA)[iii] ضمن سياقات تكوين المترجمين القانونيين من الإنجليزية إلى العربية، ومحاولة الإجابة عن السؤالين البحثيين الآتيين:

1. إلى أي مدى يتقاطع تقييم الذكاء الاصطناعي للترجمات القانونية الذي يؤديه الطلبة في مقرر جامعي مكرس للترجمة القانونية مع تقييمات الأساتذة؟
2. ما أبرز الفوائد والتحديات المترتبة عن توظيف الذكاء الاصطناعي في تقييم مهام الترجمة القانونية ضمن سياقات تكوين المترجمين؟

سعت هذه الدراسة إلى رأب الفجوة القائمة في البحوث العلمية التي تناولت

توظيف الذكاء الاصطناعي في تقييم الترجمة، ولا سيما في مجال الترجمة القانونية من الإنجليزية إلى العربية، وقد انصبت غايتها على تقصي ما إذا كان بالإمكان اتخاذ التقارب بين التقييم القائم على الذكاء الاصطناعي وتقييم الأساتذة مؤشرا لموثوقية هذه الأدوات، فإذا ما ثبت أن تقييمات الذكاء الاصطناعي تتسق بانتظام مع أحكام الأساتذة، فإن ذلك من شأنه أن يدل على أهليتها لتكون أداة فاعلة وموثوقة في التقييم. كما هدفت الدراسة إلى تحري ما ينطوي عليه استخدام الذكاء الاصطناعي في تقييم الترجمة القانونية من مزايا وتحديات قد تعين أساتذة الترجمة وطلبتها على

إدماج التكنولوجيا إدماجا أنجع ضمن المقررات الدراسية. ويمكن أن تسهم نتائج هذه الدراسة في مساعدة مطوري أنظمة الذكاء الاصطناعي على تحسين أداء هذه النماذج اللغوية الكبيرة وزيادة الاستفادة منها في مختلف السياقات.

باختصار، سعت هذه الدراسة إلى استكشاف إمكانات أدوات الذكاء الاصطناعي، مثل تشات جي بي تي 4o، كونه أداة لتقييم الترجمة القانونية ضمن سياق تعليمي. ومن خلال تقويم مدى توافق أدائه مع تقييمات الأساتذة ورصد نقاط قوته ومواطن قصوره، هدفت هذه الدراسة إلى التحقق مما إذا كان بالإمكان توظيف هذه الأدوات وسيلة موثوقة لتقييم أعمال الطلبة في مجال الترجمة داخل الأوساط الأكاديمية. ومع استمرار تطور تكنولوجيات الذكاء الاصطناعي يظل من الضروري تحليل دورها في التعليم وفي الترجمة المهنية تحليلا متأنيا حتى تكون عاملا معززا للخبرة البشرية لا بديلا عنها.

يتميز ميدان الترجمة القانونية بتعقيده، فأداة التقييم القائمة على الذكاء الاصطناعي لا يكفيها أن تحسن المعالجة اللغوية فحسب، بل يتعين عليها فوق ذلك أن تفلح في استيعاب دقائق الاصطلاح القانوني ومراعاة مقتضيات السياق والوفاء بمتطلبات الدقة الوظيفية التي يفرضها هذا الصنف من النصوص، وإذا ما ثبتت موثوقية هذه الأدوات في تقييم الترجمة، سيمكن اعتمادها من تيسير إجراءات التصحيح ومد الطلبة

بارتجاع[iv] ذو قيمة تعليمية كبيرة، وإعانة الأساتذة على صون جودة التقييم وضبط اتساقه. أما إذا اعترت النتائج مظاهر تباين أو اختلال، فإن في ذلك ما يعين على تشخيص حدود التقييم المؤتمت[v] في مجال الترجمة، ويبرز المواضع التي تستدعي مزيدا من التطوير.

وبناء على ذلك، فإن هذه الدراسة تسهم في إثراء الأدبيات العلمية المتصلة بدور للذكاء الاصطناعي الآخذ في التحول في تكوين المترجمين، فمن خلال فحص اعتمادية أدوات التقييم القائمة على الذكاء الاصطناعي، مثل تشات جي بي تي 4o، وتقييم جدواها العملية، فقد قدمت الدراسة جملة من الرؤى التي ستحاول توجيه الممارسات البيداغوجية وتطوير أنظمة الذكاء الاصطناعي، وترسم معالم مستقبل مناهج تقييم الترجمة.

وقد نُظمت بقيت هذه الورقة على النحو الآتي: أولا، استعراض أحدث الدراسات في مجال الذكاء الاصطناعي وتقييم الترجمة وتحديات الترجمة القانونية من الإنجليزية إلى العربية، ثم يعرض الإطار المنهجي للدراسة وإجراءات التحليل المعتمدة فيها، ويلي ذلك تقديم النتائج المحصلة ومناقشتها، مع استنباط جملة من الدلالات التطبيقية، فيما سيختتم بخلاصة تعرض أبرز النتائج وتبين حدود الدراسة وتستشرف الآفاق الممكنة للبحوث المستقبلية.

2. مراجعة الأدبيات

تزايد عدد الدراسات الحديثة تزايدا ملحوظا التي تتناول موضوع تقييم النماذج اللغوية الكبيرة في مجالات معرفية مختلفة، ومن بينها الطب (مثل توظيف هذه النماذج للتحقق من المعطيات ومعالجة النصوص الطبية) والتجارة الإلكترونية (مثل تكييف صيغ الأوامر لتلائم الخصوصيات السياقية وتعزز الكفاءة الوظيفية) [10،11]. وتدل هذه الأعمال على توجه بحثي عام، يرمي إلى إخضاع النماذج اللغوية لاختبارات منهجية دقيقة بغرض فحص موثوقيتها وقدرتها على التكيف ومستوى

أدائها في بيئات متخصصة عالية المتطلبات. وفي ميدان دراسات الترجمة، أفضى إدماج الذكاء الاصطناعي إلى تحولات ملحوظة، وخصوصا فيما يتصل بتوليد الترجمات وتقييمها، ومع أن أدوات الترجمة المعززة بالذكاء الاصطناعي قد نالت اهتماما بحثيا واسعا، فإن النظر في النماذج اللغوية الكبيرة على أنها أدوات لتقييم جودة الترجمة ما يزال مجالا حديث النشأة نسبيا، وقد شرعت البحوث في هذا الإطار في التوجه إلى تقصي مدى دقة أساليب التقييم المؤتمتة واتساقها وموثوقيتها بمقارنتها بالتقييم البشري الاعتيادي وبالمقاييس المعتمدة في تقييم الترجمة. وانطلاقا من هذا المسار البحثي العام تأتي الدراسة الحالية لتسهم في معالجة مجال لم يحظ بعد بما يكفي من العناية، وهو يتمثل في تقييم جودة الترجمة القانونية، وبالتالي فإنها ترمي إلى توسيع آفاق البحث في مجال تقييم النماذج اللغوية الكبيرة ليشمل حقلا ذا أهمية بيداغوجية ومهنية بالغة. وسيخصص هذا القسم لتقديم عرض تحليلي للدراسات السابقة التي تناولت إمكانات الذكاء الاصطناعي في تقييم جودة الترجمة، مع تركيز خاص على النماذج القائمة على المحول التوليدي المدرب مسبقا (Generative Pre-trained Transformer - GPT)[vi].

1.2. الذكاء الاصطناعي في دراسات الترجمة

أحدث ظهور الذكاء الاصطناعي أثرا بالغا في حقل دراسات الترجمة، إذ أتاح آفاقا غير مسبوقة، وتلته في الوقت نفسه تحديات معقدة، وقد انصرف اهتمام الباحثين على نحو متزايد إلى تقصي الطرائق التي تعيد بها تكنولوجيات الذكاء الاصطناعي تشكيل ميدان الممارسات الترجمية وطرائق دراستها وأطرها المفاهيمية.

وقد أبرزت الدراسة المشار إليها في المرجع [12] الدور التحويلي للذكاء الاصطناعي في توسيع نطاق دراسات الترجمة وتطوير مناهجها البحثية، فمن خلال أدوات متقدمة في جمع البيانات ونقلها وتحليلها، يتيح الذكاء الاصطناعي فرصة سبر معطيات تاريخية تتصل بالمترجمين لم تكن متاحة من قبل عبر لغات وثقافات متعددة، وهو ما مكن الباحثين من إثراء المحتوى البحثي واستكشاف موضوعات جديدة داخل هذا الحقل. كما أن ما يمتلكه الذكاء الاصطناعي من قدرات في التعلم

الآلي (Machine Learning) ومعالجة اللغة الطبيعية (Natural Language Processing) والرؤية الحاسوبية (Computer Vision)[vii] يعين على إجراء تحليلات أسرع وأكثر شمولا، وهذا ينعكس مباشرة على مسار رفع كفاءة البحث العلمي، غير أن المرجع نفسه [12] نبه إلى جملة من الإشكالات الملازمة لهذه التطورات، كعدم اتساق البيانات وغياب المناهج المعيارية الموحدة والهويات الترجمية المبهمة في بعض السياقات، دون نسيان الإشكالات الأخلاقية المتزايدة. وتستدعي هذه التحديات اعتماد مقاربة أكثر تكاملا، تجمع بين مزايا الذكاء الاصطناعي والتحليل المعمق لسلوك المترجمين وممارساتهم وبيداغوجيات التكوين. ولتحقيق ذلك، اقترح المرجع [12] إطارا بحثيا جديدا يقوم على دمج المعارف المستمدة من حقول متعددة، ويشجع على التفاعل النقدي مع بيانات الذكاء الاصطناعي، ويعمل على بناء فضاء بحثي تعاوني ذي كفاءة رقمية ليساند الأدوار المتكاملة لكل من الإنسان والآلة في ميدان دراسات الترجمة دائم التطور.

وفي المنحى ذاته، تناول المرجع [3] الآثار الأشمل لتوظيف الذكاء الاصطناعي في الممارسة الترجمية، وقد برر قدرته على تحسين إتاحة خدمات الترجمة وتسريع وتيرة إنجازها وخفض كلفتها. كما أسهمت أدوات الذكاء الاصطناعي، وخصوصا أنظمة الترجمة الآلية العصبية (Neural Machine Translation – NMT)،[viii] في تعزيز التواصل العالمي بتمكين التفاعل متعدد اللغات آنيا، غير أن المقال نفسه نبه إلى حدود هذه التكنولوجيات، كقصورها عن الإحاطة بالتعابير الاصطلاحية والحساسية الثقافية والتفاصيل السياقية التي تظل الخبرة الإنسانية فيها عنصرا لا غنى عنه. وانطلاقا من ذلك، دعا المرجع [3] إلى تبني نماذج ترجمة هجينة تدمج بين الكفاءة التي توفرها تكنولوجيات الذكاء الاصطناعي والقدرات التأويلية للمترجم، في حين شدد على ضرورة تحسين البيانات التدريبية وإرساء أطر أخلاقية متينة لمعالجة ما قد تفضي إليه أنظمة الذكاء الاصطناعي من تحيزات أو تهديد محتمل للتنوع اللغوي.

وفي دراسة أنجزت في سياق مماثل، تناول المرجع [13] التحديات التي يواجهها أعضاء هيئة التدريس في الجامعات السعودية عند توظيف أدوات الذكاء الاصطناعي أثناء تدريس الترجمة، وقد اعتمدت الدراسة مقاربة نوعية، إذ وزع استبيان مكون من

14 بندا على عينة قصدية ضمت 50 أستاذا في الترجمة اختيروا من جامعات سعودية مختلفة خلال الفصل الدراسي الأول من سنة 2023، وأظهرت النتائج أن غالبية المشاركين عبروا عن مواقف محبذة لإدماج أدوات الذكاء الاصطناعي في تعليم الترجمة، وعلى الرغم من إقرارهم بوجود بعض التحديات فقد رأوا بأنها يمكن تجاوزها، وخصوصا في ظل التطور المتسارع لهذه الأدوات وتزايد أهميتها في الحقل. كما رأى أعضاء هيئة التدريس أن الذكاء الاصطناعي شرع في تقديم حلول مبتكرة لإشكالات طال أمدها في تدريس الترجمة. كما كشفت النتائج أن المستوى التعليمي كان عاملا ذا تأثير دال في تطورهم المهني وقدرتهم على التكيف مع أدوات الذكاء الاصطناعي، غير أن بعض الإجابات سلطت الضوء على مواطن قلق قائمة، إذ سجلت عبارات مثل «أستطيع استخدام أدوات الذكاء الاصطناعي دون توجيه خارجي» و«أستخدم أدوات الذكاء الاصطناعي كثيرا في التدريس» متوسطات منخفضة بلغت 2.80 و2.70 على التوالي، وهو ما يشير إلى وجود حاجة ملحة إلى مزيد من التكوين وإلى ضمان سهولة الاستخدام. وعلى المنوال نفسه، لم تتجاوز العبارة «أشرك الطلبة كثيرا في تطبيقات الذكاء الاصطناعي» متوسطا قدره 2.95، وهذا يدل على أن التطبيق الفعلي لهذه الأدوات ما يزال محدودا رغم المواقف المحبذة الظاهرة، ويرجح أن يكون ذلك راجعا إلى عوائق مثل ضيق الوقت، أو نقص الدعم المؤسسي، أو محدودية الاطلاع على هذه الأدوات.

وتبين هذه الدراسات مجتمعة أن الذكاء الاصطناعي لم يعد مجرد ابتكار تقني، بل غدا قوة فاعلة تعيد تشكيل النموذج الترجمي على مستويات البحث والممارسة والتعليم معا، فقد أبرز المرجع [12] ما أتاحه الذكاء الاصطناعي من توسيع للأدوات المنهجية وتعميق للتحليل في دراسات المترجمين، في حين ركز المرجع [3] على المنافع العملية التي يوفرها في واقع الممارسة الترجمية، مقرونة بما يثيره من قيود وإشكالات أخلاقية. واستكمالا لهذه الزوايا، كشف المرجع [13] عن تصورات أساتذة الترجمة وأنماط تفاعلهم مع أدوات الذكاء الاصطناعي في السياقات الأكاديمية، وسلط الضوء على الفجوة القائمة بين المواقف المحبذة المعلنة ومستوى التطبيق الفعلي. وعليه، دعت هذه الدراسات إلى اعتماد مقاربة متوازنة متعددة التخصصات تحسن توظيف

الكفاءة التي يتيحها الذكاء الاصطناعي مع ترسيخ الميزات البشرية الراسخة، وفي مقدمتها الحساسية الثقافية وصون التنوع اللغوي والمرونة البيداغوجية.

2.2. تحديات الترجمة القانونية

يطرح تقييم جودة الترجمة في السياقات القانونية جملة من التحديات المعقدة، إذْ تعود في جوهرها إلى الطبيعة اللغة القانونية وتداخل الأنظمة القانونية المتعددة، وما يعتري نصوصها من خصوصية ثقافية متجذرة [4]. وقد أبرزت دراسات حديثة هذه الإشكالات المتداخلة، مؤكدة الدور المحوري الذي ينهض به المترجمون القانونيون المؤهلون في صون العدالة وضمان تكافؤها في عالم تتزايد فيه مظاهر العولمة [4،8].

وقد قدم المرجع [4] تحليلا معمقا لتعقيدات الترجمة القانونية، مبينا أن هذا الحقل يتجاوز بكثير حدود النقل اللغوي البسيط، فالمترجم القانوني يؤدي وظيفة الوسيط المحايد داخل أنظمة قانونية متعددة اللغات، والهدف من ذلك ضمان عدالة التواصل والمساواة القانونية في ظل الفوارق الثقافية والاختصاصات القضائية. كما يرى الباحثون أن الترجمة القانونية متجذرة في جدلية معقدة تجمع بين اللغة والقانون والثقافة، فلكل نظام قانوني مصطلحاته وأعرافه الإجرائية وتقاليده التأويلية المميزة، ومن ثم، لا يطلب من المترجم القانوني امتلاك كفاءة لغوية رفيعة فحسب، بل يتعين عليه الإحاطة بأسس القانون المقارن ومنطق الاستدلال القانوني ودقائق السياق الثقافي. ويتمثل أحد أبرز التحديات التي حُددت في النقل الدقيق للمصطلحات القانونية، فغالبا ما تحمل دلالات خاصة بنظام قانوني معين قد تتبدل جذريا عند إساءة ترجمتها، وقد يؤدي الإخفاق في نقل هذه المصطلحات نقلا صحيحا إلى المساس بالقوة الإلزامية القانونية للوثائق المترجمة، لا سيما في السياقات الحساسة كالعقود أو النازعات القضائية. وعلاوة على ذلك، يواجه المترجمون صعوبة في التعامل مع اللغة القانونية الزاخرة بالمصطلحات وصيغها التقليدية، مع الحفاظ في الوقت ذاته على الطابع الرسمي ومقاصد النص المنقول منه (Source Text - ST). وتزيد هذه التعقيدات من صعوبة تقييم جودة الترجمة، إذ يتعين في عملية تقييم

جودة الترجمة مراعاة الدقة القانونية وسلاسة التعبير اللغوي والملاءمة الثقافية في آن واحد.

وإضافة إلى هذا المسار البحثي، ركز المرجع [8] على ترجمة المصطلحات القانونية ذات الخصوصية الثقافية كونها إحدى أعقد الإشكالات الملازمة لتقييم جودة الترجمة القانونية. ومن خلال تحليل ترجمات إنجليزية واردة في مؤلف "المترجم القانوني في الميدان" (The Legal Translator in the Field) لحاتم وآخرين (Hatem et al.)، كشفت الدراسة عن الصعوبات الجسيمة التي تعترض نقل المفاهيم القانونية المتجذرة ثقافيا والتي تفتقر في كثير من الأحيان إلى مكافئات مباشرة في لغة النص المنقول إليه (Target Text - TT). وقد أفضى التحليل إلى أن المترجم القانوني معني بإتقان ثلاث كفاءات جوهرية: المعرفة القانونية المقارنة والدقة في توظيف المصطلح القانوني وإتقان أنماط الكتابة القانونية. ويفضي غياب التكافؤ المباشر بين الأنظمة القانونية إلى إدخال المترجم في مسار معقد من اتخاذ القرار، فيوازن فيه بين الوفاء لمقاصد النص المنقول منه من جهة، وضمان الوضوح والوجاهة القانونية في لغة النص المنقول إليه من جهة أخرى. ويبرز هذا الوضع إشكالا مركزيا في تقييم الترجمة القانونية، إذ لا يقتصر دور المقيم على التحقق من السلامة اللغوية فحسب، بل يتعداه إلى فحص الوظيفة القانونية والثقافية للنص المترجم داخل النظام القانوني المستقبل.

وتظهر هذه الدراسات مجتمعة أن تقييم جودة الترجمة القانونية عملية متعددة الأبعاد، فتقتضي من المقيم مراعاة طيف واسع من المعايير، يشمل الدقة المصطلحية والوضوح التركيبي وصولا إلى التماسك القانوني والانسجام الثقافي. ومع تصاعد وتيرة التفاعلات القانونية العابرة للحدود في ظل العولمة، تتنامى الحاجة إلى آليات تقييم صارمة مراعية للسياق وقادرة على استيعاب تعقيدات هذا الحقل. ومن ثم، فإنه لا غنى لتقييم الترجمة القانونية عن خبرة بينية تتكامل فيها معارف دراسات الترجمة والقانون واللسانيات والتحليل الثقافي لصون العدالة وتحقيق الإنصاف في البيئات القانونية متعددة اللغات.

3.2. الذكاء الاصطناعي أداة لتقييم جودة الترجمة

نظرا لحداثة توظيف الذكاء الاصطناعي في مجال الترجمة فإن البحوث التي تتناول تطبيق هذه الأنظمة في تقييم الترجمة مازالت محدودة، وفي هذا الإطار، اقترح المرجع [14] تقنية وضع الأوامر (التوجيهات) (prompting) تعرف باسم "أوتو أم كيو أم" AUTOMQM، وهو اختصار لـ Automated Multi-dimensional Quality Metrics، أي المقاييس المؤتمتة متعددة الأبعاد لتقييم جودة الترجمة، إذ تقوم على تسخير قدرات النماذج اللغوية الكبيرة في الاستدلال والتعلم السياقي، وتكليفها برصد أخطاء الترجمة وتصنيفها. واعتمدت الدراسة المذكورة على اختبار نموذجين لغويين كبيرين، هما "بالم" (PaLM)و "بالم 2" (PaLM-2)، إذْ يشير PaLM إلى "نموذج لغوي معين" (Particular Language Model - PaLM). وقد استمدت الأمثلة محل التحليل من نسختين من ورشة الترجمة الآلية (Workshop on Machine Translation – WMT)، هما WMT'22 وWMT'19. وقد أظهرت النتائج وجود درجة عالية من الارتباط بين تقييمات نماذج PaLM-2 وأحكام المقيمين البشر، ولا سيما نموذج "الخط الأساس المكرر والمخزن على العقد PaLM-2 BISON (Baseline Iterated and Stored on Node) الذي أثبت قدرة تنافسية ملحوظة مقارنة بالنماذج المرجعية المتعلمة، خاصة فيما يتصل بتقييم الترجمات البديلة للجملة الواحدة. وخلصت الدراسة إلى أن مقياس "التقييم قائم على مقياس التقدير المعتمد على المحول التوليدي المسبق التدريب" GEMBA (GPT Estimation Metric Based Assessment) القائم على نماذج GPT، كما ورد في المرجع [15]، يتميز بدقة أعلى على مستوى المقاطع النصية، في حين تظهر نماذج PaLM-2 أداء أقوى على مستوى النظام ككل.

تناولت دراسة المرجع [16] الطريقة التي تعيد بها نماذج الذكاء الاصطناعي تشكيل عملية تقييم جودة الترجمة اعتمادا على آليات الارتجاع الآني، وقد بحثت الدراسة في أوجه إدماج هذه النماذج تقنيا وأثرها في رفع الكفاءة والحد من الأخطاء، إضافة إلى دورها في إحداث تحولات ملموسة في صنعة الترجمة من نواحي جودة الخدمات وكلفتها وقدرتها التنافسية. كما وسعت الدراسة نطاق التحليل ليشمل تجربة المستخدم والإشكالات الأخلاقية ومعوقات التطبيق بغية تقديم صورة شاملة

عن تأثير الذكاء الاصطناعي. كما أبرزت نتائج الدراسة جملة من المحاسن الرئيسة، وفي مقدمتها زيادة الإنتاجية والحد من الأخطاء البشرية بفضل الكشف الآني عنها، وتسجيل تحسن مستمر ناتج ارتجاعات سريعة ومتواصلة، وقد أسهمت هذه العوامل مجتمعة في زيادة سرعة عملية الترجمة ودقتها وجودتها العامة. وأسهم الذكاء الاصطناعي في توليد ترجمات أكثر سلامة وموثوقية، وهذا كان له محاسنه على رضا العملاء، في حين أنه ساعد على خفض تكاليف العمل وزاد من جدوى خدمات الترجمة من حيث التكلفة وأضفى عليها تنافسية أكثر. كما أظهرت آراء المستخدمين أن عددا كبيرا من المترجمين يقدرون ما تتيحه هذه التكنولوجيات من زيادة في الكفاءة والجودة، وإن كان بعضهم قد أظهر تحفظات تتصل بحدود الأنظمة ومدى وضوح الارتجاع المقدم، ومن ثم، فإن الدراسة قد شددت على ضرورة تعزيز سهولة الاستخدام وزيادة الدقة لتوسيع دائرة القبول. ومع ذلك، هنالك جملة من التحديات القائمة فيما يتعلق بمعالجة اللغات المعقدة أو محدودة الموارد والتعامل مع القضايا الأخلاقية مثل خصوصية البيانات والتحيز وضمان الشفافية، وبذلك اقترحت الدراسة مجموعة من الحلول، ومن بينها تحسين برامج التدريب وتطوير عملية تصميم النماذج والتواصل المفتوح. وفي استشرافها للمستقبل، توقعت الدراسة زيادة الاعتماد على الذكاء الاصطناعي في تقييم الترجمة وتوظيف تكنولوجيات مثل التحليلات التنبؤية والأدوات المعززة بالذكاء الاصطناعي للارتقاء بعملية التقييم، ولكن وتيرة هذا الاعتماد تظل مرهونة بتجاوز عوائق قائمة من قبيل غياب المعايير الموحدة ومقاومة التغيير وحماية البيانات.

تناولت دراسة المرجع [17] أثر الذكاء الاصطناعي في تقييم جودة الترجمة بجمع معطيات من عينة كبيرة ضمت 450 مشاركا، وشملت مترجمين مهنيين وطلبة دراسات عليا ومتعلمي لغات أجنبية يستخدمون أدوات الذكاء الاصطناعي بانتظام، وقد واعتمدت الدراسة استبيانا رقميا، وأساليب تحليل إحصائي، كالنسب المئوية والمتوسطات الحسابية واختبارات القيمة الإحصائية t (t-test)، لاستقصاء الطريقة التي يؤثر بها استثمار الذكاء الاصطناعي وجودة النص المنقول منه ونوع نموذج الذكاء الاصطناعي المستخدم في دقة التقييم. وهدفت الدراسة إلى تقييم مدى فاعلية الذكاء

الاصطناعي في رصد أخطاء الترجمة وتحسين آليات التقييم، مع مراعاة أثر جودة المدخلات وخبرة المقيم، وقد اختبرت ثلاث فرضيات رئيسة، مفادها: أن الذكاء الاصطناعي يسهم في تحسين دقة التقييم؛ وأن جودة النص المنقول منه تؤثر في مدى إحكام تقييم الذكاء الاصطناعي؛ وأن نوع نموذج الذكاء الاصطناعي المستخدم ينعكس انعكاسا جوهريا على مخرجات التقييم. وقد أيدت النتائج الفرضيات الثلاث جميعها، وأفادت المعطيات بأن أكثر من 78% من المشاركين عدوا الذكاء الاصطناعي أداة أساسية في تقييم جودة الترجمة، فيما أقر ما يزيد على 79% بدوره في تعزيز الدقة، ورأى 75% أنه يسهم في الحد من التحيز. كما أشار 61% من المشاركين إلى أن جودة النص المنقول منه تؤثر في أداء الذكاء الاصطناعي، في حين اعتبر 63% أن أدوات الذكاء الاصطناعي تكون أكثر موثوقية عندما تكون النصوص المنقول منه مصاغة بإحكام. وفيما يتصل بنوع النموذج، أفاد 65% بأنه يؤثر في دقة التقييم، بينما شدد 73% على أهمية خبرة المقيم، وأبرز 59% دور الكفاءة اللغوية في هذا السياق. وخلصت الدراسة إلى تأكيد الأثر البالغ للذكاء الاصطناعي في تعزيز عملية تقييم جودة الترجمة، غير أن فاعلية هذه الأدوات تظل مرهونة بعوامل حاسمة، وفي مقدمتها جودة النص المنقول منه ونوع نموذج الذكاء الاصطناعي المعتمد ومستوى خبرة المقيم. ومن ثم، يتعين التعامل مع هذه العوامل بعناية منهجية إذا أريد تعظيم الفوائد الممكنة لتوظيف الذكاء الاصطناعي في تقييم الترجمة.

بينت الدراسات المستعرضة أن الذكاء الاصطناعي أداة فعالة تساعد على تعزيز عملية تقييم جودة الترجمة، وإن ظل توظيفه في هذا المجال في مراحله الأولى ومقترنا بجملة من التحديات المعقدة. فقد أوضح المرجع [14] كيف يمكن للنماذج اللغوية الكبيرة أن تحاكي أحكام المقيمين البشريين في تصنيف أخطاء الترجمة بدقة، في حين شدد المرجع [16] على الأثر الأشمل للذكاء الاصطناعي في ميدان الترجمة بالاستفادة من الارتجاع الآلي ودوره في رفع الكفاءة ودعم موجه للأساتذة. كما قدم المرجع [17] دلائل تجريبية تؤكد قدرة الذكاء الاصطناعي على تحسين دقة التقييم، مع إبراز أثر عوامل حاسمة كجودة النص المنقول منه ونوع نموذج الذكاء الاصطناعي المستخدم وخبرة المقيم. وتؤكد هذه النتائج مجتمعة تنامي موثوقية الذكاء الاصطناعي

وجدواه في تقييم جودة الترجمة، شريطة معالجة حدوده من خلال تحسين تدريب النماذج وتعزيز الضوابط الأخلاقية وترسيخ قدر أكبر من الشفافية. ومع تطور هذه التكنولوجيات، يظل اعتماد مقاربة متوازنة تجمع بين سرعة الذكاء الاصطناعي واتساقه من جهة، والبصيرة النقدية الإنسانية من جهة أخرى، عاملا أساسا لضمان أطر تقييم صارمة وقابلة للتكيف.

4.2. تشات جي بي تي أداة لتقييم جودة الترجمة

بدأت البحوث في مجال تقييم جودة الترجمة تستكشف الطريقة التي يقارن بها تشات جي بي تي بنماذج ذكاء اصطناعي أخرى في تقديم تقييمات دقيقة، وفي هذا السياق، هدفت الدراسة المشار إليها في المرجع [5] إلى تقصي مدى قدرة النماذج اللغوية الكبيرة على الاضطلاع بدور مقيمين في مجال تصحيح الأخطاء النحوية في اللغة الإنجليزية (English Grammatical Error Correction – GEC). وقد تناولت الدراسة ثلاثة نماذج لغوية كبيرة، هي: جي بي تي 3.5، وجي بي تي 4، ولاما 2 [5]. واعتمد الباحثون على نصوص مستمدة من مجموعة بيانات أرشيف اللهجات الإنجليزية الجنوب شرقية (Southeastern English Dialects Archive – SEEDA). وجرت عملية تقييم إمكانات هذه النماذج على أنها أدوات تقييم بمقارنة أدائها بالمقاييس التقليدية المعتمدة، وقد وأظهرت النتائج أن نموذج جي بي تي 4 حقق، على مستوى النظام، درجات ارتباط مرتفعة مقارنة بالمقاييس القائمة، وهو ما يبرز فائدته في مهام تقييم تصحيح الأخطاء النحوية، في حين حافظ جي بي تي 4 كذلك على مستوى الجمل على درجة ارتباط مرتفعة نسبيا عند مقارنته بالمقاييس التقليدية، وهذا يشير إلى توافق كبير مع أحكام المقيمين البشر. وعلى النقيض من ذلك، حقق نموذج جي بي تي 3.5 درجات ارتباط متوسطة فقط، وكان أداؤه أدنى بصورة ملحوظة من جي بي تي4، في حين أظهر نموذج LLaMA-2 أضعف أداء، إذ لم يسجل سوى ارتباط محدود بأحكام البشر. وتشير هذه النتائج إلى أن نموذج جي بي تي 4 يتفوق حاليا على النماذج اللغوية الكبيرة الأخرى محل الدراسة، غير أنها تؤكد، في الوقت نفسه، الحاجة إلى مزيد من الجهود البحثية لتحسين موثوقية النماذج البديلة في مهام تصحيح الأخطاء النحوية.

هدفت ورقة بحثية أخرى، أشار إليها المرجع [9]، إلى تقييم قدرة تشات جي بي تي على تصحيح الأخطاء النحوية في النصوص المكتوبة، وذلك بمقارنته بمنتجات تجارية مخصصة لهذا الغرض، مثل غرامالي Grammarly، وبنماذج متقدمة حديثة، من بينها GECToR، وهو اختصار لتصحيح الأخطاء النحوية بارتجاع آني « Grammatical Error Correction with Real-time feedback »، وقد استخرجت الأمثلة الخاضعة للتحليل من مجموعة البيانات المعيارية لمؤتمر تعلم اللغة الطبيعية لسنة 2014 (– Conference on Natural Language Learning 2014 CoNLL2014)، وجرى تقييم الأداء بالاعتماد على ثلاثة مقاييس إحصائية، هي: الدقة (Precision)، والاسترجاع (Recall)، ودرجة F0.5. ووفقا للنتائج المتحصل عليها، سجل تشات جي بي تي أعلى قيمة في الاسترجاع، في حين حقق نموذج GECToR أعلى قيمة في الدقة، بينما أظهر Grammarly توازنا بين المقياسين، محققا أعلى درجة في مقياس F0.5. وتشير هذه النتائج إلى أن تشات جي بي تي يميل إلى تصحيح أكبر عدد ممكن من الأخطاء، وهو ما قد يفضي أحيانا إلى الإفراط في التصحيح. وعلى النقيض من ذلك، يقتصر نموذج GECToR على تصحيح الأخطاء التي يثق بصحتها، الأمر الذي يترك عددا من الأخطاء دون معالجة. أما Grammarly، فقد بدا أداؤه أكثر استقرارا، إذ يجمع بين مزايا كل من تشات جي بي تي وGECToR.

تناولت دراسة المرجع [15] مسألة مدى أهلية النماذج اللغوية الكبيرة للاضطلاع بدور فعال في تقييم جودة الترجمة، وقد جعل الباحثون من نموذج جي بي تي 4 محورا رئيسا للتحليل، مع توسيع نطاق المقارنة ليشمل عددا من النماذج الأخرى، من بينها GPT-2، وAda GPT-3، وBabbage GPT-3، وCurie GPT-3، وDavinci-002 GPT-3.5، وChatGPT، وDavinci-003 GPT-3.5.1، وGPT-3.5-turbo. واعتمدت الدراسة في تصميمها المنهجي على مقاييس ورشة الترجمة الآلية لسنة 2022، إذ استمدت بيانات الاختبار من مجموعة مقاييس الجودة متعددة الأبعاد لسنة 2022. وشملت هذه البيانات ترجمات من الإنجليزية إلى الألمانية، ومن الإنجليزية إلى الروسية، ومن الصينية إلى الإنجليزية، وضمت مخرجات ترجمية صادرة عن 54 نظاما مختلفا، سواء أكانت ترجمات بشرية أم آلية. وفي هذا السياق، اقترح الباحثون مقياسا

جديدا أطلقوا عليه اسم GEMBA، وهو يقوم على تقييم مقاطع الترجمة كلا على حدة، ثم استخلاص التقييم العام للنظام من خلال احتساب المتوسط الحسابي لدرجات المقاطع. وقد بينت النتائج أن نموذج جي بي تي 4 أظهر أعلى درجات الاتساق مع أحكام المقيمين البشر عبر أزواج اللغات الثلاثة جميعها، تلاه مباشرة نموذجا Davinci-003 وGPT-3.5-turbo. وفي المقابل، سجلت النماذج الأقدم، مثل GPT-2 وCurie GPT-3، أداء ضعيفا بوضوح، إذ جاءت نتائجها في كثير من الحالات أقرب إلى التخمين. وتكشف هذه النتائج عن الإمكانات الكبيرة التي تنطوي عليها النماذج اللغوية الكبيرة المتقدمة في مجال تقييم جودة الترجمة، في الوقت الذي تبرز فيه محدودية النماذج الأقدم.

وكونه نموذجا متقدما من نماذج الذكاء الاصطناعي، حظي تشات جي بي تي [18] باهتمام متزايد في الأدبيات الحديثة نظرا لدوره المتوقع في عملية تقييم جودة الترجمة، وقد قدم المرجع [19] دراسة قارنت بين تقييم بشري وتقييم أجراه تشات جي بي تي 3.5 لترجمة مقطع علمي من الإنجليزية إلى البرتغالية، وخلال ذلك اختار الباحثون مقتطفا من مقال علمي، وقيموا أربع ترجمات مختلفة إلى البرتغالية، وأنجزت اثنتان من هذه الترجمات بجوجل للترجمة Google Translate وديبل DeepL، وثالثة ولدها تشات جي بي تي، وأما الترجمة الرابعة فكانت الترجمة البرتغالية الرسمية. واعتمدت عملية التقييم على ثلاثة معايير محددة، هي: الفصاحة والملاءمة والدقة، مع استخدام مقياس ليكرت يمتد من 1 إلى 5 لكل معيار، وطلب من المقيمين الالتزام بالمعايير نفسها وتقدير الترجمات وفق السلم ذاته. وأظهرت النتائج توصل المقيمين وتشات جي بي تي إلى الأحكام نفسها في تحديد الترجمتين التي استحقتا أدنى الدرجات وأعلاها، إذ اتفق الطرفان على تصنيف الترجمات ذاتها بوصفها الأفضل والأسوأ من بين الترجمات محل التقييم. كما كشفت النتائج أن تقييمات تشات جي بي تي اتسمت بدرجة أعلى من الاتساق مقارنة بتقييمات البشر.

تناولت دراسة أخرى أشار إليها المرجع [20] مسألة إمكانية توظيف تشات جي بي تي مقياسا لتقييم جودة الترجمة، وسعت في الوقت نفسه إلى استكشاف الكيفية التي يمكن من خلالها صياغة أوامر تمكن المستخدم من الحصول على تقييمات

موثوقة، إذا ما ثبتت صلاحية تشات جي بي تي للقيام بهذا الدور. وانطلق الباحثون من مقارنة تقييمات تشات جي بي تي القائمة على توجيه تحليل الأخطاء بجملة من المقاييس المعتمدة، من بينها مقياس بلو BLEU (Bilingual Evaluation Understudy – مقياس التقييم ثنائي اللغة المرجعي)، وبيرت سكور BERTscore (وهو مقياس يستند إلى تمثيلات BERT، أي Bidirectional Encoder Representations from Transformers – تمثيلات المرمزات ثنائية الاتجاه القائمة على المحولات)، وBLEURT (Bilingual Evaluation Understudy with Representations from Transformers) - مقياس التقييم ثنائي اللغة القائم على تمثيلات المحولات)، وكوميت COMET (Crosslingual Optimized Metric for Evaluation of Translation) - مقياس محسن عابر للغات لتقييم الترجمة). واعتمدت الدراسة على مجموعة اختبار مأخوذة من مهمة المقاييس المشتركة لورشة الترجمة الآلية لسنة 2020 (WMT20 Metric shared task)، وشملت زوجين لغويين هما: الصينية إلى الإنجليزية، والإنجليزية إلى الألمانية، واستندت استراتيجية وضع الأوامر داخل السياق (in-context prompting) إلى منهج سلسلة التفكير (chain-of-thought) وتحليل الأخطاء، إذ أظهرت النتائج أن هذا النمط من الأوامر يسهم في تحسين أداء تشات جي بي تي في مهام التقييم. وفي المقابل، كشفت الدراسة عن جملة من الإشكالات المحتملة التي ينبغي على الباحثين التنبه إليها عند استخدام تشات جي بي تي أداة للتقييم.

كما تناول المرجع [6] أثر توظيف تشات جي بي تي 4o في تقييم جودة الترجمة البشرية بالاستناد إلى مقاييس الجودة متعددة الأبعاد (Multidimensional Quality MQM-Metrics). وقد فحصت الدراسة ترجمات أداها طلبة ماستر الترجمة والترجمة الشفوية (Master of Translation and Interpreting - MTI) لنص أدبي وآخر غير أدبي، مع إجراء مقارنة بين تقييمات تشات جي بي تي 4o وتقييمات المقيمين البشر. وانصب التحليل على ثلاثة محاور رئيسة هي: الدقة والأمانة والفصاحة، مع وضع عدد الأخطاء والدرجات الإجمالية للجودة في الحسبان. وأظهرت نتائج الدراسة تقاربا واضحا بين تقييمات تشات جي بي تي 4o وتقييمات البشر، تمثل في اتساق الدرجات المقدرة وتشابه الارتجاعات النوعية.

الفجوة البحثية

أظهرت مراجعة الأدبيات السابقة، الملخصة في الجدول (1) أدناه، وجود ثلاث فجوات بحثية رئيسة، وأولها، تمحور اهتمام معظم الدراسات حول أزواج لغوية غير زوج الإنجليزية والعربية، وهو ما أدى إلى إغفال الخصوصيات اللغوية والقانونية التي تميز هذا الزوج اللغوي وما ينطوي عليه من تعقيدات خاصة (مثل [5،6]). ثانيا، غلب الطابع الكمي على البحوث المنجزة، إذ ركزت في الأغلب على مؤشرات الأداء الخاصة بالذكاء الاصطناعي والنماذج اللغوية الكبيرة، دون الالتفات الكافي إلى الأبعاد البيداغوجية أو إلى التطبيقات المرتبطة مباشرة بدور الأستاذ في سياقات التكوين [14،17]. ثالثا، لوحِظ أن هنالك تركيزا مفرطا على الأخطاء النحوية، في مقابل اهتمام محدود بالقضايا الأعمق، مثل الاتساق القانوني والسياق الثقافي، وهما عنصران حاسمان في تقييم الترجمة القانونية [8،9]. وعلى خلاف الدراسات السابقة التي انصب جل اهتمامها على تحسين أداء أنظمة الذكاء الاصطناعي ذاتها، تسعى هذه الدراسة إلى تحويل بؤرة الاهتمام نحو دعم المتكونين في مجال الترجمة والأساتذة المشرفين عليهم، فمن خلال استقصاء التحديات المرتبطة بتوظيف الذكاء الاصطناعي في تقييم الترجمة القانونية، تهدف الدراسة إلى الإسهام في بلورة مقاربات أكثر فاعلية وأشد التزاما بالاعتبارات الأخلاقية لإدماج التكنولوجيا في تكوين المترجمين.

جدول 1. عرض موجز للأدبيات المستعرضة.

المراجع	محور الدراسة	المنهجية / البيانات	أبرز النتائج
[12]	دور الذكاء الاصطناعي في دراسات الترجمة	استكشاف نظري لتوظيف الذكاء الاصطناعي في بحوث الترجمة	يوسع الذكاء الاصطناعي الأفق المنهجي ويحسن الكفاءة، ومن تحدياته: عدم الاتساق والإشكالات الأخلاقية وغياب المعايير الموحدة
[3]	الذكاء الاصطناعي في الممارسة الترجمية	تحليل تصوري للذكاء الاصطناعي والترجمة الآلية العصبية أثناء التطبيق	فوائد تشمل السرعة وجدوى التكلفة وسهولة الوصول؛ وحدود تتصل بالتعابير

			الاصطلاحية والحساسية الثقافية والأخلاقيات
[13]	تصورات الأساتذة حول أدوات الذكاء الاصطناعي في الجامعات السعودية	استبيان شمل 50 أستاذا (14 بندا)	مواقف محبذة عموما تجاه الذكاء الاصطناعي؛ محدودية الاستخدام الفعلي في الصف بسبب نقص التدريب أو الدعم
[4]	تحديات الترجمة القانونية	تحليل نظري وقانوني	الترجمة القانونية معقدة ومرتبطة بالنظام القانوني؛ وتتطلب خبرة قانونية وثقافية متخصصة
[8]	المصطلحات القانونية ذات الخصوصية الثقافية	دراسة حالة لترجمات قانونية إنجليزية	صعوبة ترجمة المفاهيم القانونية/الثقافية المضمنة؛ ضرورة تقييم الوظيفة القانونية
[14]	النماذج اللغوية الكبيرة في تقييم جودة الترجمة الآلي (PaLM، (PaLM-2	مجموعتا بيانات WMT'19وWMT'22مع توجيهAUTOMQM	نماذج PaLM-2 تظهر ارتباطا عاليا بأحكام البشر؛ ومقياس GEMBA أدق على مستوى المقطع
[16]	إعادة تشكيل تقييم جودة الترجمة في الصناعة بواسطة الذكاء الاصطناعي	تحليل مختلط: الكفاءة، تقليص الأخطاء، تجربة المستخدم	تحسين السرعة والدقة وخفض التكلفة؛ مع مخاوف تتعلق بالتحيز، وقابلية الاستخدام، والأخلاقيات
[17]	أثر الذكاء الاصطناعي في تقييم جودة الترجمة من منظور المستخدمين	استبيان شمل 450 مترجما ومتعلما	يرى معظم المشاركين الذكاء الاصطناعي أساسيا؛ وتتأثر الفاعلية بجودة النص المنقول منه ونوع النموذج وخبرة المقيم
[5]	النماذج اللغوية الكبيرةGPT-3.5) ، GPT-4، LLaMA-2)	مجموعة بيانات SEEDA مقارنة بالمقاييس	GPT-4 يظهر ارتباطا قويا بأحكام البشر

	في تقييم عملية تصحيح الأخطاء النحوية		
[9]	مقارنة ChatGPT و Grammarly و GECToR في تصحيح الأخطاء النحوية	مجموعة بيانات CoNLL2014؛ مقاييس الدقة/الاسترجاعF0.5/	ChatGPTمرتفع الاسترجاع (يميل للإفراط في التصحيح)؛ Grammarly متوازن؛ GECToR محافظ
[15]	مقياس GEMBA لتقييم جودة الترجمة باستخدام نماذج جي بي تي	مجموعة اختبار MQM لسنة EN-DE2022) ، EN-RU، ZH-EN)	GEMBA المعتمد على GPT-4 يتفوق على المقاييس الآلية الأخرى
[19]	مقارنة تشات جي بي تي والبشر في تقييم ترجمة علمية من الإنجليزية إلى البرتغالية	مقارنة أربع ترجمات (جوجل، ديبل، تشات جي بي تي، الرسمية)	توافق تشات جي بي تي مع البشر في تحديد الأفضل والأسوأ؛ واتساق أعلى من التقييم البشري
[20]	استراتيجيات وضع الأوامر لتشات جي بي تي في تقييم جودة الترجمة	مجموعة بيانات WMT20؛ مقارنة مع مقاييس BLEU و BERTscore و BLEURT وCOMET	أوامر سلسلة التفكير وتحليل الأخطاء تحسن أداء تشات جي بي تي في التقييم
[6]	تشات جي بي تي 4o في تقييم الترجمة الأدبية وغير الأدبية	ترجمات طلبة ماجستير الترجمة والترجمة الشفوية واعتماد مقاييس الجودة متعددة الأبعاد	توافق وثيق بين تشات جي بي تي 4o والمقيمين البشر؛ اتساق الدرجات والارتجاع.

وتسعى هذه الدراسة إلى سد الفجوات المشار إليها بتقصي إمكانات تشات جي بي تي 4o كأداة لتقييم جودة الترجمة في سياق تكوين المترجمين في مجال الترجمة

القانونية من الإنجليزية إلى العربية. وتهدف، في هذا الإطار، إلى دعم المتكونين والأساتذة على حد سواء، بإدماج تقييم قائم على الذكاء الاصطناعي إدماجا فاعلا وأخلاقيا، يوازن بين ما تتيحه التطورات التقنية الحديثة، وبين المتطلبات المهنية والبيداغوجية الخاصة بتكوين المترجم القانوني.

ويعرض في القسم الموالي الإطار المنهجي المعتمد في هذه الدراسة، يتبعه تقديم إجراءات التحليل والنتائج المتحصل عليها.

3. المنهجية

يتمحور هدف هذه الدراسة حول تقييم مدى اتساق تقييمات الذكاء الاصطناعي مع تقييمات الأساتذة لترجمات الطلبة في مقرر للترجمة القانونية، واستكشاف ما ينطوي عليه توظيف الذكاء الاصطناعي أداة لتقييم جودة الترجمة من مكاسب وتحديات. ولتحقيق هذه الأهداف اعتمدت دراسة حالة وصفية قائمة على المنهج المختلط، واستنادا إلى ما أورده المرجع [21]، أتاح اعتماد دراسة الحالة الوصفية إجراء تحليل معمق لعمليات التقييم ضمن سياق مقرر للترجمة القانونية، من غير الركون إلى تعميمات نظرية مجردة. كما أسهم الجمع بين المقاربتين النوعية والكمية في تقديم صورة أكثر شمولا، إذ مكنت المعطيات النوعية من رصد تصورات الأساتذة وخبراتهم التقييمية، في حين أتاحت المعطيات الكمية مقارنة الدرجات المسندة لترجمات الطلبة من الذكاء الاصطناعي والمقيمين، وقد عزز هذا التكامل المنهجي رصانة النتائج وشموليتها [21].

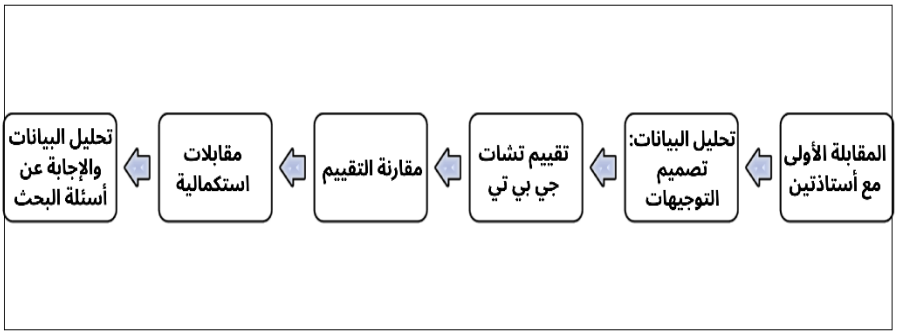
1.3. المشاركون والسياق

اختيرت أستاذتين حاملتين لشهادة الدكتوراه في الترجمة، وتمتلك الأستاذة (أ) خبرة تدريسية تمتد عشر سنوات، تركزت أساسا في الترجمة الشفوية والترجمة العسكرية، وقد انتقلت حديثا إلى تدريس الترجمة القانونية بعد أكثر من سنة من الممارسة المهنية في هذا المجال. أما الأستاذة (ب)، فلها خبرة تدريسية تقارب ثلاثين سنة، من بينها ثلاث سنوات في تدريس الترجمة القانونية. وقد عملت الأستاذتان في

إحدى الجامعات السعودية، وتولتا تدريس عدد من مقررات الترجمة، من ضمنها مقرر ترجمة النصوص القانونية. ويعد هذا المقرر جزءا من برنامج ليسانس الترجمة بكلية علوم اللغات، جامعة الملك سعود (King Saoud University - KSU) [22]. وهو مقرر إجباري يدرس في الفصل الأخير من البرنامج، ويهدف إلى تنمية قدرة الطلبة على ترجمة النصوص القانونية المعقدة والمنتقاة بعناية لتمثل مجالات قانونية متنوعة. ويدرب في إطار هذا المقرر خمسة وعشرون طالبا على التعرف إلى أنواع الوثائق القانونية المختلفة وخصائصها المميزة، وترجمة النصوص القانونية بين الإنجليزية والعربية بدقة، وفهم المضامين والمصطلحات القانونية، واستيعاب المقابلات القانونية المناسبة في اللغة المنقول إليها، وتوظيف الموارد الترجمية ذات الصلة. وجاء اختيار الترجمة القانونية محورا لهذه الدراسة الحالة بالنظر إلى ما تطرحه من صعوبات خاصة أمام المتكونين في مجال الترجمة.

2.3. الإجراءات

بعد الحصول على الموافقة الأخلاقية اللازمة لإجراء هذه الدراسة من اللجنة الدائمة لأخلاقيات البحث العلمي بجامعة الملك سعود (رقم: 810-24)، مرت الدراسة بعدة مراحل متتابعة، كما هو موضح في الشكل (1) أدناه.



شكل 1. مراحل البحث

روسلت الأستاذتان وطلبت منهما موافقة خطية للمشاركة في هذه الدراسة،

وقد أُعلمتا بأهداف الدراسة وإجراءاتها، مع التأكيد على أن المشاركة طوعية بالكامل، وأن المعطيات التي تقدم ستستخدم لأغراض البحث العلمي فحسب، مع ضمان الحفاظ على سريتها ومعالجتها على نحو يضمن عدم الكشف عن الهوية.

3.3. المقابلات الأولية شبه المنظمة

أجريت مقابلات أولية شبه منظمة مع كل أستاذة على حدة عبر منصة زوم (نحو ساعة لكل مقابلة)، وهدفت هذه المقابلات إلى الوقوف على مؤهلات الأستاذتين وخبرتهما في تدريس الترجمة القانونية، والغرض من ذلك تأطير آرائهما وفهمها في سياقها المهني والأكاديمي. وتناولت الأسئلة مقاربتهما العامة لتدريس الترجمة القانونية وتقييمها وأنواع النصوص المسندة إلى الطلبة والمعايير المعتمدة في تقييم جودة الترجمة والآليات المستخدمة في تحديد درجة خطورة الأخطاء. كما طرحت أسئلة تتعلق بأكثر الصعوبات التي يواجهها الطلبة والأخطاء المتكررة في أعمالهم، وسبل التعامل معها، وما إذا كانت بعض أنواع الوثائق القانونية أشد تعقيدا من غيرها من حيث التقييم، وتطرقت المقابلات كذلك إلى كيفية تعامل الأستاذتين مع اعتراضات الطلبة على الدرجات أو الارتجاع، ومدى فاعليته في تحسين أداء الطلبة، والتقنيات المفضلة لديهما لضمان فهم الطلبة للملاحظات المقدمة. وفي ختام المقابلات، استحصلت الموافقة على استخدام أعمال الطلبة بصورة مجهولة الهوية لأغراض البحث العلمي. وعلى هذا الأساس، جمعت من الأستاذتين عشر نسخ مجهولة الهوية من ترجمات امتحان منتصف الفصل من الإنجليزية إلى العربية، تعود إلى عشرة طلبة (خمسة عن كل أستاذة)، وتمثل مستويات أداء مرتفعة ومتوسطة وضعيفة، وقد تولت الأستاذتان مهمة تصحيح هذه الترجمات، وهما اللتان درستا الطلبة في مقرر الترجمة القانونية، ونظرا إلى أن الترجمات كانت مكتوبة بخط اليد، فقد حولت إلى صيغة رقمية باستخدام خاصية التعرف البصري على الحروف (Optical Charcater Recognition - OCR)[ix] في أداة جوجل للترجمة، وخضعت النصوص بعدها لعملية تدقيق يدوي قام بها باحثان اثنان، إذ قارن كل منهما النسخ الرقمية بالنصوص الأصلية المكتوبة بخط اليد على نحو مستقل. وعولجت أي فروق أو أخطاء محتملة

بالتوافق بين الباحثين قبل الشروع في التحليل. ورغم أن معدلات الخطأ لم تسجل تسجيلا كميا، فإن هذه الإجراءات كفلت أن تكون النصوص النهائية المعتمدة في التقييم خالية من أخطاء خاصية التعرف البصري على الحروف الظاهرة.

4.3. تصميم الأوامر (Prompt Design)

اختير تشات جي بي تي 4o ليكون أداة الذكاء الاصطناعي المعتمدة في هذه الدراسة، ويعد هذا النموذج أول أداة ذكاء اصطناعي أتاحتها شركة أوبن أي آي (OpenAI) للعامة في ماي 2024، إذ يوفر قدرات متقدمة في الاستدلال متعدد الوسائط آنيا والسرعة وسعة الانتشار [23]. ويعكس اعتماده الواسع الذي تجاوز عدد مستخدميه 400 مليون مستخدم أسبوعيا بحلول فبراير 2025 [24] مدى جدواه في مجالات معرفية متعددة، فعند توظيف أدوات الذكاء الاصطناعي بوجه عام، وتشات جي بي تي على وجه الخصوص، يعد تصميم الأوامر عنصرا حاسما، إذ تبين دراسات سابقة (مثل [25،26]) أن الأوامر التي تحدد معايير التقييم بوضوح، كالدقة والفصاحة والملاءمة الثقافية، تفضي إلى تقييمات أكثر موثوقية واتساقا مع الأحكام البشرية. وانطلاقا من ذلك، وجه تشات جي بي تي إلى إجراء تقييم منهجي لترجمات الطلبة القانونية بمقارنة الترجمة كاملة بالنص المنقول منه، مع التأكد من عدم وجود حذف أو إسقاط، وطلِب من النموذج الإشارة إلى الأخطاء داخل النص باستخدام أقواس مربعة (مثل: استخدم [المصطلح الخاطئ])، ثم إدراج قائمة بالأخطاء أسفل كل فقرة، مرفقة بتصنيفها وتقديم تعليقات تشرح طبيعة كل خطأ، وتقترح تصحيحات مناسبة، وتحدد مقدار الخصم المترتب عليه. وأما الأخطاء المتكررة فكان يشار إليها في جميع مواضعها، مع احتسابها مرة واحدة فقط. وفي حال وجود أجزاء مفقودة، كان يطلب تقديم الترجمة الصحيحة مع تطبيق الخصم المناسب، في حين تسجل الترجمات غير المكتملة بخصم قدره 0.5 نقطة عن كل جملة ناقصة. وفي الختام، قدم ملخص منظم يتضمن اقتطاع نقاط أخطاء المحتوى وغيرها، مع إظهار الدرجة النهائية المتبقية.

ولضمان مزيد من الوضوح، يقدم الجدول (2) أدناه خلاصة نظام التنقيط

المعتمد لكل أستاذة.

جدول 2. أنواع أخطاء التقييم المعتمدة في تقييم الذكاء الاصطناعي وقيم الخصم الخاصة بكل أستاذة.

نوع الخطأ	الخصم – الأستاذة (أ)	الخصم – الأستاذة (ب)
مصطلحات متخصصة غير صحيحة	0.5–عن كل مصطلح	0.25–عن كل مصطلح
كلمات غير مترجمة	0.25–عن كل كلمة	—
جمل غير مترجمة	0.5–عن كل جملة	0.5–عن كل جملة
عناوين / عناوين فرعية مترجمة ترجمة غير صحيحة	—	0.5–عن كل حالة
أخطاء أخرى (نحوية، بنيوية، إملائية، ترقيمية، أدوات الربط، ترتيب الكلمات، ترجمات غير دقيقة)	0.25–عن كل حالة	0.25–عن كل زوج من الأخطاء من النوع نفسه

وفيما يخص الأستاذة (ب)، فقد قسمت اقتطاعات النقاط بالتساوي بين أخطاء المحتوى (بحد أقصى 7.5 نقاط) والأخطاء الأخرى (بحد أقصى 7.5 نقاط)، وكفل هذا الأمر أن تأتي التقييمات متسقة ومفصلة ومنسجمة مع معايير التقييم التي يعتمدها كل من الأستاذتين في تدريس الترجمة القانونية.

5.3. تقييم تشات جي بي تي والمقارنة

استنادا إلى معايير التقييم التي استخلصت من المقابلات الأولية، خضعت صيغ الأوامر متعددة للاختبار على نحو تكراري إلى أن حصل أمر أمثل، وأدخل إلى تشات جي بي تي 4o مرفقا بالنصوص المنقول منها وترجمات الطلبة، وقد قيم تشات جي بي تي 4o كل ترجمة على حدة، ثم جمعت المخرجات بغرض التحليل. وأجريت بعد ذلك مقارنة يدوية دقيقة، مقطعا بمقطع، بين تقييمات تشات جي بي تي 4o وتقييمات المقيمين البشر، إذ صنفت الأخطاء حسب أنواعها، وقيس مدى الاتساق من خلال

مقارنة الدرجات إحصائيا للتحقق مما إذا كانت هنالك فروق كبيرة.

6.3. المقابلات الاستكمالية

أجريت جولة ثانية من المقابلات شبه المنظمة (نحو ساعة و44 دقيقة لكل مقابلة) عبر منصة جوجل ميت، وخصصت لعرض نتائج التقييم ومناقشتها واستجلاء تصورات الاتساق والمكاسب المحتملة لتوظيف الذكاء الاصطناعي وما يطرحه من تحديات. وسجلت الإجابات صوتيا، ثم أفرغت وخضعت لتحليل موضوعاتي ضمن أربع فئات رئيسة: الاتساق والفوائد والتحديات والملاحظات إضافية، مع استخراج اقتباسات مباشرة لتجسيد أبرز الآراء.

7.3. المصداقية والموثوقية

تعزيزا لمصداقية هذه الدراسة، اعتمدت الباحثتين على مصادر بيانات متعددة، مع تطبيق مبدأ التثليث المنهجي بتوظيف أكثر من مصدر للبيانات، شمل مقابلات مع الأستاذتين وتقييمات الأداء، بغية التحقق المتبادل من النتائج وضمان فهم شامل لسياق البحث [21]. وإضافة إلى ذلك، أنجز هذا البحث في إطار رسالة ماجستير للباحثة الأولى، تحت إشراف أستاذة بدرجة أستاذ كامل بكلية علوم اللغات، وأسهمت المشاورات المنتظمة وجلسات الارتجاع مع المشرفة في تعزيز الصدق التفسيري، عبر تنقيح التحليل والحد من التحيز البحثي.

4 .التحليل والنتائج

يعرض هذا القسم مرحلة التحليل والنتائج التي أسفرت عنها الدراسة، ويتناول تقييم مدى الاتساق بين تقييمات تشات جي بي تي 4o وتقييمات الأساتذة في الحكم على جودة الترجمة القانونية، وقد سعت الدراسة إلى تحديد درجة التقارب بين تقييمات تشات جي بي تي 4o وتقييمات أساتذة ذوي خبرة، من خلال فحص قدرته على رصد الأخطاء واتساق التنقيط ووضوح بنية الارتجاع ومقروئيته. وإلى جانب بحث الاتساق، يستكشف هذا القسم كذلك المكاسب والتحديات المحتملة لإدماج تشات

جي بي تي 4o في عملية التقييم، ولا سيما في سياق تقييم الترجمة القانونية.

ولتحقيق ذلك، زود تشات جي بي تي 4o بمعايير التقييم نفسها التي يعتمدها الأساتذة، مع مراعاة أن لكل أستاذة نظام تنقيط خاصا بها، وقد جمعت هذه المعايير خلال المقابلات الأولية شبه المهيكلة، في حين حُللت النتائج تحليلا نوعيا وكميا بهدف مقارنة تقييمات تشات جي بي تي 4o بتقييمات الأساتذة، مع التركيز على الفروق في الدرجات من جهة، وأنماط الحكم النوعي من جهة أخرى. كما أبرز التحليل مواضع الاتفاق والاختلاف على نحو تفصيلي، وهذا يمكن من إظهار مكامن القوة ومواطن القصور في توظيف الذكاء الاصطناعي في تقييم الترجمة.

وتناول هذا القسم كذلك مسألة اتساق تشات جي بي تي 4o ومقروئية مخرجاته عبر مختلف التقييمات، وذلك بتقييم ما إذا كان الارتجاع المنظم الذي يقدمه يسهم في تعزيز موثوقية التقييم وقيمته التعليمية. وتسهم النتائج المتحصل عليها في إثراء النقاش الدائر حول دور الذكاء الاصطناعي في مجال الترجمة، ولا سيما في الحقول المتخصصة، مثل الترجمة القانونية، إذ تحتل الدقة والفهم السياقي مكانة محورية.

1.4. تحليل المقابلات الأولية

انصب الهدف من الجولة الأولى من المقابلات على جمع معطيات تتعلق بالتحديات التي يواجهها الأساتذة عند تقييم الترجمة القانونية، فضلا عن استخلاص المعايير التي يعتمدونها في عملية التقييم. ولتحقيق ذلك، سجلت المقابلات ودونت ملاحظات في مجرياتها، ثم جمعت العناصر المتصلة بتحديات الترجمة القانونية ومعايير تقييمها في ملف واحد. وبعد ذلك، صنفت الإجابات من خلال فصل التحديات عن المعايير، واستثمرت هذه المعطيات في تصميم توجيه التقييم المعتمد في هذه الدراسة. كما نظمت البيانات ورتبت بما يبرز الأنماط المتكررة والموضوعات المشتركة.

وأظهرت المعطيات المستخلصة من المقابلات أن من أبرز التحديات التي يواجهها الأساتذة عند تقييم الترجمات القانونية غموض الصياغة، وسوء التنسيق أو اضطراب البنية الشكلية للنص، وهو ما يتطلب جهدا إضافيا ويستغرق وقتا أطول.

وقد أشارت إحدى الأستاذتين إلى أن غياب التمييز بين الحروف الكبيرة والصغيرة، أو كثرة الأخطاء الإملائية، قد يعرقل عملية التقييم، إذ يضطر المقوم إلى إعادة قراءة الترجمة مرارا لفهم ما قصده الطالب. ومن التحديات الأخرى التي طرحت مسألة «أنسنة[x] التقييم»، أي إبداء قدر من التعاطف مع الطالب دون الإخلال بمقتضيات الموضوعية أو التساهل المفرط. فعلى سبيل المثال، قد يتغاضى الأستاذ عن بعض الهفوات الأسلوبية الطفيفة إذا كان الطالب يظهر فهما متينا للمفاهيم القانونية، وبالتالي الموازنة بين الموضوعية والعاطفة.

استندت معايير التقييم إلى سلم منضبط قسمت فيه الأخطاء إلى فئتين رئيستين: أخطاء المحتوى والأخطاء اللغوية أو البنيوية الأخرى. وقد أدرجت الأستاذتان ضمن أخطاء المحتوى كلا من سوء الترجمة، واستعمال المصطلحات المتخصصة غير الدقيقة، وحالات الحذف، في حين صنفت الأخطاء النحوية والاختلالات البنيوية والأخطاء الإملائية ضمن فئة الأخطاء الأخرى. غير أنه يجدر التنبيه إلى وجود فروقٍ طفيفة في نظام التنقيط المعتمد لدى كل من الأستاذتين. فالأستاذة (ب) جعلت أخطاء المحتوى تمثل 50% من العلامة، والأخطاء الأخرى 50% كذلك، أي بحد أقصى قدره 7.5 نقاط لكل فئة، في حين لم تعتمد الأستاذة (أ) هذا السقف عند التصحيح. كما يختلف مجموع العلامة بين النصين، إذ بلغ عند الأستاذة (ب) 15 نقطة، مقابل 20 نقطة عند الأستاذة (أ). وتتعلق الفروق الأخيرة بتفاصيل نظام التنقيط نفسه؛ إذ كانت الأستاذة (ب) تخصم 0.25 نقطة عن كل مصطلح تخصصي غير دقيق أو كلمة غير مترجمة، و0.5 نقطة عن كل عنوان أو عنوان فرعي أسيء ترجمته، أو عن كل جملة غير مترجمة. كما كانت تخصم 0.25 نقطة عن كل زوج من الأخطاء المتشابهة عند تعلقها بالنحو أو البنية أو الإملاء أو علامات الترقيم أو أدوات الربط أو ترتيب الكلمات أو الترجمة غير الدقيقة.

أما الأستاذة (أ)، فكانت تخصم 0.25 نقطة عن كل كلمة غير مترجمة، و0.5 نقطة عن كل مصطلح متخصص غير دقيق، وعن كل جملة غير مترجمة. كما تخصم 0.25 نقطة عن كل خطأ مستقل في النحو أو البنية أو الإملاء أو علامات الترقيم أو أدوات الربط أو ترتيب الكلمات. ويبين الجدول (3) أدناه معايير التقييم التي اعتمدتها

الأستاذتين (أ) و(ب).

جدول 3. معايير التقييم

الجانب	الأستاذة (أ)	مقدار الخصم	الأستاذة (ب)	مقدار الاقتطاع
أخطاء المحتوى	مصطلحات متخصصة غير دقيقة	0.5- نقطة عن كل مصطلح	مصطلحات تخصصية غير دقيقة	0.25-نقطة عن كل مصطلح
	كلمات غير مترجمة	0.25- نقطة عن كل كلمة	عناوين / عناوين فرعية مترجمة ترجمة خاطئة	0.5-نقطة عن كل حالة
	جمل غير مترجمة	0.5- نقطة عن كل جملة	جمل غير مترجمة	0.5-نقطة عن كل جملة
الأخطاء الأخرى	أخطاء نحوية، بنيوية، إملائية، علامات الترقيم، أدوات الربط، ترتيب الكلمات، عدم دقة الترجمة	0.25- نقطة عن كل خطأ	أخطاء نحوية، بنيوية، إملائية، علامات الترقيم، أدوات الربط، ترتيب الكلمات، عدم دقة الترجمة	0.25-نقطة عن كل زوج من الأخطاء من النوع نفسه
المجموع الكلي للعلامة	20نقطة		15نقطة	
سقف الخصم	لا يوجد سقف للاقتطاع		اعتماد سقف للاقتطاع (50% لأخطاء المحتوى و50% للأخطاء الأخرى)	

تعود الأسباب الكامنة وراء الفروق المشار إليها أعلاه إلى اختلاف ما تراه

الأستاذتين أولى بالاعتبار في عملية التقييم، فعلى سبيل المثال، ترى الأستاذة (أ) أن الخطأ في ترجمة المصطلحات المتخصصة أشد أثرا من الخطأ في ترجمة الألفاظ العامة، ولذلك خصصت له خصما قدره 0.5 نقطة. وأما الأستاذة (ب)، فقد أوضحت في المقابلة أنها تنظر إلى أخطاء ترجمة المصطلحات، بوجه عام، على أنها متساوية الوزن، من غير تمييز بين التخصصي وغيره. كما عمدت الأستاذة (ب) إلى وضع سقف قدره 50% لكل فئة من فئات الأخطاء (أخطاء المحتوى والأخطاء الأخرى)، وذلك للحيلولة دون تضخم الخصم الناتج عن تكرار النوع نفسه من الأخطاء. في المقابل، لم تعتمد الأستاذة (أ) هذا السقف، مبينة أنها أرادت للامتحان الأول أن يؤدي وظيفة تنبيهية، وأن يدفع الطلبة إلى التعامل مع أعمالهم الترجمية بجدية أكبر.

وقد أفضت المناقشات المتعمقة التي أجريت مع الأستاذتين خلال المقابلات الأولية إلى الانتقال نحو المرحلة التالية من هذه الدراسة، التي تمثلت في تصميم التوجيه الملائم لاستخدامه مع تشات جي بي تي 4o في عملية التقييم، وهو ما سيجري تفصيله في القسم التالي.

2.4. عملية تصميم الأوامر

صيغت الأوامر الموجهة إلى تشات جي بي تي 4o لتقييم الترجمات القانونية عبر سلسلة من الجلسات والمحاولات المتكررة، وقد كشفت المراحل الأولى عن عدد من الإشكالات المنهجية، إذ أظهرت الصيغ الأولية للأوامر أن النموذج يواجه صعوبة في التعرف على عناوين البنود على أنها عناوين فرعية، وغالبا ما كان يستبعدها من نطاق التقييم. كما أنه، في حال لم تدرج الأوامر صراحة تقضي بإدراج عناوين البنود ضمن التقييم، كان النموذج يتغاضى عنها ولا يعدها جزءا من النص محل التحليل. كما برزت إشكالية أخرى عند التعامل مع الفقرات المطولة، إذ أن عدم التنصيص في الأوامر على إمكانية اختصار المقاطع الطويلة باستعمال علامات الحذف (...) عند الاقتضاء، كان يؤدي إلى مخرجات مقتضبة على نحو مخل، مما يضعف عمق التحليل ودقته. كذلك، فإن غياب تعريف واضح لمفهوم «الترجمات المبتورة» في الأوامر أدى إلى تجاهل المقاطع الناقصة في نهاية النصوص، وعدم وسمها على أنها أخطاء. وفي ذات

السياق، فإن استخدام عبارة «ترجمات ناقصة» بدل «نصوص غير مترجمة» حال دون تمكن النموذج من رصد الكلمات أو الجمل التي تركها الطلبة دون ترجمة.

وتبرز هذه الملاحظات الدور الحاسم للدقة والوضوح في صياغة الأوامر، فالأوامر المحكمة ينبغي أن تكون محددة لا لبس فيها، وأن تزود تشات جي بي تي 4o بتعليمات واضحة بشأن ما ينبغي تقييمه وكيفية ذلك. كما يستلزم الأمر تضمين تعريفات وأمثلة عند الحاجة، لتوجيه فهم النموذج لطبيعة الأخطاء المقصودة. ويغدو اختيار المصطلحات أمرا بالغ الأهمية، إذ قد تفضي فروق طفيفة في الصياغة إلى اختلافات جوهرية في قدرة النموذج على اكتشاف أنواع معينة من أخطاء الترجمة. فضلا عن ذلك، ينبغي أن تتسم الأوامر بالشمول، فتفصل جميع العناصر الواجب إدراجها في التحليل.

وبناء على التحديات التي ذكرت أعلاه، فقد تبين أن الأوامر الفعالة الموجهة إلى تشات جي بي تي 4o لتقييم الترجمات القانونية ينبغي أن تمكن النموذج من:

- الالتزام بجميع الأوامر التزاما كاملا دون حذف.
- تقييم الترجمة تقييما شاملا من بدايتها إلى نهايتها.
- التقيد الصارم بمعايير التنقيط المعتمدة في التقييم البشري.
- رصد جميع الأجزاء غير المترجمة، سواء كانت كلمات أم جملا أم مقاطع ناقصة.
- التعرف على العناوين العامة والرئيسة والفرعية، بما في ذلك عناوين البنود، وإدراجها ضمن نطاق التقييم.
- شرح الأخطاء المرصودة وبيان أسباب عدها أخطاء، مع اقتراح تصحيحات مناسبة لها.

استوعبت الأوامر المعتمدة في هذه الدراسة نظام التنقيط نفسه الذي اعتمده المقيمون (الأساتذة)، فهذا يضمن إمكانية المقارنة المباشرة بين تقييمات تشات جي بي تي 4o والتقييمات البشرية. وقد نص في هذه الأوامر بوضوح على إلزام النموذج بمقارنة الترجمة كاملة بالنص الأصلي، ووسْمِ الأخطاء باستعمال أقواس مربعة، وتصنيفها وفق المعايير المحددة سلفا، مع تقديم شرح واضح لكل خطأ يرصد. كما ألزم

النموذج باقتراح تصحيحات مناسبة، وتطبيق سلم تنقيط منظم يقتضي خصم قيم نقطية محددة تبعا لنوع الخطأ. واشتملت المخرجات النهائية على تفصيل بنود الخصم والعلامة الإجمالية.

ومن خلال إدراج هذه العناصر في الأوامر، فقد كان متوقعا أن يقدم تشات جي بي تي 4o تقييما منظما مستوفيا لمقتضيات الدقة والشمول، ويعرض في القسم التالي تحليل أداء تشات جي بي تي 4o مقارنة بتقييمات الأساتذة.

3.4. مرحلة التحليل

كما ورد في قسم المنهجية، قدمت الأستاذتان نسخا مغفلة الهوية من ترجمات امتحانات منتصف الفصل من الإنجليزية إلى العربية، كانت قد صححت مسبقا ومرفقة بارتجاع من قبلهما. وقد اختير من كل أستاذة خمس ترجمات، ليبلغ مجموع مدونات الدراسة عشر ترجمات، روعي في انتقائها أن تمثل مستويات أداء متفاوتة: ضعيفة ومتوسطة ومرتفعة. وتضمنت الامتحانات نصين قانونيين: مقتطفا من عقد توكيل (Power of Attorney)، ومقتطفا من عقد عمل (Employment Contract). وبحسب إفادة الأستاذتين، تدرج هذه النصوص عادة في بدايات المقرر، لكونها تعد أقل تعقيدا نسبيا وأكثر ملاءمة للمستوى الأولي في الترجمة القانونية. وكما يبين الجدول (4) أدناه، بلغ عدد كلمات نص عقد التوكيل 250 كلمة، في حين بلغ نص عقد العمل 343 كلمة. وقد تراوحت علامات طلبة الأستاذة (أ)، الذين كلفوا بترجمة مقتطف عقد العمل، بين 8 و16.75 من مجموع 20 نقطة. أما طلبة الأستاذة (ب)، الذين ترجموا مقتطف عقد التوكيل، فقد تراوحت علاماتهم بين 6 و13 من مجموع 15 نقطة. ويذكر أن النصين كانا في أصلهما مكتوبين باللغة الإنجليزية، وقد طلِب من الطلبة نقلهما إلى العربية.

جدول 4. النصوص وعدد الكلمات وعلامات الطلبة

النص	عدد الكلمات	علامات الطلبة (من 15 نقطة)				
		ط 1	ط 2	ط 3	ط 4	ط 5
مقتطف من عقد توكيل (Power of Attorney)	250	6	7.5	10.5	12.75	13
النص	عدد الكلمات	علامات الطلبة (من 20 نقطة)				
		ط 1	ط 2	ط 3	ط 4	ط 5
عقد عمل (Employment Contract)	343	8	8.25	11.5	15.75	16.75

انطلقت جلسات التقييم بإدخال ترجمات الطلبة إلى تشات جي بي تي 4o واحدة تلو الأخرى، ومرفقة بالأوامر المحضرة والنص الأصلي، وبعد ذلك جمعت التقييمات الصادرة عن تشات جي بي تي 4o في ملف Word، أنشئ فيه جدولان لتدوين جميع معطيات التقييم. وقد اشتملت هذه الجداول على النص الأصلي، وترجمات الطلبة الخمسة، وتقييمات الأساتذة، وتقييمات تشات جي بي تي 4o، معروضة جنبا إلى جنب.

ويجدر التنبيه إلى أن هذه الدراسة لم تشهد تدريبا أو ضبطا دقيقا لأي نموذج، فقد استخدم تشات جي بي تي 4o على أنه نموذج لغوي كبير مدرب مسبقا طورته شركة أوبن أي آي OpenAI [27]، وولجنا إليه عبر الأوامر فقط، إذ استخدم النموذج بصورته الأصلية، من غير أي تدريب إضافي، وانصب التركيز على تصميم أوامر مخصصة توجه عملية تقييم نصوص الترجمة القانونية. وعليه، فإن مصطلح «التدريب» في هذه الدراسة يحيل إلى التحسين المتكرر للأوامر، لا إلى تدريب النموذج اللغوي نفسه.

وبعد استكمال ملء الجداول بالمقاطع المترجمة وتقييماتها، انتقلت الدراسة إلى مرحلة فحص ومقارنة تقييم تشات جي بي تي 4o لكل مقطع مترجم بتقييم الأستاذ المقابل له. وفي سبيل هذا التحليل المقارن، اعتمِد نظام ترميز محدد، كما هو مبين فيما يأتي:

الاتفاق (Agreement): يشير هذا المفهوم إلى المواضع التي تلاقى فيها تقييم الأساتذة وتقييم تشات جي بي تي 4o. وقد جرى تفعيل مفهوم الاتفاق بالاستناد إلى رصد النوع نفسه من الخطأ، لا إلى تطابق مقدار الخصم النقطي. وبعبارة أخرى، إذا حدد كل من الأداة والأساتذة الخطأ نفسه (كالأخطاء النحوية أو الترجمات الخاطئة أو أخطاء المصطلحات)، عد ذلك اتفاقا، حتى وإن اختلف مقدار الخصم المترتب على هذا الخطأ. أما الأحكام الأسلوبية أو التقديرية، فلم تحتسب ضمن الاتفاق إلا إذا ارتبطت بخطأ واضح قابل للتحديد.

الخلاف (Disagreement): يدل هذا المفهوم على الجوانب التي لم يتحقق فيها توافق بين تقييم الأساتذة وتقييم تشات جي بي تي 4o. وقد تميز نوعان من الاختلاف:

- أخطاء رصدها الأستاذ فقط: وتشمل الأخطاء التي تجاهلها تشات جي بي تي 4o أو لم يلتفت إليها في حين نبه إليها الأساتذة.
- أخطاء رصدها تشات جي بي تي 4o فقط: وتشمل الأخطاء التي لم ينتبه إليها الأساتذة، بينما كشف عنها تشات جي بي تي 4o.

ملاحظات أخرى (Other Observations): تبرز هذه الفئة أنماطا أو ملاحظات لافتة برزت أثناء المقارنة، ولا تندرج على نحو دقيق ضمن فئتي الاتفاق أو الاختلاف.

وإضافة إلى ذلك، شمل التحليل مقارنة كمية بين الدرجات التي منحها كل من الأساتذة وتلك التي أسندها تشات جي بي تي 4o. ويعرض وصف نتائج هذه المقارنة في القسم اللاحق.

4.4. مقارنة الأستاذة (أ) مع جي بي تي 4o

1.4.4. مواضع الاتفاق

أظهر التحليل وجود اتفاق واضح بين الأستاذة (أ) ومع جي بي تي 4o في رصد عدد كبير من الأخطاء الواردة في الترجمات، ومن أمثلة هذه الأخطاء، على سبيل التمثيل لا الحصر، ما يلي: «مارتش» (Martsh كونها نقلا صوتيا غير صحيح لكلمة March)، و«مفعولة الدخل» (affected income)، و«صدر» (issued)، و«الأتفاقية»

(the agreement مع خطأ في الهمزة)، و«الموظف» (the employee)، و«ويمثل» (and represents). وعلى الرغم من هذا الاتفاق في تحديد طبيعة الخطأ، فقد اختلف الطرفان أحيانا في مقدار الخصم المترتب عليه. فعلى سبيل المثال، خصم مع جي بي تي 4o نصف نقطة عن ترجمة March إلى «مارتش» بوصفها خطأ في مصطلح، في حين اكتفت الأستاذة بخصم ربع نقطة، إذ لم تعده مصطلحا تخصصيا. ومثال آخر يتمثل في ترجمة العبارة made entered into إلى «صدر»، حيث خصمت الأستاذة نصف نقطة لعدها سوء ترجمة مؤثرا في المعنى القانوني، بينما اكتفى مع جي بي تي 4o بخصم ربع نقطة فقط.

2.4.4. مواضع الخلاف

على الرغم من اتساع دائرة الاتفاق بين الأستاذة ومع جي بي تي 4o في رصد الأخطاء، فقد سجلت أيضا مواضع لم يتحقق فيها هذا التوافق، وقد تعود هذه الاختلافات إما إلى رصد الأستاذة أخطاء لم يلتفت إليها مع جي بي تي 4o، أو إلى العكس، أي اكتشاف مع جي بي تي 4o أخطاء لم تسجلها الأستاذة. ولتوضيح هذه الفروق، خصص القسمان الآتيان لعرض الأخطاء التي رصدتها الأستاذة وحدها، وتلك التي كشف عنها جي بي تي 4o دون غيره.

3.4.4. أخطاء رصدتها الأستاذة (أ) فقط

من بين الأخطاء التي رصدتها الأستاذة وحدها، تكرار غياب علامات الترقيم، إذ بدا أن GPT 4o لا ينتبه إلا إلى سوء استعمال علامة الترقيم، لا إلى غيابها أصلا. كما شملت هذه الفئة بعض الأخطاء الإملائية التي لم يلتقطها النموذج، مثل: «احمد» (Ahmed)، و«هويه» (identity)، و«اليه» (mechanism/procedure)، و«الاردنيه» (Jordanian). ومن الأمثلة اللافتة أيضا أن أحد الطلبة ترجم عنوان Employment Contract إلى «عقد العمل»، مضيفا أداة التعريف «ال»، وهو ما عدته الأستاذة خطأ يستوجب خصم نصف نقطة، في حين لم يسجله GPT 4o بوصفه خطأ. وتظهر هذه الأمثلة أن الأستاذة أولت عناية خاصة بعلامات الترقيم الإملاء والنحو.

4.4.4. أخطاء رصدها GPT 4o فقط

رصد جي بي تي 4o عددا من الأخطاء التي لم تنتبه إليها الأستاذة. فعلى سبيل المثال، اعتبر ترجمة «بين كلا من» (between both of) و«فيما بعد» (in what after) أخطاء لغوية، واقترح تصحيحها إلى «بين كل من» (between each of) و«فيما بعد» (hereinafter). كما لاحظ جي بي تي 4o أن أحد الطلبة ترجم الاسم العلم Wisam Khalil إلى «وسام خالد»، وهو خطأ لم تسجله الأستاذة. ومن الأمثلة الأخرى أن أحد الطلبة ترجم المصطلحين First Party وSecond Party إلى «القسم الأول» و«القسم الثاني»، فوسم جي بي تي 4o بوسْم هذا الاستعمال على أنه غير دقيق قانونيا، وخصم النقاط المترتبة عليه. وتظهر هذه الأمثلة أن جي بي تي 4o منح أولوية خاصة للصحة النحوية، والدقة الإملائية، والضبط المعجمي، والاستعمال السليم للمصطلحات القانونية.

ولوحظ كذلك أن إحدى الطالبات وقعت في تكرار لفظي بقولها «الموافق الذي يوافق»، وقد تنبه كل من الأستاذة وجي بي تي 4o إلى هذا الخلل، غير أن الخصم النقطي اقتصر على تقييم جي بي تي 4o وحده. كما أن عنوان النص ترجم إلى «عقد توظيف» من قِبل طالبين؛ ففي إحدى الحالتين عد جي بي تي 4o الترجمة مقبولة، بينما خصم في الحالة الأخرى نقاطا بدعوى اعتبارات أسلوبية. ويلاحظ أيضا أن جي بي تي 4o بدا متسامحا مع معظم الأخطاء الإملائية، في مقابل حساسية أوضح تجاه الأخطاء البنيوية والأسلوبية التي لم تسجلها الأستاذة.

وخلاصة هذا التحليل أن قدرا معتبرا من الاتفاق تحقق بين التقييمين في تشخيص العديد من الأخطاء، مع تسجيل فروق في مقدار الخصم النقطي. كما برزت مواضع اختلاف، تمثلت في رصد الأستاذة أخطاء أغفلها جي بي تي 4o، ورصد الأخير أخطاء لم تدرجها الأستاذة. وأظهرت الملاحظات الإضافية أن بعض الإشكالات أقِر بها في كلا التقييمين، غير أن الخصم اقتصر أحيانا على التقييم الآلي، فضلا عن تسجيل قدر من التفاوت في الأحكام الأسلوبية الصادرة عن جي بي تي 4o.

5.4. مقارنة الأستاذة (ب) مع جي بي تي 4o

1.5.4. مواضع الاتفاق

أظهر التحليل وجود اتفاق واسع بين الأستاذة (ب) وجي بي تي 4o في رصد عدد كبير من الأخطاء، شملت، على سبيل المثال لا الحصر، عدم دقة الصياغة الترجمية، والأخطاء المصطلحية والنحوية. فقد اتفق الطرفان، على سبيل المثال، على أن عنوان Power of Attorney لا يصح نقله إلى «الوكالة» فحسب، بل ينبغي ترجمته إلى «وكالة قانونية» لما يقتضيه السياق القانوني من تخصيص ودقة. كما أجمع الطرفان على عد ترجمات عدد من المصطلحات القانونية أخطاء، من ذلك ترجمة Legal إلى «قضائية»، وSafety Deposit Boxes إلى «صناديق الإدخار»، وVaults إلى «سراديب»، و Real Estate Transactions إلى «الصفقات العقارية»، وقد عدت هذه الترجمات غير دقيقة من حيث المعنى القانوني والمجال المصطلحي من كلا الطرفين.

2.5.4. مواضع الخلاف

ومع أن دائرة الاتفاق كانت واسعة، فقد سجلت بعض المواضع التي لم يتحقق فيها التوافق بين الأستاذة (ب) وجي بي تي 4o، ويعود ذلك إما إلى رصد الأستاذة أخطاء لم ينتبه إليها النموذج، أو إلى العكس، أي اكتشاف جي بي تي 4o أخطاء لم تدرجها الأستاذة في تقييمها. ولإبراز هذه الفروق، خصص القسمان الآتيان لعرض الأخطاء التي رصدتها الأستاذة وحدها، وتلك التي كشف عنها جي بي تي 4o.

3.5.4. أخطاء رصدتها الأستاذة (ب) فقط

يلاحظ أن ثمة أخطاء عدتها الأستاذة وحدها أخطاء، في حين لم يصنفها جي بي تي 4o كذلك، فعلى سبيل المثال، رأت الأستاذة أن ترجمة الفعل handle إلى «يتحكم» تمثل خطأ دلاليا، لما تنطوي عليه من تحميل زائد لمعنى السيطرة، في حين لم يعد جي بي تي 4o هذا الاستعمال خطأ. بل إن النموذج نفسه اقترح لفظ «التحكم» كونه المقابل الأنسب عندما ترجم أحد الطلبة كلمة control إلى «سيطرة»، وهو ما يظهر اختلافا في تقدير الفروق الدلالية الدقيقة بين هذه الألفاظ. كما اعتبرت الأستاذة

عبارتي «شؤوني التالية» و«الأمور التالية» صيغتين غير سليمتين أسلوبيا أو دلاليا في السياق القانوني، في حين لم يسجل جي بي تي 4o أي ملاحظة بشأنهما. ومن الأمثلة الأخرى ترجمة banking transactions إلى «المعاملات البنكية»، إذ عدت الأستاذة لفظ «البنكية» غير دقيق اصطلاحيا، واقترحت استبداله بـ«المصرفية». وتظهر هذه الأمثلة أن الأستاذة أولت أولوية خاصة للدقة الدلالية، وحسن الاختيار الأسلوبي والمفرداتي، والاستعمال السليم للمصطلحات.

4.5.4. أخطاء رصدها جي بي تي 4o فقط

يمكن القول من المعطيات المعروضة في الجدول أن جي بي تي 4o أولى اهتماما ملحوظا بالأخطاء الإملائية، مثل تلك الواردة في «إسم»، و«الإستثمارات»، و«الإتفاقيات»، و«التي امتلكها»، فبادر إلى تصويبها إلى «اسم»، و«الاستثمارات»، و«الاتفاقيات»، و«التي أمتلكها»، في حين منحت الأستاذة هذه الأخطاء وزنا أقل في التقييم. وقد يعزى تجاهل الأستاذة لهذه الأخطاء إلى كونها أنواعا أخرى من الأخطاء أكثر تأثيرا في جودة الترجمة، أو إلى رغبتها في عدم التشديد المفرط، ولا سيما عندما يكون النص مثقلا أصلا بعدد من الأخطاء. ومن الأمثلة الأخرى التي انفرد جي بي تي 4o برصدها ترجمة كلمة hospitalization إلى «التنويم»، وهي ترجمة عدتها الأستاذة مقبولة، في حين اعتبرها جي بي تي 4o غير دقيقة، واقترح بدلا منها «الدخول إلى المستشفى» تحقيقا لمزيد من الضبط الدلالي. وتبرز هذه الأمثلة أن جي بي تي 4o منح أولوية لمسائل الإملاء، كما كشف عن اختلاف مصطلحي مع الأستاذة في بعض الحالات.

كما لوحظ أن الأستاذة (ب) غالبا ما لم تكن تفصح عن نوع الخطأ ولا تقدم مقترحات تصحيحية لجميع الأخطاء التي ترصدها. وقد عزت ذلك إلى كثرة أوراق الامتحان الموكلة إليها للتصحيح، وضيق الوقت المتاح للتقييم، أو إلى تفضيلها، في بعض الأحيان، أن يراجعها الطلبة مباشرة لطلب الارتجاع عند استلام أوراقهم المصححة. وفي المقابل، كان جي بي تي 4o يميل، في جميع الحالات تقريبا، إلى تحديد نوع الخطأ، وبيان سبب عده خطأ، وتقديم اقتراحات أو بدائل ترجمية مناسبة. ويعزى

ذلك إلى طبيعة الأوامر التي صاغها الباحث، والتي طلبت من النموذج صراحة تبرير أحكامه وتقديم مقترحات تصحيحية. ومن جهة أخرى، لوحِظ أن تعامل جي بي تي 4o مع الأخطاء الإملائية اتسم بعدم الاتساق؛ فكما سجل سابقا في ترجمات طلبة الأستاذة (أ)، أغفل النموذج عددا من الأخطاء الإملائية، في حين عمد، في ترجمات طلبة الأستاذة (ب)، إلى معاقبة حتى الاختلالات الإملائية الطفيفة. وقد يعزى هذا التفاوت إلى كيفية اندماج الخطأ في السياق، إذ قد يؤثر السياق المحيط في درجة حساسية النموذج تجاه الخطأ. وتطرح تفسيرات أخرى محتملة، من بينها أن بعض الأخطاء الإملائية قد تكون ناتجة عن سوء تفسير على مستوى الوحدات الرمزية (tokens) أو عن انحيازات كامنة في بيانات التدريب. فعلى سبيل المثال، قد تتكرر بعض الصيغ غير القياسية في الاستعمال الرقمي أو في العربية غير الرسمية، مما يدفع النموذج إلى التعامل معها بوصفها متغيرات مقبولة. كما قد يميل النموذج إلى إعطاء الأولوية للأخطاء ذات الأثر الأشد في المعنى على حساب الهفوات الإملائية البسيطة، ولا سيما عندما تتعدد أنواع الأخطاء في المقطع ذاته.

6.4. اتساق الدرجات بين التقييم الآلي والتقييم البشري

يبين الجدول (5) أدناه الدرجات التي أسندتها الأستاذة (أ) وجي بي تي 4o لخمسة طلبة جرى تقييمهم على سلم من 20 نقطة، كما يعرض الدرجات التي منحتها الأستاذة (ب) وتشات جي بي تي 4o لخمسة طلبة آخرين جرى تقييمهم على سلم من 15 نقطة.

جدول 5. درجات الطلبة كما حددها الأستاذتان وتشات جي بي تي 4o.

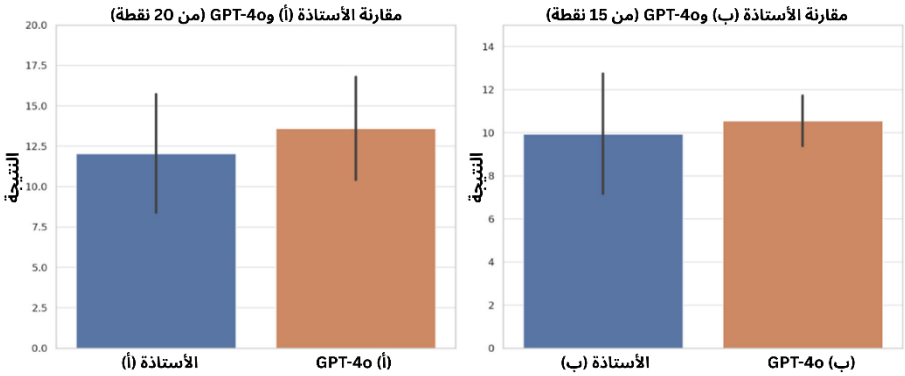
الطالب	الأستاذة (أ) (20 نقطة)	جي بي تي 4o (20 نقطة)	الفرق أ – جي بي تي 4o
1	11.5	10.5	1.0
2	8.0	9.5	−1.5
3	8.25	14.25	−6.0
4	15.75	18.0	−2.25

5	16.75	15.75	1.0
الطالب	**الأستاذة (ب) (15 نقطة)**	**جي بي تي 4o (15 نقطة)**	**الفرق ب – جي بي تي 4o**
1	7.5	9.75	−2.25
2	10.5	9.25	1.25
3	12.75	12.5	0.25
4	13.0	11.25	1.75
5	6.0	10.0	−4.0

للتحقق مما إذا كانت الفروق بين الدرجات التي منحها الأساتذة وتلك التي أسندها تشات جي بي تي 4o ذات دلالة إحصائية، أجري اختبار (t) للعينات المزدوجة، وحسِبت قيمة p، كما هو مبين في الجدول (6) أدناه. ووفقا لـ[28]، يعد اختبار (t) صالحا للاستخدام حتى مع أحجام عينات صغيرة، ولا سيما في التصاميم المزدوجة، شريطة أن تكون فروق الأزواج موزعة توزيعا يقارب التوزيع الطبيعي. كما يوضح الشكل (2) مخططا عموديا يقارن متوسط الدرجات بين تقييمات الأساتذة وتقييمات تشات جي بي تي 4o لكل مجموعة من الطلبة.

جدول 6. فروق الدرجات لدى الطلبة العشرة.

البند	**الأستاذة (أ) مقابل تشات جي بي تي 4o (20 نقطة)**	**الأستاذة (ب) مقابل تشات جي بي تي 4o (15 نقطة)**
متوسط الفروق	−1.55	−0.60
الانحراف المعياري للفروق	2.89	2.45
قيمة إحصاء t	−1.20	−0.55
درجات الحرية	4	4
قيمة p	0.296	0.613



شكل 2. مخطط عمودي يقارن متوسط الدرجات بين تقييمات الأساتذة وتقييمات تشات جي بي تي 4o لكل مجموعة.

أظهرت النتائج وجود قدر من التباين بين تقييمات الأساتذة البشر وتقييمات الذكاء الاصطناعي. غير أن قيمة p كانت أكبر من 0.05، وهو ما يدل على عدم وجود فروق ذات دلالة إحصائية بين الدرجات التي منحها الأساتذة وتلك التي قدمها نموذج GPT 4o.

وعلاوة على ذلك، حُسب حجم الأثر باستخدام معامل d لكوهين (Cohen's d) للعينات المترابطة، وذلك بالاعتماد على الصيغة الآتية:

$$d = \frac{\bar{d}}{s_d}$$

نظرا لصغر حجم العينة، يسهم احتساب أحجام الأثر في تقييم مدى الفروق بين درجات الأساتذة ودرجات نموذج جي بي تي 4o بدقة كما هو موضح في الجدول 7 أدناه.

جدول 7. أحجام الأثر المحسوبة باستخدام معامل d لكوهين (Cohen's d).

	معامل كوهين d	تفسير حجم الأثر
الأستاذ (أ) مقابل تشات جي بي تي 4o	−0.54	متوسط
الأستاذ (ب) مقابل تشات جي بي تي 4o	−0.25	صغير

وأظهرت النتائج حجم أثرٍ متوسطا لدى الأستاذة (أ) (d = −0.54)، وهو ما يدل على أن هذه الأستاذة كانت في المتوسط تمنح الطلبة درجات أعلى قليلا من تلك التي منحها لهم تشات جي بي تي 4o. وفي المقابل، كان حجم الأثر لدى الأستاذ (ب) صغيرا (d = −0.25)، وهذا يشير إلى تقارب أوثق بين درجاته ودرجات تشات جي بي تي 4o مع فروق طفيفة.

7.4. التقييم العام لأداء جي بي تي 4o

من أبرز ما يلفت النظر في أداء تشات جي بي تي 4o درجة الاتساق التي اتسمت بها عملياته في تقييم ترجمات الطلبة؛ إذ التزم النموذج على امتداد جميع التقييمات بالفئات نفسها للأخطاء، وبالآليات ذاتها في اقتطاع الدرجات وبنية موحدة للارتجاع، وهو ما يزيد من موثوقية تقييماته. وقد جرى كل تقييم وفق نسق منهجي واضح، يبدأ بفحص دقة المصطلحات، ثم ينتقل إلى رصد الأخطاء النحوية، فمواطن الخلل البنيوي، ليختتم بملخصٍ إجمالي يتضمن الدرجة العددية. ويسهم هذا الاتساق في جعل تقييمات تشات جي بي تي4o سهلة القراءة وواضحة الدلالة وميسرة التأويل، ولا سيما عند إجراء المقارنات بين عدد من الترجمات.

وإلى جانب ذلك، اتسمت تفسيرات تشات جي بي تي 4o في الغالب بالوضوح والإيجاز، إذ كان يقرن التصويبات المباشرة بتعليلات موجزة تبين سبب وسم الخطأ بوصفه خطأ. فعند رصد خطأ نحوي، لا يقتصر النموذج على الإشارة إلى عدم سلامة الصيغة، بل يقدم البديل الصحيح ويرفقه بتبرير مختصر يوضح دواعي التعديل. ومن شأن هذا المستوى من الشرح أن يعزز القيمة التعليمية لتقييمات تشات جي بي تي 4o، ويجعلها أداة فعالة في دعم تعلم الطلبة وخدمة الأغراض التدريسية.

ولوحظ في بعض الحالات أن هذا الاتساق لم ينسحب على تنسيق المخرجات بالدرجة نفسها، فبينما عرضت تقييمات الطلبة الأوائل ضمن البنية نفسها، لم يحافظ على هذا النسق مع الطلبة اللاحقين. فعلى سبيل المثال، تمكن تشات جي بي تي4o في تقييمات الطلبة الثلاثة الأوائل من عرض نص المنقول منه، والنص المترجم، وأنواع الأخطاء، والتعليقات التفسيرية، ومقدار الاقتطاعات لكل فقرة، ثم اختتم

التقييم بحساب الدرجة النهائية وتقديم ملاحظات عامة. غير أن هذا التنسيق لم يتبع إلا في تقييمات الطلبة الثلاثة الأوائل فقط. وأما في حالة الطالب الرابع، فقد أغفل النموذج عرض النص المنقول منه، مكتفيا بإدراج نص الترجمة ومتبوعا بالأخطاء والتعليقات والاقتطاعات، ثم الدرجة النهائية والملاحظات العامة. وفي تقييم الطالب الخامس، تغير التنسيق مرة أخرى، إذ اقتصر العرض على الجمل التي تضمنت أخطاء في كل فقرة، تليها أنواع الأخطاء والتعليقات والاقتطاعات، ثم التقييم النهائي والملاحظات العامة. بعبارة أخرى، وعلى الرغم من أن محتوى التقييم ومعاييره ظل متسقا في جميع الحالات، فإن تنسيق المخرجات شهد تباينا ملحوظا في الجلسات اللاحقة، غير أنه، وعند إعادة إدخال الأمر نفسه وترجمات الطلبة اللاحقين في يوم آخر، تبين أن مشكلة التنسيق المشار إليها قد زالت، وعاد النسق الموحد إلى الظهور، ويحتمل أن يشير ذلك إلى أن التفاوت في تنسيق المخرجات لدى الطلبة المتأخرين يعود إلى حدود الاستخدام المرتبطة بـ جي بي تي 4o.

8.4. رأي الأستاذتين

يتناول هذا القسم الفوائد والتحديات المحتملة لاستخدام الذكاء الاصطناعي في تقييم جودة الترجمة، إلى جانب آراء أخرى لم تندرج ضمن المحورين السابقين. وقد استندت هذه الآراء إلى تصورات الأستاذتين، وقد جمعت خلال المرحلة الثانية من المقابلات شبه الموجهة.

1.8.4. آراء حول درجة التو افق بين GPT 4o والأستاذتين

عبرت كلتا الأستاذتين عن موقف محبذ لمستوى التوافق بين تقييماتهما والتقييمات التي قدمها جي بي تي 4o، فقد لاحظت الأستاذة (ب) أن الفروق بين تقييمها وتقييم الذكاء الاصطناعي كانت طفيفة، ولا تختلف كثيرا عن التباينات المعتادة التي تظهر بين المقيمين البشر داخل المجال نفسه. وذكرت أن ارتجاع الذكاء الاصطناعي بدى لها وكأنها «رأي أستاذ آخر»، وأبدت دهشتها الإيجابية من كثرة حالات التطابق بين تقييماتها وتقييمات جي بي تي 4o، وأضافت قائلة: «يمكننا الاعتماد عليه،

ولكن ليس بنسبة 100%، ليس بعد»، مؤكدة أن الاعتماد الكامل عليه قد لا يكون مناسبا في هذه المرحلة، غير أن درجة الاتساق وضآلة هامش الخطأ تعد مؤشرات واعدة. وحسب رأيها، فإن هذا المستوى من التوافق يشكل دليلا على أن الذكاء الاصطناعي يمكن أن يكون أداة مفيدة لكل من الأساتذة والطلبة في سياقات التقييم الأكاديمي، وقد خلصت إلى أن النتائج الرقمية ومستوى التوافق العام مشجعان ويدعمان إمكانية دمج هذه الأداة في البيئات التعليمية.

من جهتها، أقرت الأستاذة (أ) هي الأخرى بوجود مستوى عالٍ من التوافق بين تقييماتها وتقييمات الذكاء الاصطناعي، إذ علقت قائلة: «وجدت الأمر مثيرا للاهتمام جدا، وله إمكانات واعدة»، وأضافت: «أعتقد أن هناك مستوى جيدا من التوافق». وعزت هذا الانطباع إلى جملة من العوامل، من بينها جودة صياغة الأوامر الموجهة، ودقة معايير التقييم المقدمة للنموذج، فضلا عن الشواهد التي أظهرتها النتائج. كما أشارت الأستاذة (أ) إلى أن اتساق الذكاء الاصطناعي يعد نقطة قوة أساسية أسهمت في تحقيق هذا التوافق، وعززت من جدواه في السياق الأكاديمي.

2.8.4. المكاسب المحتملة

أقرت الأستاذتان بوجود جملة من المكاسب المشتركة المترتبة على توظيف الذكاء الاصطناعي في تقييم جودة الترجمة، فقد اتفقتا على أن الذكاء الاصطناعي لا ينبغي أن يكون بديلا عن التقييم البشري، بل مكملا له، وصرحت الأستاذة (أ): «نعم، هما يكملان بعضهما»، فيما أكدت الأستاذة (ب) المعنى نفسه بقولها: «يمكن للذكاء الاصطناعي أن يكمل عمل الإنسان». وتتيح هذه العلاقة التكاملية للأساتذة الاحتفاظ بالإشراف الأكاديمي والحكم المهني، مع الاستفادة في الوقت نفسه من اتساق الذكاء الاصطناعي وسرعته. كما لاحظت الباحثة أن تشات جي بي تي 4o يسهم إسهاما ملموسا في تقليص الزمن اللازم لتقييم الترجمات، بفضل قدرته على إنتاج تقييمات مفصلة في وقت وجيز، وهو ما يعزز دوره في ترشيد عملية التقييم وتيسيرها.

وشدد الطرفان كذلك على قدرة الذكاء الاصطناعي على تعزيز استقلالية المتعلمين ودعم مسار التعلم، فقد أفادت الأستاذة (أ) بأن بعض طلبتها لاحظوا تحسنا

في أدائهم عند استخدام أدوات الذكاء الاصطناعي في تقييم اللغة، قائلة: «أريدهم أن يكونوا مستقلين... أن توفر لهم مصدرا إضافيا للإرشاد فيما يتعلق باللغة وبناء الجملة». وعلى المنوال نفسه، رأت الأستاذة (ب) أن «أي ارتجاع يمكن أن يكون مفيدا، سواء صدرت عن الأستاذ أو عن الذكاء الاصطناعي». وتماشيا مع هذه الآراء، لاحظت الباحثة أن الشروح التفاعلية التي يقدمها تشات جي بي تي 4o قد تساعد الطلبة على استكشاف أخطائهم وفهمها بصورة ذاتية، وهذا يسهم في تنمية استقلاليتهم التعلمية.

ومن المكاسب الأخرى التي أجمعت عليها الأستاذتان إمكانية تقليص الجهد والوقت المبذولين في تقييم أعمال الطلبة، فقد قالت الأستاذة (ب): «هذا ما نحتاجه، جهد أقل ونتائج دقيقة»، في حين نظرت الأستاذة (أ) إلى الذكاء الاصطناعي على أنه نقطة انطلاق مفيدة يمكن اعتمادها خطوة أولى، ثم يعقبها تحقق بشري لاحق من التقييم. كما اتفقتا على أن جودة صياغة الأوامر تؤدي دورا حاسما في إنتاج مخرجات ذات قيمة، فقد أشارت الأستاذة (أ) إلى أن الذكاء الاصطناعي استطاع «تحديد الأخطاء وتبرير عدها أخطاء وشرحها»، بينما رأت الأستاذة (ب) أن «الأوامر ساعدت الذكاء الاصطناعي على محاكاة الطريقة التي يفكر بها الإنسان عند التعامل مع الترجمات». وقد أكدت الباحثة بدورها أن الأوامر المحكمة كانت عاملا حاسما في موثوقية المخرجات، إذ أتاح التطوير الدقيق والتكراري لها تحقيق نتائج متسقة ودقيقة.

وبعد الاطلاع على التقييمات التي أنجزها الذكاء الاصطناعي أفادت الأستاذتين بحدوث تحول ملحوظ في نظرتهما إليه، وقد علقت الأستاذة (أ) بقولها: «ثمة إمكانات حقيقية... وكان التقييم مقنعا»، فيما أشارت الأستاذة (ب) إلى تغير موقفها بوضوح قائلة: «تغير رأيي... وأصبحت محفزا أكثر على استخدامه».

كما رصدت الباحثة فائدة إضافية تتمثل في قدرة تشات جي بي تي 4o على دعم توحيد ممارسات التقييم بين الأساتذة المختلفين، فاعتماد أوامر موحدة، منسجمة مع شبكات تقييم مشتركة من شأنه الإسهام في تحقيق تقييمات أكثر موضوعية واتساقا، وقد عدت واجهة الأداة السهلة الاستخدام عنصرا عمليا داعما لاعتمادها لأنها تتيح للمقيمين إدخال استفساراتهم دون جهد تقني يذكر. ، كما أن الإتاحة

المجانية لتشات جي بي تي 4o تزيد من جاذبيته في السياقات الأكاديمية، ولا سيما في المؤسسات التي تعاني من محدودية الموارد، على الرغم من وجود حدود يومية للاستخدام.

3.8.4. التحديات المحتملة

على الرغم من هذه المزايا، فقد أبرزت الأستاذتان جملة من التحديات المشتركة، وشددتا على أهمية التكوين، إذ تركيز الأستاذة (أ) على الحاجة إلى «تدريب مخصص للاستعمال المهني»، في حين أكدت الأستاذة (ب) أن «استخدامه يتطلب تدريبا للاعتياد عليه»، واتفقتا على أن صياغة الأوامر تمثل عنصرا حاسما في توجيه الأداة توجيها سليما. فقد نبهت الأستاذة (أ) إلى أن جودة المخرجات تبقى رهينة كون الأوامر مصممة أصلا بما يتلاءم مع معايير تقييمها، بينما أوضحت الأستاذة (ب) بقولها: «الأمر يعتمد على كيفية توجيه الذكاء الاصطناعي... فالنتائج ستظل في النهاية مرتبطة بالعنصر البشري». وإلى جانب ذلك، طرحت إشكالات عملية، ولا سيما ما يتعلق بصيغة أعمال الطلبة. فقد أشارت الأستاذة (أ) إلى أن العملية تصبح «مضيعة للوقت إذا لم تقدم الترجمات في صيغة رقمية»، بينما رأت الأستاذة (ب) أن اعتماد إجابات مكتوبة رقميا «يسهل إدخالها إلى الذكاء الاصطناعي». وقد لاحظت الباحثة بدورها أن تحويل ترجمات الامتحانات المكتوبة بخط اليد إلى نصوص رقمية كان خطوة ضرورية، لكنها مستهلكة للوقت، من أجل تهيئة النصوص للإدخال إلى الأداة. ورغم أن مصطلح «الرقمنة» لم يذكر صراحة من قبل الأستاذتين، فإن هذه العملية شكلت عائقا عمليا قد يحد من الاستثمار السلس لأدوات الذكاء الاصطناعي.

فضلا عن ذلك، أكدت الأستاذتان أن الذكاء الاصطناعي يفتقر إلى بعض أبعاد الحكم البشري والتفاعل التربوي، فقد أوضحت الأستاذة (ب) أنه «عند النقاش والجدل حول الأخطاء مع الطلبة، قد لا يكون الذكاء الاصطناعي قادرا على القيام بذلك»، أي إن هذه الأداة قد تعجز عن دعم مستوى عالٍ من التفاعل المستمر والحوار التربوي بين المقيم والطلبة.

تحد آخر لاحظته الباحثة يتمثل في عدم انتظام تنسيق المخرجات، فبينما

التزمت التقييمات الأولى ببنية واضحة تضم النص الأصلي والنص المترجَم والأخطاء المرصودة والتعليقات وقيم التنقيط المخصومة والنتيجة النهائية، انحرفت التقييمات اللاحقة عن هذا النسق رغم إعادة استعمال الأوامر نفسها. ويبدو أن هذا التذبذب مرتبط بحدود الاستخدام أو بسلوك الخوادم؛ إذ إن إعادة إدخال المعطيات نفسها في يوم مختلف أعادت البنية الأصلية المنتظمة. ومع أنّ مضمون التقييم ظلّ ثابتًا في جوهره، فإن هذا التفاوت في العرض يثير إشكالًا عمليًا، ولا سيما عند توظيف أدوات الذكاء الاصطناعي في سياقات تقييم رسمية أو مقارنات منهجية.

وإلى جانب رصد المزايا والتحديات، عبرت الأستاذتان عن انفتاحهما على مزيد من التجريب في توظيف الذكاء الاصطناعي، فقد أشارت الأستاذة (أ) إلى رغبتها في «الإبقاء على إمكانية استخدامه كمرحلة ثانية»، في حين أكدت الأستاذة (ب): «علينا أن نجرب كل شيء ثم نقرر بناء على النتائج». كما أقرتا بالتطور المتسارع لأدوات الذكاء الاصطناعي؛ إذ علقت الأستاذة (ب) بقولها: «أعتقد أن الذكاء الاصطناعي سيتفوق علينا يوما ما»، بينما اكتفت الأستاذة (أ) بالقول: «هناك إمكانات». وتعكس هذه التصريحات موقفا يتسم بالحذر، لكنه ينطوي في الوقت نفسه على اهتمام متزايد بتبني الذكاء الاصطناعي، ولا سيما حين يوظف بوصفه أداة داعمة تعزز الخبرة البشرية، لا بديلا عنها، في سياق التقييم الأكاديمي.

على الرغم من التفاؤل والانفتاح اللذين عبرت عنهما الأستاذتان تجاه توظيف تشات جي بي تي 4o، فإن تأملاتهما، مقرونة بما كشف عنه من حدود وإشكالات، تؤكد أن هذه الأداة تؤدي دورها على الوجه الأمثل كونها وسيلة مسانِدة لا بديلا عن الحكم الخبير. وتدعم تأملات الباحثة هذا الاستنتاج؛ إذ يتبين أن الموضع الأنسب لـ تشات جي بي تي 4o هو أن يدمج في منظومة التقييم بوصفه أداة دعم، فرغم ما يحضى به من مزايا لافتة كسرعة الإنجاز واتساق النتائج وسهولة الولوج وجودة الارتجاع من حيث الوضوح والبناء، فإن هذه المكاسب تظل رهينة الموازنة الدقيقة بينها وبين ضرورة الإشراف البشري، وحسن الإدماج البيداغوجي، واعتماد بروتوكولات استعمال منظمة. وعندما يحسن توظيف أدوات الذكاء الاصطناعي وتدمج بعناية، فإنها يمكن أن تتحول إلى رافد فعال يُسهم في دعم الأساتذة ويُثري الممارسات التدريسية ويعزز

خبرات التعلم لدى الطلبة.

ويخصص القسم الآتي لمناقشة النتائج التي عرضت في هذا المحور.

5. المناقشة

سعت هذه الدراسة إلى استقصاء أمرين رئيسين: (1) مدى الاتساق بين تقييمات أدوات الذكاء الاصطناعي، وخاصة تقييمات تشات جي بي تي o4، وتقييمات أساتذة ذوي خبرة في مقرر للترجمة القانونية، و(2) ما ينطوي عليه توظيف هذه الأدوات في هذه السياقات من فوائد وتحديات محتملة. وتناقش هذه الفقرة النتائج المتحصل عليها في ضوء الدراسات السابقة مع إبراز دلالاتها البيداغوجية.

1.5. الاتساق مع تقييمات الأساتذة

أظهرت النتائج في مجملها مستوى مرتفعا من الاتساق بين تقييمات تشات جي بي تي 4o وتقييمات الأساتذة، ولا سيما في رصد أخطاء المحتوى الأساسية والأخطاء اللغوية في ترجمات الطلبة، فحين صيغ الأمر صياغة محكمة، أبان تشات جي بي تي 4o قدرة واضحة على التعرف المنتظم إلى أخطاء الترجمة وتصنيفها، بما في ذلك الترجمات الخاطئة والمصطلحات المتخصصة غير الدقيقة والحذف ومظاهر عدم الاتساق النحوي. وأكد التحليل الإحصائي عدم وجود فروق ذات دلالة إحصائية بين الدرجات التي أسندها الأساتذة وتلك التي قدمها تشات جي بي تي 4o، وهو ما يدل على تقارب ملحوظ في نتائج التقييم. وتتسق هذه المعطيات مع ما خلصت إليه دراسات سابقة، مثل [6،19،29]، التي أبلغت عن درجات ارتباط قوية بين التقييمات الآلية والتقييمات البشرية في سياقات ترجمية متنوعة.

وما يلفت النظر، على وجه الخصوص، أن المرجع [19] توصل إلى أن تقييمات تشات جي بي تي كثيرا ما تعكس أحكام المقيمين البشر، بل وتتفوق عليها أحيانا من حيث الاتساق، وهو ما يتوافق مع ما لاحظته هذه الدراسة من اعتماد تشات جي بي تي 4o مقاربة منظمة ومنهجية في التقييم. ولكن، في حين ركزت تلك الدراسة أساسا على سياقات ترجمية عامة، توسع الدراسة الحالية نطاق هذه النتائج من خلال إبراز

قدرة تشات جي بي تي 4o على الأداء بكفاءة في مجالات متخصصة، مثل الترجمة القانونية، متى توافر تصميم توجيه دقيق ومفصل. وبذلك، تسهم هذه الدراسة في إثراء الأدبيات القائمة، وتدعم الرأي القائل بأن أدوات الذكاء الاصطناعي تنطوي على إمكانات معتبرة في تقييم جودة الترجمة تتجاوز نطاق الاستعمال اللغوي العام، وهو ما يتقاطع أيضا مع ما ناقشه المرجع [29].

وعلى الرغم من هذا الاتساق العام، كشفت التحليلات النوعية عن فروق ملحوظة بين تقييمات تشات جي بي تي 4o وتقييمات الأساتذة، فقد أغفل تشات جي بي تي 4o في بعض الحالات أخطاء رصدها المقيمون، في حين أبدى تشددا زائدا في رصد بعض الهفوات البنيوية أو الأسلوبية الطفيفة، فعلى سبيل المثال، اقتطع نقاطا بسبب عدم وجود اتساق إملائي عالجه الأساتذة إما بالتجاهل أو بتسامح محسوب، مراعاة لقصد الطالب والاعتبارات البيداغوجية. ويتوافق هذا مع ما أشار إليه كل من [9،20] من أن الارتجاع الآلي قد يتسم أحيانا بالصرامة المفرطة، ويعجز عن التمييز بين الأخطاء الجوهرية وغير الجوهرية، وهذا قد يضعف الأثر التعليمي إذا لم يدعم بحكم بشري.

وتبرز هذه الفروق الاختلاف الجوهري بين أنظمة الذكاء الاصطناعي القائمة على القواعد والإجراءات، وبين المقيمين الذين يستندون إلى حكم مهني يتشكل في ضوء السياق وأهداف المقرر واحتياجات المتعلمين، ويشير ذلك إلى أن تكييف صيغ التوجيه لتتضمن تعليمات مخصصة للمقرر وقوائم مصطلحات قانونية متخصصة، من شأنه أن يعيد ضبط حساسية التقييم الآلي بحيث تكون أقرب إلى توقعات الأساتذة، وهو توجه تدعمه كذلك دراسات [25،26].

وفضلا عن ذلك، أبرزت النتائج الدور الحاسم لتصميم التوجيهات، فقد أفضت التوجيهات الغامضة خلال المرحلة التجريبية إلى تجاهل تشات جي بي تي 4o لبعض العناوين أو العبارات غير المترجمة، في حين أدى تنقيح التوجيهات وتحسين صياغتها إلى تحسن ملموس في أداء التقييم. ويتوافق هذا الاستنتاج مع ما خلص إليه [25،26] من أن التوجيهات المحكمة، المصممة خصيصا لمهام الترجمة، تفضي إلى تقييمات أقرب إلى الأحكام البشرية، ولا سيما في النصوص المتخصصة. لذلك، تؤكد

هذه الدراسة مجددا أن امتلاك كفاءة في هندسة الأوامر[xi] يعد شرطا أساسا لتفعيل الإمكانات الكاملة للذكاء الاصطناعي في تقييم جودة الترجمة.

ويجدر التنبيه إلى أن الأوامر المقدمة إلى أداة الذكاء الاصطناعي قد صيغت بالاستناد المباشر إلى معايير التقييم التي يعتمدها الأساتذة، وهو ما يفتح المجال لاحتمال وقوع تحيز ناتج عن التوجيه، وقد يكون هذا التقارب في الإطار المرجعي قد أسهم في ارتفاع مستويات الاتفاق بين التقييم الآلي والتقييم البشري. ورغم أن هذا الاختيار المنهجي قد ضمن انسجاما واضحا مع سلم التنقيط المعتمد في المقرر، فإنه يحد من قابلية تعميم النتائج، إذ قد تختلف المخرجات إذا ما استخدمت توجيهات بديلة أو سلالم تقييم مغايرة، ولهذا، على البحوث اللاحقة اختبار مدى ثبات تقييمات الذكاء الاصطناعي عبر صيغ توجيه متنوعة ومعايير تنقيط مستقلة.

وبوجه عام، وعلى الرغم من أن تشات جي بي تي 4o يظهر قدرات واعدة مكملة للتقييم البشري، فإنه لا ينبغي أن يحل محل حكم الأستاذ، ولا سيما في ضوء محدودية تفهمه السياقي ومرونته البيداغوجية. وعليه، وبالتوازي مع ما ذهب إليه [9،20]، تؤيد هذه الدراسة اعتماد مقاربة تقييم هجينة تسهم فيها أدوات الذكاء الاصطناعي في دعم الحكم البشري وتعزيزه، لا في الاستعاضة عنه.

2.5. المكاسب والتحديات

كشفت هذه الدراسة عن جملة من المكاسب المرتبطة بتوظيف تشات جي بي تي 4o، إذ تنسجم مع التوجهات العامة في أدبيات تقييم جودة الترجمة. ففي المقام الأول، برز تعزيز استقلالية المتعلم بوصفه فائدة أساسية، كما أشار إلى ذلك الأساتذة المشاركون، إذ بسهم الارتجاع الآلي المنظم والفوري في تعزيز التأمل الذاتي لدى الطلبة ومساعدتهم على تصحيح أخطائهم بأنفسهم، وهو ما يتقاطع مع ما أكده المرجع [7] بشأن دور الأدوات الرقمية في تنمية التعلم الذاتي في تعليم اللغات.

وثانيا، تبينت النجاعة الزمنية بجلاء، فبحسب الأساتذة، تتيح قدرة تشات جي بي تي 4o على تقديم تقييمات مفصلة في زمن وجيز تقليص العبء الواقع على عاتق المدرس، وهو ما يتماشى مع ما أورده المرجع [16] حول إمكانات الذكاء الاصطناعي في

توفير ارتجاع آني ورفع الإنتاجية. وعلى الرغم من أن المرجع [14] لم يتناول الكفاءة الزمنية تناولا مباشرا، فإن ما خلص إليه من تقارب بين التقييمات الآلية والبشرية يوحي بإمكان تبسيط إجراءات التقييم وتسريعها، وهو ما تؤكده نتائج هذه الدراسة في سياق الترجمة القانونية.

وثالثا، شكل الاتساق في تطبيق معايير التقييم إحدى أبرز نقاط القوة، إذ التزم تشات جي بي تي 4o بالإطار التقييمي نفسه عند معالجة جميع العينات، وهذا حد من الأحكام الذاتية. ويتوافق هذا الاستنتاج مع ما توصل إليه المرجع [17] من أن التقييمات المعتمدة على الذكاء الاصطناعي، متى وجهت بأوامر واضحة، تسهم في تحسين الدقة والحد من التحيز، غير أن هذا الاتساق، كما بينت هذه الدراسة وأكده المرجع [20]، يظل منوطا بدقة تصميم الأوامر ووضوحها.

وإلى جانب ذلك، تيسر سهولة الوصول إلى تشات جي بي تي 4o وواجهته البسيطة انتشاره على نطاق أوسع، ولا سيما في البيئات محدودة الموارد، وهو ما يدعم ما ذهب إليه كل من [3،13] بشأن أهمية أدوات الذكاء الاصطناعي السهلة الاستخدام والمنخفضة العوائق في السياقات التعليمية.

وعلى الرغم من هذه المكاسب، ما تزال ثمة تحديات قائمة، إذ يعد العائق التقني المرتبط بتصميم التوجيهات من أبرزها، فقد تؤدي فروق طفيفة في الصياغة إلى اختلافات كبيرة في المخرجات. ويتماشى هذا مع ما أكده [20،25] من أن صياغة الأوامر تعد عاملا حاسما في فاعلية التقييم الآلي، وكما تقترح هذه الدراسة، يمكن للحد من مثل هذه الصعوبات بتطوير قوالب توجيه معيارية، مع توفير تدريب موجه للأساتذة في مجال هندسة الأوامر.

ومن التحديات الأخرى التي أبرزها الأساتذة مسألة الرقمنة، ولا سيما عندما تكون أعمال الطلبة مكتوبة بخط اليد، فمع أن أدوات التعرف البصري على الحروف تقدم عونا في هذا الصدد، فإن ما يكتنفها من قيود زمنية ودقية قد يحد من جدوى الذكاء الاصطناعي، وهو ما ينسجم مع ما خلص إليه [13] من أن الصعوبات التقنية كثيرا ما تعيق إدماج الذكاء الاصطناعي، على الرغم من المواقف المحبذة له. ومن سبل معالجة هذه الإشكالية تشجيع الجامعات والمؤسسات الأكاديمية على اعتماد

امتحانات وواجبات ترجمية مكتوبة رقميا، وهذا يمكن أدوات الذكاء الاصطناعي من معالجة ترجمات الطلبة بدقة أكبر.

كما كشفت هذه الدراسة عن محدودية المرونة التأويلية للذكاء الاصطناعي، إذ قد يفتقر تشات جي بي تي 4o إلى القدرة على شرح مسوغاته شرحا ديناميكيا أو الانخراط في نقاشات دقيقة، وهو ما يعكس المخاوف التي أثارها [3] بشأن قصور الذكاء الاصطناعي في استيعاب الحساسية السياقية والفروق الثقافية. ويضاف إلى ذلك، كما لاحظ [9]، أن الأنظمة الآلية تميل في الغالب إلى تطبيق قواعد التصحيح بصرامة، وهو إشكالية في الترجمة القانونية تتطلب قدرا من المرونة والتأويل السياقي.

وأخيرا، رصدت الدراسة بعض مظاهر عدم الاتساق المؤقت في تنسيق المخرجات، وربما يعود ذلك إلى تباين الجلسات أو إلى عوامل نظامية، وهو ما يردد الدعوة التي أطلقها [16] إلى ضرورة تحقيق أداء أكثر استقرارا وقابلية للتنبؤ أثناء توظيف الذكاء الاصطناعي في البيئات التعليمية الواقعية.

6. الخاتمة

يتمثل هدف هذه الدراسة في استقصاء مدى اتساق تقييم الذكاء الاصطناعي لترجمات الطلبة القانونية مع تقييمات الأساتذة، مع بحث ما يكتنف توظيف الذكاء الاصطناعي في تقييم الترجمة القانونية ضمن سياقات تكوين المترجمين من فوائد وتحديات.

وتشير النتائج المحصلة إلى مستوى مرتفع من التوافق بين تقييمات الأساتذة وتقييمات تشات جي بي تي 4o في رصد أخطاء المحتوى والأخطاء اللغوية الجوهرية في ترجمات الطلبة. كما تظهر النتائج عدم وجود فروق ذات دلالة إحصائية بين الدرجات التي أسندها الأساتذة وتلك التي قدمها تشات جي بي تي 4o، وتدل هذه المعطيات على أن الذكاء الاصطناعي يمكن أن يوظف في تقييم جودة الترجمة داخل بيئات تكوين المترجمين كأداة تقييمية مسانِدة تدعم الحكم البشري ولا تحل محله. وتكمن قيمة هذه الأدوات فيما تتيحه من اتساق في التقييم وسرعة في المعالجة وارتجاع مفصل، وهي عناصر من شأنها الإسهام في تطوير بيداغوجيا الترجمة متى أدمجت إدماجا واعيا

ومدروسا. كما يمكن توظيفها لتعزيز استقلالية المتعلمين لما توفره من ارتجاع تفاعلي ومفصل قادر على الاستجابة للاحتياجات الفردية للطلبة.

ولكن الاستثمار الناجح لهذه الأدوات في بيئات تكوين المترجمين يقتضي معالجة جملة من القضايا الأساسية، وأولها أن مهام الترجمة، كالامتحانات والواجبات المنزلية وغيرها، ينبغي أن تكون مدخلة بصيغة مكتوبة لجعل عملية المعالجة التي تؤديها أنظمة الذكاء الاصطناعي أكثر سلاسة، فإن رقمنة الترجمات المكتوبة بخط اليد باستخدام تكنولوجيات التعرف البصري على الحروف تعد من الإجراءات التي تستهلك كثيرا من الوقت، وقد يشكل هذا عائقا أمام الأساتذة والطلبة أثناء اعتماد هذه الأدوات. وثانيها، أن كلا من الأساتذة والطلبة يحتاجون إلى تكوين منهجي في هندسة الأوامر، قصد تحسين جودة مخرجات الذكاء الاصطناعي وتعزيز فاعليته، ويجب تشجيع مؤسسات تكوين المترجمين على تطوير أوامر معيارية منسجمة مع سلالم التقييم المعتمدة لديها لضمان موثوقية التقييم وتعزيز الاتساق والإنصاف.

وعلى الرغم من النتائج الواعدة التي خلصت إليها الدراسة، فإن توظيف أدوات الذكاء الاصطناعي ينبغي أن يظل خاضعا لإشراف بشري، وأن يكون أداة مساندة، فبإمكان الأساتذة، على سبيل المثال، الاستفادة من الارتجاع الذي يقدمه الذكاء الاصطناعي كونه منطلقا تشخيصيا أوليا، يعقبه تقييم بشري معمق يهدف إلى تعزيز التفاعل البيداغوجي للطلبة وتطورهم.

وفي الختام، يتطلب الإدماج الفعال لأدوات الذكاء الاصطناعي في تقييم جودة الترجمة استثمارا استراتيجيا يوازن بين الأتمتة والبصيرة الإنسانية، فمع حسن التصميم والتدريب الملائم والإشراف الرشيد يمكن للذكاء الاصطناعي أن يضطلع بدور ذي قيمة في دعم بيداغوجيا الترجمة الحديثة.

كما تعد هذه الدراسة خطوة استكشافية تصبو إلى فهم الطريقة التي يمكن من خلالها للذكاء الاصطناعي، وبوجه خاص تشات جي بي تي 4o، أن يتقاطع مع أحكام المقيمين البشر في تقييم جودة الترجمة، فعلى الرغم من أن النتائج تشير إلى إمكانات واعدة على الصعيد البيداغوجي، فإنها تستند إلى عينة محدودة ومقيدة بسياق معين، الأمر الذي يحول دون تعميمها خارج هذا الإطار. ومن ثم، فإن بناء استنتاجات أوسع

يقتضي دراسات لاحقة تعتمد مجموعات بيانات أكبر، وتتحقق من النتائج عبر مجالات متعددة، وتستند إلى تحليلات إحصائية أكثر متانة.

وعلى الرغم من الرؤى التي أفضت إليها هذه الدراسة، فإنه لا بد من الإقرار بجملة النقائص التي طالتها. فأولا، اقتصرت الترجمات المعتمدة في التقييم على عشر مهام ترجمية مأخوذة من امتحان في الترجمة القانونية بإحدى الجامعات السعودية، كما انحصر اتجاه الترجمة في الزوج الإنجليزي العربي. ويستدعي تعميم نتائج هذه الدراسة توظيف عينات أكثر تنوعا، مأخوذة من مقررات ترجمة مختلفة، ومن اتجاهات ترجمية متعددة.

إضافة إلى ذلك، تولت الأستاذة (أ) تقييم ترجمات خمسة طلبة، في حين قيمت الأستاذة (ب) ترجمات مجموعة أخرى من خمسة طلبة مختلفين. ويعكس هذا التصميم التوزيع الطبيعي لعملية التصحيح في المقرر، إذ كانت كل أستاذة مسؤولة عن تقييم أعمال طلبتها وفق السلم التقييمي نفسه. ومع ذلك، نقر بأن تقييم الأستاذتين للمجموعة نفسها من الترجمات العشر كان من شأنه أن يوفر أساسا أقوى لقياس الاتساق بين المقيمين، ولتعزيز المقارنة بين تقييماتهما وتقييم تشات جي بي تي.

كما اقتصرت هذه الدراسة على أداة واحدة من أدوات الذكاء الاصطناعي، وهي تشات جي بي تي 4o، ومن المستحسن في بحوث لاحقة إجراء مقارنات بين أداء عدد من الأدوات المختلفة، مثل جيميناي Gemini من جوجل Google وكوبايلت Copilot من Microsoft، ولا سيما فيما يتعلق بتصميم التوجيهات ورصد الأخطاء ومستوى التفصيل في الارتجاع المقدم. وبالنظر إلى أن توظيف الذكاء الاصطناعي في تقييم جودة الترجمة ما يزال مجالا حديث النشأة، فإن الباحثين مدعوون إلى استكشاف الإمكانات التي تتيحها هذه التقنية الثورية في حقل دراسات الترجمة بوجه عام، وما تحدثه من أثر في منظومة تكوين المترجمين.

إضافة إلى ما سبق، لم تحسب في هذه الدراسة معدلات أخطاء التعرف البصري على الحروف حسابا كميا، ولم تحسب مؤشرات الاتساق بين المقيمين إحصائيا، وذلك لأن المخرجات الأولية صححت في مرحلة التحقق ولم تحفظ بصيغتها الخام. ورغم أن باحثين اثنين راجعا، كل على حدة، جميع النصوص الناتجة

وتصحيحها، ما يقلل من احتمال بقاء أخطاء متبقية، فإنه لا يمكن استبعاد احتمال أن تكون بعض الهفوات الطفيفة غير المرصودة قد أثرت في النتائج.

كما يجدر التنبيه إلى أن مفهوم «الاتفاق» في هذه الدراسة طبق إجرائيا على أساس التوافق في تحديد أنواع الأخطاء فحسب، دون النظر في الفروق المتعلقة بدرجة خطورتها أو بقيم اقتطاع النقاط المترتبة عليها. وعلى الرغم من أن هذا الإجراء يخدم الطابع الاستكشافي للدراسة، الرامي إلى اختبار قدرة تشات جي بي تي 4o على رصد فئات أخطاء مماثلة لتلك التي يحددها المقيمون البشر، فإنه لا يعكس التعقيد الكامل الذي ينطوي عليه تقييم جودة الترجمة، ومن ثم ينبغي التعامل مع مؤشرات الاتساق الواردة في هذه الدراسة بوصفها نتائج أولية، وعلى البحوث اللاحقة اعتماد مقاييس اتفاق متعددة الأبعاد تراعي درجة الخطأ، والتداخل الجزئي بين التصنيفات والفروق في التنقيط، وهذا يوفر فهما أدق لطبيعة التوافق بين التقييم الآلي البشري.

أخيرا، وبما أن التوجيهات قد طورت بالاستناد إلى معايير التقييم التي يعتمدها الأساتذة أنفسهم، فإن الإطار المشترك لربما أسهم جزئيا في ارتفاع مستوى التوافق الملحوظ بين تقييمات الذكاء الاصطناعي وتقييمات البشر. ورغم أن هذا الاختيار المنهجي يعزز الصلة بالسياق الصفي، فإنه قد يكون أدى إلى تضخيم درجة الاتساق، بحكم أن تشات جي بي تي 4o وجه إلى العمل ضمن الإطار التقييمي نفسه الذي يعتمده الأساتذة، وبناء عليه، قد لا تكون النتائج قابلة للتعميم الكامل على مقيمين آخرين أو على سلالم تقييم بديلة. لذا، يستحسن أن تستكشف الدراسات المستقبلية مدى متانة تقييمات الذكاء الاصطناعي عبر صِيغ توجيه متنوعة وأطر تنقيط مستقلة.

وتشمل المقترحات الموجهة للبحوث المستقبلية تطوير حلول آلية لتحويل الترجمات المكتوبة بخط اليد إلى نصوص رقمية، وهذا من شأنه تحسين كفاءة تقييم الذكاء الاصطناعي ودقته. كما نحث الباحثين على اختبار صيغ التوجيه المعتمدة في هذه الدراسة في مجالات ترجمية أخرى، مثل الترجمة الطبية أو الأدبية أو التقنية، وتقييم أدائها باستخدام نماذج لغوية كبيرة مختلفة.

وأخيرا، يمكن للبحوث اللاحقة أن تعيد تطبيق هذه الدراسة على مجموعات بيانات أوسع، وبروتوكولات تقييم أكثر تنظيما، بغية اختبار قابلية تعميم النتائج.

كما أن استقصاء تصورات الطلبة إزاء الارتجاع المولد آليا، وأثره في مخرجات التعلم، من شأنه أن يقدم رؤى قيمة حول القيمة البيداغوجية للذكاء الاصطناعي في تكوين المترجمين. وإلى جانب ذلك، يمكن للدراسات الطولية أن تفحص الكيفية التي يؤثر بها التعرض المستمر للتقييم القائم على الذكاء الاصطناعي في تنمية مهارات الترجمة لدى الطلبة مع مرور الوقت. كما قد تسهم الدراسات المقارنة عبر سياقات تعليمية ولغات مختلفة في التحقق من مدى قابلية تعميم طرائق تقييم جودة الترجمة القائمة على الذكاء الاصطناعي.

ويمكن للبحوث المستقبلية أيضا أن تستكشف سبل إدماج التقييم الآلي ضمن نماذج تقييم هجينة تجمع بين التنقيط المؤتمت والحكم البشري لتحقيق التوازن بين الكفاءة والإنصاف. وإلى جانب ذلك، تستدعي الاعتبارات الأخلاقية، مثل حماية البيانات والتحيز المحتمل في الارتجاع الآلي، وخطر الإفراط في الاعتماد على الأتمتة، دراسة منهجية معمقة، لتوجيه اعتماد مسؤول للذكاء الاصطناعي في برامج تكوين المترجمين.

وأخيرا، من بين المقترحات البحثية ذات الأهمية نجد إدراج تقييمات متداخلة من قبل أكثر من أستاذ للنصوص الطلابية نفسها، لأن هذا يساعد على احتساب درجة الاتفاق بين المقيمين البشر كونهما مرجعية بشرية، ومن شأن هذا الإجراء أن يوفر إطارا تفسيريا ضروريا لفهم طبيعة التوافق بين التقييم الآلي والبشري، وأن يوضح ما إذا كان الأداء المرتفع للذكاء الاصطناعي يعكس مستوى اتفاق حقيقيا، أم أنه لا يتجاوز حدود التباين الطبيعي القائم بين المقيمين البشر أنفسهم.

تعليقات المترجم

[i] النماذج اللغوية الكبيرة (Large Language Models): نماذج حاسوبية متقدمة في مجال الذكاء الاصطناعي، تُعنى بفهم اللغة الطبيعية وتوليدها على نحوٍ يُحاكي الاستعمال البشري، وهي مدربة مسبقا على مدونات نصية ضخمة لاستخلاص العلاقات الإحصائية بين الألفاظ والتراكيب، وغالبا ما تقوم على معماريات حديثة، في مقدمتها المحولات (Transformers) التي تتيح لها استيعاب السياقات اللغوية الواسعة. أنظر:

Jurafsky, D., & Martin, J. H. (2023). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition (3rd ed., draft manuscript).

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.-Y., & Wen, J.-R. (2023). A survey of large language models (arXiv:2303.18223). arXiv.

https://doi.org/10.48550/arXiv.2303.18223

[ii] تكنولوجيات الذكاء الاصطناعي (Artificial Intelligence Technologies): هي الأنظمة والتكنولوجيات القائمة على الخوارزميات الحاسوبية التي تروم محاكاة القدرات الذهنية البشرية أو أتمتة المهام التي تستلزم قدرا من الذكاء الإنساني، وهي تشمل نماذج التعلم الآلي والشبكات العصبية والتعلم العميق وسائر البرمجيات أو التي تمكن الآلة من حل المشكلات واتخاذ القرار وأداء وظائف كالتعرف على الأنماط وفهم اللغة على نحو يقارب الأداء البشري. أنظر:

مجمع اللغة العربية بالقاهرة. (2003). معجم الحاسبات (بالعربية والإنجليزية) (الطبعة الثالثة). القاهرة: مجمع اللغة العربية بالقاهرة.

الكيلاني، تيسير، والكيلاني، مازن. (2001). معجم الكيلاني لمصطلحات الحاسب الإلكتروني: إنجليزي-إنجليزي-عربي موضح بالرسوم (الطبعة الثانية). بيروت: مكتبة لبنان ناشرون.

[iii] تقييم جودة الترجمة (Translation Quality Assessment – TQA): إجراء منهجي يُعنى

بفحص النص المترجم لقياس مدى دقته وقبوله وقابليته للقراءة بالعودة إلى النص الأصلي. ويعد هذا المفهوم من المحاور المركزية في دراسات الترجمة، إذ تستخدم فيه نماذج ومعايير مضبوطة لتقييم سلامة المعنى وصحة اللغة وملاءمة الخطاب ثقافيا. أنظر:

Postan, L. (2023, October 2). Quality assessment of translations. BLEND. https://www.getblend.com/blog/quality-assessment-translations/

[iv] ارتجاع (Feedback): اختير هذا اللفظ سلامته لغويًا واشتقاقه من الجذر (ر ج ع)، الذي يدل على الرد والعودة إلى الأصل، وهو معنى يوافق حقيقة مفهوم feedback كونه إرجاعا للأثر أو المعلومة. وقد فُضل على مصطلح «التغذية الراجعة» لما يمتاز به من إيجاز وصحة المبنى وتجنبه التراكيب المطولة والترجمات الحرفية، مع بقاء دلالته بينة عند المتخصصين وغيرهم. أنظر:

البعلبكي، منير، والبعلبكي، رمزي. (2008). المورد الحديث: قاموس إنكليزي - عربي. بيروت: دار العلم للملايين.

شابسيغ، عمر، والدكاك، أميمة، والعوا، نوار، وورقوزق، هاشم. (2016). معجم مصطلحات الهندسة الكهربائية والإلكترونية والاتصالات (بالعربية والإنجليزية). دمشق: مجمع اللغة العربية بدمشق.

[v] الأتمتة (Automation): إسناد إنجاز المهام إلى التكنولوجيات والبرمجيات أو الآلات، فتؤدَّى تلقائيا مع تقليل التدخل البشري أو الاستغناء عنه، وهي تقوم على برمجة مسبقة أو خوارزميات ذكية تُسيِّرُ العمليات ذاتيا، والهدف من ذلك رفع الكفاءة والدقة والحد من التكلفة والأخطاء البشرية. ويُستعمل المصطلح للدلالة على طيف واسع من التطبيقات، إذ يمتد من تشغيل الآلات الصناعية آليا إلى تنفيذ البرمجيات للمهام المتكررة داخل المؤسسات. وفي الاستعمال الحديث، تعد «الأتمتة» لفظًا عربيًا مُحدَثًا، أُخذ عن «الأوتوماتيكية»، وحل محل تعابير أقدم مثل «التشغيل الآلي».

البعلبكي، منير، والبعلبكي، رمزي. (2008). المورد الحديث: قاموس إنكليزي - عربي. بيروت: دار العلم للملايين.

الهيئة السعودية للبيانات والذكاء الاصطناعي. (2024). معجم البيانات والذكاء الاصطناعي: إنجليزي-عربي (ط. 2).

حمدان، محمود أحمد. (2007). قاموس دار العلم الهندسي الشامل (بالعربية والإنجليزية) (الطبعة الثانية). بيروت: دار العلم للملايين.

IBM. (دون تاريخ). ما هي الأتمتة؟ :

https://www.ibm.com/ae-ar/think/topics/automation

[vi] المحول التوليدي المدرب مسبقا (Generative Pre-trained Transformer - GPT): يُطلق مصطلح GPT على فئة من النماذج اللغوية المتقدمة في الذكاء الاصطناعي، وهو اختصار لعبارة Generative Pre-trained Transformer، أي «المحول التوليدي المدرب مسبقا». ويقوم هذا النموذج على بنية المحولات (Transformers)، وقد خضع لتدريب مسبق على مدوّنات نصية واسعة، وهذا يمكنه من توليد نصوص تحاكي الاستعمال البشري، وأداء طيف متنوع من مهام معالجة اللغة الطبيعية. أنظر:

Narayan, J., & Larsen, B. (2023, January 9). Generative AI: A game-changer that society and industry need to be ready for. World Economic Forum. https://www.weforum.org/stories/2023/01/davos23-generative-ai-a-game-changer-industries-and-society-code-developers/

[vii] التعلم الآلي (Machine Learning) ومعالجة اللغة الطبيعية (Natural Language Processing) والرؤية الحاسوبية (Computer Vision): من أبرز فروع الذكاء الاصطناعي وتطبيقاته المعاصرة، فالتعلم الآلي يعنى بتطوير خوارزميات تمكن الحواسيب من استنباط الأنماط من البيانات وتحسين أدائها تلقائيا دون برمجة صريحة لكل مهمة، لأن هذا يتيح لها التنبؤ واتخاذ القرار اعتمادا على خبرة مكتسبة من بيانات التدريب. وتعد معالجة اللغة الطبيعية حقلا يهدف إلى تمكين الآلات من فهم اللغة البشرية المكتوبة والمنطوقة وتحليلها وتوليدها، ويشمل مهاما كالتصنيف والترجمة واستخلاص المعنى، مع اعتماد واسع على تقنيات التعلم الآلي.

الهيئة السعودية للبيانات والذكاء الاصطناعي. (2024). معجم البيانات والذكاء الاصطناعي: إنجليزي-عربي (ط. 2).

Eisenstein, J. (2019). Introduction to natural language processing. The MIT Press. https://mitpress.mit.edu/9780262042840/introduction-to-natural-language-processing/

Klette, R. (2014). Concise computer vision. Springer.

[viii] الترجمة الآلية العصبية (Neural Machine Translation – NMT): مقاربة حديثة في الترجمة الآلية تعتمد على شبكات عصبية عميقة لترجمة النصوص بين اللغات. وعلى خلاف الأساليب الأقدم، ولا سيما الترجمة الإحصائية القائمة على وحدات جزئية، تعالج هذه المقاربة الجملة على أنها وحدة دلالية متكاملة، وهذا يمكنها من استيعاب السياق على نحو أدق وإنتاج ترجمة أكثر طلاقة وانسجاما. أنظر:

Popel, M., Tomkova, M., Tomek, J., Kaiser, Ł., Uszkoreit, J., Bojar, O., & ... Žabokrtský, Z. (2020). Transforming machine translation: A deep learning system reaches news translation quality comparable to human professionals. Nature Communications, 11, 4381.

https://doi.org/10.1038/s41467-020-18073-9

[ix] التعرف البصري على الحروف (Optical Charcater Rrecognition): تكنولوجيا برمجية لاستخلاص النصوص من الصور أو الوثائق الممسوحة ضوئيا وتحويلها إلى نصوص رقمية قابلة للمعالجة، وتعتمد هذه التكنولوجيا على تحليل المحتوى البصري للتعرف على الحروف والكلمات، سواء كانت مطبوعة أو مكتوبة بخط اليد، ثم إخراجها في صيغة نصية قابلة للبحث والتحرير، وهو ما يجعلها أداة أساسية في رقمنة الوثائق والمخطوطات والأرشيفات. أنظر:

Mithe, R., Indalkar, S., & Divekar, N. (2013). Optical character recognition. International Journal of Recent Technology and Engineering (IJRTE), 2(1), 72-75.

[x] أنسنة Humanization: مصطلح يدل في معناه العام على إضفاء الصفات الإنسانية على شيء ما، أو توجيهه ليتمحور حول الإنسان وقيمه وحاجاته، وهو يستعمل في سياقات متعددة للدلالة على جعل الممارسات أو النظم أكثر توافقا مع البعد الإنساني أخلاقيا واجتماعيا. وفي مجال التكنولوجيا والذكاء الاصطناعي، تحيل الأنسنة إلى تصميم الأنظمة الذكية وتوظيفها بما يراعي الكرامة الإنسانية، ويعزز قابلية التفاعل والفهم والثقة، ويجعل التكنولوجيا أداة في خدمة الإنسان لا بديلا عنه. أنظر:

Fenwick, A., & Molnar, G. (2022). The importance of humanizing AI: Using a

behavioral lens to bridge the gaps between humans and machines. Discover Artificial Intelligence, 2(1), 14. https://doi.org/10.1007/s44163-022-00030-8

[xi] هندسة الأوامر (Prompt Engineering): أداة حاسمة تمكن المترجم أو الخبير اللغوي من استثمار الذكاء الاصطناعي وتكريسه لما يخدم ممارساته الترجمية واللغوية، وتعرف هندسة الأوامر بأنها «عملية تصميم الأوامر المعطاة للنماذج وتنقيحها وتحسينها لتحقيق المخرجات المطلوبة» (الهيئة السعودية للبيانات والذكاء الاصطناعي، 2024، ص. 193) أنظر:
الهيئة السعودية للبيانات والذكاء الاصطناعي. (2024). معجم البيانات والذكاء الاصطناعي: إنجليزي -عربي (ط. 2).

إخلاء مسؤولية / ملاحظة الناشر:

إن التصريحات والآراء والبيانات الواردة في جميع المنشورات تعبر حصرا عن آراء المؤلف (المؤلفين) والمساهم (المساهمين) المعنيين، ولا تمثل بالضرورة آراء دار النشر MDPI و/أو هيئة التحرير. كما تُخلي دار النشر MDPI و/أو هيئة التحرير مسؤوليتها عن أي أضرار تلحق بالأشخاص أو الممتلكات نتيجة لأي أفكار أو مناهج أو تعليمات أو منتجات يشار إليها في محتوى هذه المنشورات.

مساهمات المؤلفتين

التصور المفاهيمي: هـ.ع.، ف.أ.غ؛ المنهجية: هـ.ع.، ف.أ.غ؛ التحقق من الصحة: هـ.ع؛ التحليل الشكلي: ف.أ.غ؛ الموارد: هـ.ع.، ف.أ.غ؛ تنظيم البيانات: ف.أ.غ؛ الإشراف العلمي: هـ.ع؛ الحصول على التمويل: هـ.ع.، ف.أ.غ. وقد اطلعت المؤلفتين على النسخة المنشورة من المخطوط ووافقتا عليها.

التمويل

حصل هذا البحث على منحة رقم (392/2024) من المرصد العربي للترجمة، وهو هيئة تابعة للمنظمة العربية للتربية والثقافة والعلوم (ALECSO)، بدعم من هيئة الأدب والنشر والترجمة في المملكة العربية السعودية.

بيان المو افقة الأخلاقية المؤسساتية

حصلت الدراسة على الموافقة الأخلاقية من اللجنة الدائمة لأخلاقيات البحث العلمي بجامعة الملك سعود، الرياض، المملكة العربية السعودية، تحت رقم (24-810).

بيان إتاحة البيانات

البيانات المعروضة في هذه الدراسة متاحة عند الطلب من المؤلفتين المراسلتين.

تعارض المصالح

تصرح المؤلفتان بعدم وجود أي تعارض في المصالح.

الاختصارات

الاختصار	المصطلح	المقابل العربي
AI	Artificial Intelligence	الذكاء الاصطناعي
AUTOMQM	Automated Multi-dimensional Quality Metrics	المقاييس الآلية متعددة الأبعاد للجودة
BISON	Baseline Iterated and Stored on Node	الخط الأساس المكرر والمخزن على العُقد
BLEU	Bilingual Evaluation Understudy	مقياس التقييم ثنائي اللغة المرجعي
BLEURT	Bilingual Evaluation Understudy with Real-Time	مقياس التقييم ثنائي اللغة المرجعي في الزمن الحقيقي
ESP	English for Specific Purposes	الإنجليزية لأغراض خاصة
GEC	Grammatical Error Correction	تصحيح الأخطاء النحوية
GECToR	Grammatical Error Correction with Real-time feedback	تصحيح الأخطاء النحوية بارتجاع آني
GEMBA	GPT Estimation Metric Based Assessment	التقييم القائم على مقياس التقدير المعتمد على نماذج المحول التوليدي المدرب مسبقا
GPT	Generative Pre-trained Transformer	المحول التوليدي المدرب مسبقا
ICTs	Information and Communication Technologies	تكنولوجيات المعلومات والاتصال
KSU	King Saud University	جامعة الملك سعود
LLMs	Large Language Models	النماذج اللغوية الكبيرة
MQM	Multidimensional Quality Metrics	مقاييس الجودة متعددة الأبعاد
MTI	Master of Translation and Interpreting	ماستر الترجمة والترجمة الشفوية

NMT	Neural Machine Translation	الترجمة الآلية العصبية
OCR	Optical Character Recognition	التعرف البصري على الحروف
PaLM	Particular Language Model	نموذج لغوي معين
SEEDA	Southeastern English Dialects Archive	أرشيف اللهجات الإنجليزية الجنوبية الشرقية
ST	Source Text	النص المنقول منه
TQA	Translation Quality Assessment	تقييم جودة الترجمة
TT	Target Text	النص المنقول إليه
WMT	Workshop on Machine Translation	ورشة الترجمة الآلية

قائمة المراجع

1. Alkhawaja, L. Unveiling the new frontier: ChatGPT-3 powered translation for Arabic-English language pairs. *Theory Pract. Lang. Stud.* 2024, *14*, 347–357. [Google Scholar] [CrossRef]
2. Putera, L.J.; Sujana, I.M. A comparative analysis of accuracy between Google Translate and Bard in translating abstracts of scientific journals. *Didakt. J. Ilm. PGSD STKIP Subang* 2024, *10*, 35-44. [Google Scholar] [CrossRef]
3. Shahmerdanova, R. Artificial intelligence in translation: Challenges and opportunities. *Acta Glob. Humanit. Linguarum* 2025, *2*, 62-70. [Google Scholar] [CrossRef]
4. Alhalalmeh, A.H.; Al-Tarawneh, A.; Al-Badawi, M. Navigating the complexities of translation studies in the legal field. In *From Machine Learning to Artificial Intelligence*; Musleh Al-Sartawi, A.M.A., Al-Okaily, M., Al-Qudah, A.A., Shihadeh, F., Eds.; Springer: Cham, Switzerland, 2025; Volume 572, pp. 1-12. [Google Scholar] [CrossRef]
5. Kobayashi, M.; Mita, M.; Komachi, M. Large language models are state-of-the-art evaluator for grammatical error correction. *arXiv* 2024. [Google Scholar] [CrossRef]
6. Wang, J. Exploring the potential of ChatGPT-4o in translation quality assessment. *J. Theory Pract. Humanit. Soc.Sci.* 2024, *1*,18-30. Available online: https://woodyinternational.com/index.php/jtphss/article/view/27 (accessed on 25 June 2025).
7. Selama, S.A. ESP Online learning in the post-COVID19 pandemic era: A step towards autonomy, intercultural competence, and critical thinking. *J. Lang. Transl.* 2025, *5*, 46–55. [Google Scholar]
8. Shamsan, Y. Translation of cultural terms in legal texts: An evaluation of the quality of English translation in the book by Hatem et al. (The Legal Translator in the Field). *Ibn Khaldoun J. Stud. Res.* 2025, *5*, 593–627. [Google Scholar] [CrossRef]
9. Wu, H.; Wang, W.; Wan, Y.; Jiao, W.; Lyu, M. ChatGPT or Grammarly? Evaluating ChatGPT on grammatical error correction benchmark. *arXiv* 2023, arXiv:

2303.13648. [Google Scholar]

10. Hang, C.N.; Yu, P.-D.; Tan, C.W. TrumorGPT: Graph-Based Retrieval-Augmented Large Language Model for Fact-Checking. *arXiv* 2025, arXiv:2505.07891. [Google Scholar] [CrossRef]
11. Zeng, J.; Dai, Z.; Liu, H.; Varshney, S.; Liu, Z.; Luo, C.; Li, Z.; He, Q.; Tang, X. Examples as the Prompt: A Scalable Approach for Efficient LLM Adaptation in E-Commerce. *arXiv* 2025, arXiv:2503.13518. [Google Scholar]
12. Qingliang, R. Development of translator studies in the era of artificial intelligence: Opportunities, challenges and the road ahead. *Pak. J. Life Soc. Sci.* 2024, *22*, 24108–24114. [Google Scholar] [CrossRef]
13. Alhaj, A.A.M. Exploring the challenges and obstacles encountered in utilizing artificial intelligence tools (AITs) in translation teaching from the perspectives of faculty members at Saudi universities. *Migr. Lett.* 2023, *21*, 398-412. [Google Scholar]
14. Fernandes, P.; Deutsch, D.; Finkelstein, M.; Riley, P.; Martins, A.F.T.; Neubig, G.; Garg, A.; Clark, J.H.; Freitag, M.; Firat, O. The devil is in the errors: Leveraging large language models for fine-grained machine translation evaluation. *arXiv* 2023, arXiv:2308.07286. [Google Scholar] [CrossRef]
15. Kocmi, T.; Federmann, C. Large language models are state-of-the-art evaluators of translation quality. *arXiv* 2023, arXiv:2302.14520. [Google Scholar]
16. Mohammed, S.Y.; Aljanabi, M. Advancing translation quality assessment: Integrating AI models for real-time feedback. *EDRAAK* 2024, *2024*, 1-7. [Google Scholar] [CrossRef] [PubMed]
17. Koka, N.A.; Akan, M.F.; Kana'n, B.H.I.; Khan, M.R.; Zulfiquar, F.; Jan, N. Impact of artificial intelligence (AI) on translation quality: Assessment and evaluation. *J. Southwest Jiaotong Univ.* 2023, *58*, 907-918. [Google Scholar] [CrossRef]
18. OpenAI. Introducing ChatGPT. 30 November 2022. Available online: https://openai.com/blog/chatgpt (accessed on 19 April 2025).
19. Araújo, S.; Aguiar, M. Comparing ChatGPT's and human evaluation of scientific texts' translations from English to Portuguese using popular automated translators.

In Proceedings of the CLEF 2023: Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 18–21 September 2023; pp. 2908-2917. [Google Scholar]

20. Lu, Q.; Qiu, B.; Ding, L.; Xie, L.; Tao, D. Error analysis prompting enables human-like translation evaluation in large language models: A case study on ChatGPT. *Preprints*, 2023; *preprint*. [Google Scholar] [CrossRef]
21. Yin, R.K. *Case Study Research and Applications: Design and Methods*, 6th ed.; Sage Publications: Thousand Oaks, CA, USA, 2018; Available online: https://uk.sagepub.com/en-gb/eur/case-study-research-and-applications/book250150 (accessed on 2 April 2024).
22. King Saud University, College of Languages and Translation. Study Plan for the Translation Program الخطة الدراسية لبرنامج الترجمة. Available online: https://cols.ksu.edu.sa/ar/node/2939 (accessed on 10 June 2023).
23. OpenAI. GPT-4o: The First Omni Model. 13 May 2024. Available online: https://openai.com/index/hello-gpt-4o/ (accessed on 19 December 2023).
24. Rooney, K.; OpenAI Tops 400 Million Users Despite Deepseek's Emergence. CNBC. 20 February 2025. Available online: https://www.cnbc.com/2025/02/20/openai-tops-400-million-users-despite-deepseeks-emergence.html (accessed on 15 March 2025).
25. Lau, Y.L.; Chee, S.P.; Chua, R.H.H.; Yong, Z.H.; Yong, I.K.; Tan, J.C.; Yong, H.W.; Abu Bakar, A.L.A. The datasets of human and AI translation. *J. Open Humanit. Data* 2024, *10*, 45. [Google Scholar] [CrossRef]
26. Sadiq, S. Evaluating English-Arabic translation: Human translators vs. Google Translate and ChatGPT. *J. Lang. Transl.* 2025, *12*, 67–95. [Google Scholar] [CrossRef]
27. OpenAI. ChatGPT (Version 4) [Large Language Model]. 29 August 2024. Available online: https://chat.openai.com/ (accessed on 22 March 2024).
28. Field, A. *Discovering Statistics Using IBM SPSS Statistics*, 5th ed.; SAGE Publications: Thousand Oaks, CA, USA, 2018; Available online: https://uk.sagepub.com/en-gb/eur/discovering-statistics-using-ibm-spss-statistics/book285130 (accessed on 22 March 2024).

29. Zhang, M. A study on the translation quality of ChatGPT. *Int. J. Educ. Curric. Manag. Res.* 2024, *5,* 153-163. [Google Scholar] [CrossRef]

التعريف بالمؤلفتين

- **فاطمة الغامدي** (Fatimah Alghamdi): باحثة متخصصة في ميدان دراسات الترجمة، وهي حاصلة على درجة الماجستير من كلية علوم اللغات بجامعة الملك سعود. وتنصرف اهتماماتها البحثية إلى تكنولوجيا الترجمة والذكاء الاصطناعي، ولا سيما فيما يتصل بتقاطع هذين المجالين وأثره في تطوير العملية الترجمية والارتقاء بجودتها.
faeahmed@hotmail.com

- **هند العتيبي** (Hind Alotaibi): بروفيسورة بكلية علوم اللغات بجامعة الملك سعود. نالت درجة الدكتوراه في التربية من جامعة مانشستر في تخصص دقيق في تعلم اللغة بمساعدة الحاسوب، وقد تولت في السنوات الأخيرة عددا من المناصب الأكاديمية والإدارية البارزة، من بينها منصب نائب المدير العام للمركز الوطني للتعلم الإلكتروني، وعمادة كلية العلوم الإنسانية بجامعة الأمير سلطان، ومنصب نائبة عميد كلية اللغات والترجمة بجامعة الملك سعود. وتشارك الأستاذة الدكتورة حاليا في عدد من المشاريع البحثية، من أبرزها تطوير تطبيقات تعليمية موجهة لطلبة اللغات والترجمة. وتشمل اهتماماتها البحثية تكنولوجيات المعلومات والاتصال في التعليم والتعلم الإلكتروني والتعلم بمساعدة الأجهزة المحمولة والترجمة بمساعدة الحاسوب.
hialotaibi@ksu.edu.sa
https://orcid.org/0000-0003-4215-086X