



أنظمة الذكاء الاصطناعي واللغة العربية

محمد زكي خضر

رئيس المعهد العالمي لحوسبة القرآن والعلوم الإسلامية

khedher@gmail.com

الملخص: أطلق اصطلاح الذكاء الاصطناعي على البرمجيات الحاسوبية التي تحتوي على عمليات ذكية تحاكي ذكاء المخلوقات الحية من نبات أو حيوان أو إنسان، لكن اصطلاح الذكاء الاصطناعي مؤخرًا يشير غالبًا إلى التعاملات الحاسوبية التي تقوم بالتعلم الآلي ثم تستعمل بعد تعلمها لحل المسائل المصممة لها. لذلك فإن التعلم الآلي أصبح هو محور عمليات الذكاء الاصطناعي حاليًا. يحتاج التعلم الآلي إلى بيانات ضخمة لكي يتعلم منها فيكون خوارزمياته (التي غالبًا ما تستعمل الشبكات العصبية) لكي يستعملها فيما بعد على بيانات جديدة لحل مسائله التي صمم من أجلها. تستعمل أنظمة الذكاء الاصطناعي الحالية المعلومات المتوفرة على شبكة الإنترنت والكتب والصحف والوثائق المتوفرة لكي تستخلص منها خوارزميات استقبال النصوص اللغوية وتحليلها وفهم معناها ثم القيام بالعمليات المطلوبة منها حاسوبياً وتحويل نتائج تلك العمليات إلى نصوص مكتوبة أو منطوقة أو عمليات صناعية تستخدم في الإنسان الآلي (الروبوت) أو الآلات الأخرى أو غير ذلك من تطبيقات. سنقوم في هذا البحث بتقديم بعض ما يتوفر من برمجيات الذكاء الاصطناعي واستخدامها وفحص قابليتها للتعامل مع اللغة العربية لتبيان الشوط الذي قطعه وما تشكو من نقائص مع بعض التركيز على الترجمة الآلية، ومن ثم تقديم بعض المؤشرات والمقترحات بشأن الخطوات المستقبلية اللازمة لتحسين تعامل تلك البرمجيات مع اللغة العربية.

الكلمات الجوهرية: أنظمة الذكاء الاصطناعي، معالجة اللغة العربية طبيعياً، تقنيات التعليم. التعلم الآلي واللغة العربية.

1. تعاريف أولية:

الذكاء الاصطناعي (Artificial intelligence): هو حقل جديد من التطبيقات الحاسوبية والذي فيه تحاكي البرامج الحاسوبية القدرات الذهنية البشرية وطرق عملها بحيث تستطيع التعلم والإستنتاج ورد الفعل لحالات لم تبرمج أصلا عليها. وحيث أن تعريف الذكاء البشري أصلاً محل مناقشات دون الوصول إلى تعريفه بدقة فكذا الذكاء الإصطناعي ، حيث أن مستقبل المجالات التي سيدخلها والطرائق التي يستعملها مفتوحة على كثير من التطورات المستقبلية.

معالجة اللغة الطبيعية (Natural Language Processing – NLP): معالجة اللغات الطبيعية هي فرع من فروع الذكاء الإصطناعي تسمح للأنظمة بفهم وتفسير وإنتاج اللغة البشرية بشكل طبيعي. تستخدم في تطبيقات مثل الترجمة بين اللغات وتوليد النصوص والمحادثات مع الروبوتات وغيرها .

النظام الخبير (Expert System): هو برنامج خاص يحاول محاكاة الخبرة البشرية، فهو يستقبل خبرات البشر ويحاول محاكاتها ومن ثم يمكنه أن يعطي أحكاماً أو اتخاذ قرارات. فالخبراء من البشر في أي حقل يملكون كمية هائلة من المعرفة المتخصصة في مجالات عملهم، لذا فإن النظم الخبيرة تستند عادة إلى قواعد معرفة تتضمن عدداً هائلاً من قواعد البيانات التي تحوي معلومات المعرفة. فعالم اللغة هو خبير في علم اللغة والطبيب خبير في تشخيص الأمراض وهكذا . ويمكن لكل منهم أن ينقل خبرته إلى النظام الخبير بل يمكن للنظام الخبير أن يستقبل خبرات عدد من الخبراء في الحقل الواحد وعندها يمكن أن يكون أكفأ من أي واحد منهم.

التعلم العميق (Deep Learning): يعتمد التعلم العميق على شبكات عصبية اصطناعية عميقة بحيث يتم تدريبها على مجموعات ضخمة من البيانات. ويستخدم في مجالات مثل التعرف على الصوت والصورة وتوليد النصوص.

الشبكة العصبية الإصطناعية (Artificial Neural Network): هي إحدى طرائق الذكاء الاصطناعي التي يمكن بواسطتها تعلّم أجهزة الحاسوب معالجة البيانات بطريقة مشابهة لما يحدث في الدماغ البشري. وهي نوع من عمليات التعلم العميق. تستخدم الشبكات العصبية عُقدًا أو عصبونات مترابطة فيما بينها في بنية مكونة من طبقات. وكلما كانت المسائل المراد حلها أعقد كانت هناك حاجة إلى عدد كبير من الطبقات والعقد.

2. التعلم الآلي (Machine Learning)

التعلم الآلي هو فرع من الذكاء الاصطناعي يهتم بتطوير أنظمة قادرة على تعلم وتحسين أداء المهام بناءً على البيانات المتاحة، بدلاً من البرمجة اليدوية المباشرة. يعتمد التعلم الآلي على تطوير النماذج الرياضية والخوارزميات التي تمكن الأنظمة من استيعاب البيانات وتحليلها واستخلاص المعرفة منها، وهو يتيح للأنظمة أن تتعلم وتتكيف من البيانات المتاحة دون أن تكون مبرمجة بشكل صريح. تستخدم التقنيات والأساليب في التعلم الآلي لتحقيق مجموعة متنوعة من الأهداف، مثل التصنيف والتنبؤ والتجميع والتحليل والترجمة والاستدلال وغيرها. يتم ذلك من خلال تحليل البيانات المتاحة واكتشاف الأنماط والعلاقات والتراكيب المختلفة الموجودة فيها.

في سياق الذكاء الاصطناعي، يتم استخدام التعلم الآلي كأساس لتطوير نماذج وأنظمة ذكاء اصطناعي قادرة على التعلم والتكيف واتخاذ القرارات. يتم تدريب هذه النماذج على البيانات المتاحة وتعديل معلوماتها لتحسين أدائها مع مرور الوقت. ونتيجة لذلك يمكن للأنظمة أن تتعلم من البيانات المتاحة وتحليلها وفهمها واستخلاص المعرفة منها، مما يمكنها من تنفيذ المهام المعقدة وتحقيق أداء متفوق في مجالات مثل التعرف على الصوت، التعرف على الصور، الترجمة الآلية، التحليل الضخم للبيانات، عمليات التصنيف، التنبؤ، التجميع، التحليل، الاستدلال وغيرها من التطبيقات الذكية.. ويجري ذلك من خلال تحليل ما يتوفر من بيانات واكتشاف الأنماط والعلاقات والتراكيب المختلفة الموجودة فيها.

محاكاة تعلم الأطفال: لنأخذ مثالاً عن كيفية تعلم الطفل اللغة العربية في البداية حينما كانت البادية موطناً ثرياً للغة العربية الفصحى. يتعلم الطفل أصواتاً من والديه أو محيطه ثم يتعلم أن تلك الأصوات التي تتكون من حرف واحد أو مقطع صوتي واحد أو حروف متعددة تشير إلى شيء ما، فيتعلم تسمية الأشياء بكلمات ثم يتعلم تسمية المعاني بكلمات محددة وذلك من خلال الأمثلة فيتدرج إلى أن يتعلم تركيب الجمل، ثم الفرق بين المفرد والمثنى والجمع من الأمثلة التي تمر عليه فيتعلم جمع المذكر السالم مثلاً: جمع نائم - نائمون.... وجمع راكب - راكبون. ويتعلم أن جمع سماء - سماوات... ويبدأ بالتفريق بين المذكر والمؤنث فيعرف أن جالسة هي كلمة مؤنثة وكلمة جالس تشير إلى مذكر وجمع كل منهما جمع مؤنث سالم (لكلمة جالسة هي جالسات) أو جمع مذكر سالم (لكلمة جالس هي جالسون) ويبدأ بتطبيق ذلك على كلمات لم يسمع جمعها فيما سبق. وفجأة يجد أن جمع ناقة هو نوق وليس ناقات، وجمع رجل هو رجال وليس رجلون. فيستنتج أن هناك قواعد أخرى (هي جمع التكسير دون أن يعلم تسميتها) ويبدأ بتعلم بعض قواعد جمع التكسير مثلاً جمع قائم هو قيام ولكن يمكن أن يكون قائمون ثم يتعلم التمييز في الإستعمال بين قائمون وقيام وأين يستعمل هذه وأين يستعمل تلك. وكذلك يتعلم من السياق أين يستعمل راكبون وأين يستعمل راكبين دون أن يعرف الموقع من الإعراب إن كان مضافاً إليه

مجرورا أو مفعولا به منصوبا ولكن يتعلم ذلك من كثرة الترداد فيبدأ بالتمييز أين يستعمل الحروف (ون) وأين يستعمل (ين) وهكذا.

تتعلم الآلة بالطريقة نفسها وذلك لأن الأسس النظرية للشبكات العصبية تحاول أن تحاكي طريقة تعلم الإنسان... فتستنتج القاعدة من كثرة الأمثلة التي تمر عليها وقد تخطئ عند الشذوذ عن القاعدة ويمكن لأنظمة الذكاء الاصطناعي الحديثة مثل "جي بي تي" أن تقبل تصحيح أخطائها إذا ما أعطيت الجواب الصحيح الذي يمكن أن تستعمله بالمرّة القادمة لتلك الحالة. أما إذا أريد إضافتها كقاعدة جديدة فهي بحاجة إلى أمثلة عديدة لكي تستنج منها القاعدة الجديدة ولا تكتفي بمثال واحد... وهكذا فإنّ وفرة المعلومات التي تتعلم منها يزيد من الدقة في الأداء تدريجياً. كذلك فإنّ تشكيل الكلمة (بالضمّة والفتحة والكسرة والتتوين والشدة) يتعلمها الطفل العربي تدريجياً بحيث لا يلاقي صعوبة في الكلمات التي على وزن واحد. أما الكلمات التي تتعدد تشكيلاتها فيحتاج إلى السياق. وهي مسألة أخرى أصعب مما سبق. فيحتاج إلى أمثلة أكثر لكل سياق من السياقات. وكذلك الآلة فإنها تحتاج إلى العديد من الأمثلة لتمييز السياق سواء فيما يتعلق بالتشكيل أو المعنى. وبالطبع فإن الآلة لا تستغني عن القواعد اللغوية والمفاهيم العامة للمساعدة في حل هذه المسائل من ناحية التشكيل أو المعنى. يبين الشكل (1) في الملحق (1) مقارنة بين مدخلات ومخرجات البرامج الحاسوبية العادية وبرامج التعلم الآلي.

3. أنظمة الذكاء الاصطناعي

نستعرض في هذا القسم بعض أنظمة الذكاء الاصطناعي ومنها نظم الترجمة الآلية: ترجمة جوجل وترجمة سيستران وترجمة بينج. تستند أنظمة الترجمة الآلية الحديثة إلى استعمال الشبكات العصبية وهي تحتاج بيانات ضخمة (ثنائية اللغة أو متعددة اللغات) تقوم بالتدرب عليها ابتغاء الوصول إلى خوارزميات داخلية تستعمل لترجمة النصوص المراد ترجمتها.

انتشرت في العقود الثلاثة الأخيرة أنظمة عديدة للترجمة من العربية وإليها. ولغرض تقييم بعض هذه الأنظمة لمعرفة نقاط الضعف فيها جرت محاولات كثيرة منها مقارنة الأنظمة الثلاثة المشار إليها وهي ترجمة جوجل وسيستران وبينج على نصوص علمية وإخبارية وأدبية لغرض معرفة ما تشير إليه من نقاط الضعف في بعض تلك الأنظمة (1). إن أحد الأسباب الفنية في ضعف الترجمة الآلية في الوقت الحاضر هو قلة النصوص (ثنائية اللغة) المتوفرة للتدريب لكل من اللغة المترجم منها واللغة المترجم إليها. فتدريب المترجم الآلي على بيانات ضخمة يزيد من دقة الترجمة، كما أن من الضروري في ترجمة النصوص المتخصصة في حقل معين توفر نصوص للتدريب مناسبة لذلك الحقل، فوجود نصوص في

البيئة والزراعة والغابات والمناطق السياحية مثلا يجعل ترجمة كلمات ذات علاقة بتلك الحقول أكثر دقة (1).

قمنا بتجربة لمعرفة التقدم في أشهر ترجمة متوفرة الآن وهي ترجمة جوجل بمحاولة الحصول على ترجمة لجملة معينة عام 2016 وأخرى مؤخرًا خلال عام 2023. يبين الجدول (1) المبين في الملحق (2) الجملة الأصلية وترجمتها خلال العامين المذكورين مع الترجمة الدقيقة لها. ويلاحظ بوضوح الزيادة في دقة الترجمة مؤخرًا.

كما أجرينا محاولات على ترجمة جوجل أيضا على جمل بسيطة وأخرى أكثر تعقيدًا لمعرفة مكامن الخطأ ومدى التقدم مع الزمن. يلاحظ من الترجمات في الجدول (2) في الملحق (3) أن الجملة الأولى وهي جملة بسيطة كانت الترجمة دقيقة، وحينما تقدمت فيها كلمة الطالب لتصبح جملة إسمية كانت ترجمتها خاطئة قبل عامين لكنها أصبحت سليمة الآن كما يتبين في الجملة الثالثة. وبمقارنة الجملتين في الحقلين الرابع والخامس نجد أنها كانت خاطئة فأصبحت سليمة. أما الجملتان الخامسة والسادسة فالفرق بينهما التأكيد على اللون الأحمر بإضافة ألف لام التعريف فنجد أن الترجمة قد أخذت ذلك بعين الاعتبار لحد ما. أما الجملة السابعة التي هي جملة طويلة، نجد أن الترجمة ركيكة لحد ما وربما تحتاج إلى إضافة علامات لكي تستقيم. وهكذا يمكن من هذين الجدولين أن نستنتج أن الترجمة الآلية في تحسن لا بأس به لكنها تحتاج المزيد.

وقد أجريت فحوصات أخرى على ترجمة جوجل فوجد أنها لم تميز الفرق بين (رغب في) و (رغب عن) واعتبرت كليهما (رغبة في)، كما وجد أن المطابقة بين العدد والمعدود وفيما يتعلق بالجنس كانت قبل سنوات فيها أخطاء كثيرة... الترجمة اليوم أكثر دقة في هذا الصدد. هذا بالإضافة إلى أن المتلازمات اللفظية كانت كثيرة الأخطاء لكنها في تحسن، مثلا: "وضعت الحرب أوزارها" - ترجمة جوجل كانت دقيقة، وترجمة "بادئ الأمر" و"أول الأمر" و"بادئ ذي بدء"، كلها ترجمت بمعنى واحد.

لم تستطع الترجمة الآلية التمييز بين الكلمات التي تغير معناها عبر التاريخ فمثلا لم تميز الترجمة لكلمة "سيارة" بين المعنى الوارد في القرآن الكريم في سورة يوسف: "وجاءت سيارة فأوردوا واردهم" وبين المعنى الحديث بل ترجمت حسب المعنى الحديث ولم تتعرف على المعنى القديم.

نظام "جي بي تي" GPT (2):

يعتبر "جي بي تي" مصدراً قوياً للمعلومات والإجابات التلقائية للأسئلة المختلفة، ويمكن استخدامه للعديد من الأغراض المختلفة، مثل الدعم الفني والتسويق والتعليم. وبعد التدريب على العديد من المجموعات الكبيرة من النصوص، يمكن لـ "جي بي تي" العمل بشكل فعال ودقيق في الإجابة على الأسئلة المختلفة التي يمكن أن تطرحها المستخدمون من خلال: "شات جي بي تي CHAT GPT". يعتبر "جي بي تي" أداة قوية للعديد من الأغراض التي يمكن استخدامها لتحسين العملية العامة للعديد من الأعمال. بالإضافة إلى ذلك، يعتبر "جي بي تي" أداة قوية للتعليم والتدريب، ويمكن استخدامه لتوفير الوقت والجهد باللغة العربية.

إن البيانات المتوفرة لهذا النظام باللغة الإنكليزية أكثر بكثير من اللغة العربية. لذلك فإنه في بعض الأحيان إذا ما كان السؤال باللغة العربية فإنه يقوم بترجمته إلى اللغة الإنكليزية والبحث في المعلومات المتوفرة لديه باللغة الإنكليزية ثم يعيد ترجمة النتيجة إلى العربية. وبالطبع فإن ذلك قد يتسبب في أخطاء ملموسة إما بسبب نقص المعلومات المتوفرة في الحقل المسؤول عنه باللغة الإنكليزية وإما بسبب عمليتي الترجمة.

نظام جوجل للذكاء الاصطناعي (3):

يقوم مقدمو الخدمات السحابية ببناء أعداد كبيرة من وحدات معالجة الرسومات لتوفير المزيد من القوة الحاسوبية للذكاء الاصطناعي. أعلنت شركة جوجل مؤخراً عن حاسوب عملاق يعمل بالذكاء الاصطناعي يحوي 26000 وحدة معالجة للرسومات (بالمقارنة مع أسرع حاسوب عملاق متوفر الآن - فرونتر - الذي يحوي 37000 وحدة معالجة للرسومات). يعد هذا الحاسوب العملاق محاولة من الشركة من أجل التفوق في حقل الذكاء الاصطناعي بالتنافس مع شركة مايكروسوفت.

نظام شركة ميتا (4):

أطلقت شركة ميتا Meta نموذجها الخاص بها الذي أسمته LLaMA وهو نموذج لغة كبير أساسي حديث مصمم لمساعدة الباحثين على تطوير عملهم في الذكاء الاصطناعي. تعمل النماذج الأصغر على تمكين الباحثين الذين ليس لديهم إمكانية الوصول إلى كميات كبيرة من البنية التحتية لدراسة هذه النماذج، مما يؤدي إلى إضفاء توسيع للبحث العلمي والتطبيقات للوصول إلى مزيد من التقدم في هذا المجال المهم سريع التغيير. ولذلك يعد تدريب نماذج الأساس الأصغر من لاما أمراً مرغوباً أكثر لعدم حاجته إلى مساحة نموذج اللغة الكبيرة. ومع كل التطورات الحديثة في نماذج اللغات الكبيرة، لا يزال الوصول الكامل إلى البحث محدوداً بسبب الموارد المطلوبة لتدريب وتشغيل مثل هذه النماذج الكبيرة.

أدى هذا الوصول المقيد إلى الحد من قدرة الباحثين على فهم كيف ولماذا تعمل هذه النماذج اللغوية الكبيرة، مما أعاق التقدم في الجهود المبذولة لتحسين متانتها وتخفيف المشكلات المعروفة، مثل التحيز وإمكانية توليد معلومات مضللة. وهناك شركات أخرى تقوم بمحاولاتها في هذا الحقل مثل أمازون وغيرها.

4. تعامل أنظمة الذكاء الاصطناعي مع اللغة العربية:

تتكون الذخيرة اللغوية التي يعمل عليها الذكاء الاصطناعي من المعلومات المستقاة من الإنترنت من نصوص ومقالات وكتب ومنتديات وكذلك المكتبات الخاصة ويجري التدريب عليها بواسطة التعلم العميق والشبكات العصبية لفهم وتوليد اللغة الطبيعية. وفي أثناء عملية التدريب يتعلم النموذج تركيبات اللغة والقواعد النحوية والعلاقات اللفظية والمعاني اللغوية من النصوص المقدمة له. بعد ذلك، يمكن استخدام النموذج للرد على الأسئلة وتوليد النصوص وتقديم المعلومات في مجموعة متنوعة من المواضيع وتتم معالجة ما يتقدم من سؤال أو استفساراً من خلال إدخال النص حيث تجري معالجته بواسطة الشبكة العصبية المدربة. ويتم تحليل السياق والتعبيرات والأنماط في السؤال، ثم يتولى النموذج إعادة صياغة السؤال وفهمه وإنتاج إجابة مناسبة.

تعتمد جودة ودقة الإجابات التي يقدمها النظام على عدة عوامل، بما في ذلك فهم السؤال واستيعاب السياق، ونوعية البيانات التي تم تدريب النموذج عليها. وقد يكون هناك بعض الحالات التي يمكن أن يقدم فيها النموذج إجابات غير دقيقة أو غير ملائمة للسؤال المطروح.

يتطلب تدريب نماذج الذكاء الاصطناعي موارد كبيرة وقد يستغرق وقتاً طويلاً لتحقيق الأداء العالي. وقد يبلغ حجم البيانات التي يحتاجها للتدريب مئات الملايين بل مليارات من الكلمات.

وتعتبر كمية وتنوع البيانات المستخدمة في التدريب عاملاً مهماً في تحقيق أداء أفضل لنماذج الذكاء الاصطناعي. ويفضل استخدام مجموعات بيانات واسعة ومتنوعة تشمل نماذج مختلفة من اللغة العربية لضمان تغطية شاملة لمختلف جوانب اللغة والمفاهيم المتعلقة بها. وتتوزع مصادر البيانات على ما يتوفر على الإنترنت والكتب والصحف والوثائق المتوفرة وتستخلص منها خوارزميات استقبال النصوص اللغوية وتحليلها وفهم معناها ثم القيام بالعمليات المطلوبة منها حاسوبياً. تمتد نتائج تلك العمليات إلى نصوص مكتوبة أو منطوقة أو عمليات صناعية تستخدم في الإنسان الآلي (الروبوت) أو الآلات الأخرى أو غير ذلك من تطبيقات.

الترجمة الآلية واللغة العربية:

قمنا بتجربة مقارنة بين ترجمتي جوجل و "جي بي تي" على ترجمة 12 جملة تحوي كل منها كلمة جذرها (عين) بمعنى مختلف. وهذه الجمل مع نتائج ترجماتها مبينة في الجدول (3) في الملحق (4) كمثال لفحص دقة ترجمة نظامي الترجمة (جوجل و"جي بي تي"). وبالطبع فإن دقة التعرف من السياق على ترجمة كلمة (العين) يحتاج إلى عدد كاف من الأمثلة لكل معنى من المعاني ولا يكفي مثال واحد.

نلاحظ أن ترجمة الجملة الثانية لكلا الترجمتين كانت قريبة من الدقة بسبب أن كلمة العين البشرية هي أكثر المعاني شيوعاً وأكثرها وروداً في أية بيانات تدرب عليها النظام. أما الجملة السادسة عن الإصابة بالعين اكتشف "جي بي تي" معناها بينما فشل جوجل بذلك. وفي الجملة الثانية عشر اكتشف معنى قريباً وهو المكانة الإجتماعية المالية وهي قد تكون صحيحة جزئياً ربما لأن الأمثلة التي تدرب عليها تتعلق بالحياة الإقتصادية والمال، فقد اختار لترجمة عبارة (ملء العين) (beyond measure) والتي قد تترجم إلى (فوق العادة). تعتمد جودة برامج الترجمة الآلية على دقة النماذج اللغوية التي تدرب عليها البرنامج وعلى جودة المدخلات اللغوية (النص المطلوب ترجمته). وبالطبع كلما كانت القدرة على فهم الصيغ المجازية أو المسكوكات اللغوية أكبر كلما كانت دقة الترجمة أفضل. كما أن كثيراً ما يكون للمتحدثين في حقل معين إشارات يستخدمها المتحدثون بشكل شائع في نطاق تخصصهم معروفة فيما بينهم وقد يساء فهمها أو تكون غامضة لغير المتخصصين في ذلك الحقل. وهكذا نجد أن غالبية ترجمة الجمل في الجدول قد فشل النظامان في إدراك معناها الدقيق، مما يشير إلى قلة البيانات المستخدمة للتدرب عليها. وهذه النتائج تشير بوضوح إلى المستوى الذي وصلته الترجمة الآلية حالياً.

التشكيل الآلي:

أجريت محاولات للتشكيل الآلي للنصوص العربية بينها ما قام به عدد من الباحثين في الجامعة الأردنية (5) من خلال تدريب الشبكة العصبية على عدد كبير من النصوص المشكولة. يقوم البرنامج بتشكيل الحروف سواء أول الكلمة أو أوسطها أو آخرها، لكن الأخطاء في تشكيل الآخر أكثر وذلك لأن تشكيل الآخر يعتمد على السياق ويحتاج إلى مادة لغوية ضخمة لكي يتدرب عليها ومع هذا فقد بلغت نسبة الأخطاء في تشكيل آخر الكلمة 5.82% بينما بلغت فيما عداها داخل الكلمة 2.09%.

ولمقارنة ذلك مع تشكيل نظام "جي بي تي" قمنا بتجربة طلب تشكيل آلي لنص يتكون من 77 كلمة فوجدنا في النتيجة أن عدد الكلمات المشكولة الآخر كان 44، كان بينها 3 أخطاء في تشكيل الآخر، وكان عدد الكلمات التي لا تشكل عادة 6 كلمات، وكان هناك نقص أو خطأ في استعمال الشدة والهمزة

في 10 كلمات وكان هناك أخطاء أخرى. ويمكن الإستنتاج من هذه الأرقام التقريبية أن عملية التشكيل الآلي بواسطة هذا النظام لا تزال دون متوسط قابلية الإنسان العادي غير الخبير باللغة.

وزن الشعر العربي:

كانت أول محاولة جادة لوزن الشعر العربي بالحاسوب من قبل الدكتور محمد طارق الكاتب، حيث نشر جهوده عام 1971 في كتابه المعنون: "موازين الشعر العربي باستعمال الأرقام الثنائية" (6). وقد كانت تلك الجهود تستند إلى إدخال الأوزان الشعرية الستة عشر إلى الحاسوب بتحويل الحركة والسكون إلى الرقم 1 وصفر ومن ثم وزن الأبيات الشعرية بالقواعد التي أدخلت إلى الحاسوب وفق نماذج لكل وزن من الأوزان مع ما يطرأ عليها من زخافات، ويمكن للحاسوب أن يعطي وزن بيت الشعر المدخل من بين الأوزان الستة عشر إن كانت موافقة لأحدها أم لا.

أما المنهج المستند إلى الذكاء الإصطناعي فيقوم بتدريب الآلة على عدد كبير من الأبيات الشعرية فيتعلم البرنامج عملية الوزن وكأنه شاعر متمرس يعرف البيت غير الموزون من البيت الموزون ومن ثم يعطي من أي بحر هو. في البحث (7) الذي ساهم فيه المؤلف، جرى التدريب بإدخال ما يزيد على مليون و600 ألف بيت من الشعر المعروفة أوزانها، ومن ثم بعد التدريب يمكن الحصول على وزن البيت المدخل الجديد لمعرفة وزنه من الأوزان الستة عشر. وبعد فحص البرنامج على عدد من الأبيات الشعرية الجديدة من غير المتدرب عليها أمكن الحصول على نسبة دقة تبلغ 97.27% ويمكن زيادة هذه النسبة إذا ما كان البيت من الشعر مشكولاً. ويمكن استخدام برامج التشكيل الآلي قبل القيام بوزن الشعر ومن ثم الحصول على تلك الدقة العالية.

5. تجارب على نظام جي بي تي <https://chat.openai.com>

كيف يقوم "جي بي تي" بالإجابة على الأسئلة باللغة العربية؟

لا تزال النصوص العربية التي يتعامل معها هذا النظام محدودة، لذلك فحين يتلقى سؤالاً باللغة العربية يحاول فهم السؤال ثم يجيب عليه استناداً إلى النصوص باللغة العربية التي تم تدريبه عليها وما يشابهها، فإذا لم يجد الجواب يتحول للبحث والعثور على معلومات ذات صلة بلغات أخرى فيقوم بترجمتها لإعطاء الإجابة المناسبة. وبالطبع فإن الوضع العام للإجابة بالوقت الحاضر ليس دقيقاً وقد يحوي الكثير من الأخطاء. وحيث أن اللغة الإنكليزية هي أكثر اللغات التي تدرّب عليها، لذلك فإن كمية المعرفة المتوفرة داخل المنظومة يتوفر أغلبها باللغة الإنكليزية وما تحتاجه اللغات الأخرى يترجم إليها من اللغة الإنكليزية وأحياناً من لغات أخرى. أما موضوع الترجمة بين اللغات فهو يجري بالرجوع إلى البيانات متعددة اللغات

التي تدرب عليها والتي تحتوي على أزواج من الجمل متعددة اللغات والموارد اللغوية الضخمة التي استعملت لتدريبه عليها.

فيما يأتي عدد من الأمثلة لتجارب أجريت على "جي بي تي" لفحص أولي لمعرفة مستوى النظام باللغة العربية:

المثال الأول: لفحص مستوى المعرفة بتاريخ الأدب العربي... من هو المتنبي؟

جواب "شات جي بي تي" هو:

"أبو الطيب أحمد بن الحسين بن علي المتنبي، المعروف بالمتنبي، هو شاعر عربي من أعظم شعراء العربية الكلاسيكية. وُلد في العام 915 ميلادي في المدينة المنورة في الحجاز (المملكة العربية السعودية الحالية)، وتوفي في العام 965 ميلادي في مدينة الأهواز بالعراق. المتنبي اشتهر بأسلوبه الرقيق وقوة تعبيره وحسن استخدامه للعديد من الأشكال الشعرية مثل الغزل والمديح والفخر والوصف. يُعتبر المتنبي من الشعراء الذين أثروا في الأدب العربي وتركوا أثرًا بارزًا في تاريخ الشعر والأدب العربي. تتميز قصائده بعمق الفكر وتنوع المواضيع التي يتناولها، بدءًا من الحب والغزل وصولًا إلى السياسة والفلسفة والحكمة. يعتبر المتنبي رمزًا للشاعر الجاهلي والثوري، وقد تركت قصائده بصمة عميقة في تاريخ الأدب العربي واستمر تأثيرها على الشعراء والأدباء لعدة قرون بعد وفاته".

يلاحظ أن محل ولادة المتنبي خاطئ فهو ولد في الكوفة وكذلك وفاته التي كانت قرب الصافية شرق نهر دجلة في العراق. هذا بالإضافة إلى عدم الدقة في بعض الأوصاف التي أسندت إليه.

المثال الثاني: ما هو الفرق بين الشعر الحر والشعر العمودي؟

جواب "شات جي بي تي" هو:

"الشعر الحر (المنثور) والشعر العمودي (المقفى) هما نمطان أساسيان في الشعر يختلفان في هيئتهما وتنظيمهما. إليك الفرق بينهما:

1. الشعر الحر (المنثور):

• يتميز بالحرية في التنظيم والتكوين، حيث لا يلتزم بالأنماط الشعرية التقليدية أو القيود الصارمة للقافية والوزن.

• يتناول المواضيع المتنوعة بشكل واسع، ولا يقتصر على موضوع معين.

• يعتمد على العاطفة والتعبير الشخصي والحرية الإبداعية.

• قد يكون غير منظم من الناحية الهيكلية، ويمكن أن يكون طويلًا أو قصيرًا حسب رغبة الشاعر.

2. الشعر العمودي (المقفى):

• يتميز بترتيب محكم ومنظم، حيث يتبع نمطًا قويًا في التنظيم والقافية والوزن.

- يتعامل مع موضوع محدد واحد، وغالبًا ما يكون موجزًا ومختصرًا.
 - يحترم القواعد والأنماط الشعرية التقليدية مثل البيت الشعري والقافية والوزن.
 - يعتمد على الشكل والتوزيع المنتظم للأبيات والقوافي والوزن لإنشاء تناغم وجمالية.
- الفرق الرئيسي بين الشعر الحر والشعر العمودي يتمثل في التنظيم والقواعد المتبعة. الشعر الحر يمنح الشاعر حرية أكبر في التعبير والتكوين، في حين يتطلب الشعر العمودي الالتزام بقواعد محددة ومنظمة". ويلاحظ الوصف العام في هذا الجواب غير الدقيق في معظم عباراته.

المثال الثالث: لمعرفة مكانة اللغة العربية بين لغات العالم: ما هي الأمور المشتركة بين اللغة العربية واللغات السامية الأخرى؟

جواب "شات جي بي تي"

"اللغة العربية تنتمي إلى عائلة اللغات السامية، وهناك عدة أمور مشتركة بين اللغة العربية واللغات السامية الأخرى. إليك بعض هذه الأمور:

1. النظام اللغوي: اللغات السامية تتشابه في هيكلها اللغوي ونظامها الصرفي والتصريفي. تتميز بوجود أصوات مشابهة وجذور لغوية مشتركة.
 2. الجذور اللغوية: اللغات السامية تتشارك في العديد من الجذور اللغوية الأساسية والتي يمكن تتبعها في الكلمات المشتركة بينها.
 3. النظام الثلاثي: اللغات السامية تعتمد على النظام الثلاثي في تصريف الأفعال، حيث يتم تصريف الأفعال وفقًا للزمن والشخص والعدد.
 4. التراكيب اللغوية: اللغات السامية تتشابه في استخدام التراكيب اللغوية المشتركة مثل الجملة الاسمية والجملة الفعلية والجملة الحالية.
 5. الألفاظ المشتركة: هناك العديد من الكلمات المشتركة بين اللغات السامية، والتي يمكن أن تشير إلى الصلة اللغوية بينها.
- ومع ذلك، يجب ملاحظة أن اللغات السامية الأخرى تختلف في العديد من التفاصيل والتطورات اللغوية عن اللغة العربية، ولذلك فإن كل لغة سامية لها خصوصيتها الفريدة ومميزاتها الخاصة". وهذه الإجابة العامة كذلك فيها كثير من عدم الدقة.

المثال الرابع: لمعرفة مدى القابلية على التشكيل الآلي

المطلوب تشكيل الآية: إنما يخشى الله من عباده العلماء

جواب "شات جي بي تي"

إِنَّمَا يَخْشَى اللَّهَ مِنْ عِبَادِهِ الْعُلَمَاءُ

فكان التشكيل دقيقاً رغم تأخر الفاعل إلى نهاية الجملة وتفصله عن الفعل ثلاث كلمات.

المثال الخامس: فحص المستوى الأدبي للنظام وذلك بطلب كتابة مقالة أدبية تصف "احتفاء الأم بتخرج ولدها الأصغر من الجامعة".

الجواب في الملحق (5) والذي يمكن الإستنتاج منه بأن الذكاء الإصطناعي قد بلغ شأواً بعيداً رغم وجود ركافة بعض الكلمات التي أشير لها بوضع خط تحت تلك الكلمات.

6. المناقشة والإستنتاجات:

إن التقدم الذي شهده الذكاء الاصطناعي في العقود الأخيرة كان بطيئاً في بداياته، إلا أنه تسارع بشكل مذهل خلال السنوات الأخيرة وخاصة الأشهر الماضية التي أعلن فيها عن أنظمة جاهزة. وهذا التسارع ينبئ بتقدم أسرع وأدق خلال السنوات القليلة القادمة. ويصدق ذلك على دخول اللغة العربية في هذا الحقل الهام من بين الكثير من الحقول التي دخلها أو سيدخلها الذكاء الإصطناعي.

وبالرغم من أنه يمكن ملاحظة العديد من المثالب وعدم الدقة فيما يتعلق بالتعامل مع اللغة العربية فهماً أو ترجمة أو تعبيراً، إلا أن المقارنة بين الوضع الحالي وما كان عليه الوضع قبل بضعة أعوام ينبئ بتخطي الكثير من المثالب في المستقبل. فما يتوقع هو المزيد من الدقة في فهم المضمون والترجمة من العربية وإليها وحسن التعبير والنطق والذي سيكون بالإمكان استعماله في المحادثة والتعليم والأغراض المختلفة الأخرى.

ويؤمل تطور التعليم الآلي للغة العربية سواء التعليم الجماعي أو التعليم الفردي لكل حسب قابليته وتغطية اللهجات المحلية وما يدخل معها من كلمات أجنبية. ويؤمل أن يجري في المستقبل مواكبة التطور في دخول كلمات ومعان جديدة ضمن سياقاتها وتوليد كلمات جديدة آلياً للأشياء وللمعاني التي ستظهر في المستقبل. إن التقدم في البلاغة يمثل أحد التحديات المهمة التي تواجه الذكاء الإصطناعي.

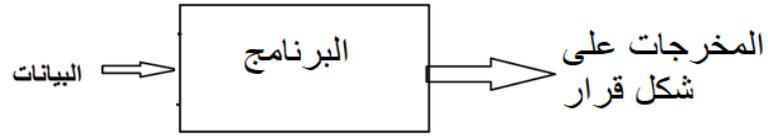
ما تحتاجه أنظمة الذكاء الاصطناعي فيما يخص اللغة العربية هو المزيد من توفير البيانات المحوسبة باللغة العربية لغرض التدريب والمعاجم اللغوية المحوسبة. كما هناك ضرورة للتعمق في أبحاث المعالجة الحاسوبية للغة العربية بالتعاون بين اللغويين واللسانيين والحاسوبيين.

7. المصادر والمراجع

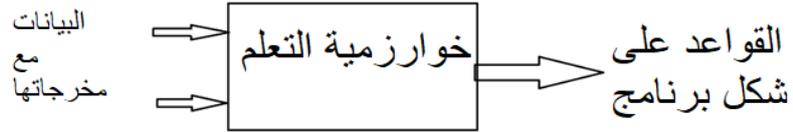
1. محمد زكي خضر، مشكلات الترجمة الآلية،الموسم الثقافي التاسع والثلاثون لمجمع اللغة العربية الأردني 6 - 2021/12/19، منشورات مجمع اللغة العربية الأردني 1443 هـ / 2022م صفحة 255-283.
2. <https://it-solutions.center/tag/%D8%AA%D8%B9%D8%B1%D9%8A%D9%81-chat-gpt/2>.
3. <https://www.hpcwire.com/2023/05/10/googles-new-ai-focused-a3-supercomputer-has-26000-gpus/3>.
4. <https://ai.facebook.com/blog/large-language-model-llama-meta-ai/>
5. Gheith A. Abandah _ Alex Graves _ Balkees Al-Shagoor _ Alaa Arabiyat _ Fuad Jamour _ Majid Al-Tae , Automatic diacritization of Arabic text using recurrent neuralNetworks,International Journal on Document Analysis and Recognition (IJDAR) June 2015, 18(2):183-197
- 6- محمد طارق الكاتب، موازين الشعر العربي باستخدام الأرقام الثنائية، مطبعة المواني العراقية البصرة 1392هـ/1971م
7. Gheith A. Abandah, Mohammed Z. Khedher, Mohammad R. Abdel-Majeed, Hamdi M Mansour,Salma F Hulliel, Lara M Bisharat, Classifying and Diacritizing Arabic poems using deep recurrent neural networks, Journal of King Saud University – Computer and Information Sciences 34 (2022) 3775-3788
8. <https://chat.openai.com>

الملحق (1)

البرامج التقليدية



التعلم الآلي



الشكل (1) الفرق بين البرامج التقليدية والتعلم الآلي

الملحق (2)

مقارنة ترجمة جملة بواسطة جوجل بين عامي 2016 و 2023

Lawyers for the four Iraqis confirmed the first suit was registered at the US District Court in Seattle, Washington	العبارة الإنكليزية
أكد المحامون عن العراقيين الأربعة أن الدعوى الأولى قد سجلت في الولايات المتحدة لدى المحكمة في سياتل بواشنطن	الترجمة الدقيقة
محامون لأربعة عراقيين وأكدت الدعوى الأولى المسجلة في الولايات المتحدة أمام المحكمة المحلية في سياتل، واشنطن	ترجمة جوجل عام 2016
وأكد محامو العراقيين الأربعة أن الدعوى الأولى سُجلت في المحكمة الجزئية الأمريكية في سياتل بواشنطن	ترجمة جوجل عام 2023

الملحق (3)

مقارنة ترجمة جمل متدرجة التعقيد بواسطة جوجل

الترجمة الإنكليزية	الجملة العربية	
The student entered the classroom very quickly	دخل الطالب الصف بسرعة كبيرة (قبل عامين ومؤخرًا)	1
Grade student entered the classroom very quickly	التالي دخل الصف بسرعة كبيرة (قبل عامين)	2
The student entered the classroom very quickly	التالي دخل الصف بسرعة كبيرة (مؤخرًا)	3
The student entered the classroom quickly carrying a red bag friend	دخل الطالب الصف بسرعة حاملًا حقيبة صديقه حمراء اللون (قبل عامين)	4
The student entered the classroom quickly carrying his friend's red bag	دخل الطالب الصف بسرعة حاملًا حقيبة صديقه حمراء اللون (مؤخرًا)	5
The student entered the classroom quickly carrying a bag of red color of his friend	دخل الطالب الصف بسرعة حاملًا حقيبة صديقه الحمراء اللون (مؤخرًا)	6
Translation from Arabic and into it is one of the means into which a great service lies in the renaissance of the nation and its future generations	الترجمة الآلية من اللغة العربية وإليها إحدى الوسائل التي تكمن فيها خدمة كبيرة في نهضة الأمة وأجيالها المستقبلية (مؤخرًا)	7

الملحق (4)

ترجمة جمل تحوي جذر "عين" بواسطة جوجل و "جي بي تي"

الجمل المحتوية على كلمة "عين"	جوجل ترجمة	"جي بي تي" ترجمة
في القرية عين زلال	In the village of Ain Zilal	In the village of Ain Zalal
أصيببت عينه بالرمد	His eye was infected with ophthalmia	His eye was infected with conjunctivitis
عين الفعل سَعَد مكسورة	The eye of the verb saad is broken	The verb 'Ain Sa'd' is broken
العين من حروف الإظهار	The eye is one of the letters of manifestation	'Ain' is a letter of indication
أعاد البضاعة عينها	Returned the same goods	He returned the goods in kind
هل تؤمن بالإصابة بالعين؟	Do you believe in eye infection?	Do you believe in the evileye?
عَيَّن السهم صوب الهدف	Point the arrow at the target	The arrow is aimed to the target
زيد يعين أباه في عمله	Zaid helps his father in his work	Zaid assists his father in his work
عَيَّن الخريج في الشركة	The graduate is assigned to the company	The graduate was appointed in the company
كان يعمل عيَّنًا للعدو	He was working as an eye for the enemy	He used to work as a spy for the enemy
هو الآن عضو في مجلس الأعيان	He is now a member of the Senate	He is now a member of the council of the elders
كان ملء العين ضمن أسرته	Fill in the eye was within his family	He was wealthy beyond measure within his family

الملحق (5) مقالة أدبية

فحص المستوى الأدبي لنظام "شات جي بي تي" وذلك بطلب كتابة مقالة أدبية تصف "احتفاء الأم بتخرج ولداها الأصغر من الجامعة".

"في يوم مشرق من أيام الربيع كانت الأم تستعد بفرح وحماس لإحتفال خاص. فقد حان اليوم المنتظر، يوم تخرج ابنها الصغير من الجامعة. كانت الأم تشعر بالفخر والسعادة العارمة لما تحقّقه ابنها من نجاح وتحصيل علمي.

إستعدت الأم بأفضل ما لديها لتحضير الحفل البسيط لهذه المناسبة الفريدة. قامت بتزيين المنزل بالزهور الجميلة والستائر اللامعة. أعدت أشهى المأكولات المفضلة لابنها وأفراد العائلة. كانت ترغب في أن يكون هذا اليوم يوماً لا يُنسى للجميع.

وصل الوقت المناسب وحانت لحظة دخول ابنها المنزل. لم يكن هناك شك في الفرح الذي أضاء وجهه والنجاح الذي انبعث من كل كلمة قالها. كانت الأم لا تستطيع أن تكبح دموع الفرح والسعادة التي تنهمر من عينيها.

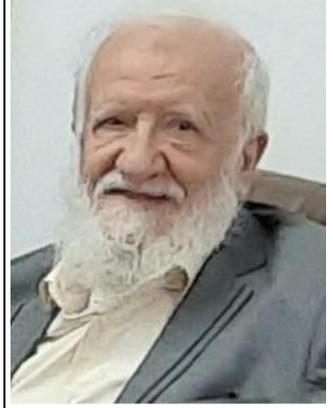
تتأثر التهناني والتبريكات من أفراد العائلة، وانطلقت الزغاريد والتصفيق في أرجاء المنزل. كانت الأم تقف في وسط الحشد العائلي المبهج وتحضن ابنها بقوة وهي تتوالى عبارات الفخر والإمتنان له. قالت الأم بصوت مؤثر: " يا حبيبي، أنت تمثل الأمل والمستقبل لنا جميعاً. أنا فخورة جداً بك وبكل ما حقّفته. لقد عملت بجد ومثابرة، والآن حان وقتك لتعيش حياتك وتحقق أحلامك. أعلم أن لديك المقدرة على تحقيق المزيد من النجاحات في المستقبل. أنت فرحي وسعادتي الأبدية".

كانت هذه المناسبة تجسد حب الأم واهتمامها العميق بنجاح ابنها. لقد كانت فرصة لتعبيرها عن فخرها ودعمها الكامل لابنها في مرحلة جديدة من حياته. وتذكرت الأم أثناء الاحتفال كل اللحظات الصعبة التي تجاوزها معاً وكل الصبر والتضحيات التي قدمتها لتشجيعه ودعمه في رحلته الأكاديمية. كما أبدت الأم في رغبتها في مستقبله المشرق وأن يتحقق له كل ما يطمح إليه. وعازمة على دعمه ومساعدته في كل خطوة يخطوها في الحياة.

كان هذا الاحتفال بمناسبة تخرج الابن هو بمثابة تكريم لمسيرته التعليمية وثمره للجهود التي بذلها على مدار السنوات العديدة. ولم تكن هناك أفضل طريقة للأم للتعبير عن حبها وفخرها ودعمها له من خلال هذا الاحتفال الخاص.

8. السيرة الذاتية للمؤلف

الأستاذ الدكتور محمد زكي خضر حصل على شهادة البكالوريوس في الهندسة الكهربائية من جامعة بغداد عام 1965، عمل في جامعة الموصل كمحاضر منذ عام 1965 ثم التحق بالدكتوراه وحصل عليها من جامعة شيفلد في بريطانيا في عام 1972. شغل عدد من المناصب الجامعية في جامعة الموصل حيث أسس مركز الحاسب عام 1974 وكان أول مديراً له وترأس قسم الهندسة الكهربائية بين عامي 1972-1977. والتحق بجامعة مانشستر في بريطانيا عام 1977-1978 ثم عاد إلى جامعة الموصل وحصل على لقب أستاذ في عام 1981. وعمل في الصناعة بين عامي 1981 و 1992، ثم التحق بالجامعة الأردنية كأستاذ وعمل فيها بين عامي 1992 و 2015. له عدد من الأبحاث والكتب العلمية المؤلفة والمترجمة في حقول الهندسة والحاسوب، وهو عضو دائم في جمعية مهندسي الكهرباء والإلكترونيات منذ عام 1968 ومؤسس فرع الجمعية في الأردن. له العديد من المؤلفات الإسلامية المرتبطة بالحاسوب. يشرف على مشروع مداد البيان في خدمة القرآن الكريم، ويرأس المؤتمر الدولي للتطبيقات الإسلامية في علوم الحاسوب والتقنية - إيمان منذ إنطلاقه في عام 2013. ويعمل كرئيس تحرير للمجلة الدولية للتطبيقات الإسلامية في علوم الحاسوب وتقنياته - إجازات منذ تأسيسها عام 2013. ويرأس حالياً المعهد العالمي لحوسبة القرآن والعلوم الإسلامية. أشرف الأستاذ الدكتور محمد زكي على العديد من طلبه الدراسات العليا وحاز على عدد من الجوائز الدولية ولديه العديد من المؤلفات في الحقل العلمي والإسلامي. وحصل مؤخراً على لقب عالم من منصة أريد للباحثين العرب.



Artificial Intelligence Systems and Arabic Language

Mohammed Zeki Khedher

President of International Institute for Quran and Islamic Sciences Computing
president@qurancomputing.org

Abstract: The term artificial intelligence is used to refer to computer software that contains intelligent processes that simulate the intelligence of living creatures such as a plant, animal, or human being. The term artificial intelligence has recently often use to mean computer transactions that perform machine learning and are then used after learning to solve the problems designed for them. Therefore, machine learning has become the focus of artificial intelligence operations today. Machine learning needs huge amounts of data in order to learn from it and build its algorithms (which often use neural networks) in order to use them later on new data to solve the problems for which it was designed. Current artificial intelligence systems use information available on the Internet, books, newspapers, and documents available to extract from them algorithms for receiving linguistic texts, analyzing them, understanding their meaning, then carrying out the operations required of them computerized, and converting the results of those operations into written or spoken texts or industrial processes used in robots or other machines or other applications. In this paper, we will present some of the available artificial intelligence software and its use and examine its capabilities to deal with the Arabic language in order to show the status it reached. It also shows the shortcomings, with some focus on machine translation, and then provides some highlights and proposals regarding the future steps necessary to improve the handling of that software with the Arabic language.

Keywords: Artificial intelligence systems, Arabic NLP, Technology Education, Machine learning & the Arabic language.