

المؤتمر الدولي الأول للغة العربية في زمن الذكاء الاصطناعي

مجمع اللغة العربية الأردني

٢٠٢٣/١٢/٢٨-٢٧

تحديات استعمال اللغة العربية في أنظمة الذكاء الاصطناعي

Challenges of Using Arabic Language in

Artificial Intelligence Systems

أ. د. محمد زكي خضر

رئيس المعهد العالمي لحوسبة القرآن والعلوم الإسلامية

president@qurancomputing.org

khedher@gmil.com

المخلص

تستعمل أنظمة الذكاء الاصطناعي الحالية المعلومات المتوفرة على شبكة الشابكة (الإنترنت) والكتب والصحف والوثائق المتوفرة لكي تستخلص منها خوارزميات استقبال النصوص اللغوية وتحليلها وفهم معناها ثم القيام بالعمليات المطلوبة منها حاسوبيا 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 بإعطاء مقدمة عن بعض ما يتوفر من برمجيات الذكاء الاصطناعي واستخداماتها وفحص قابليتها للتعامل مع اللغة العربية لتبيان الشوط الذي قطعتة وما تشكو من نقائص.

تتوزع التحديات في استعمال اللغة العربية في أنظمة الذكاء الاصطناعي على عدد من الجوانب : تحدي الكتابة الإملائية والتشكيل وتحدي النحو والصرف وتحدي الفهم الآلي والذي تعتمد عليه عدة أمور منها الترجمة الآلية وتحدي التعبير والبلاغة وتحدي فهم الكلام وسلامة النطق والكلام وسيتضمن البحث إجراء تجارب لأمتثلة تعكس استجابة بعض أنظمة الذكاء الاصطناعي لجوانب من هذه التحديات ، ومن ثم تقديم بعض المؤشرات والمقترحات بشأن الخطوات المستقبلية اللازمة لتحسين تعامل تلك البرمجيات مع اللغة العربية بشكل أفضل مما هو عليه الحال في الوقت الحاضر.

كلمات مفتاحية: أنظمة الذكاء الاصطناعي، الذكاء الاصطناعي واللغة العربية، الفهم الآلي للغة العربية ، التعلم الآلي واللغة العربية.

تعريف أولية

الذكاء الاصطناعي (Artificial intelligence): هو حقل جديد من التطبيقات الحاسوبية والذي تحاكي فيه البرامج الحاسوبية القدرات الذهنية البشرية وطرق عملها بحيث تستطيع التعلم والإستنتاج ورد الفعل لحالات لم تبرمج أصلاً عليها. وحيث أن تعريف الذكاء البشري أصلاً محل مناقشات دون الوصول إلى تعريفه بدقة ، فكذلك الذكاء الاصطناعي ، 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 الاصطناعي تسمح للأنظمة بفهم وتفسير وإنتاج اللغة البشرية بشكل طبيعي كي 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. فاللغوي 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 الاصطناعية (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 عُنْدًا أو عصبونات مترابطة فيما بينها في بنية مكونة من طبقات. وكلما كانت المسائل المراد حلها أعقد كان هناك حاجة إلى عدد أكبر من الطبقات والعقد.

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 المباشرة. يعتمد التعلم الآلي على تطوير النماذج الرياضية والخوارزميات التي تمكن الأنظمة من استيعاب البيانات وتحليلها واستخلاص المعرفة منها، وهو يتيح للأنظمة أن تتعلم 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.

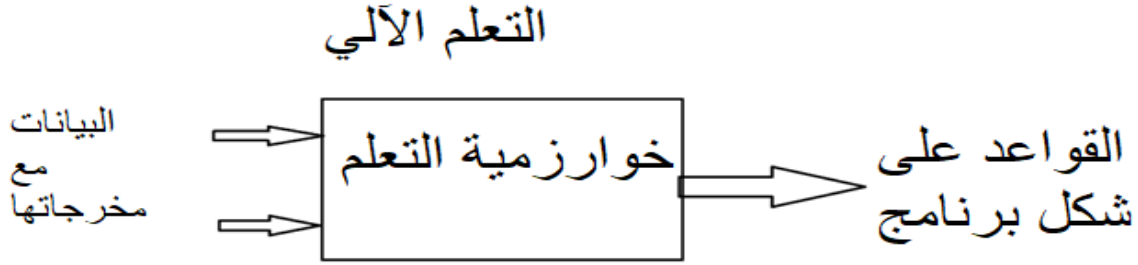
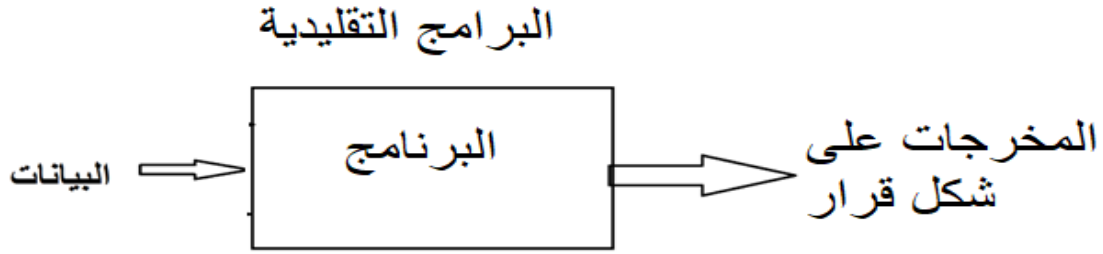
في سياق الذكاء الاصطناعي، يتم استخدام التعلم الآلي كأساس لتطوير نماذج وأنظمة ذكاء اصطناعي قادرة على التعلم والتكيف واتخاذ القرارات. يتم تدريب هذه النماذج على البيانات المتاحة وتعديل معلوماتها لتحسين أدائها مع مرور الوقت. ونتيجة لذلك يمكن للأنظمة أن تتعلم من البيانات المتاحة وتحليلها وفهمها واستخلاص المعرفة 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: كيف كان يتعلم الطفل اللغة العربية في البداية حينما كانت البداية موطنًا ثريًا للغة العربية الفصحى.

يتعلم الطفل أصواتًا 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 فينتدرج 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. لكنه يتعلم أن جمع سماء – سماوات... ويبدأ بالتفريق بين المذكر والمؤنث

فيعرف أن جالسة هي كلمة مؤنثة وكلمة جالس تشير إلى مذكر وجمع كل منهما مثلا جمع مؤنث سالم (الكلمة 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 مثلا 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 مضافا 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 تستنتج منها القاعدة الجديدة ولا تكتفي بمثال واحد... وهكذا فإن وفرة المعلومات التي تتعلم منها يزيد من الدقة في الأداء تدريجياً. كذلك فإن تشكيل الكلمة (بالضمة والفتحة والكسرة والتنوين 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 (١) مقارنة بين مدخلات ومخرجات البرامج الحاسوبية العادية وبرامج التعلم الآلي.



الشكل (١) الفرق بين البرامج التقليدية والتعلم الآلي

أنظمة الذكاء الاصطناعي

نستعرض فيما يأتي بعض أنظمة الذكاء الاصطناعي ومنها نظم الترجمة الآلية كترجمة جوجل مثلا. 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.

إنتشرت في العقود الثلاثة الأخيرة أنظمة عديدة للترجمة من العربية وإليها. ولغرض تقويم بعض هذه الأنظمة لمعرفة نقاط الضعف فيها جرت محاولات كثيرة منها مقارنة أنظمة ترجمة جوجل وسيستران وبينج على نصوص علمية وإخبارية وأدبية لغرض معرفة ما تشير إليه من نقاط الضعف في بعض تلك الأنظمة (١)

إن أحد الأسباب الفنية في وجود ضعف في 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 البيئـة والزراعة والغابات والمناطق السياحية مثلا يجعل ترجمة كلمات ذات علاقة بتلك الحقول أكثر دقة (١).

قمنا بتجربة لمعرفة التقدم في أشهر ترجمة متوفرة الآن وهي ترجمة جوجل (٢) بمحاولة الحصول على ترجمة لجملة معينة عام ٢٠١٦ وأخرى مؤخرًا خلال عام ٢٠٢٣. يبين الجدول (١) الجملة الأصلية وترجمتها خلال العامين المذكورين مع الترجمة الدقيقة لها. ويلاحظ بوضوح الزيادة في دقة الترجمة مؤخرًا.

جدول (١) مقارنة ترجمة جملة بواسطة جوجل بين عامي ٢٠١٦ و ٢٠٢٣

العبرة الإنكليزية	Lawyers for the four Iraqis confirmed the first suit was registered at the US District Court in Seattle, Washington
الترجمة الدقيقة	أكد المحامون عن العراقيين الأربعة أن الدعوى الأولى قد سجلت في الولايات المتحدة لدى المحكمة في سياتل بواشنطن
ترجمة جوجل عام ٢٠١٦	محامون لأربعة عراقيين وأكدت الدعوى الأولى المسجلة في الولايات المتحدة أمام المحكمة المحلية في سياتل، واشنطن
ترجمة جوجل عام ٢٠٢٣	وأكد محامو العراقيين الأربعة أن الدعوى الأولى سُجلت في المحكمة الجزئية الأمريكية في سياتل بواشنطن

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 (٢) 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.

جدول (٢) مقارنة ترجمة جمل متدرجة التعقيد بواسطة جوجل

الجملة العربية	الترجمة الإنكليزية
----------------	--------------------

The student entered the classroom very quickly	دخل الطالب الصف بسرعة كبيرة (قبل عامين ومؤخرًا)	١
Grade student entered the classroom very quickly	التالي دخل الصف بسرعة كبيرة (قبل عامين)	٢
The student entered the classroom very quickly	التالي دخل الصف بسرعة كبيرة (مؤخرًا)	٣
The student entered the classroom quickly carrying a red bag friend	دخل الطالب الصف بسرعة حاملًا حقيبة صديقه حمراء اللون (قبل عامين)	٤
The student entered the classroom quickly carrying his friend's red bag	دخل الطالب الصف بسرعة حاملًا حقيبة صديقه حمراء اللون (مؤخرًا)	٥
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)	٦
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)	٧

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 (٣)

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 يطرحها 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. ظهرت من هذا النظام طبعات متسلسلة، فكانت الطبعة ٣ ثم ٣,٥ ثم ٤ وهي في تحسن مستمر.

نظام جوجل للذكاء الاصطناعي (٤)

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 ٢٦٠٠٠ وحدة معالجة للرسومات (بالمقارنة مع أسرع حاسوب عملاق متوفر الآن - فرونتير - الذي يحوي ٣٧٠٠٠ وحدة معالجة للرسومات). يعد هذا الحاسوب العملاق محاولة من الشركة من أجل التفوق في حقل الذكاء الاصطناعي بالتنافس مع شركة مايكروسوفت.

نظام شركة ميتا (٥)

أطلقت شركة ميتا Meta نموذجها الخاص بها الذي أسمته لاما LLaMA وهو نموذج لغة كبير أساسي حديث هو 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 كبيرة لنموذج اللغة. 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.

نظام فالكون (٦)

هذا النظام قام بتطويره مركز أبو ظبي للأبحاث في دولة الإمارات العربية المتحدة ويعتبر نظام ذكاء اصطناعي متقدم فهو يتصدر نماذج اللغة المفتوحة المصدر بالقدرات حالياً ويغلب نظام اللاما-٢ وهو يقارب نظام "جي بي تي ٤" وقد جرى تدريبه على ٣,٥ ترليون رمز يسمى (Token)، ولكنه ثقيل المتطلبات بالحوسبة ويمكن تطوير التطبيقات وتوزيعها أو بيعها بدون رسوم للمركز الذي طوره. ومع كل تلك المزايا فهو لا يتكلم باللغة العربية بخلاف الأنظمة السابقة رغم أنه طور في دولة عربية. وهناك شركات أخرى تقوم بمحاولاتها في هذا الحقل مثل أمازون وغيرها.

نظام أنصاري (٧)

هو نظام طور حديثاً ويستعمل نظام "جي بي تي" لكنه متخصص بالمعلومات الإسلامية وخاصة القرآن الكريم وهو لا يزال قيد التطوير ويحوي الكثير من الأخطاء وسنكشف عن بعض قابلياته في الأمثلة التي سترد في هذا البحث بمقارنته مع نظام "جي بي تي"

تعامل أنظمة الذكاء الاصطناعي مع اللغة العربية

تتكون الذخيرة اللغوية التي يعمل عليها الذكاء الاصطناعي من المعلومات المستقاة من الشبكة (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 والبحث عن الإجابة ثم صياغتها بأسلوب مناسب.

تعتمد جودة ودقة الإجابات التي يقدمها النظام على عدة عوامل، بما في ذلك فهم السؤال واستيعاب السياق، ونوعية البيانات التي تدرب 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.

وتُعتبر 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 الشبكة (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 ١٢ جملة تحوي كل منها كلمة جذرها (عين) بمعنى مختلف. وهذه الجمل مع نتائج ترجماتها مبينة في الجدول (٣) 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.

جدول (٣) ترجمة جمل تحوي جذر "عين" بواسطة جوجل و "جي بي تي"

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

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 إختار 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 (٨) 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 ٥,٨٢% بينما بلغت فيما عداها داخل الكلمة ٢,٠٩% . وسنأتي على مقارنة ذلك مع تشكيل نظام "جي بي تي" فيما بعد

وزن الشعر العربي

كانت أول محاولة جادة لوزن الشعر العربي بالحاسوب من قبل الدكتور محمد طارق الكاتب، حيث نشر جهوده عام ١٩٧١ في كتابه المعنون: "موازن الشعر العربي باستعمال الأرقام الثنائية" (٩). وقد كانت تلك الجهود تستند إلى إدخال الأوزان الشعرية الستة عشر إلى الحاسوب بتحويل الحركة والسكون إلى الرقم ١ وصفر ومن ثم وزن الأبيات الشعرية بالقواعد التي أدخلت إلى الحاسوب وفق نماذج لكل

وزن من الأوزان مع ما يطرأ عليها من زحافات، ويمكن للحاسوب أن يعطي وزن بيت الشعر المدخل من بين الأوزان الستة عشر إن كانت موافقة لأحدها أم لا.

أما المنهج المستند إلى الذكاء الاصطناعي 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 (١٠) الذي ساهم فيه المؤلف، جرى التدريب بإدخال ما يزيد على مليون و ٦٠٠ 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 ٩٧,٢٧% 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.

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 كما سنرى في الأمثلة. وحيث أن اللغة الإنكليزية هي أكثر اللغات التي تدرب 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.

تجارب على استجابة نظام "جي بي تي" للغة العربية (١١)

فيما يأتي عدد من الأمثلة لتجارب أجريت على "جي بي تي" بطبعة ٣,٥ لفحص أولي لمعرفة مستوى النظام باللغة العربية:

١- تجربة عن حرفي الضاد والظاء

السؤال: المطلوب خمس أمثلة على كلمات تحوي حرف الضاد وخمس كلمات أخرى تحوي حرف الظاء ومعنى كل كلمة.

الجواب: بالطبع، إليك خمس كلمات تحتوي على حرف الضاد وخمس كلمات تحتوي على حرف الظاء مع معانيها:

كلمات تحتوي على الضاد:

١. ضوء: يشير إلى النور أو الإضاءة.
٢. ضرب: الفعل الذي يشير إلى الخلق أو اللكم.
٣. عضد: يشير إلى الدعم أو السند.
٤. استدار: يعني الدوران أو الاستدارة.
٥. مضاد: شيء مضاد يقابل شيئاً آخر، أو عكسه.

كلمات تحتوي على الظاء:

١. ظل: الظل هو الظل الذي يسببه جسم ما بفعل الضوء.
٢. ظريف: يعني لطيف أو راقى.
٣. ظهر: الجزء الخلفي من جسم الإنسان.
٤. مظهر: الشكل الخارجي أو الشكل الظاهري لشيء ما.
٥. ظرف: وهو كلمة تستخدم في اللغة العربية لتعني "زمان" أو "زماناً" وتستخدم في ترتيب الأحداث.

ملاحظة: لوحظ أن التمييز بين الضاد والظاء دقيق فيما عدا كلمة واحدة (إستدار) أدرجت ضمن

حرف الضاد وهذا خطأ واضح ، مما يشير إلى إمكانية إجابة النظام إجابات غير منطقية وغير

مبررة (ربما الأصل الذي وجده هو "إضطر" فحرفت نتيجة ترجمة إلى الإنكليزية هكذا). أما المعاني

فكانت على العموم لا بأس بها رغم أنه اختار أحد المعاني ربما الأكثر شيوعاً ضمن ما تدرّب عليه (مثل عضد حين أعطى معنى المساندة أو الدعم ولم يشر إلى عظم العضد من اليد)

٢- تجربة كتابة الهمزة

أعطيت بعض الكلمات التي تحوي الهمزة بطريقة ما وطلب إعادة الكتابة بطريقة صحيحة فكانت النتائج كما هي في الجدول (٤)

جدول (٤) كلمات تحوي الهمزة

الكلمة المعطاة	النتيجة	ملاحظات
عبيء	باء	خطأ ربما لعدم التعرف على حرف العين بالإنكليزية
امروء	إنسان	بدل تصحيح الكتابة أعطي المرادف لكلمة أمرؤ
إمرءة	إمرأة	صحيح
بءر	بئر	صحيح
وءام	وعمام	خطأ
إختبء	اختبئ	خطأ - رغم إمكانية أن يقصد المبني للمجهول إختبئ
أخطئو	أخطئوا	خطأ - رغم إمكانية أن يقصد المبني للمجهول أخطئوا
شماءل	شمال	خطأ
موأل	موال	خطأ
يألف	يلف	خطأ
عبأة	عباءة	صحيح

ملاحظة: يلاحظ أن كتابة الهمزة بشكل صحيح كان قليلاً ويعني ذلك أن البيانات التي تدرّب عليها لم تحتوي عدداً كافياً من الهمزات المكتوبة بشكل صحيح خاصة مع احتمال التشكيل المختلف

٣- تجربة العدد والمعدود

السؤال (١) : كتابة ما يأتي كتابة وليس أرقاماً: ٧ رجال ٨ شجرة ١١ طالب ١٢ حقيبة

الجواب : سبعة رجال، ثمانية شجرات، أحد عشر طالباً، اثنتا عشرة حقيبة

السؤال (٢) : كيف يقرأ العدد التالي باللغة العربية ١٣٧٢٩٩٥

الجواب: العدد ١,٣٧٢,٩٩٥ يُقرأ في اللغة العربية كالتالي:

مليون وثلاثمائة واثنان وسبعون ألف وتسعمائة وخمسة وتسعون.

السؤال (٣) : ٢٧ رجلا و ١٢٢ طالبة

الجواب : العبارة "٢٧ رجلا" تُقرأ باللغة العربية كالتالي: سبعة وعشرون رجلا

ومئة واثنان وعشرون طالبة

ملاحظة: هناك بعض الأخطاء في التوافق بين العدد والمعدود في التذكير والتأنيث. كما أن قراءة الأعداد الكبيرة قرئت كما هو متداول وليس بالطريقة العربية الصحيحة

٤- تجربة إسمي الفاعل والمفعول للأفعال

المطلوب إعطاء إسم الفاعل وإسم المفعول للأفعال التالية
قال أكل بدأ عفا ملأ تيمم إطلع أدبر أسس عني عفى قلى

جدول (٥) أسماء الفاعل والمفعول

الفاعل	إسم	ملاحظات	إسم المفعول	ملاحظات
قال	الفاعل	صحيح مع ألف لام	القول	خاطئ أعطى المصدر
أكل	الآكل	صحيح مع ألف لام	الأكل	خاطئ أعطى المصدر
بدأ	البادئ	صحيح مع ألف لام	البدء	خاطئ أعطى المصدر
عفا	العافي	صحيح مع ألف لام	العفو	خاطئ أعطى المصدر
ملأ	المملي	غير صحيح	الملء	خاطئ أعطى المصدر
تيمم	المتيمم	غير صحيح	التيمم	خاطئ أعطى المصدر
إطلع	المطلع	صحيح مع ألف لام	الإطلاء	خاطئ تماما
أدبر	المدبر	صحيح مع ألف لام	الإدبار	خاطئ أعطى المصدر

أسس	المؤسس	صحيح مع ألف لام	الإسساء	خاطئ تماماً
عني	المعني	صحيح مع ألف لام	العناء	خاطئ أعطي المصدر
عفى	العافي	صحيح مع ألف لام	العفو	خاطئ أعطي المصدر
قلى	القائل	خاطئ تماماً	القول	خاطئ تماماً

ملاحظة: يلاحظ أن كافة الأجوبة عن إسم المفعول كانت خاطئة وذلك يشير إلى قلة الأمثلة التي تتوفر عن إسم المفعول للنظام بل ربما انعدامها وأعطيت بدلا من ذلك المصادر التي ربما تتوفر أمثلة عنها في التدريب. أما إسم الفاعل فقد كانت الأجوبة فكان أغلبها صحيحاً بها مع بعض الأخطاء وعدم التمييز بين قلى وقال وهذا ربما يشير إلى الترجمة من وإلى الإنكليزية.

٥- تجربة صيغ الجمع

السؤال: ما هي صيغ جمع الكلمات الآتية:

رجل امرأة فيل زاوية ميل معلم أرض كتاب ساعة وادي
الجواب : كما في الجدول (٦)

جدول (٦) صيغ الجمع

الكلمة	الجمع	ملاحظات
رجل	رجال	صحيح
إمرأة	نساء	يلاحظ النقاط المعنى
فيل	أفيال	لا بأس ويمكن أن يكون فيلة
زاوية	زوايا	صحيح
ميل	أميال	صحيح
معلم	معلمون	صحيح
أرض	أراضي	لا بأس ويمكن أن تكون أرضون
كتاب	كتب	صحيح
ساعة	ساعات	صحيح
وادي	أودية	لا بأس يمكن أن تكون وديان

ملاحظة: الأجوبة صحيحة

٦- تجربة البحث عن اشتقاقات لبعض الكلمات لمعرفة الحصيلة اللغوية التي يحتويها النظام

السؤال : ماذا يسمى أخو الأب وأخت الأب وأخو الإم وأخت الأم وأب الأب وإم الأب وابن الإبن وبنت البنت

الجواب : أخو الأب يُسمى "عَمّ"، وأخت الأب تُسمى "عَمَّة"، أما أخو الأم فيُسمى "خَال"، وأخت الأم تُسمى "خَالَة". أب الأب يُسمى "جَد"، وإم الأب تُسمى "جَدَّة". أما ابن الإبن فيُسمى "وَلَد الوَلَد"، وبنت البنت تُسمى "بنت البنت".

ملاحظة: لوحظ أن الإجابة لا بأس بها رغم خلوها من استعمال الأحفاد أو الأسباط ربما لندرة الأمثلة التي تدرب عليها النظام.

٧- تجربة الصفات الصرفية للأفعال

قمنا بمحاولة اختبار المعرفة الصرفية للنظام وذلك بإعطائه تسعة أفعال هي:

قال وهب تلى كرر هزأ سأل أكل وهي كوى

وكان المطلوب تسمية هذه الأفعال من الناحية الصرفية فكان الجواب سيئاً حيث لم يفهم النظام معنى الصرف أصلاً . ولغرض تجربة إمكانية التعلم لدى النظام فقد قمنا بتوضيحات متتالية. فأوضحنا للنظام تسمية بعض الأفعال مثل "قال" أنه يسمى فعل "أجوف" وأن "وهب" يقال عنه "مثال" وطلب منه معرفة تسمية بقية الأفعال ففشل في ذلك. واستمرت محاولة التعليم ست مرات ففشل في الوصول إلى إجابات صحيحة ، مما يدل على الضعف الشديد في البيانات الصرفية التي تدرب عليها النظام أو إنعدامها.

٨- تجربة تشكيل النصوص

المطلوب (١) تشكيل الآية: إنما يخشى الله من عباده العلماء

الجواب : إِنَّمَا يَخْشَى اللَّهَ مِنْ عِبَادِهِ الْعُلَمَاءُ

فكان التشكيل دقيقاً رغم تأخر الفاعل إلى نهاية الجملة وأنه تفصله عن الفعل ثلاث كلمات. ربما لأن الآية متوفرة بالتدريب بدقة.

تجربة تشكيل العبارات التي وردت في خلاصة هذا البحث فكان الجواب كالآتي:

"تَسْتَعْمَلُ أَنْظُمَةَ الذِّكَاةِ الاصْطِنَاعِيِ الْحَالِيَّةِ الْمَعْلُومَاتِ الْمُتَوَفَّرَةِ عَلَى شَبَكَةِ الْإِنْتَرْنِتِ وَالْكَتُبِ وَالصُّخْفِ وَالْوَثَائِقِ الْمُتَوَفَّرَةِ لِكَيْ تَسْتَخْلُصَ مِنْهَا حَوَارِزِمِيَّاتِ اسْتِقْبَالِ النُّصُوصِ اللَّغَوِيَّةِ وَتَحْلِيلِهَا وَفَهْمَ مَعْنَاهَا ثُمَّ تَقُومُ بِالْعَمَلِيَّاتِ الْمَطْلُوبَةِ مِنْهَا حَاسُوبِيًّا وَتَحْوِيلِ نَتَائِجِ تِلْكَ الْعَمَلِيَّاتِ إِلَى نُصُوصِ مَكْتُوبَةٍ أَوْ مُنْطُوقَةٍ أَوْ عَمَلِيَّاتِ صِنَاعِيَّةٍ تَسْتَحْدِمُ فِي الْإِنْسَانِ الْآلِي (الرُّبُوتِ) أَوْ الْآلَاتِ الْأُخْرَى أَوْ غَيْرِ ذَلِكَ مِنْ تَطْبِيقَاتِ. سَنَقُومُ فِي هَذَا الْبَحْثِ بِإِعْطَاءِ مُقَدِّمَةٍ عَنِ بَعْضِ مَا يَتَوَفَّرُ مِنْ بَرْمَجِيَّاتِ الذِّكَاةِ الاصْطِنَاعِيِ وَاسْتِخْدَامَاتِهَا وَفَحْصِ قَابِلِيَّاتِهَا لِلتَّعَامُلِ مَعَ اللُّغَةِ الْعَرَبِيَّةِ لِتَبْيِينِ الشُّوْطِ الِذِي قَطَعْتَهُ وَمَا تَشْكُو مِنْ نُقْصٍ".

ملاحظة: يلاحظ أن تشكيل الآخر كان معدومًا تقريبًا بالإضافة إلى أن تشكيل الحروف ضمن الكلمة كان خاطئًا بنسبة كبيرة (وضعت خطوط تحت الكلمات الخاطئة) ويشير ذلك إلى أن نسبة النصوص المشكولة في التدريب كانت قليلة. وهذا بالفعل هو الحال بالنسبة للنصوص المتوفرة على الشبكة. وذلك يستدعي العمل الحثيث على وضع نصوص ضخمة مشكولة على الشبكة.

٩- تجربة الصياغة العامة للجملة العربية نحويا وصرفيا

السؤال : هل الجملة الآتية صحيحة لغويا؟ إذا لم تكن كذلك هل يمكن تصحيحها؟:
"هناك بضعة معلمون لم يستلمون مرتباتهن بينهما ثلاثة أختين"

الجواب : الجملة التي قدمتها غير صحيحة لغويا. يمكن تصحيحها كالتالي:

"هناك بضعة معلمين لم يستلموا مرتباتهم، من بينهم اثنان من الإخوة".

ملاحظة: في الجملة المصححة، تم تصحيح عدة أمور منها استخدام صيغة صحيحة للفعل والضمائر المناسبة وتمييز حالة رفع جمع المذكر السالم. وذلك يعتبر إنجازاً جيداً

١٠- تجربة إعراب جمل

(١) جاء الحق وزهق الباطل إن الباطل كان زهوقاً

الجواب: "جاء الحق وزهق الباطل" هي جملة تتكون من جملتين اثنتين، وهما:

"جاء الحق": الفاعل هو "الحق" وهو في محل رفع فاعل.
"زهق الباطل إن الباطل كان زهوقاً": هنا يأتي الجملة الثانية مع استخدام "إن" للتأكيد، وتحتوي على جملة فعلية واسمية. "زهق الباطل"، الفعل "زهق" مضارع وهو ماضٍ عاقب وهو في محل جزم. "إن الباطل كان زهوقاً"، هنا الفاعل هو "الباطل" وهو في محل رفع فاعل، و"كان" فعل ماضٍ في محل نصب، و"زهوقاً" خبر "كان".

وبالتالي، تعني الجملة "جاء الحق واندثر الباطل، إن الباطل كان هالكاً".

(٢) وعد الله الذين آمنوا ليستخلفنهم في الأرض

الجواب: إليك إعراب الجملة "وعد الله الذين آمنوا ليستخلفنهم في الأرض":

- "وعد": اسم مجرور بالضممة.
- "الله": مضاف إليه ومجرور بالضممة.
- "الذين": اسم موصول مبني على الضمير المتصل في محل رفع مفعول به.
- "آمنوا": فعل ماضٍ مبني على الضمير المتصل في محل رفع فاعل.
- "ليستخلفنهم": فعل مضارع منصوب وعلامة نصبه الضمة الظاهرة، وهو متعلق بالضمير الذي يُشير إلى "الذين" ويعمل كمفعول به.
- "في": جار ومجرور بالكسرة.
- "الأرض": مضاف إليه ومجرور بالكسرة.

الجملة تعني: "وعد الله الذين آمنوا أن يجعلهم خلفاء في الأرض".

(٣) قول معروف ومغفرة خير من صدقة يتبعها أذى

الجواب : إليك إعراب الجملة "قول معروف ومغفرة خير من صدقة يتبعها أذى":

- "قول": مبتدأ مرفوع وعلامة رفعه الضمة الظاهرة.
- "معروف": خبر مقرر ومرفوع وعلامة رفعه الضمة الظاهرة.
- "و": عطف بين المبتدأ والخبر.
- "مغفرة": خبر مقرر ومرفوع وعلامة رفعه الضمة الظاهرة.
- "خير": نعت مرفوع وعلامة رفعه الضمة الظاهرة.
- "من": حرف جر وظرف زمان في محل جرّ.
- "صدقة": مضاف إليه مجرور وعلامة جرّه الكسرة.
- "يتبعها": فعل مضارع مرفوع وعلامة رفعه الضمة الظاهرة، وهو متعلق بـ"صدقة".
- "أذى": خبر مرفوع وعلامة رفعه الضمة الظاهرة.

الجملة تعني: "الكلمة الطيبة والمغفرة أفضل من الصدقة مع تبعها لأذى".

ملاحظة: يتضح من أجوبة الإعراب أعلاه الأخطاء الشديدة فيها لحد عدم التمييز بين الإسم والفعل في بعضها وذلك بسبب قلة النصوص المتوفرة أثناء التدريب. وذلك يستدعي العمل على تزويد الشابكة بكمية ضخمة من الجمل المتنوعة مع إعرابها لكي تستعمل للتدريب مستقبلاً.

١١ - فهم المعنى: إشرح معنى البيت:

ذو العقل يشقى في النعيم بعقله وأخو الجهالة في الشقاوة ينعم

الجواب : هذا البيت يعبر عن مفهومين متناقضين:

"ذو العقل يشقى في النعيم بعقله": هنا يشير البيت إلى أن الشخص الذي يمتلك الحكمة والعقل يمكن أن يعاني أو يشقى في الأوقات التي يكون فيها محاطاً بالرفاهية والنعيم. قد يكون ذلك بسبب وعيه العميق بالأمور، قد يكون أكثر حساسية للمشاكل أو الجوانب السلبية في الحياة، مما يجعله يشعر بالتوتر أو القلق رغم كل ما حوله من نعيم.

"وأخو الجهالة في الشقاوة ينعم": هذا الجزء يعني أن الشخص الجاهل، الذي يفتقر إلى الحكمة والعقل، قد يشعر بالراحة أو السعادة في الأوقات التي يكون فيها محاطاً بالمصاعب والشدائد. فهو قد لا يكون على علم كاف بتلك المصاعب، أو قد يفتقر للوعي الكافي ليدرك الجوانب السلبية في الأمور، مما يجعله ينعم في الظروف التي يبدو فيها الآخرون على أنها شقاء.

باختصار، البيت يعبر عن الفكرة العامة بأن الشخص الحكيم قد يعاني في الظروف الجيدة بينما الجاهل قد يجد السعادة في الأوقات الصعبة.

ملاحظة : الشرح لا بأس به ويمكن الاستنتاج أن النظام قد فهم المعنى العام لبيت الشعر وكان التعبير عن ذلك لا بأس به من الوضوح.

١٢ - تجربة فحص البلاغة

المطلوب شرح معنى البيت التالي (بدون أي تشكيل) ثم مع التشكيل (شدة الحزم) و (شدة الحزم)

كانهم في ظهور الخيل نبت ربي من شدة الحزم لا من شدة الحزم

الجواب : هذا البيت يعبر عن فكرة مقارنة بين اثنين يظهران بمظهر مشابه ولكن بوجهين مختلفين:

"كانهم في ظهور الخيل نبت ربي": البيت يقارن بين شخصين أو مجموعتين بقدرة تشابههم مع بعضهم البعض، حيث يُشبهون أو يظهران مثل نبتة واحدة. تشبيه ظهورهم بنمو النبات يُظهر كيف يظهران متشابهين بصورة واضحة.

"من شدة الحزم لا من شدة الحزم": هذا الجزء يوضح أن التشابه الذي يظهره ليس بسبب نشأتهم من جذور مشتركة أو صفات مماثلة، بل بسبب العوامل الظاهرية. يعني أنهم يبدو متشابهين في المظهر أو السلوك، لكن هذا التشابه ليس ناتجاً عن وجود صفات مشتركة بينهم، بل يعود إلى الوضع الخارجي أو الظروف السطحية.

باختصار، البيت يشير إلى التشابه الظاهري بين أشخاص معينين وكيف أن هذا التشابه ليس ناتجاً عن وجود صفات مشتركة بينهم، وإنما ناتج عن ظروف أو مظاهر خارجية تظهر الشبه بينهم.

ثم أعيد الطلب لمعنى البيت ولكن هذه المرة مع تشكيل كلمات (شدة الحزم) و (شدة الحزم)

كأنهم في ظهور الخيل نبت ربي من شدة الحزم لا من شدة الحزم

الجواب : البيت يعبر عن فكرة التشابه الظاهري والاختلاف الحقيقي بين أشخاص معينين أو مجموعات من الناس.

"كأنهم في ظهور الخيل نبت ربي"، يُستخدم هنا تشبيه ظهور الأشخاص بنمو النبات، مما يعطي

انطباعًا بالتشابه أو الوحدة الظاهرية بينهم، كما لو كانوا نبتة واحدة.

"من شدة الحزم لا من شدة الحزم"، هذا الجزء يوضح الاختلاف الحقيقي بين هؤلاء الأشخاص. "شدة الحزم" تشير إلى القسوة أو الصرامة، بينما "شدة الحزم" تُشير إلى التوتر أو الضغط. فالبيت يشير إلى أن التشابه الظاهري بين هؤلاء الأشخاص ليس بسبب الشدة الحقيقية، بل هو بسبب الظروف السطحية.

باختصار، البيت يشير إلى أن الأشخاص قد يظهرون متشابهين من الخارج، لكن الفروقات الحقيقية في شخصياتهم وسلوكهم تكمن في عوامل أكثر عمقًا ما وراء المظاهر الظاهرية.

الملاحظة: لم يستطع البرنامج إدراك المعنى الدقيق سواء كان البيت مشكولاً أم لا رغم أن المعنى عند التشكيل كان أفضل. ويستنتج من ذلك أن إدراك المعنى البلاغي العميق لا يزال بعيد المنال على أنظمة الذكاء الاصطناعي.

١٣ - تجربة : ماذا تعرف عن مجمع اللغة العربية الأردني

الجواب : المجمع اللغة العربية الأردني هو مؤسسة تعنى بقضايا اللغة العربية في الأردن. يتمثل دورها في الحفاظ على اللغة العربية وتطويرها ودعم استخدامها في مختلف المجالات. المجمع يشمل عادة فريقاً من اللغويين والباحثين في مجال اللغة العربية والأدب، ويعمل على تنفيذ مجموعة من الأنشطة والمبادرات لتحقيق أهدافه. هذه الأنشطة قد تشمل:

١. تطوير اللغة العربية: المجمع يسعى إلى تطوير اللغة العربية والمساهمة في إثرائها من خلال إصدار القواميس والمعاجم والمصطلحات العربية والأبحاث اللغوية.
٢. الدعم للمؤلفين والأدباء: يمكن أن يقدم المجمع دعماً للكُتّاب والأدباء الذين يكتبون باللغة العربية ويسعون إلى نشر أعمالهم.
٣. النشر والبحث: يمكن أن يكون للمجمع دور في نشر الأبحاث والدراسات اللغوية والأدبية، ويمكن أن يعزز البحث في ميدان اللغة العربية.
٤. تقديم استشارات: المجمع قد يقدم استشارات وتوجيه للجهات الرسمية والمؤسسات التعليمية بشأن قضايا اللغة العربية.
٥. التعاون مع المؤسسات الدولية والمنظمات ذات الصلة .

ملاحظة: يلاحظ أن الوصف أعلاه قد استخلص بتصريف من موقع المجمع على الشبكة ومن غيره وهذا يشير إلى أن قابلية النظام على عملية التلخيص مقبولة بشكل عام لكنها ليست دقيقة دائماً.

١٤ - قطعة أدبية : رجع طفلك البالغ من العمر ١٠ سنوات من المدرسة وهو يحمل

نتيجة علاماته السنوية التي كانت سيئة. المطلوب كتابة قطعة أدبية عما تقول له.

الجواب : في حينما عاد الطفل إلى المنزل وفي يديه نتائج دراسته السنوية، ظهرت ملامح التوتر على وجهه الصغير. كان يحمل عبئاً ثقيلاً من النقاط الضعيفة، وكان كنفه تحملان أكثر مما يناسبهما. رأيت في عينيه ليس فقط خيبة الأمل، ولكن جرعة كبيرة من القلق.

جلست معه بلطف وبدأت الكلمات تتساقط كمطر الربيع على أرض يابسة. "أولاً، أريدك أن تعلم أنني أهتم بك، بغض النظر عن النتائج التي تحملها يديك. أنا هنا لأفهم ، لذا هل يمكنك أن تشاركني مشاعرك؟

ومع كل كلمة ، ذكرته بأنه ليس وحده في هذا الطريق " إفهم أن الدراسة قد تكون تحدياً، فهل يوجد شيء معين أو تحدي يمكننا معالجته معاً؟"

تحدثت بلطف عن قوته وقدراته، وطرحت أسئلة تساعد على تحديد أهداف وخطط للتحسن. " هل يمكننا وضع خطة معاً؟ ربما هناك هدف صغير يمكن تحقيقه في وقت قصير"

وفي هذا الحديث الدافئ تأكدت من تذكيره بقيمة المثابرة والعمل الجاد، وأن النجاح يأتي مع الوقت والجهد. " الحياة تحمل تحديات ، ولكننا سنتعلم منها وننمو. أنت ذكي وقادر على التحسن"

ختمت الحديث بتأكيد الحب والدعم الدائم. " مهما كانت النتائج، أنا هنا لدعمك وأحبك كما أنت. الدرجات ليست مقياسًا لقيمتك، أنت لديك القدرة على النجاح والتفوق، وسنساعد بناء هذا المسار معًا"

ملاحظة : مع وجود قليل من الأخطاء الإملائية وبعض التعبيرات التي كان يمكن أن تكون أفضل مما هي عليه، فإن القطعة على العموم لا بأس بها من ناحية الأفكار والمحتوى والسلاسة ، لكن ربما يمكن ملاحظة أثر الترجمة من الإنكليزية على بعض المفاهيم التي ليست شائعة كثيرًا في اللغة العربية.

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 الاصطناعي.

وبالرغم من أنه يمكن ملاحظة العديد من المثالب وعدم الدقة فيما يتعلق بالتعامل مع اللغة العربية فهما 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 الاصطناعي. ما تحتاجه أنظمة الذكاء الاصطناعي فيما يخص اللغة العربية هو المزيد من توفير البيانات المحوسبة باللغة العربية لغرض التدريب والمعاجم اللغوية المحوسبة. وتشمل البيانات التي يجب توفيرها النصوص المشكولة والنصوص مع إعرابها نحويًا والمعلومات الصرفية . كما أن هناك ضرورة للتعلم في أبحاث المعالجة الحاسوبية للغة العربية بالتعاون بين اللغويين واللسانيين والحاسوبيين.

المصادر

١. محمد زكي خضر، مشكلات الترجمة الآلية، الموسم الثقافي التاسع والثلاثون لمجمع اللغة العربية الأردني ٦ - ١٩/١٢/٢٠٢١، منشورات مجمع اللغة العربية الأردني ١٤٤٣ هـ / ٢٠٢٢م صفحة ٢٥٥-٢٨٣ .
2. <https://translate.google.com/>
3. <https://it-solutions.center/tag/%D8%AA%D8%B9%D8%B1%D9%8A%D9%81-chat-gpt/2>.
4. <https://www.hpcwire.com/2023/05/10/googles-new-ai-focused-a3-supercomputer-has-26000-gpus/3>.
5. <https://ai.facebook.com/blog/large-language-model-llama-meta-ai/>
6. <https://www.albayan.ae/uae/news/2023-05-30-1.4671616>
7. <https://waleedkadous.github.io/ansari/>
8. Gheith A. Abandah _ Alex Graves _ Balkees Al-Shagoor _ Alaa Arabiyat _ Fuad Jamour _ Majid Al-Tae , Automatic diacritization of Arabic text using recurrent neural Networks, [International Journal on Document Analysis and Recognition \(IJDAR\)](#) June 2015, 18(2):183-197
٩. محمد طارق الكاتب، موازين الشعر العربي باستخدام الأرقام الثنائية، مطبعة المواني العراقية البصرة ١٣٩٢هـ/١٩٧١م
10. 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
11. <https://chat.openai.com>

Artificial Intelligence Systems and Arabic Language
Professor Mohammed Zeki Khedher
President of International Institute for Quran and Islamic Sciences
Computing
president@qurancomputing.org

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 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 provide

some highlights and proposals regarding the future steps necessary to improve the handling of that software with the Arabic language.

المصطلحات

المصطلح	المقابل بالإنكليزية
الذكاء الاصطناعي	Artificial intelligence
معالجة اللغة الطبيعية	Natural Language Processing – NLP
النظام الخبير	Expert System
التعلم العميق	Deep Learning
الشبكة العصبية الاصطناعية	Artificial Neural Network
نظام "جي بي تي"	GPT