مستقبل اللغة العربية في ظل الذكاء الاصطناعي ونماذج اللغة الكبيرة ...

تحسين الشيخلي

مداخلتي في الندوة الافتراضية التي نظمتها كلية الأداب واللغات في جامعة بجاية في الجزائر يوم أمس

شهدت العقود الأخيرة بروز (هندسة اللغة) كحقل متعدد التخصصات يجمع بين اللسانيات، علم الحوسبة، والذكاء الاصطناعي بهدف تطوير تقنيات تتيح للآلات فهم النصوص البشرية ومعالجتها بكفاءة.
في جوهرها، تهتم هندسة اللغة بتحليل بنية اللغات الطبيعية، وتصميم أنظمة قادرة على توليد المحتوى، وتحسين التفاعل بين الإنسان والآلة، مما يعزز من قدرة التطبيقات الذكية على معالجة النصوص المكتوبة والمنطوقة بشكل أكثر دقة وفاعلية. ومع تنامي حجم البيانات اللغوية المتاحة رقميًا، أصبح من الضروري تطوير أدوات تعتمد على الذكاء الاصطناعي لضمان تحقيق أقصى استفادة من هذه الموارد، سواء في الترجمة، التدقيق اللغوي، أو حتى إنتاج المحتوى التلقائي.
في موازاة ذلك، برز مفهوم (صناعة اللغة) ليشمل جميع العمليات التقنية والإبداعية التي تستهدف إنتاج المحتوى، تطوير الأنظمة اللغوية، وتحسين استخدامات اللغة في البيئات الرقمية. لم تعد اللغات مقتصرة على وسائل التواصل التقليدية، بل أصبحت جزءًا من الاقتصاد الرقمي، حيث تلعب دورًا محوريًا في التجارة الإلكترونية، الإعلام، وخدمات التعليم الذكي. في هذا الإطار، أصبحت الحاجة إلى تكامل الأدوات التقنية مع المهارات البشرية أكثر إلحاحًا لضمان استدامة الصناعات اللغوية في عصر التحولات الرقمية.
يشكل الذكاء الاصطناعي التوليدي أحد أهم التطورات في هندسة وصناعة اللغة، حيث يعتمد على نماذج تعلم عميق قادرة على توليد نصوص جديدة انطلاقًا من بيانات تدريبية ضخمة. هذه النماذج، المعروفة باسم نماذج اللغة الكبيرة(LLMs) ، مثلGPT-4 ، Gemini، وMistral، Deepseekلا تقتصر على فهم النصوص وتحليلها، بل تستطيع إنتاج محتوى مترابط بأساليب متنوعة، وتقديم ترجمات سياقية، بل وحتى محاكاة الأنماط الأسلوبية المختلفة. يعتمد عملها على الشبكات العصبية الاصطناعية التي تتعلم من مليارات الكلمات، مما يجعلها قادرة على التنبؤ بالكلمات التالية في الجملة، وفهم العلاقات السياقية، وإعادة صياغة المحتوى وفقًا للحاجة.
ومع أن هذه النماذج تشكل نقلة نوعية في التفاعل بين البشر والآلات اللغوية، فإن استخدامها يثير تساؤلات حول دقة المحتوى، الإنصاف اللغوي، وحماية الهوية الثقافية للغات المختلفة. في هذا السياق، تصبح هندسة اللغة وصناعتها مجالين حاسمين لضمان أن تكون التطورات التقنية خادمة للغة، لا مهيمنة عليها، وأن تسهم هذه الابتكارات في تعزيز التواصل البشري بدلاً من فرض معايير لغوية موحدة تفتقر إلى الخصوصية والتنوع.
اللغة العربية كحال جميع اللغات الطبيعية تشهد تحولات غير مسبوقة بفعل التطورات المتسارعة في مجال الذكاء الاصطناعي، وخاصة الذكاء الاصطناعي التوليدي ونماذج اللغة الكبيرة. لم تعد هذه التقنيات مجرد أدوات مساعدة، بل أصبحت عاملاً رئيسيًا يعيد تشكيل علاقتنا باللغة وصياغة مستقبل استخدامها في الترجمة، التوطين، وصناعة المحتوى. وبينما تحمل هذه التطورات وعودًا بتمكين العربية من التوسع رقميًا وتعزيز مكانتها في الفضاء الإلكتروني العالمي، فإنها في الوقت ذاته تفرض تحديات معقدة تتعلق بالدقة، الهوية الثقافية، وإمكانية الوصول. في هذا السياق، يصبح السؤال الأساسي:
كيف يمكننا الاستفادة من إمكانيات الذكاء الاصطناعي مع الحفاظ على أصالة اللغة العربية وضمان تطورها بطريقة تحترم خصوصيتها اللغوية؟
على الرغم من التقدم الملحوظ في تقنيات الذكاء الاصطناعي، لا تزال اللغة العربية تواجه عقبات كبيرة فيما يتعلق بالمعالجة الآلية. فبنيتها النحوية والصرفية المعقدة، وتعدد أشكال الكلمات وفقًا للإعراب والتصريف، تشكل تحديات كبيرة أمام الخوارزميات التي تم تطويرها في الأساس بناءً على اللغات اللاتينية

مثل الإنجليزية. معظم نماذج اللغة الكبيرة مثلGPT-4 ، Mistral، Gemini ، ، Claudeو Deepseek تجد صعوبة في فهم السياقات العربية الدقيقة، مما يؤدي أحيانًا إلى أخطاء في المعنى، أو تراكيب لغوية ركيكة، أو فقدان التشكيل الذي يؤثر على الدلالة اللغوية. كما أن العربية تعتمد على الاشتقاق والتجذير في تكوين الكلمات، وهي ميزة لا تزال خوارزميات الذكاء الاصطناعي تجد صعوبة في التعامل معها.
إحدى المشكلات الأساسية التي تواجه الذكاء الاصطناعي عند التعامل مع العربية هي نقص البيانات اللغوية المتنوعة. إذ تعتمد هذه النماذج على بيانات ضخمة يتم جمعها من الإنترنت، ولكن المحتوى العربي المتاح رقميًا لا يزال أقل بكثير من نظيره في اللغات الأخرى مثل الإنجليزية أو الصينية. هذا يؤدي إلى انحياز النماذج اللغوية تجاه أنماط لغوية معينة، مثل الفصحى الرسمية، بينما تبقى اللهجات العامية والمستويات الأسلوبية المختلفة خارج نطاق المعالجة الدقيقة. كما أن ندرة البيانات المهيكلة عالية الجودة تجعل من الصعب على الذكاء الاصطناعي تقديم ترجمات دقيقة أو إنتاج محتوى عربي سلس وطبيعي.
التحيز اللغوي يعد من القضايا التي تستدعي اهتمامًا خاصًا، حيث إن نماذج الذكاء الاصطناعي غالبًا ما تميل إلى إنتاج نصوص تتأثر بالبيانات التي دربت عليها. في حالة العربية، يؤدي هذا التحيز إلى اعتماد تراكيب لغوية غير طبيعية، أو إلى فقدان بعض الخصائص الأسلوبية التي تميز اللغة عن غيرها. كما أن بعض التطبيقات الذكية قد تقوم بفرض أساليب لغوية معينة على المحتوى العربي دون مراعاة التنوع الثقافي والجغرافي الذي تتميز به المنطقة العربية.
أحد أكثر المجالات تأثرًا بالذكاء الاصطناعي التوليدي هو الترجمة الآلية، حيث أصبحت الأدوات الذكية تقدم ترجمات أسرع وأكثر سلاسة مقارنة بالماضي. إلا أن جودة هذه الترجمات لا تزال غير ثابتة، حيث يمكن أن تعاني من مشاكل في فهم المعنى الدقيق أو في التعبير بأسلوب عربي طبيعي. قد تنتج هذه النماذج جملًا تبدو صحيحة لغويًا لكنها تفقد المعنى المقصود، أو تعاني من أخطاء تركيبية تجعل النص غير متماسك.
على الرغم من التحسن الكبير في نماذج الترجمة العصبية، إلا أن المترجمين البشريين لا يزالون يلعبون دورًا جوهريًا في مراجعة وضبط الترجمة، مما يعزز مفهوم (الإنسان في الحلقة) او كما يدعى بالانكليزية (Human-in-the-Loop - HITL)وهو مصطلح شائع في مجال الذكاء الاصطناعي، التعلم الآلي، وهندسة اللغة. يجمع بين الذكاء الاصطناعي والخبرة البشرية لتحقيق نتائج أكثر دقة واتساقًا.
يشير مفهوم (الإنسان في الحلقة)إلى نموذج تشغيلي وتفاعلي في الذكاء الاصطناعي يتم فيه دمج العنصر البشري في دورة عمل الأنظمة الذكية لضمان تحسين الأداء، رفع جودة النتائج، وتقليل الأخطاء التي قد تنتج عن المعالجة الآلية المستقلة. يعتمد هذا النموذج على التفاعل المستمر بين الذكاء الاصطناعي والخبرة البشرية، حيث يقوم الإنسان بمراقبة، تصحيح، أو توجيه مخرجات الذكاء الاصطناعي لضمان دقة المحتوى، تحسين الترجمة، وضبط السياق الثقافي واللغوي في التطبيقات اللغوية.و يحقيق التوازن بين الأتمتة والدقة البشرية حيث يوفر الذكاء الاصطناعي كفاءة وسرعة في معالجة المهام، بينما يضمن التدخل البشري أن تكون النتائج متوافقة مع المعايير اللغوية، الإبداعية، والأخلاقية. حيث يساهم التدخل البشري في جعل المحتوى أكثر طبيعية، مرونة، وأقرب إلى الأسلوب البشري الحقيقي.
فيما يتعلق بالتوطين، تواجه اللغة العربية تحديات إضافية بسبب التنوع الجغرافي والثقافي. فالمحتوى الذي يناسب الجمهور الخليجي قد لا يكون ملائمًا للجمهور في شمال إفريقيا، والعكس صحيح. لهذا السبب، فإن عمليات التوطين يجب أن تأخذ في الاعتبار الاختلافات الثقافية واللهجية لضمان تقديم محتوى مناسب لكل منطقة. في هذا السياق، فإن الذكاء الاصطناعي قد يساعد في تحليل الفروق بين اللهجات والتكيف معها، لكنه لا يزال يحتاج إلى تطويرات أكثر تقدمًا ليكون قادرًا على إنتاج محتوى دقيق وسياقي.
من الجوانب الإيجابية لاستخدام الذكاء الاصطناعي في صناعة اللغة العربية هو تحسين إمكانية الوصول والشمولية، حيث باتت التقنيات الذكية توفر أدوات لمساعدة الفئات التي كانت تواجه صعوبة في التفاعل مع المحتوى العربي، مثل، تحويل النصوص إلى صوت باستخدام تقنيات الذكاء الاصطناعي، مما يسهل على المكفوفين وضعاف البصر الوصول إلى المعلومات.أضافة الى تحسين تقنيات الترجمة الفورية والدبلجة الآلية، مما يساعد في كسر الحواجز اللغوية أمام الناطقين بلغات أخرى.وقدرته على تحليل النصوص لضمان سهولة القراءة والفهم، مما يجعل المحتوى أكثر ملاءمة للمتعلمين الجدد أو الأشخاص ذوي الإعاقة الإدراكية.

على الرغم من هذه الفوائد، فإن الحاجة إلى تحسين دقة هذه التقنيات لا تزال قائمة، خاصة فيما يتعلق بتمييز الأصوات المختلفة في اللغة العربية وتحليل الجمل المركبة بشكل أكثر دقة.
مع ازدياد استخدام الذكاء الاصطناعي في صناعة المحتوى والترجمة، تتزايد الحاجة إلى تشريعات واضحة تنظم استخدامه في اللغة العربية. في هذا السياق، بدأت بعض الجهات التنظيمية في وضع إرشادات لحوكمة الذكاء الاصطناعي لضمان عدم إساءة استخدام هذه التقنيات في نشر معلومات مغلوطة أو إنتاج محتوى غير دقيق. كما أن قوانين مثل قانون الذكاء الاصطناعي في الاتحاد الأوروبي ستؤثر على كيفية استخدام نماذج اللغة الكبيرة في معالجة العربية، مما قد يفرض معايير جديدة تتعلق بالشفافية، مكافحة التحيز اللغوي، وضمان استخدام التكنولوجيا بطرق عادلة ومنصفة.
إن مستقبل العربية في العصر الرقمي يعتمد على مدى قدرتنا على الاستفادة من الذكاء الاصطناعي بطريقة تحافظ على أصالة اللغة، دون أن نفقد خصوصيتها اللغوية والثقافية. فالرهان ليس على استبدال الإنسان بالآلة، بل على تحقيق تكامل ذكي بين القدرات البشرية والإمكانات التقنية لضمان استمرار العربية كلغة حية ومتطورة.
هناك حاجة ماسة إلى استثمارات بحثية ومبادرات أكاديمية تدعم تطوير نماذج ذكاء اصطناعي متخصصة في اللغة العربية، قادرة على التعامل مع تعقيداتها وفهم سياقاتها المختلفة. كما أن الشركات العربية والمؤسسات اللغوية يجب أن تلعب دورًا أكبر في إثراء المحتوى الرقمي العربي، والمساهمة في بناء منصات ذكاء اصطناعي تعتمد على بيانات لغوية دقيقة وشاملة.
علينا ان ندرك إن التحدي الحقيقي ليس في تبني التكنولوجيا، بل في كيفية توظيفها لخدمة اللغة العربية بطريقة مستدامة. فالذكاء الاصطناعي قد يكون فرصة ذهبية لتعزيز العربية وانتشارها عالميًا، لكنه قد يكون أيضًا خطرًا يهدد هويتها إذا لم يتم التعامل معه بحكمة ووعي. المستقبل بين أيدينا، ويجب أن نقرر كيف نريد أن تتطور العربية في العصر الرقمي.