# مؤتمر "الذكاء الاصطناعيّ وخصائص اللغة العربيّة"

**الدوحة، 22 و23 /2025/12**

## الورقة المرجعيّة للمؤتمر

**أوّلا- الإطار العام للموضوع:**

شهد العالم في العقود الأخيرة ثورة تقنية غير مسبوقة، مدفوعةً بالتطور السريع في مجالات الحوسبة وتكنولوجيا المعلومات، والتي بلغت ذروتها مع ظهور الذكاء الاصطناعي التوليدي (Generative AI). وقد شكّل الذكاء الاصطناعي في عمومه نقلة نوعية في التعامل مع اللغات الطبيعية، إنتاجا وفهما وتحليلا وترجمة؛ حيث ركزت التقنيات عبر مسارات تطورها على جوانب لغوية مختلفة شملت الترجمة الآلية، والتعرف على الكلام، وتحليل النصوص، وتوليد المحتوى، إلخ. ومع ظهور خوارزميات التعلم العميق (Deep Learning) وتطور الشبكات العصبية الاصطناعية (Neural Networks)، برزت قدرة هذه التقنيات على معالجة اللغات بشكل أكثر كفاءة مما سبق. ولم يكن هذا التطور وليد اللحظة، بل كان تتويجا لمسيرة طويلة من البحث والتطوير بدأت جذورها في منتصف القرن العشرين، وتجلت ملامحها مع نشر آلان تورينج Alan Turing مقالته الشهيرة (Computing Machinery and Intelligence) "الآلات الحاسبة والذكاء"، والتي قدم فيها "اختبار تورينج" (Turing Test) معيارًا لتحديد ما إذا كان يمكن اعتبار

الآلة ذكية بناء على قدرتها على محاكاة المحادثة البشرية. وهذا الاختبار لم يضع فقط معيارًا للذكاء الاصطناعي، بل سلط الضوء أيضًا على أهمية اللغة باعتبارها واجهة للتفاعل بين الإنسان والآلة.

وبشكل عام، ترتبط اللغة والتقنية بعلاقة وثيقة ومتبادلة، فمنذ اختراع الكتابة مرورًا بابتكار الطباعة وحتى الوصول إلى العصر الرقمي، كانت اللغة حاضرة بقوة أداةً وموضوعًا للتطور التقني. وفي عصر الذكاء الاصطناعي، برزت تقنيات معالجة اللغات الطبيعية (Natural Language Processing) فرعًا أساسيًّا وأصيلًا من فروع الذكاء الاصطناعي يركز على تمكين الحاسوب من فهم اللغة والتعاطي معها، مما أدى إلى ظهور النماذج اللغوية التوليدية ( Generative Language Models) التي تعتمد على بيانات لغوية ضخمة لفهم النصوص البشرية وتوليدها. وقد سلك التطور في مجال الذكاء الاصطناعي اللغوي مسارات متعددة، بدايةً من النماذج الرمزية القديمة إلى النماذج الإحصائية، وصولًا إلى النماذج القائمة على التعلم العميق مثل GPT وغيره.

وقد استفادت اللغة الإنكليزية وعدد من اللغات الكبرى من هذا التطور، مدعومة بمواردها الرقمية، ومعاجمها التاريخية، وتوصيفاتها اللغوية المعيارية، ومراكزها البحثية المتخصصة. وقد نتج عن ذلك ما نراه اليوم من تطور كبير ووفرة في الأدوات والموارد المتخصصة التي شكلت رافعة قوية ومُسرِّعا فعّالا في بناء النماذج اللغوية الضخمة مثل (ChatGPT وGemini وClaude، وغيرها) وتعزيز كفاءتها؛ حيث يبدو الفرق جليا من حيث الكفاءة بين تعاطيها مع اللغة العربية وهذه اللغات، ليس فقط على مستوى الفهم والتحليل والتوليد والترجمة، وإنما أيضا على مستوى الوعي الثقافي والتحيز الفكري.

لقد أحدَث الذكاء الاصطناعي والنماذج اللغوية الضخمة بذلك تحولًا في إنتاج المعرفة وتوزيعها، ويعيدان تشكيل المشهد المعرفي على المستوى العالمي. وتأتي هذه التطورات مشفوعة بتحيزات لغوية ومعرفية وثقافية كامنة، تَظهر بوضوح خاص في اللغات والثقافات ذات التمثيل الرقمي المحدود، ومنها اللغة العربية التي تواجه تحدّيات عدّة. وتشمل هذه التحديات صعوبات التجزيء (Tokenization)، وتعقيد البنى اللغوية، ونقص البيانات، ومحدودية الموارد الحاسوبية.

ونتيجة لذلك، تواجه الثقافة العربية خطر أن تصبح ضعيفة التمثيل أو أن تُصوّر بصورة غير دقيقة، مما يزيد من التحيزات الثقافية وعدم المساواة في منظومات المعرفة الرقمية. والتصدي لهذه التحديات لا يقتصر على الجوانب التقنية فحسب، بل يبدأ بتمثيل الخصائص اللغوية ويرتبط ارتباطًا عميقًا بالحفاظ على الثقافة، وتعزيز الهوية، وينتهي بالتنمية الاجتماعية والاقتصادية.

والواقع أن للغة العربية في هذا السياق أهمية خاصة؛ فهي لغة عريقة، تمتلك تاريخا غنيًّا وحضارة ثرية ممتدة لأكثر من عشرين قرنا من الزمن؛ وكانت دائما الوعاء الذي نقل وحفظ المعارف الإنسانية في مختلف ميادين العلوم والفلسفة والأدب عبر العصور. ورغم هذه المكانة الحضارية والتاريخية، فإن نصيب اللغة العربية من التطورات التقنية الراهنة، وخاصة في مجال الذكاء الاصطناعي، لا يزال محدودا نسبيا مقارنة باللغات العالمية الكبرى. ويُعزى ذلك إلى عدة عوامل، من أبرزها ندرة الموارد اللغوية الرقمية عالية الجودة، وتنوع المستويات اللغوية المنطوقة والمكتوبة (الفصحى، المعاصرة، واللهجات المحلية)، والمميزات النحوية والصرفية الفريدة، إضافة إلى ضعف الاستثمارات المالية والبشرية في مجال المعالجة اللغوية العربية الآلية بشكل عام.

وانطلاقًا من ذلك، يبدو مشروعًا أن يتولى هذا المؤتمر دعوة الباحثين للإسهام في معالجة الأسئلة الآتية:

1- ما الذي جعل اللغة الإنكليزية -ومثيلاتها- ذات نجاعة عالية في استخدامات الذكاء الاصطناعي؟
2- ما الأسباب التي جعلت لغات أخرى، ومنها اللغة العربية، أقل نجاعة في استخدامات الذكاء الاصطناعي؟
3- كيف تعاملت اللغات المماثلة للغة العربية مع ضعف النجاعة في استخدامات الذكاء الاصطناعي؟
4- هل حقًّا لخصائص اللغة العربية، واللغات المماثلة لها، دورٌ في ضعف النجاعة تلك؟
5- كيف يمكن اللحاق بمستوى اللغة الإنكليزية، ومثيلاتها، في استخدامات الذكاء الاصطناعي؟

لا شكّ أنّ الاهتمام يتزايد في عالمنا العربي والجهود تتضاعف لتجاوز العقبات وتعزيز مكانة اللغة العربية في الفضاء الرقمي وإنتاج الموارد اللغوية اللازمة لذلك. وقد برزت مؤخرا مساهمات

عربية مهمة في مجال إنتاج النماذج اللغوية الكبرى مثل نموذج "فالكون 40ب" (Falcon 40B) في الإمارات العربية المتحدة، ونموذج "علام" في المملكة العربية السعودية، ونموذج "فنار" في قطر، ونموذج "دلة (DALLA) " الذي يسعى المركز العربي للأبحاث ودراسة السياسات بالدوحة لتطويره، مستثمرًا البيانات التي يوفّرها مشروعه الرائد "معجم الدوحة التاريخي للغة العربية" ومدونته اللغوية المهيكلة والموثقة والمؤرخة.

وعلى الرغم من أهمية هذه المبادرات العربية إلا أنها لا زالت تواجه تحديات يقتضي تجاوزها جهودًا بحثية وتطويرية مكثفة وشراكات قوية مع إيمان راسخ بأن الخصائص الفريدة للغة العربية، كثرائها الاشتقاقي ومرونتها البلاغية، تقدم فرصًا واعدة لتطوير نماذج ذكاء اصطناعي أكثر دقة وفعالية. ولعلّه من الضروري اعتماد نهج موحد بين الأطراف المعنية في العالم العربي (الدول والمؤسسات) لبناء قواعد بيانات شاملة ومستنيرة ثقافيًا. وسيكون الحوار المركّز والتعاون الاستراتيجي ضروريين لحماية اللغة والثقافة العربية في العصر الرقمي، وضمان المشاركة العادلة في صياغة المستقبل المدفوع بالذكاء الاصطناعي.

وفي هذا السياق، يسعى هذا المؤتمر إلى استكشاف العلاقة المركبة بين خصائص العربية والنماذج اللغوية الاصطناعية، سواء من حيث التحديات أو الإمكانات، في أفق بناء منظومة معرفية عربية متكاملة في مجال الذكاء الاصطناعي اللغوي. وهو يهدف إلى أن يكون منصة لتبادل الخبرات والأفكار بين الباحثين والمطورين والمهتمين، لتعميق البحث في الموضوع، والخروج باقتراحات وتوصيات تسهم في تسريع وتيرة التقدم في هذا المجال الحيوي.

**ثانيًا: أهداف المؤتمر**

1. الوقوف على أثر الخصائص اللسانية للغة العربية في المعالجة الآلية واستخدامات الذكاء الاصطناعي.
2. تحليل تحديات النمذجة اللغوية للعربية في ظل الذكاء الاصطناعي التوليدي.
3. تقييم تجارب ومبادرات معاصرة في بناء موارد لغوية تخدم تطور الذكاء الاصطناعي العربي.
4. تحفيز التعاون بين اللسانيين وعلماء الحاسوب لبناء نماذج معرفية عربية فعالة.

5. رسم معالم خارطة طريق لتطوير تقنيات الذكاء الاصطناعي المخصصة للغة العربية.

**ثالثًا: محاور المؤتمر المقترحة**

1. الخصائص اللسانية العربية وعلاقتها بالنمذجة الآلية.
2. الموارد والبيانات اللغوية في خدمة الذكاء الاصطناعي.
3. المحاولات العربية في بناء النماذج اللغوية المتخصصة.
4. السياقات الثقافية والمعرفية في نماذج الذكاء الاصطناعي العربية.
5. بناء السياسات الرقمية وترشيد الكلفة في معالجة اللغة العربية.

**رابعًا: الجهات المستهدفة**

- اللسانيون والباحثون في علوم اللغة العربية.
- المتخصصون في المعالجة الآلية للغات الطبيعية.
- مطورو النماذج اللغوية العربية.
- صناع السياسات في مجالات اللغة والذكاء الاصطناعي.
- الطلبة والمهتمون بالتقاطع بين اللغة والتكنولوجيا.

**خامسًا- شروط المشاركة**

- لغة المؤتمر هي العربية والإنكليزية.
- تُعدّ الأبحاث المقدمة للمؤتمر ملتزمةً بمواصفات البحث الشكلية والموضوعية المعتمدة لدى المركز العربي.
- تخضع البحوث المنجزة كافةً للتحكيم المسبق، بمساعدة لجنة علمية مختصة.
- يُتوقّع أن تُنشر أعمال المؤتمر في كتاب باللغتين العربية والإنكليزية يصدر عن المركز العربي.
- تتولّى الجهة المنظمة تغطية نفقات السفر والإقامة للباحثين المشاركين في المؤتمر.

**سادسًا- مواعيد مهمة**

- تقديم اقتراح أولي يتضمن الرغبة في المشاركة ببحث في المؤتمر مع ذكر عنوان البحث والمحور الذي يندرج تحته.
- إرسال البحث كاملا وفقًا لشروط الكتابة في أجل أقصاه: 2025/09/30.
- تحكيم الأبحاث والردّ على المشاركين في أجل أقصاه: 2025/10/30.
- عقد المؤتمر في الدوحة يومي 22 و2025/12/23.

**سابعًا- التواصل**

- للتواصل مع اللجنة التنظيمية للمؤتمر، المرجو استخدام البريد الإلكتروني الآتي:

DAHDArabicAI@dohainstitute.org