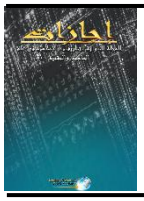


نحو تخصيص لنماذج اللغة والذكاء الاصطناعي لخدمة القرآن الكريم والعلوم الإسلامية

محمد محمد خير

المدير التنفيذي، المعهد العالمي لحوسبة القرآن والعلوم الإسلامية

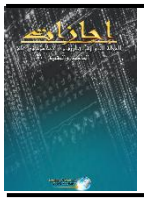
mohammad.khair@qurancomputing.org

الملخّص: تسخير نماذج اللغة لخدمة علوم القرآن الكريم والحديث والعلوم الإسلامية يتطلّب التحكّم بها للإلتزام بالواقعية والحقائق والمعلومات الموثّقة وتفادي ما يعرف بالشطط والهلوسة والخيال. ما هي الإستراتيجيات التي يمكن تطبيقها لضبط المحتوى المعلوماتي للإجابة من نماذج اللغة على أسئلة المستخدم للتعامل مع النص القرآني ونص الحديث الشريف بالأخص بدقّة وحرص على الصحّة، وكذلك توفير الحقائق والمراجع للعلوم الإسلامية.

الكلمات الجوهرية: نماذج اللغة، الذكاء الإصطناعي، الحوسبة المتوازية، القرآن الكريم، الحديث الشريف، استرجاع البيانات المكمّلة.

1. المقدّمة:

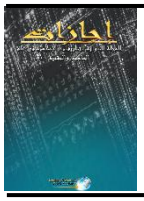
تقوم تطبيقات الذكاء الاصطناعي والتعلّم العميق لنماذج اللغة مثل ChatGPT, GPT4, Bard, Claude2, LLaMA2, Falcon وغيرها بتوفير خدمات ذات قيمة عالية استنادا على قدرة نماذج اللغة على التعامل باللغة الطبيعية مع المستخدم وفهم الأسئلة وتوفير أجوبة ذات علاقات موضوعية ومنطقية، مما يجعلها أداة مفيدة جدا لتطوير الكثير من التطبيقات الحوسبية مثل مساعدة المستخدم في البحث المعلوماتي والتلخيص وحتى التأليف أو التحرير. ولكن من المهم التمييز أن نموذج اللغة هو عبارة عن نموذج إحصائي يُمثل بالاحتمالات للربط (غير المباشر) بين مكونات وعناصر اللغة. ويتم تكوين هذه العلاقات الإحصائية من خلال عملية تدريب للنموذج على الكلمات والجمل كمدخلات فيستنبط العلاقات بينها، وذلك عن طريق إنقاص نسبة الخطأ في التنبؤ بالنتيجة الصحيحة. وبفضل هذه الاحتمالات تكون عملية الإصدار التوليدي المنتج للكلمات والجمل عند تكملة الجملة أو عند الإجابة عن سؤال المستخدم ذات تجدد في محتواها في كل مرّة تقدّم فيها نفس الجملة للتكملة، أو يسأل نفس السؤال للإجابة.



فماذا عن تطبيق نماذج اللغة في الاستفسار عن القرآن الكريم والحديث الشريف؟ ليس من المقبول أن تنتج الآيات الكريمة بشكل توليدي يعتمد على الاحتمالات، يتغير به الناتج من الكلمات في كل مرّة تصدر بها الإجابة، فتتبدّل الكلمة مكان الكلمة وتنقص كلمات أو تضاف كلمات أخرى. وكذلك بالنسبة للحديث الشريف، عن المغيرة بن شعبة رضي الله عنه أن النبي صلى الله عليه وسلم قال: (إن كذباً عليَّ ليس ككذبٍ على أحد، من كذّب علي متعمداً فليتبوأ مقعده من النار) رواه البخاري، ورواه مسلم في مقدمة صحيحه. لذلك وجب أن يكون النص المقدّم من نماذج اللغة للمستفسر عن آيات القرآن الكريم أو الحديث الشريف مطابق 100% للنص الأصلي بدون أي تغيير أو تبديل كما لا يجب أن يكون عرضة للاحتمالات التوليدية للنص، كما هو الحال في النصوص الأخرى الناتجة عن نماذج اللغة.

يقوم المعهد العالمي لحوسبة القرآن والعلوم الإسلامية QuranComputing.org حاليا بدراسة شاملة تضم التخطيط والتطوير لحلول لهذه المشاكل التي تتطلب فهما عميقا لتقنية نماذج اللغة، ومشاكلها، ونوعية الحلول المتوفرة وفاعليتها، وكيفية تطبيق الحلول المختارة عمليا، وتكلفة إنتاجها، وكيفية توزيعها وتوفيرها للمستخدم العام. وهناك عدة مشاريع مختلفة ناتجة عن هذا الجهد، وبعضها يتشارك في عوامل عدة، وهناك عدّة أهداف من هذه المشاريع، أهمها:

1- توفير قواعد معلوماتية سحابية (Cloud databases) ذات بيانات مضمنة (Data Embeddings) عن القرآن الكريم وترجماته وتفسيره واستخدامها في تخصيص نماذج اللغة بالقرآن الكريم. وسيتم التخصيص إما عن طريق استدعاء وظيفي لبرامج مساعدة (Plugin Function Calls) أو عن طريق استرجاع البيانات المكمّلة (Data Retrieval Augmented Generation – RAG) مما يضمن الدقّة في النص القرآن والحديث عند عملية الإنتاج من نموذج اللغة. [1,2,3]
2- توفير قواعد معلوماتية سحابية (Cloud databases) ذات بيانات مضمّنة (Data Embeddings) عن الصحيح من الحديث الشريف وترجماته واستخدامها في تخصيص نماذج اللغة بالحديث الشريف. وسيتم التخصيص كما هو مبيّن بالأسفل إما عن طريق استدعاء وظيفي لبرامج مساعدة (Plugin Function Calls) أو عن طريق استرجاع البيانات المكمّلة (Data Retrieval Augmented Generation – RAG) لعملية الإنتاج من نموذج اللغة.



3- توفير القاعدة المعلوماتية في العلوم الإسلامية باللغة العربية المطوّرة بصيغة JSON لتدريب نماذج اللغة. وستعتمد هذه القاعدة على معلومات من **المكتبة الشاملة** والنصوص المفتوحة المصدر لأكثر من 30000 كتاب.

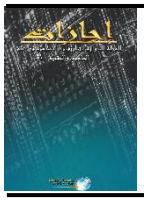
4- اختيار وتخصيص نموذج لغة مفتوح المصدر بالعلوم الإسلامية: ويهدف هذا إلى تخصيص نموذج لغة Fine-tuning LLM وذلك لتخفيف التكلفة من استخدام النماذج الربحية غير المفتوحة المصدر عند عرض الخدمات على المستوى العالمي. ويحتاج نموذج اللغة المختار أن يتكلّم العربية أو أن يكون قابلا للتدريب عليها.

5- تدريب نموذج مفتوح المصدر على الترجمة من العربية وإليها لعدة لغات أهمها الإنجليزية والفرنسية والإسبانية. بالنيّة استخدام هذا النموذج في ترجمة **المكتبة الشاملة** إلى جانب ترجمة أي مصادر إسلامية عربية أخرى إلى تلك اللغات. ومن ثم إعادة تخصيص نموذج اللغة مفتوح المصدر مرّة أخرى على العلوم الإسلامية المترجمة، إذا كان هناك فائدة واضحة من ذلك. من المفترض أنه عند تدريب نماذج اللغة تقوم عملية التدريب بتقريب القيم الرقمية المختلفة للكلمات من نفس المعنى بنفس التجمعات (Clusters) والمتجهات الشعاعية (Vector Directions) لتحصل على ترادف بالمعنى بين اللغات المختلفة، وهذا يؤدي إلى اكتساب مهارة الترجمة تلقائيا.

ويهدف المعهد إلى توفير **المصادر المفتوحة** لكل من البرامج والتطبيقات والأدوات للمعالجة اللغوية والقواعد المعلوماتية ونماذج اللغة ذاتها وأيضا أي وصلات إضافية لها.

2. التدريب السريع والفعّال لنموذج اللغة وتخفيف التكلفة:

لتسهيل وتسريع عملية التدريب التخصصي Fine-tuning لنموذج اللغة، من أسرع الأنظمة وأثرها فاعلية هو نظام (QLoRA Quantized Low Rank Adapters Fine-tuning) الذي يمكّن تسريع عملية التدريب 50 ضعفًا تقريبًا من خلال التدريب الفعّال للأوزان Parameter Efficient Fine Tuning PEFT ، والذي يعمل على المبادئ التالية:



- تخصيص الأوزان لشبكة التعلّم العميق عن طريق محوّل منخفض الرتبة LoRA – **Low Rank Adapter**
- نخصّص فقط جزءًا من الأوزان للشبكة في بعض الطبقات لنموذج اللغة التي تعتبر مهمّة للنتائج بدون الحاجة لإعادة تدريب جميع الأوزان الشبكة العميقة.
- يمكن تحميل وتخصيص الأوزان للشبكة بصيغة منخفضة الكم-4 (Quantization) bit-16 or bit-8 or bit-4بدلا من bit-32، بذلك نخفف المتطلبات على حجم الذاكرة المطلوبة للحوسبة ونزيد من سرعة الحوسبة لتجديد القيم لهذه الأوزان ، مما يخفف أيضا التكلفة للمعدات الحوسبية. وقد تبيّن من الأبحاث أن دقّة نماذج اللغة على الإجابة الصحيحة لا تتأثر كثيرًا عند تخفيف دقّة الأوزان للشبكة لمستوى bit-4 بدلا من bit-16 أو bit-32.

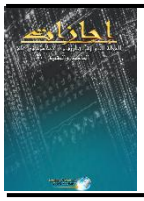
ومن أهم الأدوات المفتوحة المصدر التي تطبّق نظام PEFT QLoRA لتخصيص تدريب نموذج اللغة مع سرعة الأداء نذكر:

- **برنامج Ludwig من شركة Uber.com**

https://github.com/ludwig-ai/ludwig
https://ludwig.ai/latest/examples/nlu

- **برامج من Huggingface.co وهو أكبر موقع يختص بالبرامج والنماذج للذكاء الإصطناعي مفتوحة المصدر**

https://huggingface.co/blog/llama2#fine-tuning-with-peft
https://github.com/oobabooga/text-generation-webui
https://www.youtube.com/watch?v=lCJbO8ERZuU



3. تضمين البيانات النصية والترميز للغة العربية

تُشتق الخصائص والأنماط اللغوية من العلاقات المبنية بين المعلومات المقدّمة لنموذج التعلّم العميق مما يسمّى تضمين البيانات النصية Text Embeddings الذي ينتج البناء الرقمي للمعلومات المدخلة. وعملية الترميز Tokenization تحوّل الكلمات أو الأجزاء للكلمات المركبة إلى قيم رقمية تستعمل في حوسبة العلاقات بينها لحساب قيمة التشابه بالمعنى عند البحث عن كلمة أو جملة. ويجب مراعاة قواعد الصرف باللغة العربية عند تقطيع الكلمات والترميز لأن كل نموذج لغة يحتوي على محرك الترميز الذي استعمل لبنائه. مثلا نماذج اللغة AraBERT و CAMeLBERT المتوفرة على Huggingface.co يحتويان على محرّك ترميز (Tokenizer) ويمكن استعمالهما للغة العربية. وتحفظ جميع الرموز الرقمية للنص Text Embeddings في قواعد بيانات Vector Databases تحتوي على معلومات مختلفة التنوّع مثل النص والصوت والصور والفيديو، ومن أهمها Chroma, Pinecone, DeepLake, Faiss.

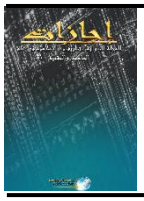
لتحقيق ضمان الدقّة للإجابة من نماذج اللغة بما يتعلّق في نص القرآن الكريم والحديث الشريف، هناك نوعان من الحلول يمكن تطبيقهما:

1- **استدعاء وظيفة البرنامج المساعد Plugin Function Calls لنماذج اللغة التجارية (غير مفتوحة المصدر)، مثل: OpenAI's ChatGPT & GPT4, Google's Bard (Palm), Anthropic's Claude2**

نتبع استراتيجية استدعاء وظيفة البرنامج المساعد Plugin Function Calls بحيث نطلب من نموذج اللغة أن يفعّل البرنامج المساعد الذي نوفره له لاستخراج النص القرآني أو نص الحديث من قاعدة بيانات موثّقة متوفرة على السحابة. وهذا التفعيل للبرنامج المساعد يحدث عند التشابه بالمعنى بين نص جملة البحث من المستخدم والنص الذي يشرح نوع فاعلية البرنامج المساعد. وللمزيد من المعلومات عن طبيعة عمل البرنامج المساعد لنماذج اللغة نرجو زيارة الرابط التالي:

https://platform.openai.com/docs/guides/gpt/function-calling

https://platform.openai.com/docs/plugins/review



ومن البرامج مفتوحة المصدر الحالية التي تستخدم استدعاء وظيفة البرنامج المساعد لل-GPT 4 Plugin برنامج Sahih AI للحديث الشريف مفتوح المصدر للأستاذ محمد الدليمي Muhammed Aldulaimi:

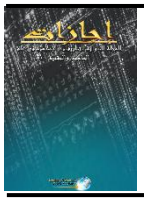
https://github.com/that-one-arab/sahih-ai

وأيضا برنامج Hadith Advice المساعد لل GPT-4 Plugin من الأستاذ فرديم منير Fardeem Munir.

2- استرجاع البيانات المكمّلة Retrieval Augmented Generation (RAG) للتحكّم بنماذج اللغة (غير مفتوحة المصدر) من خلال واجهة برمجة التطبيق Application Programming Interface (API)

إن القدرة على التحكّم بنماذج اللغة مغلقة المصدر مثل GPT-4 و ChatGPT من خلال API خاص تكمن في استخدام عدة خصائص له:

- هندسة التوجيه (Prompt Engineering) حيث يقوم المستخدم بتحديد نوع شخصية نموذج اللغة من خلال موجّه النظام (System Prompt)، مثلا يعرّف من خلالها لنموذج اللغة بأن الإجابات يجب أن يستدل عليها من المذاهب السنّية الإسلامية والحديث الشريف.
- استخدام طريقة استرجاع البيانات المكمّلة (Data Retrieval Augmented Generation - RAG) لتوفير المعلومات لنموذج اللغة التي يحتاجها للبحث وتحضير الإجابة للمستخدم، وهذا يجعل نموذج اللغة يتفادى التوليد الخلاّق للمعلومات بشكل غير صحيح أو لا أساس له عند فقدانه للمعلومات المطلوبة في بياناته التي درّب عليها بالأصل. والبيانات المسترجعة تؤخذ من قاعدة بيانات سحابية موثّقة بناءا على التشابه بالمعنى بين استعلام المستخدم ومعلومات قاعدة البيانات ومن ثم توفر تلك البيانات لنموذج اللغة لصياغتها للإجابة على المستخدم.
- للتحكم بنموذج اللغة لتفادي الهلوسة والشطط ولغرض الإلتزام بالواقعية والحقائق يجب تحديد قيمة المتغيّر الذي يُعرَف بالحراره لنموذج اللغة بالقيمة صفر، أي temperature = 0 عند الإستعلام من النموذج بسؤال ما. وهذا المتغير يأخذ قيمة من 0 إلى 2، حيث معظم الأحيان تكون قيمته 0.7 لرفع إحتمال التغيير والتجدد بالنتيجة لنفس السؤال.



من أهم محركات المحادثة التي تطبّق نظام طريقة استرجاع البيانات المكمّلة هو **محرّك أنصاري** الذي طوّره الدكتور وليد قادوس، (Chief Science Officer Anyscale.com) هو المحرّك المفتوح المصدر:

https://Ansari.chat

https://github.com/waleedkadous/ansari

ويسعي المعهد العالمي لحوسبة القرآن والعلوم الإسلامية لتوثيق إمكانيات هذا المحرّك عن طريق اتمتة فحصه على أسئلة وأجوبة من كليات الشريعة والفقة، وأيضا فحص قدرته على استرجاع آيات القرآن الكريم الـ 6236 آية بدون أي خطأ.

4. نماذج اللغة المفتوحة المصدر مثل LLaMA-2, Falcon:

ما هي الآليات المتاحة لتخصيص نماذج اللغة المفتوحة Fine-tuning عن طريق استرجاع البيانات المكمّلة Retrieval Augmented Generation (RAG) أو استدعاء وظيفة البرنامج المساعد Plugin Function Call. هذه الروابط توفّر نموذج اللغة LLaMA-2 (7B, 13B, 70B) وأيضا نموذج اللغة Falcon (7B, 40B, 180B) حيث أنها مفتوحة لتطوير الأبحاث ومجانية للتوزيع التجاري

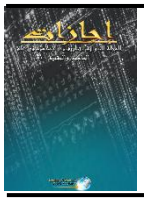
https://huggingface.co/blog/llama2

LLaMA-2: https://huggingface.co/meta-llama

Falcon: https://huggingface.co/tiiuae

تعتبر نماذج اللغة مفتوحة المصدر مثل ال LLaMA-2 and Falcon من أهم التطورات في مجال معالجة اللغة الطبيعية والذكاء الاصطناعي حيث أنها تنافس القدرات اللغوية مع ال GPT-4، ولكنها تنقصها بعض الأمور الهامة لتطبيقها في العلوم الإسلامية، مثلا ينقصها:

- التدريب على اللغة العربية: كلا الـ LLaMA-2 and Falcon لا يتعاملان مع اللغة العربية بشكل مقبول، وذلك لضعف المعلومات التدريبية باللغة العربية، ويحتاجان تخصيص التدريب لتعلّم اللغة العربية والعلوم الإسلامية خاصّة.



- القدرة على الترجمة من اللغة العربية وإليها
- القدرة على الترميز Tokenization للغة العربية بشكل فعّال
- القدرة على استدعاء وظيفة البرنامج المساعد Plugin Function Call الذي يساعد في استرجاع المعلومات الحساسة مثل آيات القرآن الكريم والحديث الشريف من قاعدة بيانات سحابية أو محلّية.

5. المعدات السحابية للحوسبة المتوازية السريعة (Cloud GPU Parallel Computing):

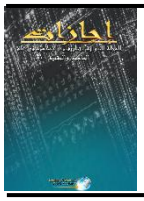
من أقل الخدمات تكلفة في الوقت الحالي التي يمكن استخدامها في الحوسبة السريعة هي الخدمات السحابية المباشرة من شركة Nvidia والتي موقعها Lambdalabs.com حيث بالإمكان استئجار H100 GPU الأسرع في عالم الحوسبة بسعر 2$ للساعة فقط، مع توفر سعة حوسبية كبيرة وفق احتياجات المستخدم. وكذلك يمكن استئجار المعدات على سحابة جوجل Google Cloud Computing (GCP) وكذلك عبر سحابة أمازون Amazon Web Services (AWS) أو من خلال سحابة مايكروسوفت Microsoft Azure ولكن بتكلفة أعلى. وهناك أيضا شركة توفّر المعدات الحوسبية في مواقع مختلفة عالمية غير مركزية (Decentralized) تسمي io.net بقيادة الأستاذ أحمد شديد Ahmad Shadid.

ما يلي، رابط لطلب منحة من شركة جوجل لإستعمال عدّة TPUs لمدّة شهر مجانا بغرض دعم الأبحاث العلمية:

https://sites.research.google/trc/about

6. برامج التدريب والخدمة للنماذج اللغوية:

برامج التدريب وإدارة عملية التدريب على الحوسبة المتوازية معقّدة وتتطلب الكثير من الجهد للتطوير، لذلك فإن أفضل أسلوب حاليا هو استخدام برامج ومكتبات وظيفية وأدوات جاهزة للاستعمال من شركة Anyscale.com التي تُطوّرها باستخدام منصّة Ray Platform مفتوحة المصدر. وتوفر خدماتها الأقل تكلفة على الإطلاق مثل Ray Data لمعالجة النصوص والمعلومات المدخلة لعملية التدريب إلى صيغة JSON، و Ray Tune لتعديل المتغيرات لتدريب الشبكة بشكل أكثر فاعلية Parameter Optimization، و Ray Train لتدريب نماذج اللغة وتخصيصها (Model Training & Fine-Tuning) ، و Ray Serve لتوفير خدمات نماذج اللغة عالميا Model Inferencing ، حيث تكلف خدمات نماذج اللغة 1$ لكل مليون رمز token عند توفيرها للعامّة. ومن الجدير بالذكر أن منصة Ray



من شركة Anyscale.com بالإمكان إستعمالها على أي سحابة حوسبة مثل Google, Nvidia, AWS. وهذه روابط برامج جاهزة توفّر نقطة إنطلاق للمستعمل للتدرب عليها وتخصيصها في تطبيقه:

https://tinyurl.com/Anyscale-Ray-Training

https://training.anyscale.com/

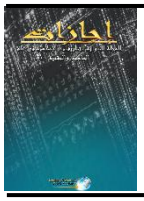
7. الوصلات الوظيفية لنماذج اللغة ودورها في تطوير قدرات نموذج اللغة لتحسين دقّة الإجابة:

في هذا المجال هناك طريقتان هما:

1- **الاستدعاء الوظيفي Function Call** : من الممكن أن نجعل نماذج اللغة الحديثة قادرة على التفاعل مع بيئتها (مثل الملفات المحليّة Excel, Word, PDF والاستعلام من قواعد البيانات والبحث في الإنترنت) من خلال تدريبها على الاستدعاء الوظيفي لبرامج خارجية خلال ردها على المستخدم، وهو ما سيساعد المستخدم في تطبيق قدرات نماذج اللغة على الاستنباط والمنطق الاستنتاجي وفهم اللغة في دائرة معلوماته الشخصية والملفات وقواعد البيانات الخاصة به.

2- **الخدمات الوظيفية Agent Call** : تمكين نماذج اللغة الحديثة من التفاعل مع بيئتها (مثل الملفات المحليّة Excel, Word, PDF، والإستعلام من قواعد البيانات بصيغة SQL، والبحث في الإنترنت) من خلال استدعاء **الخدمات الوظيفية** LangChain.com أو LLaMAIndex.ai باستخدام بيانات خاصة، والتي تسخّر استخدام النموذج اللغوي وقدراته في تطبيقات عمليّة، مثلا في التحليل والاستنتاج المنطقي لأغراض تطوير تطبيقات للذكاء الاصطناعي وفهم اللغة في دائرة المعلومات الشخصية والملفات وقواعد البيانات الخاصة بالمستخدم. يتم ترميز الملفات الشخصية والقواعد البيانية بأرقام Embeddings ثم يتم تحليل سؤال المستخدم من قبل نموذج اللغة والبحث عن معلومات متقاربة رقميا للإجابة عليه. ويُدرج الناتج في الإجابة على المستخدم.

https://github.com/openai/openai-cookbook/blob/main/examples/How_to_build_a_tool-using_agent_with_Langchain.ipynb



في الختام نؤكّد على الدور الهام الذي تلعبه نماذج اللغة في التطوّر الرقمي للإنسان المسلم المعاصر، ولكن هناك حاجة لجهد ضروري اضافي لتوجيه تقنية نماذج اللغة المتوفّرة حاليا نحو تقنية مفيدة في خدمة الإسلام وعلومه، وخاصة للإستدلال بالمعلومة من القرآن الكريم والحديث الشريف بصيغة صحيحة تامة موثّقة بدون أخطاء تنتج عن طبيعة التصميم لشبكات التعلّم العميق لجميع نماذج اللغة الحالية، مع تفاوت قدراتها بالذكاء الاصطناعي حسب حجمها وسعة تدريبها المعلوماتي. ونشجّع المهتمين بهذا المجال على التواصل مع مجلس الإدارة للمعهد العالمي لحوسبة القرآن والعلوم الإسلامية للمشاركة في بعض هذه المشاريع الهامّة.

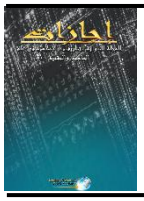
8. المراجع:

1. استرجاع البيانات المكمّلة Retrieval Augmented Generation (RAG)
https://github.com/openai/openai-cookbook/blob/main/examples/How_to_call_functions_for_knowledge_retrieval.ipynb
2. استدعاء وظيفي لبرامج مساعدة Plugin Function Call
https://github.com/openai/openai-cookbook/blob/main/examples/How_to_call_functions_with_chat_models.ipynb
3. قواعد بيانات مضمّنة
4. برنامج Sahih AI للحديث الشريف على منصّة OpenAI Plug-in للوصلات للأستاذ محمد الدليمي Muhammed Aldulaimi المفتوح المصدر:
https://github.com/that-one-arab/sahih-ai
5. محرّك أنصاري من تصميم الدكتور وليد قادوس Waleed Kadous المفتوح المصدر:
https://Ansari.chat and https://github.com/waleedkadous/ansar
7. ورقة نموذج اللغة المفتوح المصدر LLaMA-2
https://arxiv.org/pdf/2307.09288.pdf
8. ورقة نموذج اللغة المفتوح المصدر Falcon
https://huggingface.co/tiiuae

مؤسسة النهضة العلمية

9. السيرة الذاتية للمؤلف

المهندس محمد محمد خير المدير التنفيذي للمعهد العالمي لحوسبة القرآن والعلوم الإسلامية، ويعمل أيضا بهندسة الأجهزة الطبية وإدارة مشاريع تطويرها. حاصل على بكالوريوس (1989) وماجستير (1991) هندسة طبية من جامعة ماركيت، وأيضا إدارة أعمال تنفيذية من جامعة إلينوي (2007). وشارك في دراسات عليا لدرجة الدكتوراة بهندسة الكهرباء بمعهد إلينوي للتقنية (1990–1998). من إهتماماته البحثية تطوير تقنية الأجهزة الطبية للتشخيص والعلاج، وخصوصا الريادة في الخوارزميات البرمجية، ويعمل حاليا في شركة جنرال إليكتريك، ولديه 35 سنة خبرة مع شركات عالمية بالأجهزة الطبية. من إهتمامات المهندس محمد خير تطوير برامج حوسبة القرآن الكريم (ومنها برنامج الرقيم للبحث القرآني) وتطوير قواعد البيانات الإحصائية للقرآن التي تمكّن من إستكشاف التركيب الرقمي للقرآن الكريم الذي يعكس الإحكام بآيات الله تعالى. وأيضا لتمكّن من ربط العلوم وعناصر المعرفة بين القرآن الكريم والحديث الشريف، وهو ما يحقق الفهم القوي المتجدد والرؤية العميقة التي تتكامل مع الدور اللغوي لمحتواهما. ومن إهتماماته تطوير تطبيقات الذكاء الإصطناعي لعلوم القرآن والعلوم الإسلامية والمعالجة الآلية للغة العربية والترجمة الآلية منها وإليها. المهندس محمد خير حاصل على أكثر من 65 براءة إختراع وستة أوراق بحثية علمية منشورة.



مؤسسة النهضة العلمية

10. الخلاصة باللغة الانجليزية

Specialization and Fine-Tuning of AI Large Language Models for Service of the Holy Quran and Islamic Sciences

Mohammad M. Khair, MS, EMBA
CEO, International Institute for Quran and Islamic Sciences Computing
mohammad.khair@qurancomputing.org

Abstract: The application of Large Language Models (LLM) in the service of Quran, Hadith, and Islamic Sciences requires control of its output to be confined to realistic, factual, and reference information avoiding all forms of innovation, imaginative creation, or hallucination which can lead to falsification of reference facts, or falsification of the original text of the Quran scripture, or the text of the Hadith of the Prophet Muhammad (Peace and Blessings be Upon Him). We review some essential strategies that can be applied to control the behavior of the LLM application and to ensure absolute accuracy in the statement of Quran and Hadith text especially, as well as accurate, factual, and reference-based statements in Islamic Sciences in general.

Keywords: Large Language Models (LLM), Artificial Intelligence (AI), Parallel Computing, Retrieval Augmented Generation (RAG), Quran, Hadith, Islamic Sciences