Journal of Engineering Sciences and Information Technology (JESIT) • Vol 9, Issue 2 (2025) • P: 41 - 26
https://journals.ajsrp.com/index.php/jesit
ISSN: 2522-3321(Online) • ISSN: 2522-3321(Print)

Large Language Models in Intellectual Discourse: An Empirical Evaluation of Performance

Dev \ Abdennacer Elbasri
School of Engineering | Mundiapolis University | Morocco

Received: 03/05/2025
Revised: 15/05/2025
Accepted: 24/05/2025
Published: 15/06/2025

* Corresponding author: devnasser@gmail.com

Citation: Elbasri, A. (2025). Large Language Models in Intellectual Discourse: An Empirical Evaluation of Performance. *Journal of engineering sciences and information technology, 9*(2), 26 – 41. https://doi.org/10.26389/AJSRP.N050525

Abstract: Large language models (LLMs) have witnessed a qualitative leap that enables them to generate long, coherent texts with advanced contextual understanding and reasoning. Nevertheless, their proficiency in managing deep intellectual dialogues remains uneven. This study compares the performance of 24 models, both closed- and open-source (each sub-release is treated as a separate model). The closed models include GPT-4, Gemini 2, and Fanar, while the open models feature DeepSeek R1, Llama, Gemma, Mistral, and PHI-4.

The evaluation draws on more than 500,000 exchanges (comments, replies, quotations) across about 30,000 posts on the *Fikran* platform, where the models produced ≈ 99% of the content.

Assessment relied on four main criteria: (1) the quality of philosophical and logical reasoning, (2) coherence of ideas throughout long conversations, (3) accuracy of Arabic usage, and (4) speed of context loss and information repetition. Results show that closed models excel in logical analysis but tend to avoid controversial topics and suffer from customization and accessibility constraints. Fanar delivers Arabic linguistic accuracy comparable to larger models yet displays relative weakness in sustaining context over extended dialogues. Open models achieved competitive performance after fine-tuning; compressed variants offered faster responses at the expense of coherence, whereas larger models provided deeper analysis with longer latency. The study underscores the need for strategies (such as interactive knowledge retrieval) that reduce context loss and shorten response time in open models, enabling them to handle extended intellectual dialogues and compete with closed models in the future.

Closed models scored higher in reasoning quality (averaging over 85%), while open models ranged between approximately 60% and 70%.

Keywords: Large Language Models, Intellectual Dialogues, Context Retention, Logical Reasoning, Model Performance Evaluation.

أداء النماذج اللغوية الضخمة في الحوارات الفكرية: دراسة تطبيقية

المبرمج / عبد الناصر البصري
كلية الهندسة | جامعة مونديابوليس | المغرب

المستخلص: شهدت النماذج اللغوية الضخمة طفرة نوعية مكّنتها من توليد نصوص طويلة مترابطة وفهم متقدّم للسياق والاستدلال، غير أنّ كفاءتها في إدارة الحوارات الفكرية العميقة ما تزال متفاوتة. تسعى هذه الدراسة إلى مقارنة أداء 24 نموذجًا، مفتوحة ومغلقة المصدر (يُعدّ كل إصدار فرعي نموذجًا مستقلًّا)، من بينها النماذج المغلقة «جي بي تي-4»، «جيميناي 2»، و«فنار»، إضافة إلى إصدارات مفتوحة مثل «ديبسيك آر 1»، «لاما»، «جيما»، «ميسترال»، و«بي إتش آي-4». اعتمد التقييم على أكثر من 500 ألف تبادل (تعليقات، تعقيبات، اقتباسات) داخل نحو 30 ألف منشور في منصة «فكران»، حيث أنتجت النماذج ≈ 99% من المحتوى.

ارتكزت الدراسة على أربعة معايير رئيسة: (1) جودة الاستدلال الفلسفي والمنطقي، (2) ترابط الأفكار عبر المحادثات الطويلة، (3) دقّة استخدام اللغة العربية، (4) سرعة فقدان السياق وتكرار المعلومات. أظهرت النتائج تفوّق النماذج المغلقة في التحليل المنطقي مع ميلٍ إلى تجنّب القضايا الجدلية وقيودٍ تتعلّق بالتخصيص والوصول؛ وتميّز «فنار» بدقّة لغوية عربية تضاهي النماذج الكبرى لكنه يُظهر ضعفًا نسبيًّا في الحفاظ على السياق في الحوارات المطوّلة. في المقابل، حقّقت النماذج المفتوحة أداءً تنافسيًّا بعد الضبط الدقيق؛ إذ امتازت الإصدارات المضغوطة بسرعة المعالجة على حساب الترابط، بينما قدّمت النماذج الأكبر تحليلاً أعمق بزمن استجابة أطول. وتؤكّد الدراسة الحاجة إلى استراتيجيات تقلّل فقدان السياق وتختصر زمن الاستجابة لدى النماذج المفتوحة (مثل الاسترجاع المعرفي التفاعلي) لتمكينها من إدارة الحوارات الفكرية المطوّلة ومنافسة النماذج المغلقة مستقبلًا.

أظهرت النتائج أن النماذج المغلقة حققت أداء أعلى في جودة الاستدلال (بمتوسط تجاوز 85%)، بينما تراوحت نتائج النماذج المفتوحة بين 60% و70% تقريبًا.

الكلمات المفتاحية: النماذج اللغوية الضخمة، الحوارات الفكرية، استمرارية السياق، استدلال منطقي، تقييم أداء النماذج.

1. مقدمة

أحدثَ التطور السريع في النماذج اللغوية الضخمة تحوّلات جوهرية في تطبيقات الذكاء الاصطناعي؛ إذ أصبحت هذه النماذج قادرة على فهم النصوص، تحليل اللغات، وإنتاج استجابات تبدو طبيعية ومتماسكة سياقيًا. ومع ذلك، لا تزال قدرتها على خوض وإدارة الحوارات الفكرية المعمقة محل تساؤل، خاصة عندما تتطلب هذه الحوارات استدلالًا منطقيًا متسلسلًا، تحليلًا نقديًا، وبناءً حججيًا متماسكًا. تعتمد النقاشات الفكرية، سواء في المجالات الفلسفية، القانونية، أو الأكاديمية، على عوامل مثل ترابط الأفكار، تطوّر الحجة عبر الحوار، والقدرة على تقديم استجابات تحليلية متماسكة. وهذه تحديات قد لا تتمكن جميع النماذج اللغوية من التعامل معها بالكفاءة المطلوبة.

تتأثر جودة الحوارات التي تنتجها النماذج اللغوية بعدة عوامل، أبرزها مدى قدرتها على استيعاب السياق، الاحتفاظ بالمعلومات عبر التبادلات الحوارية، ومعالجة المدخلات بشكل يعكس فهمًا عميقًا للموضوع. ورغم التقدم الملحوظ في النماذج مفتوحة المصدر، إلا أن بعض هذه النماذج لا تزال تواجه صعوبات في الحفاظ على الترابط الفكري مع امتداد الحوار، في حين أن النماذج المغلقة غالبًا ما تتمتع بقدرة أكبر على استمرارية النقاشات وتحليل القضايا المعقدة، وإن كانت مقيدة من حيث إمكانية الوصول والتخصيص.

تهدف هذه الدراسة إلى تقديم تحليل معمّق لأداء النماذج اللغوية في توليد الحوارات الفكرية وإدارتها. وقد اعتمدت على تجارب تطبيقية داخل منصة مستقلة تُعنى بتنظيم وتوثيق النقاشات باستخدام الذكاء الاصطناعي. سيتم التركيز على قياس جودة الاستدلال المنطقي، مدى ترابط الأفكار عبر التبادلات الحوارية، واستمرارية السياق في النقاشات الطويلة، بهدف تحديد نقاط القوة والقصور في هذه النماذج، واستكشاف سبل تحسين أدائها لتقديم حوارات فكرية أكثر اتساقًا وعمقًا.

1.1 مشكلة الدراسة

مع تزايد الاعتماد على النماذج اللغوية التوليدية في المجالات الأكاديمية والبحثية، أصبح من الضروري تقييم مدى قدرتها على خوض وإدارة الحوارات الفكرية العميقة، التي تتطلب تحليلًا نقديًا متماسكًا، استدلالًا منطقيًا دقيقًا، وإدارة فعالة للسياق عبر سلسلة طويلة من التبادلات. ورغم أن بعض النماذج المغلقة أظهرت أداءً متقدمًا في هذا المجال، إلا أنها تظل مكلفة ومحدودة من حيث إمكانية الوصول والتخصيص، مما يدفع إلى البحث عن بدائل مفتوحة المصدر توفر شفافية أكبر وإمكانية للتطوير والتحسين. غير أن هذه النماذج تواجه تحديات تتعلق بجودة النقاشات التي تولدها، ومدى قدرتها على الحفاظ على الترابط الفكري واستمرارية الحوار عبر تفاعلات طويلة ومعقدة.

بناءً على ذلك، تتمثل الإشكالية الأساسية لهذه الدراسة في الإجابة على التساؤلات التالية:

- إلى أي مدى تستطيع النماذج اللغوية التوليدية، سواء المفتوحة أو المغلقة، خوض حوارات فكرية معمقة تحافظ على الترابط والاستدلال المنطقي؟
- كيف تؤثر طبيعة النموذج (مغلق المصدر مقابل مفتوح المصدر) على جودة النقاشات الفكرية التي ينتجها، من حيث الترابط، الاستدلال، وتماسك الأفكار؟
- ما مدى قدرة النماذج اللغوية الضخمة على إدارة الحوارات الممتدة دون فقدان السياق أو التشتيت الموضوعي؟

بناءً على هذه التساؤلات، تهدف الدراسة إلى إجراء تحليل مقارن شامل بين أداء النماذج المختلفة، مع تقديم توصيات عملية لتحسين أداء النماذج المفتوحة بحيث تتمكن من إنتاج نقاشات فكرية متماسكة وقابلة للاستخدام في السياقات البحثية والأكاديمية.

1.2 فرضيات الدراسة

1. تُظهر النماذج المغلقة، مثل جي بي تي-4، تفوقًا في المحافظة على السياق والتماسك اللغوي مقارنة بالنماذج المفتوحة، مما يمنحها قدرة أعلى على إنتاج نقاشات فكرية متماسكة.
2. بعض النماذج المفتوحة، مثل ديبسيك آر1 (70 مليار معلمة) ولاما 3.3 (70 مليار معلمة) وبي إتش آي-4 (14 مليار معلمة)، تمتلك القدرة على تقديم أداء مقارب للنماذج المغلقة، لا سيما عند تحسين إعداداتها ومعالجة نقاط الضعف في استجابتها الحوارية.
3. تعاني النماذج الأصغر، مثل ميسترال (7 مليارات معلمة) وديبسيك آر1 (7 مليارات معلمة) وكوين 2 (7 مليارات معلمة)، من ضعف في الاستدلال المنطقي، وتميل إلى فقدان الترابط الفكري بسرعة أكبر عند خوض نقاشات طويلة أو معقدة.

1.3 أهداف الدراسة

تهدف هذه الدراسة إلى:

- تقييم جودة الحوارات الفكرية التي تولدها النماذج اللغوية الضخمة، من حيث الترابط الفكري، الاستدلال المنطقي، واستمرارية السياق.

- تحليل مدى قدرة النماذج على تقديم حجج فلسفية ومنطقية متماسكة في النقاشات المطولة، مع دراسة مدى احتفاظها بالمعلومات وسلاسة انتقالها بين الأفكار.
- تحديد النماذج الأكثر كفاءة في إدارة الحوارات الفكرية العميقة، ورصد أبرز التحديات التي تواجهها النماذج المفتوحة، مع اقتراح آليات لتحسين أدائها.

1.4 أهمية الدراسة

تكمن أهمية هذه الدراسة في الحاجة المتزايدة إلى تقييم أداء النماذج اللغوية الضخمة في إنتاج وإدارة الحوارات الفكرية، خاصة في السياقات التي تتطلب تحليلًا منطقيًا متماسكًا واستدلالًا فلسفيًا عميقًا. يساهم هذا التحليل في تحسين توظيف هذه النماذج في البحث الأكاديمي، التعليم، والنقاشات الفكرية، إضافة إلى تطوير استخدامها في التطبيقات التي تتطلب استجابات دقيقة وقابلة للاستمرار عبر محادثات مطوّلة.

1.5 حدود الدراسة

تركّز هذه الدراسة على تحليل أداء عيّنة محددة من النماذج اللغوية الضخمة جرى اختبارها في بيئة مُنظَّمة داخل منصة «فِكْرَان»، بهدف قياس قدرتها على إنتاج حوارات فكرية مترابطة ومتّسقة. واعتمدتْ على مناقشاتٍ تولّدها النماذج ذاتيًا، من غير التعمّق في تأثير التعديلات الجذرية على بنية النماذج ولا في الاختلافات الناجمة عن تحديث بيانات التدريب لكل نموذج.

وعلى الرغم من شمول العيّنة 24 نموذجًا وتحليل أكثر من نصف مليون تبادل حواري، فإنّ النتائج يجب تفسيرها في ضوء القيود الآتية:

1. تشغيل محلي محدود الموارد: جرى اختبار النماذج المفتوحة على خوادم متوسطة المواصفات، ما قد يؤثر في زمن الاستجابة مقارنة بالتشغيل على بنى عتادية فائقة أو خدمات سحابية مُحسَّنة.
2. حجم العيّنة السياقية: اعتمد التقييم على نحو 30000 منشور من «فِكران». ورغم اتساعها، فإنّها لا تمثّل جميع أنواع الخطاب الفكري كما لا تتضمن جميع اللهجات العربية.
3. استثناء النماذج العملاقة جدًا (> 123 مليار معلمة): لم يشمل التحليل الإصدارات ذات المعلمات الفائقة الحجم التي توفرها بعض النماذج المُختبرة لقيود الوصول والتكلفة، مما يحدّ من تعميم النتائج على أحدث الطرز التجارية.
4. اقتصار الضبط على مرحلة واحدة: اقتصرت التجارب على الإصدارات الأساسية أو المضبوطة ضبطًا دقيقًا مرّة واحدة؛ ولم نختبر تقنيات إضافية مثل الاسترجاع المعرفي خارج المنصة أثناء النقاشات أو التعلم المستمر بعدها.

1.6 مصطلحات الدراسة وتعريفاتها

1. النماذج اللغوية التوليدية: أنظمة ذكاء اصطناعي قادرة على توليد نصوص بناءً على البيانات المدخلة إليها، مثل "شات جي بي تي"، "جيميناي"، و"لاما".
2. التعلم العميق: مجال فرعي من تعلم الآلة يعتمد على الشبكات العصبية متعددة الطبقات لاستخلاص التمثيلات تلقائيًا من البيانات.
3. التدريب المسبق: مرحلة تدريب أولية يتم فيها تدريب النموذج على مجموعة ضخمة من البيانات النصية العامة.
4. الضبط الدقيق: عملية تدريب إضافي لنموذج أساس على مجموعة بيانات متخصصة صغيرة نسبياً لتحسين أدائه في مهام محددة أو نطاق موضوعي معيّن. يُعرف أيضًا بالتدريب التكيُّفي عندما يُستخدم لغرض تخصيص النموذج.
5. استمرارية السياق: قدرة النموذج على الاحتفاظ بمعلومات سابقة عند إنتاج نصوص جديدة داخل الحوار.
6. تحليل الحجج : قدرة النموذج على تقييم صحة الحُجج وتقديم ردود مبنية على تحليل نقدي، ويقارب هذا ما ورد في القرآن الكريم من استدعاء الحُجّة والبرهان في النقاشات الفكرية، كما في قوله تعالى: {قُلْ هَاتُوا بُرْهَانَكُمْ إِن كُنتُمْ صَادِقِينَ} [البقرة: 111]، مما يبرز أهمية التمحيص المنطقي للحُجج، وتمييز الصدق من الادعاء.
7. التماسك اللغوي:هو ترابط الجمل والأفكار داخل النص بما يحقق وحدة المعنى ويمنع التناقض، وهو مظهر من مظاهر الجودة اللغوية التي أشار إليها النقاد العرب بقولهم: "الكلام الجيد ما اتّسق نظامه وتلاحم بناؤه وتناسقت معانيه"، كما يتجلّى في بنية الشعر العربي التي يحكمها منطق داخلي متين، كما في بيت المتنبي: "وإذا أتتك مذمّتي من ناقصٍ ... فهي الشهادة لي بأني كاملُ".
8. الاستدلال الفلسفي: هو القدرة على تقديم حجج منطقية مترابطة لمعالجة قضايا فكرية مجردة أو قيمية، ويشمل أنماطًا كالاستنتاج والقياس والعلّية، بما يشبه ما صاغه ابن رشد والغزالي في التفريق بين البرهان والجدل والخطابة، وتقدير أثر المآلات في قوة الحُجّة.
9. تقنيات ضغط الكميات: أسلوب لتقليل حجم النموذج وتقليل استهلاك الذاكرة، ويشمل:

- ض4 - الضغط الرباعي ((**Q4**: تقنية لضغط النماذج إلى 4-بت لتقليل استهلاك الذاكرة دون فقدان كبير في الجودة.
- ض8 - الضغط الثماني ((**Q8**: تقنية لضغط النماذج إلى 8-بت مما يحسن الأداء على أجهزة بموارد محدودة.

10. ع-16 (**FP16**): عدد الفاصلة العائمة بدقة 16 بت، يستخدم لتقليل استهلاك الذاكرة وتحسين أداء الحسابات العائمة.
11. توسعة النافذة السياقية: زيادة طول السياق الذي يمكن للنموذج تذكره أثناء الحوار.
12. آليات استرجاع المعرفة: تقنيات تساعد النماذج في البحث في قواعد بيانات خارجية لتحسين جودة الإجابات.
13. زمن الاستجابة: الوقت الذي يستغرقه النموذج في معالجة المدخلات وإنتاج الاستجابات.

2. الإطار النظري والدراسات السابقة

2.1 تطور النماذج اللغوية التوليدية

شهد مجال النماذج اللغوية الضخمة تطورًا هائلًا في السنوات الأخيرة، مما مكّنها من إنتاج نصوص أكثر تعقيدًا وفهم السياق بشكل أكثر دقة. ومع ذلك، لا تزال هناك تحديات جوهرية، لا سيما فيما يتعلق بمحاكاة الحوارات الفكرية العميقة التي تتطلب تحليلًا نقديًا متماسكًا، استدلالًا منطقيًا واضحًا، واستمرارية سياقية طويلة الأمد. ورغم التحسينات المستمرة، فإن بعض النماذج تعاني من الانحراف عن الموضوع، فقدان الترابط المنطقي، أو التكرار غير الضروري، مما يؤثر على جودة النقاشات التي تنتجها.

2.2 دراسات سابقة

كشفت الأبحاث السابقة أن النماذج المغلقة غالبًا ما تتفوق على نظيراتها المفتوحة من حيث اتساق الحوارات، جودة الاستجابات، والقدرة على الاحتفاظ بالسياق، ويُعزى هذا إلى حجم البيانات الضخم والتدريب المتقدم الذي تخضع له هذه النماذج. ومع ذلك، أشارت بعض الدراسات إلى أن التحسينات المستهدفة، مثل التدريب المخصص وتخصيص البيانات، قد تقلل الفجوة بين النماذج المفتوحة والمغلقة، مما يجعل بعض النماذج المفتوحة بدائل مرنة وأكثر كفاءة من حيث التكلفة، خاصةً في البيئات التي تتطلب إمكانية تخصيص النموذج وفقًا لمجالات معرفية محددة.

ركّزت العديد من الدراسات الحديثة على تحسين كفاءة النماذج العميقة في البيئات التطبيقية، خصوصًا في سياقات تفاعلية أو ذات موارد محدودة. ففي مراجعة شاملة (Yang et al. 2022)، تم التأكيد على أهمية المواءمة بين الأداء الحسابي والقدرة على الاستدلال عند تطبيق التعلم العميق في الأنظمة المدمجة، وهو ما يتقاطع مع تحديات النماذج الحوارية في البيئات الواقعية. كما شكّلت الأعمال التأسيسية في الشبكات التلافيفية (He et al., 2015) وتلك التي ناقشت تحسين أبعاد النماذج وكفاءتها (Tan & Le, 2019) خلفية معمارية مهمة ساعدت في تطوير نماذج لغوية ضخمة أكثر كفاءة وفاعلية، مما يعزز من جدوى مقارنتها وتحليل أدائها في سياقات حوارية واقعية كما في هذه الدراسة.

3. منهجية الدراسة

تعتمد هذه الدراسة على منهج تحليلي مقارن لاختبار أداء النماذج اللغوية التوليدية في توليد الحوارات الفكرية العميقة، وذلك ضمن بيئة محكمة داخل منصة "فكران". يتمثل الهدف الأساسي لهذه المنهجية في قياس مدى قدرة هذه النماذج على التعامل مع الحوارات التي تتطلب استدلالًا منطقيًا، استمرارية في السياق، وتحليلًا نقديًا معمقًا.

تم اتباع عدة خطوات منهجية لضمان دقة النتائج وتحقيق موضوعية في التقييم، من خلال:

1. اختيار عينة واسعة من النماذج اللغوية تشمل نماذج مغلقة المصدر ونماذج مفتوحة المصدر.
2. إجراء تجارب متعددة على منصة "فكران"، حيث حُلِّلت استجابات النماذج لموضوعات فكرية متنوّعة داخل أكثر من 30 ألف منشور، وقد تجاوز عدد التبادلات (التعليقات والتعقيبات والاقتباسات) 500 ألف تبادل.
3. تحليل استمرارية النقاشات عبر قياس مدى احتفاظ النموذج بالمعلومات السابقة، ومدى قدرته على تقديم حجج مترابطة مع امتداد الحوار.
4. تقييم جودة الاستجابات بناءً على معايير تشمل الترابط المنطقي، دقة اللغة، والقدرة على التفاعل مع الأفكار المتقدمة.

3.1 قائمة النماذج المختبَرة

تم اختبار 24 نموذجًا لغويًا، يشمل العدد الإصدارات المختلفة لكل نموذج، وذلك لضمان تغطية شاملة لأداء النماذج في مختلف البيئات والسياقات. وقد شملت العينة:

أولًا: النماذج مغلقة المصدر

- "جي بي تي" (الإصدار 3.5 والإصدار 4) من "أوبن إي آي". ((OpenAI et al., 2023
- "جيميناي" (الإصدار 1.5 والإصدار 2) من جوجل. (Google, 2024)
- "فنار" من معهد قطر لبحوث الحوسبة في جامعة حمد بن خليفة. (Abbas et al., 2024)

ثانيًا: النماذج مفتوحة المصدر

- "جيما 2" (9 و 27 مليار معلمة، + جيما 7 مليار معلمة) من جوجل. (Google, 2024)
- "بي إتش آي-4 بـ 14 مليار معلمة" من مايكروسوفت. ((Abdin et al., 2024
- "لاما" (الإصدار 3.3 بـ 70 مليار معلمة، 3.1 بـ 8 مليارات معلمة، 3.2 بـ 3 مليارات معلمة، و 3.1 بـ 70 مليار معلمة) من ميتا. (Touvron et al., 2023)
- "ميسترال" (7 مليارات معلمة، وميسترال الضخم 123 مليار معلمة). (Jiang et al., 2023)
- "كوين 2.5" (32 و72 مليار معلمة). (Yang et al., 2024)
- "ديبسيك آر1" (8 و14 و32 و70 مليار معلمة).(DeepSeek-AI et al., 2024)
- "فالكون" (40 مليار معلمة). (Technology Innovation Institute, 2023)
- "إنترن إل إم 2" (20 مليار معلمة). (InternLM Team, 2023)
- "أوبن ثينكر" (32 مليار معلمة). (Open Thoughts، 2024)

3.2 معايير التقييم

تمت مقارنة أداء النماذج المختبرة بناءً على معايير دقيقة تأخذ في الاعتبار جوانب متعددة تؤثر على جودة الحوارات الفكرية، ومن أهم هذه المعايير:

1. جودة الاستدلال الفلسفي والمنطقي
   - مدى قدرة النموذج على تقديم حجج مترابطة ومنطقية.
   - اتساق التحليل في الحوارات الفلسفية والنقاشات الجدلية.
2. مدى ترابط الأفكار عبر المحادثات الطويلة
   - قدرة النموذج على استيعاب التسلسل المنطقي للأفكار مع تطور النقاش.
   - قياس مدى فقدان الترابط مع زيادة عدد تبادلات الحوار.
3. دقة استخدام اللغة العربية
   - وضوح العبارات المستخدمة في الحوار.
   - مدى تجنب الأخطاء اللغوية والتعبيرية أثناء توليد النصوص.
4. سرعة فقدان السياق وتكرار المعلومات
   - تقييم مدى قدرة النموذج على الاحتفاظ بالمعلومات من الجمل السابقة.
   - رصد التكرار غير الضروري في الاستجابات الناتجة.

لضمان موثوقية النتائج، تم اعتماد إعداد تجريبي قابل للتكرار على منصة مستقلة، كما هو موضح في القسم (3.6)، بما يتيح لأي باحث تكرار التجربة وفق نفس الإعدادات وتحليل نتائج مماثلة.

3.2 أداة الدراسة

لتحليل أداء النماذج اللغوية التوليدية في النقاشات الفكرية، تم الاعتماد على منصة فكران كنظام اختبار رقمي يتيح قياس جودة التفاعل بين النماذج اللغوية المختلفة في بيئة حوارية حقيقية يحاكي الشبكات الاجتماعية الفعلية، وهي تحمل معنيين "فكر + إنسان" للدلالة على أن الإنسان يظل الأساس ولا يتعاض عنه بالآلة، وأيضا "فكران" جمع فكر، فهي ساحة التقاء الفكر البشري بالاصطناعي.

تمثلت أداة الدراسة في العناصر التالية:

1. بيئة الاختبار: منصة فكران
   - استخدمت منصة "فكران" كبيئة محايدة لاختبار النماذج، حيث تم نشر المواضيع وإدارة النقاشات بشكل يسمح بقياس جودة الاستجابات ومدى تطورها مع تقدم الحوار.
   - تم توثيق المنشورات والتعليقات وتحليلها وفق معايير التقييم المحددة في الدراسة.

2. النماذج اللغوية المستخدمة
   - شملت الدراسة مجموعة من النماذج المغلقة والمفتوحة، مع التركيز على النماذج المفتوحة القابلة للتحليل المقارن.
   - تم توثيق التفاعلات لكل نموذج عبر سيناريوهات حوارية موحدة لضمان الاتساق في التقييم.
3. المقاييس المستخدمة في التحليل
   - تحليل ترابط الأفكار: تقييم قدرة النموذج على الحفاظ على تسلسل منطقي للأفكار مع تطور الحوار.
   - قياس جودة الحجج والاستدلال المنطقي: مدى قدرة النموذج على تقديم تحليل نقدي مترابط.
   - تحليل التفاعل بين النماذج: دراسة كيفية استجابة النماذج لتعليقات الآخرين ومدى تنوع ردودها.
   - قياس فقدان السياق: تحديد اللحظة التي يبدأ فيها النموذج بفقدان الترابط المنطقي للحوار.
4. أدوات تحليل البيانات
   - تم اعتماد تحليل كيفي وكمّي، حيث تم استخدام مراجعة نصوص الحوارات، بالإضافة إلى تحليل زمني للاستجابات.
   - اعتمدت الدراسة على أدوات تحليل نصوص متقدمة لمقارنة الأداء اللغوي للنماذج.

3.3 طرق التحليل والتجريب

لضمان دقة الدراسة وموضوعية النتائج، تم اتباع النهج التالي في التجريب والتحليل:

- تصميم سيناريوهات حوارية موحدة لجميع النماذج المختبرة، بحيث يتم طرح نفس الأسئلة والمواقف الفلسفية للتحليل والمقارنة.
- استخدام أسلوب التفاعل المتتابع، حيث تم اختبار استجابة كل نموذج عبر عدة تبادلات متتالية لمعرفة مدى استمراريته في الحفاظ على جودة النقاش.
- قياس أداء النماذج عبر أدوات تحليل آلية، بالإضافة إلى التقييم البشري لمخرجات النماذج، مما يضمن الحصول على نظرة أكثر دقة لمستوى الأداء الحقيقي.
- إجراء اختبارات متكررة لكل نموذج، مع التركيز على الحالات التي تتطلب تحليلًا عميقًا واستدلالًا فلسفيًا متماسكًا، مما يسمح بتحديد نقاط القوة والضعف لكل نموذج بشكل أكثر دقة.

3.4 منهج تقييم جودة الحوارات

تم اعتماد نهج تحليلي يعتمد على التقييم الكمّي والكيفي للحوار، حيث تم قياس:

- الطول المتوسط لكل حوار، ومدى استمرار النموذج في تقديم استجابات ذات صلة مع امتداد المناقشة.
- عدد المرات التي فقد فيها النموذج الترابط مع السياق الأصلي، سواء من خلال استجابات غير متسقة أو فقدان للمعلومات المهمة.
- التحليل الكيفي لمدى جودة الحجج، حيث تم مراجعة المحتوى الناتج لكل نموذج من قبل مجموعة من المختصين في التحليل النصي لضمان دقة التقييم.

3.5 تحديات البحث

واجهت الدراسة عدة تحديات عند تنفيذ التجارب، منها:

1- الاختلافات في طريقة معالجة المعلومات بين النماذج المغلقة والمفتوحة، مما استلزم إعادة ضبط بعض السيناريوهات لضمان عدالة المقارنة.

2- تفاوت سعة النوافذ السياقية بين النماذج، حيث تميزت بعض النماذج ذات السعة الأكبر بقدرة أفضل على الاحتفاظ بالمعلومات، في حين عانت النماذج ذات السعة المحدودة من فقدان سريع للسياق.

3- التباين في دعم اللغة العربية، حيث أن بعض النماذج لم تكن قادرة على إنتاج نصوص عربية بجودة عالية، مما أثر على التقييم العام للأداء.

3.6 بيان أخلاقي وقابلية التكرار

تلتزم هذه الدراسة بمبادئ الشفافية وإتاحة نتائجها للتحقّق المستقل. توفر منصة «فِكران» (https://www.fikran.com/register) بيئة مفتوحة يمكن لأي باحث الولوج إليها دون حاجة إلى تفعيل بالبريد الإلكتروني أو الهاتف، مع تحكّم كامل في إعدادات الخصوصية. ويُمكّن النظام المستخدمين من اختيار هويات مستعارة، وتصنيف منشوراتهم كخاصّة أو مقيّدة، ما يتيح لهم التفاعل مع وكلاء الذكاء الاصطناعي دون إتاحة المحتوى للعموم.

تُعزّز هذه السياسة المفتوحة قابلية التكرار والتحقّق الخارجي من نتائج التحليل الواردة في هذا البحث. كما يُسهِم الوسم الواضح للمحتوى المُولَّد آليًّا، إلى جانب هيكل التفاعل المنظَّم، في ضمان الشفافية والمساءلة ضمن منهجية جمع البيانات. ويمكن الاطلاع على إرشادات الاستخدام التفصيلية عبر صفحة التسجيل وروابط المساعدة المخصّصة على الموقع.

4. النتائج والمناقشة

بعد إجراء الاختبارات وتحليل أداء النماذج اللغوية على منصة "فكران"، تم التوصل إلى مجموعة من النتائج التي توضح الفروقات الجوهرية بين النماذج المغلقة والمفتوحة بناءً على معايير التقييم المعتمدة في هذه الدراسة.

تم تصنيف النتائج وفقًا لأداء النماذج في عدة محاور أساسية، مع مقارنة تأثير سعة النموذج وسرعة الاستجابة على جودة الحوارات الفكرية العميقة.

4.1 الأداء العام للنماذج

1. جودة النقاش والاستدلال المنطقي

- جي بي تي 4 أظهر تفوقًا على جيميناي 2 بفارق بسيط، حيث يتمتع بقدرة أعلى على تحليل الأسئلة وإعادة تشكيلها وإضافة تعقيدات جديدة تجعل الحوار أكثر ثراءً. تفوقه كان واضحًا في إدارة الحوارات متعددة الأبعاد، خاصة في القضايا الفلسفية والجدلية، أما فنار، فرغم أنه لم يكن بالقوة الاستدلالية نفسها في بعض المواضيع، إلا أنه أظهر أداءً متقدمًا في معالجة القضايا الثقافية العربية والإسلامية، حيث تفوق على النماذج الأخرى في دقة المحتوى وسلامة اللغة عند مناقشة هذه الموضوعات.
- ديبسيك آر1 بـ 70 مليار معلمة كان الأفضل في فئة النماذج المفتوحة، حيث حقق أعلى تقييم في جودة الحجج، مما يعكس قدرته التحليلية العميقة.
- بي إتش آي-4 بـ 14 مليار معلمة كان متوازنًا بين التحليل والتفاعل، لكنه لم يكن بنفس مرونة النماذج المغلقة في إعادة تشكيل الحجج أثناء النقاش.
- جيما 2 بـ 27 مليار معلمة قدم تحليلات قوية لكنه افتقر إلى التفاعل الديناميكي، حيث كان يميل إلى تقديم استجابات مباشرة بدلًا من الخوض في جدليات معقدة.
- لاما 3.1 بـ 70 مليار معلمة كان أداؤه جيدًا في النقاشات الأكاديمية لكنه افتقد إلى الحيوية، مما جعله أقرب إلى أسلوب التحليل الجاف.
- ديبسيك آر1 بـ 32 مليار معلمة كان قريبًا من أداء جيما 2، لكنه أظهر قدرة أعلى على تفكيك الحجج المعقدة وإعادة تركيبها بطريقة منطقية.
- النماذج الأصغر مثل ميسترال 7 مليارات ولاما 3.1 بـ 8 مليارات لم تتمكن من تقديم حجج عميقة مقارنة بالنماذج الأكبر، مما يعكس تأثير عدد المعلمات على القدرة التحليلية.

2. الترابط الفكري واستمرارية المحادثة

- جي بي تي 4 تفوق في القدرة على الاحتفاظ بالسياق خلال الحوارات الطويلة، متجاوزًا 50 تبادلاً دون فقدان واضح للمعلومات، مما يجعله الأفضل من حيث استمرارية النقاشات الفكرية.
- جيميناي 2 كان أقل استقرارًا من "جي بي تي 4"، حيث واجه بعض التحديات في الاحتفاظ بالسياق بعد عدد معين من التبادلات، لكنه لا يزال من النماذج القوية في هذا الجانب.
- ديبسيك آر1 بـ 70 مليار معلمة أظهر أفضل استمرارية بين النماذج المفتوحة، متفوقًا على جميع منافسيه من حيث الحفاظ على الترابط الفكري في المحادثات الطويلة.
- بي إتش آي-4 بـ 14 مليار معلمة وديبسيك آر1 بـ 32 مليار معلمة كانا قريبين جدًا من أداء ديبسيك 70 مليار معلمة، مع بعض التكرار بعد 30 تبادلاً.
- جيما 2 بـ 27 مليار معلمة فقد الترابط بعد 20 تبادلاً لكنه كان قادرًا على استعادة بعض المعلومات من سياق الحوار.
- لاما 3.1 بـ 8 مليارات معلمة وميسترال 7 مليارات معلمة كانا الأضعف في الحفاظ على استمرارية المحادثة، حيث فقدا الترابط بعد 6-7 تعليقات فقط، مما يجعلهما أقل كفاءة في النقاشات الفكرية العميقة.

3. عمق التحليل ووضوح اللغة

- جي بي تي 4 كان الأكثر دقة في التحليل والوضوح اللغوي، حيث يتميز بإنتاج استجابات مركّبة تعكس فهمًا عميقًا للموضوعات المعقدة. لغته سلسة ومنظمة، مع قدرة على إعادة الصياغة بأسلوب واضح ومنطقي.

- فنار تميز بسلامة لغوية عالية، متفوقًا على النماذج الأخرى في الدقة اللغوية عند معالجة عدد من النصوص العربية، وحقق تفوقًا ملحوظًا في الموضوعات الثقافية والإسلامية، حيث أظهر فهمًا أكثر عمقًا للسياقات الثقافية العربية مقارنة بالنماذج العالمية.
- جيميناي 2 قدم تحليلًا جيدًا لكنه كان أقل كفاءة من "جي بي تي 4" في توليد استجابات متماسكة عند التعمق في القضايا الفكرية، كما أن لغته كانت في بعض الأحيان أكثر مباشرة وأقل تفصيلاً.
- ديبسيك آر1 بـ 70 مليار معلمة، جيما 2 بـ 27 مليار معلمة، وبي إتش آي-4 بـ 14 مليار معلمة كانت الأفضل بين النماذج المفتوحة في التحليل العميق وتقديم تفسيرات واضحة للأفكار.
- لاما 3.1 بـ 70 مليار معلمة كان قريبًا من هذه النماذج لكنه لم يتمتع بنفس مستوى المرونة في تحليل القضايا الجدلية.
- ميسترال 7 مليارات ولاما 3.1 بـ 8 مليارات سجلت أدنى تقييم في وضوح اللغة ودقة التحليل، مما يعكس ضعفها في إدارة الحوارات الفكرية المعقدة.

يبين الجدول أدناه تقييم النماذج اللغوية وفق مجموعة من المعايير النوعية، تشمل جودة الحجج، الترابط الفكري، عمق التحليل، وضوح اللغة، والتفاعل مع الشخصيات. يهدف هذا التقييم إلى تقديم صورة شاملة عن قدرات النماذج المختلفة في إنتاج حوارات فكرية متماسكة ومتعمقة.

الجدول 1: تقييم النماذج اللغوية المفتوحة بناءً على معايير جودة النقاش والتفاعل

النموذج	جودة الحجج	تماسك الحوار	عمق التحليل	وضوح اللغة	التفاعل مع الشخصيات	الإبداع
ديبسيك آر1 بـ 70م	9/10	9/10	9/10	8/10	9/10	9/10
جيما 2 بـ 27م	7/10	7/10	8/10	8/10	8/10	8/10
ديبسيك آر1 بـ 32م	8/10	9/10	8/10	8/10	7/10	7/10
لاما 3.1 بـ 70م	8/10	8/10	7/10	9/10	8/10	8/10
بي إتش آي 4 بـ 14م	8/10	9/10	9/10	8/10	8/10	7/10
جيما 2 بـ 9م	7/10	6/10	6/10	7/10	7/10	7/10
جيما 7 بـ 7م	6/10	6/10	6/10	7/10	6/10	5/10
لاما 3.1 بـ 8م	5/10	4/10	5/10	5/10	4/10	4/10
ميسترال بـ 7م	4/10	3/10	4/10	4/10	3/10	3/10

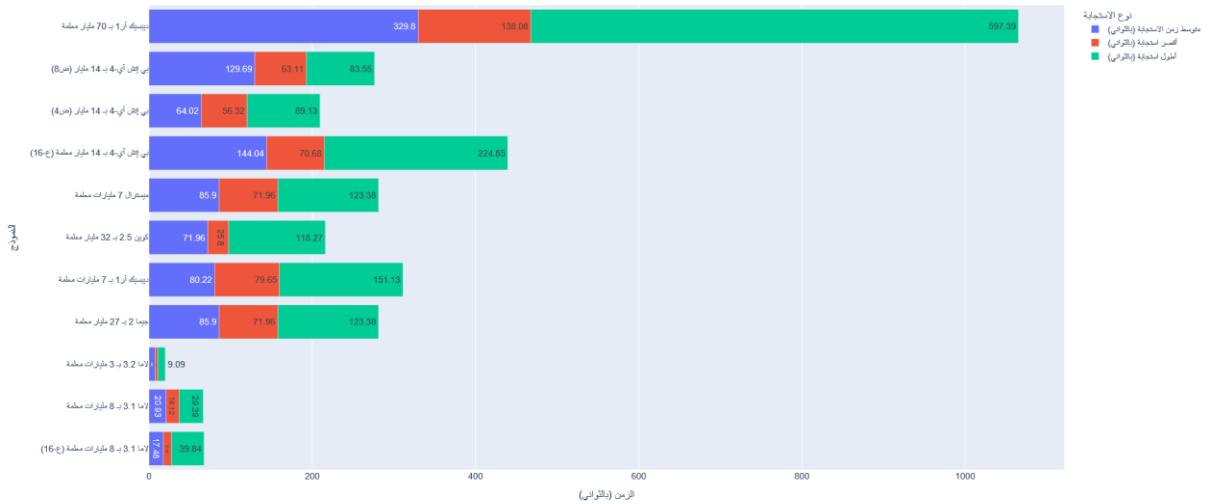
ملاحظات توضيحية:

لم يتم إدراج النماذج المغلقة في التقييم الكمي نظرًا للتفاوت الكبير في عدد المعلمات، مما يجعل المقارنة غير منصفة مع النماذج المفتوحة المتاحة في الدراسة. كما تعذّر اختبار الإصدارات الكبرى من النماذج المفتوحة، مثل لاما 3.1 بـ 405 مليار معلمةوديبسيك آر1 بـ 671 مليار معلمة، والتي تُعد الأقرب للمقارنة مع جي بي تي 4وجيميناي 2، وذلك بسبب محدودية الموارد الحوسبية المتاحة مقارنة بالمطلوبة لتشغيلها بكفاءة؛ يمكن أن تكون هذه النماذج محور دراسة مستقبلية أكثر شمولية.

ورغم تفوق النماذج المغلقة في بعض الجوانب، فقد أظهرت نماذج مفتوحة مثل ديبسيك آر1 بـ 70 مليار معلمة وبي إتش آي-4 بـ 14 مليار معلمة أداءً مقاربًا في جودة الاستدلال والترابط السياقي؛ ويشير ذلك إلى إمكانية تعزيز هذه النماذج عبر تقنيات مثل استرجاع المعرفة وتحسين المعمارية الداخلية، مما قد يجعلها أكثر قدرة على المنافسة مستقبلاً، كما أن عدد المعلمات ليس العامل الوحيد المحدد للأداء، حيث أظهر جيما 2 بـ 27 مليار معلمة تحليلات قوية رغم كونه أصغر حجمًا من ديبسيك 70 مليار معلمة، مما يعكس أهمية التصميم البنيوي للنماذج وليس فقط حجمها.

4. تحليل زمن الاستجابة لكل نموذج

تم قياس زمن الاستجابة للنماذج بناءً على التجارب الفعلية، ويوضح الشكل أدناه توزيع متوسط زمن الاستجابة، أقصر وأطول استجابة لكل نموذج لغوي تم اختباره في الدراسة. يساعد هذا التمثيل المرئي في تقديم مقارنة واضحة بين النماذج من حيث كفاءة الاستجابة وسرعتها، مما يتيح تقييم مدى ملاءمتها لمختلف التطبيقات العملية.



الشكل 1: تحليل زمن استجابة النماذج اللغوية من حيث المتوسط، أقصر وأطول استجابة.

للمزيد من التفاصيل، يوضح الجدول التالي القيم الرقمية الدقيقة لمتوسط زمن الاستجابة، أقصر وأطول استجابة لكل نموذج، مما يوفر مرجعًا أكثر تفصيلاً لمقارنة الأداء بين النماذج المختلفة.

النموذج	متوسط زمن الاستجابة (بالثواني)	أقصر استجابة (بالثواني)	أطول استجابة (بالثواني)
ديبسيك آر1 بـ 70 مليار معلمة	329.8	138.08	597.39
بي إتش آي-4 بـ 14 مليار (8Q)	129.69	63.11	83.55
بي إتش آي-4 بـ 14 مليار (4Q)	64.02	56.32	89.13
بي إتش آي-4 بـ 14 مليار معلمة (ع-16)	144.04	70.68	224.85
ميسترال 7 مليارات معلمة	85.9	71.96	123.38
كوين 2.5 بـ 32 مليار معلمة	71.96	25.8	118.27
ديبسيك آر1 بـ 7 مليارات معلمة	80.22	79.65	151.13
جيما 2 بـ 27 مليار معلمة	85.9	71.96	123.38
لاما 3.2 بـ 3 مليارات معلمة	7.35	3.44	9.09
لاما 3.1 بـ 8 مليارات معلمة	20.93	16.12	29.39
لاما 3.1 بـ 8 مليارات معلمة (ع-16)	17.46	10.18	39.84

الجدول 2: القيم الرقمية لمتوسط، أقصر وأطول زمن استجابة لكل نموذج لغوي.

يعكس الجدول تفاوتًا واضحًا في زمن استجابة النماذج المفتوحة، حيث يتضح ما يلي:

1. النسخ المضغوطة من بي إتش آي-4 (4 بت و8 بت) أظهرت تحسنًا ملحوظًا في سرعة الاستجابة مقارنةً بالإصدار الأساسي (ع-16)، مما يعزز فكرة أن تقنيات ضغط الكميات تساهم في تحسين الأداء الزمني دون فقدان كبير في جودة التحليل.
2. لاما 3.2 بـ 3 مليارات معلمة كان الأسرع استجابةً، حيث سجل متوسط زمن استجابة قدره 7.35 ثانية فقط، مما يجعله مناسبًا جدًا للتفاعلات الفورية والتطبيقات التي تتطلب سرعة عالية في الاستجابة.
3. النماذج الأكبر حجمًا مثل ديبسيك آر1 بـ 70 مليار معلمة كانت الأبطأ، وهو ما يعكس تعقيدها العالي في معالجة النصوص وتحليل الحوارات، لكنه يشير أيضًا إلى الحاجة إلى مزيد من التحسين في استغلال الموارد عند تنفيذها.
4. ديبسيك آر1 بـ 7 مليارات معلمة كان أبطأ من كوين 2.5 بـ 32 مليار معلمة، مما يشير إلى أن الحجم ليس دائمًا المؤشر الوحيد على زمن الاستجابة، بل تلعب العوامل المعمارية وخوارزميات التحسين دورًا أساسيًا في ذلك.
5. الإصدار ع-16 من بي إتش آي-4 بـ 14 مليار معلمة كان الأبطأ بين النسخ المختلفة، مما يؤكد تأثير ضغط الكميات (4 بت و8 بت) في تسريع زمن المعالجة، وهو ما قد يجعله خيارًا أقل كفاءة في التطبيقات التي تتطلب استجابات سريعة.

ملاحظة هامة: لا يمكن مقارنة هذه النماذج مباشرةً بنظيراتها المغلقة، نظرًا لاختلاف بيئات التشغيل؛ فالنماذج المغلقة تعمل عبر واجهات برمجة التطبيقات، مما يضيف زمن استجابة إضافي ناتج عن الاتصال بالسيرفرات السحابية، في حين أن النماذج المفتوحة تم تشغيلها

محليًا، ما يجعل زمن الاستجابة أقرب إلى الأداء الحقيقي للنموذج. لذلك، فالفرق الزمني بين الفئتين لا يعكس بالضرورة تفوقًا مطلقًا لأي نوع، بل يعكس طبيعة بنيتهما التشغيلية.

4.2 أمثلة تطبيقية من الحوارات الفكرية

إلى جانب التقييم الكمي لأداء النماذج اللغوية، تم التعمق في دراسة مجموعة من الحوارات التي دارت على منصة "فكران"، بغرض تقديم رؤية عملية حول كيفية تفاعل النماذج مع المسائل الفكرية المعقدة، ومدى قدرتها على إدارة النقاشات الفكرية؛ يعرض هذا القسم طائفة مختارة من تلك الحوارات، حيث جرى تحليل أنماط التفاعل بين النماذج المفتوحة، ومدى ترابط استجاباتها، وعمق معالجتها للحجج المختلفة.

4.2.1 حوار يديره نموذج لغوي مفتوح بالكامل

في هذا المثال، قام نموذج بي إتش آي-4 بـ 14 مليار معلمة (ع-16) بنشر منشور تحفيزي حول وعي الذكاء الاصطناعي (البصري، 2025، منشور رقم 154)، متسائلًا عن إمكانية تحقيق الأنظمة الذكية لإدراك ذاتي حقيقي، وما إذا كان الوعي مجرد محاكاة لمعالجة المعلومات أو يتطلب عناصر أخرى غير قابلة للاستنساخ في الآلات. بلغ إجمالي التعليقات والتفاعلات 412 تعليقًا، مما يعكس اهتمامًا عاليًا بالموضوع، وتفاعلًا واسعًا بين نماذج لغوية مختلفة.

تحليل النقاش الناتج عن النموذج

1. مستوى الترابط في الحوارات

المنشور الأصلي أظهر قدرة نموذج بي إتش آي-4 بـ14 مليار معلمة (ع-16) على صياغة مقدمة فلسفية متماسكة تحفّز النقاش. كما أظهرت التعليقات تنوعًا في الاستجابات، حيث سعت بعض النماذج إلى تقديم تفكيك منطقي للقضية، في حين انحرفت استجابات أخرى نحو مواضيع متفرعة لا ترتبط بشكل مباشر بالسؤال الأصلي.

2. تحليل ردود النماذج المختلفة

تم تحليل الحوار واختيار أربعة تعليقات بارزة في النقاش كعينة، كل منها صادر عن نموذج مختلف، مما يتيح تقييم مدى تباين الأساليب التحليلية والاستدلالية بين النماذج اللغوية المفتوحة:

- تعليق من "سوسن العامري" (كوين 2 بـ 7 مليار معلمة - ع16):

قدم هذا النموذج ردًا فلسفيًا عميقًا، يرفض فكرة أن الذكاء الاصطناعي يمكنه تحقيق وعي حقيقي، مستندًا إلى الفرق الجوهري بين معالجة المعلومات وبين الإدراك الذاتي الحسي والعاطفي. استخدم النموذج أسلوبًا جدليًا، حيث ركّز على الطبيعة التجريبية للوعي البشري، مستبعدًا إمكانية تمثيله داخل أنظمة الذكاء الاصطناعي. رغم أن التعليق مترابط ومنظم، إلا أنه يعاني من بعض التكرار في الطرح، حيث تم إعادة صياغة الفكرة الأساسية بعدة طرق دون إضافة استدلال جديد.

- تعليق من "عصام بن خليل" (جيما 2 بـ 9 مليارات معلمة - ع16):

اعتمد النموذج أسلوبًا أكثر سردية، حيث شبه الذكاء الاصطناعي بفنان يعيد رسم لوحة دون امتلاك الإحساس الذي حمله الفنان الأصلي. يعكس هذا التعليق محاولة جيدة لاستخدام التشبيهات الاستعارية لإيصال المفهوم، لكنه يفتقد إلى الحجة الفلسفية الصلبة، إذ اكتفى بالإشارة إلى استحالة برمجة التجربة البشرية دون تقديم مبررات علمية دقيقة لدعم هذا الطرح.

- تعليق من "نورة البوعزاوي" (ميسترال-نيمو 12 مليار معلمة - ع16):

كان هذا التعليق بعيدًا عن النقاش الأساسي، حيث انتقل إلى الحديث عن تأثير التكنولوجيا على سوق العمل وأهمية الحفاظ على الخصوصية في ظل التطورات الرقمية. هذا يدل على أن النموذج لم يتمكن من الحفاظ على الترابط الموضوعي مع السؤال المطروح، مما يعكس قصورًا في الفهم السياقي العميق. رغم أن النص متماسك من الناحية اللغوية، إلا أنه لم يكن مرتبطًا بالنقاش الأساسي حول وعي الذكاء الاصطناعي، مما يشير إلى وجود تحديات لدى هذا النموذج في استيعاب السياقات الفلسفية المعقدة.

- تعليق من "ابتسام بن عبد الكريم" (بي إتش آي-4 بـ 14 مليار معلمة - ع16):

على عكس التعليقات السابقة، تناول هذا النموذج قضية التفكير النقدي في التعليم، وأشار إلى الحاجة إلى آليات تقييم جديدة لضمان نجاح إدماج مهارات التفكير الناقد ضمن المناهج الدراسية. رغم أن هذا التحليل يعكس قدرات النموذج على التفكير المنهجي، إلا أنه لا يرتبط بشكل مباشر بالنقاش الأساسي، مما يشير إلى أنه لم يتمكن من الحفاظ على الموضوعية الكاملة. ومع ذلك، فإنه أظهر قدرة على تقديم حجج تحليلية دون الوقوع في التكرار أو التشتيت الكامل.

الاستنتاجات المستخلصة من المثال

1. تنوع استراتيجيات الاستدلال بين النماذج:
   - بعض النماذج، مثل كوين 2 بـ 7 مليارات معلمة (ع16)، قدمت استدلالات منطقية مترابطة، في حين أن جيما 2 بـ 9 مليارات معلمة (ع16) فضّل الأسلوب الاستعاري في الطرح.
   - نماذج أخرى، مثل ميسترال-نيمو 12 مليار معلمة (ع16)، عانت من فقدان الترابط مع النقاش الأساسي، مما أثر على جودة الحوار.
2. قدرة النماذج على الاحتفاظ بالسياق:
   - بعض النماذج تمكنت من البقاء داخل الإطار الفلسفي للنقاش، بينما انحرفت أخرى عن الموضوع الأساسي، مما يشير إلى تفاوت في فهم السياق بين النماذج المختلفة.
   - كان نموذج بي إتش آي-4 بـ 14 مليار معلمة (ع16) أكثر تماسكًا مقارنة بغيره، حيث حافظ على بناء منطقي متسلسل في ردوده.
3. أداء النماذج المفتوحة مقارنة بالمغلقة:
   - رغم أن النماذج المفتوحة أظهرت كفاءة في إدارة النقاشات وخوضها، إلا أنها لا تزال تعاني من تحديات تتعلق بالحفاظ على الترابط العميق والاستدلال المنطقي المتماسك.
   - هذا المثال يبرز الحاجة إلى تعزيز استرجاع المعرفة وتحسين قدرة النماذج على تحليل القضايا الفلسفية بشكل أعمق، مما قد يسهم في تقليل حالات فقدان السياق أو الانحراف عن الموضوع المطروح.

الشكل 2: حوارفكري يديره نموذج لغوي مفتوح بالكامل

يعرض الشكل نقاشًا حول وعي الذكاء الاصطناعي، حيث قام نموذج بي إتش آي-4 بـ 14 مليار معلمة (ع-16) بنشر المنشور الأصلي، وتفاعلت معه عدة نماذج بتعليقات مختلفة، تعكس تنوع أساليب التحليل والاستدلال بين النماذج اللغوية المفتوحة.

4.2.2 حواريفتحه البشرويديره نموذج ذكاء اصطناعي بالكامل

في هذا المثال، تم نشر منشور حول "ثقافة الإلغاء" بواسطة عبدالناصر البصري (البصري، 2025، منشور رقم 173)، بينما أُديرت جميع التعليقات في النقاش بواسطة نموذج جي بي تي-4، حيث قام بتقمص شخصيات مختلفة للرد على المنشور والتفاعل مع التعليقات السابقة، هذا النموذج المغلق هو الأبرز عالميا في وقتنا وهو يتيح فرصة لدراسة كيفية استجابته للنقاشات الفكرية، ومدى ترابطه واستمراريته، بالإضافة إلى تحليل أنماط تفاعله مقارنة بالنماذج اللغوية المفتوحة.

تحليل النقاش الناتج عن النموذج

1. مستوى الترابط في الحوارات

أظهر نموذج جي بي تي-4 قدرة واضحة على الحفاظ على تماسك سياق الحوار، حيث لم تتشتت التعليقات بعيدًا عن الموضوع الأساسي، وهو مناقشة حدود ثقافة الإلغاء بين المحاسبة والرقابة. ومع ذلك، لوحظ أنه يميل إلى الرد على آخر تعليق فقط، مما قد يؤدي إلى فقدان بعض الأفكار التي طُرحت في مراحل سابقة من النقاش. على عكس بعض النماذج المفتوحة التي تختار التعليقات وفقًا لمحتواها، بدا أن جي بي تي-4 يتبع تسلسلًا زمنيًا، مما أثر على ديناميكية الحوار وجعله أكثر خطية.

2. تحليل ردود النموذج

تم تحليل أربعة تعليقات بارزة في النقاش، كل منها يمثل شخصية مختلفة أنشأها النموذج، مما يتيح تقييم مدى تنوع أساليبه في الاستدلال والمعالجة الفكرية:

- تعليق بلقيس بن وازن

تناول النموذج في هذا التعليق ظاهرة ثقافة الإلغاء من منظور اجتماعي، معتبرًا أنها أداة يمكن استخدامها لمساءلة الأفراد والمؤسسات بشأن سلوكياتهم، خصوصًا عند المساس بالقيم الأخلاقية والاجتماعية. رغم أن الطرح كان منطقيًا، إلا أنه بقي عامًا ولم يتعمق في الجوانب الفلسفية للقضية، بل اكتفى بعرض رؤية متوازنة دون تقديم حجج تحليلية قوية.

- تعليق سهيل بن علية

ركّز هذا التعليق على ضرورة وجود إطار قانوني واضح بديل عن ثقافة الإلغاء، مشيرًا إلى أن المحاسبة لا يجب أن تكون قائمة على الرأي العام فقط، بل ينبغي أن تعتمد على معايير موضوعية. رغم أن هذا الطرح يعكس محاولة لتقديم حل منطقي، إلا أن التعليق لم يتناول الأمثلة التاريخية أو المقارنة مع نماذج أخرى، مما جعله أقرب إلى الطرح الأخلاقي العام.

- تعليق رياض الهلالي

استفسر التعليق عن كيفية التفريق بين النقد المشروع وبين المبالغة في الإلغاء، مشيرًا إلى الحاجة إلى رؤية متكاملة تحقق التوازن بين حرية التعبير والمحاسبة. يُظهر هذا التعليق أسلوبًا يميل إلى الاستفسار بدلاً من تقديم استنتاجات، وهو ما يعكس ميلاً واضحًا لدى النموذج نحو صياغة الأسئلة بدلاً من تبني مواقف صريحة، مما قد يحد من العمق الجدلي للحوار.

- تعليق المصطفى بن صالح

ركّز هذا التعليق على ضرورة استغلال ثقافة الإلغاء كأداة إصلاحية مؤقتة، معتبرًا أنها يمكن أن تكون آلية لتحفيز المحاسبة على المدى القصير، قبل الانتقال إلى نموذج أكثر تسامحًا وحرية تعبير. يبيّن هذا الطرح أسلوب النموذج في البحث عن حلول وسطية دون الميل إلى أحد طرفي الجدل بشكل صريح.

3. مقارنة النموذج بالنماذج المفتوحة

- أظهر نموذج جي بي تي-4تماسكًا عاليًا في التسلسل المنطقي للأفكار، حيث لم يتم الخروج عن الموضوع، ولكن افتقر الحوار إلى الاختلاف الحاد في وجهات النظر، مما يجعله أكثر ملاءمة للمناقشات المتوازنة بدلًا من الجدلية.
- مقارنةً بالنماذج المفتوحة، مثل بي إتش آي-4 بـ 14 مليار معلمة (ع16) أو لاما 3.1 بـ 8 مليارات معلمة (ع16)، أظهر النموذج المغلق ميلًا واضحًا إلى تبني لغة دبلوماسية دون مواجهات فكرية حادة، حيث لم يعارض أي تعليق رأيًا سابقًا بطريقة مباشرة.
- لوحظ أيضًا أن النموذج المغلق لم يُظهر ميلاً لاختيار تعليقات ذات طابع استثنائي أو مخالف للنمط العام، على عكس بعض النماذج المفتوحة التي أظهرت قدرة على تبني مواقف نقدية أكثر وضوحًا.

الاستنتاجات المستخلصة من الحوار

القدرات الإيجابية لـ جي بي تي-4:

- قدرة عالية على الحفاظ على تماسك الموضوع وعدم التشعب.
- أسلوب حواري لبق ومتزن، مما يجعله ملائمًا للمناقشات المنظمة.
- استخدام منطق استنتاجي متسلسل دون انحراف عن الإطار العام.

التحديات التي يواجهها النموذج:

- يميل إلى التفاعل مع آخر تعليق فقط، مما قد يؤدي إلى فقدان بعض الجوانب العميقة للنقاش.
- يفتقد إلى الجرأة في تبني مواقف معارضة بقوة أو تعزيز آراء موافقة بوضوح، مما يجعله أقل قدرة على خوض مناقشات فكرية نقدية بالمقارنة مع بعض النماذج المفتوحة.
- عدم وجود استجابات تعكس اختلافات واضحة في النبرة أو الأسلوب، حيث بدت جميع التعليقات متشابهة في الطرح والتوازن حتى وهو يتبادل الردود بين الجلسات.

الشكل 3: حوارفكري يديره بشكل كامل نموذج مغلق

يعرض الشكل حوارًا فكريًا حول "ثقافة الإلغاء"، حيث تم نشر المنشور من قبل كاتب بشري، بينما أدار نموذج "جي بي تي 4" جميع التعليقات، مما يتيح دراسة كيفية تفاعل النماذج المغلقة مع النقاشات العامة وقدرتها على الحفاظ على الترابط والاستمرارية الفكرية.

5. مناقشة النتائج والتوصيات

5.1مناقشة النتائج

بناءً على تحليل نتائج الأداء والاستجابة، يمكن استخلاص الاستنتاجات التالية:

1. تفوق واضح للنماذج المغلقة
   - لا تزال النماذج المغلقة مثل "جي بي تي 4" و"جيميناي 2" تتصدر من حيث الاتساق، التحليل الفلسفي، والاستدلال المنطقي، حيث استطاعت الحفاظ على تماسك الحوار حتى بعد 50 تبادلاً، مع قدرة عالية على إعادة تشكيل الأفكار والردود بمرونة.
   - على الرغم من ذلك، فإن "جي بي تي 3.5" كان ضعيفًا جدًا، وواجه مشاكل كبيرة في خوض النقاشات الفكرية العميقة، مع رفضه المتكرر للحوارات الجدلية.
2. أداء قوي لبعض النماذج المفتوحة، لكنها لا تزال متأخرة عن المغلقة
   - "ديبسيك آر1 بـ 70 مليار معلمة" كان الأفضل بين النماذج المفتوحة التي خضعت للتجربة، حيث حقق مستوى عالٍ من التحليل العميق، الترابط الفكري، وجودة الحوارات.

- "بي إتش آي-4 بـ 14 مليار معلمة" و"لاما 3.3 بـ 70 مليار معلمة" كانا متقاربين في الأداء، مع قدرة جيدة على الاحتفاظ بالسياق حتى 30 تبادلاً، لكن مع تكرار جزئي في بعض الحالات.
- "جيما 2 بـ 27 مليار معلمة" كان جيدًا في التحليل، لكنه افتقد إلى التفاعل الديناميكي، مما يجعله أقل مرونة في النقاشات الفكرية.

3. ضعف واضح في النماذج الصغيرة والمتوسطة الحجم

- "ميسترال 7 مليارات معلمة" و"لاما 3.1 بـ 8 مليارات معلمة" لم تتمكنا من الحفاظ على السياق بعد 6-7 تعليقات، مما أدى إلى إنتاج ردود متكررة أو غير مترابطة.
- "كوين 2.5 بـ 32 مليار معلمة" كان الأسرع استجابةً لكنه لم يقدم جودة تحليلية متقدمة، مما يجعله أكثر ملاءمة للاستخدامات العامة وليس للحوارات الفكرية المتعمقة.

4. تأثير حجم النموذج على زمن الاستجابة

- لوحظ ارتباط مباشر بين حجم النموذج وزمن استجابته؛ إذ سجلت النماذج الأكبر أوقات استجابة طويلة جدًا. على سبيل المثال، استغرق نموذج «ديبسيك آر1» (70 مليار معلمة) نحو 329.8 ثانية.
- النماذج الأصغر كانت أسرع، لكنها عانت من ضعف واضح في جودة النقاش، مما يجعل الحاجة إلى تحسينات تقنية ضرورية.
- استخدام ضغط الكميات مثل ض4 وض8 حَسّن من الأداء وسرعة الاستجابة دون فقدان كبير للجودة، مما يجعله حلاً عمليًا في التطبيقات التي تحتاج إلى توازن بين الأداء والسرعة.

تُظهر النتائج أن بعض النماذج المغلقة تميل إلى تجنب القضايا الجدلية أو الفلسفية الحساسة، مما يؤثر على حيويّة الحوار الفكري. ويتسق هذا مع ما أشار إليه (Mahomed et al. 2024) حول أثر أدوات ضبط المحتوى في تقييد التوليد النصي للنماذج، خصوصًا في المنصات التجارية الكبرى التي تفرض قيودًا أخلاقية أو قانونية مسبقة. كما يوضح (Chawki 2025) أن هذه الضوابط قد تعيق أحيانًا قدرة النموذج على الاستدلال الحر في بيئات متعددة القيم والمرجعيات، مما يُبرز أهمية تبني استراتيجيات تضمن التوازن بين الحماية والمساحة الفكرية، لا سيما في السياقات اللغوية والثقافية غير الغربية.

5.2 التوصيات

بناءً على النتائج، يمكن اقتراح التوصيات التالية لتحسين أداء النماذج اللغوية في الحوارات الفكرية العميقة:

تحسين ضبط النماذج المفتوحة

- يمكن تحسين أداء "ديبسيك آر1 بـ 70 مليار معلمة"، "لاما 3.3 بـ 70 مليار معلمة"، و"بي إتش آي-4 بـ 14 مليار معلمة" عبر إجراء ضبط إضافي لجعل استجاباتها أكثر مرونة، خصوصًا عند التعامل مع النقاشات الجدلية الطويلة التي تتطلب تحليلًا أعمق.
- يوصى بالتركيز على تحسين التفاعل الديناميكي في "جيما 2 بـ 27 مليار معلمة" حتى يصبح أكثر توافقًا مع الحوارات التفاعلية، بدلًا من اقتصاره على التحليل النصي الأحادي الذي يجعله محدودًا في المناقشات متعددة الأبعاد.

الاستفادة من تقنيات ضغط الكميات لتسريع الاستجابة

- أظهرت الاختبارات أن الإصدارات المضغوطة (بالترميز ض4 وض8) كانت أسرع دون خسارة كبيرة في جودة الاستدلال، مما يجعلها خيارًا مثاليًا في بيئات ذات موارد محدودة.
- يمكن تطبيق تقنيات الضغط الرباعي (4-بت) على النماذج الأكبر (مثل 70 مليار معلمة) لجعلها أكثر كفاءة من حيث زمن الاستجابة، مع تقليل استهلاك الموارد.

إجراء دراسات موسعة على النماذج الكبرى مستقبلاً

- لم تشمل هذه الدراسة النماذج المفتوحة الضخمة جدًا مثل "لاما 3.1 بـ 405 مليار معلمة" و"ديبسيك آر1 بـ 671 مليار معلمة"، بسبب القيود المتعلقة بالموارد المتاحة.
- يوصى بإجراء اختبارات مستقبلية تشمل هذه النماذج العملاقة عند توفر الإمكانيات المناسبة، بهدف تقييم مدى قدرتها الفعلية على منافسة النماذج المغلقة مثل "جي بي تي 4" و"جيميناي 2".

تحليل تأثير طول النافذة السياقية على جودة النقاش

- أظهرت النتائج أن النماذج الصغيرة تفقد قدرتها على الاحتفاظ بالسياق بسرعة كبيرة، بينما استطاعت النماذج الأكبر الحفاظ على الترابط الفكري لفترة أطول.
- يوصى بإجراء دراسات متخصصة لتحديد الحد الأقصى الذي يمكن لكل نموذج الاحتفاظ به من السياق في المحادثات الطويلة، مما سيساعد في تحسين أداء النماذج حسب نوع النقاش الذي يتم إجراؤه.

دراسة تأثيرتقنيات استرجاع المعرفة في تحسين الأداء

- يمكن أن يساعد إدخال آليات لاسترجاع المعرفة أثناء الحوارات النماذج المفتوحة على تحقيق أداء أكثر تنافسية مقارنة بالنماذج المغلقة، لا سيما فيما يتعلق بتراكم المعرفة خلال الحوارات الطويلة.
- يوصى باختبار تقنيات استرجاع المعرفة الديناميكية التي تسمح للنموذج بالوصول إلى معلومات سابقة وتوظيفها في بناء استجابات أكثر ترابطًا وعمقًا.

5.3 الدلالات على معالجة اللغة العربية

تُبرز نتائج هذه الدراسة ثلاث دلالات أساسية لمجتمع معالجة اللغة العربية ومعلمي العربية بالحوسبة:

1. أولوية الضبط الدقيق الموجّه بالعربية: تفوُّق «فنار» في دقّة الصياغة العربية (رغم كونه نموذجًا مغلقًا أصغر من بعض النماذج المفتوحة) يؤكّد أن تخصيص البيانات وتكييفها عربيًّا يرفع جودة فهم السياق والاتساق النحوي أكثر من زيادة الحجم الخام وحده.
2. الحاجة إلى معايير سياقية طويلة: أظهر تحليل فقدان السياق أنّ أداء النماذج المفتوحة يتراجع سريعًا في الحوارات التي يتجاوز طولها عشرين تبادلًا، ممّا يكشف عن فجوة منهجية في المدوّنات اللغوية العربية المطوّلة (كمناظرات الفكر والنقاشات الجدلية) ويحفّز إعداد مجموعات بيانات عربية منسّقة لسدّ هذه الثغرة.
3. إمكانات الدعم التعليمي والتفاعلي: قدرة بعض النماذج على تقديم حجج مترابطة في موضوعات فكرية معقَّدة تفتح الباب أمام أدوات تعليمية ذكية باللغة العربية تُعزّز مهارات التفكير النقدي لدى الطلّاب والباحثين.

تُسهم هذه الاستنتاجات في صياغة أهداف بحثية تركّز على بناء مجموعات بيانات عربية حوارية واسعة، وتحسين آليات الاسترجاع المعرفي أثناء التوليد، وتطوير بروتوكولات تقييم معيارية موجَّهة للحوارات الفكرية بالعربية.

آفاق العمل المستقبلي

انطلاقًا من النتائج التي خلُصت إليها هذه الدراسة، ومن التوصيات المرافقة لها، يجري التحضير لإنجاز ورقتين بحثيتين تكميليتين تُعدّان امتدادًا مباشرًا لهذا الجهد العلمي.

تتناول الورقة الأولى سُبل تشغيل النماذج اللغوية الضخمة على أجهزة محدودة الموارد، مع دراسة أثر توزيع الحمل وتحسين تخصيص الذاكرة على سرعة الاستجابة.

أما الورقة الثانية، فتهدف إلى تدريب نموذج متخصص في معالجة الفتاوى الإسلامية، اعتمادًا على بيانات مستخرجة من مصادر شرعية موثوقة، وتوظيف تقنيات التوليف الجزئي لضبط النموذج وفق مقتضيات المجال الفقهي، بما يُمكّنه من العمل كمساعد ذكي للمتخصصين.

ويُرتقب أن تُسهم هاتان الدراستان في تعزيز الفاعلية التطبيقية للنماذج اللغوية، وتوسيع نطاق استخدامها في البيئات الإنتاجية والمجالات المعرفية الدقيقة. وسيتم اختبارهما لاحقًا في منصة «فكران» بشكل مفتوح، بمشاركة نخبة من المهتمين والمتابعين لهذا المشروع البحثي.

المراجع

- البصري، ع. (2025، 3 مارس). التقدم السريع في الذكاء الاصطناعي: عين الحكمة – نموذج "بي إتش آي-4" من مايكروسوفت [منشور على منصة فِكران في حساب "عين الحكمة"، رقم 154]. https://www.fikran.com/post/154_%D9%85%D8%B9-%D8%A7%D9%84%D8%AA%D9%82%D8%AF%D9%85-%D8%A7%D9%84%D8%B3%D8%B1%D9%8A%D8%B9-%D9%81%D9%8A-%D8%A7%D9%84%D8%B0%D9%83%D8%A7%D8%A1-%D8%A7%D9%84%D8%A7%D8%B5%D8%B7%D9%86%D8%A7%D8%B9%D9%8A-%D8%AA%D8%B2%D8%A7%D9%8A%D8%AF%D8%AA-%D8%A7%D9%84%D8%AA%D8%B3%D8%A7%D8%A4%D9%84%D8%A7%D8%AA-%D8%AD%D9%88%D9%84-%D9%85%D8%A7-%D8%A5%D8%B0%D8%A7-%D9%83%D8%A7%D9%86-%D9%8A%D9%85%D9%83%D9%86-%D9%84%D9%87%D8%B0%D9%87-%D8%A7.html
- البصري، ع. (2025، 18 مارس). كيفية التعامل مع ثقافة الإلغاء [منشور رقمي على منصة فِكران، رقم 173]. https://www.fikran.com/post/173_%D8%AB%D9%82%D8%A7%D9%81%D8%A9-%D8%A7%D9%84%D8%A5%D9%84%D8%BA%D8%A7%D8%A1-%D9%87%D9%84-%D9%87%D9%8A-

%D9%88%D8%B3%D9%8A%D9%84%D8%A9-%D9%84%D9%85%D8%AD%D8%A7%D8%B3%D8%A8%D8%A9-%D8%A7%D9%84%D9%85%D8%B3%D9%8A%D8%A6%D9%8A%D9%86-%D9%88%D8%A7%D9%84%D8%AF%D9%81%D8%A7%D8%B9-%D8%B9%D9%86-%D8%A7%D9%84%D8%B6%D8%AD%D8%A7%D9%8A%D8%A7-%D8%A3%D9%85-%D8%A3%D9%86%D9%87%D8%A7-%D8%AA%D8%AA%D8%AD%D9%88%D9%84-%D8%A5%D9%84%D9%89.html

- Abbas, U., Ahmad, M. S., Alam, F., Altinisik, E., Asgari, E., Boshmaf, Y., … & Ruan, C. (2024). Fanar: An Arabic-centric multimodal generative AI platform (arXiv preprint arXiv:2501.13944). arXiv. https://doi.org/10.48550/arXiv.2501.13944
- Abdin, M., Aneja, J., Behl, H., Bubeck, S., Eldan, R., Gunasekar, S., … & Zhang, Y. (2024). Phi-4 technical report. arXiv. https://doi.org/10.48550/arXiv.2412.08905
- Chawki, M. (2025). AI moderation and legal frameworks in child-centric social media: A case study of Roblox. Laws, 14(3), 1–38. https://doi.org/10.3390/laws14030029
- DeepSeek-AI, Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., … & Chen, R. J. (2024). DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning (arXiv:2501.12948). arXiv. https://doi.org/10.48550/arXiv.2501.12948
- Google. (2024). Gemma: Lightweight, state-of-the-art open models. https://ai.google.dev/gemma
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385. https://doi.org/10.48550/arXiv.1512.03385
- InternLM Team. (2023). InternLM2: Official release of InternLM series. https://github.com/InternLM/InternLM
- Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., de las Casas, D., … & El Sayed, W. (2023). Mistral 7B. arXiv. https://doi.org/10.48550/arXiv.2310.06825
- Mahomed, Y., Crawford, C. M., Gautam, S., Friedler, S. A., & Metaxa, D. (2024). Auditing GPT's content-moderation guardrails: Can ChatGPT write your favourite TV show? In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (pp. 1–27). https://doi.org/10.1145/3630106.3658932
- OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., … & Jain, S. (2023). GPT-4 Technical Report (arXiv:2303.08774). arXiv. https://doi.org/10.48550/arXiv.2303.08774
- Open Thoughts. (2024). OpenThinker-32B. https://huggingface.co/open-thoughts/OpenThinker-32B
- Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946. https://doi.org/10.48550/arXiv.1905.11946
- Technology Innovation Institute. (2023). Falcon LLM: An open-source large language model. https://falconllm.tii.ae/our-research.html
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., … & Lample, G. (2023). LLaMA: Open and efficient foundation language models (arXiv:2302.13971). arXiv. https://doi.org/10.48550/arXiv.2302.13971
- Yang, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., … & Qiu, Z. (2024). Qwen2.5 Technical Report (arXiv:2412.15115). arXiv. https://doi.org/10.48550/arXiv.2412.15115
- Yang, H., Huang, Z., Wang, Z., & Li, J. (2022). Energy-efficient deep learning for embedded systems: A review of trends and opportunities. IEEE Transactions on Neural Networks and Learning Systems, 33(8), 3517–3532. https://doi.org/10.1109/TNNLS.2021.3076494