

Retrieval-Augmented Generation based Q&A Model for Infectious Disease in Arabic Language

Ongoing experiments

Yesim Selcuk

11.30.2023

Outline

01 Recap

02 About Models

03 Experimental Set-up

04 Experimental Results

05 Inference Results

06 Future Work

Recap

1. Recap: Motivation

- Infectious diseases disrupt communities and affect to the public health systems negatively.
- Furthermore, recent years have seen that some infectious diseases such as Middle East Respiratory Syndrome (Mers-Cov) drastically increased in the Middle East region.
- For this reason, it is important to track infectious diseases closely.

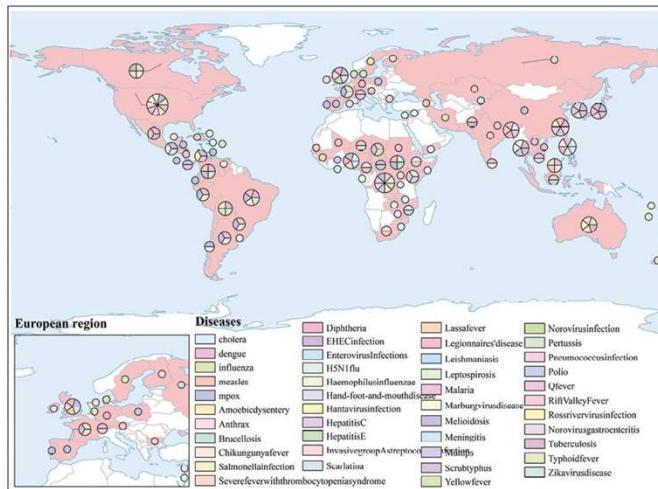


Figure 1: Infectious disease outbreaks in the world, April 2023

Yufan Wu, Jiazen Zou and Yinfu Sun et al. Global Infectious Diseases in April 2023: Monthly Analysis. *Zoonoses*. 2023. Vol. 3(1). DOI: 10.15212/ZOONOSES-2023-1005

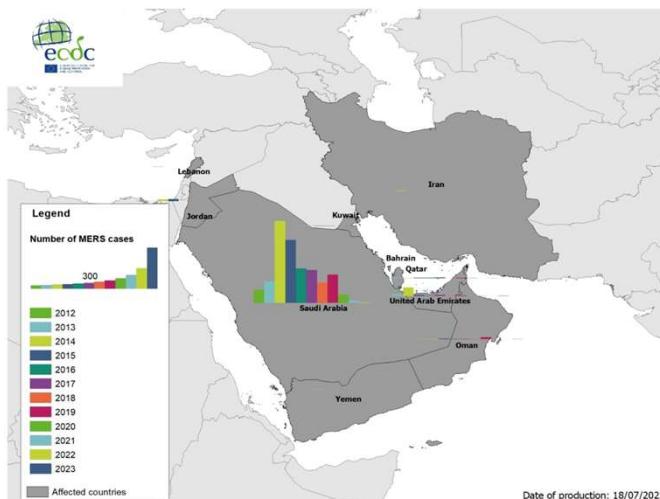


Figure 2: The year distribution for Mers-Cov , 2023

<https://www.ecdc.europa.eu/en/publications-data/geographical-distribution-confirmed-mers-cov-cases-reporting-country-april-2012-4>

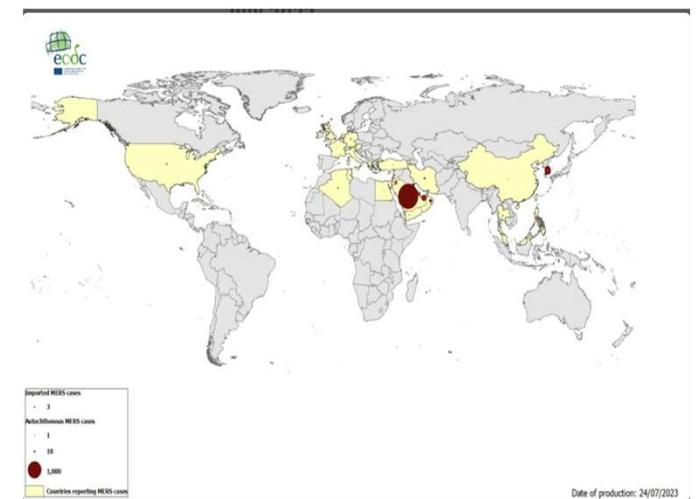


Figure 3: The number of cases for Mers-Cov, 2023

<https://www.ecdc.europa.eu/en/publications-data/geographical-distribution-confirmed-mers-cov-cases-reporting-country-april-2012-4>

1. Recap: The Goal of Study

- To achieve this goal, the main objective of this study is to implement **retrieval augmented based question & answering model for infectious diseases in Arabic language**.
- In order to perform the **question and answering task**, **Llama-2** is aimed to utilize which is an **open-source, large language model**.
- In addition to Llama-2 model, **Low rank adaptation model** called **LoRA** is also aimed to apply in order to reduce the number of trainable parameters in the model.

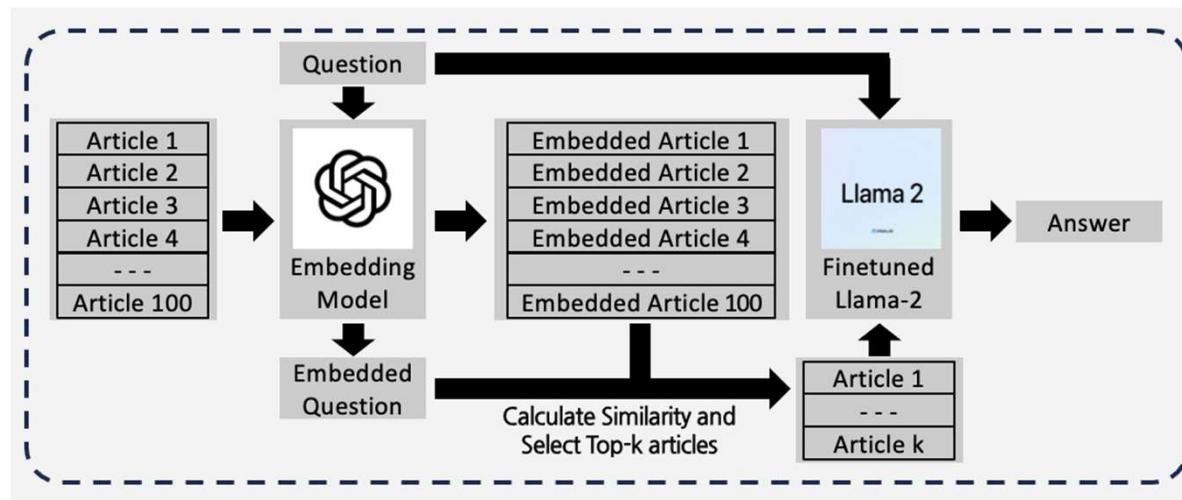


Figure 4: Retrieval Augmented-Generation based Q&A Model by using Llama-2 model

1. Recap: The Research Flow

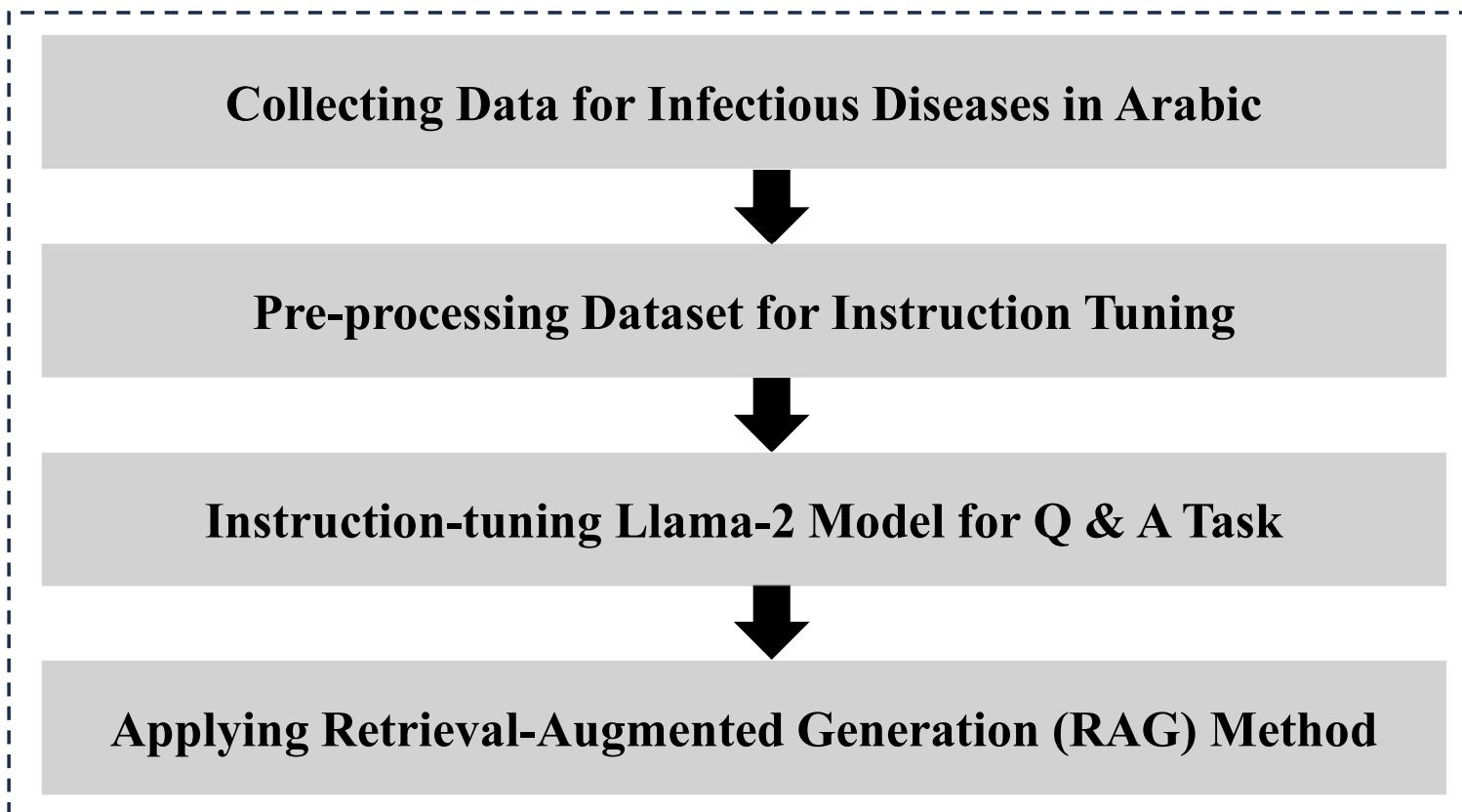


Figure 5: Research Flow

1. Recap: The Data Pre-processing Pipeline

- The pre-processing dataset process contains five steps respectively as shown in Figure 12: dataset collection, summarization, Q&A generation, translation to Arabic, structuring the data as Dolly format.
- At the end of the pre-processing phase, the 45k data was organized in the Dolly format.

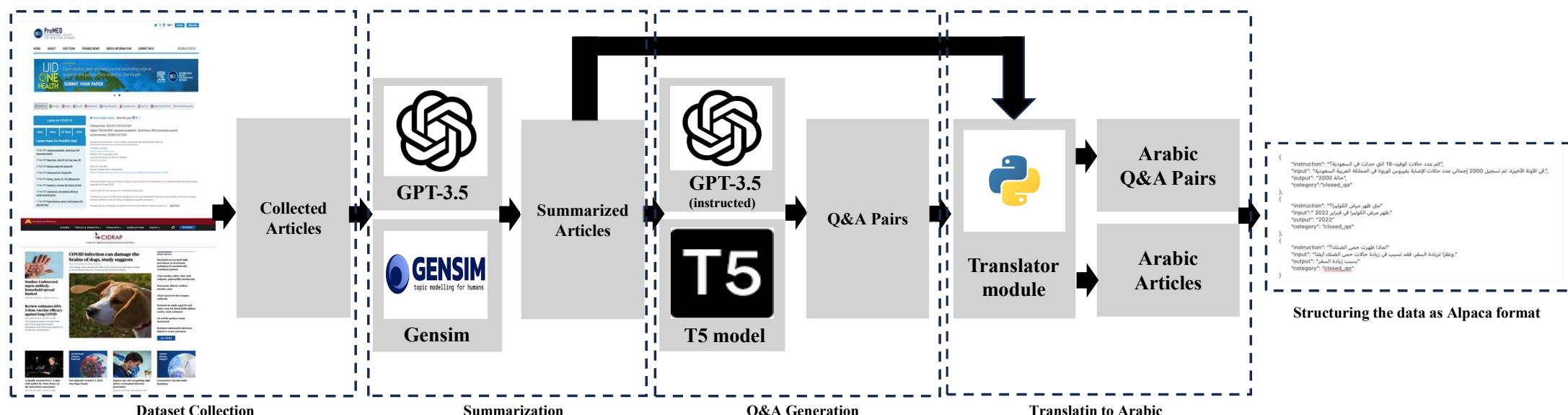


Figure 6: Pre-processing Dataset for Instruction-Tuning

1. Recap: The Instruction Tuning Pipeline

Llama-2 Model:

- For pre-trained LLM, among the open-source LLM models, Llama-2 model was chosen.
- Llama model was firstly introduced in February 2023.
- The main purpose was to introduce the best open-source model which was pre-trained with publicly available datasets.



Figure 7: The main process of instruction tuning

1. Recap: QLoRA - PEFT Method

Quantized LoRA (QLoRA):

- QLoRA applies frozen, 4-bit quantized pre-trained model, and backpropagates the gradients into LoRA.
- Without sacrificing from performance, in order to reduce the memory usage, QLoRA brought some optimizations as listed below:
 1. NF4 Data type
 2. Double Quantization
 3. Paged Optimizers

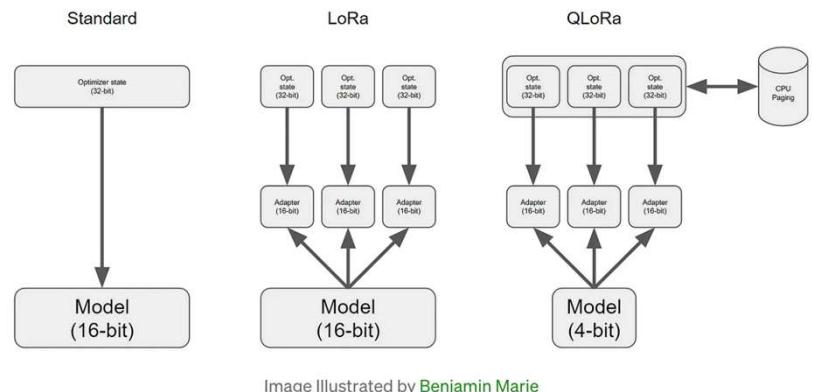


Image Illustrated by Benjamin Marie

Figure 8: QLoRA

<https://medium.com/@zahmed333/qlora-black-boxed-a-brief-summary-b9cf6386473a#:~:text=QLoRA%2C%20or%20Quantized%20LLMs%20with,when%20working%20with%20these%20models.>

1. Recap: Retrieval-Augmented Generation

- In summary, if RAG-based model is applied on instruction-tuned Llama-2 model, the process of this study should be expected as shown in Figure 9.

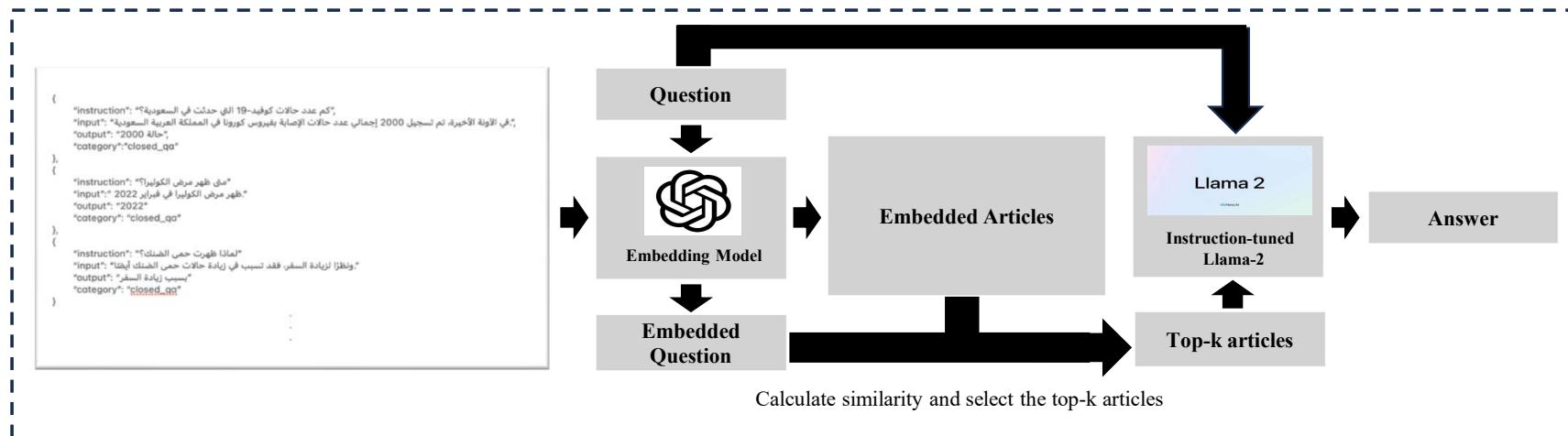
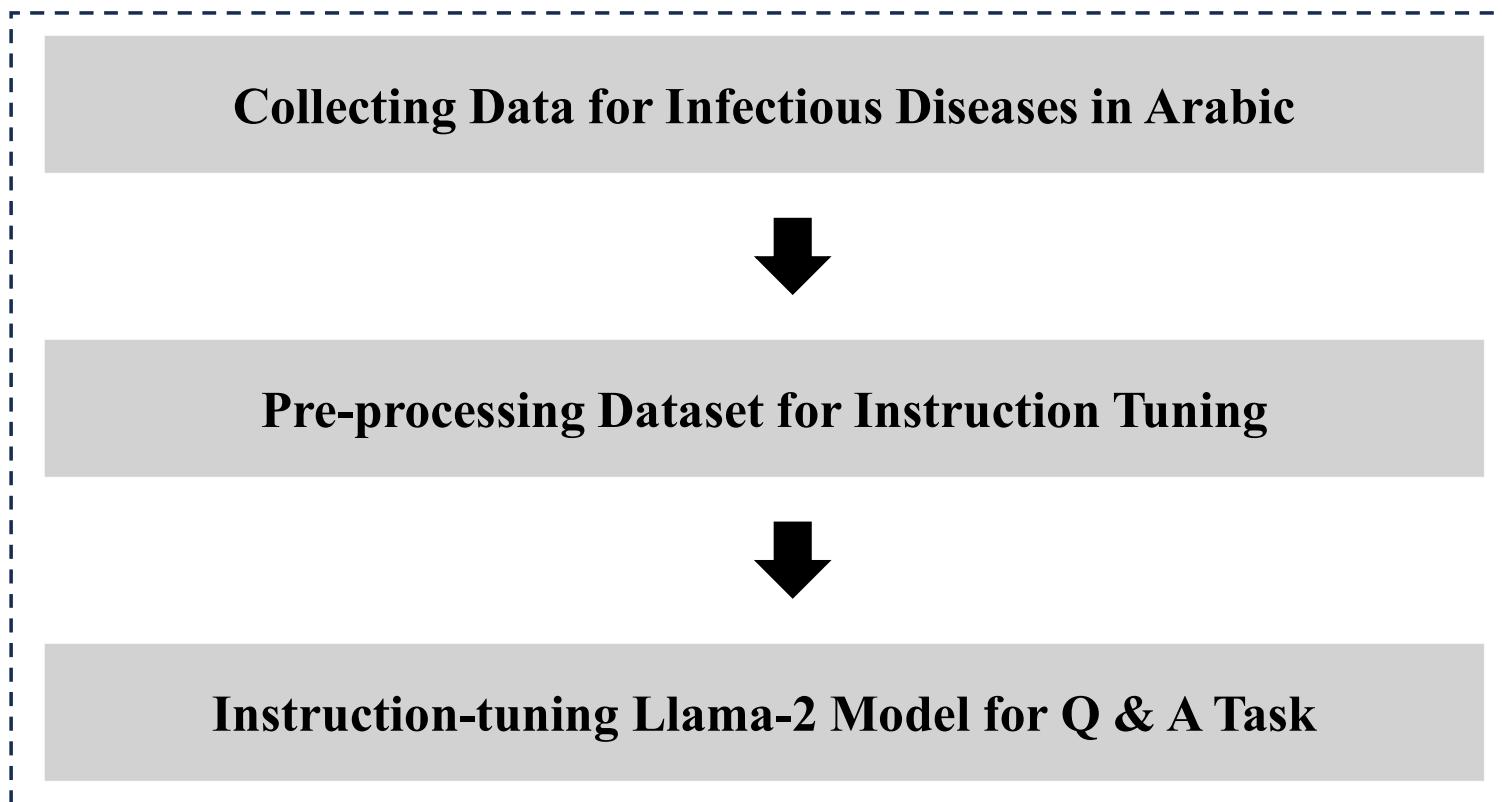


Figure 9: The RAG based Q&A Task by using Llama-2 model

<https://neo4j.com/developer-blog/knowledge-graphs-llms-multi-hop-question-answering/>

1. Recap: Experiment Flow



1. Recap: Experiment Lists

Model	Data	Tokenizer	Pre-training	# of Epoch	Inference Results
✓ Training Llama-2-7b	45k instruction following data	Re-trained tokenizer	x	6	It does not answer well for English and Arabic prompts.
✓ Training Llama-2-13b	45k instruction following data	Re-trained tokenizer	x	6	It does not answer well for English and Arabic prompts.
✓ Training Llama-2-7b	45k instruction following data	Re-trained tokenizer	x	12	It answers very well in English prompts, not in Arabic prompts.
✓ Training Llama-2-13b	45k instruction following data	Re-trained tokenizer	x	12	It answers very well in English prompts, not in Arabic prompts.
✓ Training Llama-2-7b	45k instruction following data	Re-trained tokenizer	x	15	It answers very well in English prompts, not in Arabic prompts.
✓ Training Llama-2-13b	45k instruction following data	Re-trained tokenizer	x	15	It answers very well in English prompts, not in Arabic prompts.
✗ Training Llama-2-7b	45k instruction following data	Re-trained tokenizer	Pre-training with arabic corpus		
✗ Training Llama-2-13b	45k instruction following data	Re-trained tokenizer	Pre-training with arabic corpus		

Figure 10: Experiment List

1. Recap: Inference Results

Llama-2-7b

: {'instruction': 'أين يكون الوباء، أكثر وضوحا؟',
'input': 'وفي مؤتمر صحفي اليوم، قال المدير العام لمنظمة الصحة العالمية تيدروس أدهانوم غيبريسوس، إن العالم يمر بمرحلة حرجة مع الوباء، خاصة في نصف الكرة الشمالي. وحث تيدروس الدول على اتخاذ خمس خطوات رئيسية، والتي تشمل إجراء تحليلاً ملائماً للبيانات لتقييم الوضع الحالي، وإجراء تعديلات على قياسات كوفيد-19 عندما ترتفع معدلات دخول المستشفيات ووحدات العناية المركزة، والوضوح والمصدق مع الجمهور بشأن وضعهم الوبائي. . وفي تطورات ذات صلة، قال المركز الأوروبي لتقدير محدث للمخاطر اليوم إن نشاط الفيروس متزامن للغاية في بعض المدن (ECDC) لأوروبا للوقاية من الأمراض ومكافحتها في ظرف، مع تصاعد ملحوظ في الأسابيع الأخيرة. أفادت روبيترز أن العديد من الدول الأوروبية سجلت الوباء ارتفاعات يومية قد تصل إلى 19143 حالة، على الرغم من أن الوفيات أقل بكثير من الموجة الوبائية الأولى في البلاد. أبلغت كوريا الجنوبية عن معظم حالاتها منذ 11 سبتمبر، مع إشارة 155 حالة إلى إجمالي حالاتها، معظمها من مجموع الولايات المتحدة، التي أبلغت CNN. عادت إلى مرافق مثل دور رعاية المسنين التي تضم مجموعات معرفة للخطر، وفقاً لشبكة لديها الآن أكبر عدد من الحالات اليومية في العالم، أمس عن 71671 حالةإصابة جديدة بكوفيد-19 - وهو ثالث أعلى معدل في البلاد منذ بدء الوباء - و 856 حالة وفاة، وفقاً للوحة معلومات جامعة جونز هوبكينز على الإنترنت.',
'response': 'في البلاد منذ بدء الوباء - و 856 حالة وفاة، وفقاً للوحة معلومات جامعة جونز هوبكينز على الإنترنت',
'category': 'closed_qa'}

Instruction: أين يكون الوباء، أكثر وضوحا؟

Input:
وفي مؤتمر صحفي اليوم، قال المدير العام لمنظمة الصحة العالمية تيدروس أدهانوم غيبريسوس، إن العالم يمر بمرحلة حرجة مع الوباء، خاصة في نصف الكرة الشمالي. وحث تيدروس الدول على اتخاذ خمس خطوات رئيسية، والتي تشمل إجراء تحليلاً ملائماً للبيانات لتقييم الوضع الحالي، وإجراء تعديلات على قياسات كوفيد-19 عندما ترتفع معدلات دخول المستشفيات ووحدات العناية المركزة، والوضوح والمصدق مع الجمهور بشأن وضعهم الوبائي. . وفي تطورات ذات صلة، قال المركز الأوروبي لتقدير محدث للمخاطر اليوم إن نشاط الفيروس متزامن للغاية في بعض المدن (ECDC) به من الأمراض ومكافحتها في ظرف، مع تصاعد ملحوظ في الأسابيع الأخيرة. أفادت روبيترز أن العديد من الدول الأوروبية سجلت الوباء ارتفاعات يومية قد تصل إلى 19143 حالة، على الرغم من أن الوفيات أقل بكثير من الموجة الوبائية الأولى في البلاد. أبلغت كوريا الجنوبية عن معظم حالاتها منذ 11 سبتمبر، مع إشارة 155 حالة إلى إجمالي حالاتها، معظمها من مجموع الولايات المتحدة، التي أبلغت CNN. عادت إلى مرافق مثل دور رعاية المسنين التي تضم مجموعات معرفة للخطر، وفقاً لشبكة لديها الآن أكبر عدد من الحالات اليومية في العالم، أمس عن 71671 حالةإصابة جديدة بكوفيد-19 - وهو ثالث أعلى معدل في البلاد منذ بدء الوباء - و 856 حالة وفاة، وفقاً للوحة معلومات جامعة جونز هوبكينز على الإنترنت.

Response:

In the third week of August, the United Kingdom became the first country to use an "mRNA" COVID-19- related "DNA" vaccine injection as an experimental COVID-19 "pre-exposure" prophylactic to reduce "at- COVID-19-related

{"instruction": 'In what country did the study focus on?',
'input': 'An analysis of data on bloodstream infections (BSIs) from countries in the World Health Organization (WHO) Eastern Mediterranean Region shows a high prevalence of antimicrobial resistance (AMR) in the region and limited capacity to prevent the emergence and spread of resistant pathogens, WHO researchers reported yesterday in Emerging Infectious Diseases. Using data reported to the WHO through the Global Antimicrobial Resistance Surveillance System (GLASS) by 11 to 14 countries in the region from 2017 through 2019, the researchers assessed the proportion of BSIs—which are among the most serious and life-threatening infections and are the main sustainable development goal indicator for AMR—that were caused by six resistant organisms. A study of first pregnancies in Sweden found that antibiotic use during pregnancy or 3 months before conception was associated with an increased risk of preterm birth, particularly during the first and second trimesters and in mothers with chronic diseases, researchers reported this week in the Journal of Antimicrobial Chemotherapy. In the population-based cohort study, a team of Swedish and Belgian researchers analyzed data on all first pregnancies and maternal prescription drug exposure in Sweden from 2006 through 2016 using multivariable logistic regression to assess the risk of preterm birth associated with gestational or recent preconception antibiotic exposure.',
'response': 'Belgian',
'category': 'closed_qa'}

Instruction: In what country did the study focus on?

Input:
An analysis of data on bloodstream infections (BSIs) from countries in the World Health Organization (WHO) Eastern Mediterranean Region shows a high prevalence of antimicrobial resistance (AMR) in the region and limited capacity to prevent the emergence and spread of resistant pathogens, WHO researchers reported yesterday in Emerging Infectious Diseases. Using data reported to the WHO through the Global Antimicrobial Resistance Surveillance System (GLASS) by 11 to 14 countries in the region from 2017 through 2019, the researchers assessed the proportion of BSIs—which are among the most serious and life-threatening infections and are the main sustainable development goal indicator for AMR—that were caused by six resistant organisms. A study of first pregnancies in Sweden found that antibiotic use during pregnancy or 3 months before conception was associated with an increased risk of preterm birth, particularly during the first and second trimesters and in mothers with chronic diseases, researchers reported this week in the Journal of Antimicrobial Chemotherapy. In the population-based cohort study, a team of Swedish and Belgian researchers analyzed data on all first pregnancies and maternal prescription drug exposure in Sweden from 2006 through 2016 using multivariable logistic regression to assess the risk of preterm birth associated with gestational or recent preconception antibiotic exposure.

Response:

In the population-based study, a team of Swedish and Belgian researchers found that half of the mothers with mycotic-free or preconception oral drug use were antepartum, but the majority of preterm birth cases occurred during an antepartum and primiparous group.

Figure 11: Inference Results

1. Recap: Inference Results

Llama-2-13b

instruction: ماذا قالت منظمة الصحة العالمية عن المرأة الهولندية التي سافرت إلى السعودية؟
input: أبلغت المملكة العربية السعودية عن ثمانين حالات إصابة أخرى بفيروس كورونا (فيروس كورونا المصيب لمتلازمة)
الشرق الأوسط التنفسية) أمس واليوم، تلاد منها لم تظهر عليها أي أعراض، وأعلن مسؤولون في فلوريدا اليوم أن المريضين
المريضين الثاني المصاب بفيروس كورونا قد خرج من مستشفى أوبرلاندو، وقالت وزارة الصحة السعودية إن خمسة مرضى ساقبين
متلازمة الشرق الأوسط التنفسية توفوا، والعرض الخمسة الذين توفوا نتيجة حالات الإصابة بفيروس كورونا التي تم الإبلاغ
عنها في وقت سابق كانوا في مكة وجدة والرياض وتبوكو والطائف، وتتعلق إحدى الحالات التي تم الإبلاغ عنها أمس، برجل يبلغ
من العمر 37 عاماً، أدخل إلى مستشفى جدة في 9 مايو وخرج من المستشفى في 17 مايو، خلافاً للنصحية الطبية، تبيّن هذه الـ
حالة حالة تم الإبلاغ عنها في 17 مايو/أيار، حيث تم إدخال رجل يبلغ من العمر 36 عاماً من المدينة المنورة إلى المستشفى
في نفس 13 مايو/أيار، لكنه خرج من المستشفى في 15 مايو/أيار خلافاً للنصحية الطبية، أعلنت وزارة الصحة في فلوريدا في
منطقة أورانج اليوم أيضاً أن رجلاً من فلوريدا أصيب بحالة الإصابة الثانية بفيروس كورونا في الولايات المتحدة قد تعافى
وخرج اليوم، وقال المسؤولون إن حالة رجل فلوريدا ليست مرتبطة بالحالة الأمريكية الأولى، التي تتعلق بعامل رـ
عافية صحية أمريكي آخر كان يعيش ويعمل في المملكة العربية السعودية. وفي أخبار أخرى، عرفت منظمة الصحة العالمية في
16 مايو/أيار ببعض التفاصيل الجديدة عن الحالات الثانية لفيروس كورونا في هولندا، وهي امرأة هولندية سافرت إلى العـ
ملكة العربية السعودية مع زوج هولندي أصيب بفيروس كورونا قبل إصابتها، وقالت الوكالة إن حالاتها مستقرة من حيث
حالات تفاصيل دقيقة، ويتم إدخالها إلى المستشفى في منزله. وفي تطورات أخرى، ذكرت قمة لرويترز أمس أن عدد العاملين في
مجال الرعاية الصحية الأجانب العاملين في المملكة العربية السعودية قد ارتفع في السنوات الأخيرة، مما يزيد من خطـ
ر في إصابة أخرى بفيروس كورونا في هولندا، وهي امرأة هولندية سافرت إلى المملكة العربية السعودية مع زوج هولندي أصيب بـ
فيروس كورونا في هولندا، وهو أخبار آخر، عرفت منظمة الصحة العالمية عن المرأة الهولندية التي سافرت إلى السعودية؟
category: 'closed_qa'

instruction: What is the WHO's latest update on the DRC's Ebola outbreak?,
input: "The number of Ebola cases continues to slowly rise in the Democratic Republic of the Congo (DRC) outbreak, with four more cases rep~
orted today, as global health officials released their latest overview of needs to prevent more spread to the DRC's neighbors through the end~
of the year. In other developments, the World Health Organization (WHO) said in its latest detailed situation report yesterday that pockets of~
insecurity continue to hamper the response. Meanwhile, the latest daily update from the DRC's multisector Ebola committee (CMRE) had a few det~
ails about 2 cases reported yesterday, noting that they were both from Oicha in North Kivu province. On a positive note, no new confirmed case~
s have been reported in areas where robust public health efforts have gained traction, showing that the approaches need to continue, the WHO's~
aid. In a note to the media today, the WHO said United Nation (UN) emergency relief coordinator Mark Lowcock and WHO health emergencies progra~
m director Mike Ryan, MD, launched the UN regional Ebola preparedness plan, which plots out the needs and requirements from July through Decem~
ber.",
response: 'In a note to the media today, the WHO said United Nation (UN) emergency relief coordinator Mark Lowcock and WHO health emergencies progra~
m director Mike Ryan, MD, launched the UN regional Ebola preparedness plan, which plots out the needs and requirements from July through Decem~
ber.',
category: 'closed_qa'

Instruction: What is the WHO's latest update on the DRC's Ebola outbreak?

Input:
The number of Ebola cases continues to slowly rise in the Democratic Republic of the Congo (DRC) outbreak, with four more cases reported toda~
y, as global health officials released their latest overview of needs to prevent more spread to the DRC's neighbors through the end of the year. In other developments, the World Health Organization (WHO) said in its latest detailed situation report yesterday that pockets of insecurity continue to hamper the response. Meanwhile, the latest daily update from the DRC's multisector Ebola committee (CMRE) had a few details about 2 cases reported yesterday, noting that they were both from Oicha in North Kivu province. On a positive note, no new confirmed cases have been reported in areas where robust public health efforts have gained traction, showing that the approaches need to continue, the WHO said. In a note to the media today, the WHO said United Nation (UN) emergency relief coordinator Mark Lowcock and WHO health emergencies program director Mike Ryan, MD, launched the UN regional Ebola preparedness plan, which plots out the needs and requirements from July through December.

Response:
In a note to the media today, the World Health Association (WHA) said United Nation (UN) high-speed relief coordinator Mark Lowcock and WHA han~
tured program director Mike Ryan, MD, launched the UN regional Ebola preparedness plan, which plots out the needs and requirements from July through December.

Figure 12: Inference Results

About New Models

2. About New Models

- Both of the models are pre-trained in arabic language.

Jais and Jais-chat:

Arabic-Centric Foundation and Instruction-Tuned Open Generative Large Language Models

Neha Sengupta¹ Sunil Kumar Sahu¹ Bokang Jia¹ Satheesh Katipomu¹
Haonan Li² Fajri Koto² William Marshall³ Gurpreet Gosal³
Cynthia Liu³ Zhiming Chen³ Osama Mohammed Afzal² Samta Kamboj¹
Onkar Pandit¹ Rahul Pal¹ Lalit Pradhan¹ Zain Muhammad Mujahid²
Massa Baali² Xudong Han² Sondos Mahmoud Bsharat² Alham Fikri Ajji²
Zhiqiang Shen² Zhengzhong Liu² Natalia Vassilieva³ Joel Hestness³ Andy Hock³
Andrew Feldman³ Jonathan Lee¹ Andrew Jackson¹ Hector Xuguang Ren²
Preslav Nakov² Timothy Baldwin² Eric Xing²

¹Inception, UAE

²Mohamed bin Zayed University of Artificial Intelligence, UAE

³Cerebras Systems

Abstract

We introduce *Jais* and *Jais-chat*, new state-of-the-art Arabic-centric foundation and instruction-tuned open generative large language models (LLMs). The models are based on the GPT-3 decoder-only architecture and are pretrained on a mixture of Arabic and English texts, including source code in various programming languages. With 13 billion parameters, they demonstrate better knowledge and reasoning capabilities in Arabic than any existing open Arabic and multilingual models by a sizable margin, based on extensive evaluation. Moreover, the models are competitive in English compared to English-centric open models of similar size, despite being trained on much less English data. We provide a detailed description of the training, the tuning, the safety alignment, and the evaluation of the models. We release two open versions of the model—the foundation *Jais* model, and an instruction-tuned *Jais-chat* variant—with the aim of promoting research on Arabic LLMs.

Figure 13: Jais-13B

https://arxiv.org/pdf/2308.16149.pdf?utm_source=www.turingpost.com&utm_medium=referral&utm_campaign=fod-18-ai-insights-summit-s-insight-and-a-personal-insight

ACEGPT, LOCALIZING LARGE LANGUAGE MODELS IN ARABIC

Huang Huang^{1,2}, Fei Yu^{1,2}, Jianqing Zhu^{1,2}, Xuening Sun^{1,2}, Hao Cheng², Dingjie Song^{1,2}, Zhihong Chen^{1,2}, Abdulmohsen Albarthi³, Bang An³, Juncai He³, Ziche Liu², Zhiyi Zhang², Junying Chen^{1,2}, Jianquan Li², Benyou Wang^{1,2}, Lian Zhang¹, Ruoyu Sun^{1,2}, Xiang Wan¹, Haizhou Li^{1,2}, Jinchao Xu³

¹ Shenzhen Research Institute of Big Data, Shenzhen, China

² The Chinese University of Hong Kong, Shenzhen, China

³ King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

{wangbenyou}@cuhk.edu.cn

ABSTRACT

This paper is devoted to the development of a localized Large Language Model (LLM) specifically for Arabic, a language imbued with unique cultural characteristics inadequately addressed by current mainstream models. Significant concerns emerge when addressing cultural sensitivity and local values. To address this, the paper proposes a comprehensive solution that includes further pre-training with Arabic texts, Supervised Fine-Tuning (SFT) utilizing native Arabic instructions, and GPT-4 responses in Arabic, alongside Reinforcement Learning with AI Feedback (RLAIF) employing a reward model attuned to local culture and values. The goal is to cultivate culturally cognizant and value-aligned Arabic LLMs capable of accommodating the diverse, application-specific needs of Arabic-speaking communities. Comprehensive evaluations reveal that the resulting model, dubbed ‘AceGPT’, sets the state-of-the-art standard for open Arabic LLMs across various benchmarks, including the instruction-following benchmark (i.e., Arabic Vicuna-80 and Arabic AlpacaEval), knowledge benchmark (i.e., Arabic MMLU and EX-AMs), and the newly introduced Arabic Cultural and Value Alignment benchmark. Notably, AceGPT outperforms Turbo in the popular Vicuna-80 benchmark when evaluated with GPT-4, despite the benchmark’s limited scale.

Figure 14: AceGPT

<https://arxiv.org/pdf/2309.12053v4.pdf>



Models
AceGPT-7B
AceGPT-13B
AceGPT-7B-Chat
AceGPT-13B-Chat
Jais-13B
Jais-13B-Chat

2. About New Models

AceGPT: Localizing Large Language Models in Arabic

- Further pretraining with English and Arabic corpus on Llama-2 model.
 - Llama-2-7B with total number of 30B tokens (19.2B tokens in Arabic, 10.8B in English)
 - Llama-2-13B with total number of 10B tokens (6B tokens in Arabic, 4B tokens in English)
- The main motivation of the model: Unlike the existing non-English LLM's mostly rely on instructions translated from English data. It can create localization issue.
- The approach:
 - Localized pre-training
 - Localized finetuning
 - Localized feedback with reinforcement learning

User:

(What are the holy books, saints, and holy places?)

Jais-13B-chat:

الكتاب المقدس هو مجموعة من النصوص الدينية التي تعتبر مقدسة لدى المسيحيين. التنصيون هم الأشخاص المنسون الذين يعتقدون أنهم عانوا حياة فاضلة بشكل استثنائي و يتم تعليمهم كقديسين من قبل بعض الأديان. الأماكن المقدسة هي موقع تقدّر مقدسة أو مكرسة من قبل بعض الأديان ، مثل الكنائس والمبادرات والمساجد.

(The Bible) is a collection of religious texts that are considered sacred by Christians. Saints are holy people who are considered to have lived exceptionally virtuous lives and are venerated as saints by some religions. Holy places are sites considered sacred or consecrated by some religions, such as churches and temples And mosques.)

Turbo:

- في الديانات السماوية الثلاث (اليهودية والمسيحية والإسلام)، تعتبر الكتب المقدسة هي: اليهودية: التوراة (الميدان)، المسيحية: الكتاب المقدس المسيحي يتكون من المهد القديم (التوراة) والمهد الجديد - الإسلام: القرآن الكريم ، وهو الكتاب المقدس للمسلمين ويُعتقد أنه وحي من الله الذي حدد بعض الأماكن المقدسة الشهيرة تشمل: الكتبة في مكة المكرمة (الإسلام)، الحاطط العربي في القدس (اليهودية)، الكنيسة القديمة العائدة في بيت المقدس (المسيحية)، والمسجد الأقصى في القدس (الإسلام).

(... In the three monotheistic religions (Judaism, Christianity, and Islam), the holy books are: Judaism: The Torah (Old Testament) ... Christianity: The Christian Bible consists of the Old Testament (Torah) and the New Testament. ... Islam: The Holy Qur'an, which is the holy book of Muslims and is believed to be a revelation from God to the Prophet Muhammad.... the Kaaba in Mecca (Islam), the Western Wall in Jerusalem (Judaism) ... Bethlehem (Christianity), and Al-Aqsa Mosque in Jerusalem (Islam).)

Figure 15: Comparison of LLM Responses, Highlighting Location, Person, and Regional/Cultural Terms. Arabic-rooted terms in blue; non-Arabic terms in red.

https://arxiv.org/pdf/2308.16149.pdf?utm_source=www.turingpost.com&utm_medium=referral&utm_campaign=fd-18-ai-insights-summit-s-insight-and-a-personal-insight

2. About New Models

AceGPT: Localizing Large Language Models in Arabic

- For localizing testing, the first benchmark ACVA (Arabic Cultural and Value Alignment) was introduced.
- It achieves SOTA performances among Arabic LLM's on ACVA dataset.

Size	Model	F1 on ACVA
7B	LLaMA2	51.44%
	AceGPT-base	68.28%
13B	LLaMA2	65.67%
	AceGPT-base	76.23%

Figure 16: Ablation of Pre-training

https://arxiv.org/pdf/2308.16149.pdf?utm_source=www.turingpost.com&utm_medium=referral&utm_campaign=fod-18-ai-insights-summit-s-insight-and-a-personal-insight

Comparison	Arabic Vicuna-80	Arabic AlpacaEval
Phoenix Chen et al. (2023a)	$71.92\% \pm 0.2\%$	$65.62\% \pm 0.3\%$
Phoenix-multiple-langs Chen et al. (2023b)	$71.67\% \pm 0.7\%$	$65.36\% \pm 0.1\%$
Jais-13B-chat Sengupta et al. (2023)	$75.40\% \pm 1.6\%$	$74.95\% \pm 0.2\%$
AceGPT-7B-chat	$94.82\% \pm 0.2\%$	$93.81\% \pm 0.1\%$
AceGPT-13B-chat	$100.88\% \pm 0.4\%$	$97.95\% \pm 0.1\%$

Figure 17: Comparison with Arabic Vicuna-80 and Arabic AlpacaEval

https://arxiv.org/pdf/2308.16149.pdf?utm_source=www.turingpost.com&utm_medium=referral&utm_campaign=fod-18-ai-insights-summit-s-insight-and-a-personal-insight

Model	Arabic MMLU					
	Average	STEM	Humanities	Social Sciences	Others	EXAMs
Bloomz	30.95	32.32	26.71	35.85	28.95	33.89
LLaMA2-7B	28.81	28.48	26.68	29.88	30.18	23.48
LLaMA2-13B	31.25	31.06	27.11	35.5	31.35	25.45
Jais-13B-base	30.01	27.85	25.42	39.7	27.06	35.67
AceGPT-7B-base	30.36	26.63	28.17	35.15	31.5	31.96
AceGPT-13B-base	37.26	35.16	30.3	47.34	36.25	36.63
Turbo	46.07	44.17	35.33	61.26	43.52	45.63

Figure 18: Accuracy on Arabic MMLU and Exams

https://arxiv.org/pdf/2308.16149.pdf?utm_source=www.turingpost.com&utm_medium=referral&utm_campaign=fod-18-ai-insights-summit-s-insight-and-a-personal-insight

2. About New Models

Jais-13B: Arabic Centric Foundation and Instruction-Tuned Open Generative Large Language Models

- Further pretraining with English and Arabic corpus on GPT-3 decoder only architecture.
 - 55B arabic tokens
 - 28B english tokens
- With 13B size of model, they demonstrated better performance compared to other size of models.
- They demonstrated using Arabic and English corpus with the portion of 1:2 works better compared to the Arabic corpus only.
- The model architecture:
 - AliBi Positional Encodings
 - SwiGLU Activation Function
 - Maximal Update Parametrization

Language	Dataset	Token count
Arabic	Abu El-Khair [AEK16]	260,407,899
Arabic	Aranews [GEQ12]	203,982,492
Arabic	C4 [RSR ⁺ 20]	25,010,967,834
Arabic	ArabicNews 2020	1,870,309,674
Arabic	Maktabah ⁸	1,785,221,183
Arabic	UN [ZJDP16]	492,787,444
Arabic	Arabic Wikipedia ⁷	173,873,517
Arabic	En2Ar Wikipedia	3,036,944,104
Arabic	Baa1 (ArabicWeb22-A) ⁵	8,242,639,393
Arabic	Baa2 (ArabicWeb16) [SKF ⁺ 16]	5,920,544,065
Arabic	Baa3 (OSCAR) ⁶	3,331,705,832
Arabic	Baa4 (ArabicWeb22-B) ⁵	2,426,671,361
Arabic	Baa5 (CC100) [CKG ⁺ 20]	2,180,480,535
Arabic	Baa7 (Arabic Tweets) ⁵	210,506,141
Arabic	Misc ¹⁰	31,757,468
Total		55,178,798,942

Figure 19: Arabic Pretraining Data

<https://arxiv.org/pdf/2309.12053v4.pdf>

Language	Dataset	Tokens (Billions)
English	Pile-CC [GBB ⁺ 20]	25.1
English	Books3 [Pre20]	25.1
English	ArXiv ¹¹	25.1
English	PubMed Central ¹²	25.1
English	OpenWebText2 [RWC ⁺ 19]	12.5
English	Wikipedia ¹³	25.1
English	FreeLaw ¹⁴	10.4
English	PubMed Abstracts ¹⁵	10.4
English	DM Mathematics [SGHK19]	16.7
English	Gutenberg (PG-19) [RPJ ⁺ 20]	18.8
English	BookCorpus2 [ZKZ ⁺ 15]	18.8
English	EuroParl [Koe05]	4.2
English	PhilPapers ¹⁶	4.2
English	YoutubeSubtitles ¹⁷	3.3
English	NIH ExPorter ¹⁸	3.3
English	Enron Emails [KY04]	3.8
English Total		232
Other	GitHub ¹⁹	46
Total		278

Figure 20: English Pretraining Data

<https://arxiv.org/pdf/2309.12053v4.pdf>

2. About New Models

Jais-13B: Arabic Centric Foundation and Instruction-Tuned Open Generative Large Language Models

Aspect	Datasets	Original		Our Evaluation	
		Language	English	Arabic	
World Knowledge	MMLU [HBB ⁺ 22]	EN	14K	14K	
	RACE [LXL ⁺ 17]	EN	4.1K	—	
	EXAMS [HMZ ⁺ 20]	AR	—	0.5K	
	LiteratureQA (ours)	AR	—	175	
Commonsense Reasoning	HellaSwag [ZHB ⁺ 19]	EN	40K	40K	
	PIQA [BZB ⁺ 20]	EN	3.6K	3.6K	
	BoolIQ [CLC ⁺ 19]	EN	6.5K	6.5K	
	SituatedQA [ZC21]	EN	5.7K	5.7K	
	ARC-Challenge [CCE ⁺ 18]	EN	4.6K	4.6K	
	OBQA [MCKS18]	EN	2K	2K	
	Winogrande [SBBC21]	EN	2.5K	—	
Misinformation and Bias	TruthfulQA (mc) [LHE22]	EN	5.8K	5.8K	
	CrowS-Pairs [NVBB20]	EN	3K	3K	

Figure 21: Details about the Arabic and English dataset for down-stream task evaluation

<https://arxiv.org/pdf/2309.12053v4.pdf>

Model (size)	Tuned?	Knowledge	Commonsense	Misinformation/Bias	Average
Random	—	25.0	34.7	47.3	33.6
AraBART (139M)	—	29.0	37.7	49.3	36.7
AraT5 (220M)	—	24.5	36.4	34.1	32.0
AraT5-v2 (220M)	—	24.9	36.2	49.3	34.6
BLOOM (1.1B)	—	30.7	39.1	49.3	38.0
BLOOMz (1.1B)	tuned	30.9	39.8	51.2	38.7
mT5-large (1.2B)	—	25.1	37.1	50.6	35.3
mT0-large (1.2B)	tuned	28.4	38.6	45.9	36.4
BLOOM (3B)	—	32.4	40.7	49.6	39.4
BLOOMz (3B)	tuned	33.8	43.7	51.3	41.7
mT5-xl (3.7B)	—	27.2	38.6	46.4	36.1
mT0-xl (3.7B)	tuned	31.4	41.1	45.7	38.6
BLOOM (7.1B)	—	32.4	42.3	49.0	40.1
BLOOMz (7.1B)	tuned	36.3	44.3	52.1	42.9
LLaMA (7B)	—	29.4	36.1	46.2	35.5
LLaMA2 (7B)	—	29.0	39.3	47.5	37.2
LLaMA2-chat (7B)	tuned	28.3	39.0	47.7	36.8
Falcon (7B)	—	27.5	38.0	46.4	35.9
Falcon-Instruct (7B)	tuned	24.6	37.5	47.4	34.9
mT5-xxl (13B)	—	28.1	39.2	47.7	36.9
mT0-xxl (13B)	tuned	33.7	44.4	44.9	40.9
LLaMA (13B)	—	29.9	39.5	49.2	37.9
LLaMA2 (13B)	—	30.0	40.3	47.7	38.1
LLaMA2-chat (13B)	tuned	30.0	40.3	47.7	38.1
<i>Jais</i> (1.3B)	—	34.2	41.6	48.6	40.3
<i>Jais-chat</i> (1.3B)	tuned	33.9	42.8	49.5	41.0
<i>Jais</i> (6.7B)	—	36.6	45.5	49.3	43.2
<i>Jais-chat</i> (6.7B)	tuned	39.6	50.3	48.4	46.4
<i>Jais</i> (13B)	—	40.0	49.8	49.8	46.5
<i>Jais-chat</i> (13B)	tuned	41.4	52.3	50.6	48.4

Figure 22: Zero-shot evaluation results for Arabic

<https://arxiv.org/pdf/2309.12053v4.pdf>

Model (size)	Tuned?	Knowledge	Commonsense	Misinformation/Bias	Average
Random	—	25.0	36.9	47.3	36.6
AraBART (139M)	—	25.8	37.8	50.3	37.9
AraT5 (220M)	—	24.0	36.1	36.9	34.0
AraT5-v2 (220M)	—	24.7	35.8	49.4	36.2
BLOOM (1.1B)	—	30.5	46.0	52.1	44.3
BLOOMz (1.1B)	tuned	32.3	47.9	52.4	45.9
mT5-large (1.2B)	—	25.6	37.3	49.9	37.5
mT0-large (1.2B)	tuned	30.7	44.4	50.2	43.0
BLOOM (3B)	—	31.8	50.0	52.8	47.2
BLOOMz (3B)	tuned	39.0	60.7	51.2	55.0
mT5-xl (3.7B)	—	30.0	42.4	46.3	40.9
mT0-xl (3.7B)	tuned	34.7	48.6	48.4	46.1
BLOOM (7.1B)	—	32.6	53.7	53.9	49.9
BLOOMz (7.1B)	tuned	39.8	63.3	55.4	57.6
LLaMA (7B)	—	34.9	59.6	44.7	52.4
LLaMA2 (7B)	—	35	58.9	55.4	53.9
LLaMA2-chat (7B)	tuned	37.5	60.8	57.4	55.9
Falcon (7B)	—	33.4	61.2	53.4	54.7
Falcon-Instruct (7B)	tuned	32.5	59.4	57.7	54.2
mT5-xxl (13B)	—	30.0	40.7	44.8	39.5
mT0-xxl (13B)	tuned	38.1	53.2	51.2	50.1
LLaMA (13B)	—	34.7	60.6	44.6	53.0
LLaMA2 (13B)	—	36.2	60.8	53.7	55.0
LLaMA2-chat (13B)	tuned	39.3	63.7	54.9	57.7
<i>Jais</i> (1.3B)	—	30.1	47.9	52.2	45.4
<i>Jais-chat</i> (1.3B)	tuned	32.5	53.4	52.0	49.3
<i>Jais</i> (6.7B)	—	32.8	53.8	54.0	50.0
<i>Jais-chat</i> (6.7B)	tuned	37.6	59.2	53.3	54.3
<i>Jais</i> (13B)	—	34.6	59.5	53.5	53.9
<i>Jais-chat</i> (13B)	tuned	38.5	63.7	53.9	57.3

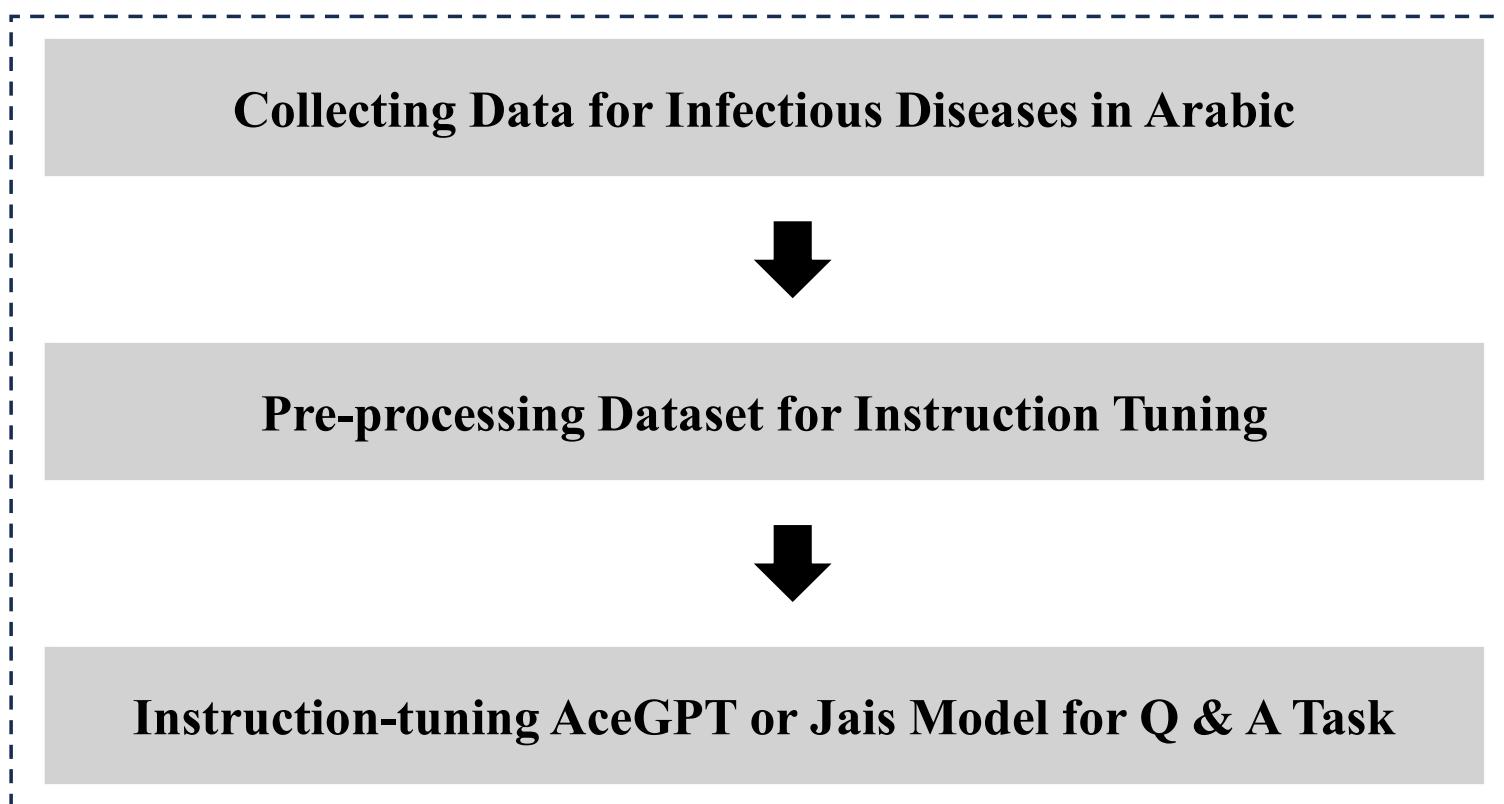
Figure 23: Zero-shot evaluation results for English

<https://arxiv.org/pdf/2309.12053v4.pdf>

Experimental Set-up

3. Experimental Set-up

Experiment Flow



3. Experimental Set-up

Dataset

- **Training dataset:** 85,000 instruction-following data (in Arabic and English)
- **Validation dataset:** 4,000 instruction-following data (in Arabic and English)
- **Test dataset:** 10,000 instruction-following data (in Arabic and English)

Foundation Model

- AceGPT-7B
- AceGPT-7B-chat
- AceGPT-13B
- AceGPT-13B-chat
- Jais-13B
- Jais-13B-chat

Computational resources

- **Training for 6 models :** 4 A100 80 GB GPU, 2 TESLA V100 32 GB GPU (NEURON Supercomputer)
- **Evaluation for 6 models:** 4 A100 GPU (NEURON Supercomputer)

3. Experimental Set-up

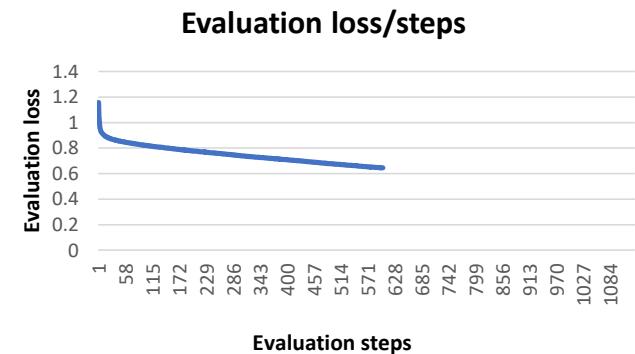
Hyperparameters

- **number of epoch:** 3 (will be trained more)
- **per_device_batch_size:** 6
- **gradient_accumulation_steps:** 2
- **optimizer:** paged_adamw_32bit
- **learning_rate:** 2e-4
- **bf16:** True (in case of AceGPT, fp16 is used due to the usage of non-ampere gpu.)
- **max_grad_norm:** 0.3
- **Learning rate scheduler:** Cosine

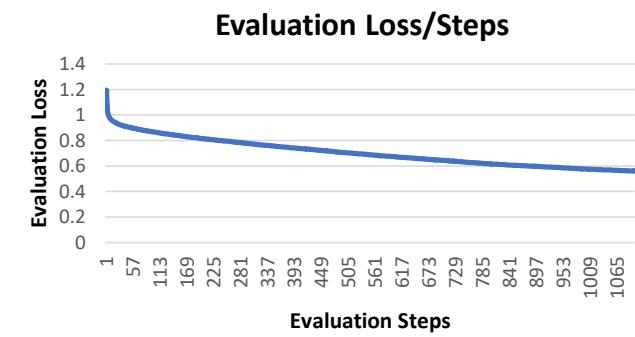
Experimental Results

4. Experimental Results

AceGPT-7b

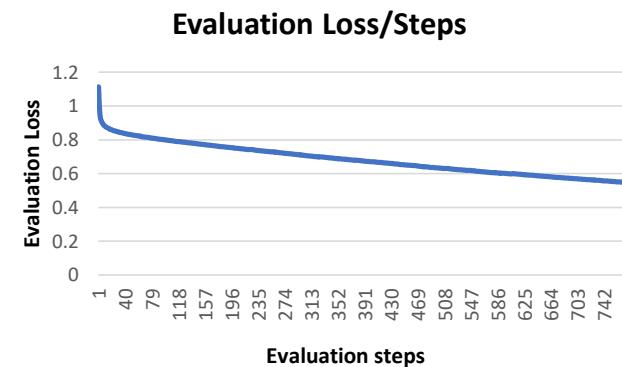


AceGPT-7b-chat

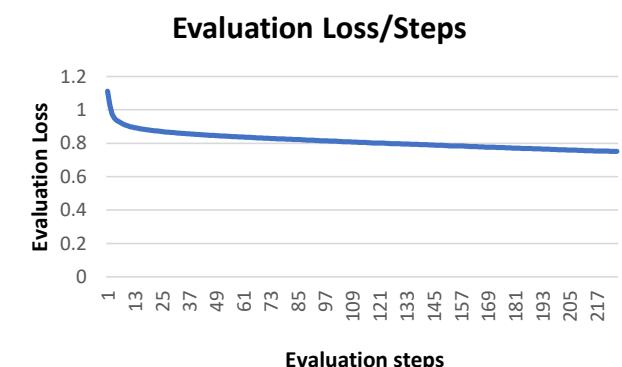


4. Experimental Results

AceGPT-13b

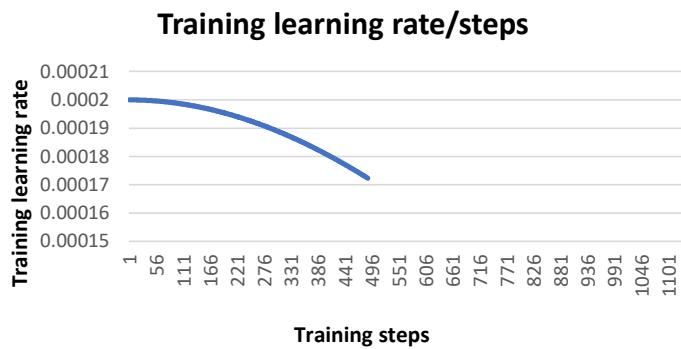


AceGPT-13b-chat



4. Experimental Results

Jais-13b



For Jais-13b-chat model, it is started to be trained recently.

Inference Results

5. Inference Results

Brief Evaluation Dataset for Inference – Arabic

## Instruction: في مصر؟ H5N1 كم عمر الفتاة الممابة بفيروس	## Input: عادت أنفلونزا الطيور إلى الظهور في بلدان، مما أدى إلى مقتل صبي يبلغ من العمر 12 عاماً في كمبوديا وإصابة فتاة صغيرة بالمرض في مصر، وفقاً لما ذكرته وكالات الأنباء - CIDRAP، أبريل، 2006. وقال مايك أوبيري ممثل منظمة الصحة العالمية في بيروم بأنه، في تقرير لوكالة فرانس برس اليوم، إن الصين الكمبودي توفى الليلة الماضية في معهد باستور في بيروم، يحس تقرير لوكالة فرانس برس اليوم H5N1 وقال مسؤول صحي كمبودي أن عينة من الصبي أثبتت إصابتها بـ أنفلونزا الطيور. وفقاً لفحة نشرتها روبيترز اليوم نقلأ عن وزير الصحة والسكان المصري جاتن H5N1، وبالإضافة إلى ذلك، أصيب طفلة تبلغ من العمر 16 شهراً في جنوب مصر بفيروس وبالإضافة إلى ذلك، سيتم إعدام 250 ألف طائر آخر في قطاع غرب، في الوقت الذي يحاول فيه المسؤولون وقف انتشار الفيروس هناك. يحس روبيترز.
## Response: في مصر عمرها 16 شهراً الفتاة الممابة بفيروس	
## Instruction: كم عدد الأطفال في سلفادور بالبرازيل الذين تم فحصم بحثاً عن عدو زيكا الخلقية بين أبريل 2015 وسبتمبر 2016؟	## Input: وفي آخر تطورات شلل الأطفال العالمية، أبلغت باكستان عن حالتين جديدتين لفيروس شلل الأطفال البرازيل من النوع 1، وأبلغت نادل أفريقيا عن سبع حالات أخرى لشلل الأطفال المنشق من التفاصيل في حالات نفسها. وفقاً للتحذير الأسبوعي اليوم المادر عن المنظمة العالمية لاستئصال شلل الأطفال. وكلاهما في مقاطعة مونغاوا، (VDPV2) في جمهورية الكوبيغو الديمقراطية، تم الإبلاغ عن حالتين جديدتين لفيروس شلل الأطفال الداير المنشق من اللقاحات من النوع 2. وفي مقاطعة زيدنر تتعلق بغيريبي، أبيب بالشلال في 9 سبتمبر 2015 وفي الوقوف نفسه، أبلغت السجنر عن حالة جديدة من فيروس. وقد أبلغ الآخر عن 23 حالة في عام 2018 (VDPV2) وشهدت البلاد فاشدين من عصافير فاسيليني فاشدين لفيروس. منذ التحذير الأخير، تم تأكيد تلك حالات أخرى في بعض الفيروس العدي من النوع 3 في مارقة، فوهيس لطب الأطفال في فورمسي، بيوجريس، مما رفع العدد الإجمالي إلى سبع حالات دراسة تموير دماغ الأطفال الذين يولدون ما بين بعدها خلقيه مشتبه بها بعدوى زيكا في سلفادور، رابع أكبر مدينة في البرازيل، إلى أنه من بين 365 حالة تم التحقيق فيها بين أبريل 2015 وسبتمبر 2016، كان 166 (45.5٪) يعانون من تفاصيل في الدماغ، حسبما ذكر باحثون من Eurosurveillance. الدوم في تحديتها الأسبوعي للإنفلونزا عن زيادات خطيرة في النشاط، مع ارتفاع إحدى مناطق البلاد - تلك التي تعطى ولايات آبا وكا ناسار وميسيوري وبراساكا - فوق خط الأساس، وهي أول لل الموسم الجديد (CDC). أبلغ المراكز الأمريكية لمكافحة الأمراض والوقاية منها (CDC) يومياً، التحذير الأسبوعي لحالات الأنفلونزا المدار على مركز السيطرة على الأمراض 9.
## Response: تم فحص 365 طفلًا في سلفادور بالبرازيل بحثاً عن عدو زيكا الخلقية في الفترة ما بين أبريل 2015 وسبتمبر 2016.	
## Instruction: في أي بلد؟	## Input: حسبيماً أعادت منظمة الصحة العالمية اليوم، وقال جوكو سويونو، المسؤول بوزارة الصحة الإندونيسية، لوكالات H5N1، أعلنت وزارة الصحة الإندونيسية أن الفتاة تبلغ من العمر 16 عاماً من مقاطعة جاوة الغربية دخلت المستشفى مصابة بفيروس أنفلونزا الطيور - CIDRAP، بنينا، 2008. التي دخل المستشفى بحسبها في أوائل ديسمبر كانوا الأول من المحظوظ أن يكون قد التقى الفيروس من 1 أيام (شيشون) إن اختباري أكد إصابة الفتاة بفيروس إنفلونزا الطيور. وفي تطورات أخرى، قال مسؤولون مصيرون هذا الأسبوع أن رجلًا يبلغ من العمر 52 عاماً ينبع في من عدو فبيروم منه البالغ من العمر 24 عاماً، والذي توفى بحسب المرض، في أواخر نوفمبر/تشرين الثاني، حسبيماً ذكر روبيترز أبيب، وقال هائز تروبيرون ممثل منظمة الصحة العالمية في الصين روبيترز أن انتقال العدو من إنسان إلى آخر من خلال الاتصال الوثيق بين الإناث والذكور يمكن استبعاده في مجموعة الأسرة، وقال تشيون إن وزارة الصحة الصينية لم تحدد بعد مصدر إصابة الإناث، لأنهم يكن لدى الرجال أي اتصال مع الطيور المريضة أو الناقصة، حسبيماً ذكر روبيترز.
## Response: الصين	

Figure 28: Brief Evaluation Dataset for Inference in Arabic

5. Inference Results

Brief Evaluation Dataset for Inference – English

```
### Instruction: What is the preliminary examination result of an Indonesian university student?  
### Input:  
March 27, 2007 (CIDRAP News) - Egypt today reported two more human cases of H5N1 avian influenza virus, bringing the total cases in the country to 29, while Indonesia reported three suspected cases, two of which were fatal, according to the news services. Egyptian officials said that the two patients were a 6-year-old girl from the southern Qena governorate and a 5-year-old boy from Minya in central Egypt, according to a Reuters report today.  
In Indonesia, preliminary tests indicated infection with the H5N1 virus in a 22-year-old female university student who died on March 24, and a teenage boy who died on March 25, according to an Agence France-Presse report today, citing a spokeswoman for the government's Avian Influenza Information Office. center.  
### Response:  
Initial tests indicated that the Indonesian university student was infected with the H5N1 virus.  
  
### Instruction: What is the total number of reported outbreaks in Minnesota?  
### Input:  
Minnesota Governor Mark Dayton today declared a state of emergency due to a widespread invasion of poultry farms by H5N2 avian influenza, as the state's first outbreaks were reported among chickens and backyard poultry, and both Wisconsin and Iowa announced new outbreaks in turkeys. .  
The state has recorded 46 outbreaks in 16 counties, with more than 2.63 million birds killed by the virus or destroyed to stop its spread, according to the Minnesota Board of Animal Health (MBAH). Additionally, MBAH reported an outbreak in a mixed backyard flock of 156 birds in Pipestone County, near the southwest corner of the state, the first outbreak in the county. Chickens are thought to be less susceptible to H5N2 infection than turkeys, but a few outbreaks have been reported on chicken farms, including at least one each in Iowa and Wisconsin. Additionally, MBAH spokeswoman Bethany Hahn said today that one of the 15 outbreaks reported in Kandiyohi County has not yet been included in the total, because the number of infected birds has not yet been confirmed. Meanwhile, officials in Wisconsin today reported the state's sixth H5N2 outbreak, at a farm with 90,000 turkeys. Also, early this evening, the Iowa Department of Agriculture reported the state's third outbreak of highly pathogenic avian influenza (HPAI), involving a commercial turkey flock of 34,000 birds in Sac County in the western part of the state. In other developments, an article in the journal Emerging Infectious Diseases suggests that H5N2 and other strains of the H5N8 avian influenza virus that emerged in Asia in 2014 may pose a long-term threat to poultry and wildlife in the Northern Hemisphere.  
### Response:  
The total number of reported outbreaks in Minnesota is 46 in 16 counties.  
  
### Instruction: How does FAO hope to reduce the risk of H5N1 infection in humans?  
### Input:  
26 February 2007 (CIDRAP News) - Following an inspection trip to Nigeria, the United Nations Food and Agriculture Organization (FAO) said today that the H5N1 avian influenza virus is still spreading there, and called on the country to increase surveillance measures on farms and markets. .  
Nigeria's confirmation of the first human case and death in late January shows that high-risk practices such as handling sick or dead chickens and poor hygiene during slaughter expose humans to continued risk of infection with the H5N1 virus, said Joseph Dominic, the WHO's chief veterinary officer.  
The FAO said the virus was likely to spread along Nigeria's main trade routes, putting neighboring countries at risk due to the cross-border movement of birds.  
The FAO said that discouraging home slaughter is one measure that can reduce the risk of diseases being transmitted to humans.  
The FAO said control measures in poultry markets, such as veterinary inspection and improved hygiene, could also reduce the risk of H5N1 infection in humans.  
In other news, Pakistani authorities closed a zoo in Islamabad after discovering the H5N1 avian influenza virus in peacocks and geese, AFP reported today.  
### Response:  
The FAO hopes to reduce the risk of H5N1 infection in humans by discouraging home slaughter and implementing control measures in poultry markets, such as veterinary inspections and improved hygiene.
```

Figure 29: Brief Evaluation Dataset for Inference in English

5. Inference Results

Instruction: في مصر؟ H5N1 كم عمر الفتاة المصابه بفيروس

Input:

عادت أنفلونزا الطيور إلى الظهور في بلدان، مما أدى إلى مقتل صبي يبلغ من العمر 12 عاماً في كمبوديا وإصابة فتاة صغيرة بالمرض في مصر، وفقاً لما ذكرته وكالات الأنباء - CIDRAP، 2006. وقال مايك أوليري ممثل منظمة الصحة العالمية في بيوم بيته، في تقرير لرويترز اليوم، إن الصبي الكمبودي توفي الليلة الماضية في معهد باستور في بيوم بيته، بحسب تقرير لوكاله فرانس برس اليوم H5N1، وقال مسؤول محلي كمبودي إن عينة من الصبي أثبتت أنها إيجابية. وفقاً لقصة نشرتها رويترز اليوم نقلأ عن وزير الصحة والسكان المصري حاتم الجبالي H5N1، وبالإضافة إلى ذلك، أصيبت طفلة تبلغ من العمر 16 شهراً في جنوب مصر بفيروس. وبالإضافة إلى ذلك، سيتم إعدام 250 ألف طنان آخر في قطاع غرة، في الوقت الذي يحاول فيه المسؤولون وقف انتشار الفيروس هناك. بحسب روبيترز.

Response:

في مصر عمرها 16 شهراً الفتاة المصابه بفيروس

AceGPT-7b

Response:
16

AceGPT-7b-chat

Response:
شهر 16

AceGPT-13b

Response:
بواسطة الاختبارات المعملية H5N1 تم استبعاد إصابة الأفراد السنّة بفيروس

AceGPT-13b-chat

Response:
شهرًا 16

Jais-13b

Input:
وفقاً لتقرير لوكاله فرانس برس، H5N1، مايو 2006 - أعلنت وزارة الصحة المصرية اليوم أن الاختبارات المعملية أثبتت إصابة صبي يبلغ من العمر 12 عاماً في مدينة الزقازيق بمحافظة الشرقية بفيروس 19 آخبار CIDRAP، 19 مايو 2006. وقال وزير الصحة المصري حاتم الجب

5. Inference Results

كم عدد الأطفال في سلفادور بالبرازيل الذين تم فحصهم بحثاً عن عدوى زيكا الخلقية من أبريل 2015 ويوليو 2016
Instruction:
وفي آخر تطورات شلل الأطفال العالمية، أبلغت باكستان عن حالتين جديدتين لفيروس شلل الأطفال المترافق من النوع 1، وأبلغت نلت دول أخرى مماثلة عن سبع حالات أخرى لشلل الأطفال المترافق من اللقاح في حالات بعضها، وفقاً للتحديث الأسبوعي للموسم الصادر عن المديرية العالمية لاستئصال شلل الأطفال.
Input:
وكلاهما في مقاطعة مونغالا (VDPV2) في جمهورية الكونغو الديمقراطية، تم الإبلاغ عن حالتين جديدتين لمفهوم شلل الأطفال الدائري المترافق من النوع 2 في مقاطعة زيندر تتعلق بميريس أسيب بالمثل في 9 سبتمبر 2018 وفي الوقت نفسه، أبلغت السباحة عن حالة جديدة من مفهوم VDPV2، وقد أبلغت الآن عن 23 حالة في عام 2018، وقد فوجئوا بظهور فاشدين منفصلين لفيروس.
توصلت دراسة تموير دماغ الأطفال الذين يولدون مصابين بعذري حلقيه مشتبه بها بعدوى زيكا في سلفادور، رابع أكبر مدينة في البرازيل، إلى أنه من بين 365 حالة تم التتحقق فيها بين أبريل 2015 ويوليو 2016، كان 166 (45.5٪) ينبعون من تفوهات حلقيه في الدماغ، فيما ذكر الوالدان المتوجه أهداهم في اليوم في تحدثها الأسبوعي للإنفلونزا عن زيارات طفيفة في النشاط، مع ارتفاع إحدى مناطق البلاد - تلك التي تعطن ولادات أبوها وكائنات ومسورى ونيراسكا - فوق خط الأسام، وهي أول لموسم الحذف (CDC) أبلغت المراكز الأمريكية لمكافحة الأمراض والوقاية منها (CDC)، توصيات التحديث الأسبوعي حالات الأطفال المتأثر عن مركز المطهرة على الأمراض.
تم فحص 365 طفلًا في سلفادور بالبرازيل بحثاً عن عدوى زيكا الخلقية في الفترة ما بين أبريل 2015 ويوليو 2016
Response:

AceGPT-7b

Response:
تم تأكيد 365 حالة تفشي من فاشيات فيروس زيكا المترافق من اللقاح من النوع 1 والمشتق من اللقاح 2 في سلفادور بالبرازيل خلال أبريل ويوليو 2016.

AceGPT-7b-chat

Response:
كان عدد الحالات المؤكدة لعدوى زيكا في سلفادور بالبرازيل 365 حالة بين أبريل 2015 ويوليو 2016.

AceGPT-13b

Response:
تم فحص 365 حالة من حالات التشوهات الخلقية في الدماغ في سلفادور، بالبرازيل، بين أبريل 2015 ويوليو 2016.

AceGPT-13b-chat

Response:
وهي الأولى في عام 2018 ، (c) أبلغت مقاطعة مونغالا في الكونغو الديمقراطية عن حالي إصابة جديدة بفيروس شلل الأطفال الدائري من النوع الثاني.

Jais-13b

وكلاهما في مقاطعة مونغا (VDPV2) وفي الوقت نفسه، أبلغت جمهورية الكونغو الديمقراطية عن حالتين جديدتين لفيروس شلل الأطفال المترافق من اللقاحات من النوع 2، في مقاطعة زيندر تتعلق بميريس أسيب بالمثل في 9 سبتمبر 2018 وفي الوقت نفسه، أبلغت السباحة عن حالة جديدة من فيروس، وقد أبلغت الآن عن 23 حالة في عام 2018، وقد فوجئوا بظهور فاشدين منفصلين لفيروس () في تحدثها مادر عن منظمة الصحة العالمية

5. Inference Results

Instruction: ای بلڈ؟

Response:
الصين

AceGPT-7b

Response:
صين

AceGPT-7b-chat

Response:

AceGPT-13b

Response:

AceGPT-13b-chat

Response:

Jais-13b

وفي اواخر الالسنواه الماضى، اعلنت وزارة الصحة الصينيه أنها اكدت إصابه رجل يبلغ من العمر 38 عاماً بفيروس H5N1، وهو فيروس مميت، وأبلغ ممثلو الصحة العالمية اليوم عن حالاتين اخرتين من فيروس H5N1.

حسماً دكتور روينر زيمبر، مدير H5N1، وفي بيانه، أبلغ مسؤولو الصحة عن حالاتين اخرتين من فيروس H5N1، وفي تطورات اخر، أبلغ المسؤولون عن ثالث حالات اخر في فنيان، وفي احدث الحالات الثلاث التي تم الإبلاغ عنها هذا الالسنواه.

5. Inference Results

Instruction: What is the preliminary examination result of an Indonesian university student?

Input:

March 27, 2007 (CIDRAP News) - Egypt today reported two more human cases of H5N1 avian influenza virus, bringing the total cases in the country to 29, while Indonesia reported three suspected cases, two of which were fatal, according to the news services.

Egyptian officials said that the two patients were a 6-year-old girl from the southern Qena governorate and a 5-year-old boy from Minya in central Egypt, according to a Reuters report today.

In Indonesia, preliminary tests indicated infection with the H5N1 virus in a 22-year-old female university student who died on March 24, and a teenage boy who died on March 25, according to an Agence France-Presse report today, citing a spokeswoman for the government's Avian Influenza Information Office. center.

Response:

Initial tests indicated that the Indonesian university student was infected with the H5N1 virus.

AceGPT-7b

Response:

The 22-year-old Indonesian university student was likely to have the H5N1 virus.

AceGPT-7b-chat

Response:

Preliminary tests indicated infection with the H5N1 virus in a 22-year-old university student in Indonesia who died on March

AceGPT-13b

Response:

The preliminary examination result of an Indonesian university student is infection with the H5N1 virus.

AceGPT-13b-chat

Response:

Preliminary tests indicate that the 22-year-old university student is infected with the H5N1 virus.

Jais-13b

Tests also indicated infection with H5N1 in a 38-year-old man who was hospitalized in Jakarta, according to the Associated Press report today. The Jakarta man was receiving intensive treatment, but had shown no signs of improvement, according to Agence France-Presse.

5. Inference Results

```
### Instruction: What is the total number of reported outbreaks in Minnesota?  
### Input:  
Minnesota Governor Mark Dayton today declared a state of emergency due to a widespread invasion of poultry farms by H5N2 avian influenza, as the state's first outbreaks were reported among chickens and backyard poultry, and both Wisconsin and Iowa announced new outbreaks in turkeys. .  
The state has recorded 46 outbreaks in 16 counties, with more than 2.63 million birds killed by the virus or destroyed to stop its spread, according to the Minnesota Board of Animal Health (MBAH). Additionally, MBAH reported an outbreak in a mixed backyard flock of 150 birds in Pipestone County, near the southwest corner of the state, the first outbreak in the county. Chickens are thought to be less susceptible to H5N2 infection than turkeys, but a few outbreaks have been reported on chicken farms, including at least one each in Iowa and Wisconsin. Additionally, MBAH spokeswoman Bethany Hahn said today that one of the 15 outbreaks reported in Kandiyohi County has not yet been included in the total, because the number of infected birds has not yet been confirmed. Meanwhile, officials in Wisconsin today reported the state's sixth H5N2 outbreak, at a farm with 90,000 turkeys. Also, early this evening, the Iowa Department of Agriculture reported the state's third outbreak of highly pathogenic avian influenza (HPAI), involving a commercial turkey flock of 34,000 birds in Sac County in the western part of the state. In other developments, an article in the journal Emerging Infectious Diseases suggests that H5N2 and other strains of the H5N8 avian influenza virus that emerged in Asia in 2014 may pose a long-term threat to poultry and wildlife in the Northern Hemisphere.  
### Response:  
The total number of reported outbreaks in Minnesota is 46 in 16 counties.
```

AceGPT-7b

```
### Response:  
Minnesota has reported 46 outbreaks in 16 counties, with more than 2.63 million birds lost or destroyed to stop the virus.
```

AceGPT-7b-chat

```
### Response:  
The total number of reported outbreaks in Minnesota is 46, with more than 2.63 million birds killed by the virus or destroyed to stop its spread.
```

AceGPT-13b

```
### Response:  
46
```

AceGPT-13b-chat

```
### Response:  
46 outbreaks
```

Jais-13b

```
### Response:  
46 outbreaks
```

5. Inference Results

```
### Instruction: How does FAO hope to reduce the risk of H5N1 infection in humans?  
### Input:  
20 February 2007 (CIDRAP News) - Following an inspection trip to Nigeria, the United Nations Food and Agriculture Organization (FAO) said today that the H5N1 avian influenza virus is still spreading there, and called on the country to increase surveillance measures on farms and markets. . Nigeria's confirmation of the first human case and death in late January shows that high-risk practices such as handling sick or dead chickens and poor hygiene during slaughter expose humans to continued risk of infection with the H5N1 virus, said Joseph Dominic, the WHO's chief veterinary officer. The FAO said the virus was likely to spread along Nigeria's main trade routes, putting neighboring countries at risk due to the cross-border movement of birds. The FAO said discouraging home slaughter is one measure that can reduce the risk of diseases being transmitted to humans. The FAO said control measures in poultry markets, such as veterinary inspection and improved hygiene, could also reduce the risk of H5N1 infection in humans. In other news, Pakistani authorities closed a zoo in Islamabad after discovering the H5N1 avian influenza virus in peacocks and geese, AFP reported today.  
### Response:  
The FAO hopes to reduce the risk of H5N1 infection in humans by discouraging home slaughter and implementing control measures in poultry markets, such as veterinary inspections and improved hygiene.
```

AceGPT-7b

Response:
The FAO believes that discouraging home slaughter, veterinary inspection and better hygiene in markets can reduce the risk of H5N1 transmission to humans.

AceGPT-7b-chat

Response:
The FAO believes that discouraging home slaughter, veterinary inspection and better hygiene in markets can reduce the risk of H5N1 transmission to humans.

AceGPT-13b

Response:
The Food and Agriculture Organization hopes that by discouraging home slaughter and implementing control measures in poultry markets, such as veterinary inspection and improved hygiene, the risk of H5N1 infection in humans can be reduced.

AceGPT-13b-chat

Response:
FAO said controlling measures in poultry markets, such as veterinary inspection and improved hygiene, could help reduce the risk of H5N1 infection in humans.

Jais-13b

term threat to poultry and humans in the U.S.
The article reports that researchers found the H5N8 virus in the United States, in an Arkansas broiler farm.
Furthermore

6. Future Work

- Further training for instruction tuning with more epochs.
- Evaluation of Q&A Task with Some Metrics.
 - F1 score
 - Exact-match score
 - BLEU score
 - ROUGE-N score
- Implementation of RAG.

6. Future Work

Evaluation Metrics

1. F-1 Score: F1 score is a metric that combines precision and recall, providing a balanced measure of a model's accuracy.

$$F_1 = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Figure 24: F1 Score Formula

https://qa.fastforwardlabs.com/no%20answer/null%20threshold/bert/distilbert/exact%20match/f1/robust%20predictions/2020/06/09/Evaluating_BERT_on_SQuAD.html

2. Exact-match Score: Exact match is a binary metric used in question answering tasks, indicating whether the model's output exactly matches the reference answer.

Exact Match is a strict evaluation metric that only gives two scores (0 or 1).
EM score is 1 if the answer provided by the annotator is precisely the same
as the predicted answer; else, it gives 0 (EM score for “the US” and “the
United States” is 0). EM might be a good metric for very short or concise
factual answers, but it isn't suitable for long answers.

Figure 25: Exact-match Definition

<https://kaushikshakkari.medium.com/open-domain-question-answering-series-part-4-answer-quality-metrics-evaluating-the-reader-f17fa20736bf#:~:text=Exact%20Match%20is%20a%20strict,United%20States%E2%80%9D%20is%200>

6. Future Work

Evaluation Metrics

3. BLEU Score: BLEU (Bilingual Evaluation Understudy) is a metric commonly used for machine translation evaluation, measuring the precision of n-grams overlap between the generated and reference text.

$$\text{BLEU} = \text{BP} * \exp(\sum p_n)$$

Figure 26: BLEU Formula

<https://medium.com/@sthanikamsanthosh1994/understanding-bleu-and-rouge-score-for-nlp-evaluation-1ab334ecadcb>

4. ROUGE-N Score: ROUGE-N (N-gram Overlap) is a metric assessing the overlap of n-grams between the generated and reference text.

$$\begin{aligned} & \text{ROUGE-N} \\ &= \frac{\sum_{S \in \{\text{ReferenceSummaries}\}} \sum_{\text{gram}_n \in S} \text{Count}_{\text{match}}(\text{gram}_n)}{\sum_{S \in \{\text{ReferenceSummaries}\}} \sum_{\text{gram}_n \in S} \text{Count}(\text{gram}_n)} \quad (1) \end{aligned}$$

Figure 27: Rouge-N Formula

<https://stackoverflow.com/questions/72758470/rouge-scores-are-different-when-using-package-datasets-and-rouge-score>

Q&A

감사합니다!