
Cross-lingual Question Answering in the Age of LLMs

Applications and Challenges

Casimiro Pio Carrino,
NLP Research Scientist at Nuclia, PhD candidate at UPC

Cross-lingual Question Answering: Datasets, Evaluation and Methods

Challenges



Motivation: Information
extraction from
multilingual documents



Tackling
cross-lingual QA:
from BERT to LLMs



CHALLENGES



Use case:
QA over
documents



Motivation:
Information extraction from multilingual
documents

On average, how many participants attend La Tomatina?



ラ・トマティバ・フェスティバルには平均して何人の参加者が参加しますか？

ラ・トマティバ・フェスティバルは、スペインのトマティーナとしても知られるトマト祭りです。この祭りは毎年8月の最終水曜日にスペインのブニョールで開催されます。参加者はトマトを使って大規模な食物戦を繰り広げます。

ラ・トマティバ・フェスティバルの参加者数は年々変動しますが、平均的には数万人から数十万人の間で推移しています。具体的な参加者数は公式には発表されていないため、正確な数字を提供することはできませんが、この祭りは非常に人気があり、世界中から多くの人々が訪れることが知られています。

この情報は、一般的な参加者数に関する情報を提供するために検索結果スニペットを使用しています。

translate the answer to english

The La Tomatina Festival, also known as the Tomato Festival, is an annual event held in Buñol, Spain, on the last Wednesday of August. Participants engage in a large-scale food fight using tomatoes.

The number of participants in the La Tomatina Festival varies from year to year, but on average, it ranges from tens of thousands to hundreds of thousands of people. Since the exact number of participants is not officially announced, it is not possible to provide an accurate figure. However, this festival is highly popular and attracts many people from around the world.

This information is based on search result snippets to provide an estimate of the average number of participants.

On average, how many participants attend La Tomatina?



ラ・トマティバ・フェスティバルには平均して何人の参加者が参加しますか？

ラ・トマティバ・フェスティバルは、スペインのトマティーナとしても知られるトマト祭りです。この祭りは毎年8月の最終水曜日にスペインのブニョールで開催されます。参加者はトマトを使って大規模な食物戦を繰り広げます。

ラ・トマティバ・フェスティバルの参加者数は年々変動しますが、平均的には数万人から数十万人の間で推移しています。具体的な参加者数は公式には発表されていないため、正確な数字を提供することはできませんが、この祭りは非常に人気があり、世界中から多くの人々が訪れることが知られています。

この情報は、一般的な参加者数に関する情報を提供するために検索結果スニペットを使用しています。

translate the answer to english


The La Tomatina Festival, also known as the Tomato Festival, is an annual event held in Buñol, Spain, on the last Wednesday of August. Participants engage in a large-scale food fight using tomatoes.

The number of participants in the La Tomatina Festival varies from year to year, but on average, it ranges from tens of thousands to hundreds of thousands of people. Since the exact number of participants is not officially announced, it is not possible to provide an accurate figure. However, this festival is highly popular and attracts many people from around the world.

This information is based on search result snippets to provide an estimate of the average number of participants.




On average, how many participants attend La Tomatina?




En promedio, ¿cuántos participantes asisten al Festival La Tomatiba?


La Tomatina 2022: Origen, precio y...

 www.antena3.com


•▶ Todo lo que debes saber sobre l...

 motordeviajes.com


Festival de la Cultura Wayúu -...

 es.wikipedia.org


FESTIVALES DE MÚSICA POPULAR...

 repositorio.uniandes.edu.co

Facultad de Comunicación y...

 repository.javeriana.edu.co

Untitled

 primolatvija.hosted.exlibrisgroup.com

En promedio, al Festival La Tomatina **asisten alrededor de 22,000 personas** ¹. Es considerada como la pelea de comida más grande del mundo y se lleva a cabo en Buñol, Comunidad Valenciana, España. Los participantes arrojan tomates unos a otros en esta festividad tradicional ¹

 antena3 ¹

Related web results

 [La Tomatina 2022: Origen, precio y cuántos tomates se utilizan](#)

[antena3.com](#)

Aug 31, 2022 ... El festival de peleas de comida más grande del mundo, retomará su actividad tras el parón que tuvo en el año 2020 por la pandemia del Covid-19.




•▶ [Todo lo que debes saber sobre la fiesta de la Tomatina 2023](#)

[motordeviajes.com](#)

La Tomatina es un festival de lanzamiento de tomates en el que a los participantes no solo se les permite, sino que se les anima a arrojar tomates con dos ...




 [Festival de la Cultura Wayúu - Wikipedia, la enciclopedia libre](#)

[es.wikipedia.org](#)


El festival de la Cultura Wayúu celebrado en Uribia-La Guajira, es el evento más significativo para los indígenas wayúu en La Guajira Colombiana e incluso ...

On average, how many participants attend La Tomatina?




En promedio, ¿cuántos participantes asisten al Festival La Tomatiba?


La Tomatina 2022: Origen, precio y...

 www.antena3.com


•▶ Todo lo que debes saber sobre l...

 motordeviajes.com


Festival de la Cultura Wayúu -...

 es.wikipedia.org


FESTIVALES DE MÚSICA POPULAR...

 repositorio.uniandes.edu.co


Facultad de Comunicación y...

 repository.javeriana.edu.co



Untitled



 primolativija.hosted.exlibrisgroup.com


En promedio, al Festival La Tomatina **asisten alrededor de 22,000 personas** ¹. Es considerada como la pelea de comida más grande del mundo y se lleva a cabo en Buñol, Comunidad Valenciana, España. Los participantes arrojan tomates unos a otros en esta festividad tradicional ¹

 [antena3](#) ¹

Related web results

 [La Tomatina 2022: Origen, precio y cuántos tomates se utilizan](#)
[antena3.com](#)
Aug 31, 2022 ... El festival de peleas de comida más grande del mundo, retomará su actividad tras el parón que tuvo en el año 2020 por la pandemia del Covid-19.


 [▶ Todo lo que debes saber sobre la fiesta de la Tomatina 2023](#)
[motordeviajes.com](#)
La Tomatina es un festival de lanzamiento de tomates en el que a los participantes no solo se les permite, sino que se les anima a arrojar tomates con dos ...


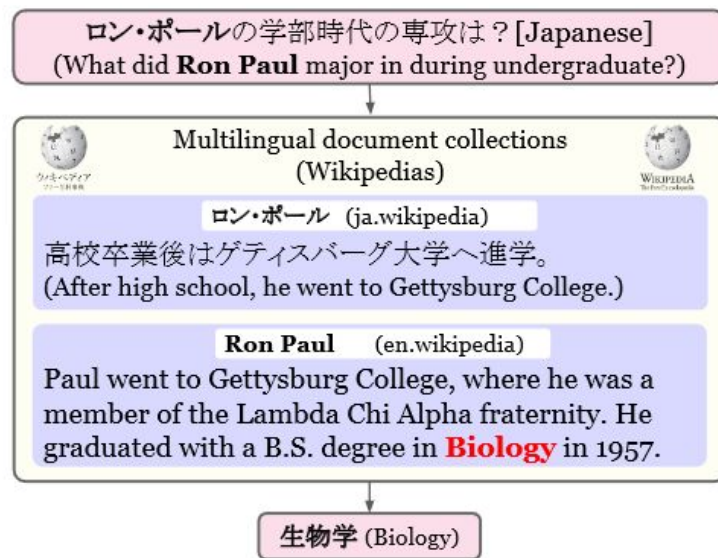
 [Festival de la Cultura Wayúu - Wikipedia, la enciclopedia libre](#)
[es.wikipedia.org](#)
El festival de la Cultura Wayúu celebrado en Uribia-La Guajira, es el evento más significativo para los indígenas wayúu en La Guajira Colombiana e incluso ...



Information asymmetry and scarcity

Information asymmetry: where questions reference concepts from other cultures

Information scarcity: where languages have few reference articles



Semantic alignment across languages

Current systems suffer from same-language bias, a tendency toward aligning semantically irrelevant examples sharing the same language rather than semantically relevant examples with different languages

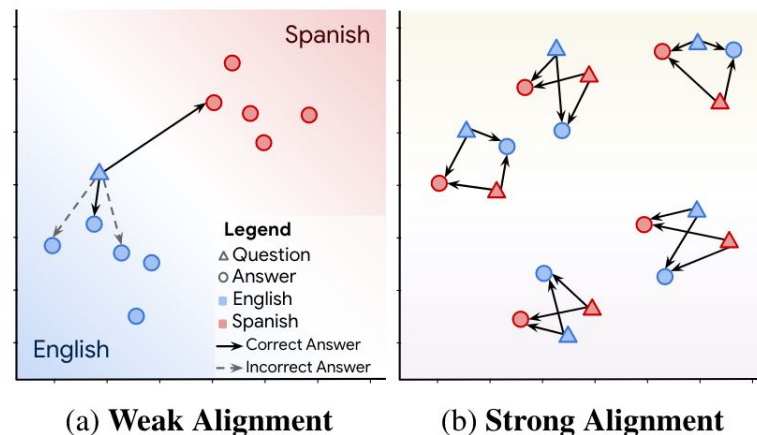


Figure 1: A weakly aligned multilingual embedding space enables zero-shot transfer between languages, but incorrect answers in the same language are preferred over correct answers in a different language. A strongly aligned embedding space “factors out” language, so the most semantically relevant pairs are always the closest, regardless of language.

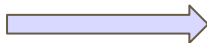
Multilingual information extraction applications

Q (ja): ラ・トマティバ・フェスティバルには平均して何人の参加者が参加しますか？

search



retrieval



generator



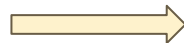
A (ja): ラトマティーナ フェスティバルには平均して約22,000 人が参加します。

generator



A (ja): ラトマティーナ フェスティバルには平均して約 22,000 人が参加します。

reader



A (ja): 22,000人

Definitions of cross-lingual QA systems

Consider a set of languages $\{L_1, \dots, L_r\}$, given a question q in language L_i , let M be the set of relevant candidate $\{C\}$ sentence answers for Q in any language. A cross-lingual system uses the top- k ranked answers in M — regardless of language — to **generate/extract an answer A in L_i**

Retrieval
+
Extractive/
Abstractive

Consider a set of languages $\{L_1, \dots, L_r\}$, given a **question Q in language L_i** , a cross-lingual system **generates an answer A in L_i**

Abstractive

Consider a set of languages $\{L_1, \dots, L_r\}$, given a **question Q and a context C in language L_i** , a cross-lingual system **extracts an answer A in L_i**

Extractive



Cross-lingual Question Answering: Datasets, Evaluation and Methods

Evaluation Benchmarks

Dataset	Task Type	Languages	Total Examples
XQuAD	Extractive	11	13K
XQA	Extractive	9	90K
MLQA	Extractive	7	46K
XQuAD-R	Retrieval+Extractive	11	13K
XOR-TYDI QA	Retrieval+Extractive	11	40K
MKQA	Retrieval+Extractive	23	260K
TYDI QA	Retrieval+Extractive	11	204K
GenTyDiQA	Abstractive	5	3K

En	During what time period did the Angles migrate to Great Britain?
The name "England" is derived from the Old English name Engaland [...] The Angles were one of the Germanic tribes that settled in Great Britain during the Early Middle Ages . [...] The Welsh name for the English language is "Saesneg"	

De	Während welcher Zeitperiode migrierten die Angeln nach Großbritannien?
Der Name England leitet sich vom altenglischen Wort Engaland [...] Die Angeln waren ein germanischer Stamm, der das Land im Frühmittelalter besiedelte. [...] ein Verweis auf die weißen Klippen von Dover.	

Ar	في أي حقبة زمنية هاجر الأنجل إلى بريطانيا العظمى؟
والتي تعني "أرض الأنجل". والأنجل كانت واحدة، Engaland، يشتق اسم "إنجلترا" من الكلمة الإنجليزية القديمة من القبائل الجرمانية التي استقرت في إنجلترا خلال العصور الوسطى . [...] وقد سماها العرب قديماً الإنكتار	

Vi	Trong khoảng thời gian nào người Angles di cư đến Anh?
Tên gọi của Anh trong tiếng Việt bắt nguồn từ tiếng Trung. [...] Người Angle là một trong những bộ tộc German định cư tại Anh trong Thời đầu Trung Cổ . [...] dường như nó liên quan tới phong tục gọi người German tại Anh là Angli Saxones hay Anh - Sachsen.	

(a)

En	What are the names given to the campuses on the east side of the land the university sits on?
The campus is in the residential area of Westwood [...] The campus is informally divided into North Campus and South Campus , which are both on the eastern half of the university's land. [...] The campus includes [...] a mix of architectural styles.	

Es	¿Cuáles son los nombres dados a los campus ubicados en el lado este del recinto donde se encuentra la universidad?
El campus incluye [...] una mezcla de estilos arquitectónicos. Informalmente está dividido en Campus Norte y Campus Sur , ambos localizados en la parte este del terreno que posee la universidad. [...] El Campus Sur está enfocado en la ciencias físicas [...] y el Centro Médico Ronald Reagan de UCLA.	

Zh	位于大学占地东半部的校园名称是什么？
整个校园被不正式地分为 南北两个校园 ，这两个校园都位于大学占地的东半部。北校园是原校园的中心，建筑以意大利文艺复兴时代建筑闻名，其中的包威尔图书馆（Powell Library）成为好莱坞电影的最佳拍摄场景。[...] 这个广场曾在许多电影中出现。	

Hi	विश्वविद्यालय जहाँ स्थित है, उसके पूर्वी दिशा में बने परिसरों को क्या नाम दिया गया है?
जब 1919 में यूसीएलए ने अपना नया परिसर खोला, तब इसमें चार इमारतें थीं। [...] परिसर अनौपचारिक रूप से उत्तरी परिसर और दक्षिणी परिसर में विभाजित है, जो दोनों विश्वविद्यालय की जमीन के पूर्वी हिस्से में स्थित हैं। [...] दक्षिणी परिसर में भौतिक विज्ञान, जीव विज्ञान, इंजीनियरिंग, मनोविज्ञान, गणितीय विज्ञान, सभी स्वास्थ्य से संबंधित क्षेत्र और यूएलसीए मेडिकल सेंटर स्थित है।	

(b)

Lang	Context paragraph w/ answer spans	Questions
en	<p>The heat required for boiling the water and supplying the steam can be derived from various sources, most commonly from [burning combustible materials]₁ with an appropriate supply of air in a closed space (called variously [combustion chamber]₂, firebox). In some cases the heat source is a nuclear reactor, geothermal energy, [solar]₃ energy or waste heat from an internal combustion engine or industrial process. In the case of model or toy steam engines, the heat source can be an [electric]₄ heating element.</p>	<ol style="list-style-type: none"> 1. What is the usual source of heat for boiling water in the steam engine? 2. Aside from firebox, what is another name for the space in which combustible material is burned in the engine? 3. Along with nuclear, geothermal and internal combustion engine waste heat, what sort of energy might supply the heat for a steam engine? 4. What type of heating element is often used in toy steam engines?
es	<p>El calor necesario para hervir el agua y suministrar el vapor puede derivarse de varias fuentes, generalmente de [la quema de materiales combustibles]₁ con un suministro adecuado de aire en un espacio cerrado (llamado de varias maneras: [cámara de combustión]₂, chimenea...). En algunos casos la fuente de calor es un reactor nuclear, energía geotérmica, [energía solar]₃ o calor residual de un motor de combustión interna o proceso industrial. En el caso de modelos o motores de vapor de juguete, la fuente de calor puede ser un calentador [eléctrico]₄.</p>	<ol style="list-style-type: none"> 1. ¿Cuál es la fuente de calor habitual para hacer hervir el agua en la máquina de vapor? 2. Aparte de cámara de combustión, ¿qué otro nombre que se le da al espacio en el que se quema el material combustible en el motor? 3. Junto con el calor residual de la energía nuclear, geotérmica y de los motores de combustión interna, ¿qué tipo de energía podría suministrar el calor para una máquina de vapor? 4. ¿Qué tipo de elemento calefactor se utiliza a menudo en las máquinas de vapor de juguete?
zh	<p>让水沸腾以提供蒸汽所需热量有多种来源，最常见的是在封闭空间（别称有 [燃烧室]₂、火箱）中供应适量空气来 [燃烧可燃材料]₁。在某些情况下，热源是核反应堆、地热能、[太阳能]₃ 或来自内燃机或工业过程的废气。如果是模型或玩具蒸汽发动机，还可以将 [电]₄ 加热元件作为热源。</p>	<ol style="list-style-type: none"> 1. 蒸汽机中让水沸腾的常用热源是什么？ 2. 除了火箱之外，发动机内燃烧可燃材料的空间的别名是什么？ 3. 除了核能、地热能和内燃机废气以外，还有什么热源可以为蒸汽机供能？ 4. 玩具蒸汽机通常使用什么类型的加热元件？

Spelling variation in Arabic transliteration

Q: من هو موزارت ؟
? mwzArt hw mn
Who is Mozart ?

A: فولفغانغ أماديوس موتسارت (27 يناير 1756 - 5 ديسمبر 1791) ولد في 27 يناير 1756 في سالزبورغ بالنمسا
bAlnmsA sAlzbwrg fy 1756 ynAyr 27 fy wld (1791 dysmbr 5 - 1756 ynAyr 27) mwtsArt A#mAdyws fwlfgAng
Wolfgang Amadeus Mozart (January 27, 1756 - December 5, 1791) was born on January 27, 1756 in Salzburg, Austria

This Arabic example demonstrates variation in the spelling of non-native names. Both spellings of Mozart are correct and refer to the same entity across the QA pair.

Script switching in Russian

Q: Кто изобрел телефон ?
Kto izobrel telefon ?
who invented telephone ?
*Who invented the
telephone ?*

A: Сам Рейс назвал сконструированное им устройство Telephone .
Sam Reis nazyval skonstruirovannoe im ustroistvo Telephone .
self Reis called constructed him device Telephone .
Reis himself called the device he created the Telephone .

This Russian example demonstrates how some entities of non-native origin may maintain the original Latin-script spelling, especially when the term has been directly borrowed into Russian and is phonologically similar to the original. Here the question about the inventor of the telephone contains the more common Cyrillic rendition of the term, 'телефон'. However, the answer passage has it in the original English spelling as 'Telephone'.

Evaluation Metrics

Retrieval + Extractive

Sentence/documents level

- Recall (@k)
- Normalized Discounted Cumulative Gain (NDCG)

Abstractive

Word n-grams level

- BLEU
- ROUGE-L

Extractive

Token level

- Token-level F1

F1 score: precision is the ratio of the number of shared words to the total number of words in the prediction, and recall is the ratio of the number of shared words to the total number of words in the ground truth

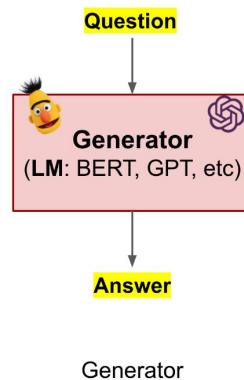
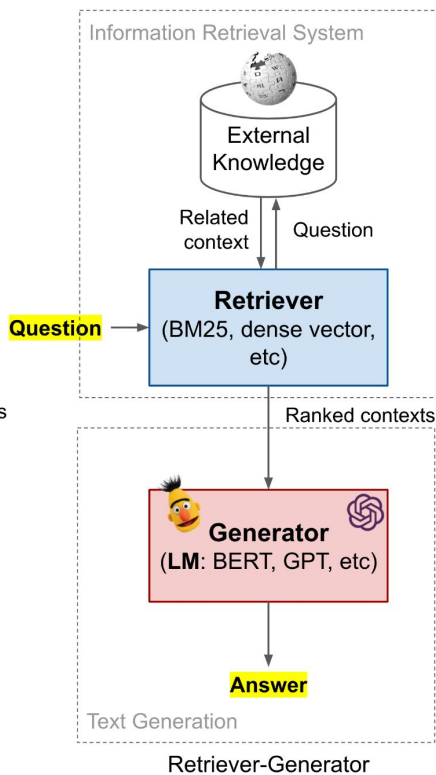
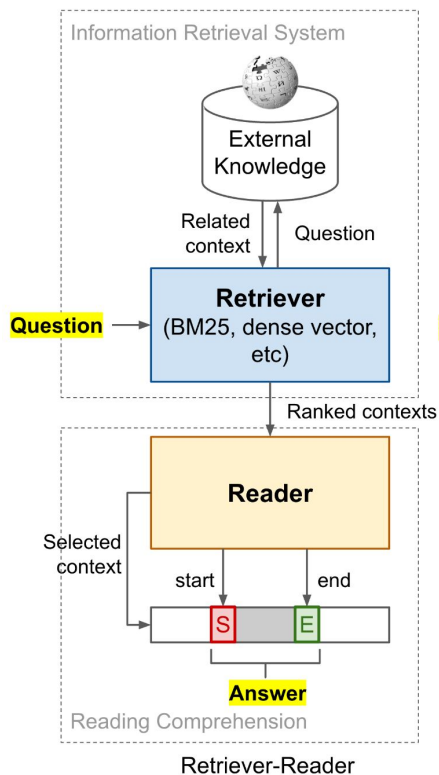
- Exact Match

EM score: string match between the prediction and the ground truth



Tackling cross-lingual QA: from BERT to LLMs

Modeling the task: fine-tuning



2018

2021



LLMs approaches: in-context learning into play

Instruction-tuned + alignment with human preferences → **In-context learning!**

Task Instruction I:

You are an NLP assistant whose purpose is to solve reading comprehension problems. You will be provided questions on a set of passages and you will need to provide the answer as it appears in the passage.

Input:

{context}

Q: {question}

A: {answer}

...

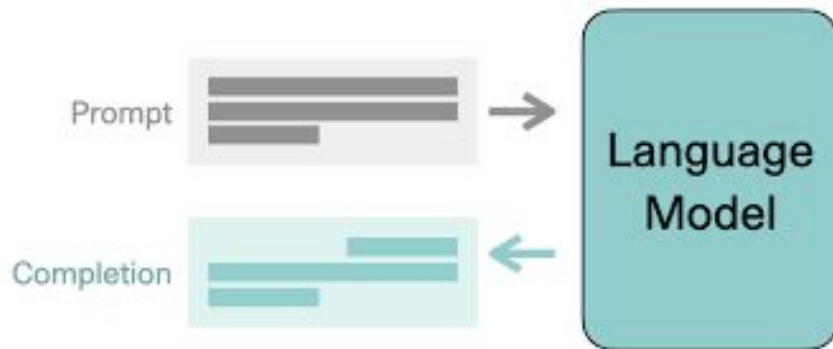
...

{context}

Q: {question}

Few-shot
examples $\times N$

Test
example



Fine-tuning vs prompting (in-context learning)

Translate-train

Translate data from English to the target language and train on the resulting synthetic data

Translate-test

Translate the data from target language to English during inference, retrieve the English answer and translate back to the target language

Zero-shot transfer

Evaluate a pre-trained QA model on an unseen language

Monolingual Prompting

Few-shot examples in the same language as test examples

Zero-Shot Cross-Lingual Prompting

Use few-shot examples from a pivot language (e.g., English)

Translate-Test Prompting

Use few-shot examples from a pivot language (e.g., English) and translate the test examples to English.

Assessing Cross-lingual QA performances

Model	Classification				Question Answering			Sequence Labelling		Summarization
	XNLI	PAWS-X	XCOPA	XStoryCloze	XQuAD	TyDiQA-GoldP	MLQA	UDPOS	PAN-X	XLSum
Metrics	Acc.	Acc.	Acc.	Acc.	F1 / EM	F1 / EM	F1 / EM	F1	F1	ROUGE-L
<i>Fine-tuned Baselines</i>										
mBERT	65.4	81.9	56.1	-	64.5 / 49.4	59.7 / 43.9	61.4 / 44.2	71.9	62.2	-
mT5-Base	75.4	86.4	-	-	67.0 / 49.0	57.2 / 41.2	64.6 / 45.0	-	55.7	28.1 [†]
XLNet-Large	79.2	86.4	69.2	-	76.6 / 60.8	65.1 / 45.0	71.6 / 53.2	76.2	65.2	-
TuLRv6 - XXL	88.8 [†]	93.2 [†]	82.2 [†]	-	86 / 72.9 [†]	84.6 / 73.8 [†]	81 / 63.9 [†]	83.0 [†]	84.7 [†]	-
<i>Prompt-Based Baselines</i>										
BLOOMZ	54.2	(82.2) [‡]	60.4	76.2	(70.7 / 58.8) [‡]	(75.2 / 63.2) [‡]	-	-	-	-
XGLM	47.3	54.5	62.0	66.5	-	-	-	-	-	-
<i>Open AI Models</i>										
gpt-3.5-turbo	62.1	70.0	81.0	87.7	60.4 / 38.2	60.1 / 38.4	56.1 / 32.8 [‡]	60.2 [‡]	40.3 [‡]	17.9
gpt-3.5-turbo (TT)	64.3	67.2	83.5	93.8	-	-	46.3 / 27.0	-	-	-
text-davinci-003	59.27	67.08	82.5	74.7	40.5 / 28.0	49.7 / 38.3	44.0 / 28.8	-	-	-
text-davinci-003 (TT)	67.0	68.5	83.8 [‡]	94.8 [‡]	-	-	54.9 / 34.6	-	-	-
gpt-4-32k	75.4 [‡]	-	-	-	68.3 / 46.6	71.5 / 50.9	-	-	-	19.3 [‡]

Table 3: Average performance across languages in each of the different tasks that we study for our benchmarking exercise. TT suffix refers to the translate-test prompting strategy discussed in Section 2.2, without any suffix we refer to the monolingual strategy by default (except for XQuAD and IndicQA where it refers to cross-lingual setup). Numbers in **bold** with [†] symbol indicate best performing Fine-tuned model and the ones with [‡] refer to the best prompt-based generative model. The best overall numbers are underlined. For BLOOMZ the values in parenthesis indicate that the model was fine-tuned on the task during multi-task training (Muennighoff et al., 2022)

Conclusions

- Specialized models performs much better than LLMs prompting!
- Cross-lingual QA Evaluation is underexplored
- Evaluation addresses only on closed API and open-source LLMs



Challenges

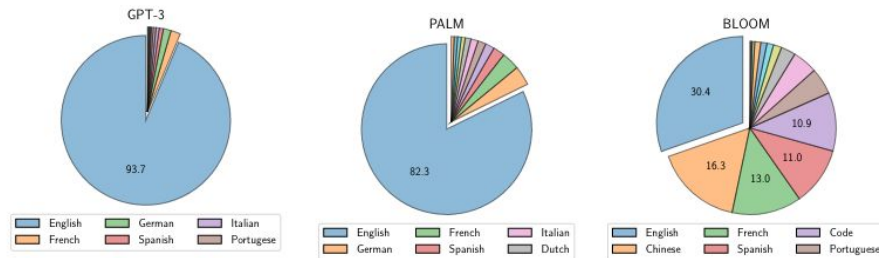
Achieving Multilingualism

LLMs training data are predominantly in English

- Multilingual task-specific fine-tuning
- Pre-training from scratch or continuous pretraining with multilingual vocabulary



Bias towards English-based prompts



I want you to act as a question answering expert for Japanese .
Question: フルハウスでスティーブが出てきたのはどのシーズンですか？
You should retell the question in English.
You should answer the question in English in one or a few words.
You should step-by-step answer the request.
You should tell me the answer in one or a few words in Japanese in this format 'Answer:'.

Question: what season does steve come in full house?
English Answer: 5.0 seasons.
Japanese Answer: 5.0 時節

Evaluation

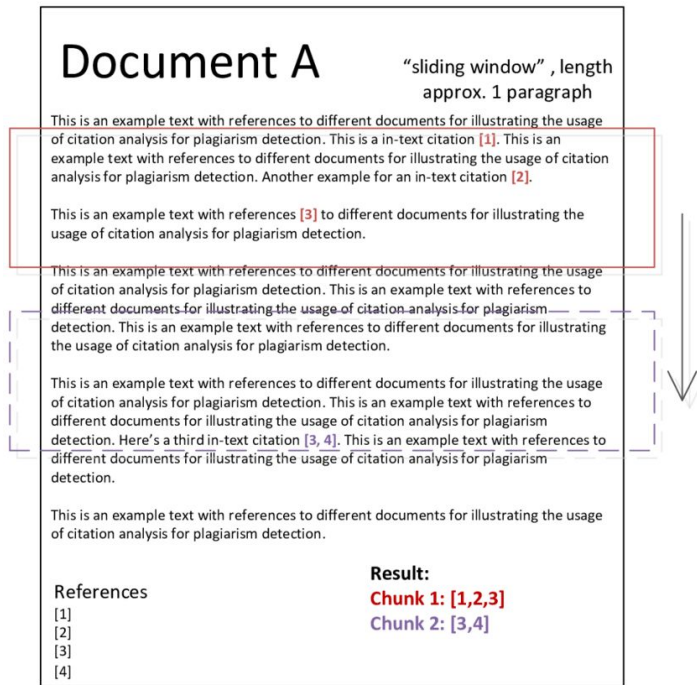
Need of Extensive evaluation

- Lack of studies to evaluate abstractive QA
- Design better metrics for automatic evaluation
- Measure factuality and reduce hallucinations

Retrieval

Improving retrieval

- Achieving strong alignment with cross-lingual embeddings
- Design methods to better chunk the information in informative contexts

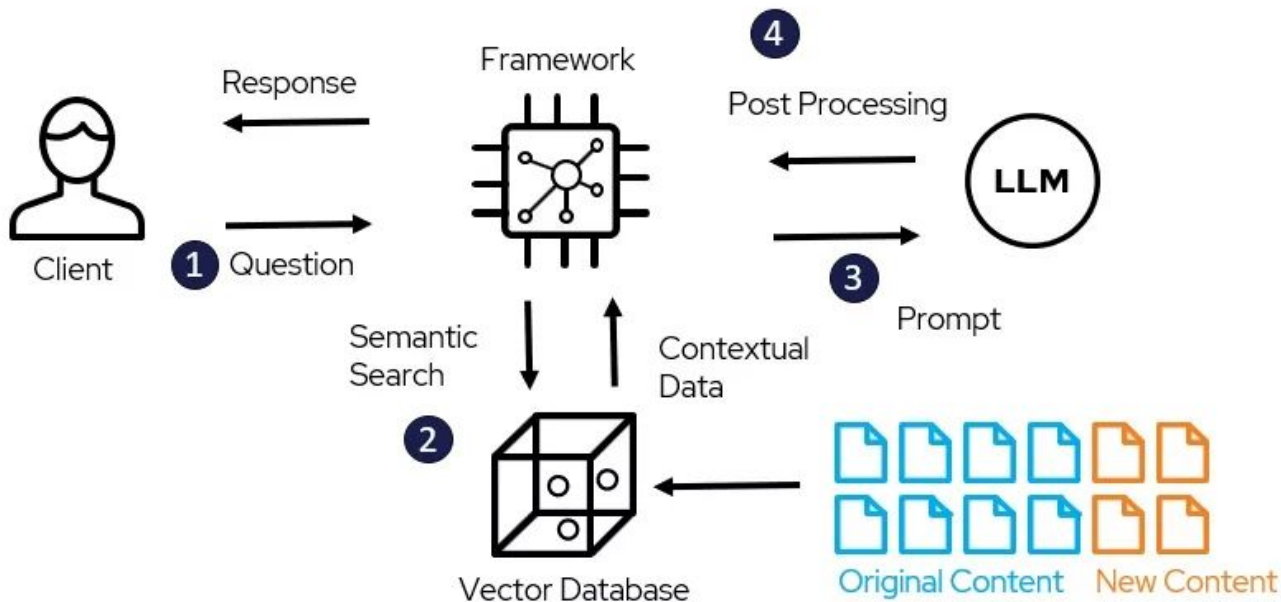




Applications: Cross-lingual QA with Retrieval Augmented Generation

RAG Pipeline

RAG Architecture Model



References (some)

1. Roy, Uma, et al. "LAREQA: Language-Agnostic Answer Retrieval from a Multilingual Pool." Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, 2020, pp. 5919–30. ACLWeb, <https://doi.org/10.18653/v1/2020.emnlp-main.477>.
2. Lewis, Patrick, et al. "MLQA: Evaluating Cross-Lingual Extractive Question Answering." Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2020, pp. 7315–30. ACLWeb, <https://doi.org/10.18653/v1/2020.acl-main.653>.
3. Asai, Akari, et al. "XOR QA: Cross-Lingual Open-Retrieval Question Answering." Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, 2021, pp. 547–64. ACLWeb, <https://doi.org/10.18653/v1/2021.naacl-main.46>.
4. Clark, Jonathan H., et al. "TyDi QA: A Benchmark for Information-Seeking Question Answering in Typologically Diverse Languages." Transactions of the Association for Computational Linguistics, vol. 8, July 2020, pp. 454–70. Silverchair, https://doi.org/10.1162/tacl_a_00317.
5. Longpre, Shayne, et al. "MKQA: A Linguistically Diverse Benchmark for Multilingual Open Domain Question Answering." Transactions of the Association for Computational Linguistics, vol. 9, 2021, pp. 1389–406. ACLWeb, https://doi.org/10.1162/tacl_a_00433.
6. Muller, Benjamin, et al. "Cross-Lingual Open-Domain Question Answering with Answer Sentence Generation." Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), edited by Yulan He et al., Association for Computational Linguistics, 2022, pp. 337–53. ACLWeb, <https://aclanthology.org/2022.aacl-main.27>.
7. Huang, Haoyang, et al. Not All Languages Are Created Equal in LLMs: Improving Multilingual Capability by Cross-Lingual-Thought Prompting. arXiv:2305.07004, arXiv, 22 Oct. 2023. arXiv.org, <https://doi.org/10.48550/arXiv.2305.07004>.
8. Ahuja, Kabir, et al. MEGA: Multilingual Evaluation of Generative AI. arXiv:2303.12528, arXiv, 19 May 2023. arXiv.org, <https://doi.org/10.48550/arXiv.2303.12528>.
9. Artetxe, Mikel, et al. "On the Cross-Lingual Transferability of Monolingual Representations." Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2020, pp. 4623–37. ACLWeb, <https://doi.org/10.18653/v1/2020.acl-main.421>.
10. Tanwar, Eshaan, et al. "Multilingual LLMs Are Better Cross-Lingual In-Context Learners with Alignment." Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), edited by Anna Rogers et al., Association for Computational Linguistics, 2023, pp. 6292–307. ACLWeb, <https://doi.org/10.18653/v1/2023.acl-long.346>.
11. Guo, Zishan, et al. Evaluating Large Language Models: A Comprehensive Survey. arXiv:2310.19736, arXiv, 31 Oct. 2023. arXiv.org, <https://doi.org/10.48550/arXiv.2310.19736>.



Practical Use-case

QA over documents



+



+



Hugging Face