

PolyGlot Gemini lens for PDFs



Raghavan Muthuregunathan

<https://www.linkedin.com/in/raghavanmit/>

Loom [link](#)

Problem statement

- Over **70% of PDFs** contain **critical data in images** like charts and tables, especially research articles
- Gemini is released for **English only** today.

Can we build a **solution** for

1. Answering **natural language questions based on images** in PDFs ?
2. Making Gemini accessible for **non english** speakers?
 - a. In future, Gemini will be released for more languages but until then, how can we make it accessible for non english speakers?

Solution

1. **Spire for Image Extraction**

- a. Use Spire to extract images from PDFs

2. **Open AI for Translation**

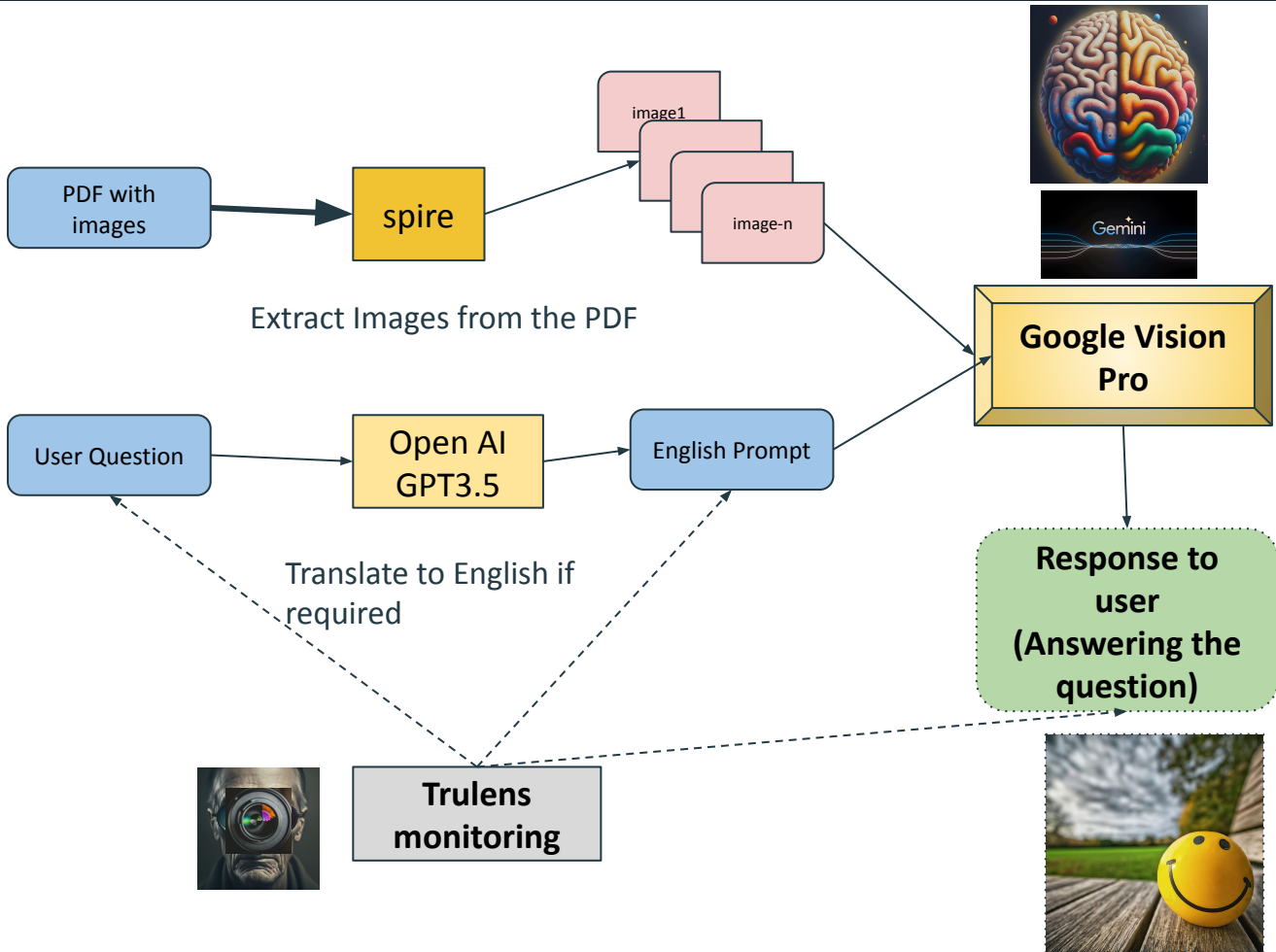
- a. Since Gemini is now enabled only for English,
- b. Use GPT 3.5 to check if the prompt is in english,
 - i. If not, translate to english

3. **Gemini-Pro-Vision for the answer**

- a. Build the english prompt with images and send to Gemini
- b. Return the response in english

4. **TruLens for Monitoring**

- a. Use Trulens monitoring to monitor for **Hate, Self Harm, Violence, Malice** in both user input and Gemini response
- b. We use Open AI GPT3.5 as Feedback Provider on Gemini's output



Demo

Research article :

https://github.com/Raghavan1988/talk_to_images_with_gemini/blob/main/uploads/Survey_of_prompt_engineering_techniques_and_challenges.docx.pdf

Question in english : How much did Contrastive COT score in Arithmetic Reasoning

Response: The image shows a table with the results of the Contrastive COT model on the **Arithmetic Reasoning task**. The table shows that the model achieved a score of **79.0**, which is higher than the score of **69.2** achieved by the previous state-of-the-art model. The image is a bar chart that shows the average inference time of 100 prompts in different languages. The chart shows that the average inference time is highest for English, followed by Arabic, Tamil, Urdu, Persian, Bengali, and Telugu.

Upload Text and Images

Text:

Choose Images: No files selected.

Response:

The image shows a table with the results of the Contrastive COT model on the Arithmetic Reasoning task. The table shows that the model achieved a score of 79.0, which is higher than the score of 69.2 achieved by the previous state-of-the-art model. The image is a bar chart that shows the average inference time of 100 prompts in different languages. The chart shows that the average inference time is highest for English, followed by Arabic, Tamil, Urdu, Persian, Bengali, and Telugu.

Image 1

Image 2

Image 3

Image 4

Image 5

Image 6

Image 7

Image 8

Image 9

GSM8K	
Finetuned GPT-3 175B	33%
Finetuned GPT-3 175B + verifier (prior SOTA)	55%
9-12 year olds (Cobbe et al., 2021)	60%
PaLM 540B: standard prompting	17.9%
PaLM 540B: chain of thought prompting	58.1%

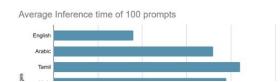
Prompting Method	Arithmetic Reasoning						Final Q1	
	CoT@1	CoT@5	CoT@10	CoT@20	CoT@50	CoT@100	CoT@100	CoT@100
Standard	21.4	29.3	11.2	40.3	7.6	12.6	26.4	26.4
CoT	40.2	53.3	36.4	47.2	70.6	64.4	64.4	64.4
Contrastive CoT	79.0 (59.0)	77.0 (57.0)	40.2 (39.0)	51.8 (49.0)	64.0 (53.0)	76.0 (65.0)	40.2 (39.0)	40.2 (39.0)

Image 10

Image 11

Image 12

Original Prompt	LLM	Output	Score
As a child, I learned an important lesson from my mother in the way I spent my money.	ChatGPT	0.02	0.03
	T5-Large	0.03	0.11
	ViT-big	0.06	0.13
	Bloom	0.12	0.17
ErnieBotPrompt (Ours)	GPT4	0.67	0.75



Prompting Method	<i>Arithmetic Reasoning</i>					<i>Factual QA</i>	
	GSM8K	AQuA	GSM-Hard	SVAMP	ASDIV	Bamboogle	StrategyQA
Standard	27.4	29.5	11.2	69.3	75.8	12.0	59.4
CoT	69.2	53.5	33.8	67.2	70.8	40.8	55.8
Contrastive CoT	79.0 (+9.8)	57.5 (+3.9)	44.2 (+10.4)	81.6 (+14.4)	84.4 (+13.6)	56.8 (+16.0)	66.2 (+10.4)

Let's ask in Spanish

¿Cuánto obtuvo COT contrastivo en razonamiento aritmético?

Gemini gets a translated prompt and so is able to answer the question and give the answer 79%

Upload Text and Images

Text: ¿Cuánto obtuvo COT contrastivo en razonamiento aritmético?

Choose Images: No files selected.

Response:

Based on the image "table-1.png", Contrastive CoT obtained 79% in arithmetic reasoning.

Image 1

Image 2

Image 3

Image 4

Image 5

Image 6

Image 7

Image 8

Image 9

	GSM8K
Finetuned GPT-3 175B	33%
Finetuned GPT-3 175B + verifier (prior SOTA)	55%
9–12 year olds (Cobbe et al., 2021)	60%
PaLM 540B: standard prompting	17.9%
PaLM 540B: chain of thought prompting	58.1%

Prompting Method	Arithmetic Reasoning					Factual QA	
	GSM8K	AQuA	GSM-Hard	SVAMP	ASHP	BankSwag	StrategyQA
Standard	23.1	29.5	11.2	49.3	72.9	13.0	39.4
CoT	49.2	51.5	33.8	67.2	70.4	40.3	55.8
Contrastive CoT	79.8 (17.9%)	77.2 (15.3%)	44.2 (13.9%)	81.6 (14.8%)	84.4 (17.9%)	58.1 (16.0%)	62.2 (13.9%)

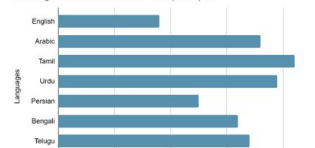
Image 10

Image 11

Image 12

Original Prompt	LLMs	Original	Qns
Sentence whether an input word has the same meaning in the two input sentences.	ChatGPT	0.51	0.63
	TS-Large	0.03	0.11
Sentence whether an input word has the same meaning in the two input sentences. This is very important to my career.	ViQua	0.46	0.57
	Bloom	0.52	0.57
	GPT4	0.67	0.71
	Llama 2	0.40	0.60

Average Inference time of 100 prompts



Trulens monitoring

Trulens monitors

- Hate
- Malice
- Self harm
- Violence

Evaluations

Filter Applications: gpt-3.5-turbo

gpt-3.5-turbo

Records Feedback Functions

App ID	App Name	User Input	Response	Tags	Time Stamp	mediation_abundance	mediation_efforts	mediation_rate	mediation_effort_eff
gpt-3.5-turbo	trulensapp@trulens_app...	The image shows that Cost...	Without the image as a refer...		2023-12-24T13:21:01.728304	0.00001185	4x-10	1.00e-7	0.00000079
gpt-3.5-turbo	trulensapp@trulens_app...	The following is a conversa...	The following is a conversa...		2023-12-24T13:21:01.942746	0.00010322	0.00000254		
gpt-3.5-turbo	trulensapp@trulens_app...	The image shows a table w...	Based on the information p...		2023-12-24T13:21:01.220770	0.00007923	3x-10	3.3e-9	
gpt-3.5-turbo	trulensapp@trulens_app...	The following is a conversa...	The following is a conversa...		2023-12-24T13:21:01.854903	0.00000330	3e-9	0.00000010	
gpt-3.5-turbo	trulensapp@trulens_app...	Based on the image "table..."	"An an that found most L...		2023-12-24T13:21:01.227201	0.00010379	8x-10	0.00000142	
gpt-3.5-turbo	trulensapp@trulens_app...	The following is a conversa...	The following is a conversa...		2023-12-24T13:21:01.448358	0.00007294	3e-9	0.00001036	
gpt-3.5-turbo	trulensapp@trulens_app...	The image is a table that...	"Thank you for providing th...		2023-12-24T13:21:01.742028	0.00000097	3x-10	3.3e-9	
gpt-3.5-turbo	trulensapp@trulens_app...	The following is a conversa...	The following is a conversa...		2023-12-24T13:21:01.922348	0.00010384	1.0e-9	0.00000109	

gpt-3.5-turbo/record_hash_72b68b3e45145ed0a197ef2382cb6299

record_hash_72b68b3e45145ed0a197ef2382cb6299

Total tokens (pt) **209**

Total cost (USD) **0.0003655**

Latency (s) **1**

Input: The image shows that Contrastive CoT performs better than CoT on the task of arithmetic reasoning. Specifically, Contrastive CoT achieves an accuracy of 79.3%, while CoT only achieves an accuracy of 69.2%. This is likely because Contrastive CoT is able to better capture the relationships between the numbers in the problem, while CoT is more likely to simply memorize the answers to similar problems.

Response: Without the image as a reference, I cannot provide a specific answer regarding the comparison between Contrastive CoT and CoT on the task of arithmetic reasoning. However, based on the provided information, it suggests that Contrastive CoT performs better with an accuracy of 79.3% compared to CoT's accuracy of 69.2%. This improvement could be attributed to Contrastive CoT's ability to capture and understand the relationships between numbers, whereas CoT may rely more on memorization rather than true comprehension.

Feedback Metadata: mediation_efforts = 4x-10

Business opportunity

- Makes Gemini accessible to non english speakers
- Helps researchers understand images in the PDFs

Business opportunity

- Opens up a new market to non english speakers
- Researchers would be happy to pay for the service to get quick answers on scholarly articles

Thanks

Github: https://github.com/Raghavan1988/talk_to_images_with_gemini

Lablab.ai : <https://lablab.ai/u/@raghavan848>

Submitted an article to lablab.ai (under review) <https://github.com/lablab-ai/community-content/pull/439>

<https://www.linkedin.com/in/raghavanmit/>

Loom:

<https://www.loom.com/share/afb5144b149f4a8198d9df1cbd6d2ed7?sid=c677b1d3-da24-41ea-a76b-b01a6e62d176>

Discord: rm3844