PolyGlot Gemini lens for PDFs

Raghavan Muthuregunathan
https://www.linkedin.com/in/raghavanmit/
Loom link

Problem statement

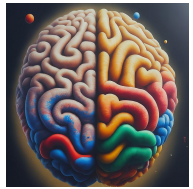
- Over **70% of PDFs** contain **critical data in images** like charts and tables, especially research articles
- Gemini is released for **English only** today.

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



PDF with images

spire

image1

image-n

Extract Images from the PDF

Google Vision Pro

User Question

Open AI GPT3.5

English Prompt

Translate to English if required

Response to user (Answering the question)

Trulens monitoring

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 Arithematic 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.

127.0.0.1:5000

Upload Text and Images

Text: How much did Contrastive COT score n Arithematic Reasoning

Choose Images: Browse... No files selected.

Submit

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

T

Image 3

Image 4

Image 5

Image 6

Image 7

	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%**

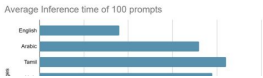
Image 8

Image 9

Image 10

Image 11

Image 12



Prompting Method	*Arithmetic Reasoning*					*Factual QA*	
	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%

127.0.0.1:5000

Upload Text and Images

Text: ¿Cuánto obtuvo COT contrastivo en razonamiento aritmético?

Choose Images: Browse... No files selected.

Submit

Response:

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

Image 1

Image 2

T

Image 3

Image 4

Image 5

Image 6

Image 7

	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%**

Image 8

Image 9

Prompting Method	*Arithmetic Reasoning*					*Factual QA*	
	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)

Image 10

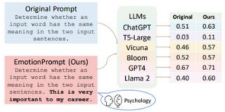
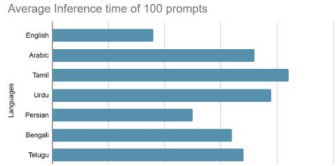


Image 11

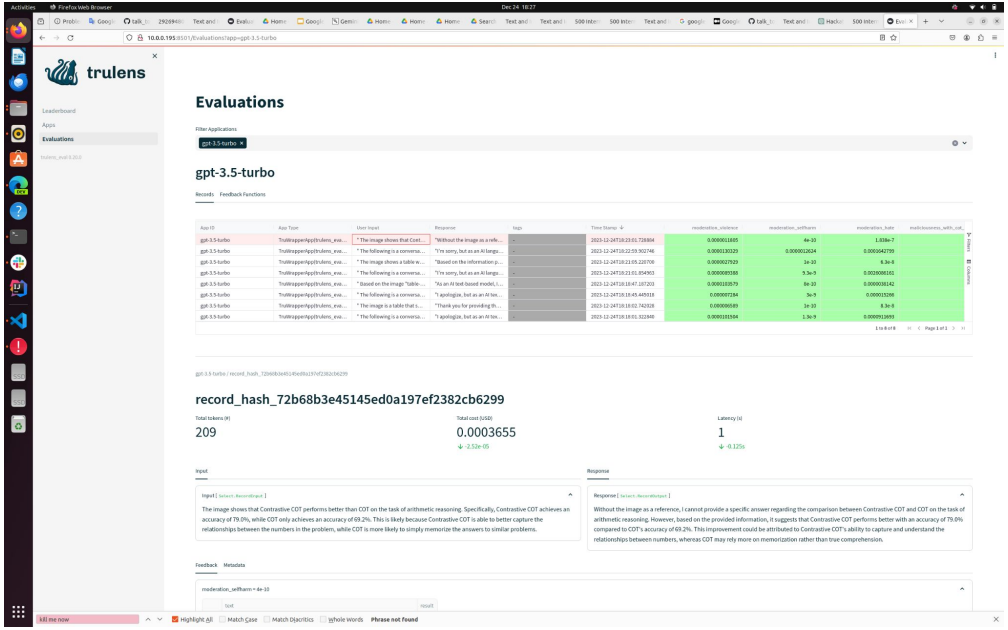
Image 12



Trulens monitoring

Trulens monitors

- Hate
- Malice
- Self harm
- Violence



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