



**GEMINI
VS.
CHATGPT**

An Introduction to Gemini Ultra and ChatGPT4

Gemini Ultra :

"Gemini is a family of multimodal large language models developed by Google DeepMind, serving as the successor to LaMDA and PaLM 2. Comprising Gemini Ultra, Gemini Pro, and Gemini Nano, it was announced on December 6, 2023, positioned as a competitor to OpenAI's GPT-4

ChatGPT4 :

Generative Pre-trained Transformer 4 is a multimodal large language model created by OpenAI, and the fourth in its series of GPT foundation models. It was launched on March 14, 2023, and made publicly available via the paid chatbot product ChatGPT Plus, via OpenAI's API, and via the free chatbot Microsoft Copilot.

Performance Comparison Across Tasks

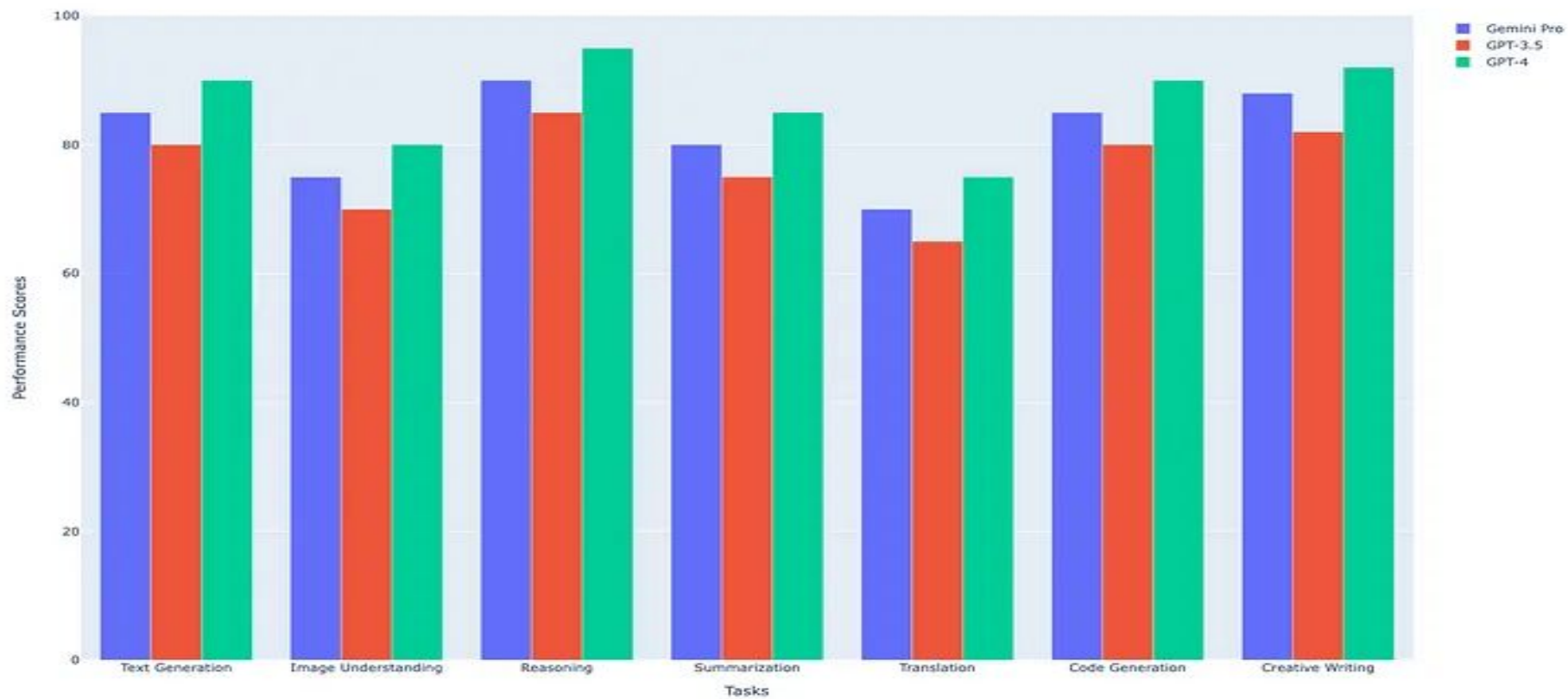
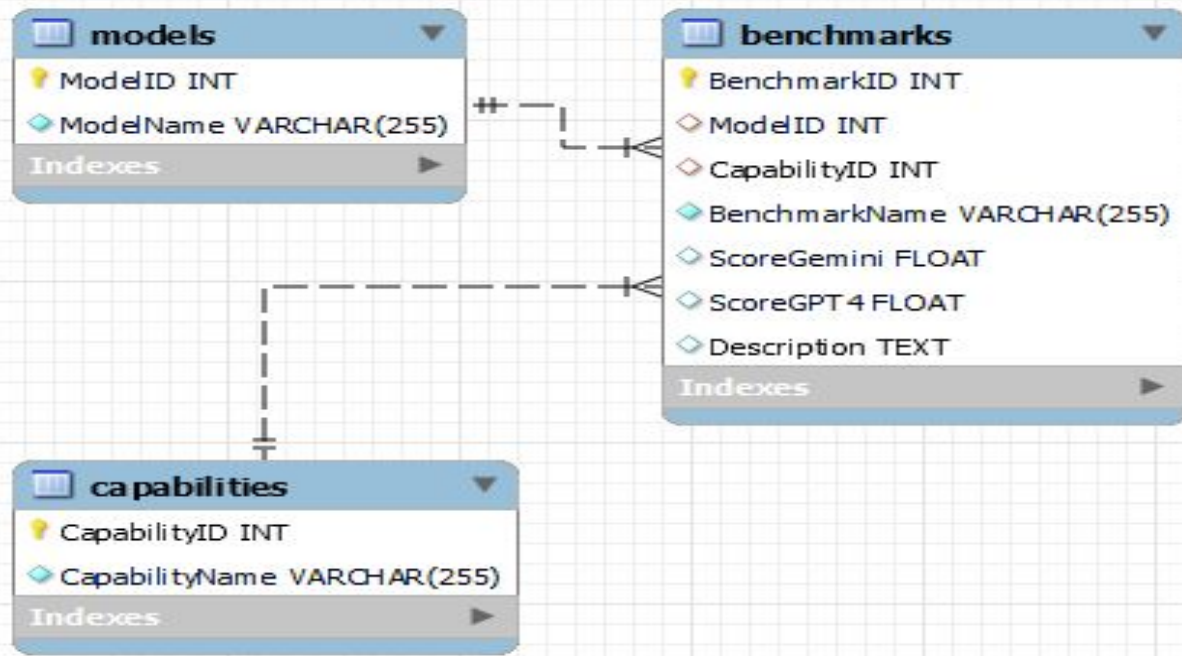


Table Schema :



Tables:

- Benchmarks
- Capabilities
- Models



Question 1 :

What are the average scores for each capability on both the Gemini Ultra and GPT-4 models

Answer :

```
select CapabilityName,  
round(avg(ScoreGemini),2) as "AvgGemini Score",  
round(avg(ScoreGPT4),2) as "AvgGPT4 Score"  
from benchmarks  
join capabilities on  
benchmarks.CapabilityID=capabilities.CapabilityID  
group by(CapabilityName);
```

CapabilityName	AvgGemini Score	AvgGPT4 Score
General	88.2	86.4
Reasoning	85.52	86.43
Math	73.13	72.45
Code	72.55	70.45
Image	73.95	70.9
Video	58.7	51.15
Audio	23.85	23.35

Question 2 :

Which benchmarks does Gemini Ultra outperform GPT-4 in terms of scores?

Answer :

select BenchmarkID,
BenchmarkName,ScoreGemini,
ScoreGPT4, Description from
benchmarks
where
ScoreGemini>ScoreGPT4;

BenchmarkID	BenchmarkName	ScoreGemini	ScoreGPT4	Description
1	MMLU	90	86.4	Representation of questions in 57 subjects
3	Big-Bench Hard	83.6	83.1	Diverse set of challenging tasks requiring multi-s...
5	DROP	82.4	80.9	Reading comprehension (F1 Score)
9	GSM8K	94.4	92	Basic arithmetic manipulations, incl. Grade Scho...
11	MATH	53.2	52.9	Challenging math problems, incl. algebra, geom...
13	HumanEval	74.4	67	Python code generation
15	Natura12Code	74.9	73.9	Python code generation. New held out dataset ...
17	MIMMU	59.4	56.8	Multi-discipline college-level reasoning problems
18	VQAv2	77.8	77.2	Natural image understanding
19	TextVQA	82.3	78	OCR on natural images
20	DocVQA	90.9	88.4	Document understanding
21	Infographic VQA	80.3	75.1	Infographic understanding
22	MathVista	53	49.9	Mathematical reasoning in visual contexts
23	VATEX	62.7	56	English video captioning (CIDEr)
24	Perception Test...	54.7	46.3	Video question answering
25	CoVOST 2	40.1	29.1	Automatic speech translation (BLEU score)
NULL	NULL	NULL	NULL	NULL

Question 3 :

What are the highest scores achieved by Gemini Ultra and GPT-4 for each benchmark in the Image capability?

Answer :

```
select CapabilityName, BenchmarkID,  
BenchmarkName,  
ScoreGemini, ScoreGPT4  
from benchmarks
```

```
join capabilities on benchmarks.CapabilityID = capabilities.CapabilityID  
where CapabilityName="Image" ;
```

	CapabilityName	BenchmarkID	BenchmarkName	ScoreGemini	ScoreGPT4
►	Image	17	MIMMU	59.4	56.8
	Image	18	VQAv2	77.8	77.2
	Image	19	TextVQA	82.3	78
	Image	20	DocVQA	90.9	88.4
	Image	21	Infographic VQA	80.3	75.1
	Image	22	MathVista	53	49.9

Question 4 :

Calculate the percentage improvement of Gemini Ultra over GPT-4 for each benchmark?

Answer :

```
select BenchmarkName,  
concat(round(((ScoreGemini-ScoreGPT4) /  
(ScoreGemini+ScoreGPT4))*100,2),'%')  
as Percentage_Improved  
from benchmarks  
group by BenchmarkName,ScoreGemini,ScoreGPT4  
having Percentage_Improved >0;
```

BenchmarkName	Percentage_Improved
MMLU	2.04%
Big-Bench Hard	0.3%
DROP	0.92%
GSM8K	1.29%
MATH	0.28%
HumanEval	5.23%
Natura12Code	0.67%
MIMMU	2.24%
VQAv2	0.39%
TextVQA	2.68%
DocVQA	1.39%
Infographic VQA	3.35%
MathVista	3.01%
VATEX	5.64%
Perception Test...	8.32%
CoVOST 2	15.9%

Question 5 :

Retrieve the benchmarks where both models scored above the average for their respective models?

Answer :

```
select  
BenchmarkName,ScoreGemini,ScoreGPT4,ModelID  
from benchmarks  
where ScoreGemini >=  
(select avg(ScoreGemini) from benchmarks) and  
ScoreGPT4>=  
(select avg(ScoreGPT4) from benchmarks);
```

Result Grid					Filter Rows:	Export:
	BenchmarkName	ScoreGemini	ScoreGPT4	ModelID		
▶	MMLU	90	86.4	1		
	Big-Bench Hard	83.6	83.1	1		
	DROP	82.4	80.9	1		
	HellaSwag	87.8	95.3	1		
	GSM8K	94.4	92	1		
	HumanEval	74.4	67	1		
	Natura12Code	74.9	73.9	1		
	VQAv2	77.8	77.2	2		
	TextVQA	82.3	78	1		
	DocVQA	90.9	88.4	2		
	Infographic VQA	80.3	75.1	1		

Question 6 :

Which benchmarks show that Gemini Ultra is expected to outperform GPT-4 based on the next score?

Answer :

```
select BenchmarkID, BenchmarkName,  
ScoreGemini, ScoreGPT4, Description  
from benchmarks  
where ScoreGemini < ScoreGPT4;
```

	BenchmarkID	BenchmarkName	ScoreGemini	ScoreGPT4	Description
▶	7	HellaSwag	87.8	95.3	Commonsense reasoning for everyday tasks
	26	FLEURS	7.6	17.6	Automatic speech recognition (word error rate)

Question 7 :

Classify benchmarks into performance categories based on score ranges?

Answer :

```
select BenchmarkName,  
case  
when ScoreGemini>ScoreGPT4 then "Good"  
when ScoreGemini<ScoreGPT4 then "Poor"  
when ScoreGemini=ScoreGPT4 then "Average"  
else "very bad"  
end as "performance_category"  
from benchmarks  
where ScoreGPT4 is not null;
```

	BenchmarkName	performance_category
▶	MMLU	Good
	Big-Bench Hard	Good
	DROP	Good
	HellaSwag	Poor
	GSM8K	Good
	MATH	Good
	HumanEval	Good
	Natura12Code	Good
	MIMMU	Good
	VQAv2	Good
	TextVQA	Good
	DocVQA	Good
	Infographic VQA	Good
	MathVista	Good
	VATEX	Good
	Perception Test...	Good
	CoVOST 2	Good
	FLEURS	Poor

Result 23 ×

Question 8 :

Retrieve the rankings for each capability based on Gemini Ultra scores?

Answer :

```
select CapabilityName,ScoreGemini,  
dense_rank()  
over(partition by CapabilityName  
order by ScoreGemini desc ) as "Rank"  
from benchmarks  
join capabilities on  
benchmarks.CapabilityID=  
capabilities.CapabilityID  
order by(CapabilityName);
```

	CapabilityName	ScoreGemini	Rank
▶	Audio	40.1	1
	Audio	7.6	2
	Code	74.9	1
	Code	74.4	2
	Code	73.9	3
	Code	67	4
	General	90	1
	General	86.4	2
	Image	90.9	1
	Image	82.3	2
	Image	80.3	3
	Image	77.8	4
	Image	59.4	5
	Image	53	6
	Math	94.4	1

Result 24 x

Question 9 :

Convert the Capability and Benchmark names to uppercase?

Answer :

```
select upper(CapabilityName)
as CapabilityName,
upper(BenchmarkName) as BenchmarkName
from benchmarks
join capabilities on benchmarks.CapabilityID=
capabilities.CapabilityID;
```

Result Grid

Filter Rows:

	CapabilityName	BenchmarkName
▶	GENERAL	MMLU
	GENERAL	MMLU
	REASONING	BIG-BENCH HARD
	REASONING	BIG-BENCH HARD
	REASONING	DROP
	REASONING	DROP
	REASONING	HELLASWAG
	REASONING	HELLASWAG
	MATH	GSM8K
	MATH	GSM8K
	MATH	MATH
	MATH	MATH
	CODE	HUMANEVAL
	CODE	HUMANEVAL
	CODE	NATURA12CODE

Result 25 ×

Question 10 :

Can you provide the benchmarks along with their descriptions in a concatenated format?

Answer :

```
select  
concat(BenchmarkName,' - ',Description)  
as "Benchmark Description"  
from benchmarks;
```

Result Grid		Filter Rows:	Export
	Benchmark Description		
▶	MMLU - Representation of questions in 57 su...		
	MMLU - Representation of questions in 57 su...		
	Big-Bench Hard - Diverse set of challenging t...		
	Big-Bench Hard - Diverse set of challenging t...		
	DROP - Reading comprehension (F1 Score)		
	DROP - Reading comprehension (F1 Score)		
	HellaSwag - Commonsense reasoning for ev...		
	HellaSwag - Commonsense reasoning for ev...		
	GSM8K - Basic arithmetic manipulations, incl. ...		
	GSM8K - Basic arithmetic manipulations, incl. ...		
	MATH - Challenging math problems, incl. alge...		
	MATH - Challenging math problems, incl. alge...		
	HumanEval - Python code generation		
	HumanEval - Python code generation		
	Natura12Code - Python code generation. Ne...		
	Natura12Code - Python code generation		
	MIMMU - Multi-discipline college-level reasoni...		
	VQAv2 - Natural image understanding		

Result 26 ×



THANK YOU