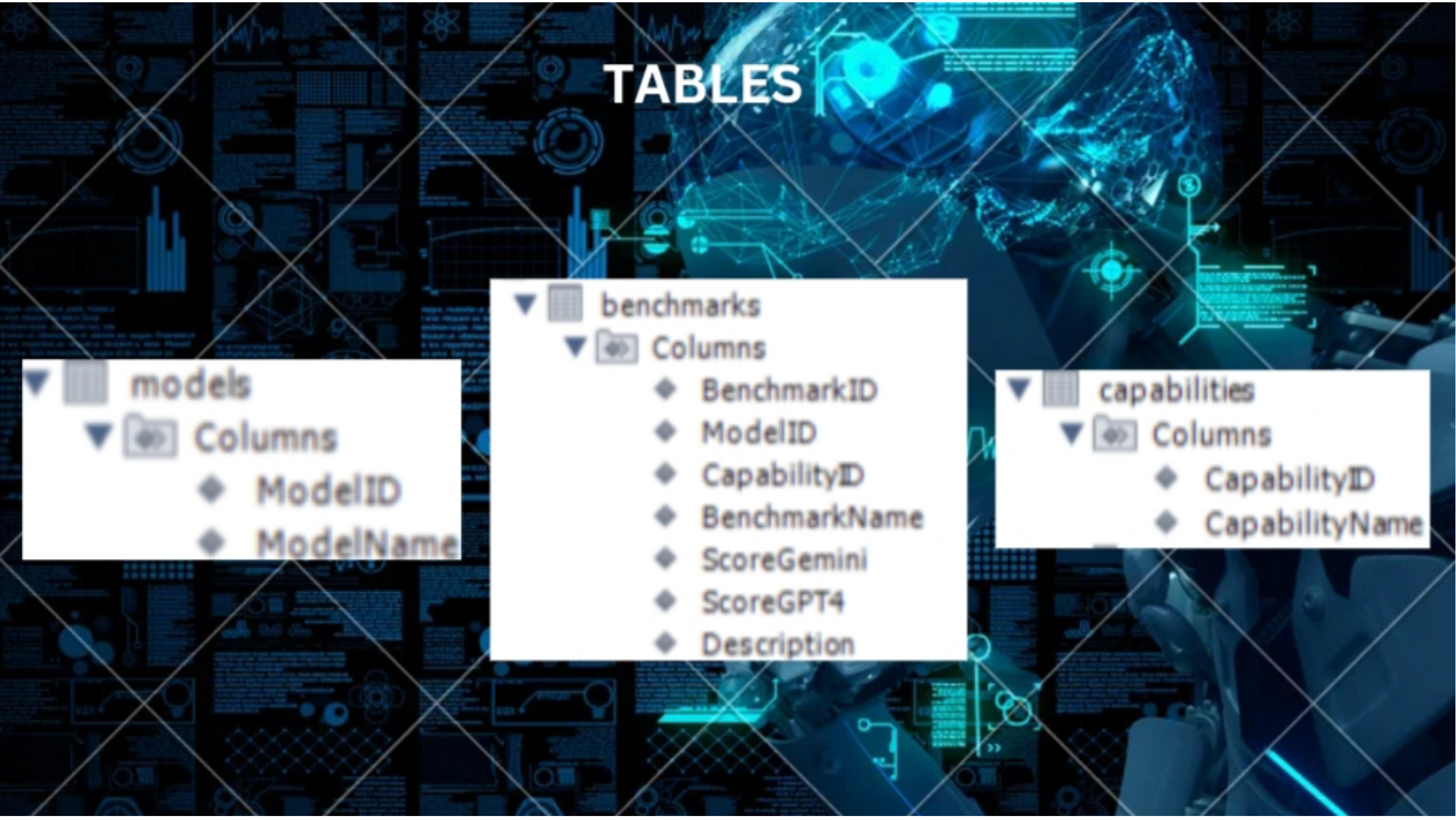

Gemini
VS
ChatGPT

TABLES


models
Columns
ModelID
ModelName
benchmarks
Columns
BenchmarkID
ModelID
CapabilityID
BenchmarkName
ScoreGemini
ScoreGPT4
Description
capabilities
Columns
CapabilityID
CapabilityName

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

```sql
SELECT
    ROUND(AVG(B.ScoreGPT4), 2) AS GPT4,
    ROUND(AVG(B.ScoreGemini), 2) AS Gemini_Ultra,
    CapabilityName
FROM benchmarks B
JOIN capabilities C USING (CapabilityID)
GROUP BY CapabilityName;
```

GPT4	Gemini_Ultra	CapabilityName
86.4	88.2	General
86.43	85.52	Reasoning
72.45	73.13	Math
70.45	72.55	Code
70.9	73.95	Image
51.15	58.7	Video
23.35	23.85	Audio

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

```sql
SELECT
    b1.BenchmarkName
    ,ROUND(SUM(b1.ScoreGemini), 2) AS gemini
    ,ROUND(SUM(b2.ScoreGPT4),2) AS GPT
FROM benchmarks b1
JOIN benchmarks b2 ON b1.BenchmarkName = b2.BenchmarkName
GROUP BY b1.BenchmarkName
HAVING ROUND(SUM(b1.ScoreGemini), 2) > ROUND(SUM(b2.ScoreGPT4), 2);
```

BenchmarkName	gemini	GPT
MMLU	352.8	172.8
Big-Bench Hard	333.4	166.2
DROP	326.6	161.8
HellaSwag	366.2	190.6
GSM8K	372.8	184
MATH	212.2	105.8
HumanEval	282.8	134
Natural2Code	297.6	147.8
MMMU	59.4	56.8
VQAv2	77.8	77.2
TextVQA	82.3	78
DocVQA	90.9	88.4
Infographic VQA	80.3	75.1
MathVista	53	49.9
VATEX	62.7	56
Perception Test...	54.7	46.3
CoVOST 2	40.1	29.1

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

```sql
SELECT
    ROUND(SUM(ScoreGemini), 2) AS Gemini,
    ROUND(SUM(ScoreGPT4), 2) AS GPT,
    BenchmarkName
FROM benchmarks
JOIN capabilities C USING (capabilityID)
WHERE C.CapabilityName = 'Image'
GROUP BY benchmarkname;
```

Gemini	GPT	BenchmarkName
59.4	56.8	MIMMU
77.8	77.2	VQAv2
82.3	78	TextVQA
90.9	88.4	DocVQA
80.3	75.1	Infographic VQA
53	49.9	MathVista

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

```sql
SELECT
    BenchmarkName,
    CONCAT(ROUND(((ScoreGemini - ScoreGPT4) / SUM(ScoreGemini + ScoreGPT4)) * 100,2),'%')
    AS improvement_percentage
FROM benchmarks
GROUP BY BenchmarkName , ScoreGemini , ScoreGPT4
HAVING improvement_percentage > 0;
```

BenchmarkName	improvement_percentage
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%

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

```sql
SELECT benchmarkname, ScoreGemini, ScoreGPT4
FROM benchmarks
WHERE ScoreGemini > (
    SELECT ROUND(AVG(ScoreGemini), 2)
        FROM benchmarks)
        AND
        ScoreGPT4 > (
        SELECT ROUND(AVG(ScoreGPT4), 2)
        FROM benchmarks);
```

benchmarkname	ScoreGemini	ScoreGPT4
MMLU	90	86.4
Big-Bench Hard	83.6	83.1
DROP	82.4	80.9
HellaSwag	87.8	95.3
GSM8K	94.4	92
HumanEval	74.4	67
Natura12Code	74.9	73.9
VQAv2	77.8	77.2
TextVQA	82.3	78
DocVQA	90.9	88.4
Infographic VQA	80.3	75.1

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

```
select Benchmarkname from
(select benchmarkname,
Scoregemini,
ScoreGPT4,
lead(scoregemini) over (order by scoregemini) as LeadGem
from benchmarks
where ScoreGPT4 is not null
) as NextScore
where LeadGem > ScoreGPT4;
```

Benchmarkname
FLEURS
CoVOST 2
MathVista
MATH
Perception Test MCQA
MIMMU
VATEX
HumanEval
Natura12Code
VQAv2
Infographic VQA
TextVQA
DROP
Big-Bench Hard
MMLU

7)Classify benchmarks into performance categories based on score ranges?

```sql
SELECT
    benchmarkname,
    Scoregemini,
    ScoreGPT4,
    CASE
        WHEN Scoregemini >= 75 THEN 'Excellent'
        WHEN Scoregemini >= 55 AND Scoregemini < 75 THEN 'Good'
        WHEN Scoregemini >= 45 AND Scoregemini < 55 THEN 'Not Bad'
        WHEN Scoregemini >= 35 AND Scoregemini < 45 THEN 'Poor'
        ELSE 'Very Poor'
    END AS Gemini_Performance_cat_wise,
    CASE
        WHEN ScoreGPT4 >= 75 THEN 'Excellent'
        WHEN ScoreGPT4 >= 55 AND ScoreGPT4 < 75 THEN 'Good'
        WHEN ScoreGPT4 >= 45 AND ScoreGPT4 < 55 THEN 'Not Bad'
        WHEN ScoreGPT4 >= 35 AND ScoreGPT4 < 45 THEN 'Poor'
        ELSE 'Very Poor'
    END AS GPT4_Performance_cat_wise
FROM benchmarks
WHERE ScoreGPT4 IS NOT NULL;
```

benchmarkname	Scoregemini	ScoreGPT4	Gemini_Performance_cat_wise	GPT4_Performance_cat_wise
MMLU	90	86.4	Excellent	Excellent
Big-Bench Hard	83.6	83.1	Excellent	Excellent
DROP	82.4	80.9	Excellent	Excellent
HellaSwag	87.8	95.3	Excellent	Excellent
GSM8K	94.4	92	Excellent	Excellent
MATH	53.2	52.9	Not Bad	Not Bad
HumanEval	74.4	67	Good	Good
Natura12Code	74.9	73.9	Good	Good
MIMMU	59.4	56.8	Good	Good
VQAv2	77.8	77.2	Excellent	Excellent
TextVQA	82.3	78	Excellent	Excellent
DocVQA	90.9	88.4	Excellent	Excellent
Infographic VQA	80.3	75.1	Excellent	Excellent
MathVista	53	49.9	Not Bad	Not Bad
VATEX	62.7	56	Good	Good
Perception Test...	54.7	46.3	Not Bad	Not Bad
CoVOST 2	40.1	29.1	Poor	Very Poor
FLEURS	7.6	17.6	Very Poor	Very Poor

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

```sql
SELECT
    Scoregemini,
    C.capabilityName,
    DENSE_RANK() OVER (ORDER BY Scoregemini) AS ranking
FROM benchmarks B
JOIN capabilities C USING (capabilityID);
```

Scoregemini	capabilityName	ranking
7.6	Audio	1
40.1	Audio	2
52.9	Math	3
53	Image	4
53.2	Math	5
54.7	Video	6
59.4	Image	7
62.7	Video	8
67	Code	9
73.9	Code	10
74.4	Code	11
74.9	Code	12
77.8	Image	13
80.3	Image	14
80.9	Reasoning	15
82.3	Image	16
82.4	Reasoning	17
83.1	Reasoning	18
83.6	Reasoning	19

9)Convert the Capability and Benchmark names to

```
SELECT UPPER(B.benchmarkname) AS Benchmark
      , UPPER(C.capabilityname) AS Capability
FROM benchmarks B
JOIN capabilities C USING (CapabilityID);
```

	Benchmark	Capability
▶	MMLU	GENERAL
	MMLU	GENERAL
	BIG-BENCH HARD	REASONING
	BIG-BENCH HARD	REASONING
	DROP	REASONING
	DROP	REASONING
	HELLASWAG	REASONING
	HELLASWAG	REASONING
	GSM8K	MATH
	GSM8K	MATH
	MATH	MATH
	MATH	MATH
	HUMANEVAL	CODE
	HUMANEVAL	CODE
	NATURA 12CODE	CODE
	NATURA 12CODE	CODE
	MIMMU	IMAGE
	VQAV2	IMAGE
	TEXTVQA	IMAGE

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

```
SELECT CONCAT(benchmarkname, '   ->    ', description) AS benchmark_descriptions
FROM benchmarks
```

benchmark_descriptions
MMLU -> Representation of questions in 57 s...
MMLU -> Representation of questions in 57 s...
Big-Bench Hard -> Diverse set of challenging t...
Big-Bench Hard -> Diverse set of challenging t...
DROP -> Reading comprehension (Fl Score)
DROP -> Reading comprehension (Fl 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. alg...
MATH -> Challenging math problems, incl. alg...
HumanEval -> Python code generation
HumanEval -> Python code generation
Natura12Code -> Python code generation. N...
Natura12Code -> Python code generation
MIMMU -> Multi-discipline college-level reasoni...
VQAv2 -> Natural image understanding
TextVQA -> OCR on natural images