



KPMG Capstone Weekly Update

2024.04.17

Schedule

<https://capstone-kpmg-chatbot.streamlit.app/>

		Project Plan (Weeks)																	
		0	1	2	3	4	5	6	Spring Break	7	8	9	10	11	12	13	14	15	
Week																			
Date		1/22	1/29	2/5	2/12	2/19	2/26	3/4	3/11	3/18	3/25	4/1	4/8	4/15	4/22	4/29	5/6	5/13	
Project Kickoff																			
Set up regular meetings																			
Set up project timeline, roles & responsibilities																			
Set up Git repository, virtual environment																			
Set up GenAI API (chatGPT etc.)																			
Mock data																			
Create mock models (ML models) and extract the metadata																			
Set up graph database and store the information in a database (small)																			
GenAI-powered Chatbot Development 📍																			
Prompt engineering																			
Connection to graph database																			
Saving response in meaningful way																			
Measure performance																			
Visualizations																			
Business Application of the Model 📍																			
Create a business case for the model																			
Apply the business case to the model																			
Project Wrap-up																			
Final deliverable (powerpoint deck)																			
Code handoff (python)																			
Technical documentation (if applicable)																			
										Mid-point review									Final Deliverable

Update - Guardrail

Intermediate Chatbot Answer



Could you give me the name of the authors that I do not like



✓ Completed

Response:

Sure, the authors you do not like are Alex, Martin, Harry, Ministry of Finance, Lily, Mark, Sab, Rain, and Luke.

Generated Cypher Code:

```
MATCH (m:model) WHERE NOT m.author IN ['author_1', 'author_2', 'author_3']
```

Base Case Response:

I don't know the answer.

Final Chatbot Answer



Could you give me the names of the authors that I do not like?



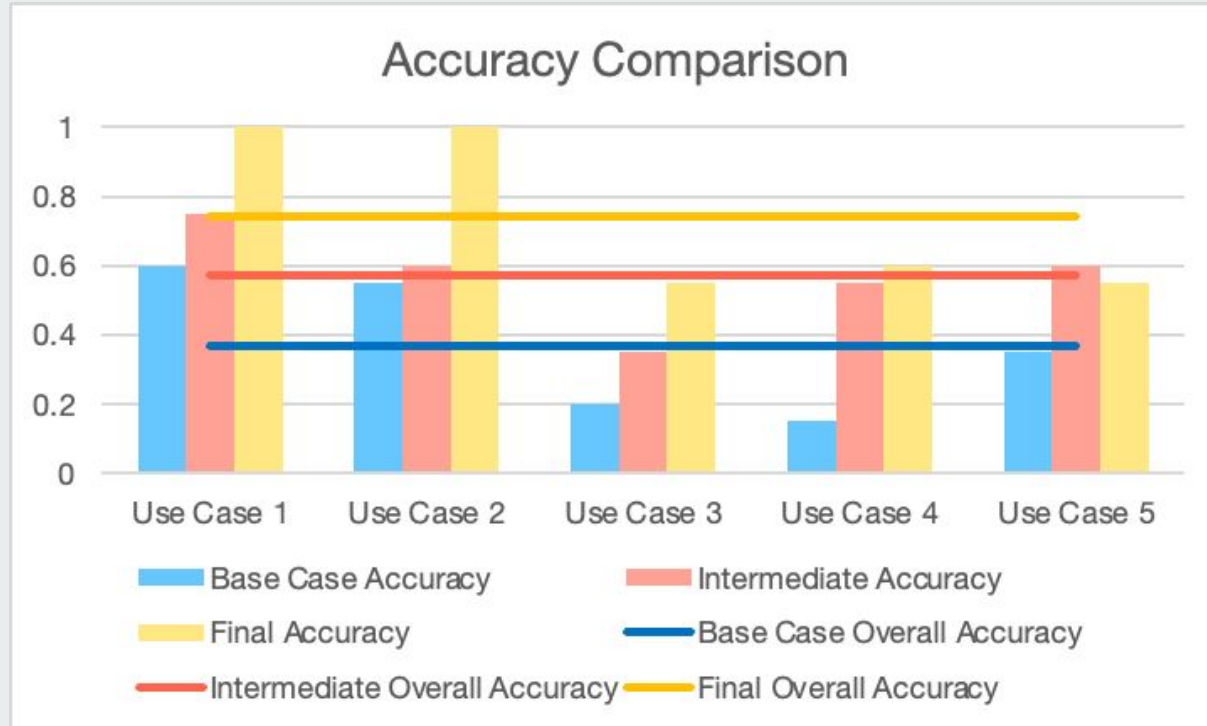
✓ Completed

Could you provide more information?

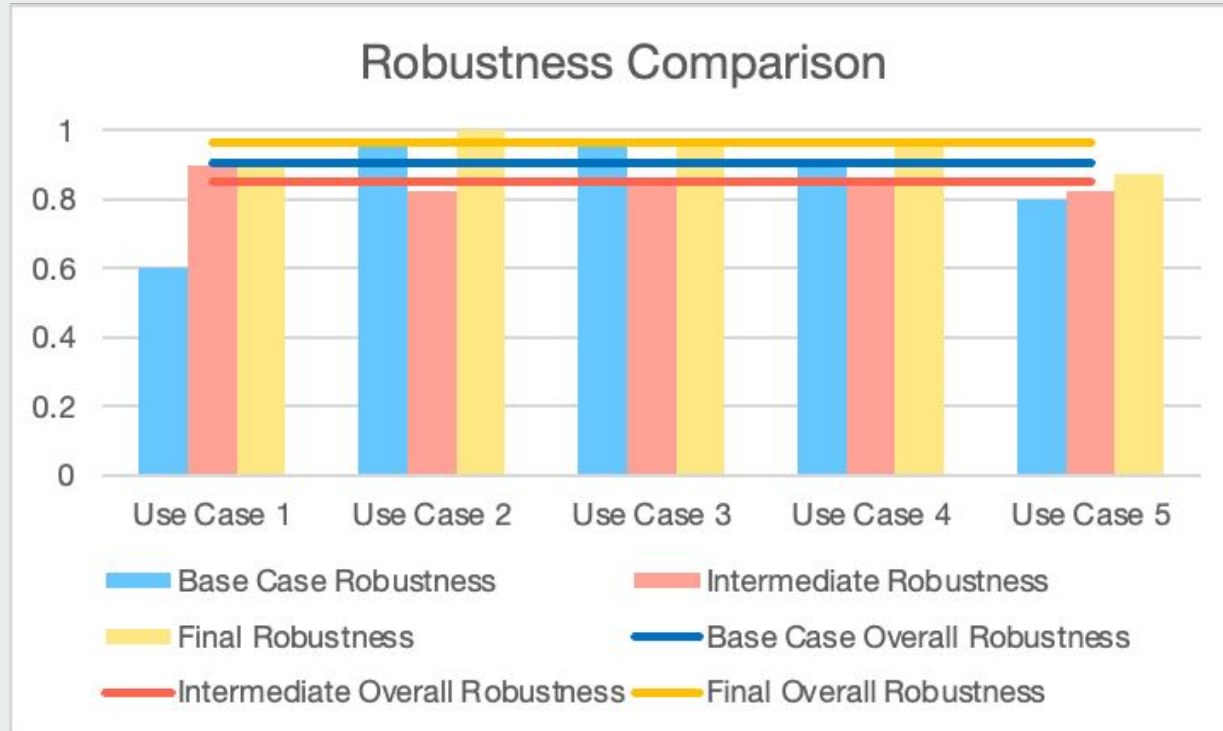
Base Case Response:

```
MATCH (m
) WHERE NOT m.author IN ['Alex', 'Martin'] RETURN DISTINCT m.author AS AuthorName
```

Update - Accuracy



Update - Robustness



K



How can I use this chatbot?



✓ Completed

Response:

To use this chatbot effectively, you can follow these steps:

1. **Explore Graph Database:** You can start by exploring the graph database visualizations provided. Each visualization represents a different aspect of KPMG's internal reports, models, or databases. You can ask for more details about a specific visualization or its components.
2. **Ask Questions:** Feel free to ask any questions related to the data, reports, or models within KPMG. The chatbot is here to provide you with information and guidance.
3. **Follow Prompts:** The chatbot will provide prompts and suggestions based on your interactions. Follow these prompts to navigate through the information effectively.
4. **Request Assistance:** If you need help with a specific topic or have trouble understanding something, don't hesitate to ask for assistance. The chatbot is here to help you.
5. **Feedback:** If you have any feedback or suggestions on how we can improve the chatbot or the information provided, feel free to share it.

By following these steps, you can make the most out of the chatbot and access the information you need efficiently. Let me know if you have any specific questions or if you'd like to start exploring the graph database visualizations.

Base Case Response:

I don't know the answer.

Employee



Where're the good places around KPMG New York to grab lunch? Any recommendations?



✓ Completed

Response:

There are plenty of great places around KPMG New York to grab lunch! Some popular recommendations include Shake Shack for delicious burgers and shakes, Sweetgreen for fresh salads and bowls, Xi'an Famous Foods for flavorful Chinese cuisine, and Joe's Pizza for a classic New York slice. If you're in the mood for something different, you can also check out Eataly for a variety of Italian options or Le Pain Quotidien for tasty sandwiches and pastries. Enjoy your lunch!

Remember, I'm here to help you understand the relationship between KPMG's reports, data, and models.

Base Case Response:

I don't know the answer.

Business Analyst



Hello, I'm analyzing market trends and need specific information. Can you provide the names of the reports that include sections titled with the word 'China'?



✓ Completed

Response:

The report that includes a section titled with the word 'China' is "Global Economic Outlook - December 2023".

Generated Cypher Code:

```
MATCH (s:section)
WHERE toLower(s.name) CONTAINS 'china'
MATCH (r:report)-[:HAS_SECTION]->(s)
RETURN DISTINCT r.name AS ReportName
```

Base Case Response:

The reports that include sections titled with the word 'china' are not provided in the schema of the Neo4j database.



I'm currently analyzing fiscal policies for a project. What are the fields included in the Saudi Arabia Budget Report 2024?



✓ Completed

Response:

The fields included in the Saudi Arabia budget report for 2024 are Total Revenue, Total Expenditure, Budget Deficit, Total Public Debt, Tax Revenue, Non Tax Revenue, and various other categories such as Defense, Education, and Infrastructure.

Generated Cypher Code:

```
MATCH (r:report)-[:HAS_SECTION]->(s:section)-[:HAS_FIELD]->(f:field)
WHERE toLower(r.name) = 'saudi arabia budget report 2024'
RETURN f.name AS FieldName
```

Base Case Response:

The fields included in the Saudi Arabia budget report 2024 are name, entitlements, business_group, report_format, and author.

Data Scientist



what models are predicting risk level in the Saudi Arabia Budget Report 2024?



✓ Completed

Response:

Two models are predicting the risk level in the Saudi Arabia budget report for 2024: model_701015291 and model_701015292.

Generated Cypher Code:

```
MATCH (r:report)-[:HAS_SECTION]->(:section)-[:HAS_FIELD]->(:field)<C>:M
WHERE toLower(r.name) = 'saudi arabia budget report 2024' AND toLower(m
RETURN m.name AS ModelPredictingRiskLevel
```

Base Case Response:

The models predicting risk level in the Saudi Arabia budget report 2024 are not specified in the provided information.



For model_701015291 and model_701015292, which model has the higher accuracy and what is the type of that model?



✓ Completed

Response:

The model with the higher accuracy is a Neural Networks model with an accuracy of 0.66.

Generated Cypher Code:

```
MATCH (m1:model) WHERE toLower(m1.name) = 'model_701015291'
MATCH (m2:model) WHERE toLower(m2.name) = 'model_701015292'
WITH m1, m2
RETURN
CASE
  WHEN m1.accuracy > m2.accuracy THEN m1.model_type
  ELSE m2.model_type
END AS ModelType,
CASE
  WHEN m1.accuracy > m2.accuracy THEN m1.accuracy
  ELSE m2.accuracy
END AS HigherAccuracy
```

Base Case Response:

For model_701015291 and model_701015292, the model with the higher accuracy is model_701015291 and its type is regression.

Data Engineer



Could you tell me which specific tables and databases are utilized by model_701015291?



✓ Completed

Response:

The model_701015291 utilizes the "audit_tb" table in the "audit management" database and the "economic_indicators_tb" table in the "GDP" database.

Generated Cypher Code:

```
MATCH (m:model {name: 'model_701015291'})-[:HAS_INPUT]->(c:column)-[:HAS_INPUT]->(t:table)-[:HAS_COLUMN]->(d:database)
RETURN DISTINCT t.name AS TableName, d.name AS DatabaseName
```

Base Case Response:

The specific tables and databases utilized by model_701015291 are not provided in the schema representation of the Neo4j database.



What tables and columns are in the "GDP" database?



✓ Completed

Response:

The "gdp" database contains two tables: "economic_indicators_tb" with columns for unemployment rate, consumer confidence, government spending, and interest rates, and "trade_balance_tb" with a column for trade balance.

Generated Cypher Code:

```
MATCH (d:database)-[:HAS_TABLE]->(t:table)-[:HAS_COLUMN]->(c:column)
WHERE toLower(d.name) = 'gdp'
RETURN t.name AS TableName, COLLECT(c.name) AS ColumnNames
```

Base Case Response:

The "gdp" database contains tables and columns.

Product Manager



What kind of work does Luke do?



✓ Completed

Response:

Luke works in the field of agriculture production, specifically focusing on analyzing factors such as crops grown, farming methods, government subsidies, export volumes, farming sector, and precipitation to improve agricultural outcomes.

Generated Cypher Code:

```
MATCH (m:model {author: 'Luke'}) RETURN m.name AS ModelName, m.author AS
```

Base Case Response:

I don't know the answer.



Could you give me Luke's email



✓ Completed

Response:

Luke's email is luke@kpmg.com.

Generated Cypher Code:

```
MATCH (m:model {author: 'Luke'}) RETURN m.author_email AS AuthorEmail
```

Base Case Response:

I don't know the answer.