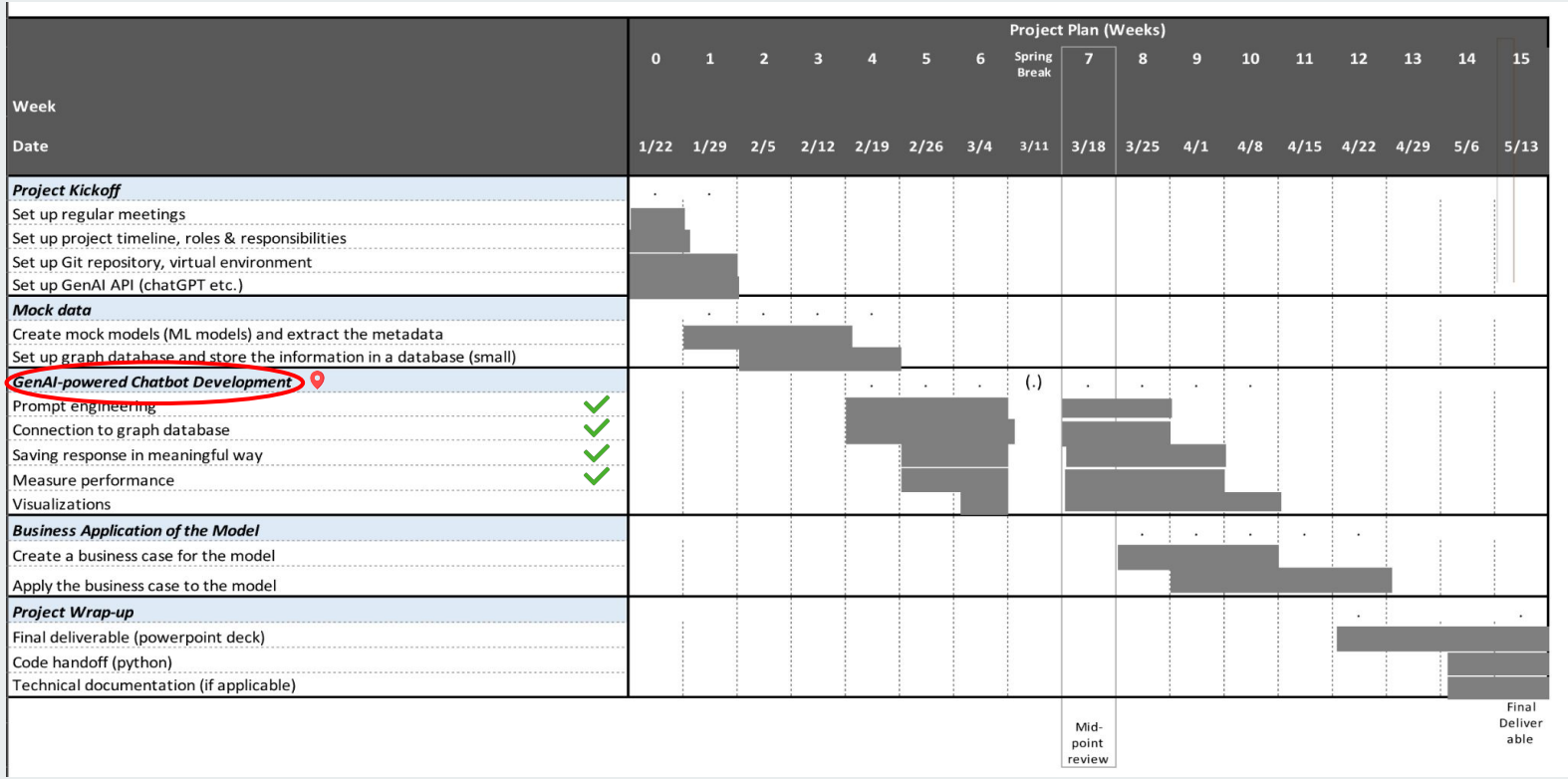
# KPMG Capstone Weekly Update

2024.3.27

# Schedule


Project Plan (Weeks)
Week
Date
0 1 2 3 4 5 6 Spring Break 7 8 9 10 11 12 13 14 15
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

1. Iteration 1: Naive Model

2. Iteration 2: 3-layer Prompt Engineering GPT Model

3. Iteration 3: Untuned Model + Streamlit setup
   - Back and forth questions
   - https://capstone-kpmg-chatbot.streamlit.app/

4. Performance Metrics Comparison

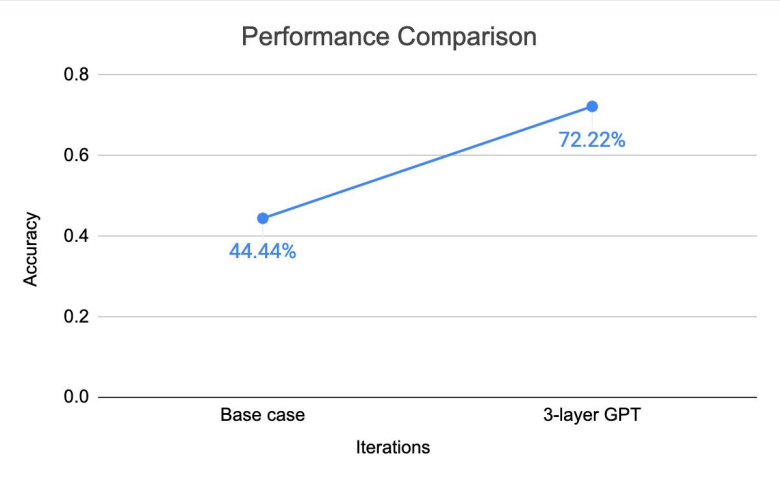
# Use Cases

1. Use Case 1: Simple Extraction (one entity)
2. Use Case 2: Simple Extraction (two or more entities)
3. Use Case 3: Complex Extraction with WHERE Clause
4. Use Case 4: Find Relationships
5. Use Case 5: YES/NO Verification Problem

https://docs.google.com/spreadsheets/d/1MIEyXIJGTAw80rXpV-S9c8Gahxqouj6ykOwSm1t7N4U/edit#gid=0

# Performance Metrics

1. **Accuracy** for response: (correct answer:1, incorrect answer:0)
https://docs.google.com/spreadsheets/d/13pAL-CoYBonbjFMU1apB_iwk-cMWQX1V1ZSlv6zS8ms/edit#gid=0




2. **Consistency** for response: we plan to run the same question 5 times and see whether the answers generated are similar/same. However, due to budget constraints, we have not yet initiated this phase of the project.

# Challenges

1. Definition of Performance metrics, such as Robustness

2. back-and-forth improvement, such as
    a. chatbot questioning ( I don't understand → I don't understand, do you mean xxx?
    b. previous question references (Q: Can you provide information for **the previous model**?)
    c. question synonym or misspelling
        (To answer question like "who is the author/owner/maintainer of the model?")

# Next week Plan

1. Tune the iteration 3 model and test the performance

2. Automate the testing pipeline

# Our Questions

- In the roadmap, what do you expect us to do about the Visualization part?
- Any advice on testing the performance?
- Business case