Embedding Model Showdown: MiniLM vs. MPNet

1. Executive Summary

- **Model A (all-MiniLM-L6-v2):** Demonstrated unexpected strength in identifying **Numerical Magnitude** for delays, finding the longest delay in the test set.
- **Model B (all-mpnet-base-v2):** Proved superior in **Conceptual Mapping** for distance, correctly identifying "Long Haul" and "Long Distance" flights where MiniLM struggled.
- **Double Failures:** Both models failed significantly on **Routing** (A to B) and **Class Constraints** (Business vs Economy), proving that vector search cannot replace Structured Graph Queries (Cypher).

2. Detailed Round-by-Round Analysis

Round 1: Numerical Magnitude (Delays)

- **Query:** `"Which passengers experienced long arrival delays?"`
- **Model A (MiniLM):** Retrieved **135 minutes** delay. (Winner).
- **Model B (MPNet):** Retrieved **42 minutes** delay.
- **Analysis:** Surprisingly, the smaller model retrieved the record that best fit the description of "long delay." MPNet's result was technically delayed, but MiniLM found the extreme case.

Round 2: Sentiment Severity (Food Complaints)

- **Query:** `"Show passengers who complained about the food quality."`
- **Model A (MiniLM):** Rated **3/10**.
- **Model B (MPNet):** Rated **1/10**. (Winner).
- **Analysis:** Both models succeeded, but MPNet correctly identified the *worst* possible experience (1/10), showing a slightly better grasp of negative sentiment intensity.

Round 3: Conceptual Distance (Long Haul)

- **Query:** `"Which passengers booked economy class but flew long distances?"`
- **Model A (MiniLM):** Retrieved a flight covering **1,024 miles**.

- **Model B (MPNet):** Retrieved a flight covering **4,834 miles**. (Clear Winner).
- **Analysis:** MPNet understands that "Long distance" implies thousands of miles. MiniLM matched the keywords but retrieved a relatively short/medium flight.

Round 4: Complex Constraints (Business Class + Long Haul)

- **Query:** `"Find business-class travelers on any long-haul flights."`
- **Result: DOUBLE FAILURE on Class.**
    - Both models returned **Economy Class** records.
    - **Distance Check:** MiniLM found 3,607 miles (Good). MPNet found 1,201 miles (Poor).
- **Analysis:** This proves that embedding models struggle with specific categorical constraints like "Business Class" when the dataset is dominated by Economy. This query **requires Cypher** `MATCH (j:Journey {passenger_class: 'Business'})`.

Round 5: Exact Route Matching

- **Query:** `"Which passengers flew from IAX TO MIX?"`
- **Result: DOUBLE FAILURE.**
    - **Model A:** IAX -> IAX (Circular/Wrong).
    - **Model B:** MIX -> IAX (Reverse Route).
- **Analysis:** Vector models treat "IAX" and "MIX" as bag-of-words. They don't understand the *directionality* (From -> To). MPNet found the reverse route because semantically, "MIX to IAX" is very similar to "IAX to MIX". **This is the strongest argument for using your Hybrid Agent.**

Round 6: Specific Vocabulary (Loyalty & Fleet)

- **Queries:** `"List passengers with Gold loyalty status"`, `"Passengers traveling on aircraft from the Boeing fleet"`
- **Verdict: Tie.**
- **Analysis:** Both models excelled at matching specific proper nouns (Gold, Boeing, B737).