

Nanyang Technological University
Nanyang Business School

BC2402 – Designing and Developing Databases

Semester 1, 2024/25

Group Project

Airline Chatbots – Is Technology the Silver Bullet?

1. INTRODUCTION

Background

The following content is extracted from a news article published on the BBC.

Maria Yagoda (23 February 2024), Airline held liable for its chatbot giving passenger bad advice - what this means for travellers. Retrieved from <https://www.bbc.com/travel/article/20240222-air-canada-chatbot-misinformation-what-travellers-should-know>

Artificial intelligence is having a growing impact on the way we travel, and a remarkable new case shows what AI-powered chatbots can get wrong – and who should pay. In 2022, Air Canada's chatbot promised a discount that wasn't available to passenger Jake Moffatt, who was assured that he could book a full-fare flight for his grandmother's funeral and then apply for a bereavement fare after the fact.

According to a civil-resolutions tribunal decision last Wednesday, when Moffatt applied for the discount, the airline said the chatbot had been wrong – the request needed to be submitted before the flight – and it wouldn't offer the discount. Instead, the airline said the chatbot was a "separate legal entity that is responsible for its own actions". Air Canada argued that Moffatt should have gone to the link provided by the chatbot, where he would have seen the correct policy.

The British Columbia Civil Resolution Tribunal rejected that argument, ruling that Air Canada had to pay Moffatt \$812.02 (£642.64) in damages and tribunal fees. "It should be obvious to Air Canada that it is responsible for all the information on its website," read tribunal member Christopher Rivers' written response. "It makes no difference whether the information comes

from a static page or a chatbot." The BBC reached out to Air Canada for additional comment and will update this article if and when we receive a response.

Gabor Lukacs, president of the Air Passenger Rights consumer advocacy group based in Nova Scotia, told BBC Travel that the case is being considered a landmark one that potentially sets a precedent for airline and travel companies that are increasingly relying on AI and chatbots for customer interactions: Yes, companies are liable for what their tech says and does.

"It establishes a common sense principle: If you are handing over part of your business to AI, you are responsible for what it does," Lukacs said. "What this decision confirms is that airlines cannot hide behind chatbots."

2. Dataset

In this project, we utilized two sources of data:

1. Flight Delay and Causes (flight_delay)
retrieved from <https://www.kaggle.com/datasets/undersc0re/flight-delay-and-causes>
2. Flight Bookings data (customer_booking)
retrieved from
<https://www.kaggle.com/datasets/minnikeswarao/british-airways-customer-booking>
3. Airline Reviews Dataset (airlines_reviews)
retrieved from <https://www.kaggle.com/datasets/sujalsuthar/airlines-reviews>
4. Customer Service Tagged Training Dataset for LLM-based Virtual Assistants (customer_support)
retrieved from <https://github.com/bitext/customer-support-lm-chatbot-training-dataset>
5. Singapore Airlines Ltd (SIAL) (sia_stock)
retrieved from <https://www.investing.com/equities/singapore-airlines-historical-data>

To reduce your workload (and demand for computational resources), you are provided with MySQL and MongoDB database instances. Do note that the databases do not contain the full dataset (from the sources) but a trimmed version (e.g., removal of special characters and some irrelevant columns).

2. Project Deliverables

The due date for the group project is 22 November 2024, 23:59 (23:59 hrs NTULearn server time)

There are two key deliverables (and one set of optional deliverables), namely

- A. 1 x project report
- B. 2 x script files (SQL and noSQL; you may include in-line command to explain the logic/ assumptions)
- C. 1 x presentation
- D. [optional] database implementations (e.g., relational database and nonrelational database). Please provide a short note to guide the deployment procedure.

PROJECT REPORT

The report should contain the following:

1. A cover page that includes a title, and names as well as matric numbers of each team member.
2. (non-technical) Discussion on the data-driven insights specific to Q6 to Q10. You may include additional aids to illustrate your discussion (e.g., charts and graphs generated based on the data output of your queries).
3. The completed Task Allocation Sheet.

PRESENTATION

Your team is expected to deliver a video-recorded presentation (which must be made available via YouTube), in which the team is expected to:

1. Present the nonrelational database design (i.e., a brief discussion on the differences in design between relational and nonrelational implementation)
2. Discussion on the data-driven insights specific to Q6 to Q10.

The entire presentation **MUST** be within 20 minutes (video duration beyond 20 minutes will be ignored). Each member is expected to contribute equally to the presentation.

DATABASE IMPLEMENTATIONS

To allow you to focus on query development, you are provided with the mySQL and noSQL databases. Depending on your team's approach to Q10, you may need to implement additional tables/collections.

You may rework the structure of the MongoDB implementation (e.g., changing the JSON structures, joining tables) as you deem necessary. It is not necessary to maintain structural consistency between the mySQL database and MongoDB database.

For both mySQL and MongoDB databases, excessive temporary/redundant tables (collections) will be penalized. Please consult with your instructor on implanting additional tables/collections.

3. SUBMISSION

A submission folder will be made available on NTULearn. Please zip the files and make a single file submission. One member will complete the submission on behalf of the group.

The following files must be submitted to complete this group project:

- A. Project report (in pdf format)
- B. YouTube URL (in a text file, please ensure the URL is viewable)
- C. Script files
- D. [optional] Database implementations

In the unlikely event that your submission file is larger than the permissible size on NTULearn, please contact your instructor to arrange for an alternative submission arrangement.

The submission must be made by 22 November 2024, 23:59. Do note that video processing and YouTube uploading can be computationally intensive and bandwidth-demanding. Please ensure ample time for processing and uploading the presentation video.

Appendix – 10 queries

1. How many categories are in [customer_support]?
TIP: You need to decide whether to clean up the data.
2. [customer_support] For each category, display the number of records that contained colloquial variation and offensive language.
TIP: Refer to language generation tags.
3. [flight_delay] For each airline, display the instances of cancellations and delays.
Hint: UNION, \$merge
4. [flight_delay] For each month, which route has the most instances of delays?
TIP: What are the first and last dates in the data?
5. [sia_stock] For the year 2023, display the quarter-on-quarter changes in high and low prices and the quarterly average price.
Note: For details on Quarter-on-Quarter, see
<https://www.investopedia.com/terms/q/qoq.asp>
6. [customer_booking] For each sales_channel and each route, display the following ratios
 - average length_of_stay / average flight_hour
 - average wants_extra_baggage / average flight_hour
 - average wants_preferred_seat / average flight_hour
 - average wants_in_flight_meals / average flight_hour

Our underlying objective: Are there any correlations between flight hours, length of stay, and various preferences (i.e., extra baggage, preferred seats, in-flight meals)?

7. [airlines_reviews] Airline seasonality.
For each Airline and Class, display the averages of SeatComfort, FoodnBeverages, InflightEntertainment, ValueForMoney, and OverallRating for the seasonal and non-seasonal periods, respectively.

Note: June to September is seasonal, while the remaining period is non-seasonal.

8. *Open-ended question; [airlines_reviews]

What are the common complaints?

For each Airline and TypeofTraveller, list the top 5 common issues.

9. *Open-ended question; [airlines_reviews] and additional data*

Are there any systematic differences in customer preferences/complaints pre- and post- COVID specific to Singapore Airlines?

Singapore Airlines hands out 8 months' bonus following record annual profit

<https://www.channelnewsasia.com/business/singapore-airlines-scoot-employees-get-nearly-8-months-bonus-4340801>

In addition to customer satisfaction, what do you think contributed to the strong performance of Singapore Airlines in recent periods?

10. *Open-ended question; [airlines_reviews], [customer_suppport], and additional data*

CAN Explains: Is Singapore Airlines obliged to compensate SQ321 passengers?

<https://www.channelnewsasia.com/singapore/sq321-compensation-singapore-airlines-turbulence-cna-explains-4404701>

How can a customer service chatbot be designed to help Singapore Airlines in such exceptional circumstances?

Possible thinking process: Consider airlines_reviews to identify the relevant issues (e.g., safety and compensations). Consider customer_support in the general chatbot responses to various lexical variations. Propose linguistic design considerations for the chatbot (e.g., apologetic tones, detailed explanations, simple sentences).

Notes: When additional datasets are considered, your team must provide formal references/sources to retrieve the original datasets.

Evaluation will be performed with attention to the coherence of your team's narrative. A coherent data narrative can be achieved using a focused dataset. A rich, diversified dataset can muddle the narrative if the data is not meaningfully integrated.