1 Overview

1.1 Project title

Travel Itineraries Powered by Recommender and Retrieval Augmented Generation

1.2 PI name and department

Alexander Mentzelopoulos, Computer Science Department ID: 219430405

1.3 Project description

We aim to create a fully automated travel itinerary generator using a recommender system coupled with a Large Language Model. Users will fill out a brief survey(yes or no questions) on their travel preferences and be given in return a personalized travel itinerary.

2 Compliance

2.1 University policy requirements

University policy or guidance

University of Bath Research Data Policy

University of Bath Code of Good Practice in Research Integrity

University of Bath Electronic Information Systems Security Policy

University of Bath Intellectual Property Policy

University of Bath Code of Ethics

2.2 Legal requirements

UK Legislation or framework

UK GDPR

2.3 Contractual requirements

3 Gathering data

3.1 Description of the data

3.1.1 Types of data

Users will be presented a series of hotels, tourist attractions, and restaurants and will be requested to answer yes/no on whether they like them or not. We will also collect their email addresses with the sole purpose of sending them their itineraries. Further, the users will be prompted with a second survey that will ask in general terms their opinion on their itinerary. The email addresses will then be deleted. Public travel data also includes data on locations, opening hours, review text(no author) and other non-personal travel data.

3.1.2 Format and scale of the data

Data will be recorded automatically into Excel format xls. Public TripAdvisor data on locations is also collected in csv format using a web scraper and public API.

100 MB

3.2 Data collection methods

The main methods used to collect data are using Google surveys and public APIs such as TripAdvisor and Google Maps.

3.3 Development of original software

All scripts are Python displayed in Jupyter Notebooks, with the appropriate machine learning libraries such as Pandas, Numpy, Scikitlearn, Keras, Gensim, Sklearn. The software is only for use in this project.

4 Working with data

4.1 Short- and medium-term data storage arrangements

(a) All research data will be stored on the University managed storage (X: or H:Drive*):* ***No***

(b) ***If No****:* Data is stored on a secure and password protected encrypted private hard-drive.

4.2 Control of access to data and sharing with collaborators

No access is given to anyone aside from myself(the sole researcher).

4.3 Documentation that will accompany the data

All documentation is accompanied by a readme file that includes instructions on how to run the files and what data is used.

5 Archiving data

5.1 Selection of data to be retained and deleted at the end of the project

All collected data including email addresses will be deleted. The travel itineraries and feedback generated may be retained, but contain no personal information. This information will be deleted directly from the password protected and encrypted personal hard drive. Access will be given to the supervisor and university staff members for assessment and administrative operations finalised to the award of the degree.

5.2 Data preservation strategy and retention period

User survey data will not be kept nor archived as this is not submitted for publication.Output travel itineraries may be preserved as they contain no personal data.

5.3 Maintenance of original software

The code will be maintained on a private GitHub repository, that does not have public access. Access is given only via a private sharing link(via SSH) submitted with the pdf thesis. This access will expire after July 30, 2024. There is no binary executable/installer file, and no one will have rights to edit the code.

6 Sharing data

6.1 Justification for any restrictions on data sharing

Travel data and trip itineraries and some select feedback which contain no personal data will be shared via the dissertation for grading purposes. No emails or personal identifiable information(email addresses) will be shared as they are deleted after completion of the feedback survey.

6.2 Arrangements for data sharing

Travel data will be shared on request from the author. No survey data will be retained or any archive.

7 Implementation

7.1 Review of the Data Management Plan

This data management plan has been reviewed every month from the beginning of the project June 2023.