# ISYS30221 Artificial Intelligence

## Coursework Documentation Template

## 1- About this submission

|  |  |
| --- | --- |
| Student Name | Giles Mwa |
| Student ID | N1357876` |
| Chatbot Topic | Premier League and Football |
| Tasks implemented in this submission | Task a  Task b  Task c    Task a extra – short title: \_\_Football.org API\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Task b extra – short title: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Task c extra – short title: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Files inventory (excluding this file) | QAPairs.csv, top3League\_classifier.h5, mybot-basic.py, mybot-basic.xml, APIFootball.py |
| Demo video URL | [Recording-20250407\_142311.webm](https://myntuac-my.sharepoint.com/:v:/r/personal/n1357876_my_ntu_ac_uk/Documents/Recording-20250407_142311.webm?csf=1&web=1&e=sir1PD&nav=eyJyZWZlcnJhbEluZm8iOnsicmVmZXJyYWxBcHAiOiJTdHJlYW1XZWJBcHAiLCJyZWZlcnJhbFZpZXciOiJTaGFyZURpYWxvZy1MaW5rIiwicmVmZXJyYWxBcHBQbGF0Zm9ybSI6IldlYiIsInJlZmVycmFsTW9kZSI6InZpZXcifX0%3D) |
| Checklist | I will submit this file separately (without compression) into DropBox  All other files are zipped and will be submitted into DropBox  The demo video is recorded as instructed, and the sharing link is inserted above  I have made sure that the demo video is shared according to the instructions, so that I allowed everybody in the university to view it.  All the sections below are populated accordingly. |

## 2- Design notes (shrink/grow as needed, add images where applicable)

|  |  |
| --- | --- |
| General explanations of the system and its goals | This system can respond to questions about the premier league and football in general. This completed by matching user input to AIML pattern giving a predetermined response. These response help to determine which response agent to use.  In the case that user inputs a pattern that does not exactly match the AIML file, td-idf vectorisation is use to calculate the cosine similarity of the input question against CSV file with Question and Answer pairs. If the similarity if 70% or over, the agent will respond using the CSV agent. These response handle facts, old data and general knowledge questions on the topic.  These allow users to ask and access static data on the topic of Football and the premier league.  An external API has been integrated to allow users to send input that match commands to fetch present/future data on teams within the premier league.  Future games, past scores, team sheets, etc. can be outputted to the user on request (“What was Manchester City last score”)  This allows user to access non-static, updated data on the topic of the chatbot.  I have integrated a logic knowledge base (kb) file. This include key relationship detailed by logic statements linking players ,clubs ,managers and rivals. Upon checking a fact or inserting a new one, the kb will check itself for contradictions. This allows users to fact check and insert new facts about football without jeopardising the truth of the kb.  Other functions include, classifying a team logo and figuring out what league the team is likely to play in. I trained a CNN image classification model on the top 5 leagues in Europe. This allowed me to send in an image of a club logo and the classifier should accurately predict what league the logo belongs to. |
| The system requirements, i.e., the list of what the system should do/have from a user’s perspective | The system is required to answer basic and intuitive questions on the premier league. The chatbot should have a predetermined response or fetch the answer from one of it external components (KB, AIML, CSV, API)  Errors should be handled gracefully to ensure seamless usage of the chatbot.  Users should receive a helpful or insightful response, instead of the code breaking. |
| The employed AI techniques, and the explanation of program codes and the supplied files. | The project employs AIML which are standard pattern and template responses. The agent will match user input exactly to pattern and give predetermined or API call answers.  Logic-Kb contains the FOL statements regarding relations between players.  Using nltk resolution prover, the chatbot can check for contradiction within its knowledge base and add and verify facts based on its output.  QAPairs this a csv file containing Q/A. I used sklearn tfi-df vectoriser and cosine similarity libraries to allow close matching input to still return the rightful answer.  APIFootball.py is a class I created to access endpoint within an external api. This file allows all of the football data gathering commands and uses Json deseriailsation and Http request to handle data and present in a readable user format.  top3League\_classifier.h5 this is the training file for the CNN classifier model I trained on 3 of the 5 top leagues. This makes use of keras preprocessing and tensorflow to train and test a image classifier model, for the purpose of predicting a league based on a logo. |

## 3- Conversation log (insert text, screenshots and/or images as required)

|  |
| --- |
| Task A:  Random response selected from list of templates in AIML files    In this case the input did not fully match an AIML pattern, therefore the answer was given by the Tf-idf cosine similarity matched response agent. If the input from the user is a direct match, the answer will be retrieved from the QA.csv, it is the same with close matches ;I have set the code to only match those with similarity >0.7. If nothing is found, the chatbot attempt to find information on the input through Wikipedia.    If the match cosine similarity is too low, the response will be a standard placeholder or, if data is found on Wikipedia that will be summarised    Task A (extra):  External football API to get live stats: standings, team sheets, managers, fixtures(present/past), top scorers    This is one of the API endpoints my chatbot can fetch data from, It is capable of taking the Team Name and processing It into a HTTP request with relative parameters to gather data on the specified team or the league itself.  Task b:    If a contradiction is found in the kb, facts will be blocked being added to the kb.    If there is no contradiction, the fact is added to the KB. Also if the KB doesn’t have the knowledge to check whether something is true, it will return that if is most likely false. However this will not always be true. Logic has not been implemented to give a guaranteed answer.  Task C:    The image classifier was only trained on 3 leagues due to lack of dataset available online.  The accuracy Is off as the premier league dataset have 10x as many images as the other. This created an unbalanced dataset for training.  The user is able to input a path string using the command “What league is this team from \*” |

(no word count is necessary)