COMP3074 - Coursework 1
AN INTERACTIVE NLP-BASED AI SYSTEM

This material is exclusively for the use of University of Nottingham students enrolled in COMP3074 (Human-AI interaction). Sharing this document outside of the university constitutes academic misconduct.

Coursework 1 – an Interactive NLP-based AI system	
Version	2022
Release date	21/10/2022 @ 15h00 (Nottingham time)
Hand-in date	25/11/2022 @ 15h00 (Nottingham time)
Late penalty	5% per working day

In this module we have seen that there are many ways that AI systems embed in our society through both user-facing and background applications. In this coursework you will be required to design and implement your own user-facing **interactive NLP-based system**. Using Python and a restricted selection of libraries[1], you will produce the following three (3) deliverables:

- **Deliverable 1**: an interactive NLP-based AI system, worth **20%** of the grade.
- **Deliverable 2**: a report of up to/around 2,000 words or **maximum** of 10 pages (excluding citations) evaluating the system, worth **70%** of the grade. There is no strict minimum number of words, but we expect that it should take you around 2,000 words.
- **Deliverable 3**: a short video of no more than 5 minutes showing your system working, worth **10%** of the grade. Not showing your code. Not showing your architecture design. Only showing the system as you use it, with you explaining what is happening in your own words.

This document will describe each of those three deliverables, explain what they should contain, and how they will be evaluated.

General notes:

1. Submissions need to be in English.
2. While I am not a stickler for formality, watch out for the language you use in your code. The use of offensive terminology in your code, report, or video will be penalised accordingly (lack of professionalism) and reported.
3. The system you submit, the report you write, and the video you record need to be about the same system.
4. While the weighting of the system might seem small (20%), the scope of your system will have an impact on your report. You won't have enough to say about if your system does not do enough, so you still need to put enough work building a system that you can write about. "How do I know if I have done enough?", you are probably thinking right now. Think backward: if you were to write a report about it, would it take you around 2,000 words, or would it fit on a couple of post-it notes?

---

[1] Python standard library, numpy, scipy, scikit-learn (and dependencies), nltk **except for** nltk.chat

Deliverable 1. Interactive NLP-based AI system

You should build a dialogue-based system which allows a user to interact with an AI, commonly known as a **chatbot**. The work will be based on what you have learned, using Python and a restricted set of allowed libraries while will be provided in due course. A lot of Python libraries can automate much of the process, so it is of the utmost importance that you do not use them. Any part of the assessment that is performed by a library will not be counted towards your grade.

The two key parts of this coursework are:

- **Natural language processing**: the system needs to be language-based so that you can apply what you have learned about processing text.
- **Interactive**: there needs to be a back-and-forth communication between the human and the machine.

The system that you develop needs to be versatile and able to do multiple things, much like a real chatbot, and part of your grade (see section on Evaluation) will depend on how versatile your system can be and how good they are at their functions.

1. Potential features

This section contains a non-exhaustive list of common tasks in interactive NLP-based systems (chatbots). They are not all equal in complexity or importance, e.g., intent matching is necessary for any multi-usage chatbot while small talk is not. **You do not have to implement all these features** but pick and choose the ones that is within the objective of your chatbot.

• Intent matching

Intent matching is a necessary part of any system which aims to do more than one task. It consists in taking in a user input and predicting the intent of the user, i.e., which function of the system the user is intending to use. Such functions could be storing/showing the user's name, outputting the current date and time, starting a transaction dialogue, or anything else described in this document. Intent detection can be done at different degrees of complexity, from simple text matching to similarity-based and machine learning-based approaches.

• Identity management

Identity management refers to being able to acquire, memorise, and use the name of the person currently using the system. It consists in functions such as detecting name statements (e.g., "my name is X", "call me X", etc.) and storing the name to be used either in basic statements addressing the user, or in explicit name output functions (e.g., "what is my name?").

• Small talk

A small talk function refers to the ability to have short discussions about shallow topics, such as "hello", "how are you", "how's the weather", etc. There is no expectation of long conversations due to them being one-off topics, but they can either be single turn (question, answer) or multi-turn (question, answer, and one or more follow-ups).

- ## Transactions

A transactional dialogue system is a tool that allows for an artificial system to dialogue with a human operator **in order to achieve a specific set of goals**, which are **transactional** in nature, such as booking a train/plane ticket, a table in a restaurant, ordering an item from a store, etc.

Here is an example of a turn-by-turn dialogue between an operator and a travel booking system, to illustrate the range of tasks that such system would need to accomplish to properly function.

Turn	Author	Text
1	Bot	Hello, welcome to the Skynet Travel Agency, how may I help?
2	User	I would like to book a flight from London to Paris
3	Bot	Would that be a return trip or a single flight?
4	User	A return trip please
5	Bot	When were you thinking of going?
6	User	From the 10th to the 19th?
7	Bot	I'm sorry, I am going to need more details than that – please use the precise dates of your trip and whether you are flexible in your dates
8	User	10/11/2021 to 19/11/2021, and I am flexible
9	Bot	Thank you, I will search my database for available tickets. Please wait.
10	Bot	I have one flight available on the 10th from London to Paris, but no flight from Paris to London on the 19th, would you like to change the return date or book a single flight?
11	User	Change the date to the 20th

- ## Information retrieval & question answering

A conversational information retrieval system is an information retrieval tool that interprets a *query* from a dialogue with a user and returns a set of one or more results which are deemed to be relevant, such as **direct answers**, **webpages**, **books**, **recipes**, **songs**, **movies**, etc. It requires having some form of database of documents and an indexing system. A question answering system can be seen as a special case of information retrieval where the documents are question-answer pairs, and only the answer is retrieved.

2. Evaluation

Your system will be evaluated on depth and breadth. A highly specialised system (e.g., purely transactional) would therefore be expected to be very deep in function, while a non-specialised system (e.g., a generalist chatbot) would be expected to have many functions (and a good intent matching system!).

Your system will be evaluated in a distinction/pass/fail manner with penalties. Generally if your system works, you will get a pass (10 out of 20 marks), subject to the following penalties/bonuses:

- There are non-system-breaking bugs: -5 marks
- There is some code that was not referenced properly: -10 marks.
- Rich and coherent set of well-implemented features: +5 marks
- Good conversational experience: +5 marks

If your system does not work on a standard installation such as the A32 computers or the virtual desktop, or if it does not actually use natural language interactions, you will get 0 marks.

Deliverable 2. **Report**

You should write a report explaining and critically analysing your system, using the following structure:

1. **Introduction** – where you **introduce your system** and **explain why you built it**.
2. **Background** – **talking about papers or projects** doing the same thing and how your project differs. This could be papers or book chapters about the techniques you are using, or papers/book chapters/blog posts about systems similar to yours or which served as inspiration.
3. **Proposed system** – technical description of your system. At least three points should be discussed: **functionality** (what it does), **motivation**: why you built it that way, **implementation**: how you built it. Make good use of images, flowcharts, and general illustrations[2]. Reference relevant bits of code in the report (if short), or in the project files (if long).
4. **Evaluation** – how you tested your project, e.g., **user testing**, **benchmarking,** etc.
5. **Discussion** – where you discuss the **results of your evaluation**, think about the **potential impact** of your system if it was deployed in the real world and reflect about the **overall fairness** and potential **sources of bias** of your AI.
6. **Conclusion** – where you **recap** your work and what you have achieved.

I strongly encourage you to follow this structure if you are not an experienced writer of scientific reports (and you probably are not). Your report will be evaluated across the following criteria

1. **Background and description of the system (30%)**
   a. Adequate coverage of the background and description
   b. Adequate level of detail in the background and description
2. **Evaluation (30%)**
   a. Completeness of the evaluation
   b. Analysis of the results
3. **Discussion (30%)**
   a. Reflection on social aspects (breadth, depth, link to literature)
   b. Reflection on ethical aspects (breadth, depth, link to literature)
   c. Reflection on results (breadth, depth, link to literature)
4. **Writing quality (10%)**
   a. Introduction and conclusion open and close the report
   b. Sensible structure and presentation
   c. Writing clarity
   d. Grammar and typos
   e. Sensible use of diagrams and screenshots

---

[2] https://www.diagrams.net is a great tool to make all sorts of charts. Hand drawn diagrams are not acceptable in 2022.

Each of those criteria will be evaluated on a scale of 0-100, with the following indicative bands:

- 0-29 -> Failing attempt.
- 30-39 -> Marginal fail.
- 40-49 -> Low pass.
- 50-59 -> Adequate to good.
- 60-69 -> Good to very good.
- 70-79 -> Very good to excellent.
- 80-89 -> Outstanding, almost worthy of being sent for publication. I might strongly suggest that you apply for graduate studies.
- 90-100 -> Exceptional, worthy of being sent for publication. I will frame them and put them on my fridge to remind me of why I became an academic. I will also strongly suggest that you pursue graduate studies.

If your report does not open or is not in a PDF format, you will get 0 marks. It is your responsibility to check that your uploaded document is readable.

If your report describes a system which is not what you actually implemented, you will also get 0 marks. This is not a creative writing exercise.

Deliverable 3. Demonstration video

You should produce a short demonstration video, of a few minutes at the longest, where you demonstrate your system working. The video should have some form of audio narration of you explaining what you are showing. In the video, you should demonstrate how your system fulfils each feature you decided to implement, with a couple of examples of your choice for each.

Your video will be evaluated in a pass/fail/distinction manner with penalties. If your video adequately shows your project working along with a narrated explanation of how it works, you will get 10 marks, subject to the following penalties:

- You did not cover all features: -5 marks
- Your video was hard to follow, or hard to understand, or there is no sound: -5 marks.

If your video does not run, you will get 0 marks. Please use a sensible video format.

While I am not a stickler for formality, swears and general offensive language in your video will yield a penalty of -10 marks.