CS6207 Advanced Natural Language Processing
AY 2023/24 Semester 2

The breakdown of the final project marks will be according to the following criteria:

Content (45%): Is the work valid and technically sound? Is the work significant and of impact? Is relevant work in the literature cited?

Presentation (30%): Is the work presented in a clear, logical, and coherent manner? (including both oral presentation and the final project report)

Novelty (25%): Is the work carried out novel? Does it advance the state of the art?

Each final project report is in the form of an ACL conference-style short paper. The main content of the paper is limited to 4 pages, with the references taking up a maximum of 1 additional page. You must use the ACL conference paper style files available at the following URL:

https://github.com/acl-org/acl-style-files

The page limit must be strictly adhered to. Marks will be deducted if the main content of your paper exceeds the 4-page limit, or the references exceed 1 page, or the font size is not Times Roman 11-point font.

Each project team must also submit the source code implemented in the project and all external (supporting) code and data. Each team will submit its presentation slides and give a 15-minute oral presentation in a video recording at the end of the course.

Possible final project topics:

1. Integrating the use of an external text database into a large language model (LLM): Current LLMs are pre-trained on a very large but static collection of texts. Devise an algorithm that allows an LLM to answer a question by integrating knowledge present in an external text database in addition to the texts on which the LLM was pre-trained.

2. Hallucination detection of a text generated by an LLM: Devise an algorithm that can detect if a text generated by an LLM is factually accurate or contains hallucination.

3. Detection of text generated by an LLM: Devise an algorithm that can accurately detect whether a document is generated by an LLM or written by a human.

4. Coreference resolution: Implement a state-of-the-art coreference resolution program and perform evaluation on the CoNLL-2012 English benchmark data set.

5. Grammatical error correction: Implement a state-of-the-art grammatical error correction program and perform evaluation on the CoNLL 2014 and BEA 2019 shared tasks.

Other project topics are also possible and in fact encouraged (subject to approval by the lecturer).