**Topic Submission Form**

This form should be submitted by the mentioned deadline.

Name:  Paresh Pradhan

Student Number:   PN1038004

Course:  36269 MS - Artificial Intelligence & Machine Learning\_\_\_\_

**Fill your topic/s below**

Project Title/Area 1:   AI-Assisted Collaborative Story Writing with Human-in-the-Loop \_

Dataset:

* WritingPrompts - <https://www.kaggle.com/datasets/ratthachat/writing-prompts>
* ROCStories - <https://cs.rochester.edu/nlp/rocstories/>
* CoAuthor - <https://coauthor.stanford.edu/>
* HANNA - <https://github.com/dig-team/hanna-benchmark-asg>
* TVRecap - <https://github.com/mingdachen/TVRecap>
* MoviePlotEvents - <https://paperswithcode.com/dataset/movieplotevents>
* Scifi TV Shows - <https://huggingface.co/datasets/lara-martin/Scifi_TV_Shows>
* WikiPlots - <https://github.com/markriedl/WikiPlots>
* OpenMEVA - <https://github.com/thu-coai/OpenMEVA#i-datasets>

Description: In recent years, neural language models have grown greatly in effectiveness. They are already being applied to real-world tasks. Large language models have shown unprecedented capabilities in natural language processing tasks. However, due to their highly context-dependent nature, large language models are difficult to grasp and make use of. State-of-the-art large language models, such as GPT-3, have achieved new levels of performance on benchmarks for language understanding and generation. These models have shown that they can perform very well on many downstream tasks without any explicit training. Story writing is one such task which continues to be a challenging problem for machines and even for humans. Existing works include neural narrative generation systems that interact with humans in different ways to generate stories. While large language models, with their improved text generation capabilities, have enabled writers to co-create stories with an AI, guiding the narrative remains a challenge. This topic explores different methodologies to facilitate iterative human-AI co-writing process in an interactive and collaborative manner.

Project Title/Area 2:   Automated Story Generation with Pre-Set Control Mechanisms \_

Dataset:

* WritingPrompts - <https://www.kaggle.com/datasets/ratthachat/writing-prompts>
* ROCStories - <https://cs.rochester.edu/nlp/rocstories/>
* HANNA - <https://github.com/dig-team/hanna-benchmark-asg>
* TVRecap - <https://github.com/mingdachen/TVRecap>
* MoviePlotEvents - <https://paperswithcode.com/dataset/movieplotevents>
* Scifi TV Shows - <https://huggingface.co/datasets/lara-martin/Scifi_TV_Shows>
* WikiPlots - <https://github.com/markriedl/WikiPlots>
* OpenMEVA - <https://github.com/thu-coai/OpenMEVA#i-datasets>

Description: In recent years, neural language models have grown greatly in effectiveness. They are already being applied to real-world tasks. Large language models have shown unprecedented capabilities in natural language processing tasks. However, due to their highly context-dependent nature, large language models are difficult to grasp and make use of. State-of-the-art large language models, such as GPT-3, have achieved new levels of performance on benchmarks for language understanding and generation. These models have shown that they can perform very well on many downstream tasks without any explicit training. Story generation is one such task which continues to be a challenging problem for machines. Existing works include neural narrative generation systems that automatically generate stories given some input prompts, keywords, or outlines. While large language models, with their improved text generation capabilities, can generate believable stories; guiding the narrative remains a challenge. This topic explores different methodologies to facilitate automatic generation of a story narrative in a controllable manner.

Project Title/Area 3\_1:  Anonymization of Speaker Voice for Privacy Protection\_

Dataset:                                                                                  \_\_ \_

Description: Collection of speech data and creation of datasets is a vital step in training speech recognition systems and other speech-based machine learning models. However, the issue of privacy protection is an increasing concern that must be addressed. Speech data carries a range of personal information, such as the speaker's identity and emotional state. These attributes can be used for malicious purposes. The rapid increase in web services and mobile apps, which collect personal data from users, has also increased the risk that their privacy may be severely compromised. In particular, the increasing variety of spoken language interfaces and voice assistants empowered by the recent breakthroughs in deep learning have prompted important concerns in the European Union in terms of preserving the privacy of speech data. For instance, an attacker can record speech from users and impersonate them to obtain access to systems that require voice identification. By extracting speaker, linguistic (e.g., dialect), and paralinguistic features (e.g., age) from a speech signal, the speaker profiles can also be hacked from users through existing technology. Current studies have addressed the topic of preserving speech privacy. One of them, the VoicePrivacy initiative, aims to promote the development of privacy preservation tools for speech technology. The task for the VoicePrivacy Challenge (VPC) is about speaker anonymization. The goal is to hide the source speaker's identity in untranscribed running speech while preserving the linguistic information. This is the problem that this topic tries to solve.

Project Title/Area 3\_2:  Anonymization of Personally Identifiable Information (PII) in Image by Replacing with Generated Fakes

Dataset:                                                                                  \_\_ \_

Description:                                                                                  \_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Fill in this section if a member of staff has agreed to be your supervisor:**

Member of Staff:                                                                                   \_\_\_\_

If you have found a supervisor then you and the member of staff who agreed to supervise your project should sign below.

\_Paresh Pradhan\_                                                                        \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student Signature                                                                         Supervisor Signature

\_18-Sep-2022\_                                                                            \_\_\_\_\_\_\_\_\_\_\_\_

Date                                                                                               Date