Extracurricular Activity Project: Exploring Recent NLP Techniques in Domain-Specific Applications

Introduction

In this extracurricular project, **you and a partner (team of 2, only students who register for this course as extracurricular)** will dive into the exciting field of Natural Language Processing (NLP). Your mission is to explore a recent NLP topic that interests you, apply it to a domain you care about, build a system, test it, and share your findings in a short research paper. There's no single "right" project—your creativity and curiosity will drive this adventure!

Project Requirements

1. Team Formation

- **Team Size**: Work in pairs (two students per team).
- **Collaboration**: Both team members must contribute equally. Use tools like GitHub for code versioning and shared documents for writing.

2. Topic Selection

- **NLP Technique**: Select a recent advancement in NLP, such as:
  - Retrieval-Augmented Generation (RAG)
  - Small Language Models (SLMs)
  - Prompt Engineering
  - Few-Shot Learning
  - Multi-Modal NLP
  - MCPs
  - Explainable AI (XAI) in NLP
  - Transfer Learning with Domain Adaptation
  - Ethical NLP (e.g., bias and fairness)
  - Or another recent topic (subject to instructor approval).

- **Domain**: Choose a domain to apply the technique, e.g.:
  - Economics (e.g., financial report analysis)
  - Healthcare (e.g., medical literature summarization)
  - Law (e.g., legal document classification)
  - Education (e.g., automated essay scoring)
  - Entertainment (e.g., review sentiment analysis)

- **Task**: Define a specific NLP task, such as text classification, summarization, or question answering.

3. System Implementation

- **Tools and Libraries**: Use open-source tools like:
  - Hugging Face Transformers
  - PyTorch
  - TensorFlow
- **Datasets**: Select or create a dataset relevant to your domain and task (e.g., from Hugging Face Datasets, Kaggle, or domain-specific sources).
- **System Scope**: Build a functional system to demonstrate the NLP technique (production-readiness not required).

4. Experimentation

- **Hypothesis**: Define a research question or hypothesis (e.g., “Does prompt engineering improve summarization in this domain?”).
- **Metrics**: Use appropriate metrics (e.g., accuracy, F1 score, ROUGE, BLEU) based on your task.
- **Baseline**: Compare your system to at least one baseline model or approach.
- **Analysis**: Provide quantitative results (e.g., tables, graphs) and qualitative insights (e.g., sample outputs, error analysis).

5. Research Paper

- **Length**: 4-6 pages (excluding references).
- **Structure**:
  - **Introduction**: Context, problem, and objectives.
  - **Related Work**: Summary of relevant research.
  - **Methodology**: System description, including technique, dataset, and implementation.
  - **Experiments**: Experimental setup, metrics, and results.
  - **Discussion**: Interpretation, limitations, and future work.
  - **Conclusion**: Summary of contributions and significance.
- **Formatting**: Use LaTeX IEEE, or ACL style templates with proper citations.

Deadline

- **Submission Date**: All deliverables are due by **May 27, 2025**, via Canvas.
- **Late Policy**:
  - Late submissions incur a 60% mark deduction unless accompanied by official documentation (e.g., medical excuse).
  - Submissions over one week late receive a zero mark.

Deliverables

1. **Source Code**: Submit a GitHub repository link with:
    - A README file (setup and execution instructions).
    - Well-organized, commented code.
2. **Research Paper**: Submit a 4-6 page PDF.
3. **Optional Presentation**: Submit a slide deck (PDF or PowerPoint) if applicable.

Evaluation Criteria

- **Functionality (30%)**: Is the system technically sound and operational?
- **Experimentation (25%)**: Are experiments well-designed with meaningful analysis?
- **Research Paper (25%)**: Is the paper clear, structured, and well-cited?
- **Creativity and Innovation (10%)**: Does the project offer unique insights?
- **Team Collaboration (10%)**: Did both members contribute effectively?

Project Discussion

- Teams will present their work (5-10 minutes) during a final discussion, covering:
    - Chosen NLP technique and domain.
    - System functionality and results.
    - Challenges faced and solutions.
    - Lessons learned and future work.

Publication Opportunity

Exceptional projects may be selected for potential publication in a relevant NLP conference or journal. To be considered, your project should demonstrate significant innovation, technical depth, and clear experimental results.

If selected, both team members will receive guidance on refining the work for publication, including feedback on the paper and suggestions for additional experiments or analyses. Both members will be involved in the publication process and listed as co-authors if the work is published.

Publication not only enhances your academic and professional profile but also allows you to contribute meaningfully to the NLP community.

Please note that publication is competitive, and not all projects will be selected. However, the process of preparing your work for potential publication is a valuable learning experience in itself.

Selected projects will be notified shortly after the project deadline and will receive support to prepare their work for submission to a suitable venue.

Good Luck!