

# CSCE 771: Computer Processing of Natural Language

## Lecture 15: Project Milestone #1 Presentations

---

PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

8<sup>TH</sup> OCTOBER, 2024

***Carolinian Creed: “I will practice personal and academic integrity.”***

# Organization of Lecture 15

---

- Opening Segment
  - Review of Lecture 14

- Main Lecture



## Main Section

- Project Updated #1; 5 mins per student

- Concluding Segment
  - About Next Lecture – Lecture 25

# Recap of Lecture 14

---

Sep 24 (Tu)	Language Model – PyTorch, BERT, {Resume data, two tasks} – <b>Guest Lecture</b>
Sep 26 (Th)	Language Model – Finetuning, Mamba - <b>Guest Lecture</b>
Oct 1 (Tu)	Language model – comparing arch, finetuning - <b>Guest Lecture</b>
Oct 3 (Th)	Language model – comparison of results, discussion, ongoing trends– <b>Guest Lecture</b>

- We looked at Language Modeling - Large Language Models, and Small Language Models
  - How they are trained
  - How they are instruction tuned
  - How they can be efficiently fine-tuned and inference
- We covered small language models as well.

# Main Segment

---

# Course Project

---

# Discussion: Course Project

**Theme:** Analyze quality of official information available for elections in 2024 [in a state]

- Take information available from
  - Official site: State Election Commissions
  - Respected non-profits: League of Women Voters
- Analyze information
  - State-level: Analyze quality of questions, answers, answers-to-questions
  - Comparatively: above along all states (being done by students)
- Benchmark and report
  - Compare analysis with LLM
  - Prepare report

- Process and analyze using NLP
  - Extract entities
  - Assess quality – metrics
    - Content – *Englishness*
    - Content – *Domain* -- election
  - ... other NLP tasks
  - Analyze and communicate overall

## Major dates for project check

- Sep 10: written – project outline
- Oct 8: in class
- Oct 31: in class // LLM
- Dec 5: in class // Comparative

# Obtaining Election Data

---

Here are a few things to do:

A) **Official data** backed by laws: state election commission

a) Find the state's election commission

b) Find the Q/As they provide. They may be as FAQs or on different web pages.

c) Collect the Q/A programmatically

B) **Secondary data** sources: non-profit

a) Find Q/As from Vote 411 which is supported by the non-profit: LWV.

For reference, for SC,

A) Official - <https://scvotes.gov/voters/voter-faq/>

B) Secondary - <https://www.vote411.org/south-carolina>

For extraction, one or more approaches:

- Manually annotating
- BeautifulSoup,
- Tika
- or other open source libraries.

# Discussion

---

- How will you use a LLMs for election data analysis ?
- When and Why? (conversely, not)



# <Student Name>

## CSCE 771 Fall 2024: Milestone#1

1. State Selected:
2. Election data sites:
  - Official site (e.g., State Election Commission) url
  - Secondary site (e.g., League of Women Voters) url
1. Report how data collected and Q/A statistics
1. Take on NLP methods you will use and why for Q/A analysis
  1. State-level (right)
  2. Comparatively: above along states being done by peers

Initial analysis of questions (Q)

\*

Initial analysis of answers (A)

\*

Initial analysis of an answer (a\_i)  
for a question (q\_i)

\*

# Concluding Segment

---

# Lecture 15: Concluding Comments

---

- Good range of states for election Q/A analysis
- Gear up for detailed analysis with language/ NLP methods discussed
  - And LLM background used

# About Next Lecture – Lecture 16

---

# Lecture 16 – Using LLMs

- Using LLMs for NLP
  - Methods learnt
    - Parsing
    - Semantics
    - ML: supervised (classification)
    - ML: unsupervised (clustering)
  - Doing with LLMs
  - Comparative advantages

13	Oct 1 (Tu)	Language model – comparing arch, finetuning - <b>Guest Lecture</b>
14	Oct 3 (Th)	Language model – comparison of results, discussion, ongoing trends– <b>Guest Lecture</b>
15	Oct 8 (Tu)	<b>PROJ REVIEW</b>
16	Oct 10 (Th)	Using lang models to solve NLP tasks
17	Oct 15 (Tu)	<b>QUIZ 2</b>
	Oct 17 (Th)	
18	Oct 22 (Tu)	Entity extraction, linking
19	Oct 24 (Th)	Events extraction, spatio-temporal analysis
20	Oct 29 (Tu)	Topic Analysis