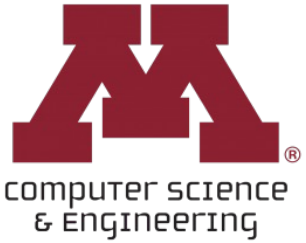
CSCI 5541: Natural Language Processing

Lecture 15: Concluding Remark

Dongyeop Kang (DK), University of Minnesota

dongyeop@umn.edu | twitter.com/dongyeopkang | dykang.github.io

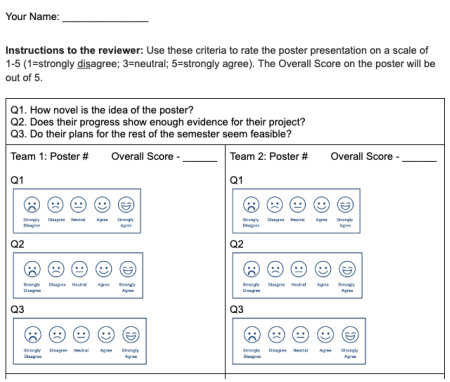


UNIVERSITY OF MINNESOTA
Driven to Discover®

Outline

- ❑ Announcement for poster presentation and final report
- ❑ Research topics in Minnesota NLP group
- ❑ Concluding remarks

Poster presentation (5 points)


Your Name: _____

Instructions to the reviewer: Use these criteria to rate the poster presentation on a scale of 1-5 (1=strongly disagree; 3=neutral; 5=strongly agree). The Overall Score on the poster will be out of 5.

Q1. How novel is the idea of the poster?
Q2. Does their progress show enough evidence for their project?
Q3. Do their plans for the rest of the semester seem feasible?

Team 1: Poster #	Overall Score - _____	Team 2: Poster #	Overall Score - _____
Q1		Q1	
Q2		Q2	
Q3		Q3	

- Upload your poster on Canvas before your presentation
  - https://canvas.umn.edu/courses/413172/assignments/3605228
- Print your "32x40" poster and present it at your assigned poster session
- Printing instructions are provided at this link ; you can request it using the form (details on how to fill out initial fields on next slide). Keep in mind, they guarantee posters submitted 2 business days in advance, but do not work on the weekends.
- Reviewers: instructors + students during poster sessions
- Expected content in presentation:

Your poster clearly describes the following information:

```
Motivation
Literature survey
Problem definition
Proposed ideas
Contribution
Experimental results and comparison with baselines
Main findings
Limitation and discussion
Plan for the final report.
```

Poster presentation (5 points)

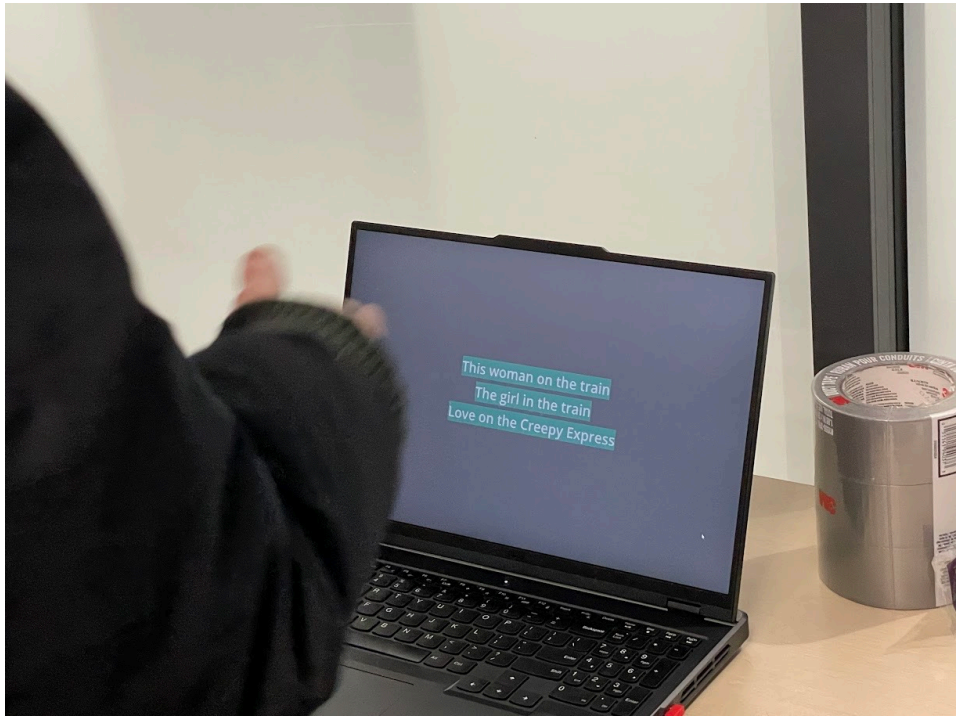
- **Location**: Shepherd 164 (aka Drone Lab)
- **Time**: April 23 (Group B), April 25 (Group A)

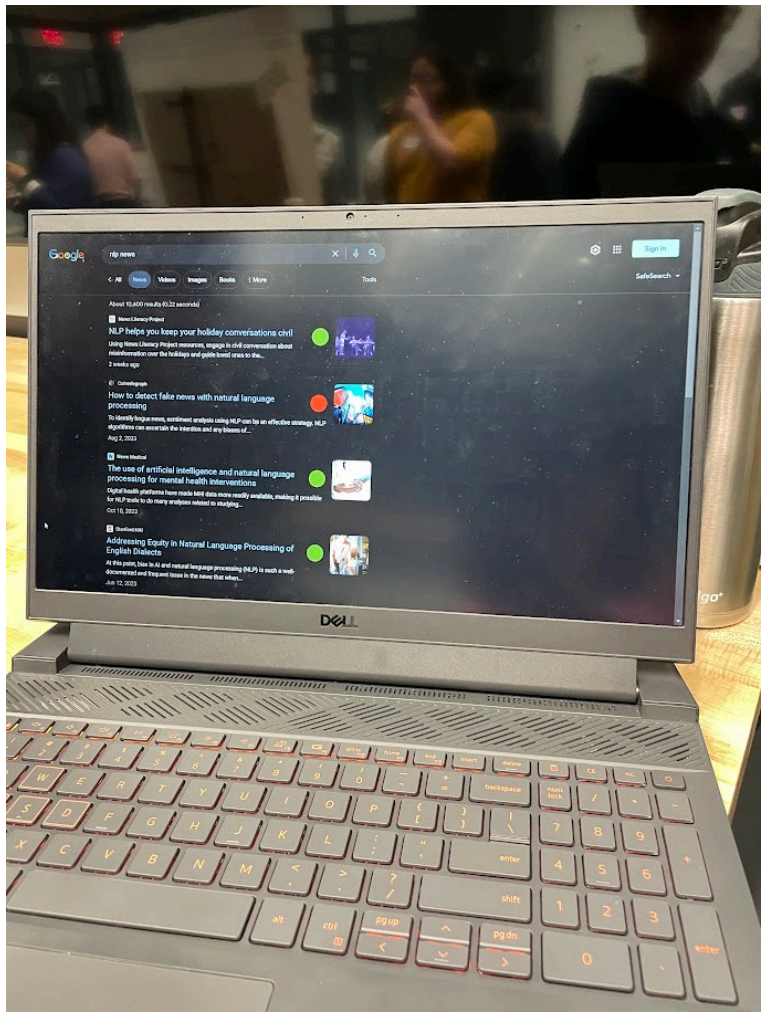
Posters for Group B

- **Caught with N-grams** (Michael Bronstein, Yichen Li, Lavanya Radhakrishnan, Wuhao Zhang) Automated Detection and Refutation of Climate-related Misinformation
- **Cybertron** (Gehna Jain, Ryan Langman, Trae Primm, Swapnil Puranik) Evaluation of Knowledge Graphs in LLMs
- **Fury GPT** (Nikil Krishnakumar, Sujeendra Ramesh, Rammesh Adhav Saravanan) User Friendly Drone Control: Fine-Tuning Language Models with ROS Commands for Real-World Application
- **NLP Ninjas** (Nirshal Chandra Sekar, Byeongchan Jeong, Fidan Mahmudova, Sheshasai Sairam) Human - Robot Interaction using LLM
- **SpotHRI** (Adam Imdieke) A Language Interface for the Spot Robot
- **Transformers** (Ritwick Banerjee, Madhan Mohan, Leyan Sayeh, Masha Volkova) Disease Diagnosis using LLM
- **NLPitch** (Dhondup Dolma, Jaeeun Lee, Yongtian Ou, Jiyoon Pyo) Transidiomation: Optimizing translation of idioms embedded in text
- **PRWZ** (Zaccheri Ciampone, Wyatt Kormick, Raymond Lyon, Preston Zhu) Research Paper Simplification
- **Too Long; Didn't Read** (Ryan Johnsen, Dylan Paulson, Logan Schaaf, Tony Zhang) Cuisine Fusion Recipe Generato

Posters for Group A

- **Lexical decryptors** (Sandeep Bhuiya, Mudit Jantwal, Hemanth Kumar Tirupati, Shreya Yashodhar) Scientific Text Simplification
- **NLPros** (Benjamin Kosieradzki, Mitchell Kosieradzki) Evaluating the Boundary between In-Context Learning and Fine-Tuning
- **pnlp fiction** (Rakshithaa Kanakarajan Selvarathinam, Vishal Kancharla, Sri Krishna Vamsi Koneru, Devansh Mishra) Identifying Bias in LLMs when using LRLs
- **Syntax Errors** (Clayton Carlson, Ryan Diaz, Charlie Rapheal, Sanjali Roy) Leveraging Language Models for Temporal Political Bias Analysis
- **Tattered-animals** (Zheng Robert Jia, Brandon Nee, Andrei Solodin, Jack Swanberg LLM Prompt Recovery
- **WordWizards** (Morgan Bozeman, Tyler Cook, Connor Holm, Derek Wong) Performance of LLMs in Various Styles
- **EdgeCaseWizards** (Apekshik Panigrahi, Anna Terzian) Terraform - Automating Infrastructure as a Service
- **Team bRockoLee** (Jithendra Jagannatha Kagathi, Arjun Thonoor, Anirudh Vasudevan, Ya-Hui Yang) Transpilation
- **Team SOTA** (Evan Way, Jerry Yin, Zaifu Zhan, Zhongxing Zhang) PaperHelper: Knowledge-Based LLM QA Paper Reading Assistant


This woman on the train
The girl in the train
Love on the Creepy Express


nlp news
NLP helps you keep your holiday conversations civil
How to detect fake news with natural language processing
The use of artificial intelligence and natural language processing for mental health interventions
Addressing Equity in Natural Language Processing of English Dialects
DELL

Final report (10 points, May 3)

- Upload your PDF **report** and **code** on Canvas
  - Maximum 8 pages with unlimited reference and appendix
  - Zipped code or link to your github
- Rubric for the final evaluation
  - 100 points will be normalized to 10 points in grading
  - Novelty / Significance / Clarity
  - Relative evaluation

```
Rubrik (100 points) for Final Report

Below are three general evaluation criteria:
(10 points) Novelty: Compared to the state-of-the-art
↪ methods/systems/datasets, how novel is your approach? Is your work
↪ publishable?
(10 points) Significance: How storng is your result? Is your finding still
↪ holding if different setups or prompting tricks?
(10 points) Clarity: How clear and easy-to-follow is your report? Do you have
↪ well organized presetnation of your results and problem definition?
Introduction / Background / Motivation:

Introduction / Background / Motivation:
(5 points) What problem do you try to solve? Describe your objectives cleraly
↪ without using any technical jargon.
(5 points) How is it done today by other researchers? What are the limitations
↪ and challenges of current practice?
(5 points) Who might be interested in your work? What kinds of impact can you
↪ make?

Approach:
(5 points) What did you do exactly? How did you solve the problem? Why did you
↪ think it would be successful? What is your hypothesis?
(5 points) What challenges did you anticipate and/or encounter during the
↪ development of your approach? Did the very first thing you tried work?
(5 points) What is scientific novel of your approach to address the
↪ challenges?
Experiments / Results / Error Analysis:

Experiments / Results / Error Analysis:
(5 points) How did you measure success? What research questions do you want to
↪ validate? What evaluation metrics and experiments were used? What were the
↪ results, both quantitative and qualitative? Did you succeed? Did you fail?
(5 points) No matter you succeed or fail, why? Which data points are
↪ incorrectly predicted by yours but previous models can't, or vice versa.
(5 points) Are there still some failure cases? Why can't your approach address
↪ them? Any potential solutions?
(5 points) Are the ideas/probelm/results presented with appropriate
↪ illustration?
Additional points:

Discussion points:
(5 points) Replicability: How easily are your results able to be reproduced by
↪ others?
(5 points) Datasets: Did your dataset or annotation affect other people's
↪ choice of research or development projects to undertake?
(5 points) Ethics: Does your work have potential harm or risk to our society?
↪ What kinds? If so, how can you address them?
(5 points) Discussion: What limitations does your model have? How can you
↪ extend your work for future research?
```

Outline

- Announcement for poster presentation and final report
- Research topics in Minnesota NLP group
- **Concluding remarks** (with pretty much of DK's personal views)

Concluding remarks

- I hope you learned how to make machines to encode (understand) and decode (generate) language as tokens in class.
  - In encoding, flexible architectures (e.g., self-attention) that do not drive any inductive biases win.
  - In decoding, current token search prevents the model from deductive reasoning and planning
- NLP has evolved so rapidly and dramatically thanks to collaboration between industry and academia, accelerated by hardware support (e.g., GPUs, Cloud computing), as well as open-source communities (e.g., HuggingFace).
- LLMs follow instructions on multiple tasks and break the boundary between tasks and languages by learning from unlabeled data (pretraining via self-supervision objectives) and labeled instruction data.
- LLM is now capable of using different tools and reasoning with better prompts, as well as self-verifying its own outputs.

Concluding remarks

- In general, scaling law seems to be working so far, but perhaps we're hitting a wall?
- There are still more rooms for other modalities and use cases though (e.g., OpenAI's Sora, coding)
- LLMs are fundamentally limited due to their training objective:
  - Next token predictions can't learn reasoning, planning, or other cognitive capabilities
  - More cognitive alignment methods beyond RL need to be developed
- NLP community needs more societal and sociotechnical evaluation
  - Not computational metrics, but metrics to measure societal values taking ethics, safety, ecosystem, and dominance into account
- Big Techs are starving from high-quality human data
  - Synthetic data is helpful but contains artifacts and biases
  - Lawsuits and copyright issues of AI-generated data

Concluding remarks

- Within the next few years, AI will be everywhere
- In the near future, 90% of online content will be generated by AI or produced in assistance with AI -> What issues?
  - Instructors using AI assistants develop education curriculums
  - Researchers write their manuscripts using chatGPT
  - Annotators annotate data for AI training using another AI like chatGPT
- Most intelligent services in different domains will have at least GPT4-level NLU/NLG abilities
- Various agents will effectively cooperate and solve the complex daily tasks for you
- Unemployment occurs in routine tasks, but AI will help create new types of jobs and most jobs will be assisted by AI

Concluding remarks

Now, you can finetune your own NLP systems and generate text from LLMs. What's next?

- After a couple of years, half of the knowledge you learned in this class will be outdated.
  - Learn yourself and build your own models. Tons of open-sourced tools and LLM services!
- NLP systems (or LLMs) will be used as a tool like Windows, Excel, or Calculator in various domains including industries and multidisciplinary academic research.
- With great power comes great responsibility
  - Don't forget sociotechnical and ethical perspectives of AI systems
- Don't obsess over achieving AGI or improving X.X% on certain benchmarks.
- Be doubtful and critical. There is no ever-lasting technology.
- PLEASE solve important problems using LLMs or other AI tools,
  - Income/education inequality, climate change,
  - Social disparities, diversity and inclusion, Safety/fairness/risks in AI,


[adult swim]