

CSCI 8980: Special Topics in LLM

Lecture 1: Class Overview

Dongyeop Kang, University of Minnesota



UNIVERSITY OF MINNESOTA

Driven to Discover®

Overview

- ❑ Instructors
- ❑ Course information
- ❑ Grading and Late Day Policy
 - Class Participation
 - Homework
 - Project
 - Reading Assignment
- ❑ Outline of Topics
- ❑ Computing resources



Instructors



Dongyeop Kang (DK)
Instructor

- ❑ Dongyeop Kang (DK) <https://dykang.github.io/>
- ❑ Office Hours
 - ❑ Mostly Slack, in-person by appointment
 - ❑ Will have regular async check-in with your project progress
 - ❑ 1-hour Office Hours with each team in the week of Feb 10 and 12.



Course Information

❑ Class meets:

- Tuesday and Thursday, 4PM to 5:15PM
- Appleby Hall 3
- Note that no recording will be available this time

❑ Class homepage

- <https://dykang.github.io/classes/csci8980/S26/index.html>

❑ Project discussion on Slack

- csci8980s26.slack.com

❑ Project submission and grading on Canvas

- <https://canvas.umn.edu/courses/553602>



Outline

- ❑ This graduate level special topics course examines **emerging topics** in large language models.
- ❑ Students will explore **state of the art research** in areas such as cognitive architectures, reasoning and planning, compositionality, social cognition, and test time scaling, as well as applications of language models in domains including law, medicine, journalism, and scientific discovery.
- ❑ Each student (or team) will select a focused topic, conduct a comprehensive literature review, and lead a seminar style lecture and discussion.
 - The course culminates in a semester long research or implementation project, presented as a final paper and in class presentation.



Prerequisite

❑ Required:

- CSCI 2041 Advanced Programming Principles
- CSCI 5521 Machine Learning Fundamentals,

❑ Recommended but not mandatory:

- CSCI 5527 Deep Learning, or
- Any other relevant course that covers fundamental machine learning algorithms.

❑ Furthermore, this course assumes:

- Background in basic **probability**, **linear algebra**, and **calculus**.
- Good coding ability, corresponding to senior undergraduate CS major.
 - ✓ **Assignments will be in Python with PyTorch.**



What we don't cover in this class (1)

❑ Modern NLP techniques and algorithms

- *Computational techniques* developed to enable computers to **understand** and **respond** appropriately to ideas expressed using natural languages.
- Take CSCI 5541 Natural Language Processing Or CSCI 4541 Introduction to NLP
 - <https://dykang.github.io/classes/csci5541/F25/index.html>

❑ Classical ML techniques and NLP algorithms

- E.g., SVM, HMM, CRF, Viterbi
- Take CSCI 5521 Machine Learning

❑ Advanced techniques for NLP models

- E.g., GNNs, Recursive Nets, Diffusion, latent variable models,
- Take CSCI 5980 NLP with Deep Learning



What we don't cover in this class (2)

❑ Human aspects of AI and NLP systems

- E.g., interaction design, formative/field study, Qual analysis, theory grounding
- Take CSCI 5115 User Interface Design or other HCI classes

❑ Linguistic aspects or computational linguistics:

- Take classes in Institute of Linguistics
<https://cla.umn.edu/linguistics/graduate/courses>

❑ Basic Python programming skills

- Basic tutorials on PyTorch and HuggingFace programming will be given by Tas



Class Slack

- ❑ csci8980s26.slack.com
- ❑ Slack will be the primary method of communication.
- ❑ “prj-TEAMNAME-with-DK”
 - With instructors
- ❑ “prj-TEAMNAME”
 - Without instructors

prj-b-zae-semanticons

Group A/B Mentor Project name

- ▼ projects
 - 🔒 prj-a-dk-golish-project
 - 🔒 prj-a-dk-lm_bros
 - 🔒 prj-a-dk-the-adversaries
 - 🔒 prj-a-dk-transformative-attentors
 - 🔒 prj-a-shirley-clickbait-analysis
 - 🔒 prj-a-shirley-pilot
 - 🔒 prj-a-shirley-verbavores
 - 🔒 prj-a-zae-word-nerds
 - 🔒 prj-b-dk-language-model-alignment
 - 🔒 prj-b-dk-tbd
 - 🔒 prj-b-shirley-title-fight
 - 🔒 prj-b-shirley-word-wizards
 - 🔒 prj-b-zae-flashcard-generator
 - 🔒 prj-b-zae-nlp-vision
 - 🔒 prj-b-zae-semanticons
 - 🔒 prj-b-zae-vjk



Canvas

canvas.umn.edu/courses/553602

☰ ▾ Project

☰ 📁 **Team formation**
Due: Sep 18 at 11:59pm | 1 pts

☰ 📁 **Project brainstorming**
Due: Sep 25 at 11:59pm | 1 pts

☰ 📁 **Proposal pitch**
Due: Oct 9 at 11:59pm | 3 pts

☰ 📁 **Proposal report**
Due: Oct 14 at 11:59pm | 5 pts

☰ 📁 **Midterm office hour participation**
Due: Nov 14 at 11:59pm | 5 pts

☰ 📁 **Poster presentation**
Due: Dec 4 at 11:59pm | 5 pts

☰ 📁 **Final report**
Due: Dec 12 at 11:59pm | 10 pts

☰ ▾ Participation

☰ 📁 **Slack Participation**
4 pts

☰ 📁 **Office Hour Participation**
4 pts

☰ 📁 **In-person Class Participation**
3 pts



Grading and Late Day Policy

- ❑ 30% Topic Presentation
- ❑ 50% Project (team)
- ❑ 20% Class Participation (individual)

- ❑ Each student will be granted **2 late days** to use for project assignment over the duration of the semester. After all free late days are used up, penalty is 1 point for each additional late day.
- ❑ The late days and penalty will be applied to **all team members** for group homework and project.
- ❑ DRC and other medical urgencies can only be accommodated. only with a *doctor's note or a letter from the DRC. They should be noted "in advance"*



Class participation (20%)

- ❑ Your class participation is **thoroughly** evaluated.
- ❑ The following metrics will be used to grade your participation:
 - Participation and discussion in class
 - **Discussion on Slack and during Office Hours for instructor
 - Discussion and QA during the presentation of the topic presentation and project presentations
- ❑ We explicitly **count the number of your offline and online participation**, and (min/max) normalize them at the end of the class.
- ❑ Your participation score will be zero if you haven't participated in class, Slack or other discussions.



Topic Presentation

- ❑ Four papers assigned per block (two lectures) and typically two papers discussed in each class lecture.
- ❑ Each paper will be presented in 30 minutes talk and 15 minutes discussion
- ❑ Each student is required to present **a total of two to three papers** over the semester. Students may indicate their topic and paper preferences using the provided [interest form](#).
 - at least one paper from the human centered track and one from the machine centered track.



Topic Presentation

- ❑ Each presenting student prepares a **structured synthesis** of the topic and **leads the class** through a lecture style presentation and discussion.
- ❑ Presentations must include a **curated reading list, summary of key technical ideas and takeaways**, and a **limitation or discussion effort** when feasible.
- ❑ Sessions follow a seminar style format, with presenters responsible for moderating discussion and encouraging critical engagement from the class.
- ❑ **Bonus Point:** Fruitful discussions and insightful questions during presentations may be rewarded with bonus participation points at the instructor's discretion.



Use of AI Tools (e.g., chatGPT)

- ❑ **Encouraged** to use AI tools like ChatGPT, CoPilot, etc.
 - ❑ You can use it to:
 - Clarify concepts learned in class
 - Get instant answers to your questions before emailing TAs and instructor
 - Test your knowledge by asking it questions
 - ❑ **Warning:**
 - Please properly mention what kinds of help you get from the tools in your homework and project (missing acknowledgement will be treated as **cheating**)
 - Use the tools with responsibility:
 - ✓ Generative AI tools often **hallucinate** and only provide **general** answers
- Again, missing acknowledgement of AI usage will be treated as cheating**



Project Goal

- A course project would be one of the following types:
 - New research results judged suitable for acceptance to a NLP or ML workshop,
 - Critical analysis of existing NLP models or datasets,
 - Collection of your own dataset on new problems or adversarial datasets that can fool the existing systems ,
 - An in-depth literature survey on emerging topics,
 - Interactive demonstration (e.g., Chrome Extension, Flask) or visualization of existing systems,
 - New open-source repository or dataset with a high impact on the community



Project Goal

- ❑ Please carefully read the project guideline later once posted on Thursday
- ❑ Make regularly weekly progress with your teammates
- ❑ Have weekly sync-up with DK over Slack.
- ❑ If needed, set up in-person office hour meetings with DK



Project Evaluation Rubric

- ❑ HWs are generously graded but the **projects are not!** Therefore, students should consider the potential contribution of the projects rather than **trying to play it safe**. Playing it safe won't give them full marks.
- ❑ Three important rubrics:
 - **Novelty:** Compared to the state-of-the-art methods/systems/datasets, how novel is your approach?
 - **Significance:** How strong is your result? Is your finding still holding if different setups or prompting tricks?
 - **Clarity:** How clear and easy-to-follow is your report? Do you have well organized presentation of your results and problem definition?
- ❑ <https://dykang.github.io/classes/csci8980/S26/rubric.html>



Project Deliveries and Dues (50%)

- ❑ Team formation (due: **Early Feb**)
- ❑ Proposal pitch during DK office hours (5%, due: **Feb 10, 12**)
- ❑ Proposal report (5%, due: **Feb 20**)
- ❑ Midterm presentation (10%, due: **March 17, 19**)
- ❑ Final presentation (15% points, due: **April 28, 30**)
- ❑ Final report (15% points, due: **May 8**)



Class Projects

The course includes a semester long research or implementation project. Projects should go beyond surface level performance gains and instead focus on **novel failure modes, under explored behaviors, or emerging risks** of large language models or agentic systems. Incremental leaderboard gains or prompt only tweaks are discouraged. Projects may be completed individually or in small teams, maximum 2 members. Each project must be **DK confirmed** during DK Office Hour week (Feb 10, 12) and grounded in a clear research question, supported by relevant literature and empirical analysis.

Project Scope

Students are encouraged to produce one of the following research artifacts:

- A benchmark or evaluation suite
- A dataset capturing non trivial behaviors or failure modes
- A measurement or diagnostic framework
- A failure taxonomy or behavioral analysis
- A mitigation, repair, or control algorithm

Deliverables

Each team must submit a written report, reproducible code or data, and presentation materials via Canvas. Reports should follow a standard conference paper format, ACL style preferred, using templates via [Overleaf](#) or [GitHub](#).

Milestones and Timeline

The following milestones are aligned with the course schedule. Canvas submission links will be provided.

- **Team formation and project idea** Due: **early February**
- **Project proposal** (problem statement, related work, plan) Due: **mid February** (Canvas link TBD)
- **Midterm presentation** In class: **March 17 to 19**
- **Final presentation** In class: **April 28 to 30**
- **Final report and code submission** Due: **May 8 (Friday)** (Canvas link TBD)

Projects are expected to be reproducible, clearly scoped, and analytically grounded. Strong projects typically combine careful problem formulation with diagnostic evaluation, stress testing, or controlled ablation studies. **Evaluation emphasis:** originality, rigor, and insight into model behavior under realistic, adversarial, or long horizon conditions rather than raw performance.

See [evaluation rubric](#) for final reports.

Selected Past Projects

Reference examples. Links will be added when available.

- *Simulating Everyone's Voice: Exploring ChatGPT's Ability to Simulate Human Annotators* (report and poster TBD)
- *Vision and Language guided Generalized Object Grasping* (report and poster TBD)
- *Generating Controllable Long dialogue with Coherence* (Published in AAAI 2024, link TBD)



VLANGO Gh: Vision and Language guided Generalized Object Grasping

CSCI 5541 Spring 2023
Nikhilanj Pelluri



Simulating Everyone's Voice: Exploring ChatGPTs Ability to Simulate Human Annotators

CSCI 5541 Spring 2023

Abdirizak Yussuf, Claire Chen, Dinesh Challa, Venkata Sai Krishna

Step 1

Scraping and filtering data.



Step 2

Human annotation.

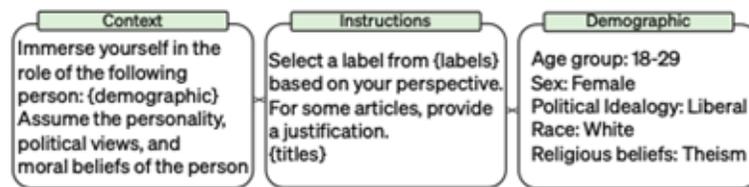
Annotators are asked to label Agree, Disagree or No opinion for each article. For 10 articles, they also provide a justification.



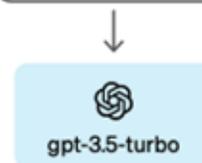
Step 3

ChatGPT annotation.

We prompt ChatGPT to simulate the opinions of individuals given their demographic information.



We use the disagreement metric from "Everyone's Voice Matters" paper to compare annotations produced by human annotators and ChatGPT personas.



Topic	Human Annotators	ChatGPT Personas
Abortion	0.22	0.32
Immigration	0.15	0.40
Social Issues	0.11	0.40
Political Issues	0.017	0.50
Racial Justice	0.19	0.40
Religion	0.18	0.36
All Topics Combined	0.15	0.42

- **Human annotators: 0.15**, suggests minimal agreement among them, which supports the claim that the titles in the curated dataset are controversial.
- **ChatGPT personas: 0.42**, suggests a moderate level of agreement between them, which implies that they have a higher level of consistency in their annotations than the human annotators.



Who is speaking? Distinguishing Artificial Intelligence Generated and Human Written Text

CSCI 5541 Spring 2023
Moyan Zhou, Mingsheng Sun, Yutong Sun

How Robotic is Your Content?

51%

🤖 Unclear if it is AI content!

62%

Predictability

29%

Probability

53%

Pattern

Want Undetectable AI Content?

Our proprietary content platform uses a mix of 3 AI engines, NLP and semantic analysis algorithms, crawls Google, and parses all the top ranking content to put research-backed, long-form, SEO driven blog posts together.

This isn't an AI writing assistant, this is a human level, long-form, blog post producing machine!

Request an Invite

In today's business world, B2B sales are increasingly becoming a crucial aspect of any company's success. The business-to-business (B2B) market involves selling products or services to other businesses, rather than to individual consumers. Whether you are just starting out in B2B sales or are looking to improve your current approach, this blog post will provide you with strategies that can help you succeed in the B2B market.

Know Your Target Market

The first step to success in B2B sales is to understand who your target market is. Who are the businesses that you want to sell to? What are their needs and wants? What are the pain points that your products or services can solve? The better you understand your target market, the easier it will be for you to create a sales strategy that appeals to them.

Build Relationships

Billions of virtual ad impressions with over 100,000,000 unique visitors. Up to 25,000 characters will be used. 1666 Characters

Want to see examples? Fully Human Human + AI

GPT-3 Chat GPT Content at Scale AI

Predicted based upon 273 words.

Check For AI Content

Results

In today's business world, B2B sales are increasingly becoming a crucial aspect of any company's success.

The business-to-business (B2B) market involves selling products or services to other businesses, rather than to individual consumers.

Whether you are just starting out in B2B sales or are looking to improve your current approach, this blog post will provide you with strategies that can help you succeed in the B2B market.

Know Your Target Market

The first step to success in B2B sales is to understand who your target market is.

Who are the businesses that you want to sell to?

What are their needs and wants?

What are the pain points that your products or services can solve?

The better you understand your target market, the easier it will be for you to

RQ1: Do people agree with each other when distinguishing AI-generated and Human-written text?

Fleiss' Kappa
0.05 (p-value = 0.017)

RQ3: How does the existing tools work?

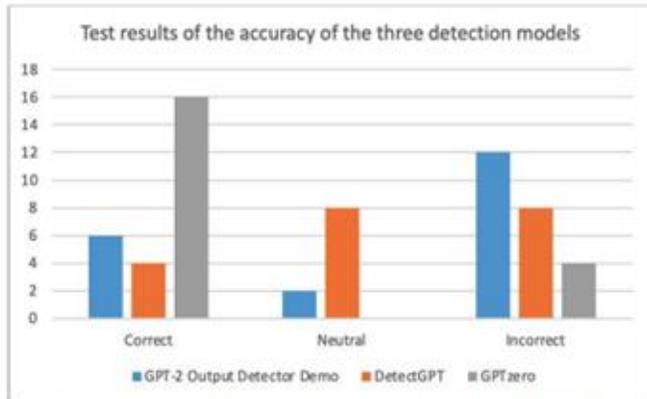


Figure 5. Test results of the accuracy of the three detection models

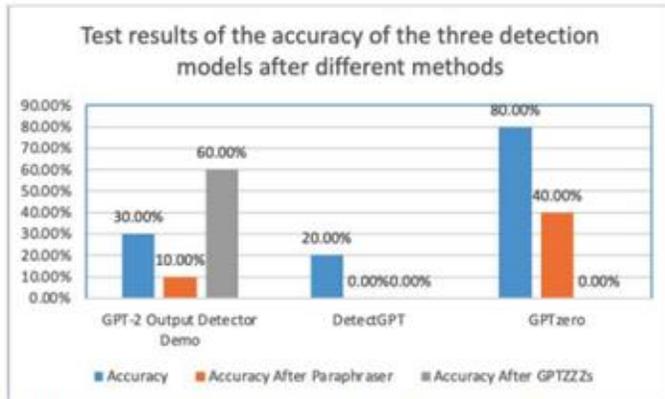


Figure 6. Test results of the accuracy after different methods



Comparing the Effectiveness of Fine-tuning vs. One-Shot Learning on the Kidz Bopification Task

CSCI 5541 Spring 2023

Jeonghoon Kim, Matthew Olson, Marco Berriodi

“I buy a new car for the b**** (For real)
I tear down the mall with the b**** (For real)”
from Bank Account by 21 Savage

“Lean with me, pop with me
Get high with me if you rock with me
Smoke with me, drink with me
F***** up liver with some bad kidneys”
from Lean Wit Me by Juice WRLD



You are a helpful assistant who is tasked with cleaning up songs. You will remove explicit material from a song's lyrics.

So we'll piss off the neighbors
In the place that feels the tears
The place to lose your fears
Yeah, reckless behavior
A place that is so pure, so dirty and raw
In the bed all day, bed all day, bed all day
Fucking and fighting on
It's our paradise and it's our war zone
It's our paradise and it's our war zone



Here's a cleaned-up version of the lyrics:

We'll tick off the neighbors
In the place that feels the tears
The place to lose your fears
Yeah, wild behavior
A place that is so real, so gritty and raw
In the bed all day, bed all day, bed all day
Arguing and laughing on
It's our haven and it's our battleground
It's our haven and it's our battleground.



Transidiomation: Optimizing translation of idioms embedded in text

CSCI 5541 Spring 2024

Dhondup Dolma, Jaeun Lee, Yongtian Ou, Jiyeon Pyo Team

NLPitch

English sentences with PIE	GPT 3.5	Google Translate	SeamlessM4T
It was raining cats and dogs last weekend.	지난 주말에는 개구리가 비 내리는 것 같았어요. It was like frogs were raining last weekend. (hallucination)	지난 주말에는 고양이와 개에게 비가 내렸습니다. Cats and dogs were rained on last weekend. (incorrect)	지난 주말에는 고양이와 개가 비를 내고 있었다. Not translatable
Getting that done was a piece of cake.	그것을 해내는 것은 쉬운 일이었다. It was easy to get that done.	그 일을 끝내는 건 아주 쉬운 일이었어 It was easy to get that done.	그 일을 처리하는 것은 쉬운 일이었습니다. It was easy to get that done.
Teaching is my bread and butter.	가르치는 것이 내 소득원이다. Teaching is my source of income.	가르치는 것이 나의 빵과 버터이다. Teaching is my bread and butter. (literal translation)	가르치는 게 내 과 버터야. Not translatable

Table 1: Examples of various NMT models translating English sentences with PIE to Korean. Text that have red Xs are incorrect translations.

“Translate the sentence into target language by following these steps:

Step1. Identify the idiom.

Step2. Find an idiom with the same meaning in the target language. If there is no equivalent idiom, give the figurative meaning of it.

Step3. Include response from Step2 to translate the sentence.”

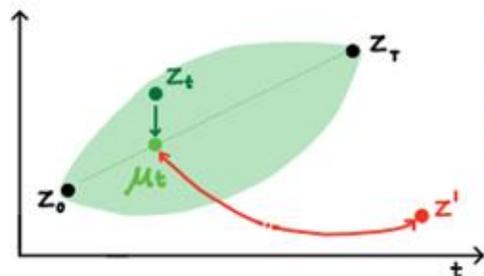


Generating Controllable Long-dialogue with Coherence

CSCI 5980 Fall 2022

Zhecheng Sheng, Chen Jiang and Tianhao Zhang

Time control in language model using Brownian bridge (Wang et al., ICLR 2022)



x_0 : [USER] Hello, I'd like to buy tickets for tomorrow.

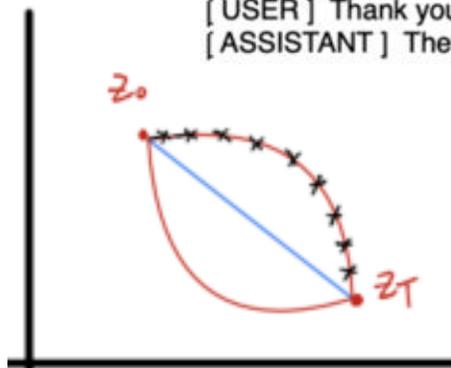
x_t : [ASSISTANT] What movie theater do you prefer?

x_T : [USER] Could you confirm my tickets just in case?

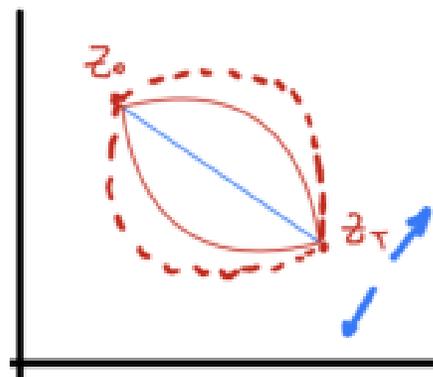
x' : [USER] Hi, I'm looking to purchase tickets for my family.

$$\mathcal{L} = -\log \frac{\exp(d(z_t, \mu_t))}{\exp(d(z_t, \mu_t)) + \exp(d(z', \mu_t))}$$

[USER] I am thinking about seeing a movie tonight, please.
 [ASSISTANT] What movie do you have in mind?
 [USER] The Elizabeth Theatres.
 [ASSISTANT] They have 2 tickets available.
 [USER] Thank you.
 [ASSISTANT] They have been waiting in the line for you.



[USER] I would love to get to the movies tonight.
 [ASSISTANT] OK. And where will you be seeing the movie?
 [USER] Creek's End, Oregon.
 [ASSISTANT] Creek's End, Oregon. Got it. Is there a particular movie you have in mind?
 [USER] No wait, the visuals are so darned.
 [ASSISTANT] No problem.
 [USER] No problem.
 [ASSISTANT] No problem.
 [USER] No problem.
 [ASSISTANT] No problem.
 [ASSISTANT] No problem.



Published in AAAI 2024, <https://arxiv.org/abs/2312.16893>



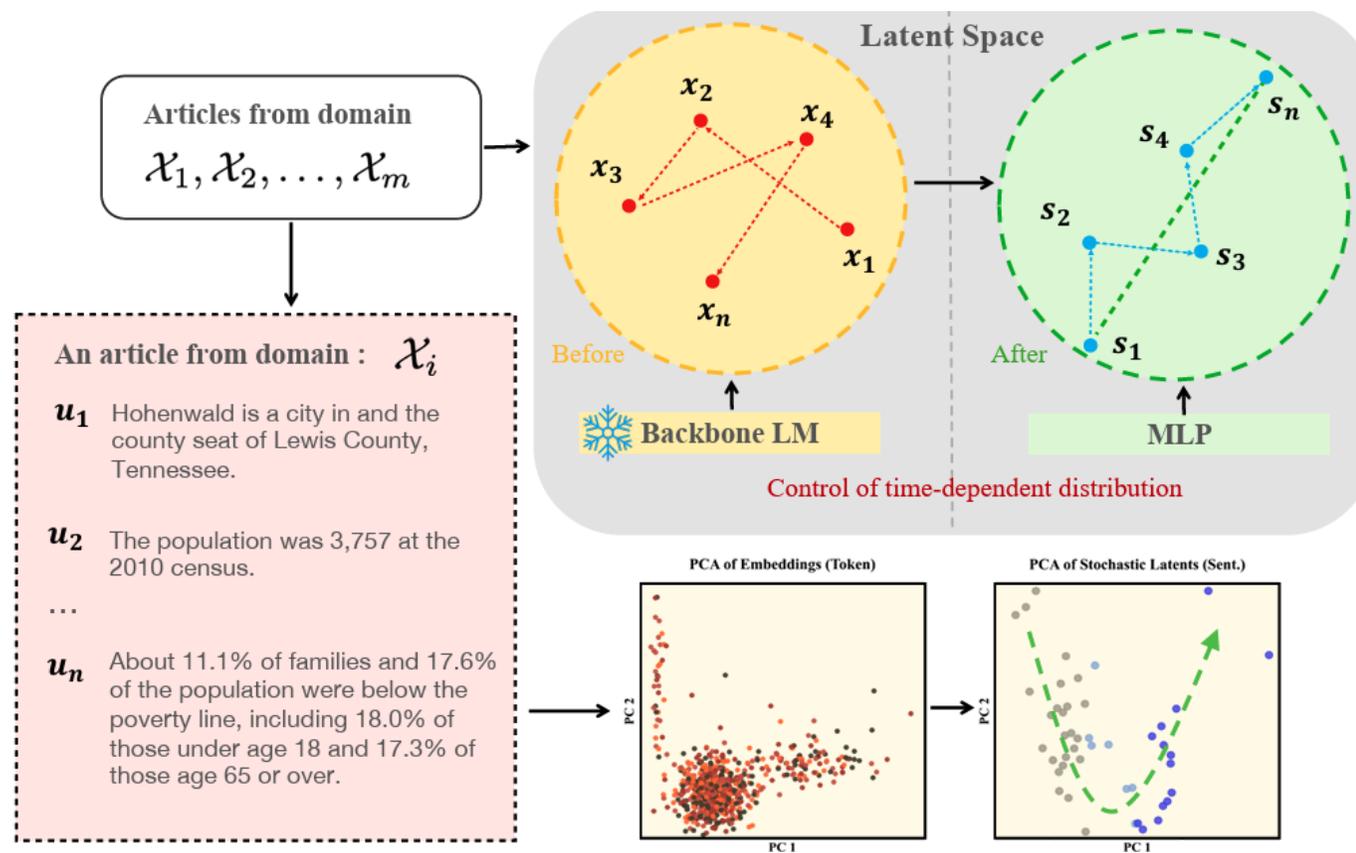


Figure 1: Schematic diagram of the Stochastic Representation in the latent space. An article from domain \mathcal{X}_i , segmented into sentences (u_1, u_2, \dots, u_n) , is processed by the encoder which consists of a pre-trained language model (LM) and a multi-layer perceptron (MLP). The encoder maps each sentence into latent space and after optimizing for the stochastic objective, the latent trajectory becomes time dependent.

The Overdose Effect of AI-Assisted Writing on Students

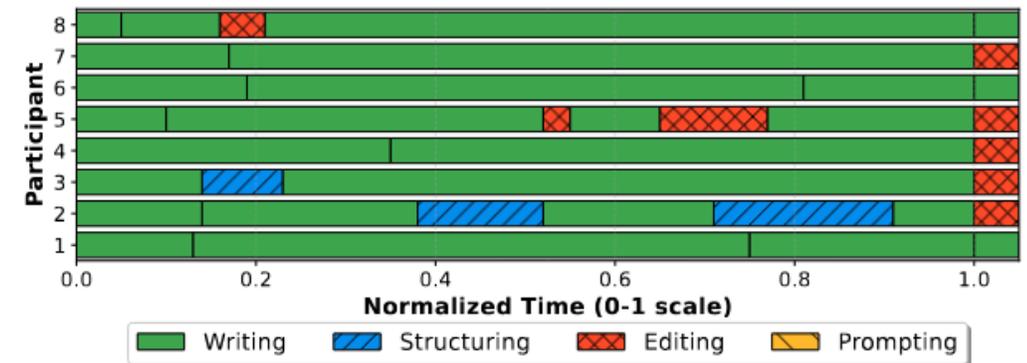
JAM: Julia Christenson, Ava Cavalco, Maddie Johns, CSCI 5541 Fall 2025

Example Essay #1

Many people state that college is the best year of your life. I believe that is because it is a time for exploration and opportunities. This is the time to take new risks and try everything that interests you. However, students in college are at varying places in the world. Some have the privilege of getting financial help, whereas others may need to work multiple jobs to stay stable while getting their degree. While career readiness is important in a selection of classes, students should balance these with classes that also interest them.

Taking classes based on interest can fuel one's desire to learn and boost motivation. When students are actually curious about a topic, they will participate more in class and engage more deeply with the material. Connecting this to my own experience, I took a class on rock music, even though I am not studying music. I found myself enjoying the class a lot more because I enjoyed the topic and learning about music history.

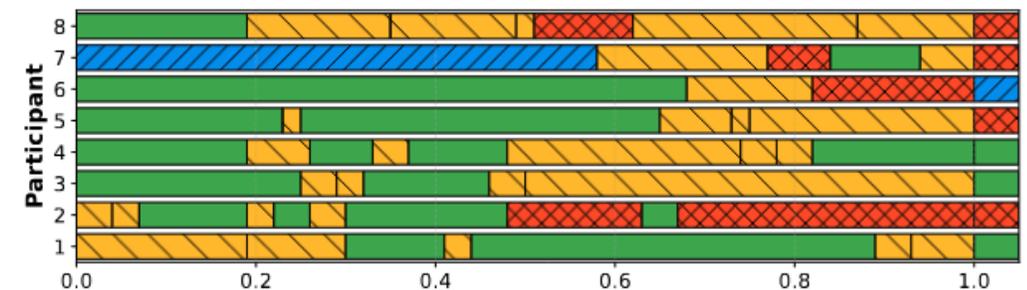
Apart from student success in college, post-college happiness is often tied to fulfillment. Research shows that while income certainly matters up to a certain point (often cited around \$75,000 to alleviate daily stress), long-term job and life satisfaction are more closely tied to intrinsic factors like meaning and interest. These classes you can take in college are much more worth exploring if it could mean you find a new passion or even a new hobby.



(a) Control group



(b) Limited AI group



(c) Unlimited AI group

Figure 5: Distribution of writing intentions across the writing process for all user groups.



Abstain-R1: Calibrated Abstention and Post-Refusal Clarification via Verifiable RL

Abstain-R1: Haotian Zhai, Jingcheng Liang, Haotian Huang, Zekang Li, CSCI 5541 Fall 2025



Unanswerable Query: "There exists a positive real number x such that $\cos(\arctan(x)) = x$. Find the value of $x^2 + y$ "

✗ Not absent and hallucinate

<thinking>

Let $\theta = \arctan(x)$, so $\tan(\theta) = x$. Using a right triangle with opposite side x and adjacent side 1, the hypotenuse is $\sqrt{x^2 + 1}$. Thus, we have $\cos(\arctan(x)) = 1/\sqrt{x^2 + 1} = x$, which simplifies to $1 = x^4 + x^2$. **Let $y = x^2$, solving this quadratic equation yields $y = (-1 \pm \sqrt{5})/2$.** Since $y = x^2$ must be positive, we take $y = (-1 + \sqrt{5})/2$. Because $y = x^2$, we have $x^2 + y = 2x^2$. Substituting $x^2 = (-1 + \sqrt{5})/2$ gives $2x^2 = -1 + \sqrt{5}$.

</thinking>

<answer>

$\boxed{-1 + \sqrt{5}}$

</answer>

✗ Absent w/ incorrect clarification

<thinking>

The problem states that there exists a positive real number x such that $\cos(\arctan(x)) = x$, and then asks for the value of $x^2 + y$. It is unclear whether the solution is unique. If multiple values of x satisfy the equation, the resulting value of $x^2 + y$ may not be well defined. Clarification is needed on whether the solution for x is unique.

</thinking>

<answer>

$\boxed{\text{I don't know.}}$ Please clarify whether the equation $\cos(\arctan(x)) = x$ has a unique positive real solution.

</answer>

✓ Absent w/ correct clarification

<thinking>

The problem states that there exists a positive real number x such that $\cos(\arctan(x)) = x$, and then asks for the value of $x^2 + y$. While the condition on x is well-defined, the variable y is never introduced or defined anywhere in the problem statement. Without a definition of y , the expression $x^2 + y$ cannot be uniquely determined.

</thinking>

<answer>

$\boxed{\text{I don't know.}}$ The problem is incomplete because the variable y is not defined. Please specify what y represents.

</answer>



CSCI 5541 Fall 2025

Natural Language Processing Course Project Showcase

Search by topic, team, title, or member.

Visible projects

16

Search projects, tags, members

All

Group A

Group B

All topics

Group B Dec 2 Quick view



Embedding Squad
Compiler Verified RL for Code Vulnerability Detection

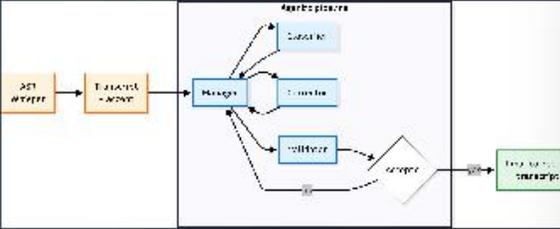
VulnRL uses compiler and analyzer feedback as reward signals to improve vulnerability detection beyond supervised labels.

Security RL Code

TB AJ NS

Open project

Group B Dec 2 Quick view



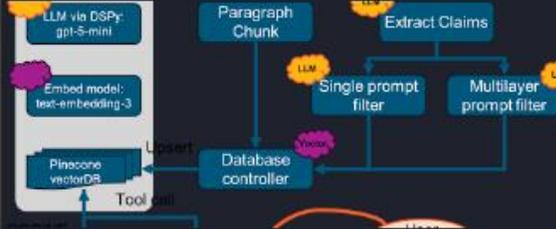
MultiAgentTeam
Multi-Agent Post-Processing Pipeline for Non-Native English ASR

A text-only multi-agent pipeline that improves non-native English ASR transcripts via error analysis, correction, and evaluation across accents.

Speech Agents Evaluation

Open project

Group B Dec 2 Quick view



NetWatch
Appropriateness Filtering of Claims for Corporate RAG Knowledge Base Storage

Pre-embedding, claim-level filtering for enterprise text using modular risk criteria to improve RAG quality and reduce inappropriate storage.

RAG Safety Filtering

Open project

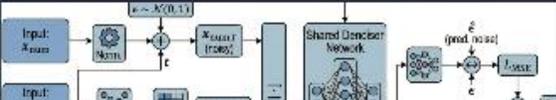
Group A Dec 4 Quick view



Group A Dec 4 Quick view



Group A Dec 4 Quick view



Topics

<https://dykang.github.io/classes/csci8980/S26/index.html>

Course Schedule

We will cover a variety of **machine-centered topics** and **human-centered topics**. Based on knowledge you gain during the class, your team will conduct **semester-long** 'research' project. Pay attention to **due dates** for projects.

Date	Topic and Focus	Presenters	Reading for Presentation	Other Reading
01-20	Class Overview			
01-22	Turntable Research Discussion			
01-27	Test Time Scaling and Self Evolving Agents inference time emergence with focus on dynamic compute allocation and self improving agent systems Subtopics: Test time compute scaling laws; Adaptive inference and early exit; Self evolving agents and organizations; Latency	0127-TTS-A- {Presenter}- {TITLE} 0127-TTS-B- {Presenter}- {TITLE}	<ul style="list-style-type: none">AlphaEvolve: A coding agent for scientific and algorithmic discovery 2025The Era of Agentic Organization: Learning to Organize with Language Models, 2025Can 1B LLM Surpass 405B LLM? Rethinking Compute Efficient Test Time Scaling, ICLR 2025	<ul style="list-style-type: none">A Survey of Self Evolving Agents: On Path to Artificial Super Intelligence 2025A Theory of Response Sampling in LLMs: Part Descriptive and Part Prescriptive, ACL 2025AgentEvolver: Towards Efficient Self Evolving Agent System, 2025
01-29	cost reliability tradeoffs Research Directions: Dynamic inference controllers; Benchmarks for adaptive inference and agent self improvement; When search beats learned reasoning; Failure modes of self evolving agents	0129-TTS-A- {Presenter}- {TITLE} 0129-TTS-B- {Presenter}- {TITLE}	<ul style="list-style-type: none">Beyond Pipelines: A Survey of the Paradigm Shift toward Model native Agentic AI, 2025	<ul style="list-style-type: none">Scaling Unverifiable Rewards: A Case Study on Visual Insights 2025s1: Simple test time scaling
02-03	Expert AI and Workflow Modeling	0203-Expert-A-	<ul style="list-style-type: none">The AI Scientist: Towards Fully	<ul style="list-style-type: none">How Does Time Horizon Vary Across



Academic Integrity

- ❑ Everything you turn in must be your own work, and you must note **the names of anyone** you collaborated with on each problem and cite **any resources** (e.g., chatGPT prompts you used) that you used to learn about the problem.
- ❑ If you have any doubts about whether a particular action may be construed as cheating, ask the instructors for clarification before you do it. Cheating in this course will result in a grade of **F** for course and the [University policies](#) will be followed.



Computing Resources

- ❑ Google Colab Pro account by CS&E department
 - Current roaster has been granted

- ❑ MSI account
 - Current roaster has been granted

- ❑ NVIDIA DGX Spark 2 nodes



Next (Jan 22)

- Project details
- Make final enrollment decisions
- Skim through reading list and assign yourself to papers you like to present
- Prepare 5m pitch of your project ideas and interests for turnaround table discussion (Thursday)

