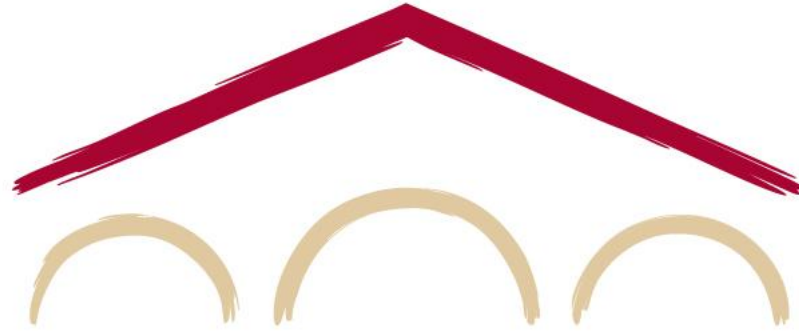


Lit review overview



Christopher Potts
CS224u: Natural Language Understanding

Rationale

The goal of this assignment (and the next one) is to enter into a productive dialogue with your teammates and with your mentor.

- Pose questions
- Identify obstacles and propose workarounds
- Find datasets and models
- Think carefully about your resources

Requirements

<https://web.stanford.edu/class/cs224u/projects.html#litreview>

Gradescope submission format

Lit review

Student Name

Search students by name or email...

Q1 Team member email addresses
0 Points

First member email:

cgpotts@stanford.edu

Second member email:

Enter your answer here

Third member email:

Enter your answer here

✓ Correct

Save Answer Last saved on May 08 at 8:06 AM

Q2 Planned topic
0 Points

In one sentence, briefly describe your planned topical area of your project. This need not be final

Causal abstraction techniques for explaining how models do in-context learning

Q3 Preferred Mentor
0 Points

Kawin Ethayarajh (data, metrics, evaluation)

Sidd Karamcheti (robot learning, grounding, scaling)

Mina Lee (Human-AI Interaction, HCI, LLMs)

Lisa Li (Diffusion models, prefix tuning, in-context learning)

Siyan Li (Human-centered NLP, semantics, digital humanities)

Tolúlope Ògúnremí (Multilingual and low-resource NLP)

Tianyi Zhang (LLMs, emergence, in-context learning)

Any of the above

✓ Correct

Save Answer Last saved on May 08 at 8:06 AM

Q4 Lit review document
10 Points

Please upload your literature review PDF here:

Currently Uploaded Files

CS224u_lit_review.pdf Download Remove

Please select file(s) Select file(s)

Save Answer Last saved on May 08 at 9:02 AM

Gradescope rubric

4: Lit review document ▾

1 of 1 graded

Total Points
8.0 / 10.0 pts

[⚙️ Rubric Settings](#)
[Collapse View ▾](#)

▾ **1** **+2.0** [⚙️ Select many](#) **✕**

Problem/task

Q **+2.0**
The general problems/tasks are clearly articulated.

W **+1.0**
It is difficult to tell what problems or tasks this lit review is focussed on.

E **+0.0**
No problems or tasks are clearly articulated.

[+ Add Rubric Item to Group](#)

▸ **2** **+3.0** [⚙️ Select many](#) **✕**

Summaries and compare and contrast

+3.0 The summaries are fairly cursory but the comparisons are useful.

Lit search tips

1. Search with keywords in the [ACL Anthology](#), [Google Scholar](#), or [Semantic Scholar](#).
2. Download relevant and/or highly-cited results and check out their abstracts and related work sections, seeking out key questions and techniques, and other highly-cited papers. *Do not try to read entire papers at this point.*
3. Download papers that you see prominently in the related work sections.
4. Return to step 1 with some new keywords.
5. Break out of the 1–4 loop when you have a sense for what you are doing and what others have done in the area.
6. Select some core papers from your downloaded set to read deeply and cover in the lit review.

Plagiarism policy

- Make sure you know [the course policy](#).
- There is no rule against using an AI assistant to help you with the lit review, but all output from the model needs to be quoted, per the policy.
- Assignments consisting of lots of quoted text will not receive good grades.
- Assignments with substantial overlap in prose will be scrutinized for plagiarism.

USER	Summarize the ELECTRA paper and explain how it relates to the Transformers paper "Attention is all you need".
ASSISTANT	<p>The ELECTRA paper, titled "ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators," introduces a new pre-training method called ELECTRA for natural language processing (NLP) models. The authors, Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning, propose a more efficient approach to pre-training that outperforms the popular masked language model (MLM) pre-training methods, such as BERT, while requiring less computation.</p> <p>ELECTRA works by training a discriminative model to distinguish between "real" and "fake" tokens in a given text. The model consists of two components: a generator that produces "fake" tokens by corrupting input text and a discriminator that attempts to identify whether tokens are original or generated. By forcing the model to focus on the differences between real and artificial tokens, ELECTRA learns to capture complex language representations. This approach is different from the MLM-based methods (e.g., BERT), which involve predicting masked tokens in the input text.</p> <p>The relation between the ELECTRA paper and the Transformers paper, titled "Attention is All You Need," which introduced the transformer architecture, lies in the underlying model used</p>

The next assignment: The protocol

This is a short, structured report designed to help you establish your core experimental framework. The required sections are as follows:

1. Hypotheses
2. Data
3. Metrics
4. Models
5. General reasoning
6. Summary of progress so far
7. References section

Goal: clarity of project goals, identification of obstacles and project risks.

Other tips and resources

<https://github.com/cgpotts/cs224u/blob/main/projects.md>