**224n Project Ideas**

1. Plagiarism detector
   1. call for paper:<https://www.frontiersin.org/research-topics/50171/plagiarism-detection-using-natural-language-processing-techniques>
2. Counter disinformation
   1. <http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=171094&copyownerid=320>
3. disaster management
   1. data processing for disasters is massive, heterogenous task due to all sorts of input sources
   2. <http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=157371&copyownerid=20876>
4. speech language tech for dravidian languages
   1. <http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=153458&copyownerid=89933>
5. CHoRaL —> massive dataset of humorous and non-humorous FB posts. Can use this for something social media related?
   1. <https://aclanthology.org/2021.emnlp-main.364/>
6. Medical report summarization?
   1. MIMIC-CXR dataset
7. Some sort of legal, accounting related summarizer or key words finder? Can greatly help expedite routine work.
8. Student code understanding, evaluation, style suggestions? Contacted Chris Piech, waiting for reply
   1. <https://arxiv.org/abs/2107.14035>

​​meeting notes 2/4

- course search tool

- code style? piech office hours 3-5 pm monday

- go at 4:30pm

- medical / legal information summarization

- policy optimization - NLP + RL

from the link:

- baby LM challenge

- redundancy in contextual representations

- automating english language

- NLP reasoning for chess / poker / other game?