

What do I want to build?

- For who: AI reading clubs @ companies (research engineers)
 - They want to find interesting new topics that might help achieve business goals
 - They are interested in hiring relevant people/working with people who share similar interests
 - They are not actively in one field (mix between research and implementation)
 - Get insight into what each team is reading (including own team) -> Summaries of what other read

[!- **Research Distribution:** Over last few decades, there's been a big push for research to be freely available online. This includes the formation of [arXiv.org](https://arxiv.org) and [PLOS](https://plos.org), and [journal editorial boards resigning](#) to start open-access journals.

Increasingly, the challenge is filtering accessible content. Karpathy's [ArXiv Sanity](<http://www.arxiv-sanity.com/>) is a lovely tool for this. Crowd curation by online communities also helps a great deal.]

-

- Personalized recommendations
- Summaries of papers
- What people are reading (based on the citations in their papers)
- Complimentary recommendations (for teams to get a wider area of knowledge)
- Suggestions for people to work with (most similar research taste)
- Based on what "similar researchers read" but also based on content (to rank)
- Information used to decide if a paper is good or not:
 - Huggingface papers: <https://huggingface.co/papers/2412.15115>
 - BlueSky:
https://www.reddit.com/r/reinforcementlearning/comments/1gwqp7s/blue_sky_researcher_starter_packs_for_mlairl/
 - Citations
 - How it is being introduced in papers (semantic analysis)
 - Papers with code: <https://paperswithcode.com/>
- Useful for:
 - Reading/journal clubs

- Research teams looking to collaborate
- Researchers curious about what others are reading
- Research engineers to find cool new papers

First things first

What do I want to build by the end of winter break?

- Downloading papers pipeline
 - Auto download when new papers gets uploaded
 - Create reading lists of each author
- Find most similar documents (using SPECTER features)
- Find most similar author

Resources

- <https://github.com/karpathy/arxiv-sanity-preserver>
- <https://github.com/karpathy/arxiv-sanity-lite?tab=readme-ov-file>
- <https://arxiv.org/abs/2004.07180> (maybe do contrastive learning in hyperbolic space? with hyperbolic transformer)
- <https://www.kaggle.com/datasets/Cornell-University/arxiv>
- <https://github.com/kaustubh187/Research-paper-recommender-system?tab=readme-ov-file>

Things to research

Things to do