



Top 10 ML & NLP Project Ideas

With Dr. Rachael Tatman

● LIVE

Can you help me come up with
an idea for a course project?



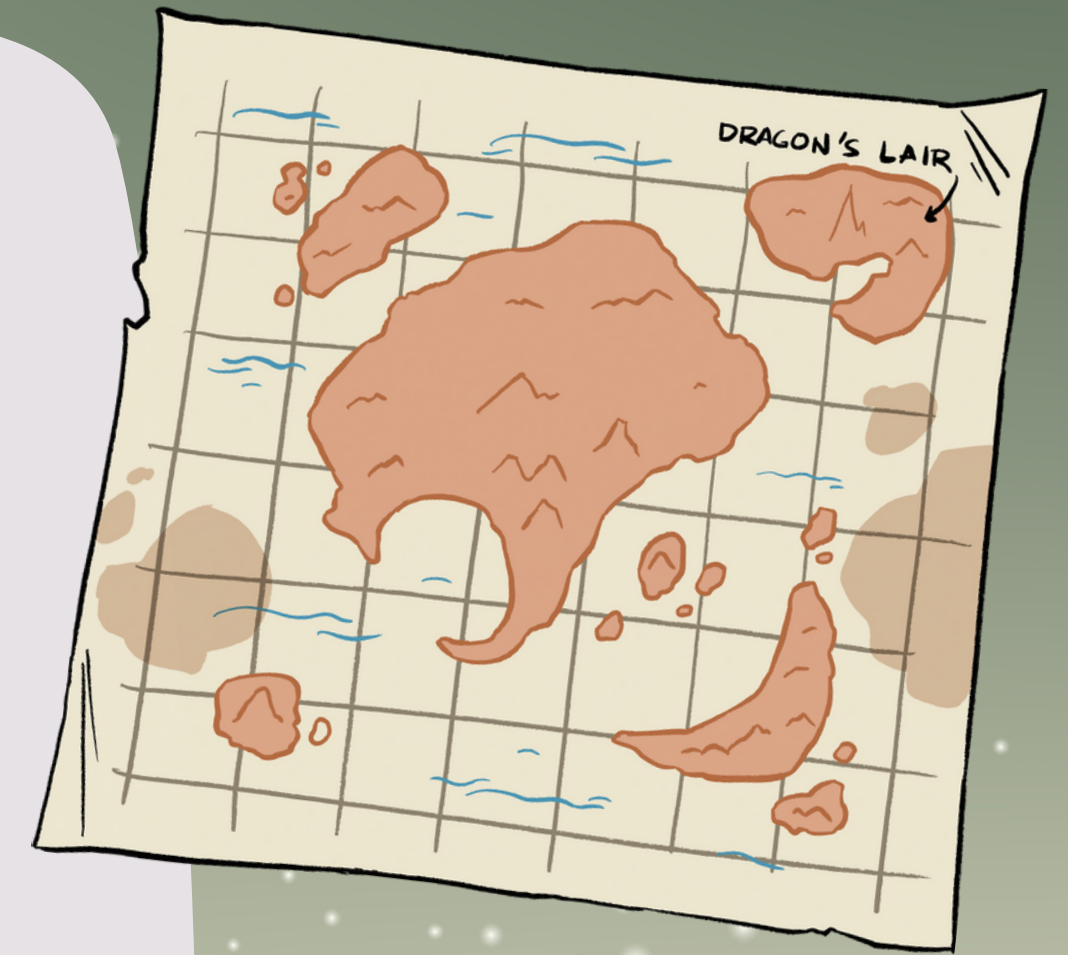
What should I do to grow
my portfolio?

I need help coming up with
an idea for my thesis!



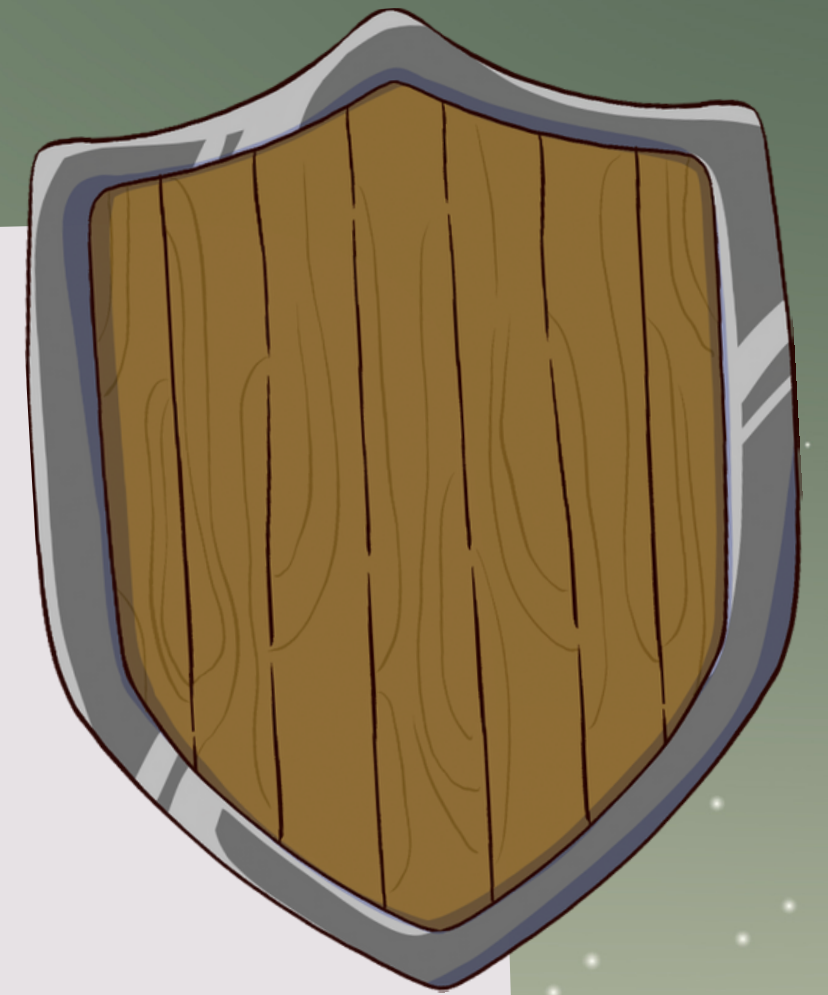
Write a survey paper

- These are overviews of what other people have done for a specific task
- It's a great way to understand where the field is now & where other work might fit in
- Bonus: they make good blog posts!



Do a system(s) evaluation

- Gives you a chance to "kick the tires" on a system or two
- Sure, it work great on a specific leader board... but what about a totally new dataset? (Make sure to check for data leakage!)
- Bonus: can help you work through deployments



Make an open source contribution

- Check for open issues on a repository
- In general, smaller and more specialized packages or tools will be easier to get started with (so probably don't try to start with numpy)
- Bonus: learn about source control & best practices (like linters)



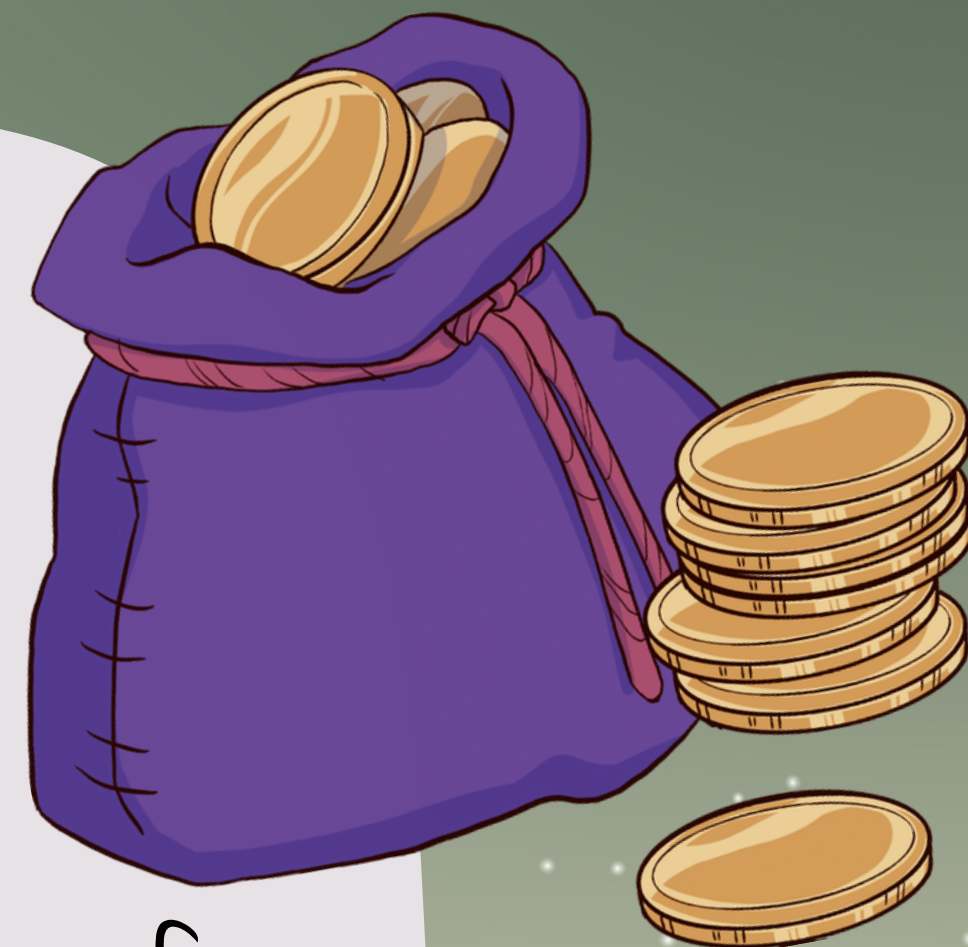
Replicate a research paper

- NLP has a fairly strong tradition of sharing code & open science
- Try looking for Github links in papers
- Check out Papers With Code or Distill
- Bonus: practice working with & understanding other people's code



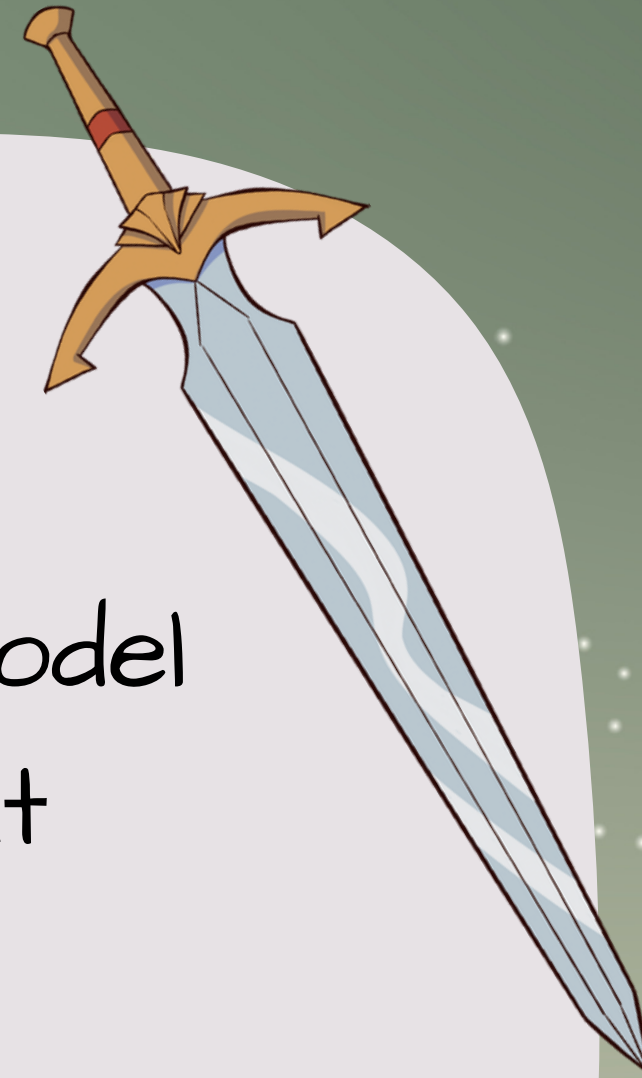
Add annotations to a dataset

- A pretty common need
- If you're feeling ambitious you can add an entirely new set of types of annotations
- You can also try out data labeling as a user of something like Zooniverse
- Bonus: data labelling is **invaluable** experience; I think everyone in ML should spend at least some time doing it



Do an ablation study

- Basically you remove bits of a model or make them smaller and see what happens
- Can also be a research project/paper
- Bonus: practice with experiment tracking/working with lots of models



Extend a research paper

- You can check "future directions" sections for ideas (or ask the authors)
- As you read more papers, you'll probably get more "but what about..." ideas
- Bonus: we're firmly into the "this could be a publishable research project" territory here



Collect a new dataset

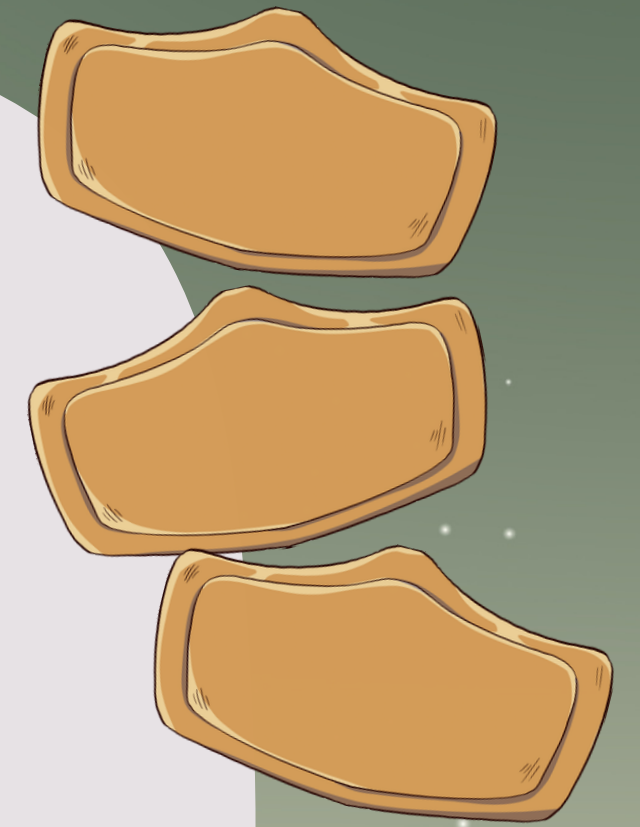
- Be warned: this can be **extrememly** time intensive
- You'll get the irreplicable experience of working with raw data
- Bonus: you can build other projects off the dataset you collected



Make something you wish existed

- This is sort of the point of learning NLP in the first place: building useful things using language technology
- A concrete idea of what you want will help motivate you to stick with it!
- Bonus: you'll have made the thing you wanted!





Recreate an NLP project in a new (human) language

- This can be pretty involved: most languages don't have as many existing NLP tools as English does
- I'd especially recommend working on increasing access to language technology in your own languages
- Bonus: You'll may have a lot of other folks who are interested in using the tooling you've built!