

# October 11 Meeting Summary

---

- **Goal:** predict perceived linkages between conditions and drugs
- **Core problem:** processing unstructured text so that it could be mapped into structured ontologies
  - Ontologies are freely available, but we need mappings to them from colloquialisms
- We have a manually annotated dataset called Inspire. We should get embeddings using Word2Vec/GloVe, train a model on the annotated data, and then test on the NPR dataset.
- Next meeting expectation:
  - Download data
  - Train embeddings on PubMed abstracts/PubMed Central papers
  - Check percentage of drug names in the NPR dataset
  - Check distances between synonyms, between unrelated terms, etc.