HW 2 Rubric

1. Running program and output files **(60 pts)**

- **1.1** Script downloads papers (5 pts)
- **1.2** Script generates authors, affiliation, etc. (15 pts)
- **1.3** GPT-2 implementation (25 pts)
- **1.4** Script generates papers' PDFs (5 pts)
- **1.5** 500 papers' PDFs (5 pts)
- **1.6** new TSV (5 pts)

2. Readme **(5 pts)**

3. Report **(40 pts)**

- **3.1** Structure **(5pts)**
- **3.2** Satisfactory answers to the following questions **(30pts)**
  - **3.2.1** What did the GPT-2 generated texts look like?
  - **3.2.2** Were they believable? And why?
  - **3.2.3** Would your associated ancillary features from assignment 1 have been able to discern what was false or not?
- **3.3** Summary **(5pts)**

4. Generating PDF using Latex **(Extra 20 pts)**

5. In the report, thinking more broadly, answer the following questions:

5.1 How much do you think media falsification is solvable using ancillary metadata features, or using actual content based techniques? Is one better than the other? **(Extra 5 pts)**
5.2 What other types of datasets could have been used to generate the falsified papers? Pick at least 2 datasets from distinct MIME types. **(Extra 5 pts)**
5.3 What other sorts of "backstopping" would be required to generate a believable paper trail for the scientific literature? **(Extra 5 pts)**