**Writing notes for Master’s thesis**

Outline:

1. Introduction
   1. What are common misinterpretations
   2. Why
2. Methods
   1. Sub-Steps
      1. Statement Detection
         1. RegEx and Papercheck
      2. Statement Classification
         1. NLP, BERT
      3. Statement Correction
         1. Groq and LLMs
   2. Sample
      1. Descriptives
   3. Evaluation with Human labels
      1. Where what was evaluated by a human and how it will be compared with the automated approach
3. Results
   1. Detection Reliability
      1. How many of the human identified statements with nonsig. p values does the automated approach get and how many false positives are there?
   2. Classification performance
      1. How well did the model predict correct labels of 0 or 1
   3. Correction performance
      1. How many of the (model predicted) incorrect statements that were corrected by the LLM were now actually correct, incorrect or otherwise problematic (as in it’s not useful as writing assistance), as assessed by a human expert?
4. Discussion
   1. Summary
   2. What worked well
   3. Why this approach is promising
   4. What did not work well
   5. What would need to be changed/improved & limits of this approach
   6. Conclusion