CS 410 Project: Spanish-English Machine Translation of AI Research

Team Members and Captain:

Name: Felipe Arias
NetID: felipea2
Role: Captain and sole team member

So far, I've completed preliminary research, designed the code structure, and implemented a basic working example. The code is divided into three submodules, each with an initial functional solution:

- PDF Parsing: Utilizing existing libraries, I'm parsing the introduction of a NeurIPS research paper. This is challenging, as it involves parsing section/subsection titles, images, equations, and addressing incorrect parsings.
- Translation: I'm currently using a Google API wrapper for translations. While successful, the translation quality could be improved. Exploring other APIs and enhancing parsing might yield better results if time permits. I will also begin looking into enforcing specific translations for technical terminology.
- Markdown Generation: The goal is to create markdown files for parsed and translated content. The necessary code is in place, but it requires integration with PDF parsing to maintain the original structure of the PDFs.

1. Which tasks have been completed?
    a. The completed tasks are described above
2. Which tasks are pending?
    a. The pending tasks are described above, but the remaining tasks are parsing the entire paper rather than only the introduction, improving the parsing quality, coupling the parsing and markdown generation to retain the paper structure, enforcing specific translations for technical terms, and improving the quality.
3. Are you facing any challenges?
    a. While I am confident that I will have a working parser/translator for the final project, I am not certain of how good the translation will be.