**PROJECT OBJECTIVES**

We present an automated text-mining tool written in Python to measure the technical responsibility of students in computer science courses.

- Our tool automatically collects reflection documents written by students from their GitHub repositories.
- Then, using natural language processing analyzes them for ethical considerations based on pre-determined questions and criteria.
- The tool helps to track the progression of student ethical understanding and sense of social responsibility by analyzing writing samples across the computer science curriculum.

**TEACHING RESPONSIBLE COMPUTING**

Teaching responsible computing is critical in developing software that produces a positive impact on our society, economy, and individuals.

- Each application course in computer science at Allegheny College integrates ethical considerations in its pedagogy.
- Broad learning categories include topics of internet health, ethics and responsible computing customized to each application course.
- The delivery of these concepts include readings, discussions, class and lab assignments with heavy software development emphasis.
- As an output, students write reflection reports to demonstrate their understanding of relevant issues, ability to analyze information, and capacity for integrating the understanding and analysis of ethical thinking into their own work.

**TEXT MINING TOOL TO DETERMINE ETHICAL PEDAGOGY**

Our tool first obtains student reflection documents (as Markdown files) stored in AWS.

- Markdown parser goes through the Markdown files and constructs a dictionary.
- Natural language pre-processing is done with SpaCy with the output stored into a pandas data frame for further analysis.
- Five categories of analysis are included that can be queried and customized. The result of each analysis is stored in a separate pandas data frame.

**FEATURES**

- Our tool can be run through a command-line or a graphical interface.
- Visualization was developed using Altair, with the generated graphs displayed using Streamlit.

**SAMPLE RESULTS**

- **Word Frequency Analysis**

  - Overall most frequent words in the directory

  - Figure: 4 Word Frequency Analysis

- **Document Similarity**

  - Similarity between each student’s document

  - Figure: 5 Document Similarity

---

![Figure: 1. This project is supported by the Responsible Computer Science Challenge, funded by Omidyar Network, Mozilla, Schmidt Futures and Craig Newmark Philanthropies.](image_url)