



# Thesis Defense

---

Computer Science Master's Program

## “Improving Automatic Transcription using Natural Language Processing”

By Anna Kiefer

### **Abstract:**

Digital Democracy is a CalMatters and California Polytechnic State University initiative to promote transparency in state government by increasing access to the California legislature. One foundational step of the project is obtaining accurate, timely transcripts of California Senate and Assembly legislative hearings. The Transcription Tool, a software application built by Cal Poly, allows human editors the ability to view and edit legislative hearing transcripts and associated information. This human upleveling process is the most costly and time-consuming step of the Digital Democracy pipeline. This thesis presents contributions to improve the Transcription Tool and decrease editor annotation time using Natural Language Processing systems and techniques. The primary features integrated into the tool include a new automatic speech recognition software, speaker diarization model, speaker identification process, and text post-processing and data changes. We quantitatively evaluate the accuracy and runtime of the new features before and after integration into the Transcription Tool. We compare the average editor duration for three versions of the tool, determining a transcriber efficiency improvement of over 40% from the baseline to final version of the tool.

**Date: Friday, March 22<sup>nd</sup>, 2024**

**Time: 3:00 PM – 5:00 PM**

**Location: 14-238b**

**Committee: Dr. Khosmood, Dr. Dekhtyar, and Dr. Kurfess**

