**Team Members: Quinn Wilson, Kenan Carames, Meareg Hailemariam, Sarah Albert**

**Project Title: Supreme Court Decisions Predicted from Oral Argument Transcripts using Deep Neural Networks**

**Project Description:** The Supreme Court serves final determination on the Constitutionality of laws and legislation in the United States. For each case brought before the Court, oral arguments are carried out for about 30 minutes per side to give an overview of the case, their arguments, and to answer any questions the judges may have. The arguments allow more open dialogue between lawyers and judges before the court sessions begin and often draw lots of attention from the public. It is common for legal experts to offer predictions on how they think the case will be decided after these opening presentations. A study found experts to be about 66% accurate. (<https://www.jstor.org/stable/4099370>) We are interested in using transcripts from oral arguments to build a model capable of making predictions better than legal experts. We are curious if the oral arguments alone are capable of providing powerful insights about the highest court of law before decisions are brought. To this end, we propose using argument transcripts to predict case outcomes either in favor of or against the Petitioners with the use of deep neural networks.

**Dataset**: Our data is collected from the Supreme Court transcript archive which consists of [oral arguments](https://www.supremecourt.gov/oral_arguments/argument_transcript/2022) and corresponding [outcomes](https://www.supremecourt.gov/case_documents.aspx)/decisions. Since we are planning to use a labeled dataset with the argument transcripts as predictors and decisions as outputs, we have to identify the relevant documents and outcomes so that we can link court orals with their corresponding outcome. Currently, the files are put in separate files and not structured in a way that would allow us to train supervised machine learning algorithms. We plan to match case transcripts to corresponding decisions by case docket number, and reference the binary outcome [as compiled by Washington University School of Law](http://scdb.wustl.edu/data.php). We are able to access cases dating back to 1970 with accuracy, encompassing on the order of 7,000 cases argued and decided.

**Deliverables**

| Phase 1 (until April 15th) | 1.1 to decide the number and years of court orals we will be working with  1.2 identifying, extracting and building labeled datasets  1.3 further preprocessing of the overall dataset  1.4 generating explanatory analysis from the dataset |
| --- | --- |
| Phase 2 (Until May 3rd) | 2.1 build deep neural networks based court oral outcome models  2.2. evaluate the performance and finetune models |

**Performance Measurement:** We will evaluate our model based on its probability of an accurate prediction. Since there are two possible outcomes for the cases, the first goal would be to achieve a better-than-guessing accuracy (>0.5). We are not sure if oral arguments will give enough information for such accuracy but this is the goal. Legal experts have been [found](https://www.science.org/content/article/artificial-intelligence-prevails-predicting-supreme-court-decisions) to have about 66% accuracy while more robust ML models using historical justice data have been able to predict the future with up to 75% accuracy. If we could get close to those numbers by just using the oral arguments, it would be quite a powerful model and this project would be a large success.