# HUMAN VS AI DISTINGUISHMENT

Final Team Project Group 2
AAI-590: Capstone Project
University of San Diego
Applied Artificial Intelligence Program

Jeremy Cryer, Jason Raimondi, and Shane Schipper

April 15, 2024

# PRESENTATION OVERVIEW

- Problem Statement
- Live Web App Demo
- Datasets and Prep
- Methodology Approaches
- Training and Evaluation
- Selection and Results
- Production Readiness

# PROBLEM STATEMENT

**Problem Statement:** Advancements in Artificial Intelligence (AI) is making it increasingly difficult to distinguish text as being human or AI-generated

**Areas of Concern:**

- News Feeds
- Academic Integrity

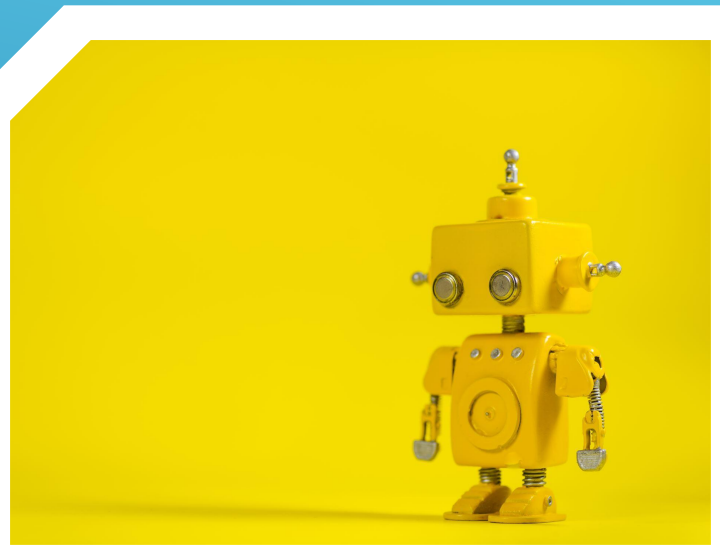
---

**Primary Goal**: Develop a Machine Learning (ML) model that can predict and provide the probability of text being human or AI-generated

**Secondary Goal**: Develop an interactive web app for users to interact with the model

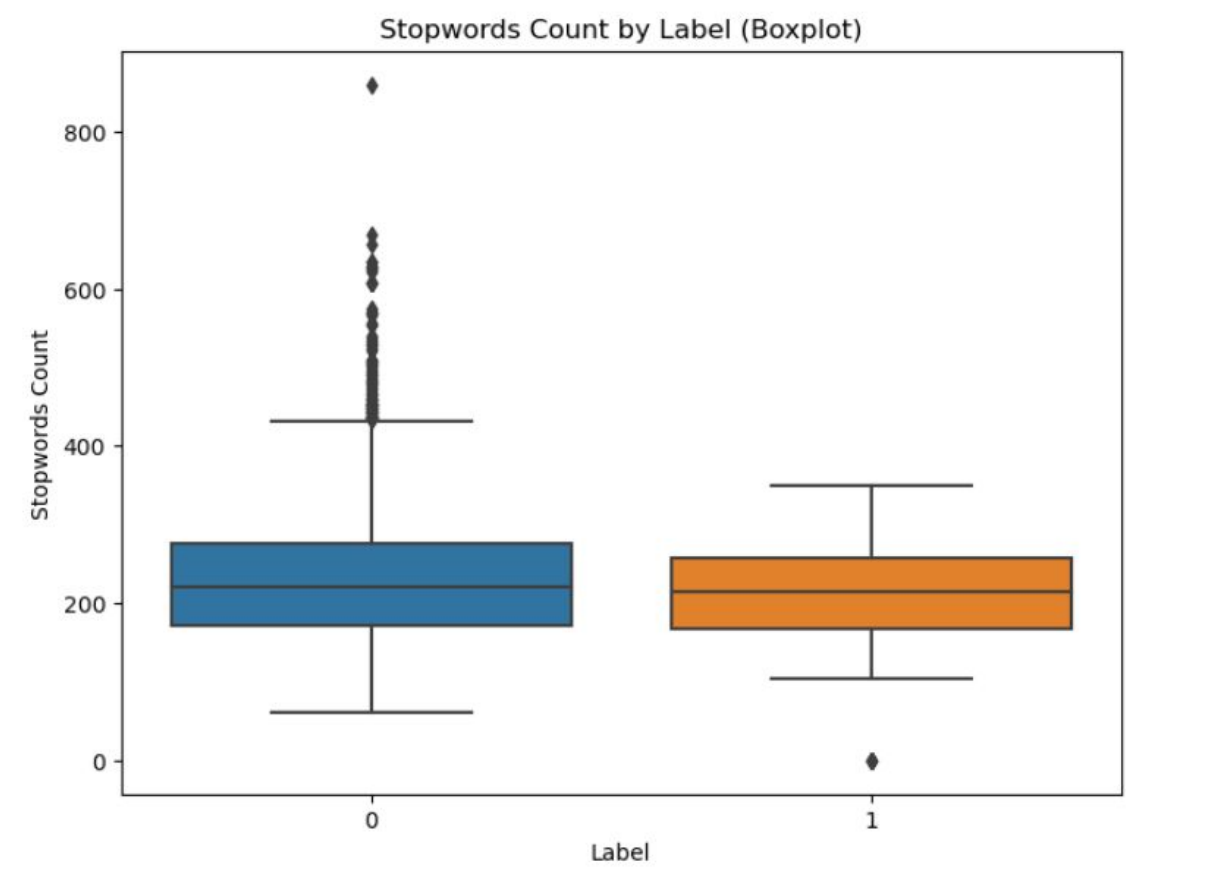
**Concepts:**

- Traditional ML Algorithms
- Transformer-based Models
- Containerization




# LIVE WEB APP DEMO

# DATASETS AND PREP




Datasets

- Training
  - AI vs Human Text
  - Nearly 500,000 Samples
- Inference
  - AI vs Student Text
  - 1,103 Samples

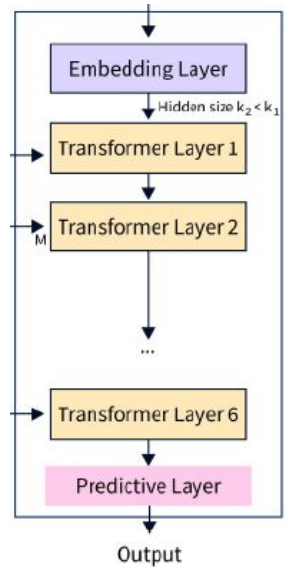
Prep

- Minimal Preprocessing
  - Remove samples missing data
  - Downsample majority class
- Tokenization, including:
  - Words/Subwords
  - Stop Words
  - Capitalization
  - Punctuation

# METHODOLOGY APPROACHES

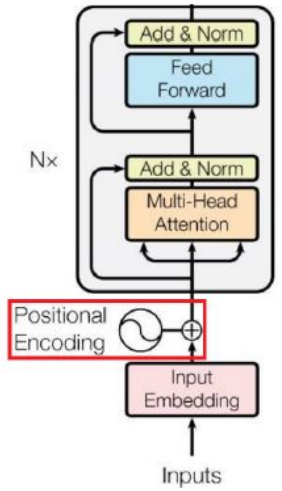
## Pretrained Transformer

- DistilBERT
- Proven Solution
- Low Development



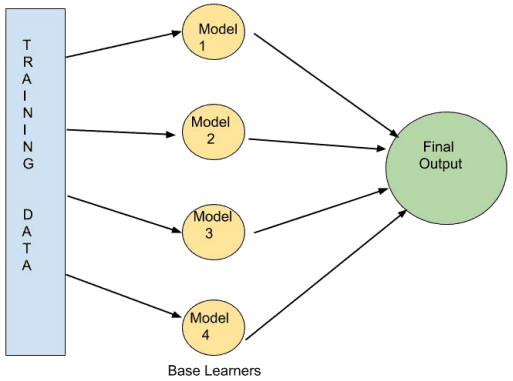

## Custom Transformer

- PyTorch Framework
- Flexible Configuration
- High Development



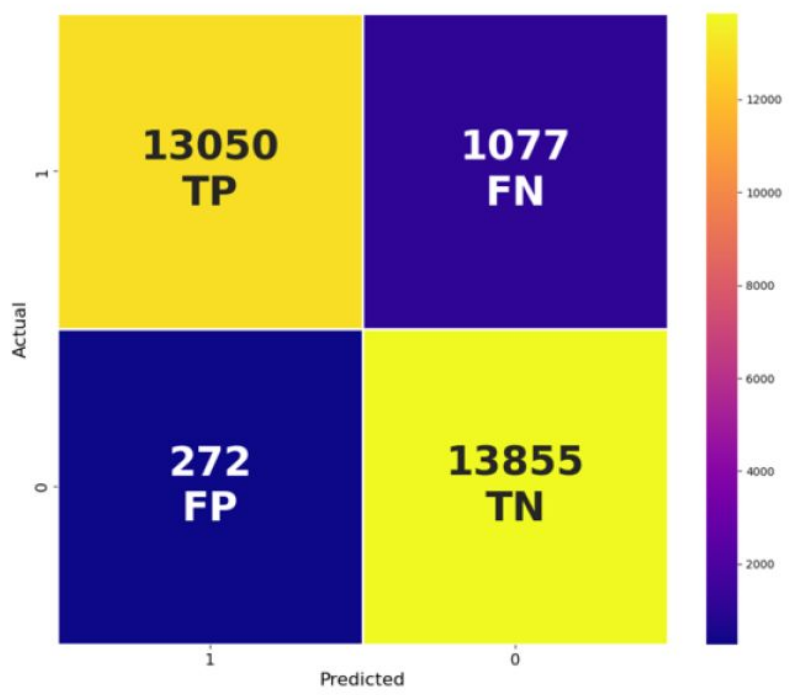

## Traditional Algorithm

- Random Forest
- Lower Complexity
- Moderate Development



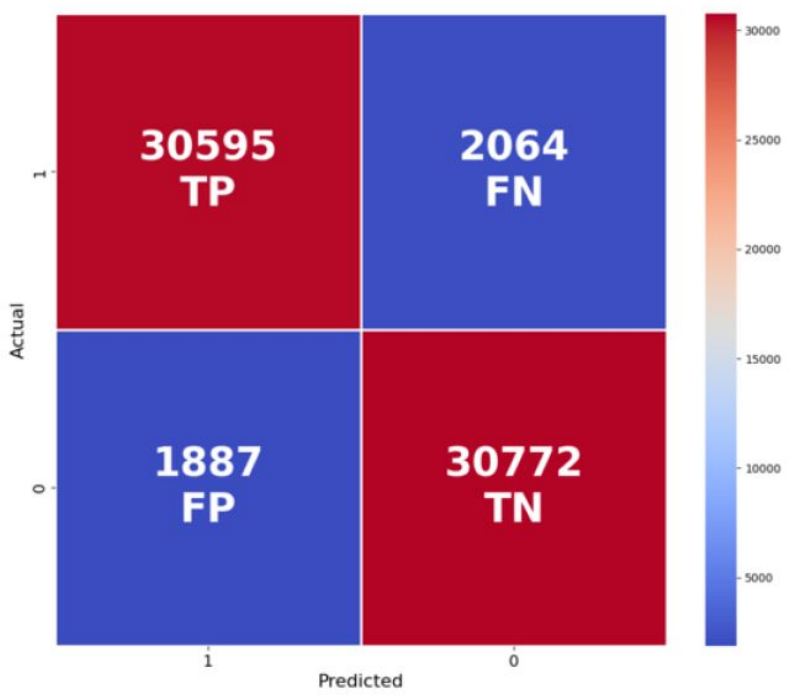
# TRAINING AND EVALUATION

## Pre-trained DistilBERT



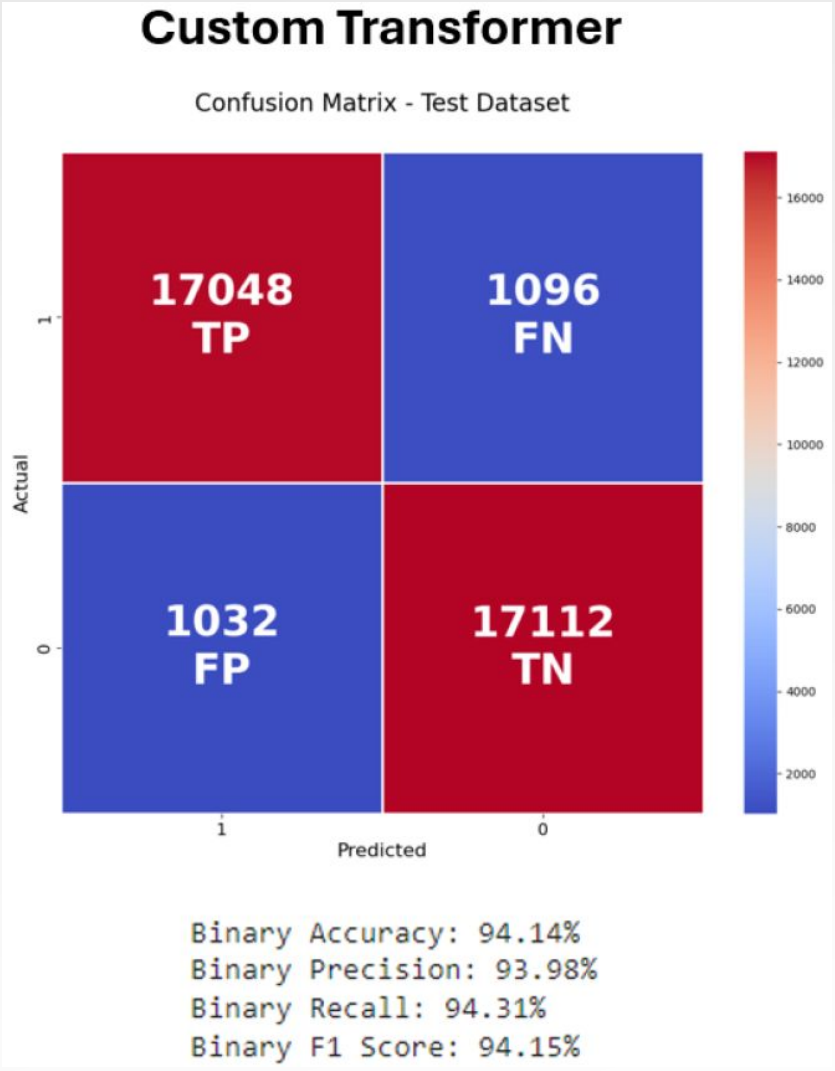

Binary Accuracy: 95.23%
Binary Precision: 92.79%
Binary Recall: 98.07%
Binary F1 Score: 95.36%

## Custom Transformer




Binary Accuracy: 93.95%
Binary Precision: 93.71%
Binary Recall: 94.22%
Binary F1 Score: 93.97%

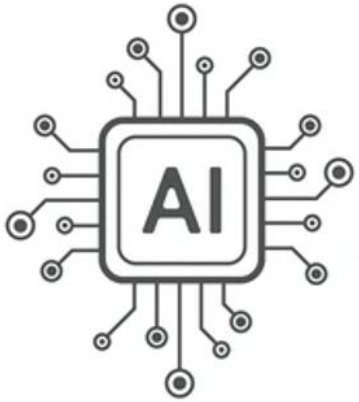
# SELECTION AND RESULTS



# PRODUCTION READINESS

## Model Robustness

- Increase training data size to > 1M samples
- Increase text diversity
  - Variable length
  - Writing styles
- Ensure AI source diversity
  - ChatGPT
  - Bard
  - Others




## Compute Resources

- Utilize High-Performance Computing (HPC) resources
- Increase model training duration
  - High memory runtimes
  - Additional GPUs
- Increase web app scalability

# THANK YOU!

Please contact us with any questions.

Jeremy Cryer
jcryer@sandiego.edu

Jason Raimondi
jraimondi@sandiego.edu

Shane Schipper
shaneschipper@sandiego.edu