

# Comparative Analysis on Traditional and Emerging Text Classification Methods

Presented by: Connor Morin, Annie Pi, Christina Yi, Jiangyuan Yuan, Jonathan Zhao

Department of Computer Science, University of North Carolina at Chapel Hill

April 22, 2025



# Background

## Text Classification

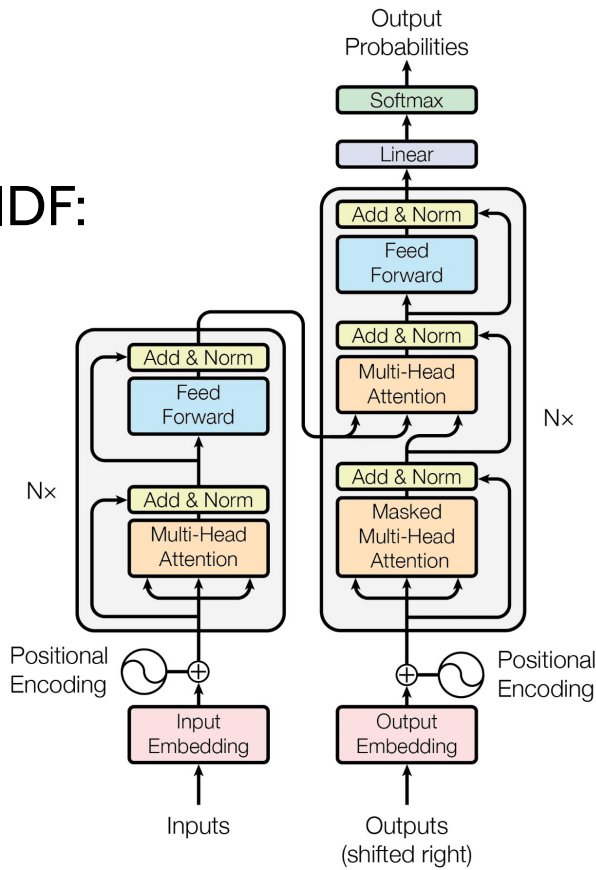
- Assigns labels (e.g., topic, sentiment, spam) to free-form text,

## Datasets

- IMDb Movie Reviews: User-written film reviews from the IMDb website, each annotated as positive or negative
- 20 Newsgroups: A curated collection of Usenet discussion posts drawn from 20 distinct newsgroups (e.g., computing, sports, science, religion)

# Approach

- Traditional Methods using TF-IDF:
  - Logistic Regression
  - SVM
  - Decision Tree
  - Random Forest
- Deep Learning
  - RNN (LSTM)
  - Transformer (BERT)
- LLM (GPT-4o)



# Results

