In our project titled "Breaking the Echo Chamber: Analyzing Political Bias in News Media," my colleagues Yash Rathi, Seshadri Kakani, and I from the Department of Computer Science at Virginia Tech have developed a system that addresses the critical issue of political bias in news media. We designed a novel tool that not only classifies political bias within news content but also provides detailed explanations for the detected biases using Large Language Models (LLMs).

The core of our system leverages the generalization and emergent abilities of LLMs to analyze and generate bias explanations from a labeled political bias dataset – BIGNEWSALGN, which has clustered news reporting of the same story from different biases. This capability is crucial in today's environment, where information overload can easily trap individuals within echo chambers, reinforcing existing opinions and dismissing opposing views as misguided or delusional.

Our system is structured to offer several interactive modalities for users: they can input queries, submit entire article texts, or provide URLs from specific news sources. These inputs are processed using various information retrieval techniques, including TF-IDF vectorization and web scraping, to fetch and classify news content effectively. Our approach not only classifies the political bias as left, lean left, center, lean right, right but goes a step further to articulate the reasoning behind each classification, thereby fostering a greater understanding of the nature of biases in news reporting.

Technically, our project involved fine-tuning task-specific LLMs using preprocessed BIGNEWSALGN dataset based using Alpaca instruction format. Despite the computational barriers of fine-tuning LLMs, our models achieved robust performance benchmarks for the political bias classification task. This underlines the effectiveness of LLMs in handling complex tasks like political bias detection. We also quantize the model to make them more accessible. Our best model Gemma 7b finetuned achieved 54.50% accuracy for 5 bias label classification and 67.50% for 3 label bias classification i.e. left, center, right.

For future work, we hope to further improve the efficiency of our models and find a way to evaluate the generated political bias analysis.