Applied AI Project Proposal

Group Members: Chetan Thakur(110213868), Krish Thakkar(110204274), Hiten Patil(110210987), Krishal Modi(110206943)
Project Title: AI-Powered News Personalization and Bias detection
Theme: ☑ **Recent News-Based AI** ☐ AI for Education

1. Introduction

Project Overview:

Our initiative is a news aggregation platform that will be an AI-based solution that offers the user tailored content, and also provides a clear way to quantify political bias. The platform will examine the articles of political news, rate them out of 0 to 100 their liberal or conservative orientation, and use the analysis to tailor a balanced and customized news feed to the user. We are also going to make use of natural language processing to create brief summaries of the articles.

Background:
The contemporary media environment is both highly fragmented and partisan in its nature, hence consumers might struggle to find objective news. The biased content is hiddenly received by many people and strengthens the beliefs that people already have, and leads to polarisation in politics. Artificial intelligence, specifically large language models (LLMs), is a potent resource to examine textual data to detect implicit biases and offer a more transparent and informative experience with news.

Motivation:

The current project is a response to the real-life issue of media bias and the consequences of this development on the discourse. We hope to enable individuals to be more critical although the information they receive by developing a tool that detects and measures political bias in news stories. We are inspired by such work as USA Today in their article titled DeepDive, but where surfacing and quantifying political bias is involved to create a more informed and well-rounded view of what is happening around us.

Objectives:

1. Create a web-based news app based on React on the front-end and Node.js on the back-end.
2. Train a large language model (LLM) on political news articles and use it to determine a bias score between 0 (becoming highly liberal) and 100 (becoming highly conservative).
3. Use a personalization algorithm that suggests news items to the user based on their

reading history whilst leaving the user to regulate the exposure to various sides of political thought.
4. Apply NLP to create summaries of each news item: short and neutral.
5. Create a user-friendly interface that easily represents the political bias of every source and article.

2. Literature Review

Prior Work:
A number of research papers have investigated how computational techniques can be used to detect bias within the news media. As an example, Baly et al. (2020) proposed the use of a multi-task learning model to estimate hyperpartisanship and news article bias, which was highly accurate and was based on the analysis of rhetoric and word choice. On the same note, Spinde et al. (2021) developed a dataset and a model of fine-grained bias detection, whereby particular sentences are analyzed, which contain bias. Moreover, commercial solutions such as Ad Fontes Media and AllSides have developed ways to rate news sources on a political scale, but tend to use human reviewers instead of scalable AI. Such projects illustrate that the detection of bias is possible but it is also a complex issue.

Identified Gaps:

Although current studies have been able to detect hyperpartisan sources efficiently, less attention is paid to it offering the user a dynamic, article-level bias rating in a real-time, personalized interface. A lot of the available solutions provide a static rating of a complete news source, which fails to consider differences at the article level. We will fill this gap in our project with the help of an advanced LLM, which will be able to analyze articles on a more granular scale, generating an article-by-article news aggregator and an educational tool at the same time, a media literacy one**.**

3. Methodology

- **AI Approach / Techniques: :** We shall mostly rely on Natural Language Processing (NLP) with a state-of-the-art Large Language Model (LLM), via an API (such as the GPT series of OpenAI). The LLM will be asked to do **two major tasks**:

  Analysis of Bias: Evaluate the text of an article to establish a political bias on a liberal-conservative spectrum and produce a score out of 0-100.

  Summarization: Provide a hypothesis, abstractive summary, of the main points of the article.

- **Data Collection:** We shall pool news articles of a wide variety of popular online sources

of political news, representing the political spectrum (e.g., CNN, Fox News, Associated Press, Reuters, MSNBC, The Blaze). It will be achieved through news aggregation APIs or creation of web scrapers. We shall also be conscious of copyright, terms of service of each source and not save the full article text long term, but rather process it and save our analysis.

- **Model Development:** A well-designed prompt system towards the external LLM will become our "model" We will create a prompt that will elaborate on what to do to make the LLM sound as an unbiased political analyst and rate articles on criteria such as word choice, framing and topic focus. The prompt will be optimised by trial and error.

- **Evaluation Metrics**: Our bias-scoring mechanism will be assessed by comparing its results to known media bias charts (such as the Ad Fontes Media Bias Chart) and also by making our own judgment on a test collection of articles.

Experiments and Evaluation:

We will conduct experiments to validate the consistency and accuracy of our bias scores. This will involve feeding the system articles from sources with known biases and checking if the scores align with expectations. We will also test the system on more neutral sources like the Associated Press to ensure it can recognize and appropriately score less-biased content. User testing will be conducted to evaluate the usability of the platform and the clarity of the bias visualization.

4. Resources

Personnel:

- **Chetan Thakur:** Backend Development (Node.js, API integration, data pipeline).
- **Krish Thakkar:** Frontend Development and ranking the articles on NLP basis (React, UI/UX design, data visualization).
- **Hiten Patil:** AI/LLM Prompt Engineering and Evaluation (Developing and refining prompts, testing model outputs).
- **Krishal Modi:** Data Collection and Prompting setup (Setting up news aggregation, timeline management, documentation).

Hardware / Software:

- **Frontend:** React.js, Tailwind CSS
- **Backend:** Node.js, Express.js
- **AI & NLP:** OpenAI API (or similar LLM provider)

- **Development Tools:** VS Code, Git/GitHub, Postman
- **Deployment:** Vercel (for frontend), Heroku or AWS (for backend)

Data Requirements:

We will require continuous access to new articles from a variety of online news sources. This will be achieved through a news API subscription (e.g., **NewsAPI.org, GNews**) or custom web scrapers.

5. Expected Outcomes

The first and most obvious output of our project will be a **fully-functioning web application that is both a personalized news aggregator and a media bias analysis too**l. The users will have access to a c**urated feed of the news and will be provided with a political bias score of each of the articles as well as a concise and AI-generated summary**. The project will also lead to a final report of our methods, findings, and how successful it was to use an LLM to accomplish this. Its possible practical outcome would be to increase media literacy and equip readers with the means of critically evaluating the information they are taking in, hopefully leading to a more sophisticated and informed discussion in the public..

6. Timeline / Milestones

Milestone	Expected Completion	Notes
Data Collection	Week 1	Set up news API and/or scrapers.
Model Development	Week 3	Finalize LLM prompt and **scoring logic**.
Experiments & Evaluation	Week 5	Test and validate the bias scoring system.
Final Report & Presentation	Week 7	Complete documentation and presentation.

7. References

Baly, R., Da San Martino, G., Glass, J., & Nakov, P. (2020). We Can Detect Your Bias: Predicting the Political Ideology of News Articles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 544-555). Association for

Computational Linguistics.

Spinde, T., Vafa, K., Jetter, F., & Staab, S. (2021). B-SHOT: A Corpus for Bias-aware Headline Generation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume* (pp. 3060-3071). Association for Computational Linguistics.

AllSides. (n.d.). *Media Bias Ratings*. Retrieved from https://www.allsides.com/media-bias/ratings

Ad Fontes Media. (n.d.). *The Media Bias Chart*. Retrieved from https://adfontesmedia.com/