



Bilkent University

Department of Computer Engineering

Senior Design Project

Impartial - Group T2324

Project Specification Report

Group Members

Nisa Yılmaz - 22001849

Mustafa Cem Gülümser - 22003430

Murat Güney Kemal - 22002692

Aslı Karaman - 21901576

Eren Duran - 22003932

Supervisor

Halil Altay Güvenir

1. Introduction	3
1.1. Description	3
1.2. High Level System Architecture & Components of Proposed Solution	4
1.3. Constraints	5
1.3.1. Implementation Constraints	5
1.3.2. Economic Constraints	6
1.3.3. Ethical Constraints	6
1.4. Professional and Ethical Issues	6
2. Design Requirements	7
2.1. Functional Requirements	7
2.2. Non-Functional Requirements	7
2.2.1. Usability	7
2.2.2. Security	7
2.2.3. Performance	7
2.2.4. Supportability	7
2.2.5. Scalability	8
3. Feasibility Discussions	8
3.1. Market & Competitive Analysis	8
3.2. Academic Analysis	8
4. References	10

1. Introduction

1.1. Description

Our project, Impartial, will initially be a web application that will obtain the news from several data sources and categorize them according to three main categories: left, right, and center. The primary purpose of Impartial is to provide a platform for the readers to perceive and analyze news from different perspectives. It will obtain the news from various resources, and without any need for editors for categorization, the news will be categorized based on left, right, and center classifications with integrated AI technologies. In addition to categorization, Impartial will be able to generate articles initially written in different categories but interpreted in other categories. Integrating AI and Generative AI technology into the application will obtain the article, perceive its category, and generate new articles based on the original article's content but with a different perspective. As a result, the generated article's content will be the same, but its perspective will differ from the original one. With these features, Impartial aims to provide a platform to users where they can be informed about different political perspectives on various political content and approach the political news with a broader perspective and richer understanding. It will combine different perspectives in one online newspaper platform by preventing biased and unfiltered content.

Initially, our target customers will be the US people who follow politics regularly. Our goal is to broaden the target customer and expand our customer base in the future phases of our development. In our application, we plan to design a user-friendly interface that will include all the application categorization and filtering features in one main effective and practical page. Also, a subscription option will be included in our application to serve better alignment and filtering options for the subscribed users. Currently, some online news platforms serve similarly to Impartial, but in these applications, the news is classified by the editors but not with an integrated technology [1]. In contrast, Impartial will classify the news into three main categories with the integrated AI technology and without any need for editors.

To sum up, we intend to develop an application named Impartial that will classify political news mainly into three categories: left, right, and center. Additionally, Impartial will generate articles whose perspective differs from the primary perspective. All of these features will operate based on the integrated AI technologies.

1.2. High Level System Architecture & Components of Proposed Solution

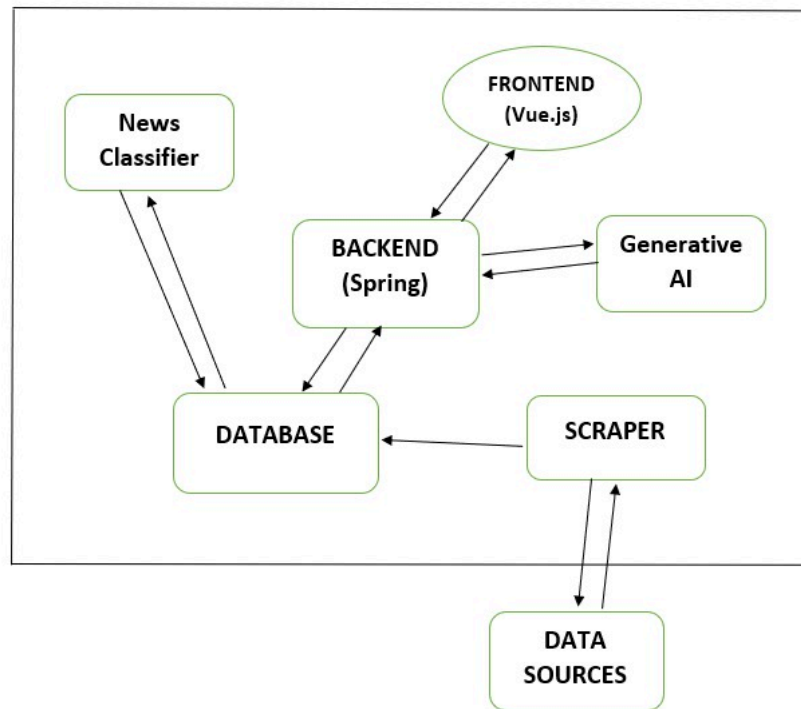


Figure 1: General system setup of the project.

The high-level system architecture of our project will mainly consist of "News Classifier," "Database," "Scraper," "Data Sources," "Generative AI," "Backend," and "Frontend." In the backend development of our project, we plan to use "Spring," which is a Java Library, and for the frontend development of our project, we will use the "Vue.js" framework. The news will be obtained from various data sources, extracted with the scraper, and stored in our project's database. The news classifier of our architecture will receive the necessary data from our database. Later, the categorized news will also be stored in the database. Another feature of our project, which is generating articles with other categories, will be implemented with our project's Generative AI system architecture component

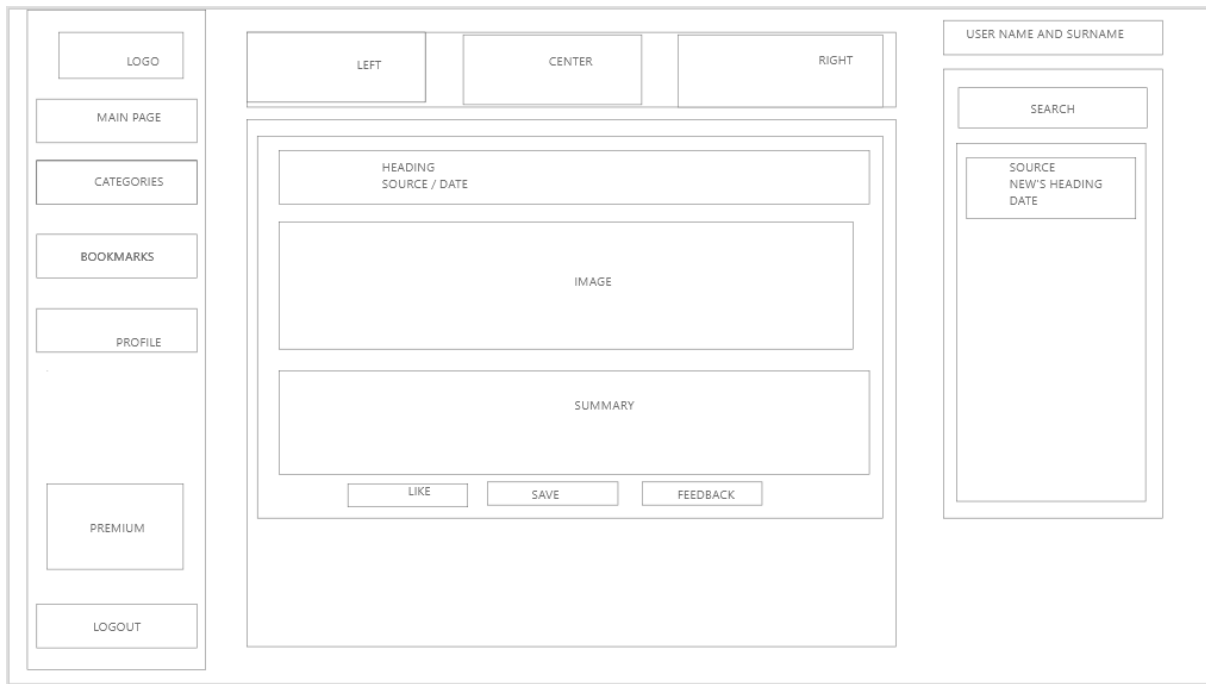


Figure 2: UI design first draft.

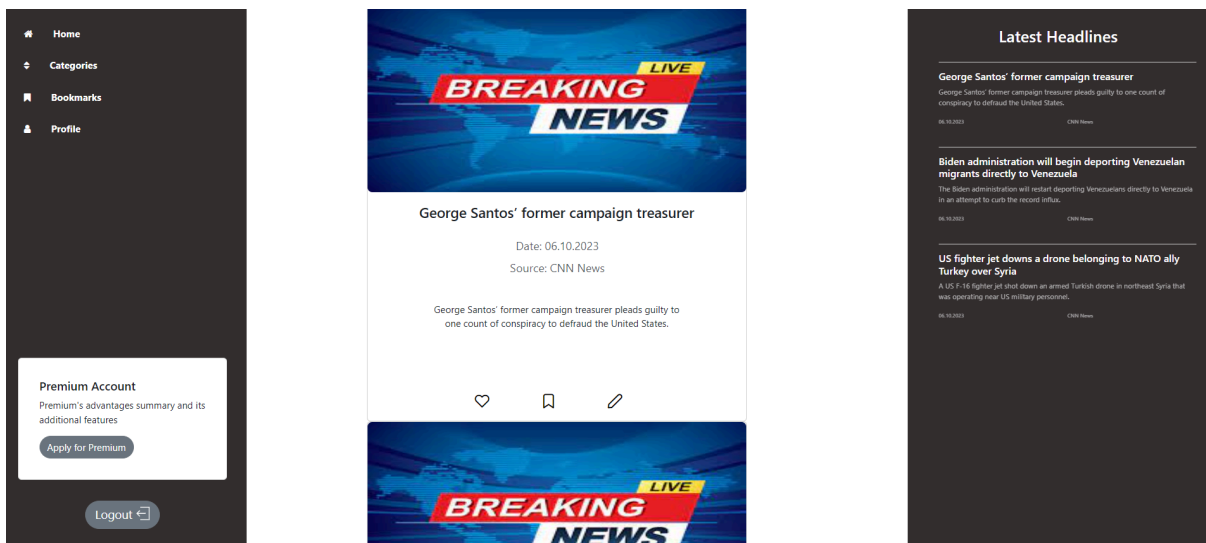


Figure 3: A prototype of the project UI.

1.3. Constraints

1.3.1. Implementation Constraints

- Github and Jira will be used to control the development process of our application.
- Vue.js will be used for the frontend development.
- Java and Spring framework will be used for the backend development.
- Pytorch and Huggingface transformers frameworks will be used to develop the AI models.
- FastAPI and Python will be used to deploy the AI models.
- PostgreSQL will be the database system.

- The backend will be deployed to AWS.
- Our machine learning models will be trained in the cloud using Google Collab.
- NewsAPI will be used to acquire new news articles [2].

1.3.2. Economic Constraints

- Backend deployment will be free using AWS EC2 free tier.
- Depending on our performance requirements, we might upgrade our Google Collab to a paid plan.
- All frameworks and libraries used in this project are free.
- NewsAPI is free in the development stage [2].

1.3.3. Ethical Constraints

- Any user data collected will be necessary and relevant to our system.
- The collected data will not be shared with 3rd parties without the users' explicit permission.

1.3.4. Sustainability Constraints

- The application and AI models will be maintained periodically.
- Any discovered bugs will be fixed in a week.

1.3.5. Language Constraints

- The language of the website will be English. More languages may be added in the future.

1.4. Professional and Ethical Issues

1.4.1. Professional Issues

- The project's repository will be private.
- Weekly meetings with our supervisor will be conducted to track the progress of the project.
- The project will be developed in alignment with Agile methodology.
- The task will be distributed equally.
- Each member will keep logs about the work or research done for the project.

1.4.2. Ethical Issues

- Only necessary user data will be collected.
- The collected data will be protected and will not be shared without the users' permission.

2. Design Requirements

2.1. Functional Requirements

- Users will be able to register/log in using their email and password.
- Users will be able to view news articles regardless of signing in.
- Users will be able to filter news articles based on their political alignment.
- Users will be able to save news articles.
- Users will be able to upvote/downvote the accuracy of classification of news articles.
- Admins will be able to add/edit/remove news manually.
- Admins will be able to perform a web scrape in case of breaking news in addition to the automated web scrapes performed hourly.
- The system will be able to scrape news from different sources.
- The system will be able to parse and categorize scraped news and save them to the database.
- The system will be able to generate a news article from a given news article that is written in other perspectives.

2.2. Non-Functional Requirements

2.2.1. Usability

The app has a non-functional usability requirement because of the diverse and vast user base. The news is something that everybody on the planet sees, and if we want to succeed in this area, the app should be user-friendly. It should have a trivial user interface and instinctive functionality.

2.2.2. Security

Since we will be getting requests from the users, we will be handling user data, and that is a big problem with security. We will design our app with the security constraints in mind and use the appropriate encryption and safety practices throughout our journey.

2.2.3. Performance

Since we will use machine learning, generative AI, microservices, and such, the new content should be processed with acceptable speed, like a minute at the top, and the generative part should be comparable to the other alternatives in the area. For example, if it is vastly slower than chatGPT's response time our application won't be preferred to better alternatives.

2.2.4. Supportability

We are building the app for the web because we want to support every device possible and get an extensive range of coverage across platforms, and to achieve that, we will optimize how our web app looks for every platform.

2.2.5. Scalability

Since all the consumers for news are a considerable number, our system should scale accordingly because of our revolutionary idea. We are expecting at least 10% of the Reuters users, and that number is in the order of millions. To accommodate that many users and handle their requests for the generative part, we need to write optimized and scalable code from the bottom up.

3. Feasibility Discussions

3.1. Market & Competitive Analysis

The primary innovation our project brings to the market is the classification of news articles according to their political alignment using machine learning algorithms. Popular news webpages do not present the political alignment of their news articles explicitly, but rather avoid using an explicit political tone, and present their opinionated articles under ‘opinion’ columns. However, news articles published in such websites are seldom unbiased, or politically unoriented. Therefore, many news sources are voted as biased by the US public, either leaning towards the Democratic or Republican side [3].

There exists another webpage entitled ‘Allsides’ which presents news articles based on their political alignment [1]. The classification process of news articles are done manually by the editorial team of Allsides, who analyze news articles for bias and determine whether the news article is left, center, or right wing oriented. They do not use any AI/ML tools for the classification of news articles [1].

In terms of generating news articles on differing political alignments using a generative AI tool, there are no other applications that have such functionality. Based on the feedback of the US public news readers, political orientation is commonly seen on different news sources, and people tend to read the news source that fits their political outlook the most [3].

3.2. Academic Analysis

As stated previously, there is no research conducted on the generative AI side of our project. When it comes to the classification of news articles on political bias, there are some literature works that are worth mentioning.

- **Detecting Bias in News Articles using NLP Models:** This research benchmarks natural language processing (NLP) algorithms to understand the effectiveness of ML techniques for identifying biased political leanings in news sources. The study starts with a TensorFlow deep neural network (DNN) using bag of words (BOW), then incorporates term frequency-inverse document frequency (TF-IDF) for weighting. Unsupervised K-Means clustering is applied to analyze patterns among articles from different sources. Additionally, SimCSE, a contrastive learning framework, is used for sentence embeddings. The study concludes that contrastive learning is the most accurate NLP model for detecting political bias in news articles [4].
- **Detecting Media Bias in News Articles using Gaussian Bias Distributions:** This research focuses on detecting media bias, emphasizing the limitations of existing methods reliant on low-level lexical information, especially for news articles where words are used in new contexts. The study explores the impact of second-order information, such as the frequency,

positions, and sequential order of biased statements within an article, using a Gaussian Mixture Model. The findings indicate that considering the frequency and positions of biased statements significantly influences article-level bias detection, with the exact sequential order being of secondary importance. Empirical evidence suggests that incorporating second-order information significantly enhances the effectiveness of article-level bias detection compared to methods that do not consider such information [5].

4. References

- [1] *Balanced News via media bias ratings for an unbiased news perspective*. AllSides. (2022, October 20). <https://www.allsides.com/>
- [2] *News API*, News API Search News and Blog Articles on the Web, <https://newsapi.org/>.
- [3] Ralph, P. (2018). *These are the most and least biased news outlets in the US, according to Americans*. Business Insider.
<https://www.businessinsider.com/most-biased-news-outlets-in-america-cnn-fox-nytimes-2018-8>
- [4] Nadeem, M. U., & Raza, S. Detecting bias in news articles using NLP models - Stanford University.
https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1224/reports/custom_116661041.pdf
- [5] Wei-Fan Chen, Khalid Al Khatib, Benno Stein, and Henning Wachsmuth. 2020. *Detecting Media Bias in News Articles using Gaussian Bias Distributions*. In Findings of the Association for Computational Linguistics: EMNLP 2020, pages 4290–4300, Online. Association for Computational Linguistics.