Latent Probabilistic Model of News Sources

Project Alpha Prototype Report

Army Cyber Institute





**Machine Learning for Media Bias**

William Hiatt, Gabriel Matthew, and Deven Biehler

02/09/2023

**TABLE OF CONTENTS**

[I. Introduction 4](#_Toc127088412)

[I.1 Project Introduction 4](#_Toc127088413)

[II. Team Members - Bios and Project Roles 4](#_Toc127088414)

[III. Project Requirements 5](#_Toc127088415)

[IV. Solution Approach 5](#_Toc127088416)

[V. Test Plan 5](#_Toc127088417)

[VI. Alpha Prototype Description 5](#_Toc127088418)

[VI.1. [Subsystem Name] 6](#_Toc127088419)

[VI.1.1. Functions and Interfaces Implemented 6](#_Toc127088420)

[VI.1.2. Preliminary Tests 6](#_Toc127088421)

[VII. Alpha Prototype Demonstration 6](#_Toc127088422)

[VIII. Future Work 6](#_Toc127088423)

[IX. Glossary 6](#_Toc127088424)

[X. References 6](#_Toc127088425)

[XI. Appendices 7](#_Toc127088426)

# Introduction

This document serves as a guide to the Latent Probabilistic Model of News Sources. It will present new and updated details of the current state of the project as well as the direction the project is going. It will include information such as the social science aspects of the project, the open source models we will be using, models we will be creating, research that we will be using and research we will be doing internally, and any requirements that we currently have. Finally this document will quickly go over the bio’s of group members on the team that are working on developing the model as well as all the stakeholders that the project will effect.

# Project Introduction

The US Army Cyber Institute needs a model used for evaluating bias within individual news articles. This model will be used to give a bias rating on a specific news article. The model will grade the article on a scale based on the different bias’s that the model looks for. With this scale the Army will be able to fight information warfare more efficiently, both on US based news sites and foreign news sites. The end goal of this project is to create a new model for the Army that will rate individual news articles on their bias and misinformation. This rating will be displayed in a way that is easy to read and follow. The model will use a mixture of currently available software as well as software created by the team.

# Background and Related Work

Concerns about media bias and its potential impact on society have grown in recent years. Misinformation, polarization, and even discrimination can result from biased news. Journalists and media companies can become more aware of their own biases and work toward creating more objective and balanced news content by using a machine learning model that detects bias in news reports. Furthermore, news consumers can use such a model to critically evaluate the news they consume and make more informed decisions.

Due to large amounts of misleading information or complete misinformation, the US military has grown increasingly interested in fighting back. With their help, we are hoping to build a model to decrease the amount of misinformation and the speed at which it spreads.

There are some sites that currently provide a similar service such as allsides.com, mediabiasfactcheck.com, and adfontesmedia.com. These sites currently just look at the media outlet instead of breaking down individual articles. These outlets also tend to use humans to help determine this bias and that brings in another potential bias.

# Project Overview

The media has a huge duty in the form of distributing news to the American public. The information plays a huge role in how people vote. The current offerings in finding media bias aren’t as detailed as they should be and still contain a bias in themselves as it’s often a human who is assigning these bias’s. For example, allsides.com has CNN ranked as far left while adsfontesmedia.com has them as slightly left leaning. This is just a single example, but it shows that there is even a bias within deciding bias. In addition one post from a media outlet could contain no bias and be factually correct while another may contain a lot of bias and not be factually correct. The goal of the team is to combat this by creating a model to find the bias, without any hint of human bias, of a single news article.

The team, with the help of mentors at the Army Cyber Institute, will create a new model that will address the issues presented in current media bias tools. Through research, designing, building, and testing the team will deliver a media bias tool to help fight information warfare.

The project is being built from the ground up. There are no limitations on how we create the model. The team has decided to use Python to create the model. Python is extremely popular for machine learning and will be a great platform to use. The code will be housed on GitHub, this allows for easy version control and collaboration.

The team is also in contact with two Social Scientists at Army Cyber Institute to help in finding the most important bias’s to use as well as the best way to display the output. The team will be using sentiment analysis and fact selection as the primary drivers in deciding a bias. As the project progresses, we plan to include additional secondary drivers.

# Client and Stakeholder Identification and Preferences

Our client is the US Army Cyber Institute with Senior Research Scientist Iain Cruickshank as our mentor and primary contact for the project. The product will be used and maintained by the Army. There are several stakeholders within the Army Cyber Institute including Iain and his colleagues.

# Team Members - Bios and Project Roles

William Hiatt is a 4th year software engineering student at Washington State University. His skills include C#, Python, SQL, Java, C++, and GoLang. He has prior experience working as a software engineer intern at Kochava as well as a junior web developer at Washington State University. For this project, William will act as a team lead, main developer and designer of the model, and work with his team to create a viable machine learning model.

Gabriel Sams is a 4th year Computer Science major at Washington State University. His skills include C#, Python, Database development/management, SQL, machine learning, software development, agile process, test-driven development, and data science. His prior experience includes mobile application development, database development and management, and unit, end-to-end, and functional testing. For this project, Gabriel will act as a main developer and designer of the model, and work with his team to create a viable machine learning model.

Deven Biehler is a 4th year Computer Science major at Washington State University. His skills include C/C++, Python, SQL, machine learning, HTML/CSS, agile process, Haskell, and data science. For this project, Deven will act as a main developer and designer of the model, and work with his team to create a viable machine learning model.

# Project Requirements

Use this section to define the requirements for your project. Lift materials from the Project Requirements document as needed, notably sections II and III. This will orient your reader to the quantified requirements expected from your final design and ensure they are ready for the Solution Approach.

While you can copy directly from your current Requirements document, make sure to have an introduction paragraph stating what this section includes.

# Solution Approach

Lift materials from your solution approach document. This should demonstrate to the reader what your project shall include, at least according to the current design. Just like the Requirements section, ensure you look over the Solution Approach document and take what fits in this narrative and update details as needed.

# Test Plan

Bring in your test plan materials. These should demonstrate how you will validate your solution to determine whether it works and if it satisfies the requirements of the project. Look over the plan to ensure your requirements all have at least some kind of unit or functional test that addresses them.

# Alpha Prototype Description

\*\*\*This is the first section that is truly new for this document\*\*\*

Describe your alpha prototype implementation. Please format this section according to what you think is the best way to describe your prototype. The following is just a suggestion.

Explain what parts/subsystems of your proposed architecture your team started to implement. Mention the current progress at each part, i.e., how much of the proposed functionality you have completed.

For the parts that you already started to work on, explain whether you have tried to integrate them with the rest of system. In other words, explain which of the interfaces in your architecture have you started to implement.

If you have performed any tests on your alpha prototype (or the subsystems of your prototype), briefly explain your findings.

I recommend to include plenty of images and pictures of the following where appropriate:

- any diagrams/figures that visualize various features of your prototype;

- the screenshots of your user interfaces;

- the screenshots of your test programs;

- pictures of your team testing and debugging the devices, programs, etc.

A well-thought and clear diagram is better than long and descriptive text.

If your document starts to be very long due to screenshots and diagrams, please put at least some of them into an appendix to this document.

For each subsystem that you have implemented in your alpha prototype, you may include the following sub-sections.

## [Subsystem Name]

### Functions and Interfaces Implemented

List and describe the implemented functionality. Explain the remaining work.

### Preliminary Tests

Report any test results for the unit and integration tests that you performed on your prototype. This subsection is a good place to include screenshot images from your tests (if applicable). A notable component here would be to include the results of your CI/CD status. Hopefully master still builds, right?

# Alpha Prototype Demonstration

Summarize the highlight of your prototype demonstration to your mentor. The items to discuss in this section may include the following. (Please include all other necessary details in addition to the following).

1. Summary of what you showed to your mentor.
2. Your mentor’s comments/suggestions on your prototype.
3. Your mentor’s questions to your team and your responses to those questions.

After testing your prototype and demonstrating it to your mentor, you will have a better idea whether the initial design you proposed earlier will work. Additionally your mentor might suggest modifications to your current design. In this section list and explain all design modifications that you plan to make based on your preliminary test results and mentor comments (if applicable).

# Future Work

List the major tasks for the second semester and briefly explain your plan to complete them.

# Glossary

Define technical terms used in the document.

# References

Cite your references here. -- Ensure you’re pulling them from your earlier works!

For the papers you cite give the authors, the title of the article, the journal name, journal volume number, date of publication and inclusive page numbers. Giving only the URL for the journal is not appropriate.

For the websites, give the title, author (if applicable) and the website URL.

Please use either Chicago or IEEE format for your citations

# Appendices

As needed, copy over your appendices for the various sections. You can have as many appendices as required. Normally, they’re numbered with letters:  
Appendix A  
Appendix B  
…  
Appendix *n*