

CS486C – Senior Capstone Design in Computer Science

Project Title: Misinformation and Credible News Analysis Tools	
	Jacob Bailly, Chief Technology Officer NOBL Media jacob@wearenobl.com

Project Overview:

The internet provides incredible opportunities for people to communicate and share ideas, and has massively increased the public's easy access to useful information ranging from the minutes of the last city council meeting to a YouTube video on how to change your car's oil. In this sense, it represents an incredible "knowledge revolution" in which the amount of publicly accessible information has expanded exponentially, leading to tremendous new opportunities and economic growth. Unfortunately, there is also a dark side to easy universal knowledge sharing that has powerful negative implications for our societies and system of government: a platform that allows anyone to publish any content they choose also makes it far too easy for bad content to be published and to spread like wildfire.

In recent years, there has been explosive growth of misinformation, disinformation, conspiracy theories, hate speech, and similar socially negative content published on the internet. Even the smallest fringe or special interest group is able to erect an internet presence that is difficult to distinguish from a legitimate news source (which we define as a news source that operates strictly and consistently under well-established journalistic practices, e.g., lack of editorializing, balanced reporting, fact-checking, cross-verification from multiple sources, etc.). This has allowed individuals, fringe groups, and special interest groups to distort and manipulate public opinion in ways never before experienced in our societies, creating fake or misleading content to feed on fears about the economy, the COVID pandemic, climate change, immigration, and countless other topics. Sometimes the goal is to stoke divisions between people for political or personal gain, but sometimes misinformation is simply a means to make money by feeding on the outrage, fear and confusion that drives advertising dollars.

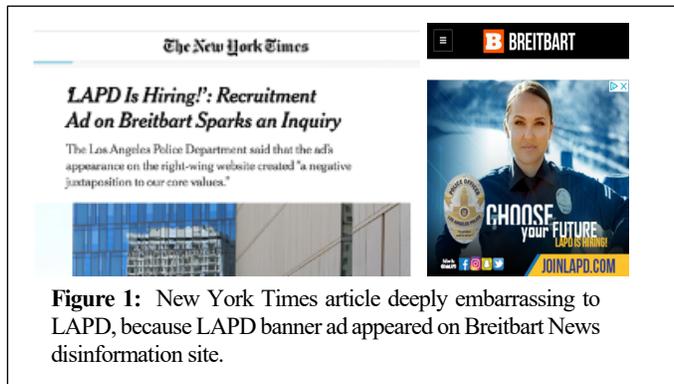


Figure 1: New York Times article deeply embarrassing to LAPD, because LAPD banner ad appeared on Breitbart News disinformation site.

A practical way forward: let's not slay the dragon, let's just start by taking its wallet!

NOBL Media believes in the power of the internet and in the importance of good, quality, fact-based information; everyone should have access to unbiased information and new reporting. The nature of the problem is essentially a matter of "smart filtering": there are still plenty of reliable news sources out there, it's just that they are literally buried under the flood of bad content, which is typically designed to be indistinguishable from real content. What is needed is for the internet (as it often has) to catch up and create powerful tools that solve this filtering problem, by helping to distinguish bad from good content. While this problem as a whole will be difficult to resolve and will take years, our goal with this project is to take a first step by focusing on just one important facet: our plan is essentially to "follow the money"!

We begin with the fact that it is advertising that funds much of the internet. Take a look at almost any site and you'll see ads that financially support that content. Whether you go to Reuters or dailygrail.com, you will be presented with banner ads for countless major brands. The same is true across the web, whether you are looking at major news publishers... or at any one of countless disinformation sites; they all make their money off of banner ads placed on

those sites. In order to meet the massive scale of the internet, the ad placement process is fully automated: Companies create ads and then contract with an internet advertising firm (Google is likely the largest); the company uses the “ad placement” tools of that firm to specify and configure the kinds of sites where their ads should appear. On the other side, site creators register their sites with Google, allowing Google to place banner ads on their pages...and Google pays them some tiny amount for each page view, and a premium if the banner ad is clicked. Thus, sites with large numbers of page views are able to make substantial amounts of money...including bad content sites. Indeed, the fundamental business strategy of many such sites has little to do with politics or conspiracy, but simply to create and post content that brings as many viewers to the site (preferably again and again) as possible. In short, many internet sites – whether good or bad actors – fund their operations and profits from banner ads dynamically placed on their sites by internet ad service providers like Google.

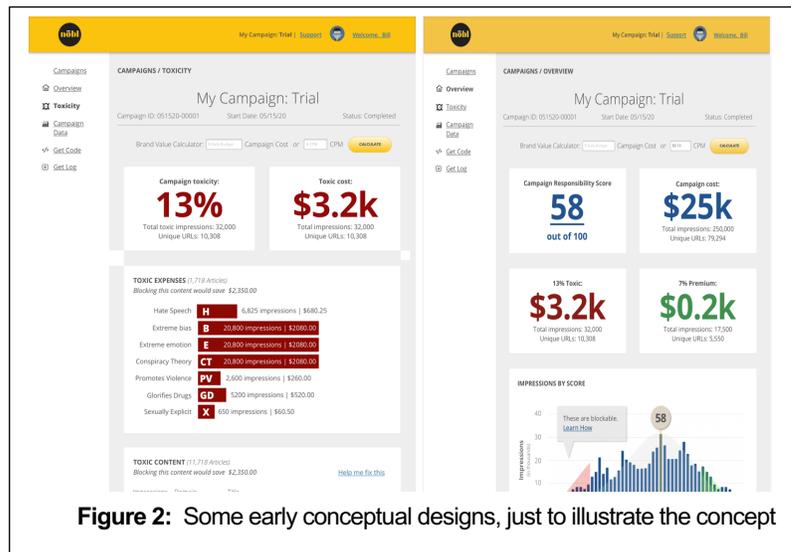
As a startup specializing in internet advertising, NOBL Media recognizes that for most firms advertising on the internet, their good public image is a vital business assets. As a result, advertisers want to reach their audiences *responsibly* and in positive ways, meaning they want their ads to appear on sites that reflect their company’s values and public image. One of the worst things that can happen to a company in this context is that their ad could be placed (by Google’s algorithms) on a “bad information” site. Unfortunately, the algorithm-driven dynamic ad placement mechanisms of the modern internet mean that major companies don’t always know exactly where their ads might appear...and sometimes there are major fails (Figure 1), with enormous public relations consequences for the company.

A Solution Vision: Disinformation Analysis and Detection

What is needed is an effective and accurate tool analyzing content on a website to allow companies more insight into the sites on which their ads are appearing, to be able to control their targeting. That’s where NOBL Media comes in. We have built technology that helps advertisers stop their media dollars from flowing to hate, disinformation, and extremism. We use AI to evaluate the content on pages where ads could appear to determine, in real time, whether that page has credible, engaging content worthy of showing an ad. By directing ads to the “good stuff”, we not only help our clients protect their brand image and make their advertising campaigns far more effective, we are helping to starve out misinformation. By preventing ads from displaying on the “bad” pages, our technology helps take away the funding for the people creating misinformation and hate speech, and helps drive them off the internet. Our advertisers win by having more effective, ethical campaigns, and we all win by cleaning up the internet!

We have focused our business on building the technology and tools that get this work done—that give advertisers transparency and control over their ads’ placement online. Our clients not only want to use the tools we are building, but they would also like to have a way to get a real-time view of how their ad campaigns are running, so that they have a clear sense of how responsible they are with their advertising dollars. With the powerful backend page analysis machinery in place, we are in need of an attractive, effective, and easy-to-use front end to help us demonstrate the benefits of the NOBL technology to new or prospective customers.

Specifically, we envision a powerful, secure web2.0 web application that provides a GUI front end for our tools. It would support both administrative users (our staff) in configuring various aspects of the interface, managing user accounts, and so on; and a public facing interface for our existing corporate clients to access their data and for potential clients to explore and try out our technology. We already have a clear set of data that we are collecting and understand the ways that our clients would like to visualize their data. What we need are the components that give our clients access to the data and those visualizations. Although we look forward to help from the team to fully specify the product, some key elements would likely include:



Level 0: The minimum viable product

- REST APIs that provide access to campaign data. You would not have to define the analytics algorithms yourself, but would need to implement them within the APIs. These APIs would be used by clients who prefer to create their own front-end visualizations of data but just need to get the raw data from NOBL.
- A graphical dashboard (a website) that gets its data from the APIs you have implemented and displays campaign data in a simple, easy-to-understand format

Level 1: A truly useful solution would also add:

- The ability to export reports (as a CSV, Excel or similar file) from the graphical dashboard
- Multiple different visualizations in the dashboard, with a menu allowing users to select which view they want to see.
- Integration with an authentication component, allowing users of the dashboard to log in to the site using a username and password that would automatically restrict their access to only their data.

Level 2: Possible stretch goals might add:

- A media planning interface that allows the media buyer to plan and model their campaign to maximize value from the service. This would basically require expanding on the functionality built into the dashboard to include extra interactivity and different views of the data
- Customizable reports in the web dashboard. That is, instead of having a single fixed-format report, clients would be able to add or remove fields, add filters and sorting, etc., and save their own custom configuration so that it would be reusable later.
- The ability to define automated alerts within the dashboard that could proactively notify clients when certain things happen to a live advertising campaign

As you work on this project, we expect that you would work with our team to help define the best technical implementation to provide these analytical tools for NOBL clients.

Broader Impacts: This project will be extremely helpful to NOBL's clients, because it will validate their partnership with NOBL. By having access to data and dashboards that clearly show how much good, credible content their ads are showing next to and how much harmful content they are avoiding, these tools will reinforce the value of the overall NOBL solution. We can clearly show that it is possible to stop the spread of hate, disinformation and conspiracy theories, fund fact-based journalism, and build a strong internet where truth matters. You can be a part of this!

Knowledge, skills, and expertise required for this project:

- Comfortable building server-side modules in Node.JS
- Basic user-level skills in Linux
- Knowledge of web design in React is useful but not required
- Basic knowledge of SQL databases (MySQL preferred), comfortable writing simple queries
- High-level understanding of API design and security using REST helpful

Equipment Requirements:

- There should be no equipment or software required other than a development platform (Mac, Windows or Linux are all fine) and software/tools freely available online.

Software and other Deliverables:

- The web application product outlined above, installed, tested and demonstrated on a platform of the client's choice.
- A strong as-built report detailing the design and implementation of the product in a complete, clear and professional manner. This document should provide a strong basis for future development of the product.
- Customer-facing API documentation detailing the functionality and implementation details needed for a customer to integrate with any APIs built as part of the project.
- Complete professionally-documented codebase, delivered as a repository in our company's GitLab server.