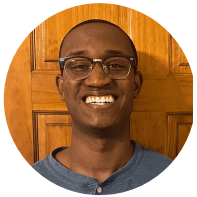
Misinformation Detector

Team Blue
CS 410
Professor El Mesalami

Table of Contents

Team Bio



Christopher Artis

Marc Ryan Carretero

Larry Teasley

Carter Gray

Ricardo Neri

Shawn Watkins

Ryan Asberry

Elevator Pitch

Misinformation continues to become a significant problem in America, which can be used for malicious intent. Misinformation creates confusion, spreads harm, and erodes public trust. The proposed solution is a web application that uses a machine learning model that will detect misinformation by:

- Scanning articles to see how much misinformation they each might have
- Search articles by topic and compare them with factual sources
- Take in consideration of attributes like sources, authors, website origin, tone, and general differences
- Rating how much misinformation might be in each article and providing additional commentary.

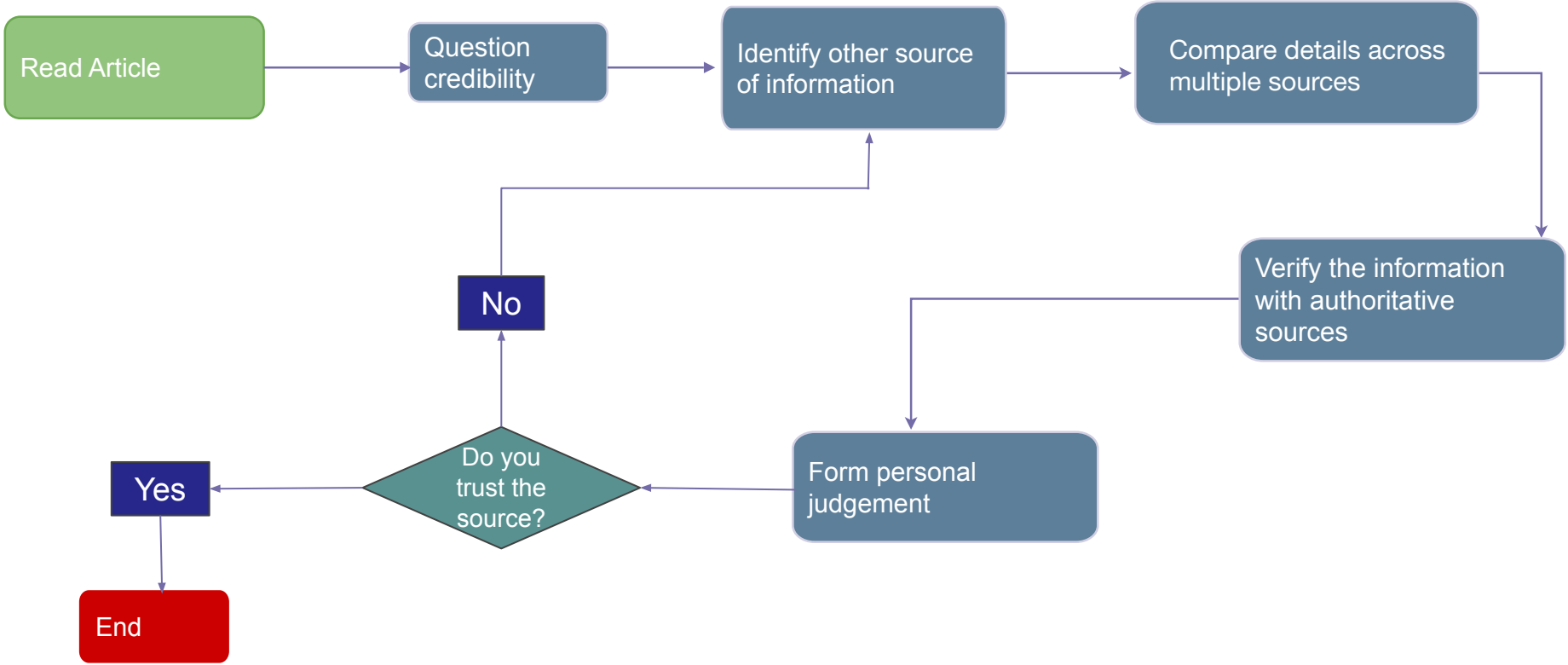
Problem

Given the spread of information due to easy access to social media, the spread of misinformation now targets a range of topics such as politics, business, and general everyday life. Trust in mass media has fallen drastically in recent years.

Problem Characteristics

- Average readers lack the tools to easily verify information
- Scale of online content makes manual checks extremely difficult
- Only 34% of Americans trust mass media to report in an unbiased, factual manner (Gallup, 2022)
- Misinformation spreads up to ten times faster than credible reporting on social media (PIRG, 2025)

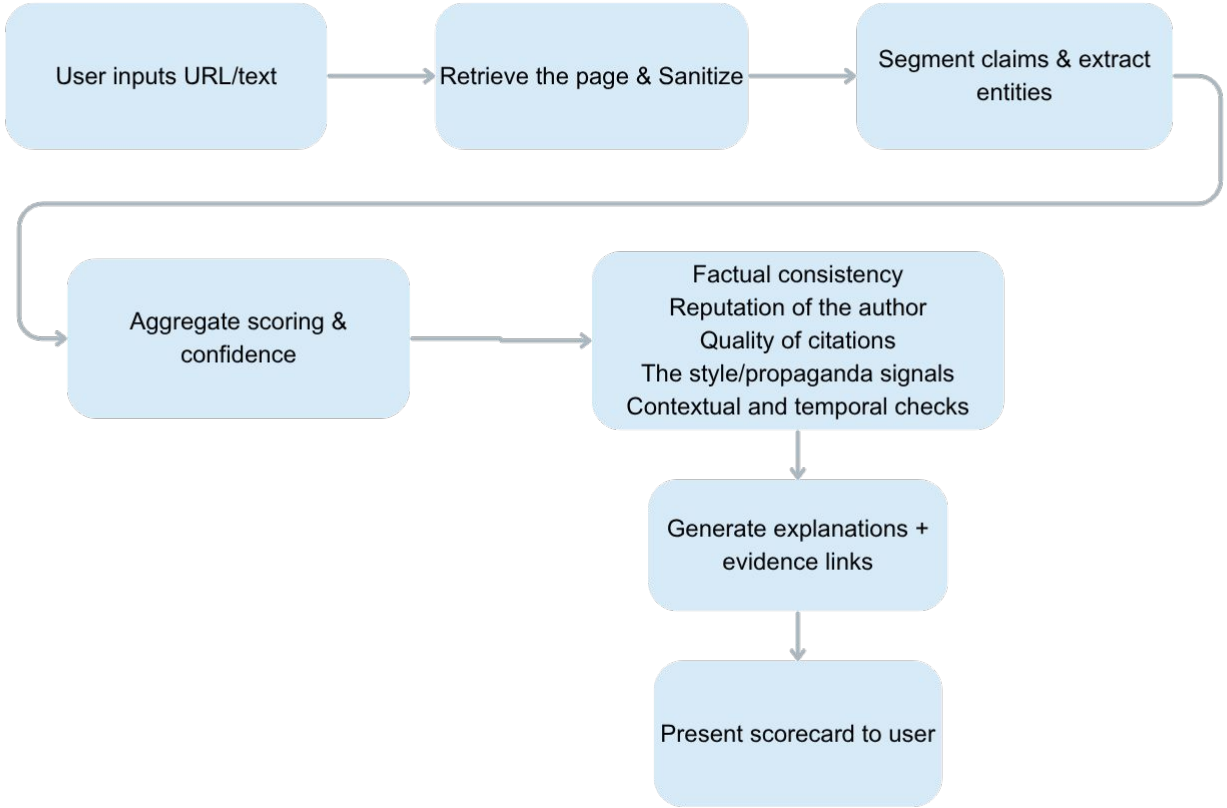
Current Process Flow



Solution

A web-based application that automatically evaluates the credibility of online articles. The tool will scan content, compare it against verified sources, and analyze factors such as authorship, tone, and factual consistency. A credibility score along with supporting notes will then be generated, enabling users to quickly determine whether or not an article contains inaccurate information.

Solution Process Flow



What It Will Do

- Compare and contrast sources of information for valid news information
- Take input from users (News article links, social media post links, or wikipedia articles), and compare the claims to trusted sources and factual articles
- Return a score of credibility
- Update regularly with new information as it is released

What It Will Not Do

- Provide resources containing opinionated or subjective content
- Analyze images for doctoring, AI or photoshopping
- The website will be built for text documentation or articles, images will not be included in the analysis part of the website

Competition Matrix

Features	Misinformation Project	Captain Fact	Lead Stories	Google Fact Check Tools
Database of trusted sources	Yes		Yes	
Image/Video Checking		Yes	Yes	Yes
Credibility scoring	Yes			
User-submitted information	Yes	Yes	Yes	Yes

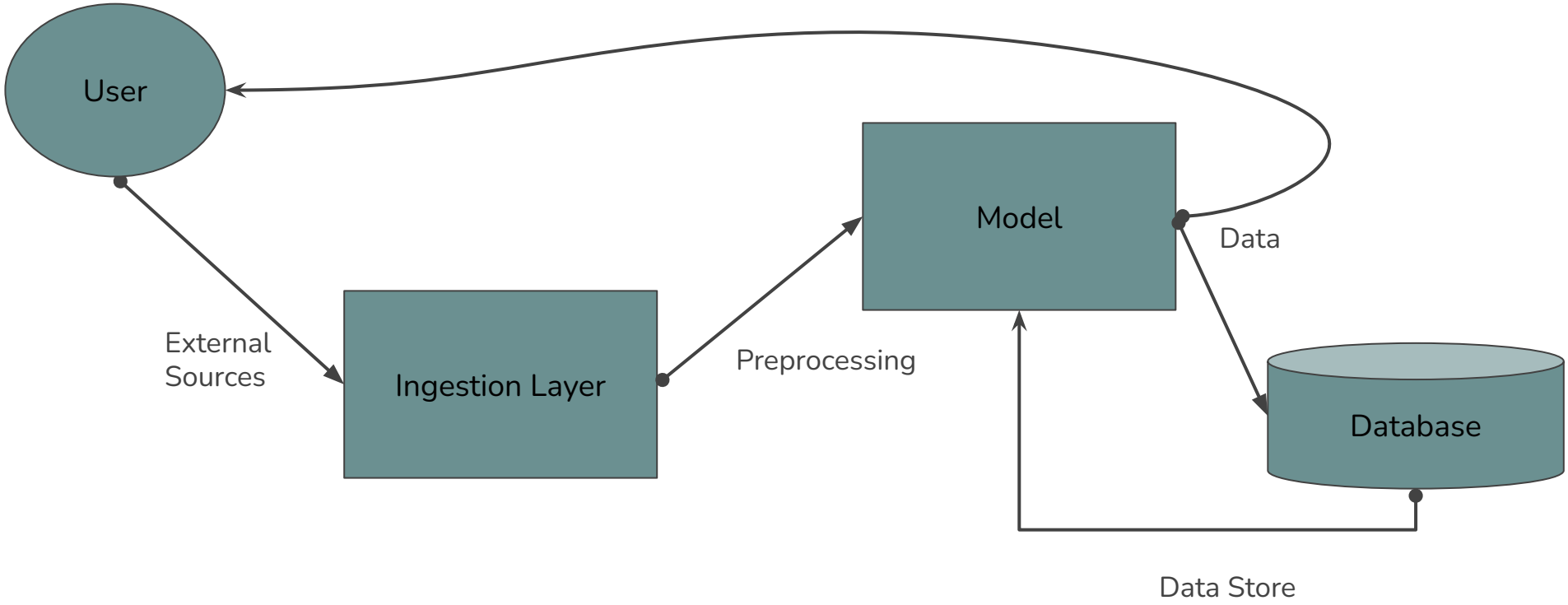
Development Tools

- Integrated Development Environment (IDE) - VSCode
- Version Control - Git through Github
- Continuous Integration (CI) - Github Actions
- Continuous Development (CD) - Github Workflows
- Selected Language (Backend) - MongoDB, ExpressJS, Node.JS, MySQL
- Selected Language (Frontend) - React, Javascript, HTML, Tailwind CSS
- Testing Framework(s) - Jest
- Documentation Tool - JSDoc

Major Functional Components

- Content Ingestion to collect data from sources
- Automated Misinformation Detection to run content through automated detectors trained by a machine learning model
- Source Credibility & Analysis to develop and maintain a registry of sources with credibility scores
- User Actions to define actions that users can execute on sources

Major Functional Components Diagram (MFCD)



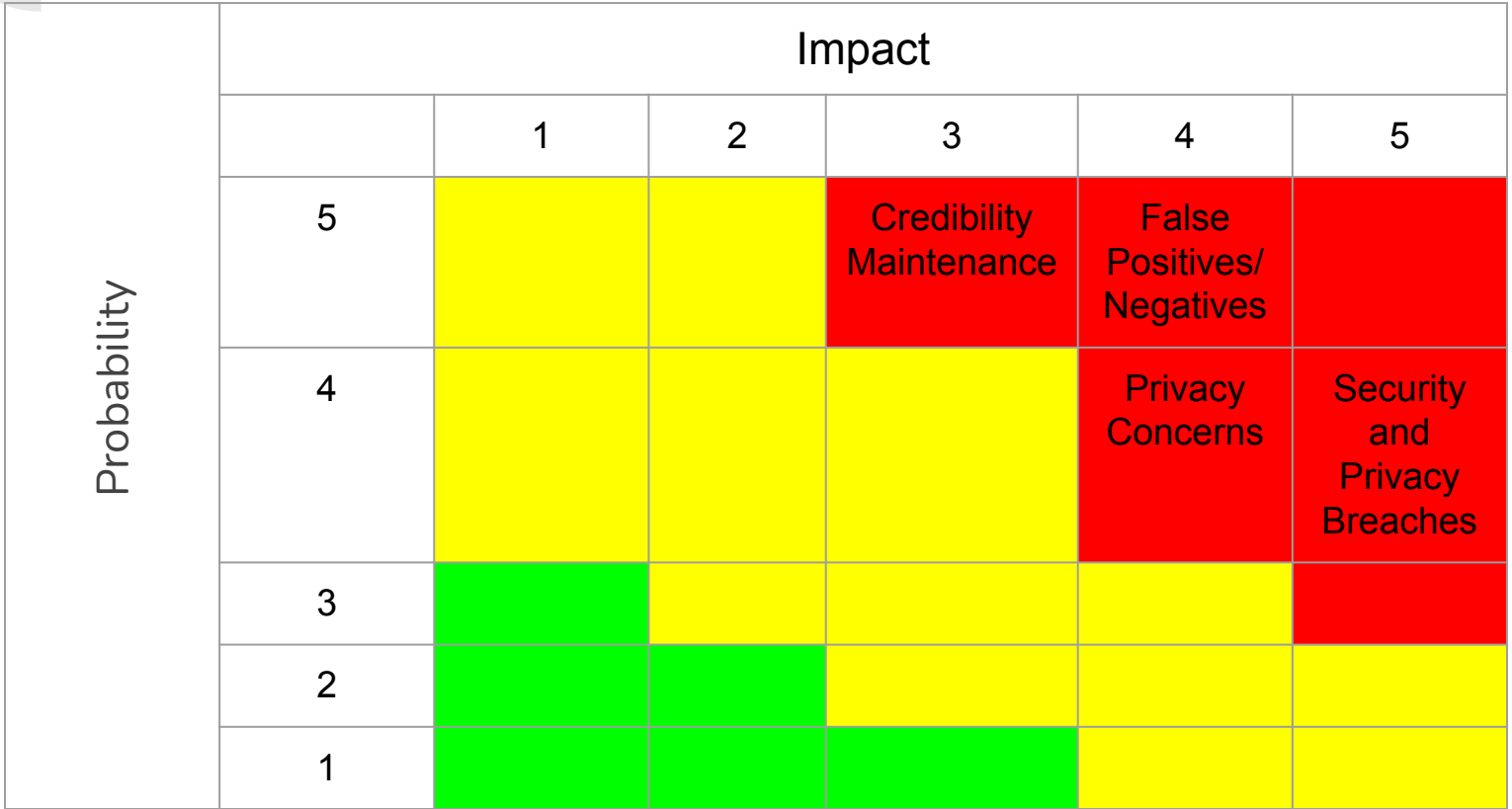
Risks

When developing a web-service, many risks can be presented during development and during operation. Hacks and outages are our main issues, with other issues pertaining to our users, legal action, or overall security presenting threats to the service and the user base.

Customer & End User Risks

- **Security and privacy breaches**: Data from unsecured or malicious sources could introduce malware or leaks
- **Privacy concerns:** Users may worry their data is being collected without consent
- **False positives/negatives:** Users may see truthful content flagged as misinformation or vice versa.
- **Credibility maintenance:** Keeping credibility data current requires ongoing analysis and human oversight

Probability of Customer & End User Risks



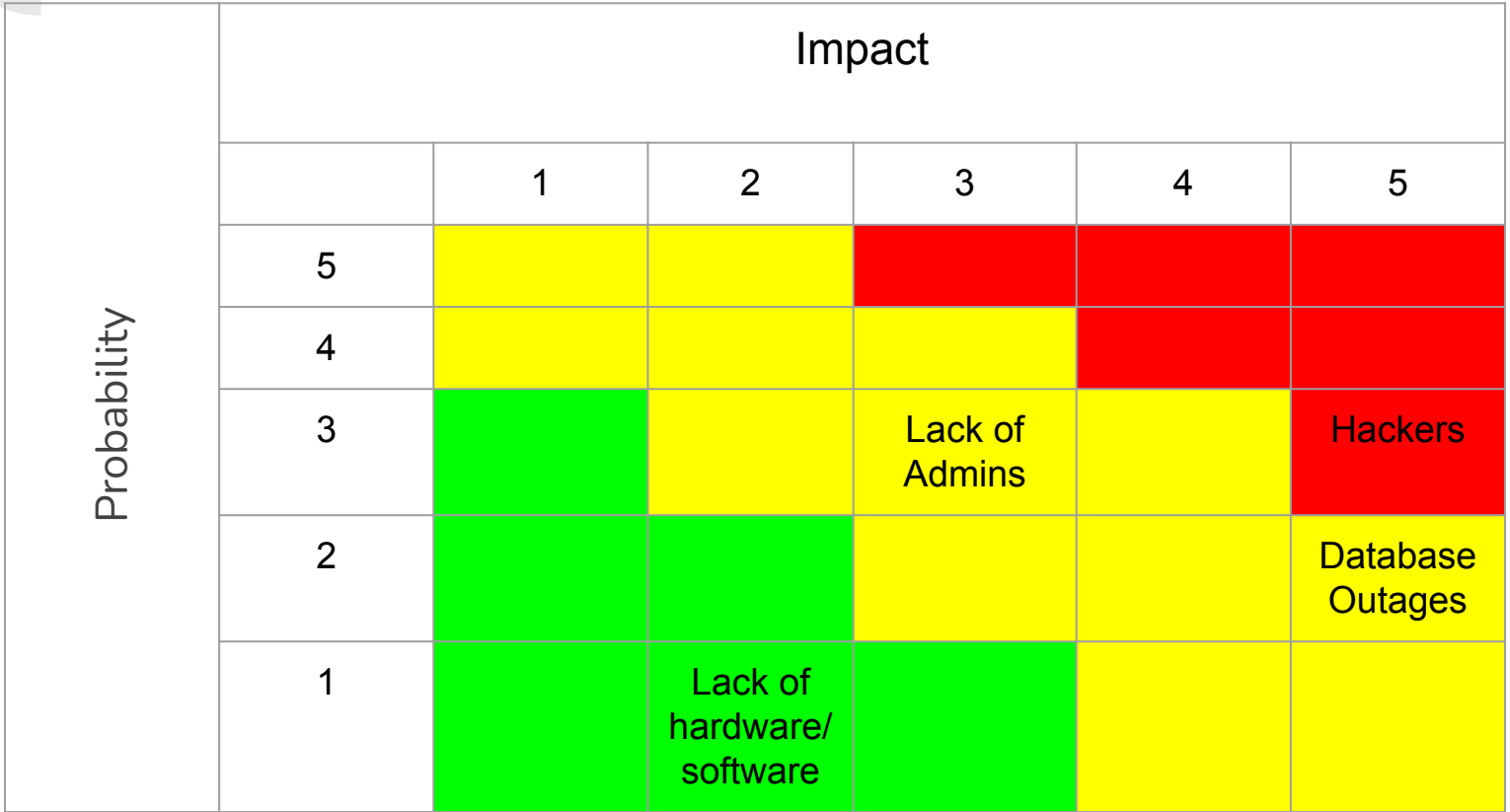
Mitigation of Customer & End User Risks

1) Implement strict data sourcing, only allowing input from vetted sources.
2) Collect only the necessary data for system functionality, and remove any personally identifiable information
3) Pair automated decision making with human fact checking for critical analysis and continuously update the models.
4) Schedule updates to regularly refresh and validated database and implement a feedback mechanism to report any errors.

Technical Risks

1) Hackers with ill intent trying to gather harmful data, or exploit the web-service for illicit or harmful actions
2) Database outages, preventing the service to access its knowledge base and provide information
3) Lacking admins with technical knowledge
4) Lacking sufficient hardware or software

Probability of Technical Risks



Mitigation of Technical Risks

1) Implementing modern security practices such as input sanitization and validation for SQL injections, and implementing smart and safe networking security
2) Utilize a third party database or cloud service, implement power backups and server backups
3) Implement clear standards for job positions and vetting policies to better ensure those applying are qualified
4) Entrusting data or services to third parties who have the necessary hardware or software resources

Security Risks

1) The submission of harmful URLs by users
2) Data poisoning - the feeding of fake “trusted” articles to make misinformation look real
3) Malicious users might flood the site with too many requests, causing server overload
4) Insiders or hackers could change data in the credibility database without approval
5) Logs or internal files could potentially be exposed on accident
6) Attackers might steal login cookies or tokens to act as legitimate users

Probability of Security Risks

Probability	Impact				
	1	2	3	4	5
5					
4				Data Poisoning	
3			Harmful URLs	Server Overload	
2		Accidentally Exposed Files	Stolen Login Cookies/Tokens		Data Change Without Approval
1					

Mitigation of Security Risks

1) Only allow safe web addresses (for example, start with https://)
2) Only accept new “trusted” sources from verified, reliable organizations
3) Add rate limits and CAPTCHAs; limit how many requests a single user or IP address can make)
4) Give editing rights only to approved staff (e.g. role-based access)
5) Store logs and backups privately and securely
6) Use HTTPS; make cookies secure and HTTP-only so scripts can’t access them; implement session timeouts

Legal Risks

1) Users who felt like they were scammed by the software or felt that the software did not do its job correctly.
2) Companies who did not want their data and/or articles to be used.
3) Individuals and/or organizations who had felt that the copyright was being abused through the software.
4) Individuals who intentionally spread misinformation might sue.
5) Governments who intentionally spread misinformation might sue and claim that the software is false news.
6) Companies who might have felt that the software and/or its ideas were plagiarized from their own similar software tools.

Probability of Legal Risks

Probability	Impact				
	1	2	3	4	5
5			Plagiarism		
4		Refusal of Use of Data by Companies	Scammed Users	Copyright Abuse	
3			Sued by Those Who Spread Misinformation		Government Issues
2					
1					

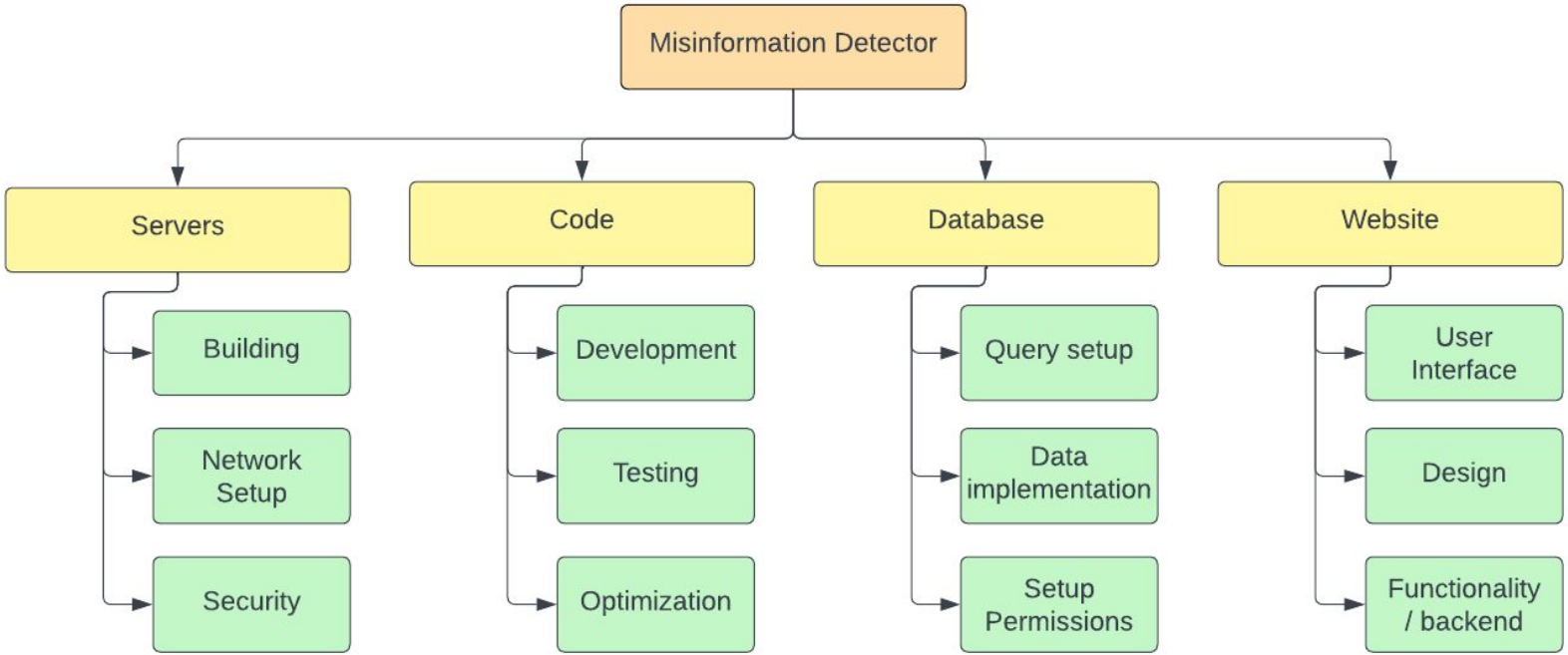
Mitigation of Legal Risks

1) To mitigate this risk, the software will be crafted to ensure that accuracy and preciseness is maintained when dealing with and analyzing potential misinformation. Additionally, apologies will be sent out to those who were affected.
2) To mitigate this risk, consent must be asked and given from each company on the use of this software on the articles. Additionally, the right to freedom of speech could be declared.
3) To mitigate this risk, the correct licenses in terms of fair use and copyright must be obtained.
4) To mitigate this risk, declaration of freedom of speech will be given beforehand.
5) To mitigate this risk, the government will be notified that this software falls under freedom of speech and that it is in the right to correct and to criticize false information.
6) To mitigate this risk, a declaration will be made that while the software may have similar ideas to other tools made in the past, those tools were not looked upon and there is no intention of plagiarizing any other software. Additionally, it is good to have competition so that the respective softwares can improve.

Real World Product vs Prototype

Features & Functionality	RWP	Prototype (Planned)
Database of trusted source	Yes	Yes
Credibility scoring	Yes	Partially
User-submitted information	Yes	Partially

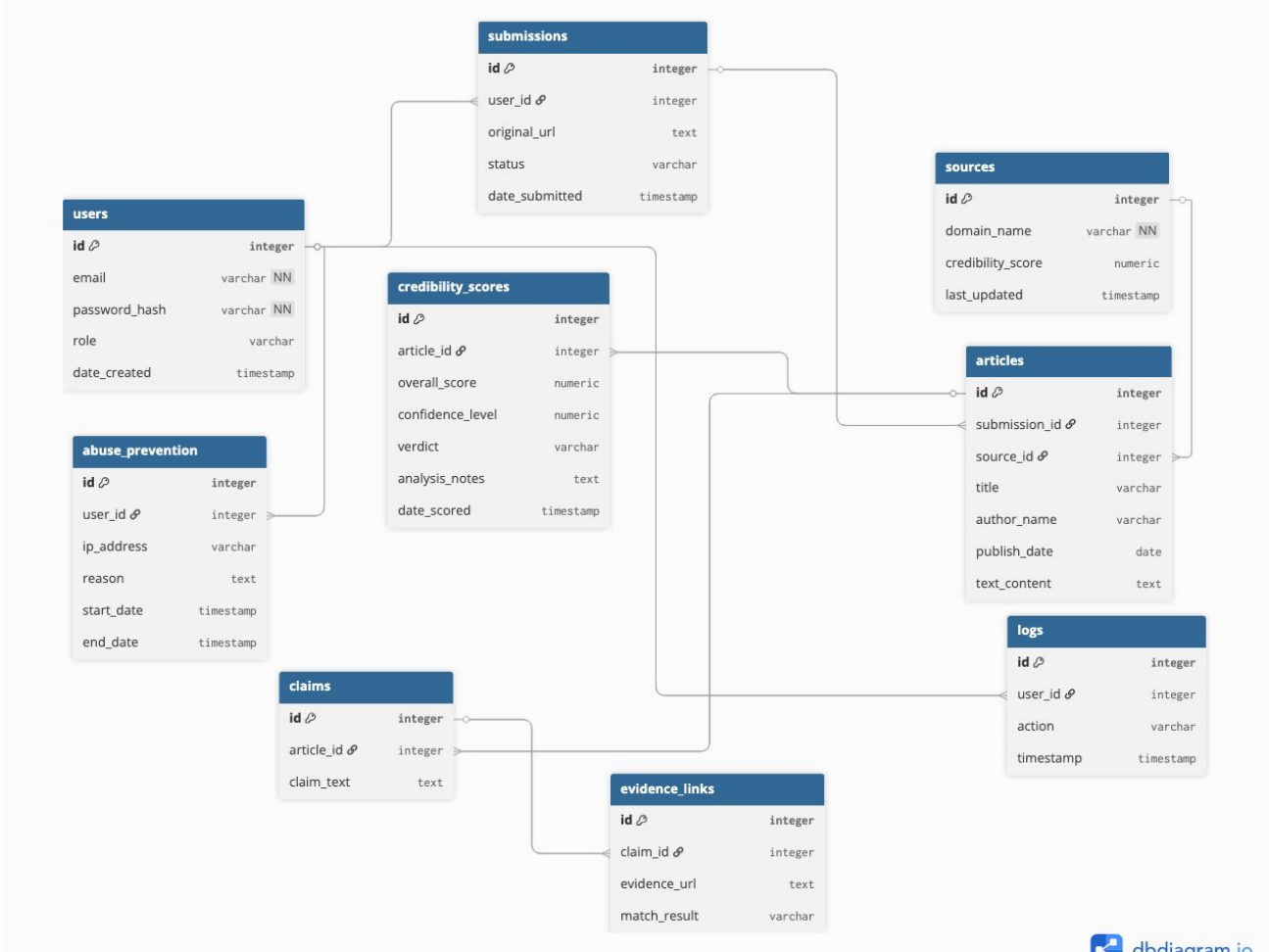
Work Breakdown Structure



Algorithms

- Vector Search - a search algorithm used to find similarities in text and context of articles, used to find similar topics to the one sent by the user
- Latent Dirichlet Allocation - A model used to cluster articles together based on topic
- Panda - an algorithm that searches for unique and quality pages, penalizing those with recurring or non-unique information or quality
- EEAT - Experience, Expertise, Authoritativeness, Trustworthiness. A set of guidelines rather than an algorithm, commonly used by google, to assess quality and trustworthiness of websites and articles

Database Schema



Development Tools

1.) Write: VS Code, Javascript, HTML, Tailwind, CSS, React, Node.js, ExpressJS, MongoDB, MySQL
2.) Test: Jest
3.) Run: Github Actions, Github Workflows
4.) Collaboration: Git

Required Libraries, Tools, & Technologies (Dependencies)

1.) Libraries: React
2.) Languages: Javascript, HTML, Tailwind, CSS
3.) Frameworks: Jest
4.) Technologies: MongoDB, ExpressJS, Node.JS, MySQL

References

1.) “Captain Fact.” *Rand.org*, https://www.rand.org/research/projects/truth-decay/fighting-disinformation/search/items/captain-fact.html. Accessed 27 Sept. 2025.

2.) “Lead Stories FactChecker.” *Rand.org*, https://www.rand.org/research/projects/truth-decay/fighting-disinformation/search/items/lead-stories-factchecker.html. Accessed 27 Sept. 2025.

3.) Gallup. “Americans’ Trust in Media Near Record Low” Gallup, 2022, https://news.gallup.com/poll/403166/americans-trust-media-remains-near-record-low.aspx

4.) PIRG. “How Misinformation on Social Media Has Changed News.” PIRG, 2025, https://pirg.org/edfund/articles/misinformation-on-social-media/

5.) “Google Fact Check Tools.” *Newsinitiative.withgoogle.com*, https://newsinitiative.withgoogle.com/resources/trainings/google-fact-check-tools/. Accessed 2 Oct. 2025.

6.) “About#faq.” *Toolbox.google.com*, https://toolbox.google.com/factcheck/about#faq. Accessed 2 Oct. 2025