Brian Mungal

CSCI 480

Bogert

1/15/25

Brief Proposal

Abstract

Web extensions are tools that help enhance the users browsing experience through tasks such as ad-blocking, password management, and content customization. A growing issue in research is evaluating credibility of articles, sources, and information to distinguish between well researched work and unreliable information. Existing tools such as NewsGuard, The Factual, Credibilator, and Captain Fact address aspects of news reliability, media bias, or misinformation, but often provide limited transparency regarding how credibility ratings are produced. This project presents SusAbility, a browser extension that computes an explainable credibility score for online articles by analyzing observable signals such as authorship, publication date, citation presence, and site-level indicators. The system presents users with both a numerical score and a breakdown of the factors contributing to that score, along with extracted references to support further verification. The extension was evaluated through time-based comparison tasks, perceived usefulness surveys, and measures of user awareness, demonstrating improved efficiency in assessing article credibility and increased user understanding of source quality.

Introduction

The rapid spread of online information has made it difficult for users to assess the credibility of articles and sources encountered during everyday web browsing. Recent changes to platform-level moderation and third-party fact-checking initiatives have further shifted the responsibility of evaluating information accuracy from institutions to individual users. As a result, readers are often left without transparent tools to help them judge the reliability of the content they consume. This project introduces SusAbility, a browser-based extension designed to assist users in evaluating the credibility of online articles by analyzing observable and explainable signals present on a webpage. SusAbility scans a given article for key indicators such as authorship, publication date, citation presence, and site-level characteristics, and uses these signals to generate an overall credibility score. Unlike existing tools that provide opaque ratings, SusAbility presents users with a clear breakdown explaining how each signal contributed to the final score, allowing users to understand and critically engage with the assessment. The

extension also extracts and displays cited sources and outbound references found within the article, enabling users to explore related material and conduct further verification when needed. By emphasizing transparency and user agency rather than automated truth judgments, SusAbility aims to support informed decision-making during online research. The primary users of SusAbility include students, researchers, and general web users who regularly encounter informational content online. Additional stakeholders include educators interested in promoting digital literacy and responsible research practices. Providing users with accessible tools for evaluating source quality is important for reducing the spread of unsupported or misleading information and encouraging more critical engagement with online content.

Literature Review

Several browser-based tools and extensions have been developed to assist users in evaluating the credibility of online information. These tools generally aim to reduce exposure to misinformation by guiding users toward higher-quality sources, though they vary significantly in methodology, transparency, and scope. NewsGuard is one of the most widely adopted credibility assessment tools and is built by a team of professional journalists who manually evaluate news and information websites. NewsGuard assigns each site a numerical score ranging from 0 to 100 and supplements these ratings with visual cues, such as green checkmarks and red exclamation points, that appear alongside links in search results. The platform also provides a "nutrition label" that details a site's ownership, funding, and editorial practices. While NewsGuard offers detailed evaluations, its approach is limited to news outlets and depends on continuous manual updates by journalists. The rationale behind numerical scores is not always immediately visible to users, and full access to the tool is restricted behind a paywall for many browsers. The Factual employs an AI-driven scoring system intended to guide readers toward higher-quality information by analyzing factors such as source diversity, author expertise, and language use. The platform also surfaces alternative articles and viewpoints on similar topics. However, The Factual provides limited transparency regarding how individual features contribute to its credibility scores, making it difficult for users to understand why a particular article is rated highly or poorly. Following its acquisition by Yahoo in 2022, the tool's public-facing focus appears to have shifted, away from recommending higher quality information, and is not limited to filtering malware. Credibilator is another AI-based tool designed to warn users about potentially inaccurate information. Its creators acknowledge that automated detection may not match the accuracy of human review, and the system relies heavily on pattern recognition and misinformation signals derived from social media content. While Credibilator visually distinguishes between credible and non-credible information, it does not clearly disclose how its models are trained or which features contribute to its classifications, reducing explainability and user trust. Captain Fact takes a crowdsourced approach to credibility assessment, allowing users to vote on the quality of sources and factual claims, particularly within video content such as

YouTube. This model emphasizes community participation and user empowerment, but its effectiveness depends heavily on sustained user engagement and moderation. As a result, coverage and reliability may vary across topics and platforms. Across these tools, a common limitation is the lack of transparent, user-facing explanations for how credibility assessments are generated. Many systems rely on proprietary scoring models, expert judgment, or ambiguous AI techniques that provide little insight into why a particular rating was assigned. SusAbility deliberately avoids automated factchecking in favor of an explainable, user-centered approach that analyzes observable credibility signals such as authorship, publication date, and citation presence then transparently reports how each signal contributes to an overall credibility score.

Deliverables

By the end of Capstone II:

1. **Fully Functional Browser Extension**
   A polished Chrome browser extension that assists users in evaluating the credibility of online articles through an explainable scoring system. The extension will analyze observable signals on a webpage and present results in a dedicated sidebar interface.

2. **Explainable Credibility Scoring System**
   A rule-based scoring model that assigns a credibility score to an article based on factors such as authorship, publication date, citation presence, and site-level indicators. The system will present a clear breakdown explaining how each signal contributed positively or negatively to the final score, along with an overall credibility tier (ex., low, medium, high).

3. **Reference and Source Exploration Interface**
   An information panel that displays extracted metadata and outbound citations found within the article, allowing users to review referenced sources and explore related material for further verification.

4. **Citation Generation Feature**
   A built-in citation generator that allows users to export formatted citations (APA or MLA) for the current article and any related sources surfaced by the extension.

5. **User Evaluation Study and Results**
   A structured user study evaluating the usability and perceived effectiveness of the extension, including task-based testing, time-to-verification measurements, survey responses, and qualitative feedback.

6. **Documentation and Presentation**
   Documentation describing the system architecture, scoring methodology, and limitations,

along with a final presentation, demo, and written report suitable for the capstone symposium.