

# Crowdsourcing Credibility: A Citizen-Science Approach to News Literacy via PublicEditor

Daniel Kim

University of California, Berkeley  
ddk@berkeley.edu

Aditya Ranganathan

University of California, Berkeley  
adityarn@berkeley.edu

## ABSTRACT

As technological, economic, and cultural barriers to online publishing have fallen, the proliferation of inaccurate or misleading news is a growing problem requiring practical solutions. By mobilizing the knowledge and interests of the crowd to assess articles, PublicEditor enables the public to evaluate the credibility of online news content and share their collectively-validated findings with the world in a politically-neutral manner. Mass collaboration software trains students and citizen volunteers to label online news and opinion reporting according to credibility criteria valued by the journalistic and scientific communities. PublicEditor uses a crowd-wisdom algorithm to aggregate these labels into credibility scores for articles, journalists, and news sites. A browser extension displays these scores next to news articles in the form of credibility badges, which indicate performance across various domains of analysis. These compact, data-rich badges not only allow readers of all skill levels to instantly assess the credibility of the content they are consuming; they provide transparent metrics of information quality rewarding journalists and newsrooms dedicated to high journalistic standards.

## KEYWORDS

News, Media Literacy, Crowdsourcing, Citizen Science, Data Mining, Credibility, Audience Engagement, Fake News, Misinformation, Journalism, Scientific Thinking, Annotation

## 1 INTRODUCTION

The Internet has democratized publishing far faster than it has democratized editing and fact-checking. As a consequence, biased and outright false information distort readers' conceptions of reality, with perilous consequences for democracy. The problem of misinformation is exacerbated in social media communities where like-minded peer groups reward punchy headlines and hasty conclusions that reinforce their worldviews. And even the most reputable publications have (at least on occasion) succumbed to the temptation to win readers and advertising revenue through 'clickbait' headlines and articles.

Motivated by a desire to help improve news literacy and restore trust in legitimate news sources, a group of experts came together in early 2017 at the Berkeley Institute for Data Science (BIDS) in order to develop a solution. Led by the Director of BIDS, physicist and Nobel laureate, Saul Perlmutter,<sup>1</sup> and BIDS research fellow Nick Adams, a team of data scientists, decision theorists, media researchers, psychologists, journalists, and natural language processing specialists began work on a new software platform. Called PublicEditor, it was developed to evaluate the credibility of news

and thereby help to improve media literacy while removing much of the confusion around news.

PublicEditor brings together the critical thinking skills of thousands of people in the vetting of daily online news content. This approach solves another major social problem, too: the public's sense of helplessness and anxiety in the face of so much low quality and misleading information. For the many people frustrated by polarizing discourses and partisanship today, PublicEditor will provide more than hope. It will engage them as members of a lasting community taking concrete and effective action to improve the quality of information all citizens consume.

## 2 BACKGROUND AND RELATED WORK

Large shifts in technology, economics, and culture have created opportunities for traditional news media, but they have also created enormous challenges. While digital and network technologies have enabled news media to reach new audiences beyond fixed geographic ranges, they have also enabled amateurs to create online news sites with confusingly similar aesthetics to those of legitimate news organizations. Although the emergence of click-based advertising models and distribution platforms have introduced alternative revenue streams for traditional newspapers, they have also encouraged and emboldened producers of low-quality or even misleading news who have become their competition for advertising revenues. Less obvious cultural changes have also accompanied technological and economic changes, including shifts in taste, habits, and etiquette that clash with some practices and values of serious journalism while complementing others.

These tensions and contradictions have intensified with the rise of third-party news aggregators and social media feeds. Today, the average news consumer who receives most of his or her news through social media platforms can be inundated with articles of varying degrees of quality and unknown provenance alongside articles from award-winning news organizations written by award-winning journalists. The concern cuts across political affiliations. A recent Pew poll reveals that nearly two-thirds of Americans agree that fabricated news articles create "great confusion" while another quarter believe it causes "some confusion."<sup>[1]</sup>

However, confusion is just the start of the problem; there are additional consequences on our understanding of, and trust in, news generally. As low-quality or misleading news sources have continued to proliferate, there is growing concern about our ability to consistently obtain, identify, and consume high-quality news. As news media remains our main source for information and opinions about current events and issues, our ability to engage thoughtfully with each other around important and urgent issues and to make informed decisions becomes increasingly threatened.

<sup>1</sup>2011 Nobel Prize in Physics—for the discovery of the accelerating expansion of the universe

Not surprisingly, there have been several different approaches to addressing this new complex of perverse incentives. Fact-checking websites like Snopes and Politifact, which have been around for many years alongside news organizations like the Associated Press, assign journalism professionals to look into contentious or questionable claims made in news articles. Another approach involves panels or teams of fact-checkers who are asked to rate articles that have been flagged. This is the approach of the International Fact-Checking Institute based at the Poynter Institute, and news aggregator sites that are using multiple sources of fact-checking to communicate disputed news articles.[2, 3] Crowdsourcing is another approach that is gaining popularity as demonstrated in applications like hypothes.is and Genius, which allow readers to see annotations made by other readers overlaid over webpages. Most recently, third-party news aggregators like Google and Facebook have turned toward algorithms to manage automatic processes of de-emphasizing suspect articles and cutting off advertising revenue to repeat offenders.

While each approach is able to address aspects of the problem of inaccurate or misleading news, each is also characterized by inherent limitations. Fact-checking sites that rely on small teams of paid workers, for example, are constrained in the number of articles they can analyze, as well as the speed, which means it is common for an inaccurate article to go viral long before Snopes or Politifact—or a news aggregator that relies on them—has completed their fact-checking process on it.[6] On the other hand, open crowdsourced annotation applications are easy to manipulate or ‘game’ by partisan readers since they can generally handle only a small number of annotators.<sup>2</sup> Thus far, the use of algorithms has yielded unpredictable or even undesirable results. Facebook has not yet been able to demonstrate that their efforts have been successful at all.[6] Furthermore, other counter-strategies are appearing on the horizon, including disreputable websites publishing occasional ‘real’ news story to create additional confusion.[8] Clearly, the problem is too large and complicated to be solved by a few efforts.

## 2.1 Understanding and Mobilizing the Crowd

At the heart of PublicEditor, there are two important understandings about human agency and collaboration. The first is the ‘citizen scientist’<sup>3</sup> model of participatory knowledge creation. Using our team’s state-of-the-art mass collaboration, annotation software, we train users to recognize the credibility criteria that indicate low- and high-quality content. Trained users are then enlisted to help with labeling words and phrases in an article. With practice, different members of the PublicEditor community will become experts at applying different sets of labels.

The second is the understanding that news is a form of culture. People engage with news media for a number of reasons—for information and entertainment, to create and reinforce social or community bonds, and to demonstrate intellectual or cultural status. PublicEditor mobilizes these complex and sometimes contradictory

<sup>2</sup>Politifact gets about 200 flagged articles a day, but only a “small amount” are outright false. [7]

<sup>3</sup> Popularized by platforms like Zooniverse and Galaxy Zoo, the first recorded use of the term was in an article describing a large project conducted by 225 members of the Audubon Society to collect data on acid rain levels.[5]

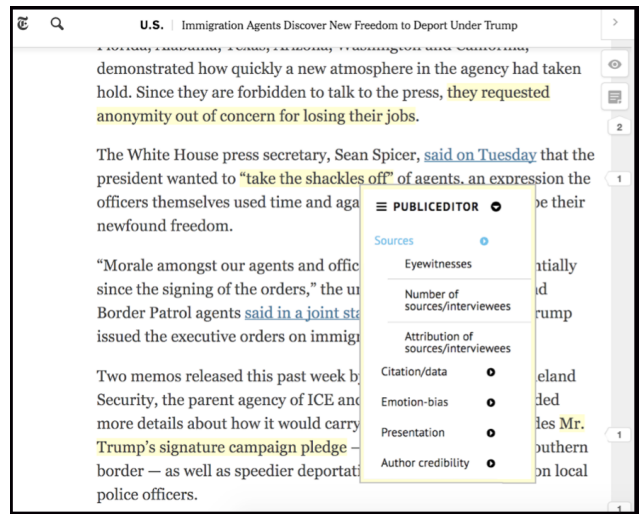


Figure 1: Example of a ‘flagger’ interface in action.

motives, and it redirects them into a process resulting in transparent and powerful news credibility metrics, increased media literacy, and rewards for newsrooms that strive to publish quality content.

## 3 THE PUBLICEDITOR PLATFORM

PublicEditor is a platform for evaluating the quality of news and combating inaccurate or misleading information. PublicEditor works by training citizen scientists in specific concepts across a set of domains such as scientific thinking, psychological biases, argumentative fallacies, and journalistic standards. The evaluation process for PublicEditor starts with identifying news that seems problematic. Articles that show signs of low-quality journalism are flagged by users via a PublicEditor browser extension. Consequently, flagged articles get sent to an advanced level user who then annotates pieces of text with conceptual tags. These sections of tagged text are then evaluated individually, and the data is recombined into a credibility score.

Neutrality is built into PublicEditor from the ground up. Unlike many news credibility platforms, PublicEditor is less concerned with the factual content of a given claim; citizen scientists focus on process and form. Based upon the scientific method, process checking confers the benefit of increased neutrality because users to assess how news articles arrive at specific claims, as opposed to asking them to make assessments about claims themselves. This is important for political news articles which can be especially partisan. That said, PublicEditor will eventually work in conjunction with fact-checking organizations.

PublicEditor as a system comprises many complex, interacting parts ranging from software to training. These can be grouped into four main categories: Annotation, Users, Concepts, and the Credibility Scores.

### 3.1 Annotation

PublicEditor’s user interface is designed to harness the power of annotation. From the initial flagging to the work that experienced

users perform, annotation is at the core of the PublicEditor user interface. In-browser annotation gives users who quickly want to highlight suspicious pieces the opportunity to do so without breaking flow or having to change their reading habits. By lowering the threshold of effort required to use PublicEditor, in-browser highlighting promotes increased use of the platform.

Operationally, annotation is powerful because it allows articles to be broken up into small chunks, allowing for a deeper level of analysis. PublicEditor uses annotation as a way to tag specific passages—paragraphs and sentences—from an article as being prone to committing a particular error. Then, only the tagged passages are passed on to the next level and evaluated. When users are presented with short passages of text, contextual information that might cloud their judgment is stripped away, which promotes a text-focused analysis. Lacking information about who the publisher is, what the overarching argument of the article is, etc., users are more likely to dissociate from their internal biases.

Additionally, the more granular the evaluation process, the richer the data that can be fed into machine-learning programs. As the number of users increases, the annotated data could provide essential information to automate parts of the detection process in the future.

### 3.2 Users

Users of PublicEditor span multiple degrees of engagement—from passive readers up through active annotators. Furthermore, PublicEditor allows users to change their role based on their level of interest and training. As users become more experienced, they are provided training in the relevant concepts at each stage. To satisfy the various user types, PublicEditor offers four unique interfaces, one for each idealized type of user role<sup>4</sup>: reader, flagger, specialist, triager.

*Readers.* Those who engage at the level of passive consumption are called ‘readers.’ Readers take advantage of PublicEditor with a simple browser extension that allows them to see credibility scores next to articles in their news feed. For readers, PublicEditor is a tool that helps them select and personally evaluate the news they read. The user interface for readers was designed to be clutter-free and offers no annotation capacity.

*Flaggers.* Users at the next level of engagement, reading with annotation, are called ‘flaggers.’ Like readers, flaggers use a simple browser extension and require no training in the concepts to use the system. Unlike readers, however, flaggers are able to mark specific articles as suspicious, either by annotating as they read or by ticking a box to mark the entire article as worthy of review. These annotations are tracked in the system, and a metric that tracks the accuracy of each flagger is factored into the credibility score algorithm. The flagger interface marks the boundary between passive consumption and active contribution to the PublicEditor system.

<sup>4</sup>Note that a single user can use different interfaces at different times. For the purpose of this paper, we are using user to describe someone who occupies a single role in the PublicEditor system, but it is very possible—likely, even—that a contributor may also be a simple consumer at a later time.



Figure 2: Example of a working Triager interface. Each color highlight corresponds to a particular section of ‘tagged text’.

*Specialists.* A ‘specialist’ deepens engagement by evaluating an article with respect to a particular concept. Specialists are users who get ‘badged’ for each concept. A badge is earned after conceptual training and practice is completed. Once a specialist is badged in a particular concept, they are sent portions of text to evaluate with respect to that concept. For example, a specialist who is trained in correlation and causation will be sent passages that potentially conflate correlation and causation; it is the specialist’s job to determine whether the passage actually commits the error. To contribute, specialists log in to their site, and select a concept they are badged on (specialists can have multiple badges). At that point, text will be presented to the specialist (from the triager), which they can vet for however long they like. Once a specialist has been badged in a concept, they can begin the training for the next concept.

*Triagers.* Specialists become ‘triagers’ after they accrue multiple badges and become qualified to discern the applicability of a concept to a given article or portion of an article they can progress to the ‘triager’ role. Like specialists, triagers log on to their own interface. But unlike specialists, who are sent specific pieces of text from within an article, triagers are sent entire articles (from flaggers). The triager is then responsible for tagging parts of the article with the relevant concepts. For example, a triager receiving an article about “how Advil cures cancer” would annotate the portions of the article that talk about correlation between Advil use and reduced cancer rates and tag that section as ‘correlation and causation.’ The selected text would then be passed on to a specialist who was badged in correlation and causation.

### 3.3 Credibility Concepts

At the core of PublicEditor’s work are over 150 credibility concepts curated across a variety of domains: scientific thinking, argumentative fallacies, psychological biases [4], and journalistic practices. Approximately half of these concepts originate from a course on scientific style critical thinking<sup>5</sup> at Berkeley co-taught by Saul Perlmutter. The remaining concepts were developed by an interdisciplinary team of Decision Theorists, Psychologists, Media Researchers, and Journalists. Together, these concepts serve as an

<sup>5</sup>For a more complete list of concepts, visit <http://sensesensibilityscience.com/>

interdisciplinary lens for news readership that can lend unique insight.

In PublicEditor, there is no explicit separation between types of articles evaluated. As a result, scientifically culled concepts such as fact and value, or signal and noise, can be applied equally to articles on a range of topics. The cross-discipline application of concepts allows PublicEditor’s triagers and specialists to go beyond the reporting norms of different fields and dig deeper into the underlying quality of the journalism.

For specialists and triagers to deploy concepts to news, proper training is essential. Training in concepts will be provided using a twofold approach that combines both an ‘online classroom’ and iterative learning. The online classroom will use animations, readings, and demonstrations. However, as we discovered in Berkeley classes that teach credibility concepts, theoretical knowledge of the material is insufficient for proficient application in news. To that end, PublicEditor utilizes a large database of expertly annotated articles for each concept. Trainees annotate these articles on their own, and then are shown the expert results. Once a trainee shows sufficiently close results to that of experts, they earn a badge in the given concept.

### 3.4 Credibility Scores

PublicEditor’s commitment to working with the actual people who create and consume journalistic content sets it apart from other attempts to generate media credibility scores. While it is possible to quickly generate measures of information quality using purely automated methods (at least for articles and news sites if not for content), real credibility is only achieved through a social process. Journalists and readers must be included in an open and transparent approach if credibility scores are to become widely trusted and adopted. Thus, the primary goal for the PublicEditor credibility score is to maximize transparency and minimize effects of individual bias.

Once each article has been reviewed by a complement of specialists, PublicEditor combines their individual responses into a credibility score. Credibility scores are generated using a crowd-wisdom algorithm that accounts for both the trustworthiness of each user’s review and the relevance of the concepts used to evaluate the article. The weight of each user’s rating is determined by examining a user’s performance in relation to that of trained experts. Concept validity, likewise, is found by calculating the correlations between structural anomalies and the presence of misleading information in a given article.

Additionally distributions of user data can also be used to reduce the impact of biased data on the credibility score. Only normal distributions that show expected variation amongst users are included in the credibility score. For example, bimodal or flat distributions that are indicative of confusion or polarization in contributors are not used. Furthermore, ratings from individual contributors who are highly correlated with partisan positions are not used.

By splitting articles into bite-sized passages, PublicEditor is able to show a reader a wealth of information about how their peers (other users) rated a given article. PublicEditor makes the entire process for determining the credibility score open to the public. An occasional incorrect credibility score is inevitable, but with a

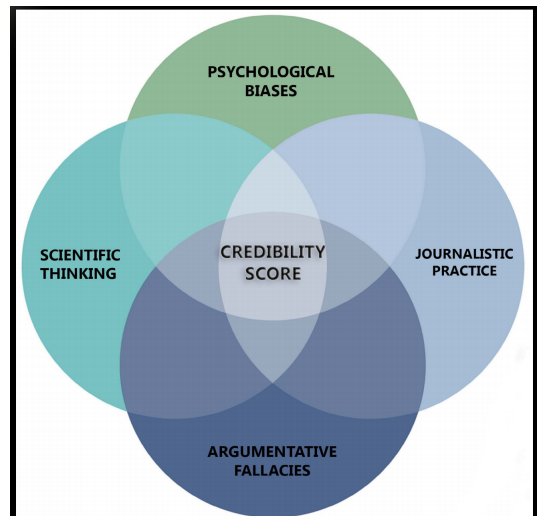


Figure 3: The credibility score spans four different conceptual domains.

transparent algorithm, we hope our users will catch mistakes in the PublicEditor algorithm just as they catch mistakes in the news. The overall credibility score attempts to capture four-dimensional data about an article in an easy-to-read fashion, but greater detail is also easily available for the interested reader.

## 4 EVALUATION AND IMPROVEMENT

In order to ensure that the credibility concepts provide a consistent and reliable indicator as to the quality of news, testing and iteration are necessary. Establishing high confidence in inter-rater and inter-item reliability is essential in creating a trusted credibility tool. It begins with the PublicEditor team’s expertise across a number of conceptual domains and expertise in survey creation and continues with a systematic approach to building a resilient, effective platform. Every concept and item will be tested in between iterations by being used to evaluate real news. Data provided by our users can help answer important questions for concept and platform refinement. Which concepts are most powerful for each type of news? What concepts do users find particularly challenging to apply? How well do different training modules work? Although large-scale platform testing will likely not be possible till the end of 2017, preliminary testing of the 25 most important concepts has already begun.

In Fall of 2017, we will begin a rigorous beta-testing process with college students and Amazon Mechanical Turk workers. This testing phase will utilize an extensive data set of both high-credibility news and low-credibility news. The credibility algorithm will be tested for its ability to accurately discern the difference between the high-credibility and low-credibility news samples. The aim is to gather correlations between each credibility concept—starting with 25-30 concepts in our first beta testing cycle—and credible journalism. Correlations between the validity of a concept and correct identifications of news credibility can be used to derive a utility score for each concept, define concepts more clearly, and help create more effective training modules.

PublicEditor has been designed with the philosophy of iteration in mind, and will continuously release and update our concepts as more users come on-board to provide us with feedback on how the platform can be more effective. And users can request concepts as well—the 150 concepts we’ve curated so far are merely an initial foray—which will increase the comprehensiveness of the PublicEditor system over time.

## 5 FUTURE WORK

In order to ensure broad and lasting adoption among a stable population of users with different motivations, experiences, and commitment levels, the PublicEditor project will require the development of additional features.

### 5.1 Gamification

By ‘gamifying’ users’ activities and encouraging more social uses of PublicEditor, we will be able to build additional incentives beyond the desire to share information. Using features like point systems, status badges, special privileges, and other encouraging forms of recognition, users will come to see that contributing to PublicEditor is not only a good deed, but a fun and rewarding activity that can be integrated into everyday life. Consistent with our understanding that news is a form of culture, PublicEditor can become a community of like-minded people who share an interest in quality journalism, media literacy, and open discourse.

### 5.2 Natural Language Processing

By incorporating natural language processing and machine learning into future iterations, PublicEditor will also begin to explore automated process-checking for simple concepts. The granularity of training data (i.e., for machine learning) that PublicEditor generates will help scale-up the flagger and triager interfaces to be able to evaluate at least the top 200 articles shared on the internet each day during the 2018 election cycle. The addition of NLP will enable news articles to be vetted by both community members and software.

## 6 CONCLUSION

We have highlighted some of the features of Public Editor, an emerging citizen-science platform that engages users from a range of backgrounds in the process of improving the quality of media. Using credibility concepts based on principles and insights from science and journalism, trained volunteers make assessments about news articles. These assessments in turn are used by an algorithm to calculate aggregate scores, which will be displayed as credibility indicators placed next to articles posted on news aggregation sites. Not only will credibility indicators help news consumers make informed decisions about their consumption choices, but they will also encourage writers and editors of news articles to adhere to higher standards of journalistic professionalism and responsibility. In time, we hope that PublicEditor will serve as a source of useful data for newsrooms and news aggregators alike. In the process, PublicEditor will foster a growing community of citizens improving their own media literacy while delivering an independent signal of news quality that news outlets can use to help differentiate their writing in an increasingly noisy information landscape.

## ACKNOWLEDGMENTS

The authors would like to thank Nick Adams, Saul Perlmutter, the PublicEditor team, the Berkeley Institute for Data Science, and the Center for Science, Technology, Medicine and Society at Berkeley.

## REFERENCES

- [1] Michael Barthel and Amy and Mitchell. 2016. *Many Americans Believe Fake News Is Sowing Confusion*. <http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion>.
- [2] Samuel Gibbs. 2017. Google to display fact-checking labels to show if news is true or false. *The Guardian* (April 7 2017). [goo.gl/UXPhuu](http://goo.gl/UXPhuu).
- [3] Elle Hunt. 2017. Disputed by multiple fact-checkers. *The Guardian* (March 21 2017). <https://www.theguardian.com/technology/2017/mar/22/facebook-fact-checking-tool-fake-news>.
- [4] Daniel Kahneman. 2011. *Thinking Fast and Slow*. Farrar, Straus and Giroux, New York.
- [5] Robert Kerson. 1989. Lab for the environment. *MIT Technology Review* (Jan 1989), 11–12.
- [6] Sam Levin. 2017. Facebook promised to tackle fake news. But the evidence shows it’s not working. *The Guardian* (May 16 2017). <https://www.theguardian.com/technology/2017/may/16/facebook-fake-news-tools-not-working>.
- [7] Laura Hazard Owen. 2017. *Facebook is paying its fact-checking partners now (and giving them a lot more work to do)*. Nieman Foundation, Harvard University. [goo.gl/knG92m](http://goo.gl/knG92m).
- [8] Jon Swaine. 2017. New fake news dilemma: sites publish real scoops amid mess of false reports. *The Guardian* (May 16 2017). <https://www.theguardian.com/media/2017/may/16/fake-news-sites-reports-facts-louise-mensch>.