Federal Register Notice 87 FR 15274, https://www.federalregister.gov/documents/2022/03/17/2022-05683/request-for-information-on-federal-priorities-for-information-integrity-research-and-development, May 15, 2022

Request for Information on Federal Priorities for Information Integrity Research and Development

NewsGuard

Request for Information on Federal Priorities for Information Integrity Research and Development

1. *Understanding the information ecosystem:* What are the key research challenges in providing a common foundation for understanding information manipulation within this complex information ecosystem?

One of the most significant challenges for researchers investigating the information ecosystem is navigating the opacity of social media platforms' algorithms: specifically how their algorithms curate, downgrade, and deplatform content and users. Until technology platforms provide researchers and industry stakeholders with access to data on the spread of information on their platforms, and the factors that trigger their algorithms to promote or demote content, researchers cannot adequately understand how misinformation spreads online, and thus are hindered in their ability to develop effective strategies to combat the spread.

NewsGuard is an anti-misinformation internet trust service powered by human intelligence, not artificial intelligence, that protects brands, consumers, and democracies from misinformation. Our trained journalists rate the online news and information sources that account for 95% of engagement with the news, assigning each outlet a point score out of 100, a Green or Red rating depending on its score, and a detailed "Nutrition Label" review describing the source and its adherence to standards of credibility and transparency. Together, these nearly 8,000 Nutrition Labels and point scores form our Reliability Ratings dataset of source-level evaluations.

Our second dataset is the Misinformation Fingerprints catalogue — a collection of hundreds of human- and machine-readable myth entries chronicling the most prominent false claims circulating on the internet, spanning topics ranging from the Russia-Ukraine war to COVID-19 misinformation. Each entry contains a description of the myth, a debunk of the myth, citing authoritative sources, variations of the myth, associated keywords and hashtags, and a list of websites we've rated that have perpetuated the myth. This data can be used to seed AI/machine learning tools to track misinformation online at scale, and to aid content moderation efforts on social media platforms, among other use cases.

From our perspective at NewsGuard, having greater access to data from the major platforms would enable us to leverage our expertise to yield insights about how misinformation spreads online and make recommendations about how to mitigate that spread. For example, it would be beneficial to know the proportion of posts shared on platforms like Facebook and Twitter that contain links to news sources that are "Red-

rated" and deemed generally untrustworthy by NewsGuard (scoring below 60/100 on our nine apolitical journalistic criteria of transparency and credibility) – or content matching one of our Misinformation Fingerprint entries. Likewise, it would be useful to understand the level of engagement on these posts, the individuals chiefly responsible for perpetuating the misinformation in these posts, and the number of users who were exposed to posts containing such misinformation or content from unreliable sources. It would be of extraordinary value to know what percentage of users of each of the major social media platforms get most of their news and information from unreliable sources— we believe this number is significant and goes to the heart of the information-manipulation crisis affecting many people who are being shown false content and have no tools at their disposal to know which sources are generally reliable and which are not.

Thus far, the major social media platforms have not made these metrics widely available, thus hindering important research in the field of misinformation. Without this level of transparency, research practitioners are unable to adequately assess the full extent of the problem.

A related challenge facing researchers in this field is determining how to define the scope of "news" or "information" online. Defining the online information ecosystem has become even more complex as a result of the fact that platforms host user-generated content while also retaining algorithmic and curatorial authority over what users ultimately see on their platforms. The industry needs a baseline definition of what constitutes a news or information source to not only make disparate research efforts in the field more comparable, but also to hold platforms to account by requiring them to clearly delineate the parts of their business that publish news and authoritative information from those that publish user-generated content. Further, industry stakeholders need a common understanding of what distinguishes a reliable news and information source from an unreliable one.

Conceptualizing such a definition of news and information sources is the collective responsibility of regulators, platforms, and industry, and it is a decision that NewsGuard is well-positioned to help inform. For example, NewsGuard has already become the industry standard for source reliability classifications among academic researchers studying misinformation, used by top researchers at Stanford, Dartmouth, New York University, Northeastern University, the German Marshall Fund, and other institutions. Our work is used by entities from the World Health Organization to the Pentagon's Cyber Command and the State Department's Global Engagement Center.

In sum, researchers, publications, news aggregators, platforms, and content moderators would benefit from both (1) a baseline definition of what constitutes a news or

information source that is transparently communicated to all stakeholders and (2) an impartial, objective third-party source reliability assessment to use as a common denominator when promoting and curating information for the benefit of consumers and quality newsrooms alike.

2. ***Preserving information integrity and mitigating the effects of information manipulation:*** **Strategies for protecting information integrity must integrate the best technical, social, behavioral, cultural, and equitable approaches. These strategies should accomplish a range of objectives including to detect information manipulation, discern the influence mechanisms and the targets of the influence activities, mitigate information manipulation, assess how individuals and organizations are likely to respond, and build resiliency against information manipulation. What are the key gaps in knowledge or capabilities that research should focus on, in order to advance these objectives? What are the gaps in knowledge regarding the differential impact of information manipulation and mitigations on different demographic groups?**

When it comes to the investigation and identification of information manipulation, much attention has been paid to developing AI-based solutions that can track the provenance and spread of dubious content at scale. One example is the Content Authenticity Initiative (CAI) announced by Adobe in 2019 in partnership with Twitter and the New York Times.[1]

But such technology-based solutions are incomplete if they are not coupled with the expertise of human experts capable of parsing the subtleties of misinformation. More focus must be placed on developing human-intelligence solutions that can be paired with AI technologies in the process of information investigation. Early research conducted by NewsGuard with AI technology company Blackbird AI shows that NewsGuard's source Reliability Ratings and Misinformation Fingerprints library of false content can augment the efficacy of AI-based misinformation tracing efforts.[2] More research should be done to extend these early attempts to combine human and artificial intelligence in pursuit of holistic and effective misinformation detection.

In terms of mitigation efforts, strategies for protecting information integrity are generally most sustainable when they elicit lasting behavioral change among news consumers. Large platforms and companies with large audiences should have a responsibility to provide user empowerment tools that equip consumers with greater context and authoritative information in order to make their own decisions, centering the user

---

[1]https://contentauthenticity.org
[2]https://www.newsguardtech.com/wp-content/uploads/2020/10/NewsGuard-x-Blackbird.AI-Joint-Report-1.pdf

experience instead of censoring information or blocking content. One example of such a user empowerment intervention is NewsGuard's browser extension, which displays Red and Green shields next to links to sources, indicating the general reliability of a source at the moment a user first encounters a piece of content. From there, users can read the full "Nutrition Label" review of the source, empowering them to gain richer context behind a source without blocking them from accessing any content. Researchers and other industry stakeholders should promote research and development of tools like NewsGuard, which is recommended by the European Commission to the digital platforms to help them comply with the user-empowerment requirements of the European Commission Code of Practice on Disinformation and by the UK government as it has identified online safety tools the platforms can provide their users.

3. *Information awareness and education:* A key element of information integrity is to foster resilient and empowered individuals and institutions that can identify and abate manipulated information and create and utilize trustworthy information. What issues should research focus on to understand the barriers to greater public awareness of information manipulation? What challenges should research focus on to support the development of effective educational pathways?

Forthcoming research should focus on evaluating the efficacy of educational interventions on users' ability to discern information manipulation, such as NewsGuard's browser extension, which provides seamless source reliability ratings alongside social media posts and search results. Research that demonstrates the effectiveness of equipping online users with contextual information and source credibility indicators would help regulators and educators implement tools for students and young people to become more discerning consumers of online information.

For example, researchers and practitioners might look to NewsGuard's partnership with the American Federation of Teachers, one of the largest teachers' unions in the U.S. Through the partnership, the AFT's 1.7 million educator members receive free access to the NewsGuard browser extension for their 20 million students and their families, providing a unique media literacy intervention. Studies that test the efficacy of NewsGuard and similar tools in classrooms would accelerate the adoption of these research and contextual information tools.

Similarly, NewsGuard integrates with new media companies that are reimagining existing platforms and offering user-centric benefits like privacy, ads-free interfaces, and source reliability context. Bright, a new social media app that emphasizes user privacy and experience, displays NewsGuard's ratings within the app where news links are

posted to give users and moderators alike greater context for the content being shared.[3] Neeva, an ads-free subscription search engine founded by former Google and YouTube executives, similarly displays NewsGuard data in its search results.[4] Further research measuring the effectiveness of NewsGuard integrations in social and search platforms would help make the case for other technology companies to incorporate NewsGuard's user empowerment tools to make the online media environment a safer place.

4. *Barriers for research:* Information integrity is a complex and multidisciplinary problem with many technical, social, and policy challenges that requires the sharing of expertise, data, and practices across the full spectrum of stakeholders, both domestically and internationally. What are the key barriers for conducting information integrity R&D? How could those barriers be remedied?

One of the central barriers to conducting information integrity R&D is insufficient baseline context for the credibility of sources and individual pieces of content in the sprawling online information ecosystem. Presented with a Twitter dataset of billions of tweets, for example, a researcher is tasked with the challenge of determining which pieces of content may contain misinformation.

NewsGuard offers one solution to this quandary by providing human-curated, independent assessments of the reliability of news sources, providing researchers with "ground truth" for quickly determining which pieces of content contain information from dubious sources as opposed to reliable sources. For example, researchers at the University of Michigan use NewsGuard ratings as sources for their regular "Iffy Index," which measures the quantity of misinformation on the various digital platforms over time. Moreover, NewsGuard's "Misinformation Fingerprints" dataset, which offers a constantly-updated catalog of the top myths spreading online, can be used to match individual pieces of content to specific known false claims. By combining both datasets together, researchers can pinpoint content containing falsehoods with even greater accuracy.

Having licensed our data to dozens of academic and government researchers studying misinformation, NewsGuard would welcome the opportunity to provide its data to a greater number of independent researchers under a wider, government-funded license.

5. *Transition to practice:* How can the Federal government foster the rapid transfer of information integrity R&D insights and results into practice, for the timely benefit of stakeholders and society?

---

[3]https://www.newsguardtech.com/press/newsguard-partners-with-bright-the-new-ethical-social-media-app-to-keep-misinformation-off-the-platform/
[4]https://neeva.com/press/neeva-and-newsguard-team-up

Federal government support for turnkey solutions like NewsGuard's could equip tens of thousands of users with tools to evaluate the reliability and credibility of news websites while browsing the search engine results, social media feeds, or the open internet. NewsGuard is a ready-to-deploy solution with a proven record in schools, universities, public libraries, search engines, and social platforms.

NewsGuard's Red and Green source ratings, trust scores out of 100, and detailed Nutrition Label reviews all serve to inform users as they navigate the internet. As a browser extension, NewsGuard does not interfere with user behavior or block content: instead, it seamlessly provides source level reliability ratings alongside the browsing experience to help users decide whether they want to trust a certain news source.

A wealth of research already supports the effectiveness of NewsGuard's intervention:

- After launching the browser extension, NewsGuard and the Knight Foundation commissioned a Gallup survey to assess how the tool worked when installed on personal computers. The study, conducted in November 2018, found that 91% of respondents found the NewsGuard Nutrition Labels helpful, and 90% generally agree with the ratings and respondents trusted the ratings more because NewsGuard ratings are done by "trained journalists with varied backgrounds." The Gallup researchers concluded: "The positive results among people who accepted Gallup's invitation to download the NewsGuard browser extension suggest a desire for more information about the sources of news people see online, such as in their social media newsfeeds and in their search results. The news source rating tool offers a scalable solution to identify which news sources adhere to the basic journalistic standards of accuracy and accountability citizens expect and deserve."

- In 2017, Indiana University researchers Alan Dennis and Antino Kim ran an investigation comparing different reputation rating formats to assess their ability to influence users' belief of news articles.[5] They found that presenting source reputation ratings directly influences the extent to which users believe articles on social media. In a 2019 article in The Conversation,[6] Dennis and Kim said: "What we learned indicates that expert ratings provided by companies like NewsGuard are likely more effective at reducing the spread of propaganda and disinformation than having users rate the reliability and accuracy of news sources themselves."

---

[5]https://misq.org/says-who-the-effects-of-presentation-format-and-source-rating-on-fake-news-in-social-media.html
[6]https://theconversation.com/rating-news-sources-can-help-limit-the-spread-of-misinformation-126083

- In their article "Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence," published in May 2017 in PLOS One, Stephan Lewandowsky, a psychologist at the University of Bristol, John Cook, a researcher at the Center for Climate Change Communication at George Mason University, and Ullrich Ecker, a cognitive psychologist at the University of Western Australia, explain how, because pre-existing beliefs impact how people respond to novel information, warnings about misinformation are more effective when they are administered before misinformation is encountered rather than after.[7] NewsGuard's intervention of providing source-level evaluations when someone first encounters content can be described as "pre-bunking," supporting the findings of Lewandowsky et. al's study.

6. *Relevant activities:* What other research and development strategies, plans, or activities, domestic or in other countries, including in multi-lateral organizations and within the private sector, should inform the U.S. Federal information integrity R&D strategic plan?

Social media and technology platforms with significant consumer bases and audiences have a duty of care to protect their users from the harms of mis- and disinformation, according to the UK Online Safety Bill.[8] The Online Safety Bill aims to strengthen the UK's stance on internet regulation by holding large tech companies to account for protecting users from online harms and keeping harmful information off their platforms, with major fines and regulatory consequences for firms that fail to comply. An exemplar of a private company that has proactively addressed such harms across its product areas is Microsoft, which has forged a global, company-wide partnership with NewsGuard to safeguard employees, clients, and users from misinformation. Teams across Bing, MSN, Teams, Edge, and the Democracy Forward Initiative have relied on NewsGuard's trustworthiness indicators for news and information websites, setting a precedent for other technology platforms to forge proactive partnerships that safeguard information integrity for users. The U.S. Federal Government is well positioned to capitalize on NewsGuard's solution for both internal research purposes and broader licensing to agencies and associated companies, following the success of NewsGuard's work with private sector stakeholders like Microsoft, as well as public sector researchers and analysts at groups like the Pentagon.[9]

---

[7]https://pubmed.ncbi.nlm.nih.gov/28475576/
[8]https://www.gov.uk/government/news/world-first-online-safety-laws-introduced-in-parliament
[9]https://www.newsguardtech.com/press/newsguard-wins-pentagon-state-department-contest-for-detecting-covid-19-misinformation-and-disinformation/

In Europe, the European Commission is currently convening dozens of signatories for the revised Code of Practice on Disinformation, a self-regulatory instrument that enables technology platforms to collectively set common goals for addressing misinformation and associated KPIs for tracking progress, with self-reporting of progress. A section on "empowering users" outlines how the platforms should provide tools to their users from independent entities using apolitical, transparent criteria to alert them to news sources that publish misinformation. It is the first time that global industry stakeholders have agreed, on a voluntary basis, to self-regulatory standards to fight disinformation by setting a wide range of commitments, from transparency in political advertising to the closure of fake accounts and demonetization of purveyors of disinformation.[10] As the U.S. moves closer to formalizing legislation that holds technology platforms accountable and protects internet users from online harms, the EU's self-governing frameworks are a useful model to consider.

There also exists the opportunity to consider legislative mechanisms that promote media literacy education. Numerous U.S. states have recently considered or enacted different laws requiring media literacy education in schools, with Illinois being the first state to require a unit of media literacy education as a prerequisite for graduating high school.[11] Fourteen other states currently have some standards for media literacy, but none require a unit of instruction. Elevating and enforcing standards for media literacy education across the country — and providing resources like NewsGuard to do so — would yield more informed and discerning online citizens for generations to come.

7. *Support for technological advancement:* **How can the Federal information integrity R&D strategic plan support the White House Office of Science and Technology Policy's mission:**

- **Ensuring the United States leads the world in technologies that are critical to our economic prosperity and national security; and**
- **maintaining the core values behind America's scientific leadership, including openness, transparency, honesty, equity, fair competition, objectivity, and democratic values.**

The European Union through its Code of Practice on Disinformation and the UK through its Online Safety Bill have taken the lead on addressing misinformation by encouraging digital platforms to provide news consumers with news-literacy tools. Many online safety tools have been developed in the U.S., and the White House Office of Science and Technology can add its support to this growing industry, including by building

---

[10]https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation

[11]https://www.chicagotribune.com/opinion/commentary/ct-opinion-illinois-bill-media-literacy-fake-news-schools-20220228-p5q2xcuj3zautaedsi4hqfadga-story.html

awareness among the platforms of the benefits to them and their users from making third-party "middleware" tools available to consumers.

Ratings of news sources at the domain level has been proven an effective tool for news consumers, if these ratings are done in an open, disclosed and transparent manner, using criteria that are objective and apolitical. Restoring trust to generally trustworthy sources of news and information supports democratic values and democratic systems. A free and trustworthy press has been at the center of the American experiment since its beginning and must be restored if our democratic system is to be as strong and resilient as the people expect.