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

WITNESS

May 15, 2022

Re: WITNESS RFI Response: Information Integrity R&D (Docket Number NSF - 2022-05683)

WITNESS takes this opportunity to respond to the Networking and Information Technology Research and Development (NITRD) National Coordination Office (NCO) and National Science Foundation (NSF) Request for Information on Federal Priorities for Information Integrity Research and Development.

About WITNESS

WITNESS is an international human rights organization that helps people use video and technology to protect and defend their rights. As a key element of that we ensure that vulnerable and marginalized communities are equipped with the skills, tools and infrastructure to create information with integrity, and challenge misinformation and disinformation. Working across five regions (Asia and the Pacific, Latin America and the Caribbean, the Middle East and North Africa, Sub-Saharan Africa, and the United States) alongside those most excluded or at-risk, our teams identify gaps, design solutions, provide guidance, and co-develop strategies. We then scale this work globally on a systems level, sharing what we learn with communities facing similar issues and advocating the needs of vulnerable and marginalized communities to technology companies and other influential stakeholders to ensure they are translated into policies, governance and solutions. As a critical part of that work over the past decade we have engaged in research, prototyping and action around critical issues in information integrity - including two foci: i) appropriate ways to develop content authenticity and provenance infrastructure and ii) counter emerging forms of visual deception such as malicious deepfakes via solutions grounded in a global, human rights framework.

More context on our work in this area can be found at:
https://lab.witness.org/projects/synthetic-media-and-deep-fakes/ and
https://lab.witness.org/ticks-or-it-didnt-happen/

WITNESS is also a member of the Coalition for Content Provenance and Authenticity
(C2PA) where we support efforts to promote a human rights framework in the design of
these specifications for content provenance and information integrity. WITNESS is the
co-chair of the Threats and Harms Taskforce where it leads a harms, misuse and abuse
assessment. This submission is not in our capacity as a C2PA member, however we
reinforce aspects of the C2PA framework below.

*1. Understanding the information ecosystem: There are many components, interactions,
incentives, social, psychological, physiological, and technological aspects, and other
considerations that can be used to effectively characterize the information ecosystem.
What are the key research challenges in providing a common foundation for
understanding information manipulation within this complex information ecosystem?*

WITNESS supports research focused on two elements that are under-researched in the
current environment.

The first is focused work on the prevalence and appropriate approaches to handling
audiovisual information manipulation challenges. Existing research has a heavier focus
on text, while video and audio continue to grow in prominence as vectors for information
manipulation.

The second area is a focus on understanding how information manipulation affects
vulnerable and marginalized communities in the US and globally and how interventions
are assessed for their proportional or disproportionate impact on vulnerable
communities. This area includes understanding how information integrity interventions
that apply to US-based platforms – or are developed in US contexts – function in a
diverse set of global geographies and contexts with different standards of rule-of-law,
adherence to human rights and attention from social media platforms.

*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?

WITNESS recommends that the following be prioritized as a methodological approach to preserving information integrity and mitigating the effects of information manipulation amid a range of approaches that includes literacy, tools and architectures/standards for understanding manipulation as well as content integrity, authenticity and provenance.

1) Research that from early-on centers a human-rights based approach to understanding potential harms, misuses and unintended consequences of interventions, particularly as they impact vulnerable and marginalized populations. From our context as members of the C2PA, this coalition's efforts provide a good example of early work to identify [potential harms](#) that may come from these specifications, especially to those that may be marginalized and particularly vulnerable to information manipulation. Additional research on information interventions is critical that centers consultations with individuals, communities and institutions likely to be impacted by interventions, and to place emphasis on those who already face similar systemic harms. Broad, ongoing, multi-disciplinary consultations are a necessary basis to understand how information manipulation affects different demographic groups, and how standards and practices may avert harm and open up new opportunities to tackle mis- and disinformation and to raise the trustworthiness of authentic content.
2) As noted above, understanding of the global implications of information interventions that are unevenly applied.
3) Research into strategies for improving access to emergent solutions to information integrity issues that inherently require access restrictions (e.g. for cybersecurity and efficacy reasons) but also contain an approach to ensure that the most vulnerable have access to protections. An example is in the area of deepfakes detection where WITNESS has identified a looming problem (see https://blog.witness.org/2021/07/deepfake-detection-skills-tools-access/) for media and civil society globally, and where controlled access to tools must be balanced with global equity and need.

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?

WITNESS has identified a more transparent information ecosystem as one of the frameworks that can increase public awareness about manipulated information and facilitate educational processes, provided that this transparency can be achieved in an opt-in manner that facilitates privacy and recognizes a diverse set of global contexts for how individuals choose to share information and the regulatory and legal contexts in which they share information.

Although transparency can be understood and expressed in a myriad of ways, WITNESS has been exploring three options: the disclosure of provenance information (such as the C2PA), the inclusion of forensic traces (particularly for new forms of synthetic media where photorealistic manipulation is increasingly hard to spot) and labeling of information as it circulates. Of those, we discuss the first one here.

The same questions mentioned above could be reformulated in the context of provenance information as a potential solution: How could provenance information help increase public awareness of information manipulation? What challenges should research focus on to support the development of effective educational pathways with provenance information?

To answer these questions, there is a need to continue researching how provenance tools and standards could be made accessible to as many people, communities and entities that may need it, all the while ensuring their privacy and other human rights.

There is also a need to develop prototypes that cater to different potential content creators in order to understand what would spark adoption, what would promote privacy-oriented designs and usages, and what are the emerging threats and harms. There is also the need to research how consumers process provenance indicators, and what are the expressions that promote trust and what are those that could undermine the trustworthiness of authentic content.

WITNESS provides further insights into research dilemmas and questions in our 2019 research report 'Ticks or It Didn't Happen: Confronting Key Dilemmas in Authenticity Infrastructure for Multimedia' (https://lab.witness.org/ticks-or-it-didnt-happen/)

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?

Inclusion more broadly in prioritization of potential threats and solutions for information integrity R&D is critical. WITNESS has pursued a deliberate multi-stakeholder process

in our own work that focuses on centering a mix of lived, expert and technical expertise in relation to current and emerging information integrity issues to ensure that information integrity risks and challenges are understood in context, and solutions are prioritized within R&D based on a diverse set of inputs.

In the context of standards development bodies (SDOs), participation is restricted and they often lack the diversity of knowledge, experiences and identities that should inform the design of specifications that would have a determining role in any ecosystem. This is to a certain degree expected, since the work and discussions in these bodies tends to be of a highly technical nature. One remedy for this is to actively seek to 'translate' technical work into more accessible information aimed for broader audiences. As one example of proactive action in this regard, C2PA has made efforts to this end by publishing an 'Explainer' alongside the specifications.

Looking towards the future, more of these 'translations' are necessary, accompanied by committed efforts to socialize this information and to continuously engage with as many stakeholders as possible.

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?*

The Federal government can facilitate transfer of R&D insight into practice by active engagement with a diverse range of stakeholders designing and implementing interventions at a local level in the US as well as globally. The prioritization of vulnerable and marginalized stakeholders in this work is both important from a harm reduction perspective and ethical perspective, but also likely to lead to the most effective long-term impact.

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?*

WITNESS notes the analysis in the recent UNESCO Broadband Commission report to which we contributed that highlights critical questions around information integrity and misinformation/disinformation questions from a freedom of expression human rights perspective: 'Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression' https://www.broadbandcommission.org/publication/balancing-act-countering-digital-disinformation/ and also notes the recent/upcoming work by the UN Special Rapporteur on Freedom of Expression in this area.

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.

WITNESS has no comments on this area.