CCC
Computing Community Consortium
Catalyst

The Computing Community Consortium's Response to PCAST Working Group on Generative AI Invites Public Input

July 24, 2023

Written by: Maria Gini (University of Minnesota), Madeline Hunter (Computing Community Consortium), Sven Koenig (University of Southern California), Daniel Lopresti (Lehigh University), Rajmohan Rajaraman (Northeastern University), Ufuk Topcu (The University of Texas at Austin), Matthew Turk (Toyota Technological Institute at Chicago), and Holly Yanco (University of Massachusetts Lowell).

This response is from the Computing Research Association (CRA)'s Computing Community Consortium (CCC). CRA is an association of nearly 250 North American computing research organizations, both academic and industrial, and partners from the professional societies. The mission of the CCC is to bring together the computing research community to enable the pursuit of innovative, high-impact computing research that aligns with pressing national and global challenges.

In our response, we promote and emphasize the role of academic research in addressing risks associated with generative AI. Solutions will require multidisciplinary, multi-pronged approaches that simultaneously implement solutions in the form of policies and regulations, technological advances, and education. These will necessitate transparent communication among the various stakeholders, the use of basic computing research, major contributions and collaborations from social sciences, and a determined, well-rounded approach that will take time.

The academic research space provides for these needs with its unique capabilities to promote long-term research, interdisciplinary solutions, and unbiased advancements. This is not to say that academic research is the only component, it is only a piece of the puzzle, and requires collaborations and input from industry and the government.

Questions:

1. **In an era in which convincing images, audio, and text can be generated with ease on a massive scale, how can we ensure reliable access to verifiable, trustworthy information? How can we be certain that a particular piece of media is genuinely from the claimed source?**

Rather than aim for concrete proof that information is trustworthy, which may be an unrealistic goal, we prefer to focus on providing tools to navigate this inherently untrustworthy environment so that citizens can form their own informed judgments. Achieving this requires a multi-prong approach focusing on routinely providing full provenance of the media data (and its components) and teaching consumers how to evaluate such provenance.

Full provenance is identifying all sources and entities that have contributed to the production of a photograph, video, or news article, and how it was edited or composed. For example, with the publication of a photograph, provenance would require an explicit label stating who or what created the image, who edited or altered it (and how), and who published it (and when). This would require clear disclosure from any application or system that utilizes generative AI, including warnings (where appropriate) that the information may contain errors and may be misleading.

Achieving this could be done in part by leveraging existing laws and regulations surrounding copyright, making them more robust and applicable to current technologies such as generative AI. Additionally, these regulations and policies could be enforced through certifications and tagging. For example, Facebook or YouTube would only play videos that have an appropriate certification that clearly provides provenance of a piece of media or text. A requirement such as this would yield a combination of hardware techniques, cryptographic techniques, software techniques, etc., that protects the owner of the intellectual property (e.g., the photographer) and privacy, and discloses data authenticity.

Providing citizens with this information would allow them to make better informed choices based on what is authentic and what has been tampered with. Similar to the trustworthiness of news sources with different perceptions of reliability, citizens will have the tools to decipher between more trustworthy sources and less trustworthy sources. Further research is needed to determine what kinds of provenance are most useful and how to effectively present such information.

Current tools for tasks such as identifying disinformation and for detecting whether text was produced by generative AI models are unreliable, and short- and long-term research is needed to improve them and identify more reliable approaches. The first step to a solution is thus investing in technologies and promoting regulations and policies, but these efforts will have very little impact unless all stakeholders are properly educated about the possible issues and limitations, what to expect, how to consider (and have appropriate skepticism about) media, and their sources. Among other things, the public needs to understand AI technologies, their strengths, and their limitations much better than is currently the case.

2. **How can we best deal with the use of AI by malicious actors to manipulate the beliefs and understanding of citizens?**

Understanding how technologies can manipulate citizens' thought processes and beliefs requires an interdisciplinary and holistic research effort between computing and social science.

On top of this, the aforementioned mechanisms, such as transparency, provenance, certification and technological advancements, will help to mitigate the risks and likelihood of manipulation. This can be furthered and emphasized by explicit warnings to media consumers when appropriate. This type of implementation will require an educational piece that teaches media consumers how to use the information they are given and which things to watch for, such as warning labels and certifications. This education should target the public at large, as well as permeate every level of education starting as early as grade school.

Aside from the prevention aspect, there should be some sort of accountability for the companies that deliver AI-generated content. The recent ACM GenAI document[1] suggested that providers (all entities that deliver genAI technologies, components, systems, or applications to users or other entities) should undertake extensive impact assessments prior to the deployment of such technologies to thoughtfully ensure that the benefits to society of any such deployment outweigh its risks. They should also provide sufficient information about such systems to permit expert evaluation of their risks and impacts.

3. **What technologies, policies, and infrastructure can be developed to detect and counter AI-generated disinformation?**

We have various options for mitigating risks associated with AI-generated disinformation, including education, policy, regulation, incentives, and technology. Each may have positive and negative implications. We need to implement these options holistically. In order to do this, all stakeholders — academia, industry, government,

[1] https://www.acm.org/binaries/content/assets/public-policy/ustpc-approved-generative-ai-principles

community members — need to come together and declare dealing with disinformation, particularly from generative AI, a national priority. There needs to be an open channel of communication and resources among all entities.

Producers of technologies that proliferate AI-generated disinformation in society must be held accountable for adequately assessing their own risks and impacts before implementing these technologies. As suggested in the recent ACM GenAI document, providers of generative AI systems should create and maintain public repositories where errors made by the system can be noted and, optionally, corrections made. This will ensure accountability and transparency between stakeholders, as well as, potentially helping future providers know what to look out for in systems being developed.

There also must be continued support in research into deep fake detection and multimodal alignment detection. Additionally, the development of auditing capabilities for models, algorithms, and data (and making such information publicly available) will help in giving citizens the tools to determine verifiable information, promote accountability and mitigate risks with generative AI disinformation.

4. **How can we ensure that the engagement of the public with elected representatives—a cornerstone of democracy—is not drowned out by AI-generated noise?**

This is not a new issue - there are already many sources partaking in political discussions that provide misinformation and noise without the help of generative AI. Some mechanisms and tools are already in place, such as the public's recognition of a ".gov" URL as a reliable source, that can continue to help.

It will be important to make official government websites easy to use and accessible to all, with all forms of communication (text, video, audio, etc.) providing necessary provenance in both directions (to and from the representatives). With such trustworthy sources of information, the public will be incentivized to go there when they want to know something, or when they have doubts about what they've heard from some other source. These official sources could also be marked with the use of watermarking/digital signatures on communications from and websites of elected officials.

5. **How can we help everyone, including our scientific, political, industrial, and educational leaders, develop the skills needed to identify AI-generated misinformation, impersonation, and manipulation?**

In order to enable citizens to use the tools provided by technological and policy advancements, the message that media can be and is oftentimes misleading must be reinforced and permeate all levels of education (K-12). For example, education is required for consumers to learn how to read text differently in the age of generative AI since generative AI generates text that appears convincing to humans, despite it containing factual and reasoning mistakes. To further the example, teachers should not draw conclusions on the correctness of homework submitted by students purely from its grammar. Similarly, AI-generated photographs and videos such as deep fakes can be seemingly authentic and convincing. This requires a more careful analysis and checking of text and media than has often been done before. Changing this behavior and adopting a more careful analysis of media and text should ideally happen early during the education process. Similarly, the social norm of looking for labels, warnings and verifiable sources must be as ingrained in society as looking both ways before crossing the street.

While some things might be too time-consuming and rigorous to teach to everybody on all levels of education, such as probability and statistics, critical thinking skills related to what you are reading and seeing can easily be taught in schools and as public message campaigns. Understanding the inner workings of AI and statistics is a large undertaking, that many do not have the desire nor the time to learn. Broad concepts and topics that do not require a deep knowledge of anything complicated are easier for people to internalize and consider common knowledge. For example, teaching something like “if you see a video, question it” would not require extensive education, but in theory has the ability to change behavior. Similarly, it is important to teach people how to use the tools that generative AI provides in a responsible way, rather than disallowing their use. It is vital to give people the opportunity to understand and be literate in AI.

The AI research community could partner with our colleagues in education to help develop the materials so that everyone in our society understands the power and the limits of generative AI. This will include instilling in students a healthy sense of skepticism when they see media that might not be trustworthy and could have been fabricated by an AI to manipulate them.