Submission to the NTIA AI Accountability Policy Request for Comment from the Climate Action Against Disinformation coalition, June 9, 2023

Why the AI Accountability Policy Should Require Product Safety

Climate Action Against Disinformation (CAAD) is a coalition of 20+ leading climate and anti-disinformation groups committed to combating disinformation about climate change. We thank you for soliciting comments on concerns about generative artificial intelligence (GAI). We agree with the Federal Trade Commission Chair that artificial intelligence risks turbocharging fraud, and we agree with the Blueprint for an AI Bill of Rights that its algorithms "have been found to reflect and reproduce existing unwanted inequities or embed new harmful bias and discrimination." They also risk amplifying climate disinformation.

Without robust regulation, the unrivaled ability of AI to create credible disinformation on an unprecedented scale represents a clear danger to authentic climate information and thus creates a significant obstacle in taking the climate action that science tells us we must to protect humanity and the planet. In the comment below, we recommend five regulatory principles in order to guide the administration's approach to regulating AI, with the primary goal that **the AI Accountability Policy should require product safety before AI enters use,** so that it does not increase the spread of climate disinformation.

The AI threat to climate disinformation

CAAD and other researchers have documented extensive examples of the harms of climate disinformation on social media in the U.S. that we believe could each be further worsened by the onset of GAI, including: an overall rise in the amount of climate disinformation, early failures within ChatGPT4 to not produce climate disinformation, algorithmically enhanced lies that falsely blame oil & gas infrastructure failures on wind power, the monetization of climate disinformation, false claims linking wind power and whale deaths, and the algorithmic prioritization of climate denial narratives during global climate negotiations.

For decades, the United States has required a broad range of companies to prove their products are safe before entering use, and GAI should be no different. Pharmaceutical companies must conduct clinical trials. Vehicles are tested by the National Highway Traffic Safety Administration. Planes are certified by the Federal Aviation Authority. The Nuclear Regulatory Commission oversees the safety of nuclear power plants. GAI represents as large a risk as previous technologies, and should be similarly addressed.

We recommend the United States government assess generative artificial intelligence with **a systems-wide approach to the health, integrity, and resilience of the information ecosystem,** including any impacts that could increase disinformation on climate change.

Avoid repeating the mistakes of social media

GAI's sudden universal accessibility to the public represents the second time this century that technology will usher in widespread social change. In the mid-2000s, social media was introduced in a largely unregulated manner, with far-reaching negative consequences that we see widely today. Services soon became monetized through invasive surveillance and tracking of their users. The AI Blueprint notes that social media "data collection has been used to threaten people's opportunities, undermine their privacy, or pervasively track their activity—often without their knowledge or consent." The Federal Trade Commission believes many companies broke the law in their attempts to restrict competition. Artificial Intelligence has already contributed to discriminatory outcomes when it has been used in the areas of policing, housing, and employment.

But social media's trajectory was not inevitable; rather, it was shaped by a broad range of policy choices and federal and congressional inaction. Artificial intelligence should not be the next in a string of Silicon Valley products that is allowed to "move fast and break things", especially when those things include planet Earth and its 8 billion people. Policymakers must use this opportunity to not repeat the mistakes of the past, especially with a technology that is far more powerful than social media. The extraordinary ability of GAI to tailor-make and target disinformation to individual users raises concerns over privacy, and potentially massive disinformation generation campaigns to permanently subvert essential, science-based discourse about the imperative to take climate action.

The energy and climate impact of AI

Beyond the danger that AI disinformation presents, the creation and use of large language models is itself a climate danger. GAI systems demand a lot of energy, at a time when the world must dramatically reduce carbon output in order to prevent world temperatures from rising more than 1.5 degrees Celsius. Already, Google's servers alone used 15.5 terawatt (a trillion watts) hours of electricity in 2020—that's a 22% increase from 2019. Adding AI to its search engines will increase this by four or five times. For example, the carbon emitted to train a large AI language model is equivalent to driving from San Francisco to New York 550 times.

Regulatory recommendations

Given its scope and proven dangers, Artificial Intelligence should be subject to similar regulation as most other American industries, from airlines to pharmaceuticals. As we have seen with social media, the window to regulate is before new products enter widespread use—not after. This is especially true given AI's potential for self-improvement post-release, including its ability to spread climate disinformation more effectively. We note that proposed AI regulations in China–the often discussed global competitor supposedly driving American companies to accelerate GAI development–would require companies to provide a safety assessment before AI enters public use.

For this reason, we urge that the AI Accountability Policy require the following:

1. Safety be proven before release, including through these steps:
    a. GAI companies' release of transparent plans that identify and prevent harm, including the spread of climate disinformation and other harms, before product release.
    b. Vigorous safeguards against mass producing disinformation, fraud, and hate designed to manipulate human emotions.
    c. Explanation of how GAI models produce their information, measure their accuracy, and show their sourcing.
    d. Adherence to community content standards that include detecting misuse of AI and enforcing standards against its misuse.
    e. Assessments of the health, integrity, and resilience of the information ecosystem, conducted by a multi-agency task force that considers how the proliferation of GAI might further erode trust in science, harm young people's mental health, increase hate speech, and accelerate the power of digital gatekeepers, as well as how our regulatory systems could adapt to new GAI systems and potential threats before release.
    f. Assessment of threats to individual privacy.
2. Company GAI systems are transparent, including through these measures:
    a. Regular reporting on existing and potential harms, which should extend to updates and changes in product design.
    b. Allowing researchers and academics to access how the technology functions and is used.
    c. Publishing a description of the dataset or corpus used to train the GAI model, explaining any principles used for including or excluding source materials. And publishing any principles, constitutional directives, or guardrails used during training to align the models with human values.
3. Companies obtain consent for using copyright-protected images or community-owned data for large language model training. Government develops rules for "fair use" in cases where GAI models are summarizing copyrighted texts, that should include hyperlinked citations back to the source materials.
4. Companies and their executives are held accountable and liable for their products' harms, recognizing that GAI should not receive any liability protection offered by Section 230 of the Communications Decency Act, and that the public has recourse against companies that spread disinformation by artificial intelligence.
5. GAI systems publish regular energy usage reports for their systems' overall energy use and for aggregated individual energy use.
6. That until these essential protections are in place, a federal moratorium is placed on any new public GAI deployment.