Legal Arguments in Support of Third-Party Risk Audits

In an era of advancing technology, protecting young users from potential harms arising from social media use is paramount. Effective laws require understanding the role algorithms play in perpetuating harms on social media and should incorporate enforcement like algorithm risk audits.

What is an algorithm risk audit (ARA)?

An algorithm risk audit (ARA) helps us see how recommendation, or engagement-based, algorithms are designed to tailor a social media feed to automatically show different content to different users. The audits act as a safety check to:

- Provide transparency about the effects of algorithms on content feeds,
- Objectively compare how algorithms are designed to distribute content to users,
- Identify biased content distribution practices to hold platforms accountable.

What are some real-world examples of ARAs?

In a settlement between Meta and the US Department of Justice, Meta must address the discriminatory aspects in its algorithm design to promote housing ads equitably across race and gender groups.

Does an ARA violate the First Amendment?

Steps 1 and 2 of an ARA (identifying and measuring harms) do not infringe on the First Amendment because they do not regulate content posted on social media. To address potential constitutional concerns, the third step (reducing harms) can be optional while requiring harm identification and measurement remains mandatory. Results of the audits can be used by law enforcement to bring deceptive advertising or unfair business practice claims against social media companies.

How does an ARA work? It will:

- Identify Specific Harms: Define the harms to be measured, e.g., pro-eating disorder content.
- Measure Extent of Harm: Choose how to measure the harm caused by a biased distribution of content, e.g., number of young users (versus adult users) within a week who are prompted by algorithms to view content promoting extremely restrictive dieting plans.
- Reduce Harm: Legislative body and social media platform agree on benchmarks and reporting periods to mitigate identified harms, e.g., quarterly or monthly report.

Algorithm risk audits would be conducted by independent third-party reviewers at regular intervals to ensure impartiality. Results of these audits must be made publicly available to discourage the business practice of designing algorithms that promote extreme content in a biased way. Review the full legal rationale here.

Action Steps

Community Members: Propose to your state representatives to champion the *Social Media Algorithm Accountability Act*.

State Policymakers: File the *Social Media Algorithm Accountability Act* and work with your colleagues to pass it into law.

Have questions about our Roadmap or model legislation? **Contact us: striped@hsph.harvard.edu**