

A Trauma-Informed Approach to Auditing Algorithmic Decision-Making Systems

Professor Renée Cummings, Assistant Professor of Data Science, University of Virginia (UVA), School of Data Science, first Data-Activist-in-Residence, School of Data Science, UVA, Co-Director, Public Interest Technology University Network (PIT-UN), at UVA, Nonresident Senior Fellow, Governance Studies, Center for Technology Innovation, Brookings Institution, and Inaugural Senior Fellow, AI, Data, Public Policy, All Tech Is Human

Written Statement to the U.S. Senate AI Insight Forum: Risk, Alignment, and Guarding Against Doomsday Scenarios

December 6, 2023

Majority Leader Schumer, Senators Rounds, Heinrich, and Young, other distinguished Members of the Senate, I thank you for the invitation and opportunity to participate in this AI Insight Forum on Risk, Alignment, and Guarding Against Doomsday Scenarios.

There is no questioning the extraordinary, transformative power, and promise of AI. With AI we can reimagine, recreate, redefine, reposition, and realign ourselves and society to embrace some of the world's greatest challenges and deploy some of the most creative solutions for a resilient, sustainable, equitable, and just future.

However, in a world increasingly shaped by AI and powered by algorithms, how do we manage AI risks, the unforeseen abilities, the "hallucinations" and disturbing behaviors of algorithms, that often "emerge spontaneously without explicit programming"? How do we prepare for and mitigate the large-scale risks and doomsday scenarios before they arise and materialize? In a world of AI monetization and AI optimization how do we respond to the weaponization of AI and algorithmic oppression?

As we think about AI out-performing humans, and the challenges of ethical and responsible AI, the challenges of compliance, governance, legislation, regulation, enforcement, and redress, we are also thinking about risk mitigation, risk management, crisis management, contingency planning, and business continuity plans for AI business models and algorithms that could go rouge in real time.

How do we "do no harm" or do the least harm possible? How do we prevent wanton recklessness, misuse, and abuse of AI, and other potential dangers? What are our collective responsibilities? How do we protect civil liberties, civil rights, and human rights? How do we protect democracy, decision-making, and due process, agency, autonomy, self-determination, self-actualization, and identity development in the age of AI? How do we ensure equity, fairness, and justice, uphold human dignity, and celebrate diversity and inclusion in the age of AI?

As we imagine co-creating a future through collective augmented intelligence, we continue to witness the amplification of social injustices through algorithmic decision-making systems, the erosion of social stability through disinformation, deception, media manipulation, and deepfakes, diluting reality, truth, and trust.

We cannot overlook Generative AI's amplification of risks such as child sexual abuse, gender-based violence, harassment, hate speech, and extremism. While we continue to examine AI's ability to exacerbate abuse, conflict, threat, and violence, and AI's ability to manufacture doomsday scenarios, one word has been missing from the discussion of AI and Generative AI. Trauma and how data trauma deployed through algorithms impacts our brains, bodies, families, and communities.

Critical thinking around a trauma-informed approach to algorithmic auditing, red-teaming, stress testing and AI risk management offers an interdisciplinary, collaborative perspective that stretches the imagination of risk mitigation, risk management, impact assessments, alignment, and AI governance.

DATA TRAUMA

Increasingly, we are seeing more and more casualties of algorithmic injustice and increasing cases of data trauma attached to algorithmic injustice and institutional betrayal. Bias. Discrimination. Marginalization. Profiling. Victimization. All AI generated; all data driven. Ethical, responsible, and trustworthy AI must advocate for a trauma-informed approach to building and deploying AI and auditing algorithmic decision-making systems, conducting impact assessments and vulnerability audits.

A trauma-informed approach is a real time response and protective intervention, as we are running out of time, to reduce disparities, protect against bias, discrimination, and unfairness in AI. A trauma-informed approach to algorithmic auditing engages diverse and inclusive perspectives, interdisciplinary thinking, and expands and reimagines the range of experts, stakeholders and community partners involved in the ethical, equitable and responsible design, development, deployment, procurement, and adoption of AI and how we think about data...the lifeblood of AI.

There is no AI without data. Data sets do not have the luxury of historical amnesia. Therefore, we continue to see the intergenerational transmission of data trauma through surveillance technologies that track, trace, terrorize and traumatize and algorithmic decision-making systems that deny access, opportunities, resources and services to impacted, marginalized, vulnerable, high needs, underserved and communities.

GENERATIVE TRAUMA

Generative Trauma examines how Generative AI can (re)traumatize, (re)victimize, disempower, and disenfranchise, not only vulnerable groups but society at large, at scale, and with unimaginable speed.

A TRAUMA-INFORMED APPROACH

The goal is to bring a trauma-informed approach that reimagines and redefines how AI and Generative AI are designed, developed, deployed, adopted, procured, used, and audited. Promoting trauma-informed data curation and redaction, enhancing accuracy, accountability, auditability, transparency, explainability and fairness, through trauma-informed algorithmic auditing can force a rethinking of the meaning of duty of care and encourage duty of repair, heightened due diligence and ethical vigilance, across domains and disciplines.

By engaging stakeholders in critical public conversations on Data Trauma and Generative Trauma and the multimodal traumatic impacts of AI, we can scrutinize, more effectively and efficiently, the traumatic impacts of algorithms and the intergenerational trauma trapped in the memory of historic data sets.

This approach to knowledge production and policy development will also introduce technology to the core research and theory of poly-victimization and the negative and potentially lifelong biopsychosocial impacts of AI related trauma.

As we examine the trauma of AI systems, Data Trauma and Generative Trauma, we can start to explore, more deeply, Generative Terror, the distortion of reality, extremism, the apocalyptic imagination, and fantasies of cataclysmic destruction, all part of the doomsday script.

A trauma-informed approach to algorithmic auditing will also ensure the legacies we fought so hard to build are not erased or destroyed by an algorithm and dreams are not deferred by data.

Thank you.