Lay Stakeholder Centric Sociotechnical Mechanisms For Addressing the Impacts of Generative AI

Julia Barnett
Northwestern University
juliabarnett@u.northwestern.edu

Introduction

People today have little agency to reduce the harms they are experiencing from generative AI (genAI) systems, some of which exist even if they do not actively use them. Some *micro*-scale harms (e.g., copyright infringement from uninformed creation) affect users choosing to engage with genAI tools, while *macro*-scale harms (e.g., misinformation) affect everyone regardless of whether they opt to use these systems. In the U.S., one of the few existing mechanisms to address these issues is through the courts, which ultimately decide how genAI creators must operate.

Past literature in risk and algorithmic assessment has focused on expert-based mechanisms to evaluate and mitigate the harms experienced by laypeople, and typically excludes them from this process. These lay stakeholders are people who possess no formal professional nor technical expertise, but possess a form of situated and lived expertise through the act of experiencing the impacts of genAI on the ground floor. This lay expertise can inform risk management in a manner that is currently non-existent in current structures. This work will ultimately contribute to the larger discussion of impact assessment and mitigation of generative AI, arguing for greater inclusion of lay stakeholders in all stages of the process. It will approach all aspects of genAI risk management from an inherently sociotechnical perspective, because harms resulting from genAI are an "outcome of entangled relationships between norms, power dynamics, and design decisions" (Shelby et al. 2023). Ignoring these entanglements leads to ineffective mitigation..

As detailed in Figure 1, this work proposes three mechanisms to provide more agency to the lay stakeholders experiencing harms resulting from genAI systems through three distinct levers: (1) a **technological mechanism** enabling users of genAI technology to resist harms on a micro-scale domain, e.g., within the scope of generative music models, (2) a **policy mechanism** operating on a macro-scale integrating lay stakeholders into the policy process at an early stage of design, and (3) a **judicial mechanism** on a meso-scale assessing where lay stakeholder expertise can provide valuable insight into the court case process assessing the harms inflicted by genAI companies.

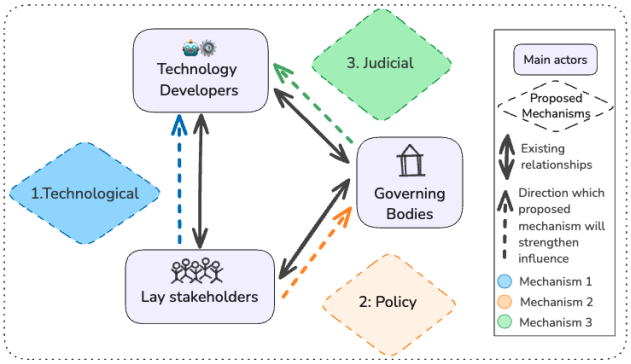

Figure 1: Main framework detailing a relationship concerning genAI between technology developers, lay stakeholders, and governing bodies. All three actors have bidirectional relationships of influence, though these relationships are not balanced. The three proposed mechanisms in this work aim to add balance to these relationships, with the (1) technological mechanism giving more agency to lay stakeholders in regard to technology developers, (2) policy mechanism giving more agency to lay stakeholders in regard to governing bodies, and (3) judicial mechanism strengthening the influence governing bodies have over technology developers.

Mechanism 1: Technological

At the *micro*-scale, (individual domains with scoped harms), I propose technical design intervention to give lay stakeholders more agency to resist harms. The lay stakeholders in this case are users of the technology; they have situated expertise but little agency to resist harms. For this mechanism I focus my case study on the generative music domain. I conducted a systematic literature review in which I identified a set of harms specific to the generative audio domain for both music and speech models (Barnett 2023). The studies presented in this work focus on two of those harms: the potential for copyright infringement and cultural appropriation. These studies identify the root cause as uninformed creation and propose a data attribution mechanism for generative music models.

Study 1 (Barnett, Garcia, and Pardo 2024), published at ISMIR in 2024, establishes an easily replicable methodology and framework to perform training data attribution for a generative music model, which has been validated in a

human-listener study and explored against robustness for audio perturbations. Study 2 (in progress) creates an interface built upon the technology established and tested in Study 1 and conducts experiments with human users to understand how this tool can help educate musicians and non-musicians about their "new" creations using generative music tools.

RQ1 *Study 1*: How can we systematically identify similar pieces of music from training data to new generations in a manner that is useful for understanding training data attribution of newly generated music audio?

RQ2 *Study 2*: To what extent does using our interface make users more informed about their creations from generative music models?

Mechanism 2: Policy

At the *macro*-scale(impacts on broader citizenry),I propose integrating lay stakeholders into the policy process to give them more agency. The lay stakeholders in this case take the form of people impacted by the use of genAI technologies by others. Here, I focus on harms to the U.S. media ecosystem. Study 3 (Barnett et al. 2025) utilizes lay stakeholder input to map mitigation and prevention strategies of AI impacts by assigning various actions to stakeholders, dubbed "stakeholder action pairs" (SAPs). We then query these lay stakeholders for prioritization and importance of the SAPs, and finally convert this information into a format useful to inform policy processes. Study 4 (in progress) transforms these SAP rankings into an optimization problem and seeks to identify optimal sets of SAPs that result in the greatest perceived mitigation of the negative impacts.

RQ1 *Study 3*: How can we systematically leverage lay stakeholder input to inform the policy process in a manner that allocates responsibility to various actors and guide the prioritization of these measures?

RQ2 *Study 4*: How can we transform this lay stakeholder input into a simulation study that produces optimal sets of allocated actions to mitigate the impacts of generative AI?

Mechanism 3: Judicial

One of the existing primary mechanisms to address genAI harms in the United States is through the courts. These harms includes those resulting from alleged copyright infringement and use of creative work to train models and then subsequently compete with those creative works through their outputs. Copyright law in particular has been shaped over many decades to adapt to new technological developments. This mechanism explores how to improve judicial decision-making in genAI copyright cases by enhancing the information available to judges and litigators, including the integration of lay stakeholder perspectives. Case studies of past tech-related copyright disputes will reveal how courts gather evidence, frame judgments, and where gaps remain. By analyzing these cases, we aim to develop a clearer framework for the types of information that courts need today in order to make more informed decisions regarding generative AI. This chapter will result in a clearer understanding of what researchers in both law and computer science need to do in order to support the judicial process surrounding these monumental cases:

RQ1 *Study 5*: What information needs to be obtained in order to make informed judicial decisions about whether generative AI companies infringe upon copyright during both training and outputs; more specifically what would determine whether these uses constitute fair use?

RQ2 *Study 5*: How can we integrate lay stakeholder input into this judicial information gathering process in order to provide a fuller picture of the impacts upon those whose copyright may have been infringed?

Contributions

This work contributes a grounded and participatory framework for assessing and mitigating genAI harms by redistributing agency toward those most affected across technological, policy, and judicial dimensions. Current AI risk mitigation frameworks are expert-driven and fail to meaningfully include lay stakeholders: people directly experiencing harms from genAI who lack technical or institutional power and expertise. This work establishes a framework that affords these lay stakeholders more agency by embedding them meaningfully into technical, policy, and judicial impact assessment and mitigation processes.

Different scales of harm require different solutions, illustrated through case studies at each level. Each of my chapters proposes a tangible product that can be used by researchers and non-researchers alike to mitigate harms from genAI. The three main themes central to this work are the need for impact assessment of genAI to be (1) sociotechnical, (2) participatory, and (3) actionable. Through specific case studies responding to mapped harms, I demonstrate different forms genAI impact assessment can take in different settings. I illustrate how this can manifest while simultaneously addressing harms society is actively experiencing due to generative AI.

References

Barnett, J. 2023. The ethical implications of generative audio models: A systematic literature review. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 146–161.

Barnett, J.; Garcia, H. F.; and Pardo, B. 2024. Exploring musical roots: Applying audio embeddings to empower influence attribution for a generative music model. In *Proceedings of the 25th International Society for Music Information Retrieval (ISMIR)*.

Barnett, J.; Kieslich, K.; Helberger, N.; and Diakopoulos, N. 2025. Envisioning Stakeholder-Action Pairs to Mitigate Negative Impacts of AI: A Participatory Approach to Inform Policy Making. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*.

Shelby, R.; Rismani, S.; Henne, K.; Moon, A.; Rostamzadeh, N.; Nicholas, P.; Yilla-Akbari, N.; Gallegos, J.; Smart, A.; Garcia, E.; et al. 2023. Sociotechnical harms of algorithmic systems: Scoping a taxonomy for harm reduction. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 723–741.