Community-led AI Audits: Methodology for Placing Communities at the Center of AI Accountability

An Extended Abstract

Authors

Gemma Galdon-Clavell, Founder of Eticas

Yung-Hsuan Wu, Algorithmic Auditor, Eticas Foundation

Greta Byrum, Board Member, Eticas Foundation

As algorithmic and AI systems[1] proliferate worldwide, we are just starting to learn about their impact. Not enough is known about the ways in which vulnerable and marginalized communities feel the disparate impacts of these systems, nor how communities and oversight authorities might prevent risks and harms that accompany the use of AI tools and systems on the ground. The Eticas Foundation has years of experience working with civil society organizations (CSOs) and communities in investigating algorithmic and AI injustice, with the aim of increasing transparency and accountability in these systems while empowering affected communities to voice their concerns and take action to demand accountability and redress. We learned that even seasoned organizers, advocates, and grassroots leaders leading important civil rights struggles are often unfamiliar with the specifics of AI technology, which creates a barrier for them to describe, evaluate, and challenge the negative impacts these systems impose in their expert domains. To tackle this conundrum, Eticas developed the methodology of community-led audits (CLAs) to explicitly center and supports community leadership and expertise in resisting AI companies' dominance and pushing for AI accountability.

1. Algorithmic auditing as an accountability instrument

Algorithmic auditing is an instrument for the dynamic appraisal and systematic inspection of algorithmic and AI systems with regard to their performance and external impacts. Auditing as a practice is well-established in aviation, finance, accounting, and information security industries where evidence in the security and functionality of complex systems is much needed; AI researchers and practitioners have since borrowed learned lessons and adapted auditing practices to the context of algorithmic and AI systems.[2] Conducting

---

[1] AI system here refers to software which generates outputs for a given set of objectives such as content, predictions, recommendations, or decisions influencing the environments they interact with. The term AI system in this guide refers to the entire technology. For a mobility service, it could be the app that integrates a Machine Learning (ML) model to predict demand and adjust pricing, including, for example, the data pipelines and protocols. In the rest of the paper, algorithmic audits and AI audits are used interchangeably, both referring to audits on algorithmic or AI systems.

[2] Ryan C. LaBrie and G. Steinke, "Towards a Framework for Ethical Audits of AI Algorithms," 2019, https://www.semanticscholar.org/paper/Towards-a-Framework-for-Ethical-Audits-of-AI-LaBrie-Steinke/c103601dbf79c05c7f72b865ce05e6f82048c1ca; Miles Brundage et al., "Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims" (arXiv, April 20, 2020), https://doi.org/10.48550/arXiv.2004.07213; Jakob Mökander et al., "Ethics-Based Auditing of Automated Decision-Making Systems: Nature, Scope, and Limitations," *Science and Engineering Ethics* 27, no. 4 (July 6, 2021): 44, https://doi.org/10.1007/s11948-021-00319-4; Jakob Mökander and Luciano Floridi, "Ethics-Based Auditing to Develop Trustworthy AI," *Minds and Machines* 31, no. 2 (June 1, 2021): 323–27, https://doi.org/10.1007/s11023-021-09557-8; Adriano Koshiyama et al., "Towards Algorithm Auditing: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms," SSRN Scholarly Paper (Rochester, NY, January 1, 2021), https://doi.org/10.2139/ssrn.3778998.

algorithmic audits can promote procedural regularity, increase transparency, and inspire proactivity in harm prevention and mitigation during the design of systems.[3]

Algorithmic audits can be broadly classified into either internal or external audits, depending on the auditors' distance from the developers or implementers of an algorithm; while Eticas has been at the forefront of developing both types of audits, this paper focuses on its achievements in empowering vulnerable peoples through incorporating community-led processes in external auditing methodologies.[4]

In external algorithmic audits, third-party auditors operate independently from the developers to examine the impact and, to the extent possible, the functioning of an algorithmic system. The auditor's independent position helps remove misaligned incentives in developer self-reporting, establish accountability, and increase overall transparency and public trust in the inspected system. External audits further hold the promise of detecting biases, inefficiencies, anomalies, and other hidden practices that could be unfair or harmful towards vulnerable communities within the society.[5]

While many proposals of external algorithmic auditing have been put forward, most constrain the relevant actors to only the auditors and system developers. Lacking in these proposals is the role that affected communities could play in co-developing and applying the product of an algorithmic audit.

In this proposal, we make a case for rendering external algorithmic auditing participatory through partnership with vulnerable communities, outline our CLA methodology, and demonstrate through multiple case studies the concrete and unique values that CLAs bring.

2. Placing communities at the center

Algorithmic audit studies have gradually acknowledged the value of involving those most affected by an AI technology in one way or another to complement an auditor's analysis. In most instances, algorithmic auditors enter a case as an independent expert external to the

---

[3] Shahar Avin et al., "Filling Gaps in Trustworthy Development of AI," *Science* 374, no. 6573 (December 10, 2021): 1327–29, https://doi.org/10.1126/science.abi7176; Jakob Mökander and Luciano Floridi, "Operationalising AI Governance through Ethics-Based Auditing: An Industry Case Study," *AI and Ethics* 3, no. 2 (May 1, 2023): 451–68, https://doi.org/10.1007/s43681-022-00171-7; Jakob Mökander et al., "Auditing Large Language Models: A Three-Layered Approach," SSRN Scholarly Paper (Rochester, NY, February 16, 2023), https://doi.org/10.2139/ssrn.4361607.
[4] Gemma Galdon-Clavell et al., "Adversarial Algorithmic Auditing Guide 2023" (Association Eticas Research and Innovation, 2023).
[5] Galdon-Clavell et al.

developers, end users, and the broader affected communities. In other words, the scope of audits is often limited by the hypotheses auditors think to test instead of those informed by end users' lived experience as they experience algorithmic impacts.[6] In an attempt to include user perspectives, some researchers proposed end-user audits or crowdsourced audits, such as tech companies' bounty challenges.[7] However, the precise audit methods in these proposals majorly transfer the onus of auditing algorithmic and AI systems onto individual end users, who may not hold the technical knowledge to investigate and build a case around their experiences. These proposals also pose a barrier to entry on vulnerable communities who have been historically excluded from digital access that would have enabled them to participate meaningfully in critical response.

To address this gap, we employ participatory research approaches along with quantitative methods, grounding our work by bringing together researchers (auditors in our case) and community members as collaborators while mutually learning about each other's knowledge and perspectives.[8] Audits established in social sciences, especially those driven by explicit concerns of social justice and racial equity, have a strong convention of requiring the direct participation of the impacted communities and are oriented around establishing accountability.[9] These participatory audits stress the importance of conducting research and analysis *with* participants, not *on* or *for* them, and serving their needs and goals.[10]

The core of Eticas' CLA approach is precisely this notion of placing communities at the center in the push for AI accountability. Our CLA design balances the technical expertise of

---

[6] Michelle S. Lam et al., "End-User Audits: A System Empowering Communities to Lead Large-Scale Investigations of Harmful Algorithmic Behavior," *Proc. ACM Hum.-Comput. Interact.* 6, no. CSCW2 (November 11, 2022): 512:1-512:34, https://doi.org/10.1145/3555625.

[7] Lam et al.; Wesley Hanwen Deng et al., "Understanding Practices, Challenges, and Opportunities for User-Engaged Algorithm Auditing in Industry Practice," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23 (New York, NY, USA: Association for Computing Machinery, 2023), 1–18, https://doi.org/10.1145/3544548.3581026.

[8] William F. Whyte, "Advancing Scientific Knowledge through Participatory Action Research," *Sociological Forum* 4, no. 3 (September 1, 1989): 367–85, https://doi.org/10.1007/BF01115015.

[9] Diana Auret and Stephanie Barrientos, "Participatory Social Auditing : A Practical Guide to Developing a Gender-Sensitive Approach," IDS Working Papers 237 (Brighton, the UK: Institute of Development Studies (IDS), 2004), https://opendocs.ids.ac.uk/articles/report/Participatory_social_auditing_a_practical_guide_to_developing_a_gender-sensitive_approach/26480614?file=48231346; Diana Auret and Stephanie Barrientos, "Participatory Social Auditing: Developing a Worker-Focused Approach," in *Ethical Sourcing in the Global Food System* (Routledge, 2006); Briana Vecchione, Karen Levy, and Solon Barocas, "Algorithmic Auditing and Social Justice: Lessons from the History of Audit Studies," in *Equity and Access in Algorithms, Mechanisms, and Optimization* (EAAMO '21: Equity and Access in Algorithms, Mechanisms, and Optimization, -- NY USA: ACM, 2021), 1–9, https://doi.org/10.1145/3465416.3483294.

[10] Vecchione, Levy, and Barocas, "Algorithmic Auditing and Social Justice"; Alice McIntyre, *Participatory Action Research* (SAGE Publications, Inc., 2008), https://doi.org/10.4135/9781483385679.

auditors and the lived, contextual, and strategic expertise of community members, allowing both to co-facilitate different steps of the way, from setting research agenda to providing on-the-ground data, reports, and metrics. The outcome of CLA is guided by the needs of the communities to support the latter to take action against unfair or unjust algorithmic and AI systems.

3. CLA step-by-step guide

Based on years of experience, Eticas has long developed and implemented a step-by-step guide for CLA.[11] The following section introduces the guide and highlights the crucial role communities play in each step, hereby shedding light on the community participation aspect of the traditionally exclusive auditing activities. While we generally found the sequence of steps commonly followed and convenient, it is worth noting that the steps may need to be adjusted according to the specific context in each case.

There are two main phases in CLA: planning and execution. It is important to note that CLA impacts tend to be more extensive and effective when there is a third phase of actionable plans, such as engaging in strategic litigation against discriminatory algorithmic or AI systems or building an alliance with other communities facing unfair algorithmic outcomes to advocate for policy changes. Hence, it is recommended that when the auditor and communities collaborate on a CLA, they jointly adjust the scope of the audit to any action plans in sight.

As the original guide provides detailed descriptions of each step of how each step can be applied in practice, here we only provide a shorter version (Table 1):

Planning	
1. Choosing a system to audit	Listening to communities to identify an AI system with social impact and an initial feasibility check for identifying possible access points to the algorithm(s) for an audit
2. Contextual analysis	Building understanding about the AI system and the community's experience of it, the context in which it operates and the possible negative impacts it may lead to, through discussions and interviews as well as technical and policy research

[11] Galdon-Clavell et al., "Adversarial Algorithmic Auditing Guide 2023."

3. Stakeholder mapping	Identifying all relevant parties to an AI system, such as the developers and implementers of the system and the communities affected directly or indirectly by it
4. Feasibility assessment	Data mapping to determine if the auditor can obtain sufficient information about an AI system via legal means within the relevant jurisdiction
5. Alliance building	Participatory research design with communities and civil society organizations to ensure that the perspectives of affected groups are incorporated in the auditing process
6. Methodology design	In consultation with communities, defining the scope of the audit, the research questions, the methods to investigate them, the utility of the results, and the timeline of the project
Execution	
7. Data collection	Safe and consentful[12] qualitative and quantitative data gathering about the inputs, outputs and societal impact of an AI system via specialized techniques for adversarial algorithmic auditing and social science research methods
8. Data analysis	Translating raw data into meaningful insights via quantitative and qualitative data analysis
9. Mitigation and recommendations	Providing actionable audit outputs, including reports, metrics, visualizations, and recommendations that serve community leadership in demanding accountability and improvement from developers, implementers, and policymakers

(Table 1. Step-by-step description of a CLA.)

3.1. Planning phase

The planning phase involves a series of steps aimed at ensuring that the audit has a clear goal and that it is well-prepared and organized. This involves the following steps: choosing a system to audit, contextual analysis, stakeholder mapping, feasibility assessment, alliance building, and methodology design.

Step 1-3. Making communities' needs visible

[12] The term "consentful" is inspired by a Design Justice practice. See more: https://alliedmedia.org/projects/consentful-tech-project

The order of the first three steps may vary depending on who initiated the CLA. While external auditors may be generally following different incidents of algorithmic and AI system misconducts, a community or a CSO may also reach out to auditors, seeking technical advice and suggesting an audit. Regardless of the initiator, these three steps heavily involve auditors and communities to come together and collectively make sense of the impacts of a contested system. It is important to let the experiences of the affected communities speak for themselves at this stage, with the auditor actively listening while performing the role of a facilitator. Auditors should predicate their understanding of system, the environments in which it operates, and any undesirable effects it creates on the ground reality communities experience and a robust technical and policy review.

Auditors, CSOs, and communities are encouraged to organize initial meetings to determine the nature and subject of the audit based on following crucial considerations (Table 2):

Considerations	Descriptions
Impact and scale	This refers to the potential impact of the system, whether it happens to a small community or extends to more groups, and even the society as a whole. The precise impacts and their scale may determine the focus of the study or the need for other alliances.
Potential access	The possibilities of accessing (a part of) the system may dictate the design of methodologies.
Bias and harms	By performing a contextual analysis and learning from the communities' firsthand experiences, the auditor may suggest possible biases and harms that a system may have. The auditor should co-construct and verify these hypotheses with the affected communities.
Actors and workflow	A stakeholder mapping will reveal the complete workflow of all actors involved in the entire lifecycle of the inspected system. Note that sometimes, end users may not be the negatively affected communities but a different stakeholder group (e.g. the user of a CV-screening algorithmic system is a human resources officer, while the affected communities may be people of minority ethnicities). It is critical to identify which actors perform which

	functions or carry out which interactions throughout the workflow of a system. Additional to a stakeholder mapping, the community and auditor may consider also carrying out a power mapping, which further identifies who has power over whom and what may influence the individuals or institutions involved in the system lifecycle.

(Table 2. Considerations for communities and auditors during the first three steps.)

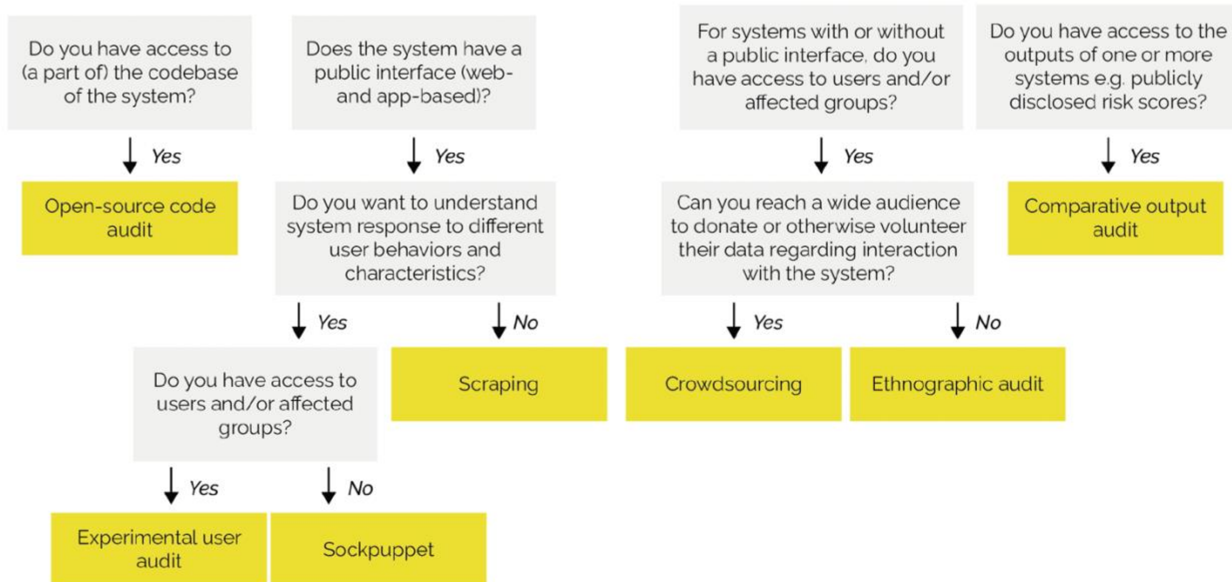
Step 4-6. Mutual learning of expertise and practical knowledge

These three steps are the heart of an audit design: feasibility assessment and methodology design are closely linked to each other. One of the most challenging aspects of CLAs is the lack of access to internal data and technical documentation about algorithmic systems. After performing a data mapping exercise along with a legal feasibility assessment to see the available “arsenal”, an auditor and the community must then concoct a creative methodological solution for a rigorous and sound audit.

The auditor and the community play a co-constructive role in these steps: While auditors hold technical knowledge on the specificity of algorithmic and AI systems, the community may possess more **practical insights** about the day-to-day operation of the system in question by having interacted with the latter extensively. For example, they may know better the bugs and quirks of the system’s behavior and the concrete effects it has on the community itself. Especially with predictive systems, they may have gained an intuition about the outcome of the system based on their input, which can be informative for the audit. Moreover, the community often possess highly **contextual information** about the system’s operational environment, such as the common traits of the communities itself, the workplace culture where the system is used, the historical evolution of system versions, the domestic regulations and rules that apply, etc.

Lastly, the community most likely come with **resources and networks** beyond the reach of an auditor; especially with vulnerable communities that tend to be closed off to outsiders, trust and alliance building are time consuming yet essential. An auditor may need to rely on the community with in-built connections to reach out to individual members or mobilize them for data collection, analysis, and further action.

During this collaborative process, communities can also learn about an auditor’s work, different methodological designs, and relevant technical insights into the system under investigation. There are methodologies more collaborative and participatory than the others. CLA combines traditional social science methods from a socio-technical perspective and specialized methods for algorithmic auditing. The following figure serves as guidance on the selection of the most appropriate auditing method depending on the feasibility assessment (Figure 1):



(Figure 1. Selection of method for conducting CLAs.)

As the original guide readily explains each method in detail, we present here only a brief run-down of the non-exhaustive list of methods, ranking them from involving most direct community participation to the least (Table 3):

Method	Description	Community participation
Experimental user	This is a systematic method for observing and recording system responses to real user behaviors under different conditions predetermined by the auditor. While the users are authentic, their interactions with the system are performed by design, rather than reflecting their normal engagement with a system (as in crowdsourcing).	Community members should be invited to co-design experiments based on their lived experiences and replicate them in experimental conditions for further analysis. The community and the auditor can also recruit identified vulnerable community members to play the role of real users here to learn from their concrete practices and reflections.

Crowdsourcing	Crowdsourcing is a method for collecting users' regular interactions with a system, which is often done via voluntary data donations.	Community members are the primary source of data provider in this method, as they offer their lived experiences in the form of data collected by algorithmic or AI systems to be the center of analysis.
Ethnography	An ethnography is a qualitative method for data collection through observation, interviews and surveys to understand and analyze how end users, particularly vulnerable groups, interact with a system.	Communities will play the role as the participants alongside ethnographers, who are participant observers. The goal of ethnography is to render community members' real-life practices and perspectives of the studied system visible.
Comparative output	A comparative output method involves comparing an algorithm's predicted outcomes with the actual outcomes or comparing the performance of one system against another, a benchmark, or a statistical measure for accuracy.	Communities and CSOs may be potential data sources, especially if they have been recording on a long timescale or in large numbers of system outputs. This may be especially helpful when official statistics is not available or sufficiently granular.
Sock puppet	Sock puppet is a systematic method for simulating real user behavior which involves the use of impersonation through (sock puppet accounts) and recording the system's response to different user characteristics and behavior(s). The sock puppet method can be executed manually by the	The role of an advisor and an observer. Alternatively, community members may also co-design sock puppet users based on archetypes most relevant to themselves (i.e., using common demographic characteristics or habitual

	researcher, or automatically by using a custom script.	traits germane to a particular community).
Scraping	Scraping is a systematic method of issuing repeated queries to a platform under different conditions and collecting the results. Scraping can be done manually by the auditor, or automatically by using a custom script.	The role of an advisor and an observer.
Open-source code audit	If white-box access to the system is possible, the auditor can perform a thorough review of the system's source code, training data, and other inputs to understand the algorithm's intentions and objectives. Open-source code audit also allows the testing of biases and fairness metrics.	The role of an advisor and an observer.

(Table 3. Brief description of each audit method.)

A few of the methods above are inherently participatory. Take the crowdsourcing method: Many datasets useful for an auditor are either proprietary or concern detailed personal information, such as age, gender, sexual orientation, financial records, medical records, personal travel history, etc. Companies and governments who collected such data from individual users to train and develop algorithmic or AI systems oftentimes do not disclose such information, nor do they share the outputs of their systems as public datasets. However, community members can exercise their right of access as defined in Article 15 of the European Union's General Data Protection Regulation (GDPR): the data subject shall have the right to obtain from data controller whether the latter processes the former's personal data, a copy of such data, and any supplementary information.[13] In other words, community members may be able to make requests to system developers to gain a copy of their data that will be essential for an auditor to test.[14] This is especially a powerful right that rests with the affected individuals, given that many countries outside of the European Union also legislated a similar version of GDPR.

[13] Council of the European Union and European Parliament, "REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)," Pub. L. No. JOL_2016_119_R_0001 (2016), https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.
[14] Vecchione, Levy, and Barocas, "Algorithmic Auditing and Social Justice."

Another example is the experimental user method: Community members can contribute in two different ways. First, they can inform experiment design based on their lived experiences with a given system. Second, they can also play the real users following a script designed by the auditors. In both cases, community members boast more experience interacting with the tested system, which allow them to provide in-depth reflections of the strengths, flaws, potential biases and unfairness a system generates and further suggest concrete mitigation strategies that best fit their needs.[15]

Overall, it is recommended that the auditor and community members consider the CLA as an opportunity for mutual learning on both sides.

Auditors, communities, and CSOs are again encouraged to conduct scoping calls to exchange knowledge and expertise and jointly define the scope and methodology of the audit. It is recommended that the auditors align the scope with any potential course of action that communities and CSOs can take up after the completion of the audit, such as strategic litigation, activism movement building, coalition formation, etc.

3.2. Execution phase

The execution phase involves carrying out the audit according to the previously designed methodology, starting with data collection, analyzing and interpreting results, presenting findings and finally providing recommendations or mitigation measures.

Depending on the methodological choice, communities may be the main lead, the collaborator, the subject, or mere observers in the data collection step. Importantly, any personal data of the community members are collected during this process must be treated with the utmost care, especially if it concerns vulnerable communities where their information may be sensitive.

In processing personal data, an auditor should consider the following (Table 4):

Considerations	Notes
Obtaining written informed consent	Human data subjects should be duly informed of:[16] • Type of data the auditor will collect from them • Where and for how long will their data be stored • What the data will be used for • Where the data will be shared (if known)

[15] Noura Howell et al., "Reflective Design for Informal Participatory Algorithm Auditing: A Case Study with Emotion AI," in *Proceedings of the 13th Nordic Conference on Human-Computer Interaction*, NordiCHI '24 (New York, NY, USA: Association for Computing Machinery, 2024), 1–17, https://doi.org/10.1145/3679318.3685411.
[16] Consider using recommendations by medical journals (which tend to have the highest standards) to draft consent forms.

	• Otherwise, how the data subjects' privacy, sensitive data, and any identifiable personal data will be protected in case when they will be shared
Identifying sensitive data	According to Articles 4, 9, and Recitals 51 to 56 of GDPR, the following data is considered sensitive: • Personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs • Trade-union membership • Genetic data, biometric data processed solely to identify a human being • Health-related data • Data concerning a person's sex life or sexual orientation
Data anonymization	• Remove anything that identifies the subject: names, addresses, workplaces, occupations, or salaries • Take out unnecessarily precise information: use age instead of date of birth • Aggregate individual-level data: use ranges, medians, or averages instead of precise numbers • Generalize when suitable: replace peoples' specific area of expertise with more general definitions; replace address and postal codes with local authority names • Use pseudonyms or personal index: use fictitious names or unique indexing • Avoid listing the upper or lower ranges of variables: this will disguise outliers, such as salary range for example.
Gated access	Especially in the case when the auditor will make the CLA open source, upon the involved parties' consent: • Share only aggregated or synthetic data snippets publicly • Allow others who have reasons to obtain original datasets to access via applications or getting in touch • Draft special data licensing agreements with those who are granted access beyond gates

(Table 4. Considerations when handling data from communities.)

While the data analysis step is most likely to be performed by the auditor in conventional external auditing, a CLA concentrates on the mutual learning and action of involved communities and technical experts. On parts that concern more about the community's expertise and knowledge, the analysis – from the formation of research questions and responses to such questions – should draw more heavily on their reflections and potentially be led by them. On parts that concern more about the specificity of the tested system and where the community may be less familiar with, the auditor should try to ensure that their analysis is also explainable to community members with clarity, allowing the latter to participate in the more technical part of the analysis and gain knowledge about the process.

This could be done through open sourcing the materials of a CLA, making the audit reproducible and its method replicable. The auditor should consider making the code scripts, datasets, step-by-step execution guide of the analysis, and any other relevant documents available to the participating communities, if not publicly online. To this end, Eticas has been working on publishing its past audits on GitHub, which is a cloud-based platform popular for open-sourced programming projects.[17]

When making an audit shareable, the auditor must consider the readability of codes, the replicability of the analysis, the clarity and quality of documentation, and the delicacy of data processing with regard to privacy issues (as outlined in Table 4).

Lastly, the mitigation and recommendation step is where the auditor and involved communities jointly consider the social, legal, and economic implications of the findings, and ways to address the biases, inefficiencies and other negative impacts. The results of the CLA should take the form of actionable outputs, such as concrete mitigation measures of any harms, policy recommendations that go beyond existing regulations, or visualizations of the impacts of the investigated system on the community. The core idea is that a CLA report or other output formats should be a key resource that serves community leadership in demanding accountability and improvement from developers, implementers, and policymakers.

4. Case studies

Over the years, Eticas has conducted multiple CLAs with different degrees of community participation. As the methodology for involving communities and CSOs mature, we see a few concrete ways in which community participation is not only beneficial but crucial to the success of the audits. In this section, we show in three CLAs ways in which communities have participated in our work.

4.1. Contextual perspectives and practical know-how

Eticas conducted an audit on RisCanvi, a predictive algorithmic system designed to assess the recidivism risks of inmates in Catalonia.[18] Being the first audit conducted on a criminal

[17] Eticas Foundation GitHub page: https://github.com/eticas-foundation.
[18] Gemma Galdon-Clavell et al., "Automating (In) Justice? An Adversarial Audit of RisCanvi" (Eticas Foundation, June 2024), https://eticasfoundation.org/automating-injustice-an-adversarial-audit-of-riscanvi/.

justice algorithmic system in Europe, Eticas had to work with little to no publicly available data on the implementation of the system and its outcomes. Eticas worked closely with Iridia,[19] a CSO that defends the civil and political rights of inmates, to leverage heavily their local, domain-specific perspectives and practical know-hows in developing the audit methodology. Iridia played an essential role in identifying key stakeholders relevant to the audit, building trust among them, and conducting the interviews.

Through this collaboration, we designed a two-part methodology, with the first part being an ethnography and the second part being a statistical analysis of the predictive power of the system using proxy data.

The ethnography was key in our analysis, as it surfaced a high level of contextual information and lived perspectives of the users and affected communities. It portrays a more complete picture of how the system was implemented and its palpable social and psychological effects on people. The ethnography concentrated on the concrete experiences of a validator of the system, inmates (affected communities) whose life was impacted by the system, prison workers[20] whose input fed into the system, and a wide array of professionals and advocates who worked in and around the prison system. Through this investigation, we found that those directly impacted by this algorithmic system were the least aware of its existence and were in a weak position to uphold their rights. Moreover, the prison workers reported not knowing how the system operated and not feeling capable of contradicting with the system's risk level estimates. Coupling many of the interviewees' insights, we were able to discover that apart from the dubious results the system was producing, it was also untrusted by the people who were interacting with it given its opacity and intractability. This valuable ethnographic data, made possible with the direct participation of communities and the assistance of Iridia, pointed to the grander lack of trust of criminal justice algorithmic systems that were critical to so many lives.

4.2. In-built trust and networks

Eticas conducted an audit on Azul, a facial recognition system created by Zurich Insurance Group in Spain, and its discriminatory impacts on people with Down Syndrome.[21] Azul's primary purpose is to scan a client's face through a webcam, assess that individual's facial features, and estimate their age, smoking status, and body mass index (BMI). Its underlying AI-based prediction algorithm will then assign an estimated price for life insurance coverage to each individual.

---

[19] https://iridia.cat/en/.

[20] Including social educators and psychologists working in prisons.

[21] Gemma Galdon-Clavell, Matteo Mastracci, and Miguel Azores, "Invisible No More: The Impact of Facial Recognition and Price Discrimination AI on People with Disabilities" (Eticas Foundation, August 2023).

Our audit followed a crowdsourcing method where we collaborated closely with another CSO, Cedown Jerez,[22] a prominent organization that supports and advocates for the rights of people with Down Syndrome. Members of vulnerable communities like people with disabilities tend to be more difficult to reach without organizations putting in a considerable amount of time and effort in mobilizing individuals for a cause, reducing barriers and transaction costs, and establishing a trusted network. Given that Cedown Jerez has been serving the community of people with Down Syndrome for years, we leveraged the trust they've built with the community to recruit participants for the study and invite them to donate their data for our testing.

4.3. Community-exclusive knowledge

In one of the current audits, Eticas studies the potential discriminatory impact of ride-hailing platforms' AI systems on Roma communities that often live in more secluded settlements in urban areas. [23] Ride-hailing platforms like Uber and Bolt automate their services with a wide range of AI models, such as surge pricing algorithms, driver-rider matching algorithms, and real-time demand-supply prediction algorithms. Given that these algorithms are very sensitive to historical mobility patterns associated with hyper local geographic areas, ride-hailing platforms may be providing different prices or quality of services to socioeconomically disadvantaged and ethnic minority neighborhoods.

Our audit highly depends on the precise geolocation of Roma settlements; nevertheless, mapping out such settlements proved to be extremely challenging. Due to their difficult history with local authorities and systemic discrimination that prevented them from escaping the vicious cycle of poverty, most Roma people were pushed to the brinks of urban areas and now live in informal, temporary settlements. The precarity of their living conditions and aversion to most government-led census meant that there is a lack of official records on the geolocation of Roma settlements.

Through collaboration with Fundación Secretariado Gitano,[24] a long-standing CSO working on the promotion and equal opportunities of Roma populations in Spain and Europe, we received their curated dataset on the geolocation of Roma communities in Spain. As they have worked with Roma communities for years, Fundación Secretariado Gitano's years of trust-building and the gained knowledge on their mobility patterns, current living conditions, and household compositions were central to the audit's basis of analysis.

[22] https://www.cedown.org/.
[23] Unpublished work in progress at the time of writing (24 January 2025).
[24] https://www.gitanos.org/.

5. Conclusion

In this extended abstract, we have highlighted the importance of community-led audits (CLAs) as a methodology for addressing algorithmic and AI injustices. By reimagining the role of communities in the auditing process, CLAs shift power dynamics, enabling communities to move beyond being passive observers or mere data points. Instead, they are positioned as active leaders who can shape, guide, and hold these algorithmic and AI system developers, designers, and policymakers accountable. This leadership role is essential in pushing back against the dominance of AI companies, which often operate with minimal oversight or accountability. By centering the voices and expertise of those most impacted, the CLA methodology not only empowers affected communities but also provides them with actionable results to address their needs and desires.

Our three case studies illustrate how community participation can play a pivotal role in resisting AI injustices. These examples demonstrate the value of engaging communities as collaborators in the auditing process, ensuring their insights directly inform efforts to identify risks and propose solutions. By focusing on lived experiences and localized knowledge, CLAs uncover harms that may be overlooked by auditors alone. Moreover, these CLAs help bridge the gap between civil society matters and the technical specificity of AI technologies, equipping grassroots leaders with the tools to advocate for accountability and redress. As AI systems increasingly shape society, embedding community expertise into auditing frameworks offers a path toward more equitable and accountable technology development.

References

Auret, Diana, and Stephanie Barrientos. “Participatory Social Auditing : A Practical Guide to Developing a Gender-Sensitive Approach.” IDS Working Papers 237. Brighton, the UK: Institute of Development Studies (IDS), 2004. https://opendocs.ids.ac.uk/articles/report/Participatory_social_auditing_a_practical_guide_to_developing_a_gender-sensitive_approach/26480614?file=48231346.

———. “Participatory Social Auditing: Developing a Worker-Focused Approach.” In *Ethical Sourcing in the Global Food System*. Routledge, 2006.

Avin, Shahar, Haydn Belfield, Miles Brundage, Gretchen Krueger, Jasmine Wang, Adrian Weller, Markus Anderljung, et al. “Filling Gaps in Trustworthy Development of AI.” *Science* 374, no. 6573 (December 10, 2021): 1327–29. https://doi.org/10.1126/science.abi7176.

Brundage, Miles, Shahar Avin, Jasmine Wang, Haydn Belfield, Gretchen Krueger, Gillian Hadfield, Heidy Khlaaf, et al. “Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims.” arXiv, April 20, 2020. https://doi.org/10.48550/arXiv.2004.07213.

Council of the European Union and European Parliament. REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), Pub. L. No. JOL_2016_119_R_0001 (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.

Deng, Wesley Hanwen, Boyuan Guo, Alicia Devrio, Hong Shen, Motahhare Eslami, and Kenneth Holstein. “Understanding Practices, Challenges, and Opportunities for User-Engaged Algorithm Auditing in Industry Practice.” In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–18. CHI ’23. New York, NY, USA: Association for Computing Machinery, 2023. https://doi.org/10.1145/3544548.3581026.

Galdon-Clavell, Gemma, Mat Mastracci, Luis Rodrigo González Vizuet, and Miguel Azores. “Automating (In) Justice? An Adversarial Audit of RisCanvi.” Eticas Foundation, June 2024. https://eticasfoundation.org/automating-injustice-an-adversarial-audit-of-riscanvi/.

Galdon-Clavell, Gemma, Matteo Mastracci, and Miguel Azores. “Invisible No More: The Impact of Facial Recognition and Price Discrimination AI on People with Disabilities.” Eticas Foundation, August 2023.

Galdon-Clavell, Gemma, Iliyana Nalbantova, Asylai Akisheva, and Luís Rodrigo Gonzalez. “Adversarial Algorithmic Auditing Guide 2023.” Association Eticas Research and Innovation, 2023.

Howell, Noura, Watson F Hartsoe, Jacob Amin, and Vyshnavi Namani. “Reflective Design for Informal Participatory Algorithm Auditing: A Case Study with Emotion AI.” In *Proceedings of the 13th Nordic Conference on Human-Computer Interaction*, 1–17. NordiCHI ’24. New York, NY, USA: Association for Computing Machinery, 2024. https://doi.org/10.1145/3679318.3685411.

Koshiyama, Adriano, Emre Kazim, Philip Treleaven, Pete Rai, Lukasz Szpruch, Giles Pavey, Ghazi Ahamat, et al. “Towards Algorithm Auditing: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms.” SSRN Scholarly Paper. Rochester, NY, January 1, 2021. https://doi.org/10.2139/ssrn.3778998.

LaBrie, Ryan C., and G. Steinke. “Towards a Framework for Ethical Audits of AI Algorithms,” 2019. https://www.semanticscholar.org/paper/Towards-a-Framework-for-Ethical-Audits-of-AI-LaBrie-Steinke/c103601dbf79c05c7f72b865ce05e6f82048c1ca.

Lam, Michelle S., Mitchell L. Gordon, Danaë Metaxa, Jeffrey T. Hancock, James A. Landay, and Michael S. Bernstein. “End-User Audits: A System Empowering Communities to Lead Large-Scale Investigations of Harmful Algorithmic Behavior.” *Proc. ACM Hum.-Comput. Interact.* 6, no. CSCW2 (November 11, 2022): 512:1-512:34. https://doi.org/10.1145/3555625.

McIntyre, Alice. *Participatory Action Research*. SAGE Publications, Inc., 2008. https://doi.org/10.4135/9781483385679.

Mökander, Jakob, and Luciano Floridi. “Ethics-Based Auditing to Develop Trustworthy AI.” *Minds and Machines* 31, no. 2 (June 1, 2021): 323–27. https://doi.org/10.1007/s11023-021-09557-8.

———. “Operationalising AI Governance through Ethics-Based Auditing: An Industry Case Study.” *AI and Ethics* 3, no. 2 (May 1, 2023): 451–68. https://doi.org/10.1007/s43681-022-00171-7.

Mökander, Jakob, Jessica Morley, Mariarosaria Taddeo, and Luciano Floridi. “Ethics-Based Auditing of Automated Decision-Making Systems: Nature, Scope, and Limitations.” *Science and Engineering Ethics* 27, no. 4 (July 6, 2021): 44. https://doi.org/10.1007/s11948-021-00319-4.

Mökander, Jakob, Jonas Schuett, Hannah Rose Kirk, and Luciano Floridi. “Auditing Large Language Models: A Three-Layered Approach.” SSRN Scholarly Paper. Rochester, NY, February 16, 2023. https://doi.org/10.2139/ssrn.4361607.

Vecchione, Briana, Karen Levy, and Solon Barocas. “Algorithmic Auditing and Social Justice: Lessons from the History of Audit Studies.” In *Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–9. -- NY USA: ACM, 2021. https://doi.org/10.1145/3465416.3483294.

Whyte, William F. “Advancing Scientific Knowledge through Participatory Action Research.” *Sociological Forum* 4, no. 3 (September 1, 1989): 367–85. https://doi.org/10.1007/BF01115015.