IPIE

International Panel on the Information Environment

Recommendations for a Global AI Auditing Framework:

Summary of Standards and Features

Summary for Policymakers 2024.2

DOI Number: 10.61452/GUYX7442

C. McKinley, P. Sylvester, B. Faveri, M. Johnson-León, W. H. K. Chun, A. Hannák, M. Mendoza, M. Broussard, R. Enikolopov, A. Nelson, C. Sandvig, 'G. Sesan, A. Sporle, R. Srihari, J. Srinivasan, C. Wilson, M. Zou, S. Valenzuela, P. N. Howard

SYNOPSIS

The growing integration of artificial intelligence (AI) into critical sectors of society, from healthcare to education, has the potential to support widespread social transformation and progress. However, AI systems also have the power to perpetuate biases, deepen inequalities, and cause environmental harm.

Accurately evaluating the risks and benefits of an AI system requires a careful audit. Current approaches to auditing, however, rarely involve independent auditors, provide sufficient evidence, or account for global impacts.

Policymakers urgently need a comprehensive global framework for AI audits that validates genuine benefits and risks. The IPIE's *Scientific Panel on Global Standards for AI Audits* set out to independently establish global, cross-disciplinary scientific consensus on what makes an audit effective and trustworthy. The Panel consisted of 16 experts from computer science, the social sciences, and the humanities, with expertise spanning generative AI systems, algorithmic auditing, indigenous data sovereignty, data journalism, and the relationship between civil society, human rights, and AI.

This Summary for Policymakers summarizes the Panel's second assessment, *A Global AI Auditing Framework* (SR2024.3), which offers a technical explanation of why a global framework for AI audits must provide:

- common standards and procedures,
- auditor independence and qualifications,
- auditor access to evidence,
- transparent audit results, and
- evaluation of AI's disparate global impacts.

The proposed framework for AI audits addresses seven features:

1. audit scope,
2. auditor role,
3. mechanisms for stakeholder engagement,
4. context-appropriate audit criteria,
5. evidence requirements,
6. audit methodologies, and
7. comprehensive post-audit activities.

A global framework must take care to map relevant stakeholders, develop participatory methods, establish criteria to minimize potential harms, develop audit methodologies, and check if the audit meets stakeholder expectations.

INTRODUCTION

Artificial intelligence (AI) systems are rapidly transforming industries, public services, and social interactions worldwide. While AI systems hold immense potential to drive progress and improve human welfare, they may also carry significant risks if not developed and deployed responsibly. New AI systems are being developed and deployed so quickly, however, that there is no trustworthy and effective means of inspecting, monitoring, and verifying that AI systems do what they claim to do, or evaluating what risks, if any, their use may create or exacerbate.

As detailed in *A Global AI Auditing Framework* (SR2024.3), AI systems can perpetuate historical biases, deepen socioeconomic inequalities, infringe on individual rights, and cause environmental harm, often in ways that disproportionately impact marginalized communities. As these systems become increasingly ubiquitous, robust auditing and public policy oversight is needed to prevent such harms.

A formal audit will evaluate an AI system and its interaction with people, communities, and organizations. It will assess whether a system is properly developed, deployed, operated, and managed, and check whether it adheres to vital social, ethical, and legal norms such as fairness, data privacy, and environmental sustainability. When an audit detects deviations from those conditions it demonstrates the AI system's risks and harms.

As investigated in the Panel's first assessment, *Global Approaches to Auditing Artificial Intelligence* (SR2024.1), current approaches to AI auditing vary widely in scope, revealing the need for:

- common standards and procedures,
- auditor independence and qualifications,
- auditor access to evidence,
- assurance that audit results are transparent, and
- auditing that addresses AI's disparate global impacts [1].

Furthermore, the heterogeneity of auditing approaches currently in practice makes it difficult to ensure consistent quality and accountability in the audit process. Most AI audits are conducted by actors in North America and Europe and may overlook how these technologies disparately impact people in the majority world, such as through violations of human rights, labor exploitation, and environmental degradation.

To address these challenges, this *Summary for Policymakers* outlines the IPIE's global AI auditing framework (SR2024.3), which provides a comprehensive set of recommendations for designing and implementing AI audits that are contextually relevant, multi-stakeholder driven, and focused on mitigating systemic harms [2].

By creating a basis upon which common standards and procedures can be developed, the framework supports a more trustworthy audit ecosystem capable of safeguarding human rights and collective wellbeing. It addresses the concerns of impacted communities, not just those of AI developers and deployers. Audits developed using this framework can also be used to evaluate the risk of biases, labor issues, and the environmental impacts of ever-growing AI use.

LESSONS FROM THE CURRENT AI AUDIT ECOSYSTEM

The current AI auditing ecosystem provides several lessons for a global AI auditing framework. Particular audit regimes, such as that mandated by New York City's pioneering Local Law 144 on the use of automated employment decision tools, reveal what it will take to set a global standard in AI audits [3].

KEY FINDING

Establishing common standards and procedures is vital to ensure effective AI audits and to build trust in the audit ecosystem.

Currently, we lack common standards and procedures for comparing the findings of various auditing approaches, from closed internal assessments to independent third-party evaluations. Ensuring that auditors demonstrate independence and appropriate qualifications is crucial.

For an audit to be effective and trustworthy, auditors require access to comprehensive evidence about an AI system's responsible development and deployment, including model and data documentation produced by the AI developer or deployer. Ensuring the public accessibility of audit reports supports stakeholders' ability to hold AI developers and deployers accountable for any harm their systems may cause.

To capture the systemic impacts of AI development and deployment, audits need to account for the disproportionate effects that these technologies can have on communities around the world. Addressing these challenges through IPIE's proposed auditing framework will help to build a more trustworthy and effective audit ecosystem to safeguard individual freedoms, mitigate systemic harms, and advance human development objectives.

FEATURES OF A GLOBAL AI AUDITING FRAMEWORK

The IPIE's *Scientific Panel on Global Standards for AI Audits* recommends a series of actions to be carried out either sequentially or in parallel to ensure effective audits.

The framework makes recommendations for defining the scope, auditor role, methods of stakeholder engagement and context identification, criteria, evidentiary requirements, methodology, and post-audit activities. Choosing which of these recommendations to apply depends on decisions made by the auditor in consultation with the auditee and other stakeholders. Any single AI audit need not apply all of them.

1. The Audit Scope

The audit scope is defined by what is to be audited and why: the objects of an AI audit should be selected according to the purpose and context of the audit.

- If the audit is motivated by regulatory or standards compliance, the objects should be defined based on the relevant jurisdictions or standards.
- To evaluate an AI system's performance, the audit object should be selected from the relevant performance characteristics, such as accuracy, reliability, fairness, transparency, interpretability and explainability, and robustness and safety.
- For audits focused on organizational risk management, objects should align with the organization's involvement in system development or use.
- For audits motivated by ethical and social considerations, the objects should be defined by the concerns of those impacted.

Once its purpose has been defined, the context within which an audit takes place further influences decisions around audit objects.

- The auditor should establish which aspect(s) of the AI system—model(s), data, governance, and outcomes—will be audited.
- The auditor should establish which stage(s) of the AI life cycle (design, development, and deployment) and which part of the supply chain (labor and material) are relevant.

2. Auditor Role

The framework provides clarity on the role of AI auditors through guidelines for demonstrating their obligations and accountability relationships, as well as providing transparency around their independence, standards, funding relationships, and qualifications. It states that audit processes will depend on:

- The type of auditor—internal, third-party, or independent auditors—which varies by the stakeholders to whom the auditor is accountable and their motivations for the audit.
- Auditor independence, which depends on who is funding their work and what standards of independence they adhere to.
- Auditor qualifications, which come from professional training, academic credentials, or domain expertise.
- Protections, such as safe harbor, which are necessary to ensure truly independent audits.

3. Context and Stakeholders

To account for an AI system's local impacts across its life cycle, auditors should identify an AI system's contexts and develop well-structured processes and mechanisms for stakeholder engagement.

- The domains with which the system interacts (application, institutional, jurisdictional, economic, cultural, and/or social) and any potential harms in those domains should be identified.
- Auditors should develop and use a stakeholder map to ensure that the audit objects reflect the concerns of those responsible for oversight and monitoring of the AI system, and those impacted by it.
- Stakeholders should be engaged through participatory methods with appropriate tools to gather input, translate this input into action items, explain to stakeholders how their input was used, and provide compensation to those who have contributed.

4. Audit Criteria

Audit criteria should capture stakeholder concerns, be relevant for the audit scope, and establish minimum conditions for compliance.

- Audit criteria should be precise, observable, contextual, and designed to assess risks and harm mitigation.
- Audit criteria should be developed with stakeholder input and by incorporating public knowledge.
- For audits of AI performance, criteria should formalize considerations of accuracy, fairness, transparency, interpretability and explainability, as well as robustness and safety.
- For auditing AI system training, testing, and validation data, criteria should address data privacy and protection, indigenous and collective data sovereignty, individual data rights, and cultural norms.
- When assessing an AI developer or deployer's commitment to fair labor practices, audit criteria should consider employment laws, fair labor

principles, and workers' rights, such as fair compensation, protection from surveillance and algorithmic management, and access to mechanisms for contesting automated decisions.
- For an evaluation of an AI system's environmental impact, criteria should allow for local environmental regulations and community needs. These criteria should account for the minimization of energy and water consumption, greenhouse gas emissions, carbon footprint, and e-waste.

Audit criteria should use stakeholder input to develop additional criteria addressing systemic harms that are locally relevant.

5. Audit Evidence

Evaluating a system against these criteria requires appropriate evidence.

- AI developers and deployers should provide auditors with the evidence they request, such as model cards, data sheets, audit trails, and data documentation.
- Auditors and auditees should agree on the safety and privacy mechanisms necessary to share sensitive or confidential data. Where possible, trusted intermediaries such as data observatories, data commons, and data cooperatives should facilitate data sharing.
- Auditors should collect relevant evidence from AI system end-users and impacted communities.
- When auditing for systemic and cumulative impacts, it is recommended to collect audit evidence over an extended period and from multiple sites.

6. Audit Methodology

The audit methodology should comprise a mix of performance evaluation, documentary analysis, stakeholder engagement, and other means as defined by the audit scope and criteria.

- Auditors must test the performance of the AI system, rather than relying solely on documentation provided by the auditee, with performance and scenario tests appropriate to the system's model(s), training, test, and validation data, its interaction domains, and its potential harms.
- The auditing methodology must test whether an audited system produces consistent outcomes within its intended domain, while accounting for systemic outliers and the limits of existing guardrails, and do so over an extended period so that long-term impacts can be captured.
- Auditors should conduct documentary analysis to evaluate the parts of an AI system on which AI developers and deployers have valuable insight.
- Auditors should gather testimonial evidence and conduct interviews and site visits to engage developer and deployer staff and contracted workers.

- The audit toolkit should employ participatory methods to assess the ways that people engage with or otherwise encounter AI systems in their everyday lives, and verify stakeholder satisfaction with the selected audit features.
- Auditors should identify the auditee responses to previous audits so as to assess the auditee's commitment to addressing identified issues.

Underpinning these methodological requirements is the need for auditors to demonstrate the value and suitability of their approach, especially when utilizing bespoke or non-standard methods and tools, or when collaborating with community and civil society stakeholders. Auditors must also satisfy any relevant ethical considerations, such as obtaining approval from their affiliated institution.

7. Post Audit Activities

The final audit report should state how the audit addressed each of the criteria above, provide justifications for deviations, and outline concrete recommendations and action items for the auditee and other stakeholders.

- The report must identify an auditee point of contact to whom stakeholders can contest aspects of the AI system, or request further information related to the audit.
- Audit results must be made publicly available through widely accessible news media; open-access, peer-reviewed journals; or searchable public repositories.
- Audit findings should be accessible for use in regulatory enforcement, legal systems, or other independent accountability structures, such as oversight boards. If an AI system is responsible for observable harms, the auditee should receive a grace period in which to address those concerns.

Where necessary, there should be follow-up audits so that the AI system's performance is continuously measured and improved.

A full explanation of the framework, with explanations of particular elements, citations to the scientific literature, and detailed case analyses, is available in the main assessment in *A Global AI Auditing Framework* (SR2024.3).

CONCLUSION

As the world grapples with the transformative potential of AI, policymakers, researchers, and civil society must work together to establish auditing mechanisms that prioritize transparency, accountability, and the protection of human rights.

Rigorous, context-sensitive AI audits can uncover biases, labor exploitation, environmental impacts, and other issues that currently undermine individual freedoms and collective wellbeing. The development of a comprehensive auditing framework is a crucial step towards building a more trustworthy and effective ecosystem for assessing the impacts of AI.

However, challenges remain in the AI audit ecosystem. Most pressing is the asymmetry between stakeholders in majority world countries and other countries that affect how AI audits can be conducted. Continued academic research into AI auditing design and practice, as well as institutional capacity-building to support auditors, is recommended. IPIE's Global AI Auditing Framework can play an important role in shaping the responsible development and deployment of AI technologies by ensuring that AI audits are responsive to the needs of stakeholders around the world.

REFERENCES

[1] International Panel on the Information Environment, 2024. *Global Approaches to Auditing Artificial Intelligence: A Literature Review*. SR2024.1. Zurich, Switzerland: IPIE.
[2] International Panel on the Information Environment, 2024. *A Global AI Auditing Framework: Assessment and Recommendations*. SR2024.3. Zurich, Switzerland: IPIE.
[3] New York City Consumer and Worker Protection, Local Law 144. 2021. [Online]. Available: https://rules.cityofnewyork.us/wp-content/uploads/2023/04/DCWP-NOA-for-Use-of-Automated-Employment-Decisionmaking-Tools-2.pdf

ACKNOWLEDGEMENTS

Contributors

Drafting authors: Catherine McKinley (Technical Writer, United Kingdom), Prem Sylvester (Consulting Scientist and Panel Manager, India/Canada), Benjamin Faveri (Consulting Scientist, United States), Maureen Johnson-León (Consulting Scientist, United States), Wendy Hui Kyong Chun (Panel Chair, Canada), Anikó Hannák (Panel Vice-Chair, Switzerland/Hungary), Marcelo Mendoza (Panel Vice-Chair, Chile), Meredith Broussard (Member, United States), Ruben Enikolopov (Member, Spain/Russia), Alondra Nelson (Member, United States), Christian Sandvig (Member, United States), ‘Gbenga Sesan (Member, Nigeria), Andrew Sporle (Member, New Zealand), Rohini Srihari (Member, United States), Janaki Srinivasan (Member, India/United Kingdom), Christo Wilson (Member, United States), Mimi Zou (Member, United Kingdom/Australia), Sebastián Valenzuela (IPIE Chief Science Officer, Chile), Philip Howard (IPIE President and CEO, Canada/United Kingdom). Independent Reviewers: Saiph Savage and Mona Sloam. Copyediting: Michelle Rosen. We acknowledge feedback from the IPIE affiliates: Luis Adrián Castro-Quiroa, Eloísa Gacía-Canseco, Joan Tirwyn Hassan, Piers Howe, Ozan Kuru, Anita Lavorgna, Wenlong Li, Nihat Muğurtay, Fredrick Ogenga, Dariia Opryshko, Tanja Pavleska, Margarita Sordo, Jun Zhang. We gratefully acknowledge support from the IPIE Secretariat: Jasmine McGhee, Egerton Neto, Donna Seymour, Alex Young.

Funders

The International Panel on the Information Environment (IPIE) gratefully acknowledges the support of its funders. For a full list of funding partners please visit www.IPIE.info. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the IPIE and do not necessarily reflect the views of the funders.

Declaration of Interests

IPIE reports are developed and reviewed by a global network of research affiliates and consulting scientists who constitute focused Scientific Panels and contributor teams. All contributors and reviewers complete declarations of interests, which are reviewed by the IPIE at the appropriate stages of work.

Preferred Citation

An IPIE *Summary for Policymakers* provides a high-level precis of the state of knowledge and is written for a broad audience. An IPIE *Synthesis Report* makes use of scientific meta-analysis techniques, systematic review, and other tools for evidence aggregation, knowledge generalization, and scientific consensus

building, and is written for an expert audience. An IPIE *Technical Report* addresses questions of methodology or provides a policy analysis on a focused regulatory problem. All reports are available on the IPIE website (www.IPIE.info).
This document should be cited as:

International Panel on the Information Environment, 2024. *Global AI Auditing Framework: Recommendations for Audit Design, Process, and Procedure.* SFP2024.2. Zurich, Switzerland: IPIE.

DOI

https://IPIE.info/10.61452/GUYX7442

Copyright Information



This work is licensed under an Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

ABOUT THE IPIE

The International Panel on the Information Environment (IPIE) is an independent and global science organization providing scientific knowledge about the health of the world's information environment. Based in Switzerland, the IPIE offers policymakers, industry, and civil society actionable scientific assessments about threats to the information environment, including AI bias, algorithmic manipulation, and disinformation. The IPIE is the only scientific body systematically organizing, evaluating, and elevating research with the broad aim of improving the global information environment. Hundreds of researchers worldwide contribute to the IPIE's reports.

For more information, please contact the International Panel on the Information Environment (IPIE) at secretariat@IPIE.info or at Seefeldstrasse 123, P.O. Box, 8034 Zurich, Switzerland.



International Panel on
the Information
Environment

Seefeldstrasse 123
P.O. Box 8034 Zurich
Switzerland

