Provocation on Expertise in Social Impact Evaluations of Generative AI (and Beyond)

Zoe Kahn
School of Information
UC Berkeley
zkahn@berkeley.edu

Nitin Kohli
Center for Effective Global Action
UC Berkeley
nitin.kohli@berkeley.edu

Abstract

Social impact evaluations are emerging as a useful tool to understand, document, and evaluate the societal impacts of generative AI. In this provocation, we begin to think carefully about the types of experts and expertise that are needed to conduct robust social impact evaluations of generative AI. We suggest that doing so will require thoughtfully eliciting and integrating insights from a range of *domain experts* and *experiential experts*, and close with five open questions.

1 Introduction

Social impact evaluations (e.g., (Solaiman et al., 2023)) are emerging as a useful tool to understand, document, and evaluate the societal impacts of generative AI; these evaluations take an important step towards the responsible development and enhanced accountability of artificial intelligence. Who develops categories of social impact—and who conducts these evaluations—shapes how generative AI is evaluated and held accountable. In this provocation, we begin to think carefully about the types of experts and expertise needed to conduct robust social impact evaluations. Without adequate expertise, there is a risk of developing evaluation criteria that fail to capture real-world harms or producing misleading evaluations that obscure critical issues. In this provocation we suggest that robust social impact evaluations require eliciting and integrating input from "domain experts" alongside "experiential experts." As working definitions, domain experts refer to people who have received training or professionalization in a particular domain such as data science, legal studies, history, among others; experiential experts refer to "people who are living the experience or those closely associated with someone living the experience" (Young et al., 2019). In our use of the term, experiential experts are not only experts in their own lived experience but also experts in culture (Abebe et al., 2021), organizational structure (Pfeffer and Leblebici, 1977), local values (Abokhodair and Vieweg, 2016), among others.

2 Airplane Design: An Illustrative Analogy

We motivate the discussion of experts and expertise in social impact evaluations of generative AI by turning attention to the design of a physical artifact: the airplane. Let's imagine we wanted to build an airplane that was safe to fly and comfortable for passengers. To do so, we would need to elicit and integrate insights from different types of experts: structural engineers to make sure the plane is strong enough to withstand the loads it could encounter; aerodynamicists to ensure the plane is as efficient and aerodynamic as possible; and experiential experts—or passengers—to make sure that the plane is comfortable. It would be important that each expert provide input on the aspects that align with their expertise—we would want structural engineers to weigh in on their particular area of specialization, but not define what constitutes passenger comfort; and we would want passengers to weigh in on the comfort of the plane, not the structural design of the plane.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

3 Attending to Expertise in Social Impact Evaluations of Generative AI

While airplane design is distinct from generative AI, attending to expertise can inform how robust social impact evaluations are conducted in practice. What we find appealing about the framing of expertise is that it treats all types of experts with dignity and provides a mechanism for broad accountability: no one kind of expert is necessarily more important than another.

Provocation 1: On the Experts Needed to Evaluate Social Impact. Each category of social impact likely requires different types of experts. Evaluating environmental costs may require individuals with expertise calculating water and energy consumption, a type of expertise not relevant to privacy and data protection evaluations. Even within a single social impact category (e.g., privacy and data protection), evaluations will likely require appropriate combinations of experts. While domain experts—such as technical experts to measure model memorization—are important, they may be insufficient on their own. Building on our prior work in rural Togo and a long lineage of work demonstrating that local norms and expectations of privacy depend on the social, political, and cultural context (Abokhodair and Vieweg, 2016; Abebe et al., 2021), we find that the data privacy concerns raised by domain experts often fail to include data privacy concerns of experiential experts; in our case, rather than concerns about the use of personal data in algorithms, individuals in rural Togo raised relational privacy harms that can arise when sensitive data is revealed to people nearby such as family members or people in the community (Kahn et al., accepted 2025). Here, and more broadly, robust social impact evaluations will likely hinge on including the appropriate combination of experts; failing to include the appropriate experts—or asking experts to weigh in on topics that do not align with their expertise—could obscure critical issues and lead to misleading evaluations.

Provocation 2: On the Construction of the Social Impact Evaluation Framework. Attending to expertise in the construction of the social impact evaluation framework reveals key strengths and potential limitations. Consider that the social impact evaluation framework was developed by a group of nearly 60 domain experts spanning industry, academia, civil society, and government (Solaiman et al., 2023). This is a strong, diverse range of expertise incorporated into its creation. It also reveals the types of experts who were not included, such as experiential experts. Such absence likely has implications for the categories of social impact that are (and are not) part of the social impact evaluation framework. For example, experiential experts in low- and middle-income countries may expand definitions of existing categories of social impact or reveal entirely new categories. Again, building on the first-author's prior work in rural Togo, new categories may be needed to capture how generative AI can shift interpersonal relationships and community dynamics (Kahn, submitted 2024). Importantly, observing which experts were part of the process positions us to ask: What experts are missing? What new categories are needed? How else might such frameworks need to shift or adapt?

4 Five Open Questions

In this provocation, we begin to think carefully about the types of experts and expertise needed to conduct robust social impact evaluations of generative AI; we close with five open questions. See *Appendix A* on our work in Togo and *Appendix B* drawing comparisons to design approaches. More broadly, while the paper focuses on social impact evaluations of generative AI we anticipate much of this thinking, including the open questions, applies to the design of sociotechnical systems writ large.

1. **Conceptualizing experts.** Beyond domain experts and experiential experts, are other categories of experts needed for a comprehensive conceptualization of expertise?
2. **Identifying experts and expertise.** How do you identify which experts and expertise are needed? How many experts of each type are enough? How do you identify the questions each type of expert is uniquely positioned to provide input on? Who decides?
3. **Enabling input.** How is meaningful input elicited from different types of experts? When is it important for experts to understand technical details? How are these technical details communicated in ways that enable meaningful input from experts with diverse expertise (particularly experts who may lack technical backgrounds)?
4. **Defining harms.** What counts as a harm? How are harms prioritized? Who decides?
5. **Resolving tensions.** How are tensions resolved when experts disagree? Tensions may arise within a type of expert (e.g., privacy experts disagree with one another) or between types of experts (e.g., privacy experts may raise concerns in tension with experiential experts).

Acknowledgments and Disclosure of Funding

Thank you to Christopher Jirucha and Geoff Martindale for providing aerospace expertise to ensure the accuracy of the airplane analogy.

References

Solaiman, I.; Talat, Z.; Agnew, W.; Ahmad, L.; Baker, D.; Blodgett, S. L.; Chen, C.; Daumé III, H.; Dodge, J.; Duan, I.; others Evaluating the social impact of generative ai systems in systems and society. *arXiv preprint arXiv:2306.05949* **2023**,

Young, M.; Magassa, L.; Friedman, B. Toward inclusive tech policy design: a method for underrepresented voices to strengthen tech policy documents. *Ethics and Information Technology* **2019**, *21*, 89–103.

Abebe, R.; Aruleba, K.; Birhane, A.; Kingsley, S.; Obaido, G.; Remy, S. L.; Sadagopan, S. Narratives and Counternarratives on Data Sharing in Africa. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 2021.

Pfeffer, J.; Leblebici, H. Information technology and organizational structure. *Pacific Sociological Review* **1977**, *20*, 241–261.

Abokhodair, N.; Vieweg, S. Privacy & Social Media in the Context of the Arab Gulf. New York, NY, USA, 2016; p 672–683.

Kahn, Z.; PERE, M. F. C.; Aiken, E.; Kohli, N.; Blumenstock, J. E. Expanding Data Privacy: Insights from Rural Togo. *Proceedings of the ACM on Human-Computer Interaction* **accepted 2025**,

Kahn, Z. Digital Social Protection and Social Cohesion: The Role of Human Intermediaries in Rural Togo. **submitted 2024**,

Aiken, E.; Bellue, S.; Karlan, D.; Udry, C.; Blumenstock, J. E. Machine learning and phone data can improve targeting of humanitarian aid. *Nature* **2022**, *603*, 864–870.

Kensing, F.; Madsen, H. *Design at work*; CRC Press, 2020; pp 155–168.

Alan Turing Institute, I. Stakeholder Impact Assessment. Alan Turing Institute. 2022.

Birhane, A.; Isaac, W.; Prabhakaran, V.; Diaz, M.; Elish, M. C.; Gabriel, I.; Mohamed, S. Power to the People? Opportunities and Challenges for pPrticipatory AI. Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization. 2022; pp 1–8.

Groves, L. Algorithmic Impact Assessment: A Case Study in Healthcare. Ada Lovelace Institute. 2022.

Ministry of the Interior & Kingdom Relations, M. Impact Assessment Fundamental rights and algorithms. 2022.

Zytko, D.; J. Wisniewski, P.; Guha, S.; PS Baumer, E.; Lee, M. K. Participatory Design of AI Systems: Ppportunities and Challenges Across Diverse Users, Relationships, and Application Domains. CHI Conference on Human Factors in Computing Systems Extended Abstracts. 2022; pp 1–4.

Delgado, F.; Yang, S.; Madaio, M.; Yang, Q. The Participatory Turn in AI Design: Theoretical Foundations and the Current tate of Practice. Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization. 2023; pp 1–23.

Pak, V. Introducing the HUDERIA. Council of Europe. N.D.

Norman, D. A.; Draper, S. W. *User centered system design; new perspectives on human-computer interaction*; L. Erlbaum Associates Inc., 1986.

Friedman, B.; Hendry, D. G. *Value sensitive design: Shaping technology with moral imagination*; MIT Press, 2019.

Appendices

A From Theory to Practice: Identifying, Eliciting and Integrating Expertise from Experiential and Domain Experts

To begin to move from theory to practice for how experts and expertise can be effectively identified, deployed, and integrated into evaluation frameworks, we provide an example from our own work leveraging experts and expertise to study data privacy in low-and middle-income countries (Kahn et al., accepted 2025) and reflect on possible implications for social impact evaluations of generative AI. While data privacy is not the same as social impact evaluations of generative AI, the structure and process may be informative for social impact evaluations of generative AI. In particular, social impact evaluations of generative AI may similarly find it useful to: identify experiential and domain experts and expertise currently represented; identify missing experiential and domain experts and expertise; initiate intentional efforts to elicit input from missing experts; ask experts for input on topics that leverage their unique expertise; and integrate across experts and expertise. This should be seen as the start of a conversation, not an articulation of best practices.

Research context. To situate our prior work, in response to the Covid-19 pandemic, the Government of Togo launched the world's first entirely digital cash transfer program that used machine learning and mobile phone metadata to determine program eligibility (Aiken et al., 2022). We, along with colleagues in data science and economics who trained the machine learning models and collaborated with the Government of Togo to implement the program, wanted to understand the data privacy risks that arise from the use of mobile phone metadata as part of development interventions. To do this work, we followed the following process. Importantly, while the process appears linear in its written form below, it was an iterative process in practice.

1. *Identify experiential and domain experts and expertise currently represented.* In our work in rural Togo, a literature review revealed that privacy and development experts and expertise were primarily represented in the literature exploring the data privacy concerns related to mobile phone metadata and its application to development. Leveraging what we learned in rural Togo, social impact evaluations of generative AI may benefit from identifying which types of experts and expertise are initially represented in evaluations. Methodologically, this could be done through a literature review or documenting in real-time which experts and expertise are represented in a given evaluation, among others.
2. *Identify important missing experiential and domain experts and expertise.* In our work in rural Togo, after documenting the experts and expertise currently represented it was notable that experiential experts were missing, namely, people living in rural Togo. Given that we were interested in data privacy, it was important to have experiential experts represented because, not only were their data used, but it is well documented that privacy norms differ across cultures. Leveraging what we learned in rural Togo, social impact evaluations of generative AI may similarly benefit from identifying the delta between the types of experts and expertise needed to conduct a given evaluation and the types of experts and expertise currently represented. It may also be that the necessary types of experts and expertise change over the course of a social impact evaluation, such as if the evaluation is taking place before or after AI release, requiring periodic re-evaluation.
3. *Initiate intentional efforts to elicit input from missing experts.* In our work in rural Togo, to understand the data privacy concerns of experiential experts the first author conducted ethnographically-informed fieldwork and semi-structured interviews in rural Togo to understand the data privacy concerns of people living in rural villages related to the collection and use of mobile phone metadata. Leveraging what we learned in rural Togo, social impact evaluations of generative AI may also benefit from intentional efforts to elicit input from missing experts. Methodologically, the most appropriate methods may depend on the type of expert (e.g., interviews, expert panels, statistical reports, requests for feedback on system design, among others).
4. *Ask experts for input on topics that leverage their unique expertise.* In our work in Togo, we initially got it wrong: we asked questions that did not align with the expertise of our participants in rural Togo. Inspired by Future Workshops (Kensing and Madsen, 2020), our initial interviews asked participants to envision different uses of mobile phone data, both

uses that should be supported and those that should be prevented. Our participants living in rural Togo often responded with blank stares, telling us they had little understanding of mobile phone data, let alone how it could be used or misused. This made us recognize our questions were perhaps better suited to a person with expertise envisioning data uses. After reflecting, we came to understand that experiential experts were especially well positioned to situate data within their everyday lives. We pivoted, first developing new methods that leverage visuals and storytelling to explain the data recorded in mobile phone metadata, helping participants with varying levels of literacy and formal education to develop an intuition for how that data could be used to draw inferences (Kahn et al., accepted 2025, accepted); then, shifting to ask participants how they would feel if their mobile phone data were shared with their spouse, household, village, the government, or researchers. This was a topic participants had a lot to say about, in part, because it aligned with their area of expertise. Leveraging what we learned in rural Togo, social impact evaluations of generative AI will likely be more robust if experts are asked to provide input on topics that align with their expertise, and in some cases it may be important to scaffold technical details to enable meaningful input.

5. *Integrate across experts and expertise*. In our work in rural Togo, we found that people in rural Togo surfaced a different set of privacy concerns than people with privacy and development domain expertise. In contrast to domain expert concerns related to surveillance, consent, and technical re-identification, individuals living in rural Togo raised a set of relational privacy concerns that could arise if mobile phone data were revealed to people nearby such as spouses, households, or the broader village. We do not argue that the privacy concerns of experiential experts are "right" and domain experts "wrong" (or vice versa). Instead, we take an integrative approach and demonstrate that addressing data privacy holistically will require reckoning with the data privacy concerns raised by both domain experts and experiential experts. Leveraging what we learned in rural Togo, social impact evaluations of generative AI will likely also encounter instances where different experts (of the same type or of different types) surface different concerns or provide different evaluations that are in tension with one another. In instances of disagreement, it will likely be important to identify constructive paths forward that take expertise into account and treat experts with dignity. Depending on the topic, this may entail deferring to one type of expert or expertise or integrating across different types of experts and expertise.

In this appendix, we provided an example from our own work leveraging experts and expertise to study data privacy in rural Togo, reflecting on how the lessons learned can be leveraged to inform social impact evaluations of generative AI. Effectively attending to experts and expertise in social impact evaluations of generative AI is an exciting opportunity for creativity and innovation. More broadly, we intuit that careful attention to experts and expertise may be applicable to sociotechnical system design writ large.

B Centering Experts and Expertise: Related To (And Distinct From) Established Design Approaches

As part of an emerging body of work examining experts and expertise in sociotechnical system design (Alan Turing Institute, 2022; Birhane et al., 2022; Groves, 2022; Ministry of the Interior & Kingdom Relations, 2022; Zytko et al., 2022; Delgado et al., 2023; Pak, N.D), we delineate how our approach builds on, but is distinct from, several well established design traditions. By attending to experts and expertise, we center *who* participates in the design process, with a particular focus on *what* expertise those individuals possess so that sociotechnical systems support a broad range of human values. In what follows we begin to articulate its relation to two established design traditions, user centered design (UCD) and value sensitive design (VSD). We caution that this discussion is preliminary and additional work is needed to properly explore experts and expertise in theory and practice.

UCD engages users so that sociotechnical systems are more useful and usable (Norman and Draper, 1986). We expand UCD in two ways:

1. *Expand who participates.* The focus on experts and expertise expands who participants in the design process beyond users to a broader group of experts. This includes domain

experts (e.g., privacy scholars, historians, lawyers) as well as experiential experts who may be users or non-users alike (e.g., people with lived expertise navigating low-bandwidth settings, people with lived expertise choosing not to use a technology, people with lived expertise who are unable to use a technology for any number of reasons).

2. *Expand the purpose of participation.* Traditionally, UCD engages users so that products are more useful and usable to end users. While these are two important values, we believe assembling an appropriate set of experts can be leveraged to envision, develop, and evaluate sociotechnical systems in ways that support a broader range of human values (e.g., dignity, autonomy, privacy, accountability, among others).

VSD engages direct and indirect stakeholders so that sociotechnical systems support a broad range of human values (Friedman and Hendry, 2019). While more work is needed to theorize experts and experts, we suspect centering experts and expertise is complementary to VSD.

1. *Complement who is impacted (stakeholders) with who constructs solutions (experts and expertise).* VSD engages direct and indirect stakeholders, identified by role, to understand how sociotechnical systems may implicate human values. By contrast, attending to experts and expertise turns attention toward whose knowledge and perspectives are needed to construct paths forward. To make this concrete, let us imagine that we are conducting a social impact evaluation of generative AI related to environmental costs and carbon emissions (Solemani et al. 2024). VSD would have us conduct a stakeholder analysis to identify direct and indirect stakeholders based on their role and stake in the environmental costs and carbon emissions of generative AI. Through such a process, we might identify a range of stakeholders including environmental advocates, people living in areas especially impacted by climate change, people living in areas near data centers and other large scale compute resources, and non-human stakeholders such as rivers, oceans, and creatures from small to large. This would help us begin to understand who may be impacted and how. By contrast, a focus on experts and expertise would have us consider the types of experts and expertise needed to conduct such an analysis and whose knowledge and perspectives are needed to construct paths forward. Through such a process, we might identity people trained to calculate carbon emissions, people with lived expertise who live nearby large infrastructure and can speak to the impact on their lives, people with historical expertise who can help understand how the introduction of infrastructure shifts place over time, and people with emerging expertise representing the perspectives of non-human stakeholders in sociotechnical design processes. This example begins to illustrate how attending to who is impacted (stakeholders) is complementary to attending to whose knowledge and perspectives are needed to construct paths forward (experts and expertise).

2. *Focus on the topics each stakeholder (or expert) is well positioned to weigh in on.* VSD does not address if there are topics that particular stakeholders are more (or less) well positioned to weigh in on. In contrast, we intuit that different experts will be well positioned to weigh in on different topics, and the successful implementation of this approach will likely hinge on engaging experts on topics that align with their expertise.

In this appendix, by delineating between existing design approaches (i.e., UCD and VSD), we begin to articulate how attending to experts and expertise in social impact evaluations of generative AI (and beyond) may provide new, constructive, and complementary paths forward.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: In the abstract, we claim to attend to experts–and expertise–in social impact evaluations of generative AI; that is the contribution of the paper.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [No]

   Justification: This provocation makes clear that attending to expertise in social impact evaluations is early work and therefore the limitations are not yet known. To that end, we end the paper with five open questions which point to areas for future exploration and research.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: This paper does not provide any theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [NA]

Justification: This paper does not include any experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [NA]

   Justification: This paper does not include experiments requiring code.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: This paper does not include any experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: This paper does not include any experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: This paper does not include any experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have reviewed the NeurIPS Code of Ethics and confirm that the paper conforms to the code.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: Given the short (2-page) length of the provocation, we do not include a separate section on broader impacts. However, the paper attends to the types of experts—and expertise—to conduct more robust social impact evaluations of generative AI, and ends with a set of questions that point to potential limitations.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: This paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: This paper does not use existing assets.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: This paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This article does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.