BABL AI Inc.
The Algorithmic Bias Lab
630 Fairchild Street
Iowa City, Iowa 52245
https://babl.ai

From: **Borhane Blili-Hamelin, Ph.D.**
Associate, AI Audit & Assurance
Ethical Risk & Impact Assessment Team
BABL AI Inc.
borhane.blilihamelin@bablai.com

To: **Federal Trade Commission**
Office of the Secretary
600 Pennsylvania Avenue NW
Suite CC-5610 (Annex B)
Washington, DC 20580

Re: **Commercial Surveillance ANPR, R111004**

Nov. 21, 2022

To Whom It May Concern:

On behalf of the team at BABL AI, I'd like to thank the Commission for providing us the opportunity to comment on Commercial Surveillance ANPR, R111004, and we would be happy to provide further clarification on any of the comments below.

BABL AI is a leading boutique consultancy focusing on responsible AI governance, algorithm risk and impact assessments, algorithmic bias assessments and audits, and corporate training on responsible AI. We also conduct research and provide education on AI governance and ethics. From our perspective, the FTC has a unique opportunity to address gaps in the current AI governance ecosystem.

Executive Summary

Our main recommendations are as follows:

1. (Questions 53, 54 and 56) **The FTC can help improve organizations' approach to testing ADS for accuracy, validity, fairness, or robustness by incentivizing the adoption of *algorithmic impact assessments* or *ethical risk assessments*. The FTC would ideally also require companies who deploy or procure ADS to disclose:**

    a. The range of contexts where the tool is likely not to work.
    b. What impact assessment or ethical risk assessment were conducted (if any).
    c. The technical tests and metrics they used in measuring error rates, and what their confidence intervals are.

2. (Question 92) **The FTC should promote and in some cases require (e.g. for high-risk applications or for algorithms that affect a very large number of US consumers) independent but cooperative third-party audits of ADS**.
    a. **Conflicts of interest** between auditor and auditee are a known problem in other sectors where we rely on audits to help ensure accountability (e.g. finance).
    b. Adversarial third-party audits of ADS often suffer from limited access to the components of the ADS.
    c. Cooperative third-party audits can achieve meaningful independence through limits on cross-selling non-audit services, access requirements, the professionalization and standardization of audit services, and ensuring auditor liability for serious violations of their professional responsibilities. (Raji et al. 2022; ForHumanity)
    d. (**Transparency**) The FTC could also help organizations balance the benefits and harms of public disclosure by providing clear standards for transparency and public disclosure of audit results.

3. (Question 69) Organizations attempting to protect consumers from algorithmic discrimination against members of protected categories would benefit from more clarity about the implications of current anti-discrimination laws and norms. **The FTC can help clarify what federal anti-discrimination and non-discrimination norms require in cases where ADS or AI contribute to such discrimination.**
    a. The FTC can help by clarifying what counts as unfair practices against protected groups in cases like proxy discrimination, discrimination due to disparity in access, and cases where algorithms discriminate by accurately replicating real-world disparate impacts.
    b. The FTC can also help by requiring organizations who deploy or procure ADS to compare ADS outcomes for members of protected groups to the outcomes of the baseline legacy systems that would be the alternative to an ADS.

Below, we provide more context for our recommendations and supporting evidence in the form of references to peer-reviewed articles, white papers, and books.

Our Perspective

BABL AI is conducting a research project on the effectiveness of currently implemented **AI governance initiatives** across the private and public sectors. We ask organizations whether they have used tools such as engaging in audits, creating documentation, or conducting analyses of risks and how effective they find these initiatives.

We notice that many organizations have moved beyond the stage of developing AI governance principles and frameworks. Yet our early findings suggest that the field **struggles** with converging upon metrics for the success of governance programs, measuring their effectiveness, and establishing agreed-upon best practices.

We believe that the FTC is in a unique position to help address these challenges by setting authoritative baselines for due diligence in protecting consumers from unfair commercial surveillance and data security practices.

Given our expertise in AI governance, risk assessment, and auditing, our public comments below focus especially on questions about automated decision-making systems (ADS) errors, algorithmic discrimination, and disclosure requirements.

1. (Questions 53, 54 and 56) We recommend requiring impact assessments and ethical risk assessments to protect consumers from algorithmic errors

The questions we are targeting with this recommendation are:

- 53 How prevalent is algorithmic error? To what extent is algorithmic error inevitable? If it is inevitable, what are the benefits and costs of allowing companies to employ automated decision-making systems in critical areas, such as housing, credit, and employment? To what extent can companies mitigate algorithmic error in the absence of new trade regulation rules?
- 54 What are the best ways to measure algorithmic error? Is it more pronounced or happening with more frequency in some sectors than others?
- 56 To what extent, if at all, should new rules require companies to take specific steps to prevent algorithmic errors? If so, which steps? To what extent, if at all, should the Commission require firms to evaluate and certify that their reliance on automated decision-making meets clear standards concerning accuracy, validity, reliability, or error? If so, how? Who should set those standards, the FTC or a third-party entity? Or should new rules require businesses to evaluate and certify that the accuracy, validity, or reliability of their commercial surveillance practices are in accordance with their own published business policies?

1.1 RISK ASSESSMENTS AND SOCIO-TECHNICAL SYSTEMS

We believe that examining algorithmic errors as an **isolated technical problem** is often a mistake when the concern is protecting consumers. Instead, where the goal is preventing harmful and prevalent errors, effective technical measurement of ADS accuracy, validity, bias, fairness, interpretability, and robustness needs to be thoroughly **informed by algorithmic impact assessments or ethical risk assessments** that: (1) "take into account the socio-technical context of the algorithm", (2) "how the algorithm is employed to serve certain purposes of an organization," and (3) "how it affects the rights and interests of stakeholders—including whether it is unfair or biased in some way." (Hasan et al. 2022)

Technical tests must be informed by risk assessments because ADS are inherently **socio-technical systems**. (Brown et al. 2021; Selbst et al. 2019) This concept from the field of Science and Technology studies captures how technologies irreducibly involve both human and technical components, not only in their impact but also in their very construction. The real-world impacts of ADS cannot be adequately evaluated without considering both (a) the technical and human aspects of ADS design, deployment, or procurement and (b) the complex range of stakeholders whose interests and rights an ADS might impact.

Consider the example of an ADS "used to help make hiring decisions by screening and ranking candidates that are qualified and fit for a particular job." (Hasan et al. 2022)[1] There are always time and resource constraints on the range of technical tests that can be performed to examine a given ADS's validity, accuracy, robustness, and fairness. "One cannot test for everything, and the initial ethical risk assessment can help one decide what sorts of questions to ask about the algorithm, and what possible errors and biases to test for. All this will in turn inform one's selection of testing data and testing metrics." (Hasan et al. 2022)" Given the abundance of options in selecting technical tests for algorithmic error, developers are often tempted to default to selecting the technical tests that are most familiar or convenient for them. One of the many benefits of algorithmic risk assessments that take a socio-technical lens is that they enable selecting and prioritizing the technical tests of algorithmic errors that are most likely to help mitigate potential harms to consumers.[2]

---

[1] "Let us take as an example software used to help make hiring decisions by screening and ranking candidates that are qualified and fit for a particular job. It might use natural language processing to parse candidate resumés (and perhaps other application materials, online profiles, etc.) that vary in style and structure and feed the information into a ranking algorithm. Knowledge of the purpose, and specifically of the required or desired qualifications for the job, is relevant to determining which features of the data are likely to be good or reliable indicators of these qualifications—and, importantly, whether they are appropriate and fair."

[2] Here are three other examples from BABL AI peer-reviewed papers.

1. "If actions are taken autonomously based on some threshold value of the output, details of how this threshold was decided upon and justified are important to the auditing process as well. Some of this information can be extracted directly from the context, while other pieces will involve detailed testing of the algorithm's response to different inputs. An example could be the automatic screening of new tenant applications in the rental housing market, where,

1.2 CONFIDENCE INTERVALS

Moreover, we believe that **confidence intervals** should be much more widely disclosed in connection with ADS error rates. ADS are statistical tools: they make probabilistic estimates or predictions based on data. In the scientific community, confidence intervals are a widely accepted way to represent and communicate uncertainty about statistical estimates, including for ML systems. (Hüllermeier & Waegeman (2021); Schwartz et al. (2022))

In practice, we are already seeing insistence on the importance of confidence intervals in emergent regulation efforts. For instance, the current version of the Stop Discrimination by Algorithms Act of 2021 of the Council of the District of Columbia recommends documenting and reporting "any performance metrics the entity uses to gauge the accuracy of the assessments, including accuracy, **confidence intervals**, and how those assessments are obtained."

1.3 RECOMMENDATIONS

The FTC can help improve organizations' approach to testing ADS for accuracy, validity, fairness, or robustness by **incentivizing** the adoption of algorithmic impact assessments or ethical risk assessments.

The FTC would ideally also **require** companies who deploy or procure ADS to disclose:

- The range of contexts where the tool is likely not to work.
- What impact assessment or ethical risk assessment were conducted (if any).

---

e.g., any application scoring lower than X is automatically rejected. It is important to find out how the cutoff of X was found and justified. One would also want to test how slight changes in the cutoff differentially affect people belonging to different socio-economic groups." (Brown et al. 2021)

2. "Let us illustrate again with a fictitious AI essay-grading algorithm example, focusing on just two metrics and two interests for the student stakeholder: privacy and non-discrimination for the interests, and societal bias and transparency of data use and collection as the two metrics. Essay grading algorithms are widely used and provide automatic assessment of students' essays. However, the details of how much the student knows and consents to what happens to their data is highly relevant to privacy, independent of what those details actually are (which would be assessed in the metric score). For non-discrimination, societal bias is highly relevant, and for the purposes of this hypothetical example we score the transparency of data use and collection as medium relevance for non-discrimination. This stems from the possibility that student data is used to update the algorithm in some way, which could lead to compounding bias." (Brown et al. 2021)
3. "For example, suppose that a facial recognition algorithm used by law enforcement sometimes mistakes innocent persons for suspects or persons of interest. Learning that the system relies heavily on a database of mugshots populated primarily by persons of color suggests that such persons may be particularly vulnerable to unfair treatment. Or, to give another example, discovering that an algorithm used by an autonomous vehicle is poor at categorizing smaller bodies, or poor at tracking their behavior once they move behind and are occluded by vehicles or other objects, could raise risks to the safety of bicyclists, pedestrians, and children." (Hasan et al 2022)

- The technical tests and metrics they used in measuring error rates, and what their confidence intervals are.

SOURCES

- B24-0558—Stop Discrimination by Algorithms Act of 2021, *Council of the District of Columbia* (2021) (testimony of Phil Mendelson). https://lims.dccouncil.gov/downloads/LIMS/48421/Introduction/B24-0558-Introduction.pdf
- Brown, S., Davidovic, J., & Hasan, A. (2021). The algorithm audit: Scoring the algorithms that score us. *Big Data & Society*, 8(1), 205395172098386. https://doi.org/10.1177/2053951720983865
- Costanza-Chock, S., Raji, I. D., & Buolamwini, J. (2022). Who Audits the Auditors? Recommendations from a field scan of the algorithmic auditing ecosystem. *Proceedings of the Conference on Fairness, Accountability, and Transparency.*. https://doi.org/10.1145/3531146.3533213
- Hasan, A., Brown, S., Davidovic, J., Lange, B., & Regan, M. (2022). Algorithmic Bias and Risk Assessments: Lessons from Practice. *Digital Society*, 1(2), 14. https://doi.org/10.1007/s44206-022-00017-z
- Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning,* 110(3), 457–506. https://doi.org/10.1007/s10994-021-05946-3
- Raji, I. D., Xu, P., Honigsberg, C., & Ho, D. (2022). Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 557–571. https://doi.org/10.1145/3514094.3534181
- Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and Abstraction in Sociotechnical Systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68. https://doi.org/10.1145/3287560.3287598
- Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). Towards a Standard for Identifying and Managing Bias in Artificial Intelligence. *National Institute of Standards and Technology*. https://doi.org/10.6028/NIST.SP.1270

2. (Questions 92 and 51) We recommend promoting independent third-party audits

The questions we are targeting with this recommendation are:

- 92. To what extent should the Commission, if at all, make regular self-reporting, third-party audits or assessments, or self-administered impact assessments about commercial surveillance practices a standing obligation? How frequently, if at all, should the Commission require companies to disclose such materials publicly? If it is not a standing obligation, what should trigger the publication of such materials?

- 51. To what extent, if at all, should the Commission require firms to certify that their commercial surveillance practices meet clear standards concerning collection, use, retention, transfer, or monetization of consumer data? If promulgated, who should set those standards: the FTC, a third-party organization, or some other entity?

2.1 INDEPENDENCE

The FTC should promote and in some cases require (e.g. for high-risk applications or for algorithms that affect a very large number of US consumers) independent but cooperative third-party audits of ADS.

Conflicts of interest between auditor and auditee are a known problem in sectors like finance, where we rely on audits to help ensure accountability and trust and to protect the public against a lack of due diligence.

Since 2020 BABL has been collaborating with ForHumanity on tackling this problem through **cooperative third-party auditing** standards and certification for ADS inspired by the Sarbanes-Oxley Act. Such audits are **cooperative:** because the audited party agrees to being audited; and because they provide auditors the assistance, access, and information they need to conduct the audits effectively. (For a helpful definition of cooperation, see Article 37 of the EU Digital Services Act.) Such audits are **independent** in the sense that: the audit provider "must receive no other remuneration [...] than the audit fees". (Carrier & Brown) BABL AI and ForHumanity also protect independence in their license agreements by stipulating that an auditor cannot provide any other services to the auditee for a period of 12 months.

Cooperative third-party audit requirements are also favored by recent peer-reviewed research, such as "Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance" by Raji et al (2022).

Much of the current AI audit ecosystem relies on (a) **internal audits** (performed by employees of the audited organization), and (b) on audits performed by parties contracted by the audited organization **without any rigorous safeguards against conflicts of interest**.

One alternative is **adversarial third-party evaluations**: conducted *without the cooperation of the audited party* by independent parties: such as civil society organizations (e.g. the 2018 ACLU evaluation of Amazon's "Rekognition" tool), journalists (e.g. ProPublica's 2016 evaluation of COMPAS), and academic researchers (e.g. Buolamwini & Gebru's Gender Shades (2018); Rhea et al.'s 2022 evaluation of Humantic AI and Crystal)). However, adversarial third-party assessments suffer from minimal access to the components of the ADS. They can thus only bring visibility to a limited range of ADS risks and failures. For that reason, as Raji et al. point out, "while third-party oversight by non-profit organizations, journalists, and academics have [...] played critical roles in exposing weaknesses to AI systems, such forms of oversight are not conventionally considered audits per se." (Raji et al. 2022)

Especially when supported by government rules and regulations, cooperative third-party audits can achieve a significant level of **independence** (and help mitigate conflicts of interests) through measures like those required by the Sarbanes-Oxley Act: including limits on cross-selling non-audit services, access requirements, the professionalization and standardization of audit services, and ensuring auditor liability for serious violation of their professional responsibilities. (Raji et al. 2022; ForHumanity)

2.2 TRANSPARENCY

Transparency is an important issue in connection with third-party audit requirement. Who should companies be required to communicate audit results to? What (if any) information about audit results should be communicated to the general public? Public disclosures can be beneficial in empowering outside actors to play an effective role in helping protect the public interest from algorithmic harms. Likewise, transparency helps build trust with outside parties. However, it is vital to recognize that transparency can also sometimes be detrimental. (In philosophical terms, transparency is arguably an instrumental good: its value depends on what further ends it serves. (Blackman (2022)) For instance, public disclosure of cybersecurity vulnerabilities can empower threat actors to exploit the vulnerability. Moreover, public disclosure of personal data can be extremely harmful to individuals and incur legal consequences for the party at fault. (For a helpful case study of the complex benefits and potential harms of transparency focused on open-source deepfake tools, see Widder et al. (2022)) **We believe that the FTC could help organizations navigate the balance between the benefits and harms of public disclosure by providing clear standards for transparency and public disclosure of audit results.**

2.3 RECOMMENDATIONS

BABL is committed to pioneering cooperative third-party algorithmic audits, even without government guidance that requires it. In December 2022, we will launch one of the first attempts at providing scalable third-party algorithmic audit services focused on audits of employments algorithm in the context of New York City Local Law 144.

However, we believe that government norms, rules, and incentives from agencies like the FTC are critical to building a thriving ecosystem of third-party audits that genuinely protect the public interest and US consumers. **The FTC is uniquely positioned to incentivize and help standardize cooperative third-party audit requirements and certification aimed at protecting consumers against unfair and deceptive ADS practices**.

Realistically, it may be impractical for the FTC to require certified third-party audits for all ADS. One possible strategy, similar to the EU AI act, would be for the FTC to mandate third-party audits for ADS that the FTC considers to pose a *high risk* to US consumers.

Another possibility, similar to the recent EU Digital Services Act (DSA), would be to set independent audit requirements only for algorithms that affect a **very large** number of US consumers. In the case of the DSA, the threshold for "very large" platforms and services is set at roughly 10% of the EU population, "calculated as an average over a period of six months". (DSA) Compliance requirements that apply no matter the size of the platform risk disproportionally benefiting very large actors, who can more easily afford to meet the requirements (and who often spend large amounts of money on algorithmic governance already.) If the FTC wants to avoid being disproportionally more disadvantageous to small companies, a strategy analogous to the EU DSA might be appropriate.

SOURCES

- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. *ProPublica.* https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
- Blackman, R. (2022). *Ethical machines: Your concise guide to totally unbiased, transparent, and respectful AI*. Harvard Business Review Press.
- Brown, S., Davidovic, J., & Hasan, A. (2021). The algorithm audit: Scoring the algorithms that score us. *Big Data & Society*, 8(1), 205395172098386. https://doi.org/10.1177/2053951720983865
- Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In S. A. Friedler & C. Wilson (Eds.), *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (Vol. 81, pp. 77–91). PMLR. https://proceedings.mlr.press/v81/buolamwini18a.html
- Carrier, R., & Brown, S. (2021). Taxonomy: AI Audit, Assurance & Assesment. *ForHumanity*. https://forhumanity.center/blog/taxonomy-ai-audit-assurance-assessment/
- Cochran, C. (2015). *ISO 9001:2015 in Plain English*. Paton Professional.
- Costanza-Chock, S., Raji, I. D., & Buolamwini, J. (2022). Who Audits the Auditors? Recommendations from a field scan of the algorithmic auditing ecosystem. *FAccT '22*. https://doi.org/10.1145/3531146.3533213
- Hasan, A., Brown, S., Davidovic, J., Lange, B., & Regan, M. (2022). Algorithmic Bias and Risk Assessments: Lessons from Practice. *Digital Society*, 1(2), 14. https://doi.org/10.1007/s44206-022-00017-z
- (DSA) Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and Amending Directive 2000/31/ec (Digital Services Act). (2022). *Official Journal of the European Union*, L 277.

https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2022:277:FULL&from=EN

- Raji, I. D., Xu, P., Honigsberg, C., & Ho, D. (2022). Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 557–571. https://doi.org/10.1145/3514094.3534181
- Rhea, A., Markey, K., D'Arinzo, L., Schellmann, H., Sloane, M., Squires, P., & Stoyanovich, J. (2022). Resume Format, LinkedIn URLs and Other Unexpected Influences on AI Personality Prediction in Hiring: Results of an Audit. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 572–587. https://doi.org/10.1145/3514094.3534189
- Snow, J. (2018, July 24). Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots | News & Commentary. American Civil Liberties Union. https://www.aclu.org/news/privacy-technology/amazons-face-recognition-falsely-matched-28
- Widder, D. G., Nafus, D., Dabbish, L., & Herbsleb, J. (2022). Limits and Possibilities for "Ethical AI" in Open Source: A Study of Deepfakes. 2022 *ACM Conference on Fairness, Accountability, and Transparency*, 2035–2046. https://doi.org/10.1145/3531146.3533779

3. We recommend providing more clarity about the implications of current anti-discrimination law and norms for algorithmic discrimination

The question we are targeting with this recommendation is:

- "Question 69 Should the Commission consider new rules on algorithmic discrimination in areas where Congress has already explicitly legislated, such as housing, employment, labor, and consumer finance? Or should the Commission consider such rules addressing all sectors?"

Organizations attempting to protect consumers from algorithmic discrimination against members of protected categories would benefit from more clarity about the implications of current anti-discrimination laws and norms. The FTC can help clarify what federal anti-discrimination and non-discrimination norms require in cases where ADS or AI contribute to such discrimination.

The FTC can help by clarifying what counts as unfair practices against protected groups in cases like proxy discrimination[3], discrimination due to disparity in access, and cases where algorithms discriminate by accurately replicating real-world disparate impacts.

[3] "An example is the fact that smartphone data oversamples those that can afford smartphones, and algorithms making use of this data may favor the preferences of those groups for which wealth is a **proxy**." Brown et al. 2021

The FTC can also help by requiring organizations who deploy or procure ADS to compare ADS outcomes for members of protected groups to the outcomes of the baseline legacy systems that would be the alternative to an ADS.

Conclusion

On behalf of the entire team at BABL AI, I would like to thank the FCT for providing us the opportunity to comment on these vital questions for algorithmic governance. We would be happy to provide further clarification on any of the above questions.

Contact

Borhane Blili-Hamelin, Ph.D.
Associate, AI Audit & Assurance
Ethical Risk & Impact Assessment Team
BABL AI Inc.
borhane.blilihamelin@bablai.com