Volume 4, No. 1 | January–February 2021

RAIL

The Journal of Robotics, Artificial Intelligence & Law

fastcase FULL COURT PRESS

RAIL
The Journal of Robotics, Artificial Intelligence & Law
Volume 4, No. 1 | January–February 2021

THE JOURNAL OF ROBOTICS, ARTIFICIAL INTELLIGENCE & LAW (ISSN 2575-5633 (print) /ISSN 2575-5617 (online) at $495.00 annually is published six times per year by Full Court Press, a Fastcase, Inc., imprint. For customer support, please contact Fastcase, Inc., 711 D St. NW, Suite 200, Washington, D.C. 20004, 202.999.4777 (phone), 202.521.3462 (fax), or email customer service at support@fastcase.com.

Publishing Staff
Publisher: Morgan Morrissette Wright
Journal Designer: Sharon D. Ray
Cover Art Design: Juan Bustamante

Cite this publication as:

The Journal of Robotics, Artificial Intelligence & Law (Fastcase)

A Full Court Press, Fastcase, Inc., Publication

Editorial Office

711 D St. NW, Suite 200, Washington, D.C. 20004
https://www.fastcase.com/

POSTMASTER: Send address changes to THE JOURNAL OF ROBOTICS, ARTIFICIAL INTELLIGENCE & LAW, 711 D St. NW, Suite 200, Washington, D.C. 20004.

Articles and Submissions

Direct editorial inquiries and send material for publication to:

Steven A. Meyerowitz, Editor-in-Chief, Meyerowitz Communications Inc., 26910 Grand Central Parkway, #18R, Floral Park, NY 11005, smeyerowitz@meyerowitzcommunications.com, 646.539.8300.

Material for publication is welcomed—articles, decisions, or other items of interest to attorneys and law firms, in-house counsel, corporate compliance officers, government agencies and their counsel, senior business executives, scientists, engineers, and anyone interested in the law governing artificial intelligence and robotics. This publication is designed to be accurate and authoritative, but neither the publisher nor the authors are rendering legal, accounting, or other professional services in this publication. If legal or other expert advice is desired, retain the services of an appropriate professional. The articles and columns reflect only the present considerations and views of the authors and do not necessarily reflect those of the firms or organizations with which they are affiliated, any of the former or present clients of the authors or their firms or organizations, or the editors or publisher.

QUESTIONS ABOUT THIS PUBLICATION?

For questions about the Editorial Content appearing in these volumes or reprint permission, please contact:

Morgan Morrissette Wright, Publisher, Full Court Press at mwright@fastcase.com or at 202.999.4878

For questions or Sales and Customer Service:

Customer Service
Available 8 a.m.–8 p.m. Eastern Time
866.773.2782 (phone)
support@fastcase.com (email)

Sales
202.999.4777 (phone)
sales@fastcase.com (email)
ISSN 2575-5633 (print)
ISSN 2575-5617 (online)

Everything Is Not *Terminator*

Assessment of Artificial Intelligence Systems

John Frank Weaver*

Many information security and privacy laws such as the California Consumer Privacy Act[1] and the New York Stop Hacks and Improve Electronic Data Security Act[2] require periodic assessments of an organization's information management systems. Because many organizations collect, use, and store personal information from individuals—much of which could be used to embarrass or impersonate those individuals if inappropriately accessed—these laws require organizations to regularly test and improve the security they use to protect that information.

As of yet, there is no similar specific law in the United States directed at artificial intelligence systems ("AIS"), requiring the organizations that rely on AIS to test its accuracy, fairness, bias, discrimination, privacy, and security.

However, existing law is broad enough to impose on many organizations a general obligation to assess their AIS, and legislation has appeared requiring certain entities to conduct impact assessments on their AIS. Even without a regulatory mandate, many organizations should perform AIS assessments as a best practice.

This column summarizes current and pending legal requirements before providing more details about the assessment process.

Federal Trade Commission and Algorithmic Accountability Acts

The Federal Trade Commission's ("FTC") authority to police "unfair or deceptive acts or practices in or affecting commerce" through rule making and administrative adjudication is broad enough to govern AIS, and it has a department that focuses on algorithmic transparency, the Office of Technology Research and Investigation.[3] However, the FTC has not issued clear guidance

Robotics, Artificial Intelligence & Law / January–February 2021, Vol. 4, No. 1, pp. 67–75.

ISSN 2575-5633 (print) /ISSN 2575-5617 (online)

regarding AIS uses that qualify as unfair or deceptive acts or practices. There are general practices that organizations can adopt that will minimize their potential for engaging in unfair or deceptive practices, which include conducting assessments of their AIS.[4] However, there is no specific FTC rule obligating organizations to assess their AIS.

There have been some legislative efforts to create such an obligation, including the Algorithmic Accountability Act,[5] which was proposed in Congress, and a similar bill proposed in New Jersey,[6] both in 2019.

The federal bill would require covered entities to conduct "impact assessments" on their "high-risk" AIS in order to evaluate the impacts of the AIS's design process and training data on "accuracy, fairness, bias, discrimination, privacy, and security."[7]

The New Jersey bill is similar, requiring an evaluation of the AIS's development process, including the design and training data, for impacts on "accuracy, fairness, bias, discrimination, privacy, and security," and must include several elements, including a "detailed description of the best practices used to minimize the risks" and a "cost-benefit analysis."[8] It would also require covered entities to work with external third parties, independent auditors, and independent technology experts to conduct the assessments, if reasonably possible.[9]

Although neither of these has become law, they represent the expected trend of emerging regulation.[10]

The Need to Require AIS Assessments

When organizations rely on AIS to make or inform decisions or actions that have legal or similarly significant effects on individuals, it is reasonable for governments to require that those organizations also conduct periodic assessments of the AIS. For example, state criminal justice systems have begun to adopt AIS that use algorithms to report on a defendant's risk to commit another crime, risk to miss his or her next court date, etc.; human decision makers then use those reports to inform their decisions.[11]

The idea is that the AIS can be a tool to inform decision makers—police, prosecutors, judges—to help them make better, data-based decisions that eliminate biases they may have against

defendants based on race, gender, etc.[12] This is potentially a wonderful use for AIS, but only if the AIS actually removes inappropriate and unlawful human bias rather than recreate it.

Unfortunately, the results have been mixed at best, as there is evidence suggesting that some of the AIS in the criminal justice system is merely replicating human bias.

In one example, an African-American teenage girl and a white adult male were each convicted of stealing property totaling about $80. An AIS determined that the white defendant was rated as a lower recidivism risk than the teenager, even though he had a much more extensive criminal record, with felonies versus juvenile misdemeanors. Two years after their arrests, the AIS recommendations were revealed to be incorrect: the male defendant was serving an eight-year sentence for another robbery; the teenager had not committed any further crimes.[13] Similar issues have been observed in AIS used in hiring,[14] lending,[15] health care,[16] and school admissions.[17]

Although some organizations are conducting AIS assessments without a legal requirement, a larger segment is reluctant to adopt the assessments as a best practice, as many for-profit companies care more about accuracy to the original data used to train their AIS than they do about eliminating the biases in that original data.[18] According to Daniel Soukup, a data scientist with Mostly AI, a start-up experimenting with controlling biases in data, "There's always another priority, it seems.... You're trading off revenue against making fair predictions, and I think that is a very hard sell for these institutions and these organizations."[19]

I suspect, though, that the tide will turn in the other direction in the near future, with or without a direct legislative impetus, similar to the trend in privacy rights and operations. Although most companies in the United States are not subject to broad privacy laws like the California Consumer Privacy Act or the European Union's General Data Protection Regulation, I have observed an increasing number of clients that want to provide the privacy rights afforded by those laws, either because their customers expect them to or they want to position themselves as companies that care about individuals' privacy.

It is not hard to see a similar trend developing among companies that rely on AIS. As consumers become more aware of the problematic issues involved in AIS decision-making—potential bias,

use of sensitive personal information, security of that information, the significant effects, lack of oversight, etc.—they will become just as demanding about AIS requirements as privacy requirements. Similar to privacy, consumer expectations will likely be pushed in that direction by jurisdictions that adopt AIS assessment legislation, even if they do not live in those jurisdictions.

The Basics of an AIS Assessment

Organizations that are looking to perform AIS assessments now in anticipation of regulatory activity and consumer expectations should conduct an assessment consistent with the following principles and goals:

1. Retain a Neutral Third Party

Consistent with the New Jersey Algorithmic Accountability Act, any AIS assessment should be done by an outside party, preferably by qualified AI counsel, who can retain a technological consultant to assist them. This performs two functions.

First, it will avoid the situation in which the developers that created the AIS for the organization are also assessing it, which could result in a conflict of interest, as the developers have an incentive to assess the AIS in a way that is favorable to their work.

Second, by retaining outside AI counsel, in addition to benefiting from the counsel's expertise, organizations are able to claim that the resulting assessment report and any related work product is protected by attorney-client privilege in the event that there is litigation or a government investigation related to the AIS. Companies that experience or anticipate a data security breach or event retain outside information security counsel for similar reasons, as the resulting breach analysis could be discoverable if outside counsel is not properly retained. The results can be very expensive if the breach report is mishandled.

For example, Capital One recently entered into an $80 million Consent Order with the Department of Treasury related to a data incident following an order from a federal court that a breach report prepared for Capital One was not properly coordinated through outside counsel and therefore not protected by attorney-client privilege.[20]

2. Identify Risks

An AIS assessment should identify, catalogue, and describe the risks of an organization's AIS.

- Does the AIS process sensitive data, including data that identify a person's membership in a federally protected group (*e.g.*, race, gender)?
- Does the organization maintain security and privacy measures appropriate to that data?
- What populations are most affected by the AIS?
- Are those populations at risk because of the legal or similarly significant effects of the AIS's decisions?
- Could the AIS's decisions produce disparate impacts or unequal outcomes that violate legal prohibitions or conflict with the organization's principles?
- Could the AIS's decisions exacerbate known forms of discriminatory activity, biases, or unfairness in the organization's field?

Properly identifying these risks, among others, and describing how the AIS impacts each will allow an organization to understand the issues it must address to improve its AIS.[21]

3. Develop Notices to Impacted Populations

Once the risks in the AIS are identified, the assessment should focus on how the organization alerts impacted populations. This can be in the form of a public-facing AI policy, posted and maintained in a manner similar to an organization's privacy policy.[22] This can also be in the form of more pointed pop-up prompts, a written disclosure and consent form, automated verbal statement in telephone interactions, etc. The appropriate form of the notice will depend on a number of factors, including the organization, the AIS, the at-risk populations, the nature of the risks involved, etc. The notice should include the relevant rights regarding AIS afforded by privacy laws and other regulations.

4. Establish Process to Accept Comments in Response to Notices

After implementing appropriate notices, the organization should anticipate receiving comments from members of the

impacted populations and the general public. The assessment should help the organization implement a process that allows it to accept, respond to, and act on those comments. This may be similar to how organizations process privacy rights requests from consumers and data subjects, particularly when a notice addresses those rights. The assessment may recommend that certain employees be tasked with accepting and responding to comments, the organization add operative capabilities that address privacy rights impacting AIS or risks identified in the assessment and objected to by comments, etc. It may be helpful to have a technological consult provide input on how the organization can leverage its technology to assist in this process.

5. Propose Remediation

The assessment should help the organization remediate identified risks. The nature of the remediation will depend on the nature of the risks, the AIS, and the organization. Any outside AIS counsel conducting the assessment needs to be well-versed in the various forms remediation can take. In some instances, properly noticing the risk to the relevant individuals will be sufficient, per both legal requirements and the organization's principles. Other risks cannot or should not be "papered over," but rather obligate the organization to reduce the AIS's potential to injure.[23] This may include adding more human oversight, at least temporarily, to check the AIS's output for discriminatory activity or bias. A technology consultant may be able to advise the organization regarding revising the code or procedures of the AIS to address the identified risks.

Additionally, where the AIS is evidencing bias because of the data used to train it, more appropriate historical data or even synthetic data may be used to retrain the AIS to remove or reduce its discriminatory behavior.[24]

Conclusion

All organizations that rely on AIS to make decisions that have legal or similarly significant effects on individuals should periodically conduct assessments of their AIS. This is true for all organizations: for-profit companies, non-profit corporations, governmental entities, educational institutions, etc. Doing so will help them avoid potential legal trouble in the event their AIS is

inadvertently demonstrating illegal behavior and ensure the AIS acts consistently with the organization's values.

Organizations that adopt assessments earlier rather than later will be in a better position to comply with AIS-specific regulation when it appears and to develop a brand as an organization that cares about fairness.

Notes

* John Frank Weaver, a member of McLane Middleton's privacy and data security practice group, is a member of the Board of Editors of *The Journal of Robotics, Artificial Intelligence & Law* and writes its "Everything Is Not *Terminator*" column. Mr. Weaver, who may be contacted at john.weaver@mclane.com, has a diverse technology practice that focuses on information security, data privacy, and emerging technologies, including artificial intelligence, self-driving vehicles, and drones.

1. Cal. Civ. Code § 1798.150 (granting private right of action when a business fails to "maintain reasonable security procedures and practices appropriate to the nature of the information," with assessments necessary to identify reasonable procedures).

2. New York General Business Law, Chapter 20, Article 39-F, §§ 899-bb.2(b)(ii)(A)(3) (requiring entities to assess "the sufficiency of safeguards in place to control the identified risks"), 899.2(b)(ii)(B)(1) (requiring entities to assess "risks in network and software design"), 899.2(b)(ii)(B)(2) (requiring entities to assess "risks in information processing, transmission and storage"), and 899.2(b)(ii)(C)(1) (requiring entities to assess "risks of information storage and disposal").

3. 15 U.S.C. § 45(b); 15 U.S.C. § 57a.

4. John Frank Weaver, "Everything Is Not *Terminator*: Helping AI to Comply with the Federal Trade Commission Act," *The Journal of Artificial Intelligence & Law* (Vol. 2, No. 4; July-August 2019), 291-299 (other practices include: establishing a governing structure for the AIS; establishing policies to address the use and/sale of AIS; establishing notice procedures; and ensuring third-party agreements properly allocate liability and responsibility).

5. Algorithmic Transparency Act of 2019, S. 1108, H.R. 2231, 116th Cong. (2019).

6. New Jersey Algorithmic Accountability Act, A.B. 5430, 218th Leg., 2019 Reg. Sess. (N.J. 2019).

7. Algorithmic Accountability Act of 2019, *supra* note 5, at §2(2) and 3(b).

8. New Jersey Algorithmic Accountability Act, *supra* note 6, at §2.

9. *Id.*, at §3.

10. For a fuller discussion of these bills and other emerging legislation intended to govern AIS, see Yoon Chae, "U.S. AI Regulation Guide: Legislative Overview and Practical Considerations," *The Journal of Artificial Intelligence & Law* (Vol. 3, No. 1; January-February 2020), 17-40.

11. *See* Jason Tashea, "Courts Are Using AI to Sentence Criminals. That Must Stop Now," *Wired* (April 17, 2017), https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/.

12. Julia Angwin, Jeff Larson, Surya Mattu, & Lauren Kirchner, "Machine Bias," *ProPublica* (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing ("The appeal of the [AIS's] risk scores is obvious… If computers could accurately predict which defendants were likely to commit new crimes the criminal justice system could be fairer and more selective about who is incarcerated and for how long.").

13. *Id.*

14. Jeffrey Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women," *Reuters* (October 9, 2018), https://uk.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUKKCN1MK08G (Amazon "realized its new system was not rating candidates for software developer jobs and other technical posts in a gender-neutral way").

15. Dan Ennis and Tim Cook, "Banking from AI lending models raises questions of culpability, regulation," *Banking Dive* (August 16, 2019), https://www.bankingdive.com/news/artificial-intelligence-lending-bias-model-regulation-liability/561085/#:~:text=Bill%20Foster%2C%20D%2DIL%2C,lenders%20for%20mortgage%20refinancing%20loans ("African-Americans may find themselves the subject of higher-interest credit cards simply because a computer has inferred their race").

16. Shraddha Chakradhar, "Widely used algorithm for follow-up care in hospitals is racially biased, study finds," *STAT* (October 24, 2019), https://www.statnews.com/2019/10/24/widely-used-algorithm-hospitals-racial-bias/ ("An algorithm commonly used by hospitals and other health systems to predict which patients are most likely to need follow-up care classified white patients overall as being more ill than black patients—even when they were just as sick").

17. DJ Pangburn, "Schools are using software to help pick who gets in. What could go wrong?" *Fast Company* (May 17, 2019), https://www.fastcompany.com/90342596/schools-are-quietly-turning-to-ai-to-help-pick-who-gets-in-what-could-go-wrong ("If future admissions decisions are based on past decision data, Richardson warns of creating an unintended feedback loop, limiting a school's demographic makeup, harming disadvantaged students, and putting a school out of sync with changing demographics.").

18. Todd Feathers, "Fake Data Could Help Solve Machine Learning's Bias Problem—If We Let It," *Slate* (September 17, 2020), https://slate.com/technology/2020/09/synthetic-data-artificial-intelligence-bias.html.

19. *Id.*

20. In the Matter of Capital One, N.A., Capital One Bank (USA), N.A., Consent Order (Document #2020-036), Department of Treasury, Office of the Comptroller of the Currency, AA-EC-20-51 (August 5, 2020), https://www.occ.gov/static/enforcement-actions/ea2020-036.pdf; *In re: Capital One Consumer Data Security Breach Litigation*, MDL No. 1:19md2915 (AJT/JFA) (E.D. Va. May 26, 2020).

21. For a great discussion of identifying risks in AIS, see Nicol Turner Lee, Paul Resnick, and Genie Barton, "Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms," *Brookings* (May 22, 2019), https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/.

22. For more discussion of public facing AI policies, see John Frank Weaver, "Everything Is Not *Terminator*: Public-Facing Artificial Intelligence Policies—Part I," *The Journal of Artificial Intelligence & Law* (Vol. 2, No. 1; January-February 2019), 59-65; John Frank Weaver, "Everything Is Not *Terminator*: Public-Facing Artificial Intelligence Policies—Part II," *The Journal of Artificial Intelligence & Law* (Vol. 2, No. 2; March-April 2019), 141-146.

23. For a broad overview of remediating AIS, see James Manyika, Jake Silberg, and Brittany Presten, "What Do We Do About Biases in AI?" *Harvard Business Review* (October 25, 2019), https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai.

24. There are numerous popular and academic articles exploring this idea, including Todd Feathers, "Fake Data Could Help Solve Machine Learning's Bias Problem—If We Let It," *Slate* (September 17, 2020), https://slate.com/technology/2020/09/synthetic-data-artificial-intelligence-bias.html, and Lokke Moerel, "Algorithms can reduce discrimination, but only with proper data," *IAPP* (November, 16, 2018), https://iapp.org/news/a/algorithms-can-reduce-discrimination-but-only-with-proper-data/.