

Comments Received by the Department of
Consumer and Worker Protection on

Proposed Rules related to Local Law 202 of 2019, Local Law 1144 of 2021,
and Local Law 37 of 2022



June 3, 2022

Hon. Vilda Vera Mayuga, Esq.
Commissioner
Department of Consumer and Worker Protection
42 Broadway #5, New York, NY 10004

Re: Local Law 37 of 2022 (Open Captioning)

Submitted Via Email to Rulecomments@dca.nyc.gov

Dear Commissioner Mayuga:

On behalf of the movie theatre trade association, NATO, Theatre Owners of New York State, Inc., thank you for the opportunity to submit written comments regarding Local Law 37 of 2022. The movie theatres will comply with the requirements of this open captioning law, and want to avoid confusion for either patrons or theatre managers. Therefore, to ensure a smooth implementation of this law, we are highlighting three areas of concern to mitigate customer complaints and costly fines. Additionally, we are providing further suggestions as additions to the frequently asked questions document to greater inform and educate the public.

The movie theatre industry wants all patrons to feel safe and welcome as they return to the theatres, including the members of the deaf and hard of hearing community. However, we do not want the complex language of the open captioning statute to lead to confusion amongst patrons as to the timing and format of movies shown, which could then result in expensive fines for the theatres. Movie theatre audiences are down almost 50%, an estimated 8% of the audience may never return, revenues are 10% what they were pre-pandemic, and overall annual box office sales are 25% of the lowest ever recorded. Avoiding confusion about the open captioning statute would help theatres avoid fines, and would also help the City's recovery, since in 2019 New York City audiences spent \$325 million at nearby bars, restaurants, and retail when seeing a movie.

First, the dynamic scheduling utilized by the movie theatre industry is an area of concern that could lead to complaints and fines against theatres. Dynamic scheduling allows theatres to increase or decrease available showtimes based upon audience demand. Consequently, if a particular movie is not drawing an audience, it will quickly be substituted for another movie in the schedule. Schedules are generally set by Monday for the following weekend's showtimes, and published on Tuesdays. Conceivably, a movie patron could look on a theatre company's website and see that a movie is playing in open captioning format for the following Sunday, and make their plans to see the movie in open captioning format. However, in the intervening week, the movie might not draw an audience, and consequently be removed from the schedule altogether. At the beginning of the week, that prospective theatre patron might see on the website that the movie would play at least 14 times, and therefore require 4 open captioning showtimes. By the end of the week, though, after a very poor performance, the movie might be played only 7 times, and then provide open captioning format for 2 showtimes. At the beginning or end of the week, that movie theatre would comply with the Local Law 37's required number of open captioning showtimes. Yet, that patron might still call 3-1-1 to complain that based on what they originally saw on the

theatre's website, they made plans for the weekend to see the movie in open captioning and it was not available. As the theatres seek to maintain compliance with the requirements of the open captioning statute, we hope that the theatre industry's use of dynamic scheduling does not lead to confusion amongst patrons, and subsequently lead to fines.


Second, we also want to raise concerns about Local Law 37's vague language that seemingly prohibits "overlap" of open captioning movies at a theatre. To contrast with the example above, in the instance when a movie is drawing an especially large and sustained audience, the theatres may decide to add more open captioning showtimes of that movie. However, in a multiplex theatre, adding more open captioning showtimes will necessarily mean that different movies will play at the same time in open captioning. If a large audience for a tent-pole superhero movie like "*Doctor Strange and the Multiverse of Madness*" necessitates more showtimes, and allows for additional open captioning showtimes, overlap with open captioning showtimes of films like "*Father Stu*" or "*Everything Everywhere All At Once*" would be unavoidable without a multiverse, divine intervention, or interdimensional assistance. Consequently, the theatres remain concerned that the statute's language on overlap could lead to complaints, citations, and costly fines, and we want to avoid situations where patrons are upset by such confusion regarding overlapping open captioning showtimes.

Third, there are many movies that are simply not made available by studios in open captioning format. Accordingly, we want to alleviate potential patron confusion about open captioning availability in the instance when certain titles are not even available in open captioning. Open captioning is entirely a studio decision, and not one that theatres can control. Currently, the majority of major studios do not offer captioning in the 3D format on their titles, and many of the so-called "art house" or smaller independent movies popular with an older audience demographic are not available in open captioning either. If possible, we would like to see greater communication and collaboration with the Mayor's Office of Media and Entertainment, the Department of Consumer and Worker Protection, the studios, and the movie theatre industry about what movies are available in open captioning format. Such collaboration might serve to quickly correct potential patron complaints about the lack of open captioning formatting for certain movies on a 3-1-1 call, instead of mistakenly leading to levying fines on theatres.

The movie theatres want to help facilitate a smooth implementation of Local Law 37, and will make a good faith effort to comply with all of the statute's requirements. We want nothing more than all audiences to return to the theatres, which remain the best place to watch a movie. We will also explore ways to coordinate with organizations representing the deaf and hard of hearing community to compliment the educational outreach that the City is making to inform consumers of their rights, and theatres of their responsibilities. However, whether a result of dynamic scheduling, an inability to avoid overlapping, the lack of open captioning formatting for a particular movie, or perhaps even mechanical or technological problems, the movie theatres anticipate that there may be unavoidable issues that will cause consumer complaints, and potentially costly fines.

We remain committed to collaborating with the City to ensure the successful implementation of Local Law 37. To this end, we are also providing suggested additions to the frequently asked questions below. We are happy to continue the discussion at your earliest convenience. Thank you for your time and the opportunity to submit testimony on this important matter.

Sincerely,



Robert Sunshine
Executive Director
Enclosure

917-940-5082

Suggested FAQ Additions

1) Why is this movie not available in open caption format?

Not all movies are made with open captions. Some studios make the decision to not make open caption available on their movies.

Movies are made by the studios and loaned to movie theatres to present to customers on the studios' behalf. Movie theatres have no control over the studios or the movie making process.

2) How do we know if a movie is available in open caption format?

Customers can go to a specific theatre's website to check to see which movies are available in open captioning format. Open caption showtimes are listed separately on the website. Some theatres will list the format as "Open Cap/Eng Sub", or "Open Caption (On-Screen Subtitles)", or "OC", depending on space.

3) Why were there more open caption showtimes on the website at the beginning of the week, but now less open caption showtimes?

Show schedules are dynamic in that they can change at any time. When a big event movie opens, movie theatres add showtimes and other movies are canceled to accommodate the more popular movie that is in demand. The ability to make changes to the number of movies shown based on supply and demand is like any other business. However, theatres will still comply with the mandate spelled out in the statute, which require the following:

	Total Showings Per Week of a Single Movie Title				
	1-3	4	5-8	9-12	13+
Minimum open-captioned showings	0	1	2	3	4

4) Why were two different movies shown in open captioning format at the same time?

Movie theatres want to give all films the best showtimes. Sometimes it is unavoidable to prevent overlapping whether a movie is shown in open captioning format, or not shown in open captioning format. When scheduling film times within the defined peak time periods of the statute, that means multiple movies will be shown during these time periods. This is due to varying length of movies, which is particularly prevalent during the Friday 6pm to 11pm window. Movie theatres are required to provide a total of two open captioning showtimes per film per week during the weekends, which include the Friday evening and Saturday or Sunday matinee or evening hours. They are also required to provide one open captioning showtime during the weekday evening hours, and one open caption showtime at theatre manager discretion. Balancing the requirements, with the number of showtimes per week is more of an art than a science, but theatres will certainly comply with the requirements and adjust schedules based upon demand.



Pesetsky & Bookman, PC

Attorneys at Law

325 Broadway, Suite 501
New York, NY 10007

(212) 513-1988 | www.PB.law

Max Bookman | Partner | max@pb.law

May 25, 2022

Commissioner Vilda Vera Mayuga
New York City Department of Consumer and Worker Protection
Via email only: rulecomments@dca.nyc.gov

Re: New York City Hospitality Alliance Comments on Proposed Rules
Implementing LL 202 of 2019

Dear Commissioner Mayuga:

We represent The New York City Hospitality Alliance, a not-for-profit trade association representing New York City's hospitality industry, including several thousand eating and drinking establishments across the five boroughs.

We submit these comments regarding the penalty schedule provision of the proposed rules implementing Local Law 202 of 2019, which our members commonly refer to as the foie gras ban.

The proposed penalty schedule is unjustifiably more punitive than the penalties specified in Local Law 202. The law enacted by the City Council provides as follows:

§ 17-1903 Enforcement. Any person who is found to violate any provision of this chapter shall be subject to a civil penalty of *not less than \$500* and not more than \$2,000 for each violation. Each such violation *may* be treated as a separate and distinct offense, and in the case of a continuing violation, each day's continuance thereof *may* be treated as a separate and distinct offense.

See LL 202 of 2019 (emphasis supplied).

1. The penalty should start at \$500, not \$1,500

Even though the City Council directed the civil penalty to start at \$500, the proposed penalty schedule assesses a \$1,500 penalty for a first violation. The Department offers no justification for why it wishes to fine small businesses \$1,000 more for a first violation than what the City Council directed.

Nor can we conceive of what an appropriate justification would be. If the proposed penalty schedule is enacted, it would amount to an act by Department staff to unilaterally overrule the judgment of the elected officials in the legislative branch as to what the appropriate minimum penalty should be.

2. Multiple first-time violations – and multiple days – should not be treated as separate and distinct offences

The proposed penalty schedule permits the Department to treat each violation – and each day of violation – all as separate and distinct offenses. Although the law gives the Department that option, it does not mandate it. Thus, as with the unjustifiably higher minimum penalty, the Department has made the decision to pursue more punitive rules than what the law requires.

3. The Department’s more punitive proposal quickly adds up

Although we expect the foie gras ban to impact an exceedingly small segment of our association’s membership, for those who are found in violation, the difference between the penalty required by the City Council and the first-time penalties proposed by the Department is significant.

A business found in possession of 10 foie gras livers for 10 days would be subject to a staggering \$150,000 civil penalty under the Department’s proposal, versus \$500 under the law enacted by the City Council. Even if the Department were to elect to treat each violation and day of violation as separate and distinct offenses, but still keep the \$500 minimum penalty enacted by the City Council, the penalty in this example would be \$50,000; which is \$100,000 less than the outcome under the Department’s proposal.

4. No historical precedent *requires* this outcome

We understand that there is a history of the Department making the same choices with other penalty schedules in connection with other statutes. In other words, we recognize that this is not the first time that the Department has proposed a higher penalty for a first violation than what the statute provides; or has elected to penalize multiple violations or multiple days of violation as separate and distinct offences when the statute does not require it.

The Department is not bound to repeat that history. Nothing in Local Law 202 of 2019 requires it. There is a new Mayor and a new Commissioner, and a new commitment within the administration to find ways to *reduce* regulatory burdens and “gotcha” fines on small businesses. We submit that first time violators should be given the benefit of the doubt, by being subject to the absolute minimum penalty that the statute requires, in this context and in all the penalty statutes that the Department enforces.

Thank you for your consideration. We look forward to working with you on a range of issues that are essential to New York’s critical bar and restaurant industry.

Very truly yours,

PESETSKY & BOOKMAN, P.C.



By: Max Bookman, Esq.

June 2, 2022

Commissioner Vilda Vera Mayuga
New York City Department of Consumer and Worker Protection
42 Broadway, 9th Floor
New York, New York 10004
RuleComments@dca.nyc.gov

Re: Comments to Proposed Rules for Local Law 144 of 2021

Commissioner Vilda Vera Mayuga:

Beckage is one of the leading privacy and data security legal practices as recognized by numerous rankings and awards, including from LAW360, CyberSecurity Docket and Net Diligence. Beckage is general counsel to the Data Privacy Alliance (“DPA”), a non-partisan coalition of for-profit and not-for-profit organizations. The DPA’s objective is to promote data privacy and security by increasing awareness and contributing to informed and appropriate legislative and regulatory standards.

In the course of counseling the DPA and its members on New York City’s Local Law 144 of 2021 (the “Law”) regarding the use of automated employment decision tools, we have identified various questions, ambiguities and issues that could be addressed through the regulatory authority of the Department of Consumer and Worker Protection (the “Department”). We have outlined some of those comments below, which are organized based on the sections of the Law.

These comments do not necessarily reflect the opinions or concerns of the members of the DPA. These comments also do not reflect a position statement by the firm.

I. Comments

Automated Employment Decision Tool

The Law’s definition of an “Automated Employment Decision Tool” is ambiguous, and the Department should issue regulations to clarify the ambiguity.

- As an example, the definition includes a tool “that issues simplified output, including a score, [or] classification” which arguably includes credit reports that contain credit scores, and criminal background reports that contain classifications, which are already regulated by the Fair Credit Reporting Act. The Department should clarify whether the Law is intended to apply to such reports and any similar background reports.
- As an additional example, the definition includes “any computational process derived from [...] data analytics.” It is unclear whether this definition would apply when an employer sorts and/or organizes a list of candidates based on certain data, such as their years of experience, highest level of education, or city of residence. The Department should clarify whether the Law is intended to apply to tools such as spreadsheets or databases that sort and organize candidates based on certain data.

- The ambiguity described above regarding spreadsheets and databases that sort data is not resolved by second sentence of the definition of automated employment decision tools. While the second sentence lists examples of tools that are not included in the definition, that list is preceded by a generic statement of the exemption that is ambiguous. Specifically, that statement employs a double negative to state that the definition “does not include a tool that does not automate, support, substantially assist or replace discretionary decision-making processes.” However, the conditions here are broad enough that the exemption is arguably nonexistent. As an example, for the exemption to apply, the tool must not “support” or “substantially assist” the decision-making process, but tools like spreadsheets and databases that sort data provide support and substantial assistance to decision makers. The Department should clarify the ambiguity created by this exemption.

Employment Decision

The Law defines an “employment decision” as “screen[ing] candidates.” This definition is ambiguous, for example “screening” candidates is arguably broader than making a “decision.” As noted above, if a tool merely sorts candidates based on certain criteria such as years of experience, highest level of education, or city of residence, does that constitute a “screen” that meets the definition of an “employment decision?” The Department should clarify the ambiguity created by this definition.

Bias Audit & Summary of Results

The Law purports to require a bias audit to evaluate disparate impact but does not establish any criteria or measures for doing so. Indeed, the Law does not say anything about the acceptable methodology or results of the audit, and instead simply requires disclosure of a summary of the results. As an example, the Law does not explain whether a bias audit that uses the four-fifths rule for measuring disparate impact as described in 29 CFR § 1607.4(D) or some other standard is sufficient. The Department should clarify the criteria that a bias audit should assess. The Department should also clarify what information should be included in the required “summary” of the audit results. The Law uses the passive voice to state that an employer cannot use an automated employment decision tool unless the tool “has been the subject of a bias audit.” The Law does not make clear who needs to procure the audit. For example, if multiple employers use the same automated employment decision tool provided by a third-party vendor or agency, and one employer has already procured a bias audit of the same tool, can the other employers rely on that bias audit for that tool without needing to produce their own? Further, if the vendor of the tool has procured its own independent bias audit of the tool, can employers using the tool rely on the vendor’s audit to comply with the Law? The Department should clarify this point.

The Law requires that the bias audit be “an impartial evaluation by an independent auditor.” The term independent is ambiguous. As an example, is an audit sufficiently independent if it is conducted by a separate department or business unit of the employer that is making the employment decision, such as a research and development department? The term impartial is also ambiguous. As an example, if the vendor of an automated employment decision tool retains an auditor to conduct a bias audit for the tool, is that audit sufficiently impartial? Or must the audit need to occur at the employer level. The Department should issue rules to clarify the ambiguity of these terms.

For an automated employment decision tool that is in continuous use, the Law suggests that the bias audit must be repeated on at least an annual basis but does not provide any clarity on this particular point. If annual audits are required, can the employer use the same auditor to conduct each annual audit? The Department should issue rules to clarify the requirements for recurring audits.

Notice

The Department should issue rules clarifying the application of the Law’s requirement to provide 10-business days’ advance notice before using an automated employment decision tool. The following scenarios demonstrate situations in which application of the law is ambiguous:

- An employer posts an available position, and a candidate submits an application for the position. After the candidate has already submitted his application, and after the employer has received an unexpectedly large number of applicants, the employer decides to use an automated employment decision tool. The employer then provides notice to the prior applicants that it intends to use an automated employment decision tool 10 business days prior to doing so, along with the other requirements of the Law. Is this notice sufficient even though it was not provided prior to the candidate submitting his initial application?
- An employer posts a notice on its website regarding an available position and includes in the announcement the required notice regarding the use of an automated employment decision tool. The announcement states that the window to apply will remain open for 20 business days, following which the tool will be used. A candidate first sees the notice on the last day of the application window and submits an application. Is the notice in the initial announcement sufficient to satisfy the 10-business day advance notice relative to the candidate applied on the last day? Or does the Law force employers to wait at least two weeks (10 business days) until the application window has closed before it can use the automated employment decision tool?

The Department should issue regulations clarifying the application of the 10-business days advance notice requirement.

Geographic Applicability

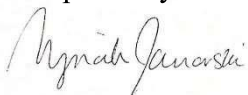
The Department should clarify whether the Law is intended to apply to non-NYC resident applicants for NYC positions.

For example, the bias audit requirement purports to require an audit where the position is located within the City, regardless of the residency of the applicant. However, the notice requirement seems to purport to apply to New York City applicants only.

II. Conclusion

We appreciate the opportunity to make comments for your consideration and look forward to participating in the formal rule-making process. If you have any questions regarding these comments, please feel free to contact us.

Respectfully submitted,



Myriah Jaworski, Esq.



Brian Myers, Esq.



From: **Shea Brown, Ph.D.** Chief Executive Officer BABL AI Inc.
sheabrown@bablai.com

To: **NYC Department of Consumer and Worker Protection** Rulecomments@dca.nyc.gov

Re: **Local Law 144 of 2021**

June 3, 2022

To Whom It May Concern:

On behalf of the team at BABL AI, I welcome the opportunity to provide public comments on Local Law 144 of 2021¹, which requires yearly “bias audits” of automated employment decision tools (AEDTs) and mandatory notifications to employees and candidates subject to such AEDTs. This law is a welcome attempt to mitigate potential harm that such systems could cause, while strengthening the market position of vendors that invest in thorough due diligence to tackle these issues head on. As a company that audits algorithms for ethical risk, effective governance, bias, and disparate impact, BABL AI believes that the spirit of this law furthers our mission to promote and protect human flourishing in the age of AI.

However, as the Department now is considering new rules that would add penalty schedules for violations of this law, we encourage the Department to clarify several ambiguities in Local Law 144 that pose barriers for companies wishing to make good-faith efforts to comply. Below we outline some ambiguous concepts in the law, as well as questions that we feel need to be answered in order to understand how to navigate effective compliance.

Independent Auditor: The new law states that a bias audit “means an impartial evaluation by an independent auditor.” In financial auditing, the notion of independence has been codified by the Sarbanes–Oxley Act of 2002 (“SOX”).² But such clarity does not exist for algorithm auditing. This lack of clarity will lead to uncertainty in the audit results, and confusion for companies trying to decide where to allocate time and resources for compliance. In particular, clarity on two topics would immediately provide substantial benefits to companies seeking to comply with the new law:

- **Internal vs. external** – Would an internal audit function be considered sufficiently independent?

¹ <http://nyc.legistar1.com/nyc/attachments/c5b7616e-2b3d-41e0-a723-cc25bca3c653.pdf>

² <https://www.govinfo.gov/content/pkg/PLAW-107publ204/pdf/PLAW-107publ204.pdf> ³ See “Taxonomy: AI Audit, Assurance and Assessment” for a detailed discussion:

<https://forhumanity.center/bok/taxonomy-ai-audit-assurance-assessment/>

- **Conflict of interest** – Should there be restrictions on non-audit remunerations as in SOX?³ Or, can a single firm provide both advisory and bias audit services to the same client?

Testing for Disparate Impact: The new law states that the bias audit must “assess the [AEDTs] disparate impact” on certain persons. In a field as new as algorithm auditing, especially in the dynamic space of AEDTs, the notion of testing for disparate impact can mean many things depending on the use-case, the data available, and who is doing the testing. Below are some of the key uncertainties that we see when trying to engage with clients on these issues:

- **Vendor vs. employer** – if the vendor conducts an audit, is it sufficient for employers to reference the vendor’s public “summary of the results” of the bias audit to satisfy their obligations under the new law? Since many vendors offer customization for large clients or when integrating into large platforms, the potential for disparate impact may lay outside the vendor’s part of the product chain. Given this, we urge the Department to consider the practical and ethical responsibilities employers and platform owners have when relying on a vendor’s public summary of audit results.
- **Access to demographic data** – in many cases, the availability of demographic data from employees and candidates that interact with these systems is extremely limited. This is also true of the data used to train, validate, and test machine learning AEDTs. The reasons for this are varied, including a lack of resources and privacy, bias, and regulatory concerns.³ In these cases, we encourage the Department to permit employers to use proxy variables and imputation of demographic data, as long as the statistical limitations of these methods are rigorously quantified and justified.
- **Actual vs. expected disparate impact** – in some cases, demographic data for the production AEDT is unavailable or woefully incomplete. In these cases, we

recommend that the Department permit the use of offline testing data (either historical, synthetic, or with variables that are proxies for protected categories) to serve as sufficient for testing the expected disparate impact.

- **Direct testing vs. independent verification of testing** – in many cases, employers or vendors have conducted disparate impact testing and monitoring of their AEDTs. Can an independent auditor verify that that testing has been done and meets a certain publicly available standard?⁴ Or, does the auditor need to directly access the system and run the disparate impact tests?

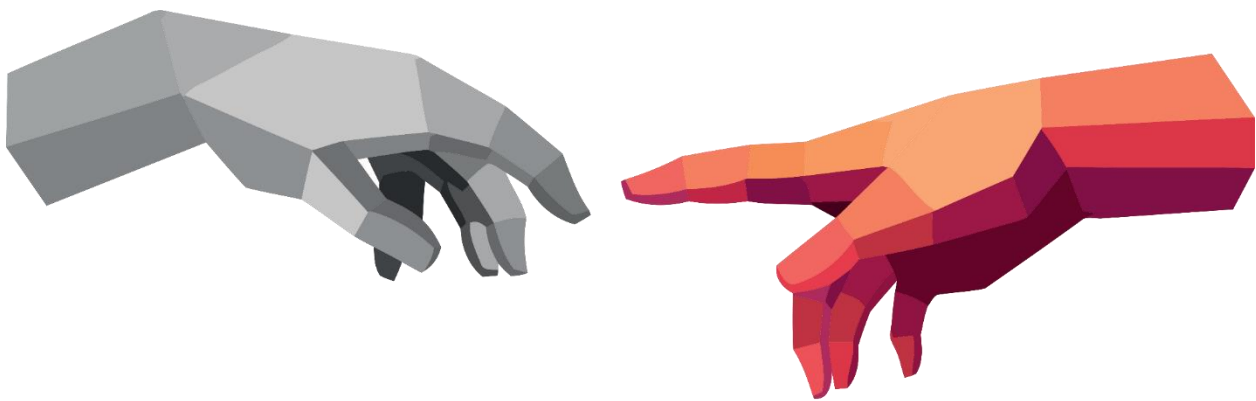
I’d like to thank the Department for providing us the opportunity to comment on Local Law 144, and we would be happy to provide further clarification on any of the above questions.

Contact

Shea Brown, Ph.D., CEO & Founder sheabrown@bablai.com

³ <https://partnershiponai.org/workstream/demographic-data/>

⁴ e.g., non-profits like [ForHumanity](#) have binary criteria-based audits for these purposes.



FORHUMANITY

ForHumanity⁵

⁵ [ForHumanity \(https://forhumanity.center/\)](https://forhumanity.center/) is a 501(c)(3) nonprofit organization dedicated to addressing the Ethics, Bias, Privacy, Trust, and Cybersecurity in artificial intelligence and autonomous systems. ForHumanity uses an open and

Government and Regulatory Services
for the governance, accountability and oversight of Artificial Intelligence, Algorithmic
and Autonomous Systems
(*Pre-Production Draft*)

Ryan Carrier
Executive Director
980 Broadway #506
Thornwood, NY 10594

Re: NYC AEDT Bias Audit law - Local Law 144 2021 Dear Chair and

Department Members:

It is ForHumanity's pleasure to submit this letter and our Government and Regulatory services in regards to Local law 144 of 2021 related to Automated Employment Decision tools (AEDT). The protection afforded by the law to candidates of AEDT are aligned with ForHumanity's mission to *examine and analyze downside risk associated with the ubiquitous advance of AI, algorithmic and autonomous systems and where possible to engage in risk mitigation to maximize the benefits of these systems... ForHumanity.*

ForHumanity (<https://forhumanity.center/>) is a 501(c)(3) nonprofit organization dedicated to addressing the Ethics, Bias, Privacy, Trust, and Cybersecurity in artificial intelligence and autonomous systems. ForHumanity uses an open and transparent process that draws from a pool of over 1000+ international contributors, from more than 70 countries to construct audit criteria, certification schemes, and educational programs for legal and compliance professionals, educators, auditors, developers, and legislators to mitigate bias, enhance ethics, protect privacy, build trust, improve cybersecurity, and drive accountability and transparency in AI and autonomous systems. ForHumanity works to make AI safe for all people and makes itself available to support government agencies and instrumentalities to manage risk associated with AI and autonomous systems.

In support of the NYC AEDT Bias Audit law, ForHumanity has regularly convened a team of volunteers (all humans are welcome in our transparent, crowdsourced process) to draft ForHumanity's NYC AEDT Bias Audit - a certification scheme aimed to satisfy Local Law 144's term -"bias audit" . It is our belief that a "bias audit" is not a widely understood and accepted term, whereby all auditors know all steps that are required to satisfy such an audit.

[transparent process that draws from a pool of over 900+ international contributors to construct audit criteria, certification schemes, and educational programs for legal and compliance professionals, educators, auditors, developers, and legislators to mitigate bias, enhance ethics, protect privacy, build trust, improve cybersecurity, and drive accountability and transparency in AI and autonomous systems. ForHumanity works to make AI safe for all people and makes itself available to support government agencies and instrumentalities to manage risk associated with AI and autonomous systems.](#)

In our conversations with auditors, AEDT providers, plaintiff-side attorneys and employers, great ambiguity remains on how audit satisfaction will be achieved. In light of this ambiguity most compliance will error on the side of minimum compliance. The ambiguity exists as a result of the law’s language copied here, “*Such bias audit shall include but not be limited to the testing of an automated employment decision tool to assess the tool’s disparate impact on persons of any component 1 category required to be reported by employers pursuant to subsection (c) of section 2000e-8 of title 42 of the United States code as specified in part 1602.7 of title 29 of the code of federal regulations*”. The “*but not limited to*” clause rightly highlights that bias is not only about disparate impact. In fact, bias exists in many forms, such as statistical bias, cognitive bias and non-response bias. Further bias manifests in data, architectural inputs and outcomes from AI, Algorithmic and Autonomous systems (AAA Systems), like AEDTs. ForHumanity agrees with the Council that we ought to maximize bias mitigation (“*but not limited to*”) in AEDTs and our audit criteria already is designed to mitigate a wider array of bias.

The law also did not appear to fully embrace all Protected Categories (the subjects of bias), such as the Disabled. AAA Systems by their very design (seeking “best-fit” conclusions) are often exclusionary, especially in the areas of Disability and neuro-divergence. ForHumanity’s audit criteria can help the Council include bias remediation in AEDTs for all New Yorkers

ForHumanity offers to assist the Council in overcoming these challenges with our expertise in drafting audit criteria and our focus on mitigating risk for AAA Systems for all humans. We offer this service for the Council’s consideration as a means to establishing uniformity, certainty and an infrastructure of trust for “bias audits” of AEDTs. This offer is not unique for ForHumanity. We have provided the UK’s Information Commissioner’s Office with a similar submission of audit criteria for the General Data Protection Regulations (GDPR) and we have been retained by CEN/CENELEC JTC 21 as a technical liaison on the conformity assessment called for in the EU’s Proposed AI Act. ForHumanity is conducting this work in numerous other jurisdictions globally as law-makers race to place guardrails around these largely ungoverned AAA Systems.

Financial audits have a series of critical elements of infrastructure, including checks and balances leading to successful governance, oversight and accountability. Those key elements are discussed in the attached document laying out a comprehensive framework establishing an infrastructure of trust and are summarized here:

- 1) Trained bias audit professionals - like CPAs
- 2) Independent third-party rules (Like Generally Accepted Accounting Principles - GAAP) - accepted and approved by the Council
- 3) A body to ensure independence, anti-collusion and uniformity of audits prevail.
- 4) A code of Ethics and Professional Conduct governing auditors and their actions

This set of criteria would dramatically enhance the impact and compliance with the law, providing a leveraged enforcement mechanism of trained auditors abiding by a set of rules the council has approved. ForHumanity provides the services, under the authority of the council for all four elements at no cost to the Council or New York City. As a non-profit, public charity, 501(c)(3) registered, the Council can be assured that our goals are aligned protecting New Yorkers from bias in Automated Employment Decision Tools.

Thank you to the Council and the City of New York for the opportunity to testify on behalf of all New Yorker's who are the beneficiaries of ForHumanity's work and mission. We hope you will consider our assistance and would welcome any opportunity to further share our work in support of Local Law 144 2021.

Kind regards,

Ryan Carrier

Executive Director, ForHumanity

Table of Contents

- [ForHumanity’s Mission Statement What is ForHumanity \(FH\)?](#)
- [FH Government and Regulatory Services](#)
- [Independent Audit of AI Systems](#)
- [Assuring Trust - IAAIS](#)
 - [Independence](#)
- [Ecosystem Explained - Roles and Responsibilities](#)
 - [Taxonomy: AI Audit, Assurance and Assessment](#)
- [ForHumanity’s Perspective](#)
- [Audit Criteria - Government Submissions](#)
- [Determination of Audit Criteria Development](#)
- [Global Harmonization](#)
 - [Relevant Legal Frameworks](#)
 - [Jurisdictional Sensitivity Criteria Mapping](#)
- [Ethics-by-Design](#)
- [Privacy-by-Design](#)
- [Bias Mitigation](#)
- [Trust](#)
 - [Transparency/Disclosure](#)
 - [Accessibility](#)
 - [Control/Safety](#)
 - [Explainability Inclusion](#)
- [Cybersecurity](#)
- [Diverse Inputs and Multi Stakeholder Feedback](#)
- [Special Committee Structure](#)
- [Biometric Data](#)
- [FH AI Risk Management Framework](#)
 - [Concept](#)
 - [Overall Framework](#)
 - [FH Risk Foundations](#)
 - [FH foundational reading on risk management \(Principles\)](#)
 - [FH Risk Guiding Documents](#)
 - [Risk Taxonomy](#)
 - [Risk Management Policy](#)

[Maximizing Risk Mitigation for Humans](#)

[Diverse Inputs & Multi-stakeholder feedback](#)

[Need for committees](#)

[FH Risk Management Process](#)

[Guidance on operationalizing risk categories from human impact perspective](#)

[Guidance on operationalizing risk categories from human impact perspective](#)

[Guidance on determining Risk Tolerance and Risk Appetite](#)

[Functional Oversight](#)

[Functional Risk Management](#)

[Responsibilities of Committees](#)

[Role of product, business and other stakeholders in risk management in AI lifecycle](#)

[Functional Risk Management reports](#)

[Residual Risk Management Internal Reviews](#)

[cAIRE Reporting](#)

[Understanding cAIRE report Risk and Control Scope](#)

[templatecAIRE residual risk log -](#)

[Threat and Risk \(Emergent & Horizon scanning\) & Systemic Societal](#)

[Feeding into Operational Risk Management at an Organizational Level](#)

[Enterprise Risk Management - Guidance for integrating AI risk with ERM COSO ERM: AI Risk Management integration](#)

[ForHumanity University](#)

[Trained and Accredited Workforce](#)

[Compliance-by-Design](#)

[Board of Directors Audit](#)

[Evaluation Methods](#)

[Data Taxonomy and Technique Documentation of AAA Systems Process Flow](#)

[Body of Knowledge - Knowledge Stores](#)

[Certifications](#)

[Certification Merits](#)

[Limitations of Certification](#)

[Engagement with a Certification Body](#)

[Auditor - Auditee agreement on Scope](#)

[Certification Warning/Certification At-risk](#)

[Withdrawal of Certification](#)

[Certification mark use standards and guidelines](#)

[Certification Steps](#)

[Define Scope](#)

[Target of Evaluation Determination Process](#)

[Conduct Pre-assessment/ Pre-audit](#)

[Identify Certification Body](#)

[Identify Auditors for cCertification](#)

[Independence Enforced via License](#)

[Anti-Collusion](#)

[Certification Issuance](#)

[ForHumanity and Accreditation Service Examinations](#)

[Audit Period of Validity](#)

[Recertification](#)

ForHumanity's Mission Statement

ForHumanity is a US 501(c)(3) tax-exempt public charity and our mission is to *examine and analyze downside risk associated with the ubiquitous advance of AI, algorithmic and autonomous systems and where possible to engage in risk mitigation to maximize the benefits of these systems... ForHumanity*

What is ForHumanity (FH)?

ForHumanity is:

1. Mission-driven, non-profit, public charity
2. Consisting of 900+ members from more than 70 countries
3. Only natural persons are permitted to join ForHumanity
4. We accept no corporate funding
5. All works performed have been conducted on an all-volunteer basis
6. All work-projects are executed transparently via crowdsourcing
7. All decisions and governance within ForHumanity are mission-aligned and executed by the Executive Director, or the Board of Directors
8. Majority of Board of Directors are elected from the community of ForHumanity Fellows and by the ForHumanity Fellows

FH Government and Regulatory Services

We offer to facilitate critical portions of the ecosystem under the authority and in cooperation with governments.

ForHumanity has developed a comprehensive ecosystem - an infrastructure of trust for Artificial Intelligence, Algorithmic and Autonomous Systems (AAA Systems) - modeled on the ecosystem of financial accounting and reporting called Independent Audit of AI Systems (IAAIS). All elements of the ecosystem adhere to common accepted practices by many governments and are meant to be accepted, adopted and integrated by government approval.

ForHumanity provides a unique toolkit for the benefit of legislators and regulators offering unprecedented secretariat services:

1. We draft audit criteria to used by third-party, independent auditors on for AAA Systems

2. We adapt law, guidelines, regulations, standards and best-practices in binary (compliant/non-compliant) criteria **submitted for approval by governments and regulators**
3. We educate and train individuals on the audit criteria - accrediting them upon examination as ForHumanity Certified Auditors (FHCAs)
4. We uphold a [Code of Ethics and Professional Conduct](#) for FHCAs
5. We maintain an open forum for crowd-sourced, transparent, all-inclusive input on the audit criteria - available to all natural persons without meaningful barriers-to-entry to volunteer contribution
6. We operate a licensing system for approved audit criteria that ensures:
 - a. Independence (see below for further definition)
 - b. Fair and level playing field
 - c. Anti-Collusion principles
 - d. Uniformity of certification practices and compliance
 - e. Post Audit Compliance Reports
 - f. Verification/Trust services (blockchain verifiable)
 - i. Verified practitioners
 - ii. Verified credentials
 - iii. Verified compliance and certifications
7. We provide oversight for certifying bodies (in the absence of a national accreditation service) and for certified individuals, including reviews of past certifications for quality control
8. We advocate for mandatory third-party independent audits on all AAA Systems that impact humans that are not excluded on the basis of low-risk of negative impacts to natural persons
9. ForHumanity provides governments and regulators access to a set of harmonized global best practices tailored and specified to the laws and regulations of your jurisdiction.

Independent Audit of AI Systems

Independent Audit of AI Systems is an all-inclusive term to describe the entire ecosystem of governance, oversight, accountability and trust for AAA Systems. It is highlighted by the following characteristics:

1. IAAIS is applied across the entire lifecycle of the AAA systems including design, development, deployment and decommissioning
2. IAAIS captures and mitigates risk to natural persons across five pillars (ethics, bias, privacy, trust and cybersecurity)
3. IAAIS is designed to identify and mitigate the unique and specialized risks occurring from the very nature of socio-technical systems (e.g. Data Entry Point Attacks⁶ and embedded instances of Ethical Choice)
4. IAAIS integrates with a comprehensive [risk management framework](#) that operates with four lines of defense for risk mitigation:
 - a. Designers, Developers, Product managers and Data Scientists
 - b. Managers, Overseers, Human-in-Command, Committees

⁶ Capitalized terms reflect a ForHumanity defined term in our audit criteria and can be found here <https://forhumanity.center/definitions/>

- c. Internal Audit, Risk Reviews
 - d. External, Independent Auditors
5. IAAIS mitigates risk to a wider collection of stakeholders beyond ISO 31000's stakeholder list inclusive of not only organizational risk, but also risks to natural persons, communities and the environment
 6. IAAIS establishes binary (compliant/non-compliant rules) and specific documentary evidence required for sufficient proof of compliance

Assuring Trust - IAAIS

Trust is assured when three characteristics come together:

- 1) Independent, third-party, objective and widely accepted criteria are universally applied
- 2) Assurance is executed by accredited, well-trained, independent experts with verifiable credentials
- 3) Independent, accredited certification bodies, acting on behalf of society and not on behalf of the auditee, assure compliance with government approved rules

Independent, external auditors provide a service to the public and the constituency of the government by assuring compliance with the rules and regulations put forward in audit criteria. Certification is a strong signal, annually verified, that compliance with the law is being maintained. The process, like with financial reporting and audit, generates a compliance-by-design approach, whereby best practices are built into the beginning of AAA system development.

Independent, external auditors provide objective certification founded upon their Code of Ethics and Professional Conduct. Additionally, their obligation to receive no other remuneration from auditees, coupled with their risk of false assurance of compliance solidifies their objectivity. The marketplace is assured that an infrastructure of trust is present and may be relied upon by organizations and individuals that need assurance.

No transparent system is foolproof. When the rules are transparent, then companies or persons seeking to commit fraud and malfeasance may succeed. However, a robust ecosystem of transparency and disclosure will eventually identify their misdeeds. Even some of the greatest financial frauds (e.g. Enron and WorldCom) were eventually caught by the system itself as a result of transparency and disclosure. In most cases, IAAIS provides a reliable infrastructure of trust for the willfully compliant.

Independence

A legal term defined by [The Sarbanes-Oxley Act of 2001](#) that requires a certifying body (an Auditor) to receive no other remuneration from an Auditee beyond reasonable audit fees. In its license agreements, ForHumanity further stipulates that a licensee holder cannot be an Auditor and an Assessor/Consultant (or provide any other form of service) to the same Auditee within a 12-month period. ForHumanity has adopted this rule and determination into the licensing agreements for certification bodies and abided by FHCAs under their Code of Ethics and Professional Conduct.

Independence and independent audit increases compliance with established laws and regulations. Time and again, human nature has proven that self-assessment is useful but insufficient, thus requiring the need for further enforcement mechanisms. However, government and regulatory enforcement requires resources to examine societal compliance. Enforcement bodies can mandate uniform criteria that satisfies compliance (e.g. the Securities and Exchange Commission mandating adherence to Generally Accepted

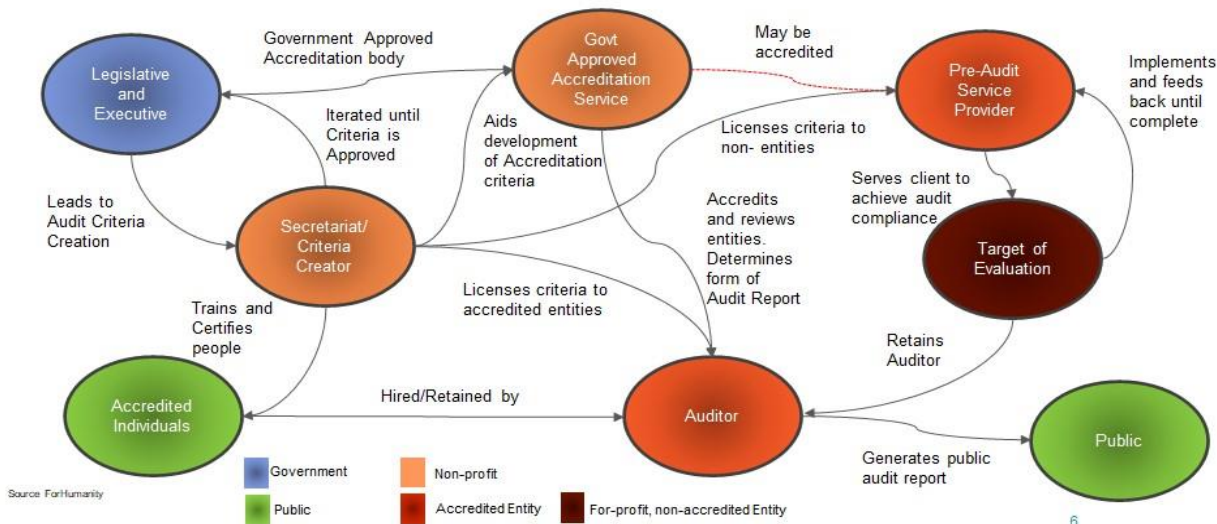
Accounting Principles GAAP for publicly traded companies in 1975). Then, Independent Audit when mandated by governmental enforcement agencies creates a leveraged, overarching compliance mechanism - examining and assuring compliance - accomplished by third-party trained practitioners, accredited robustly (and equally overseen - “watching the watchers”), using uniform rules, regularly assure compliance, at their own risk of false assurance of compliance. Under this ecosystem, conflicts are mitigated, objectivity is maximized, and trust is built.

More details on specific examples of Independence can be found in [ForHumanity’s Certified Auditor Code of Ethics and Professional Conduct v1.0](#).

Ecosystem Explained - Roles and Responsibilities

Discussed in detail in ForHumanity’s [Infrastructure of Trust for AI - Guide to Entity Roles and Responsibilities](#), the image below depicts the interactions and segregation of duties in Independent Audit of AI Systems. This ecosystem mirrors that of financial audit and reporting. The graphic includes the role of government, government-approved accreditation bodies (if appropriate), accredited entities, auditors, pre-audit service providers, auditees and the role of individuals and the public in general.

Roles and Responsibilities – for AI Audit infrastructure



This ecosystem takes a government enforcement role and transfers compliance oversight to the marketplace. Compliance oversight is conducted by accredited entities employing accredited individuals trained in audit compliance and certification. Both ForHumanity and the Accreditation Body “watch the watchers” and have authority to revoke accreditation or licensing rights for insufficient governance, accountability, and oversight. More details are available in the guide.

Taxonomy: AI Audit, Assurance and Assessment

Described and delineated in ForHumanity’s Guidance: [Taxonomy: AI Audit, Assurance & Assessment](#). ForHumanity describes differentiated roles amongst third-party services providers. This is critical, because it is the

unique characteristics of the auditor that establishes an infrastructure of trust by providing services that are truly independent and conflict-free to provide the public with confidence that compliance has been assured.

	Internal Audit	3rd-Party, Independent Audit	Assurance	Assessment	Consulting
Certified Practitioners Required?	No, Employees	Yes	Yes	No	No
Objective/Subjective	Objective	Objective	Objective	Subjective	Subjective
Independent	Yes	Yes	Yes	No	No
Known 3rd Party transparent Binary, Rules or Laws	Yes	Yes	No	No	No
Known 3rd Party transparent non-binary, Standards, Frameworks or Guidelines	No	No	Yes	No	No
Service provided for?	Management	Users, Society	Users, Society	Contracting Party	Contracting Party
Feedback Loop with the Company, iterative problem solving, teaching, tailoring	No	No	No	Yes	Yes
Consequences for False Compliance assertions	job loss	Liability	Liability	No liability	No liability
Written Report produced for the Public	No	Yes	Yes	No	No

Sources: ForHumanity, COSO, IAASB, Sarbanes-Oxley, IFAC

ForHumanity’s Perspective

ForHumanity drafts audit criteria from a specific perspective - what is best for humans. This human-centric, mission-driven focus is intentional to counterbalance the corporate-permissive focus prevailing across most western societies. AI, algorithmic and autonomous systems are socio-technical, meaning the human is “in” the system through the use of Personal Data and is the target of the outcomes of the system as well. This integration necessitates a 360-degree perspective of risk, beyond the corporate and including people, communities and the environment.

When drafting audit criteria, the guidance is simple - *“does this criteria mitigate risk to humans from the AAA System?”* ForHumanity asserts that sustainable profitability for corporations will occur when risks to humans are mitigated.

Audit Criteria - Government Submissions

All humans may contribute to the ForHumanity audit drafting process. It is transparent to all who agree to abide by the community’s [Code of Conduct](#) to assure decorum. Upon entry into the community, all contributors have full and complete access and transparency to ForHumanity work projects. ForHumanity celebrates Diverse Inputs and Multi

Stakeholder feedback in order to maximize risk assessment from a 360-degree perspective of impact. The same holds true for the creation of audit criteria. All may view, all may comment. The filter and adjudication on comments and their inclusion in final audit criteria drafts is simple - *“does a new word, definition or audit criteria mitigate risk to humans.?”* If so, it finds its way into our work.

Unless a certification scheme is deployed by organizations voluntarily, the government always has the final authority as to what audit criteria is approved.

All government approved audit criteria become available publicly under Creative Commons BY-ND-NC⁷. Licenses are offered to all qualified certification bodies and fees are due to ForHumanity upon receipt of revenue. As a non-profit, public charity, ForHumanity is restricted on revenue generation and all revenues must be put towards the operating budget and the mission. Qualified certification bodies are those entities

⁷ <https://creativecommons.org/licenses/by-nc-nd/4.0/>

employing FHCAs on the licensed criteria they deploy by contract with clients. The criteria are available for all non-revenue generation applications, such as research and academic study, freely.

Determination of Audit Criteria Development

As jurisdictions enact laws governing AAA Systems, ForHumanity will maintain pace with the law and have audit criteria drafted for and submitted for approval to governments. Additionally, ForHumanity maintains an active engagement process with the market to determine demand for new certification schemes for voluntary adoption or as guidance for policymakers and judicial settlements. ForHumanity intends to produce audit criteria for every AI, algorithmic and autonomous system that impacts a human.

Global Harmonization

In the interest of humanity, ForHumanity drafts audit criteria in accordance with local jurisdiction laws. However, having contributors from more than 70 countries around the world provides diverse inputs and broad perspectives - a lens that no single country can emulate. This provides ForHumanity with a unique perspective on global best practices, ones we offer to each country to uncover solutions and risk mitigations that might not be apparent or readily available to a country. Furthermore, since many organizations are international, a set of criteria with maximum harmonization will minimize the cost of compliance resulting from compliance requirements in multiple jurisdictions.

Relevant Legal Frameworks

ForHumanity tailors each audit criteria to the Relevant Legal Frameworks applicable to the jurisdiction. Auditors shall refer to these local laws for determination of Protected Category Variables, human rights and freedoms afforded to Data Subjects or natural persons especially in the areas of equality and anti-discrimination law, access to goods and services, children's laws, sector, platform- or service-specific law.

Jurisdictional Sensitivity

Under Independent Audit of AI Systems, nation-states retain their authority; in fact have their enforcement capabilities enhanced. Audit criteria are jurisdictionally sensitive, drawing upon local law and regulations to specify such details as, for example, Protected Categories. By focusing on local regulations, the audit avoids "legislating" compliance but instead leaves these governance questions in the hands of elected officials. Furthermore, the audit will occasionally fall back on the legal concept of "reasonable" relying on either past jurisprudence or current examples of possible solutions without being prescriptive.

Under IAAIS, proactive compliance can be achieved through the certification process - an evidentiary based proof-statement, independently verified by an objective, third-party auditor working for the public good.

An example of this jurisdictional sensitivity can be found in Protected Category Variables. Bias, in itself, is a statistical term describing a characteristic of a data set. However, when society then dictates that certain activities shall not be biased in their execution, it becomes something we need to account for in our systems. In the case of Protected

Category Variables, each jurisdiction may be different. In Scotland, for example, socioeconomic status is a Protected Characteristic, but that is not true of law in the United States.

Each jurisdiction's laws will be considered in the adaptation of the audit rules.

Criteria Mapping

ForHumanity draft criteria takes time to conduct “map” a specific service related to audit criteria that delineates the difference between two sets of audit criteria at a micro-level:

1. Definition to Definition
2. Legal Term to Legal Term (e.g. meaning of Consent)
3. Authority/Regulator to Authority/Regulator
4. Gaps from one jurisdiction to another

ForHumanity publishes official mapping documents to enable gap analysis services using official ForHumanity criteria.

Ethics-by-Design

The nature of socio-technical systems is to embed the human in the system through the use of Personal Data, while producing outcomes that have impacts on the human. The result of this interaction creates systems with a specific shared moral framework: that of the organization and/or the designers and developers. The ethics of the systems are meaningful to the human and will have significant impact on outcomes.

This interaction of corporate ethics and new/changing law governing AAA Systems often requires ethical/soft law considerations and user's ethics. This intersection requires trained experts to adjudicate the myriad instances of ethical choices embedded in AAA Systems such as:

1. Necessity
2. Proportionality
3. Adjudication of soft law
4. Statistical benchmarks for representativeness
5. KPI design for concept drift
6. Interface design choices
7. Explainability

Responsibility for managing the process ranging from the creation of a public Code of Ethics to implementing controls around instances of Ethical Choice is a standing, trained and empowered Ethics Committee, presided over by experts in algorithm ethics and applied ethics.

However, “ethics washing” and superficial applications of ethics remain a risk without governance, oversight and accountability on instances of ethical choice within organizations. IAAIS criteria require the Ethics Committee to examine and consider all instances of Ethical Choice to be documented and attested by an independent, external auditor to ensure objectivity, oversight and accountability over the design, development, deployment and potential decommissioning of AAA Systems. The implementation of ethics-by-design is discussed in ForHumanity's paper on the [Rise of the Ethics Committee](#).

Privacy-by-Design

Of ForHumanity's five pillars for AAA Systems (ethics, privacy, bias, trust and cybersecurity), privacy is the most well developed because of legal efforts on data privacy and protection, notably advanced by the General Data Protection Regulation (GDPR). ForHumanity has developed a comprehensive set of certification criteria designed to be used by certification bodies to assure compliance with the GDPR (both EU and UK versions).

The criteria provide assurance around:

1. General Governance, Accountability and Oversight
2. Necessity
3. Proportionality
4. Lawful Basis (including Consent)
5. Specified Scope/Nature/Context/Purpose
6. Data Minimization
7. Data Protection
8. Technical and Organizational Controls
9. Security
10. Cybersecurity
11. Data Subject Rights and Freedoms
12. Fairness
13. Transparency and Notice
14. Automated Decision Making (Profiling)
15. Explainability
16. Governance of Data Transfers

Certification of privacy law and regulations provides proactive compliance and fosters compliance-by-design thinking to enable economies of scale and general efficiency that can be leveraged into all AAA Systems.

Bias Mitigation

AAA Systems using Personal Data examine historical data to make inferences about people. Eradicating bias is a statistical impossibility and thus ForHumanity's goal is to maximize bias mitigations. Also, fairness and equity are enforced in different ways through law around the world.

We have identified three stages in AAA Systems where bias can be examined and mitigation rendered:

- 1) Bias in data
- 2) Bias in architectural inputs
- 3) Bias in outcomes

Bias in Data has many potential manifestations as ForHumanity explored in [Bias Mitigation in Datasets](#):

1. Source data
2. Cleaning
3. Labeling
4. Anomaly and outlier treatment
5. Training and testing/validations splits

6. Representativeness
7. Cognitive bias
8. Non-Response Bias

ForHumanity requires documentary evidence of bias mitigations in audit criteria to tackle each of these forms of bias and fight against discrimination results from the data. Many of these mitigations must be documented in a Data Transparency Document and made public allowing for a higher form of discourse, regarding the appropriate mitigations in data sets.

Data however is not the only source of bias, there are two more elements of a AAA System to examine for bias:

- 1) Architectural Inputs to models
- 2) Model Outcomes

As designers and developers construct their models, the choices they make and the methods they choose may result in bias. ForHumanity has drafted numerous audit criteria to examine and consider all appropriate mitigations to ensure fair and equitable treatment for people in decisions that designers and developers make across model architecture.

AAA Systems take on many forms with varied degrees of understanding on exactly how conclusions are reached. Some large language models will even have billions of parameters making it virtually impossible to replicate the decision process. For this reason, IAAIS examines outcomes against accepted and tested standards, like the American 4/5th rule⁸. To be certain, there is no magic to the arbitrarily-chosen 80% but at least there is a legislatively-accepted threshold that demands further analysis and justification. As these outcomes manifest as Residual Risk, they are disclosed to the natural persons prior to engagement with the AAA System. Under this disclosed, risk-based system, model designers and developers will encounter market-based feedback on the risk/reward profile of their model.

Trust

Trust is ForHumanity's catch-all category, focused on the many ways in which trust is earned, demonstrated, proved and sometimes forfeit. Below, we describe a series of sub-categories that all are captured under Trust.

Transparency/Disclosure

Often feared by corporations, transparency and public documentation (disclosure) are necessary elements of any trustworthy system. However, Intellectual property (IP) and trade secrets are protected under the infrastructure of trust that Independent Audit of AI Systems creates. IP review can be governed by Non-Disclosure Agreement with an Auditor in the rare occasion that it would be necessary for compliance. For example, the IAAIS approach to bias mitigation examines data, architectural inputs and outcomes of AAA Systems; there are no source code audits. The theory of this practice is simple - companies with problems in their models (e.g. bias, discrimination, unethical choices), are accountable for the inputs to the model as well as the outcomes. IP does not need to be examined or disclosed because the auditee is accountable for their outcomes, regardless of the root problem. In most instances, IAAIS is the best compromise between transparency and assurance of compliance for the public.

Typical transparency/disclosure under IAAIS are not intellectual property and trade secrets, instead are proof statements to the public of critical information to allow for users of the system to make informed decisions about their interactions with the AAA Systems. Transparency and disclosure notifications will describe decisions about

⁸ <https://www.ecfr.gov/current/title-29/subtitle-B/chapter-XIV/part-1607>

data around representativeness or confirmation of scope/nature/context/purpose of the AAA Systems including the description and usage of the Personal Data being collected under Consent.

When transparency/disclosure is deployed properly, the intent is to ensure that users are well informed about the scope/nature/context/purpose of the AAA System and the risks that are present for the natural purpose during the interaction with the system. This is a defining characteristic of trust: two parties agreeing to an interaction, knowing the responsibilities and expectations for each party. Transparency/disclosure mitigates an enormous amount of risk for both parties in the interaction.

Finally, transparency/disclosure is one of the last lines of defense. Even in record-setting frauds, like Enron and WorldCom, it was transparency and disclosure that finally allowed the system to catch on to the misdeeds. Under compliance-by-design infrastructure, transparency/disclosure requirements often become systemic productions. Therefore, when they are absent or malformed, their absence or anomalous compliance can become the bread crumbs leading to uncovering non-compliance. Transparency and disclosure are the cornerstones of compliance as they represent absolute accountability.

Accessibility

AAA Systems cannot be considered trustworthy if they cannot be accessed broadly. Systems that exclude groups of persons unfairly because of insufficient accessibility harm two kinds of users, those who do not have access and those who value AAA Systems that respect human dignity.

Accessibility of AAA Systems is rarely a problem of ability, but instead a problem of attention, knowledge and resource allocation. Many countries have equality and anti-discrimination laws requiring accessibility. Such laws designed to protect people in vulnerable situations proved necessary in order to provide added incentives (through legal enforcement) for organizations to ensure their services are available to all persons. AAA System accessibility is rooted in human dignity and organizations should either provide accessibility or the meaningful accommodations to meet the needs of all people.

Control/Safety

Especially in the realm of AAA Systems where physical and psychological harms are possibilities, control and safety are mission critical. For example, ensuring that a AAA System remains true to the intended scope/nature/context/purpose without concept drift is a requirement under GDPR and most informed consent lawful basis. Moreover, it would be unethical to operate a AAA system or any machine designed to benefit humanity without assurance that the system can be turned off and will remain off until the human providers of the system intentionally return the system to service. Such characteristics demonstrate control.

In regards to safety, ForHumanity advocates for a risk-based approach, similar to the US National Transportation Safety Board and other international standards, where systems are rigorously stress-tested in a broad range of conditions that challenge the system in all foreseeable environments ensuring sufficient reliability, robustness and resilience to avoid system failure, resulting in harm to humans.

Control and safety require a robust and comprehensive risk management process centered in a culture that embraces risk management across all three lines of defense (as defined by most traditional risk management frameworks) and includes Diverse Inputs and Multistakeholder Feedback to maximize risk treatment across the entire spectrum of risk inputs.

Explainability

Around the world and across numerous industries, especially when organizations wield power over natural persons, the law requires that corporate decisions are accompanied with explanations. These laws cover decisions rendered from AAA Systems and call for the outcomes to be explained in clear and plain language, however compliance with this rule of law is often a minimalist approach.

The theory behind such laws is based on human dignity. No one who receives a favorable decision is often concerned with an explanation, so this is clearly centered around persons receiving a negative result. ForHumanity's work ensures that automated decision-making explainability is accomplished commensurate with relevant legal framework requirements. However, ForHumanity recommends that AAA Systems go one step further to *Explainability+*.

The theory of *Explainability+* is simple. Being informed as to "why" an automated decision making system has produced a result, especially when that explanation is perfunctory does not help or empower the person to remedy their situation. *Explainability+* recommends the provider of the AAA System to go one extra step in support of the human and their humanity.

Explainability+ provides the natural person with the education required to achieve a favorable result from the AAA System: steps they might take to improve their situation and thus in a second iteration receive a more favorable outcome. Or, if a favorable outcome proves too challenging, other remediation services from within or outside of the organization designed to help the natural person achieve their desired goals. ForHumanity believes this will lead to great sustainable revenue for the organization, engender positive relations between the parties and celebrate human dignity with care for resolution and opportunity for satisfaction of the person's original goal. Interactions under ForHumanity's *Explainability+* create trust.

Inclusion

Most AAA Systems seek an average - a fitness across the data set - trained to explain all of the data. However, as a result of the very nature of most algorithmic models, fitness and inclusion are frequently at odds. Data scientists seeking greater accuracy often eliminate outliers and anomalies that by their very nature reduce model fitness. In socio-technical systems using Personal data, anomalies and outliers equate to people. The testing and evaluation of the algorithmic modeling process necessarily struggle with the balance between model accuracy and outlier inclusion. This tension can result in discrimination against Protected Categories and inclusion failures.

ForHumanity advocates for "edge-in thinking", a design concept that works to include edge cases, anomalies and outliers from the outset of the design and development process. From this starting point, modeling can consider appropriate and equitable accommodations for persons who would not otherwise be included in the process.

Finally, Diverse Inputs and Multi Stakeholder Feedback maximizes human inclusion in the risk input, analysis, evaluation phases of risk management. Please see the Diverse Input & Multi stakeholder Feedback section below for more details the advancement of inclusivity built into Independent Audit of AI Systems.

Cybersecurity

Cybersecurity is a fairly mature industry by comparison to artificial intelligence, algorithmic and autonomous systems, however, the socio-technical nature of AAA Systems creates new and unique vectors for cyber attacks. IAAIS incorporates existing gold-standard cybersecurity frameworks with tailored controls designed to address the new and unique attack vectors. Notably, data entry point attacks (e.g. data poisoning, model inversion and membership inference attacks) present innovative challenges that the marketplace is still grappling with. ForHumanity's audit criteria for AAA System cybersecurity is built upon the US NIST framework. Organizational solutions to AAA System cybersecurity should be tailored to specific systems and never shared publicly.

ForHumanity’s audit criteria represent a foundational governance, accountability and oversight framework for cybersecurity and helps to guide the minimum infrastructure to operate a successful cybersecurity system. However, the downside of transparency is that the rules, processes and criteria are available to bad actors as well, educating them on the methods and procedures that might lead them to find a weakness. Therefore ForHumanity strongly suggests, on behalf of the humans impacted by cyber breaches and data entry point attacks, that entities go above and beyond the foundational ForHumanity cybersecurity criteria. In regards to the specifics of a robust cybersecurity system, opacity is in the best interest of humanity.

Diverse Inputs and Multi Stakeholder Feedback

Diverse Inputs and Multi Stakeholder Feedback (DI&MSF) describes the 360-degree perspective of risk beyond ISO 31000 stakeholders (listed below):

- executive-level stakeholders
- appointment holders in the enterprise risk management group
- risk analysts and management officers
- line managers and project managers
- compliance and internal auditors
- independent practitioners

ForHumanity adds to the list of risk assessors:

- external domain experts
- natural persons (users and impacted)
- communities
- employee organizations (e.g. Unions)
- environmental representatives

Additionally, Independent Audit of AI Systems ensures diversity in the risk assessors by relying upon the Ethics Committee to determine the definition of diversity for the organization. ForHumanity recommends such a definition includes diversity of thought and lived experience in addition to the inclusion of Protected Categories and intersections thereof. DI&MSF risk assessors are trained in assessing risk in AAA Systems and provide valuable diversity in the risk input, analysis and evaluation process. For more detailed information, see ForHumanity’s paper on [Diverse Inputs and Multi Stakeholder Feedback](#) as well as the guidance in the Risk management Framework below.

Special Committee Structure

AAA Systems exist in a myriad of forms impacting numerous specific groups of people (e.g children, persons with disabilities or persons in vulnerable situations). ForHumanity recognizes the special needs of each of these groups and requires teams of trained experts to provide governance, accountability and oversight over such systems.

Each special committee expert requires deep and specific knowledge. For example, a member of the Children's Data Oversight Committee should have expertise in understanding and interpreting the UN’s Rights of the Child declarations including the ability to evaluate and adjudicate impacts and outcomes that affect their health, well-being, safety, avoidance of sexual abuse or exploitation, avoidance of economic exploitation, rights to privacy and exercising their own data privacy, supportive structures for their family relationships, elements that support their physical, psychological and emotional development, support of their right to develop their own identity, views and have their perspectives heard. These expertise are deep and specific and are necessary to ensure the Child is well represented in the design, development and deployment of AAA Systems.

Biometric Data

Biometric data is Personal Data and sometimes Special Category Data or Sensitive Personal Data, however, ForHumanity argues that it is more dangerous and sensitive than most types of data relating to humans:

1. **The immutable nature of many Biometric Data items** makes them more intrinsic to people's being than most other data, and so results in enormous risk to people from breach, theft, misuse and misappropriation
2. **The richness of information extractable from biometric data** makes drifts in scope, nature, and purpose extremely common and potentially widespread
3. **The conspicuous nature of many biometrically-scannable identifiers** (say, our faces or gaits) makes such data potentially very public, and has already led to a significant loss of privacy whenever people are in public places, either physically or virtually

There are many uses of Biometric Data already in commercial implementation. While this is usually considered benign, there is an extreme asymmetry with respect to the benefits of the use of such data given the risks outlined above. Such risks are rarely characterized with transparency and disclosure. If the risk/reward profile was more widely recognised, it is our belief that uses of Biometric Systems would be held to higher standards, and fewer usages of frivolous Biometric Systems would occur. Further, if the law required sufficient levels of governance, protection and oversight be applied to uses of Biometric Data, then the "cheap and easy" solution would begin to lose its luster.

Current adoption of Biometric Systems is built on a flawed risk/reward profile. The reward is skewed to the system operator and the risk is almost exclusively held by individuals subject to these systems. As of now, many biometric systems are being tailored to maximize profitability or efficiency, without enough regard to the possible - and often probable harms that humans may incur, including often unproven and potentially unprovable conclusions reached by these systems about us.

ForHumanity has deep concerns regarding the accuracy, validity, and assumed causality of many Biometric Systems. Many of the inferences (e.g race, mood, personality characteristics, mental state) are very hard to test against a ground truth, as they are often "fuzzy" even when being defined by humans and are frequently based on scant scientific evidence. Furthermore, most of these systems are poorly adapted to edge cases, like the disabled, neuro-divergent or other at-risk Protected Categories and intersections thereof. In consequence, ForHumanity requires a risk-based approach to accepting Accuracy, Validity, Reliability, Robustness, Resilience (AVR3). Biometric Systems call for especially high transparency, high disclosure, and an assumption of high risk. For further insight on our treatment of [Biometric Data](#), please read our published work.

FH AI Risk Management Framework

ForHumanity's Risk Management Framework is designed for AAA Systems specifically and for integration into ISO 31000 and COSO ORM and ERM applications. The information below can also be found [here](#).

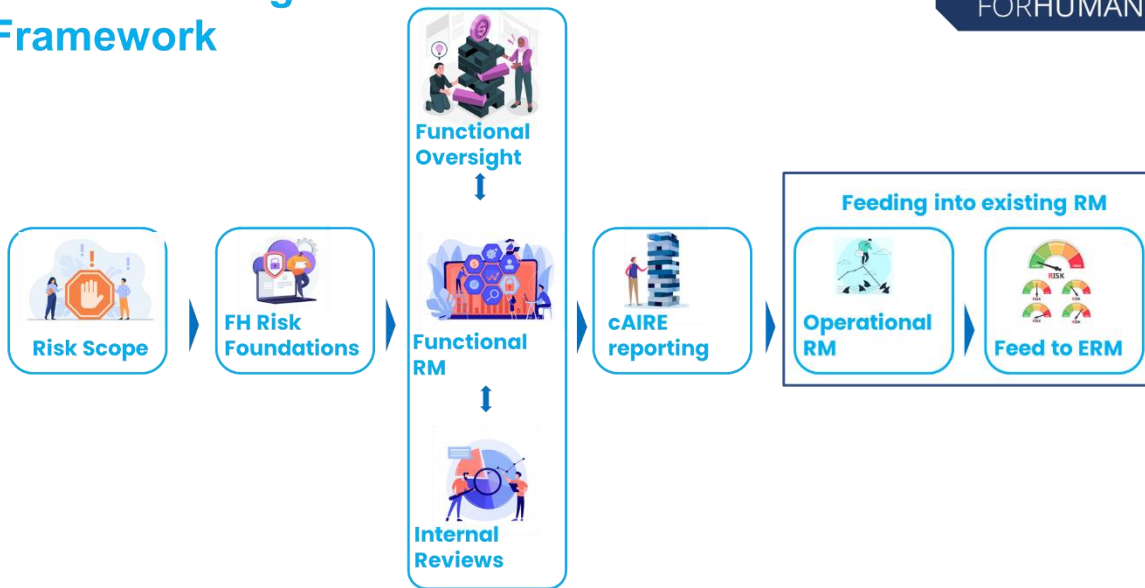
Concept

FH Risk Management Concept

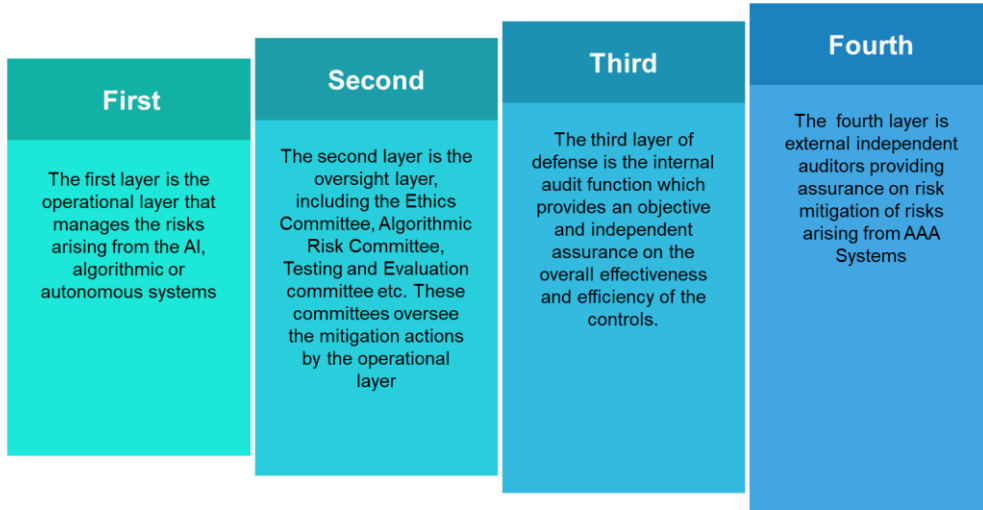


Overall Framework

FH Risk Management Framework



Four Layers of Defense



ForHumanity Risk Foundations

ForHumanity foundational reading on risk management (Principles)

ForHumanity’s mission is to mitigate downside risks posed by AI, algorithmic and autonomous systems. One of the clear ways to mitigate risk is to implement and operationalize a robust & agile Risk Management framework.

ForHumanity’s approach to risk management is centered on Ethics, Bias Privacy, Trust and Cybersecurity. Considered from a 360-degree multi stakeholder perspective, these pillars encapsulate the range of negative impacts and risk from socio-technical systems. ForHumanity wraps those pillars with a risk management framework that ensures compliance, mitigation and operability including characteristics such as: ethical, human-centric, accountable, governable, overseeable, transparent, documentable, proveable, evidence-based, and independently auditable.

FH Risk Perspective



Human Centric, Ethical & Fair



Actionable, Operational & Accountable



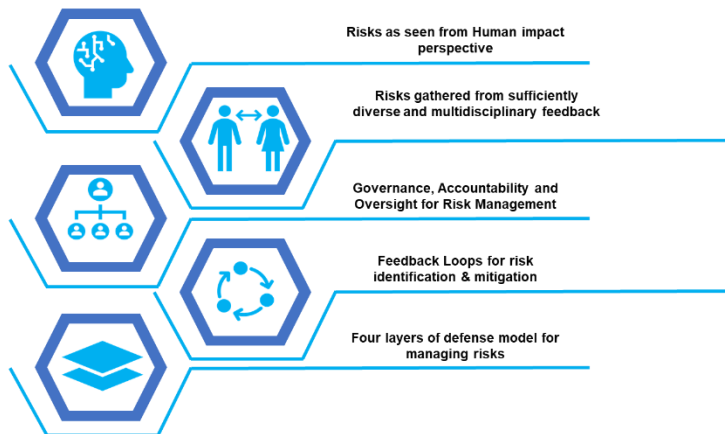
Auditable, Certain & Transparent

From a ForHumanity context, Risk management is an essential component to not just enable compliance with the criteria, but also sustainably prevent, detect and respond to emergent risks. ForHumanity advocates for a risk

management framework that is omni-directional and multivariate. Multivariate in that the framework considers corporate risk (which damages employees and shareholders), risk to humans (which damages users/clients/prospects and unwitting participants), societal risk (which damages our systems, groups, communities, markets and collectives) and environmental risks (which damages nature and sustainability considerations). As risk is never wholly removed, residual risk will always remain. These residual risks, well disclosed and considered, will empower an increased ability to identify emerging risks, support concentrated research on novel mitigations and encourage informed acceptance of consequences when residual risk manifests itself.

Foundational Principles

FH Baseline Risk Principles

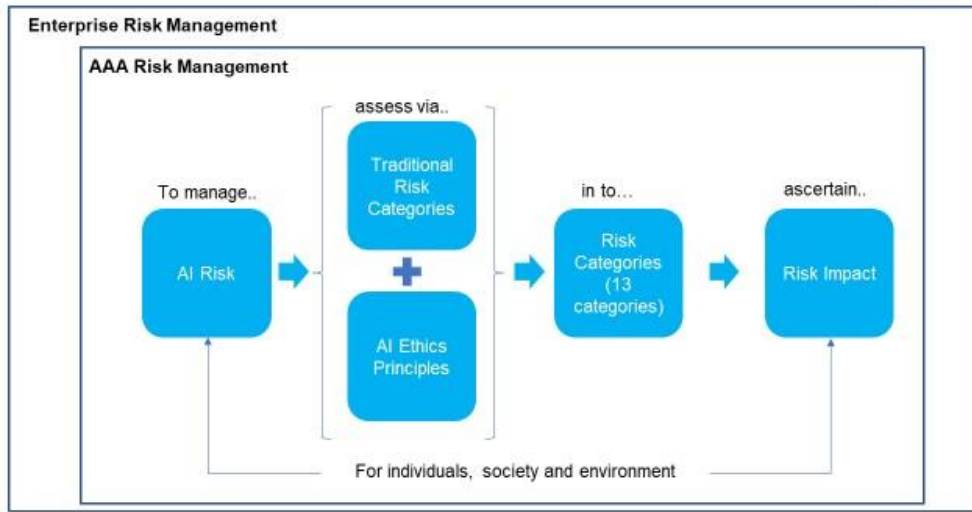


Read the complete brief here [FH foundational reading on Risk Management](#)

Also read about Low risk AAA system identification process here

[Identifying Low Risk AI, Algorithmic and Autonomous Systems](#)

Risk Taxonomy



Risk Management Policy



INCLUDE MULTI-STAKEHOLDER FEEDBACK

Explain the approach to gathering diverse inputs and multi stakeholder feedback

DEFINE FREQUENCY & REASSESSMENT CRITERIA

Establish a periodicity of reviewing risks and criteria for reassessment of risks

DEFINE RISK TOLERANCE & RISK APPETITE

Define Risk Appetite and Risk Tolerance to enable risk evaluation, risk treatment and managing residual risks.



HIGHLIGHT SIGNIFICANT RISK

Highlight risks to the key stakeholders that has an impact to people.

DEFINE THE ROLE OF COMMITTEES

Set up essential committees to ensure adequate segregation of duties, oversight and accountability.

DEFINE RISK MANAGEMENT PROCESS

Provide broader overview of the AI Risk Management process and its integration with Enterprise Risk Management

Maximizing risk mitigation for humans

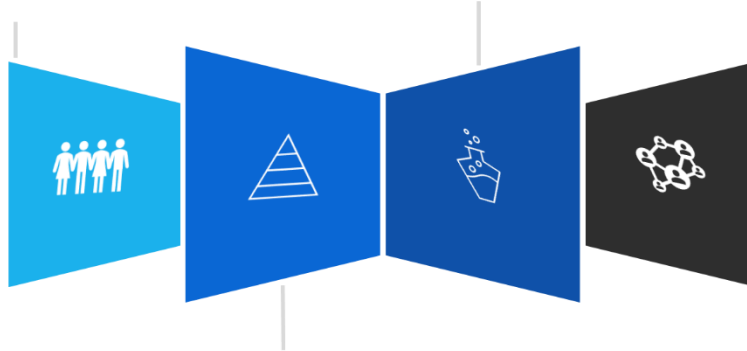
Foundational Guiding Documents

Maximizing Risk Mitigation for Humans



Risks need to be mitigated for who will get impacted by the risk

While maximizing risk mitigation, care shall be taken with reference to risk interactions



Risk Mitigation shall be maximized to a reasonable degree

Maximizing Risk Mitigation to a reasonable degree will inherently reduce organizational (accountability) risks

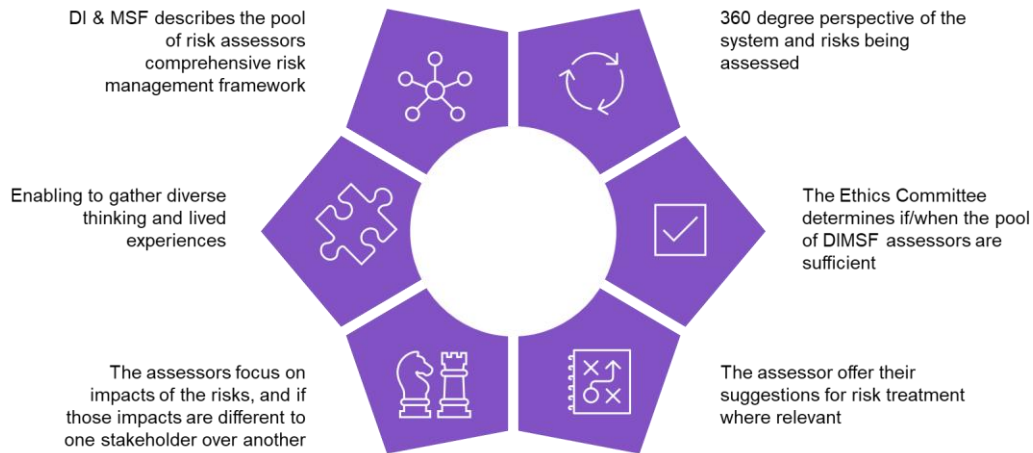
Diverse Inputs & Multi-stakeholder feedback

Diverse Inputs and Multi-Stakeholder Feedback & associated guidance

DIMSF - Guideline and template

Operationalizing Risk Management

Diverse input & Multi-stakeholder feedback



Need for committees

Functional Risk Management

Committees contribute to



Multi-disciplinary

with varied skills and experience from relevant domains

Risks & impact-

ability to assess risks and evaluate their impacts.

Establish and fairly value **Risk**

Tolerance and Risk Appetite

SOD &

Accountability have clear segregation of duties & function as a Second Line of Defense (2LOD)



Increase **Diverse Input and Multi Stakeholder Feedback** and risk input process

Manage roles and responsibilities **risk** from departures, role changes, and institutional knowledge

Reduce risk of autonomy, and bias (including cognitive biases)

Manage residual risk by treating the risks and disclose the residual risks

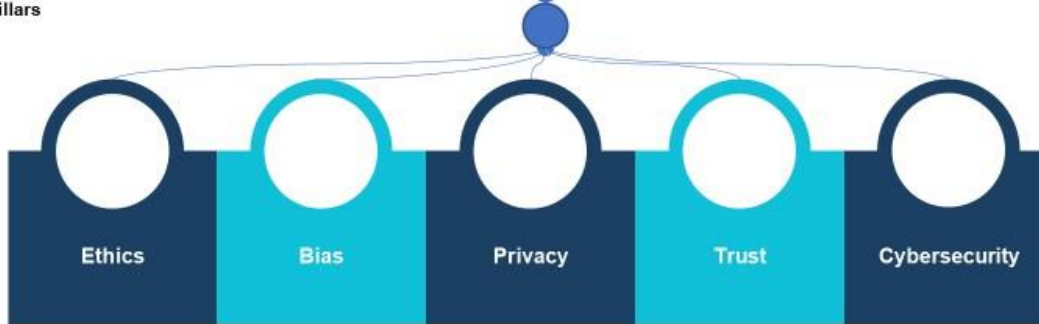
Committees & Risk Coverage



Committees (illustrative)



Pillars

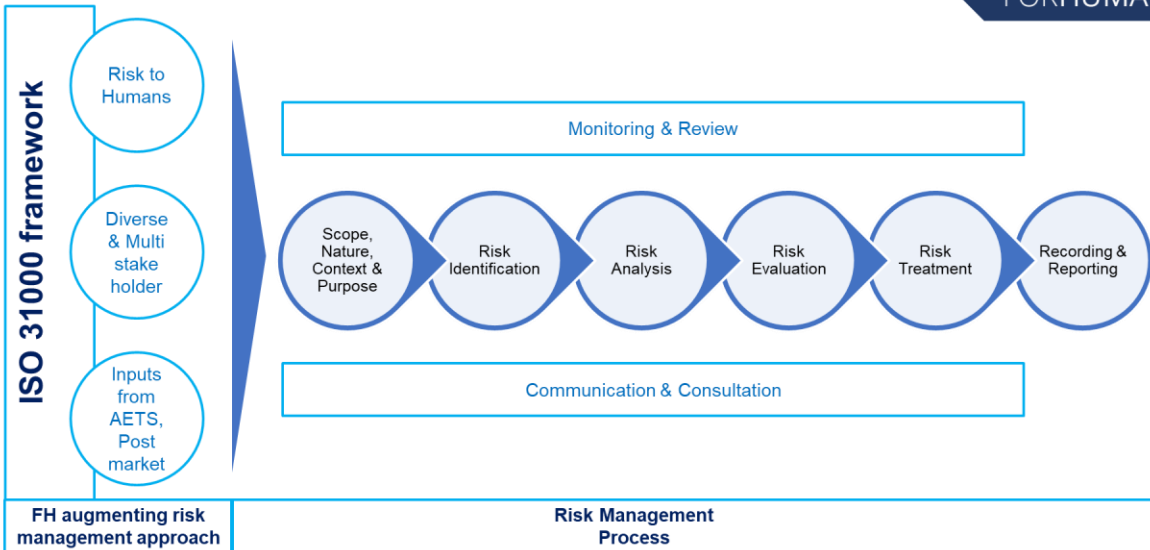


FH Risk Management Process

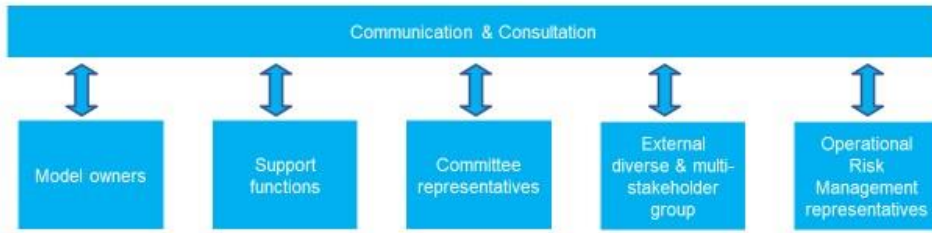
FH - AI Risk Management Processes

Foundational Guiding Documents

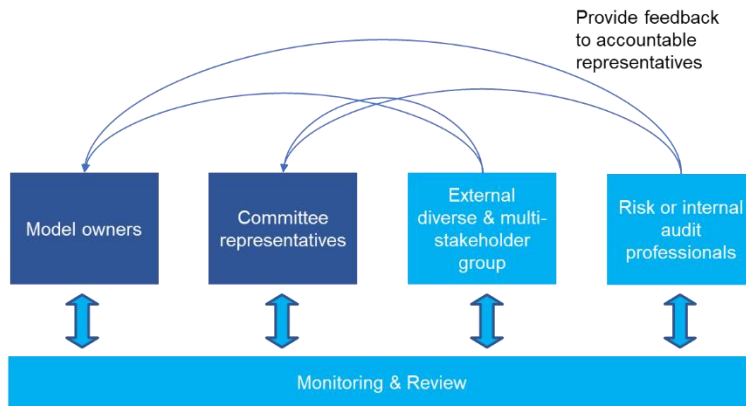
Risk Process Flow



Risk Process Flow



Risk Process Flow



Accountable

Input providers

Operationalizing Risk Management

Process associated with Risk Categories



Key elements of operationalizing process associated with Risk Categories



Establishing Principles



Establishing Policy



Deploying Process



Establishing Review, Oversight and Monitoring



Remediation & Disclosure

Functional Oversight

Lines of Defence

Functional Oversight

Technical Documentation & Reports

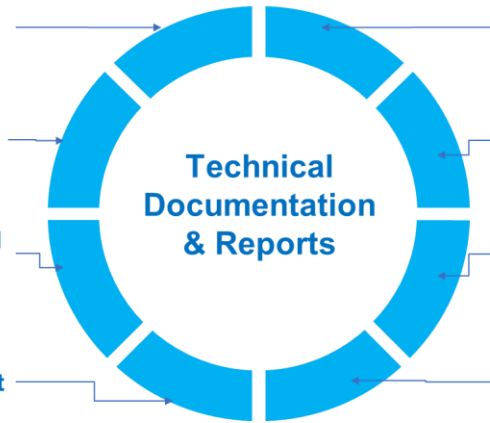


Data Management Report

Data Transparency Document

Pre-deployment Model Validation Report

HTL Integration Report



Model Health, Fitness and Monitoring

Post Deployment Model Management

Information Management Report

User Guide

Functional Risk Management








Responsibilities of Committees

Responsibilities of committees in AI lifecycle

Explaining the role of committees across the lifecycle of the AAA systems

Role of product, business and other stakeholders in risk management in AI lifecycle

Functional Risk Management reports

Committee	Subject	Guidance & Templates
Algorithmic Risk Committee (ARC)	ARC Structure and Governance	 BoK on ARC and ARA
	Algorithmic Risk Assessment Components	 ARA Components and Guidance
	Algorithmic Risk Assessment Template	 ARA-Risk template
Ethics Committee (EC)	EC Structure and Governance	 BoK on EC Structure and Resp...
	Ethical Risk Assessment Components	 ERA Component and guidance
	Ethical Risk Assessment Template	 ERA-Risk templates
Testing & Evaluation	TEC Structure and Governance	
	TEC Components	
	T&E At-Risk Report Template	
Children's Data Oversight Committee	CDOC Structure and Governance	 BoK on CDOC Structure and R...
Data Management Committee (DMC)	DMC Structure and Governance	
	Data Management Report Components	
	Data Management Report Template	
AI Governance	AI Governance Structure and Governance	
	AI Governance Components	
	AI Governance Assessment Template	

Residual Risk Management

Operationalizing Risk Management

Residual Risk Management



Internal Reviews

Lines of Defence

Internal Reviews



Note: Brief Intro [Incident Management essentials in the context of AAA systems](#)

cAIRE Reporting

Understanding cAIRE report -

[cAIRE report](#)

Monitoring & Reporting

cAIRE Report

Risk & Control Log

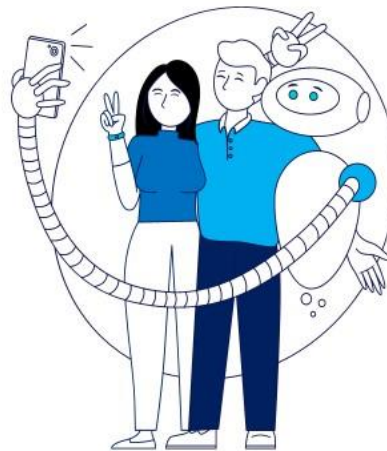
Consolidation of risks and mapped mitigating controls for risks where appropriate mitigations exist

Residual Risk Schedule

Consolidation of the residual risks from all the reports along with the treatment plan and specific impact assessments

Threat and Risk (Emergent & Horizon scanning)

List of emergent risks identified based on horizon scanning (including industry, domain, technology etc)



Risk and Control Scope template-

[Risk and Control Scope](#)

cAIRE residual risk log -

[Residual risk log](#)

Threat and Risk (Emergent & Horizon scanning) & Systemic Societal

- Threat and Risk Template - to be created
- Systemic Societal Risks - an Introduction:
- [Systemic Societal Impact Analysis](#)
- Systemic Societal Risk template - to be created

Feeding into Operational Risk Management at an Organizational Level

Enterprise Risk Management - Guidance for integrating AI risk with ERM

[COSO ERM: AI Risk Management integration](#)

ForHumanity University

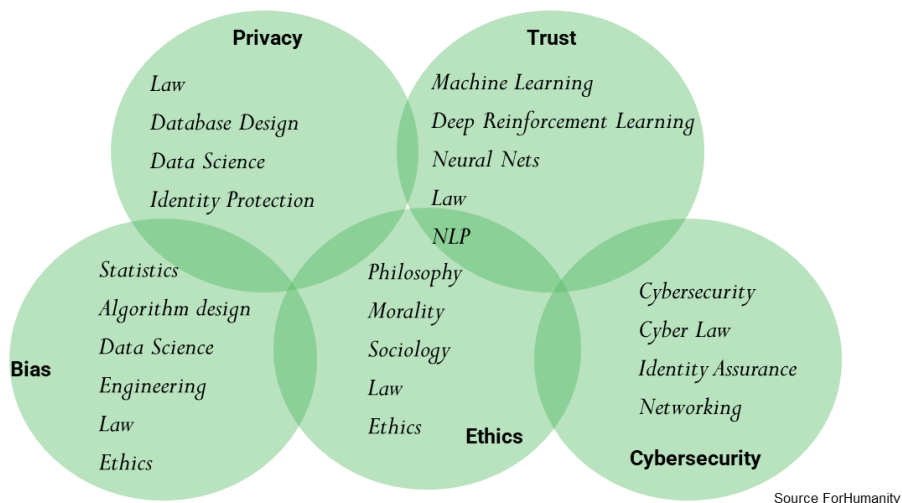
ForHumanity University is an online teaching environment designed to educate individuals on the precise skills required to execute in the ecosystem of Independent Audit of AI Systems. Coursework such as *Foundations of IA AIS* prepares the student with a broad understanding of terminology, theoretical and historical underpinnings, checks and balances, segregation of duties, the nature of audit criteria and IA AIS itself.

Building upon *Foundations of IA AIS*, ForHumanity University takes two pathways, the first is the ForHumanity Certified Auditor (FHCA) program discussed in more detail in the next section. This curriculum assures the student has a deep understanding of the audit criteria for which they are becoming accredited. Coursework covers terms and

definitions, individual criteria, and associated documentary evidence. Upon completion, the student will have demonstrated a comprehensive understanding of the audit criteria through examination.

The second pathway is the Accredited Expert program such as the certifications for ForHumanity’s Risk Management Framework and Ethics Committee. These curricula are designed for practical application by students in the field of AAA System risk management, or as a dedicated Ethics Officer on the Ethics Committee responsible for algorithm ethics. These expert certifications will provide job seekers and employers with a proven credential by which job requirements may be partially fulfilled in an area of expertise where training is still in its infancy.

These studies cannot replace the enormous amount of multi disciplinary study necessary for understanding the interplay between Ethics, Bias, Privacy, Trust and Cybersecurity. Below is a chart to illustrate some of the multidisciplinary studies that universities and institutes of higher learning can provide their students, with either generalist knowledge or deep specialist knowledge in any specific subset. ForHumanity aims to coordinate our accreditations with universities and institutes of higher learning around the world.



Examples of multi-disciplinary study in IAAIS

Trained and Accredited Workforce

The ForHumanity Certified Auditor (FHCA) accreditation process is exclusively for individuals and not organizations. This rigorous training teaches an individual about the Foundations of Independent Audit of AI Systems and the specific elements of audit criteria, documentary evidence for compliance and the critical elements of the Code of Ethics and Professional Conduct by which all FHCAs abide upon completion of their accreditation.

The accreditation process assures uniformity across all audits under the Independent Audit of AI System. Auditees can expect the same process and evaluation of compliance, regardless of the auditor or certification body.

FHCAs are required to maintain continuing education on the ever-changing landscape of laws, regulations, and best practices in order to maintain their good standing.

ForHumanity administers the continuing education and certification of good standing and provides the marketplace with verifiable credentials.

Compliance-by-Design

For Humanity's Certification Schemes are designed with the intention that most of the elements in the audit rules will promote and allow for compliance-by-design. It is anticipated that designers and developers will build the audit backup satisfaction into the system, plugging into the COSO system of internal risk control, governance and audit compliance as well as operational risk management and enterprise risk management.

Board of Directors Audit

There are audit criteria dedicated to Boards of Directors and which must be answered according to the required audit documentary evidence. It is not expected that the Board of Directors will have day-to-day responsibilities associated with audit compliance, however, the Board should have accountability for systemic failures of governance and accountability systems. Audit criteria are designed to ensure culpability, designed to ensure that the Board has adequate knowledge and oversight of key elements of the audit process. Notably, the requirement to establish Algorithmic Risk Committee, the Ethics Committee and the Children's Data Oversight Committee. All audit criteria referencing the Board are the sole responsibility of the Board of Directors. Auditors should ensure that the Board is the respondent of record to relevant audit questions.

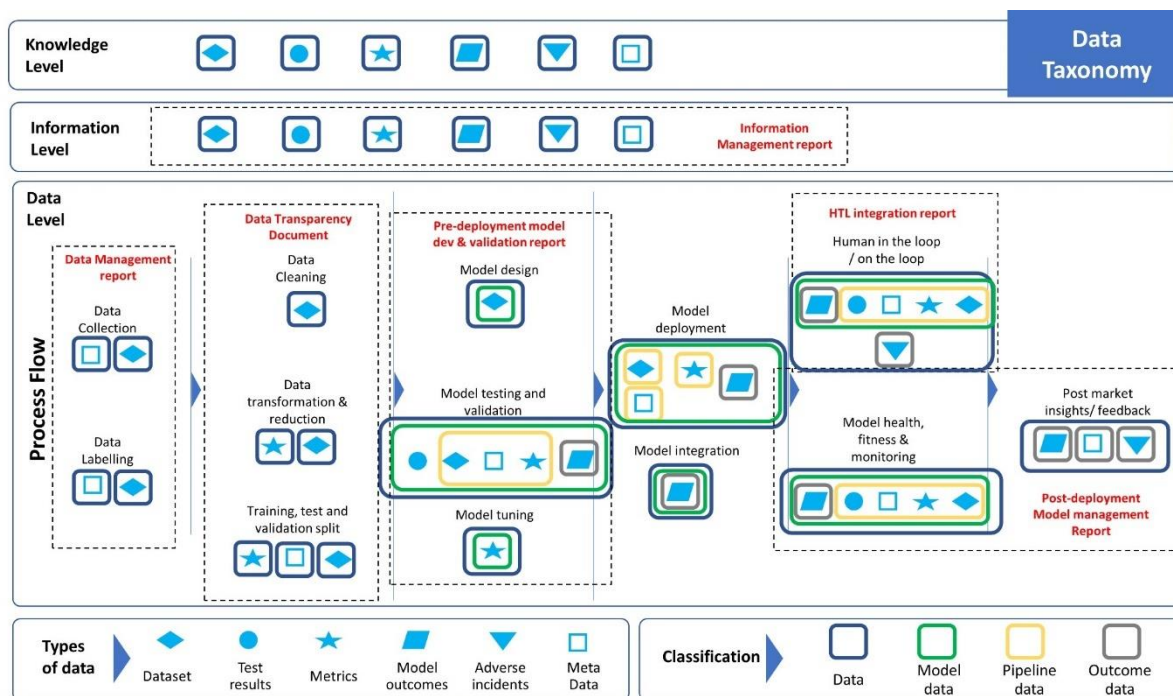
Evaluation Methods

Each of the scheme criteria identifies a type of evaluation method. The auditor may vary the evaluation method type where it provides additional assurance, but not so that it provides less. The following types are listed:

1. *Contract*. An executed contract can be examined and demonstrates compliance with the criteria.
2. *Correspondence*. Historical correspondence is available that demonstrates compliance with the criteria.
3. *Internal log, register or database*. Internal records and reports or systems with proof of authenticity can be examined by the auditor, and demonstrate compliance.
4. *Internal procedure manual*. Internal procedural documentation can be shown to the auditor that demonstrates compliance with the criteria. Note that these procedures should be of sufficient detail to show that they are up-to-date, implemented, operational and complete. High-level policies are not sufficient to demonstrate implementation.
5. *Public disclosure document*. A publicly disclosed document will demonstrate compliance. This may include comparison to other evaluation types.
6. *Physical testing*. This can refer to any of the following, at the auditors' discretion:
 - a. *Records of previous events that can be examined*. For example, if there is a clear audit trail demonstrating the response to prior Data Subject Access Request, the auditor can review this audit trail to gain confidence that the organization can comply with the criteria.
 - b. *Witnessing current events*. For example, to ensure that an organization can restore from backup, the organization can demonstrate its ability to do so to the auditor.
 - c. *Technical testing*. For example, to demonstrate that network traffic is encrypted, the auditor may inspect the traffic.

Copies of all evidence obtained during the evaluation should be stored in encrypted form by the auditor, except where this includes personal data and does not comply with the principle of data minimisation.

Data Taxonomy and Technique Documentation of AAA Systems



ForHumanity has established a data taxonomy to enable increased precision and specification of the overused word “data.” The EU Artificial Intelligence Act requires significant amounts of new documentation systems to achieve compliance with the Act. The required documentation is listed on the diagram above in red font, with the exception of the AAA Systems User Guide, a document presented to Users by providers of AAA Systems:

- 1) Data Management Report
- 2) Data Transparency Report
- 3) Information Management Report
- 4) Pre-Deployment model development and validation report
- 5) Post-deployment model management report
- 6) HTL Integration Report

This diagram is from a forthcoming paper on Data Taxonomy. Beyond satisfying legal requirements for documentation, the Data Taxonomy paper provides clarity on the word “data” in the following ways:

- 1) Process Flow - from Left to Right - Sourcing Data to Design to Development to Deployment to Integration to Decommissioning
- 2) Data Hierarchy - from bottom to top - Data to Information to Knowledge
- 3) Data Type - keyed in from shapes
 - a) Dataset
 - b) Test Results
 - c) Metrics
 - d) Model Outcomes

- e) Adverse Events
- f) Metadata
- 4) Data Classifications
 - a) Data
 - b) Model Data
 - c) Pipeline Data
 - d) Outcomes Data
- 5) Data Task - described at each step of the Process Flow, to each Data Type

Process Flow

- 1) Design
 - a) Source Data
 - b) Data Management and Preparation
- 2) Development
 - a) Training, testing/validation and transformation
 - b) Pre-deployment, Development, Validation and Functional Correctness
- 3) Deployment
 - a) Pipeline
 - b) HTL Integration (Human-in/on-the-Loop including Human-in-command)
 - c) Downstream Integration
 - d) Model health, fitness and monitoring
 - e) Post Market monitoring and Adverse Impact Reporting Systems (AIRS)
- 4) Decommission

The paper provides a detailed explanation of each term, each step and each differentiation. These definitions already exist in our GDPR audit criteria or EU Artificial Intelligence Act criteria, but are behind combined in the paper for clarity.

Body of Knowledge - Knowledge Stores

The Body of Knowledge and its specific Knowledge Stores are guidance notes for Auditors, to be applied when examining items of compliance sufficiency and maturity. They do not represent normative criteria. Instead they reflect measures, tools and thresholds that help an Auditor understand if the documentary evidence is sufficient, or even a mature level of compliance. Further, the Knowledge Stores highlight frequent insufficiencies related to documentary compliance evidence designed to draw attention to common mistakes with sufficiency. The Body of Knowledge - Knowledge Stores can be found [here](#).

Certifications

Certification Merits

Independent certification of conformity to ForHumanity's Certification scheme for AAA systems represents the highest form of compliance.

1. Certification demonstrates the organization's willingness to transparently and objectively document controls, governance, and accountability with respect to the approved certification schemes.
2. Certification helps your organization demonstrate compliance to the regulator, the public and in your business-to-business, supply chain relationships.

3. Mitigation of risks associated with AAA Systems.
4. The Certification mark received by the organization upon compliance is an outward, public expression of a commitment to uphold the law to the highest standard and to indicate to clients, customers, prospects and employees that their Personal data/Personal Information is well protected, used ethically and fairly and that interactions, interfaces and outcomes will be fair and according to the Code of Ethics and Code of Data Ethics.
5. Certification enables an opportunity to have an independent examination of the data supply chain, the AAA System life cycle, including the associated supply chain and includes robust documentation and disclosures.
6. It enables organizations to prove monitoring of ethical, bias, privacy, trust and cybersecurity risks, to implement proportional controls, and encourage a company culture upholding privacy.

Limitations of Certification

Certification does not provide immunity from regulatory scrutiny or action, and is not a guarantee of continuous compliance with the law and/or certification schemes.

Certification marks must be used in association with the pre-agreed disclaimer language to disclose the data processing purpose that is covered by the certification mark and any limitations.

Engagement with a Certification Body

The Auditee shall engage with an Auditor (Certification Body) by executing an Audit Engagement Letter. This letter will explicitly identify a Target of Evaluation (TOE) upon which the certification will be conducted. This letter shall stipulate the following:

1. Scope of the Data Processing Purpose including beginnings and ends (ToE)
2. Disclaimer for the certification mark
3. Rules and guidelines for use of the certification mark
4. Expectations from an auditee to provide documentary evidence
5. Expectations regarding ongoing and post market monitoring
6. Certification Plan, including, if applicable:
 - a. Opening meeting where the scope is verified and the names of organizations and individuals participating, and their roles
 - b. Confirmation of the authorisation of the auditors to award the certification, and their impartiality
 - c. The Target of Evaluation (ToE, as documented in the contract)
 - d. The legal basis of the AAA System and the role of the Auditee
 - e. Expected documentary evidence
 - f. Physical testing scheduling
 - g. Any site or network access required, and any special requirements for that access (e.g. permission to conduct intrusive network scanning)
 - h. Closing meeting for presentation of Certification Report, issuance of Certification or issuance of Non-Compliance Letter
7. Certification Report, including:
 - a. Clear explanation of the scope agreed in the Audit Engagement Letter and the beginnings and ends, also expressed in the disclaimer
 - b. Any deviations from the certification plan
 - c. Process narratives, walkthroughs, flowcharts, diagrams, control descriptions, codes, policies

- d. The specific software and hardware versions and assets inspected including third-party assets, as applicable
- e. The actual dates of inspection(s)
- f. A list of documentation and assets that will be retained as audit evidence, and explanation of deviations
- g. A duly authorized signatory
- h. A list of deficiencies, if certification will not be issued
- i. A determination of sufficient/mature levels of compliance
- j. Whether a certification is awarded, and its duration
- k. Sufficient deliverable for disclosure requirements
- l. Sufficient, robust and resilient ongoing monitoring systems and explicit statement that systemic failures of ongoing monitoring systems will preclude future certification

Auditor - Auditee agreement on Scope

See section Target of Evaluation Determination Process

Certification Warning/Certification At-risk

The Certifying body (Auditor) may issue a written warning to an auditee that they are not compliant with the terms of the Audit Engagement Letter. This written warning shall include a timestamp, remediation period, and the expected remedy. Failure to satisfy may result in the withdrawal of certification. Potential warnings could include:

1. Misuse or misrepresentations in use of certification mark and their stated purpose
2. Contravention to any of the contractual clauses for certification
3. Failure to maintain documentary evidence related to the certification
4. Failure to maintain post market, robust and ongoing monitoring on the data process
5. Failure to uphold agreed and documented thresholds, Key Performance Indicators on the data process
6. Concept drift and deviations from scope, nature, context or purpose of the data processing
7. At the launch of an investigation based upon a report or complaint by the FH certified auditor highlighting potential misrepresentation, falsification or fraud associated with information provided for audit
8. Reported data privacy breaches

Warnings and at-risk certification may or may not lead to revocation of certification based upon this guidance and failures to remediate in a timely fashion, at the discretion of the Auditor.

Withdrawal of Certification

Certification may be withdrawn for any of the following reasons:

1. Regulatory action related to the data process
2. Successful civil litigation of a case directly pertaining to the data process certified
3. Failure to maintain documentary evidence related to the certification
4. Failure to maintain post market, robust and ongoing monitoring on the data process
5. Failure to uphold agreed and documented thresholds, Key Performance Indicators on the data process

6. Concept drift and deviations from scope, nature, context or purpose of the data processing
7. Material change in organizational governance, accountability, oversights or controls related to the data process
8. Reported data privacy breach
9. Fraud, misrepresentation or malfeasance associated with material information related to the certification

The Auditor will notify the Auditee that certification has been withdrawn with a Letter of Withdrawn Certification and will be required to provide the auditee with the associated reason for the withdrawn certification from the list above. This may be done at their sole discretion according to the Audit Engagement Letter for any reasons listed above.

Certification mark use standards and guidelines

See the ForHumanity license agreement and website for standards and use guidelines for certification marks.

Certification Steps

Define Scope

ForHumanity designs certification schemes for specific AAA Systems. The scheme may only be applied to systems that fall within parameters and scope as defined in the certification scheme.

Target of Evaluation Determination Process

The Auditee determines the data process(es) to which the Auditor will apply the scheme and documents this agreement in a contract. The Target of Evaluation (ToE) shall be defined by an Audit Engagement Letter between the auditor and the organization (the Auditee).

The Audit Engagement Letter shall document all information required by the Auditor for a sufficient Certification Plan and shall include all of the following:

- 1) Name/identifier of the ToE, specifically noting the boundaries of the data process
- 2) Beginnings and Ends of the Data Process(es) where Personal Data is processed (including a visual representation)
- 3) Systems or organizations expected to be “in” or “out” of scope (including data Processors under contract), including a visual representation as appropriate
- 4) Description of the lawful basis for processing, as well as its scope, nature, context and purpose
- 5) Description of the data deployed in the system, specifically noting the Personal Data/Personal Information and Special Category Data/Sensitive Personal Data/Biometric Data that may be present (including Inferences and/or potential Proxy Variables)

The Auditor will only perform an audit of the documented scope. The Auditee bears the responsibility of ensuring that all relevant aspects, infrastructure, data, data processes, storage, interfaces, software, service providers and output system necessary for the proper function of the AAA System identified as a ToE undergo an Audit.

The Auditor and Auditee shall document in Audit Engagement Letter the wording of an associated disclaimer from ForHumanity to be published alongside the Certification mark to provide clarity on what has been certified.

Conduct Pre-assessment/ Pre-audit

Certification requires extensive preparation and work to ensure that requirements will be met completely. During a certification audit, there is limited ability to remedy meaningful shortcomings. Therefore, the organization is strongly advised to invest time in pre-audit compliance, designed to meet the requirements of the audit in advance, while preparing the organization for ongoing maintenance needed to maintain certification. Pre-audit compliance should establish the following key components for certification compliance:

1. Establish infrastructure for governance, accountability and oversight as specified in the certification criteria
2. Identification of certification criteria requiring documentary evidence
3. Establishing process for compiling and storing documentary evidence
4. Identification of training needs and sourcing auditable solutions
5. Drafting codes and procedure manuals
6. Identification of system requirements (hardware and software)
7. Verifying operational risk management and control processes
8. Preparation for disclosure requirements

Identify Certification Body

ForHumanity relies upon local, government-approved accreditation services (e.g United Kingdom Accreditation Service (UKAS)) when one exists. ForHumanity will provide information and resources to the accreditation service in support of their mission. In the event that a government-approved accreditation service does not exist, ForHumanity has a process for evaluating sufficiency of accreditation bodies and is willing to provide that service.

Identify Auditors for Certification

Auditors must be trained and certified in the scheme that they intend to provide. Not all persons engaged in the provision of certification services must be individually certified, however the issuance of a certification may only be provided by a named ForHumanity Certified Auditor (FHCA).

FCHAs are well-trained in the certification scheme criteria and the audit process that leads to certification. They are required to maintain their knowledge through continuing education and their current status can be checked on the ForHumanity Certified Auditor website found [here](#).

FCHAs must abide by the principle of Independence.

Independence Enforced via License

As a legal term defined in America by [The Sarbanes-Oxley Act of 2001](#), a certifying body (an Auditor) must receive no other remuneration from an Auditee beyond reasonable audit fees. ForHumanity further stipulates, in its license agreements, that a licensee cannot be an

Auditor and an Assessor/Consultant (or provide any other form of service) to the same Auditee in a 12-month period.

More details on specific examples of Independence can be found in [ForHumanity's Certified Auditor Code of Ethics and Professional Conduct v1.0](#).

Anti-Collusion

Independence is further enforced through licensing requirements enforcing anti-collusion amongst auditors and pre-auditors. As the market for data auditing matures and grows, it is impermissible for pre-audit service providers and auditors to regularly guide clients to each other excessively. This prevents pre-auditors and auditors from becoming overly comfortable with each other's processes/expectations and failing to deliver the maximum diligence and objectivity owed to the client and the public. Anti-collusion requirements ensure maximum mitigation of risk to humans and implement complete compliance.

Certification Issuance

Upon completion of the audit process, an accredited auditor in their sole discretion will either issue certification or explain why certification has been denied. This occurs after all attempts at remediation have been made by the auditee within a reasonable time period as determined by the auditor.

ForHumanity and Accreditation Service Examinations

Both government-appointed accreditation services and ForHumanity have the right and responsibility to periodically review certification reports to ensure that accredited certification bodies are conducting their work in a responsible and proper manner, consistent with the requirements of the approved certification scheme. The organization receiving the certification for a AAA System would only be notified if there were a discrepancy or shortcoming of the certification process. The auditee would have a reasonable opportunity to rectify and meet the requirements of certification.

Audit Period of Validity

A certification is good for one year. Compliance should be renewed annually and an auditee is expected to maintain compliance with the current version of the audit. In any areas where the audit has changed, the auditee will have until the next annual audit to bring their systems into compliance.

Significant changes in the nature, scope, and purpose of an AAA System should require updated certification. Significant changes to an algorithmic system may jeopardize the certification status. Some examples which may require recertification to maintain status are:

1. Acquisition/Change in Control
2. Complaint
3. Regulatory intervention
4. ForHumanity's Cause for Concern

Recertification

It is expected that organizations will want to maintain their certified status for the data processing purpose. The organization will be welcome to recertify against the current version of the certification scheme, and it is expected

that recertification will be substantially easier resulting from the investment in compliance from the original certification.



Written Comments from Merve Hickok
Regarding Local Law 144 of 2021 in relation to Automated Employment Decision Tools

TO: New York City Department of Consumer and Worker Protection (DCWP)
Rulecomments@dca.nyc.gov

6/6/2022

Dear Chair and Members of DCWP,

As the Founder of AIethicist.org, I appreciate the opportunity to provide public comments regarding Local Law 144 of 2021, amending the administrative code of the city of New York, in relation to automated employment decision (AED) tools.⁹ With this law, NYC legislature passed a globally pioneering law which requires automated employment decision tools used for employment or promotion decisions within NYC to undergo an annual independent bias audit.

As professional artificial intelligence (AI) ethicist, and a certified human resource professional with 20 years of HR experience, I am deeply interested in the development, implementation and regulation of technologies incorporating AI into HR products. Through my organization and my affiliated civil society and academic work, I provide research, training, and consulting on how to develop and use these products in a responsible way. My research and civil society activities focus on bias, accountability, governance, human rights and accessibility across AI and algorithmic decision-making tools. I also conduct research and training on AI policy and regulation at Center for AI & Digital Policy, and lecture on data science ethics at University of Michigan.

As DCWP is proposing to add penalty schedules for violations related to AED tools, I would like to take this opportunity to 1) provide feedback to strengthen the legislation's protections, and 2) request further clarification regarding the terminology. These changes would both protect individual's rights and make the expectations from the employers and vendors are clearer. In return, it would make it easier to detect the violations, and prevent different interpretations of the requirements.

Recommendations:

Inclusion of Disability: Law 144 requires AED tools to be audited for disparate impact on individuals based on their race, ethnicity, or sex. However, it notably lacks any audit requirements for people with disabilities, and consideration of impact of these systems on their access to opportunities. Recently, U.S. Equal Employment Opportunity Commission (EEOC) published a technical assistance document detailing how AED tools may violate existing requirements under Title I of the Americans with Disabilities Act ("ADA") and providing tips to employers on how to comply.¹⁰ A bias audit must include a review of AED tools with respect to their implications on people with disabilities.

Notice and Disclosure: Employers are required to provide candidates minimum 10-day notice about the future use of AEDT and include the specific job qualifications and characteristics the tool will use in determining its outcome. This is to allow a candidate to request an alternative selection process or accommodation. A template or a guidance

⁹ Local Law 144 of year 2021: <http://nyc.legistar1.com/nyc/attachments/c5b7616e-2b3d-41e0-a723cc25bca3c653.pdf>

¹⁰ U.S. Equal Employment Opportunity Commission (5/12/2022). [The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees](#)

document establishing the minimum content of such Notice is crucial and necessary. Otherwise, the content can be very vague and high-level, or too long and full of legal and technical terms for it to be beneficial for the candidate.

Summary of the results of the most recent bias audit: Employers are required to disclose the summary of the results of the annual audit on their website. Like above concern, a template or a guidance document establishing the minimum content of such Audit Result is crucial and necessary. Otherwise, such documents can be only a few sentences long and not provide any quantitative or qualitative information to be beneficial for the candidate or public interest.

Requests for Clarification:

Definition of ‘independence’: The NYC law requires an ‘impartial evaluation by an independent auditor’. The legislation does not clarify the conditions of independence or audit. Currently there exists no accreditation scheme in or outside of US of a bias audit of AED tools.

Therefore, it is necessary to clarify

- 1) whether the auditor could be internal to the vendor developing, or employer implementing the AED tool?
- 2) what are the criteria to determine independence if an internal auditor?

In addition, the scope and rules of audit must be developed independent of the auditor conducting the audit to ensure conflict of interest and integrity of audit.

Definition of ‘bias’: In the absence of a definition of bias, the audits can turn into a simple check of disparate impact by using 4/5ths Rule. This is a race to the bottom and can turn the requirements of this legislation into a rubber stamp activity. Responsible AI practices and consideration of bias must be across practical and statistical tests, and across design decisions. The assumptions, decisions and trade-offs made by vendor and employers must be included as audit criteria.

Impacted employers: The legislation suggests ‘In the city, it shall be unlawful for an employer or an employment agency to use an automated employment decision tool to screen a candidate or employee for an employment decision unless’ certain obligations are met. However, it is not clear if impacted employers should be 1)employers based in NYC, or 2)employers based anywhere and hiring NYC ‘residents’, or 3)employers based anywhere and hiring into NYC-based ‘roles.’ Reference to ‘employee or candidate that resides in the city’ is only mentioned under the Notice requirement.

Thank you for your consideration of my views. I would welcome the opportunity to discuss further about these recommendations.

Merve Hickok, SHRM-SCP

Founder of AIethicist.org, and Lighthouse Career Consulting

merve@lighthousecareerconsulting.com www.aiethicist.org/mervehickok

TO: New York City Department of Consumer and Worker Protection (DCWP)
Rulecomments@dca.nyc.gov

RE : Local Law 144 of 2021

DATE: 6/6/2022

Dear Chair and Members of DCWP,

As Credo AI, an organization focused on empowering organizations to deliver Responsible AI at scale, we welcome the chance to provide public comments regarding Local Law 144 of 2021, in relation to Automated Employment Decision Tools (AEDT).¹¹With this regulation, NYC Council has taken a lead in regulating the automated employment decision tools before any other jurisdiction in the US or Europe. It sets a precedent by which other similar government entities at city, state, federal or national levels can require employers to have similar controls. It also brings a certain level of transparency to vendor and employer practices by way of requiring audit results to be disclosed by employers.

To ensure responsible AI development and use is embedded within an AEDT, design decisions, changes to data and model, and results of tests must be documented, and a robust governance structure must be in place. At Credo AI, we have been working on these challenges for years to ensure that AI is always in service to humanity. Our mission is to ensure that an increasingly AI-embedded society will have more equitable access to healthcare, education, and employment - not less. Our Responsible AI (RAI) Governance Platform provides context-driven governance and risk assessment to ensure compliant, fair, and auditable development and use of AI. We operationalize industry best practices, standards, and regulations into actionable Policy Packs that our customers use to assess whether their ML models, datasets, and development processes are meeting business, regulatory, and ethical requirements.

In response to DCWP's request for input regarding the proposal to add penalty schedules for violations related to AEDT, we would like to provide the following recommendations and request more detail on certain terminology within the legislation.

Bias Audit: The most significant lack of guidance within legislation is what is meant by a *bias audit*. The law specifies neither *bias* nor *audit* beyond the requirement for an assessment of disparate impact for sex, race, and ethnicity.

Independent audit rules: If an auditor (internal or external) determines what the rules and content of a bias audit should be, and then conducts an audit according to these self-determined rules, a conflict arises. The set of audit rules to be used for the purposes of this legislation must be objectively and independently crafted by a party outside of the AEDT vendor and employer in question. Only then an objective examination can be guaranteed.

Independent auditor: The new law states that a bias audit means “an impartial evaluation by an independent auditor” but fails to deliver clarity on what is acceptable. For example if a company chooses to do the audit internally, would the results of the bias audit be acceptable? If a company chooses to engage with a big 4, would it need to have the advisory separate from the audit for algorithmic auditing?

Disparate Impact: Legislation requires ‘bias audit shall include but not be limited to the testing of an automated employment decision tool to assess the tool’s disparate impact’. However, since no further requirement is set for

¹¹ NYC Local Law 144 of year 2021: <http://nyc.legistar1.com/nyc/attachments/c5b7616e-2b3d-41e0-a723-cc25bca3c653.pdf>

testing, the audit might be limited to checking of outcomes against a practical test of 4/5ths Rule. Such simplistic approaches would take away from the spirit of the legislation and add no value to the protection of candidates or employees impacted by these decisions. It means no further consideration is necessary for the new risks introduced by algorithmic systems. Though some progress has been made in confronting discrimination in the hiring process in the past few decades, there is still much work to do addressing systemic bias that obstructs more equitable employment opportunities.

Other protected categories: Comparing the first draft of the bill made public by the NYC Council with the final version, the scope of class attributes covered is reduced to disparate impact based on race, ethnicity, and sex. By leaving out explicit mention of discrimination based on disability, age, or sexual orientation - many people who suffer the consequences of bias are still left unprotected and further disadvantaged. We recommend these categories to be added to a mandated bias audit.

Access to data: An AEDT could be developed by a vendor or inhouse by the employer itself. Similarly, the vendor might be providing the model to a client, but the employer could have full control over the demographic data. For a robust disparate impact analysis, an auditor requires access to both the model and the demographic data, and the outcomes across different protected categories/groups.

Disclosure of audit results: The bill also requires employers to post a *summary* of the bias audit on their website but gives no outline for what that summary must contain. We believe a detailed version of the audit results must be retained by both employers and vendors as per their data retention policies. We also recommend the summary of the audit to be detailed and actionable enough to ensure public and candidate understanding of how these systems work, and for which groups they work better.

Credo AI is honored to provide this feedback for your consideration. We would welcome the opportunity to discuss these recommendations and share our work in support of the Local Law 144 of 2021.

Navrina Singh,

Founder & CEO of Credo AI www.credo.ai

Email: navrina@credo.ai



VIA EMAIL

June 6, 2022

Hon. Vilda Vera Mayuga
Commissioner
New York City Department of Consumer and Worker Protection
42 Broadway #5
New York, NY 10004

RE: Automated employment decision tools

Dear Commissioner Mayuga:

Thank you for the opportunity to comment on proposed rules related to Local Law 144 of 2021. The local law established that it was “unlawful for an employer or an employment agency to use an automated employment decision tool to screen a candidate or employee for an employment decision.”

The Retail Council of New York State represents member companies ranging in size from the nation’s largest and best-known brands to the smallest Main Street entrepreneurs that fuel local economies. According to [New York State Comptroller Thomas DiNapoli](#), there were 300,200 retail jobs in New York City in March 2022, making the industry the second largest employer in New York City behind the “office sector.”

We respectfully submit the following comments for your consideration:

- 1.) Neither the law, nor proposed rule, provides sufficient clarity on the specific tools that are covered by the definition of “automated employment decision tool.” Employers may use assessments that are not necessarily Artificial Intelligence or machine learning, but are scored by a computer. Clarity is, therefore, necessary, so employers are not inadvertently noncompliant.

A few examples of these assessments include:

- a. Employers using multiple choice tests where a computer scores the respondents answers against a standard answer sheet
 - b. A Computer Coding Skills Assessment in which a computer evaluates whether the appropriate variables are present in the output of the candidate’s final submission
 - c. A typing test in which a computer is scoring the words typed and subsequent errors.
- 2.) Similarly, some assessments may include multiple portions — a multiple-choice component that the computer scores, a separate interview or presentation with a human being, and a written component scored by a human being. The scores from all of the sections are combined to determine if the person should move

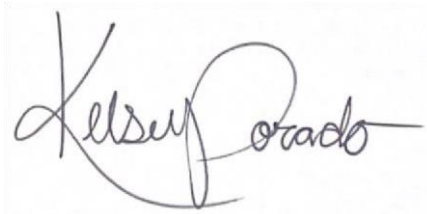
forward in the selection process. Clarity on whether this type of combined assessment, or only specific parts, would be covered by Local Law 144 of 2021 is necessary.

We urge the Department to provide clear and specific definitions for Artificial Intelligence and machine learning.

- 3.) Second and subsequent violations include “entering into a settlement agreement” within two years of the prior violation(s). This is cause for concern as it means retail employers would now incur higher per-violation penalties, regardless of whether the underlying claim had any merit. Additionally, this would result in companies being issued higher violation fines.

Thank you, again, for the opportunity to provide comment on the proposed regulations. We will remain constructive throughout the regulatory process, as always, and are available if you have any questions.

Sincerely,



Director of State and Local Government Relations Retail Council of New York State



NYU

TANDON SCHOOL OF ENGINEERING

Testimony of Julia Stoyanovich before the New York City Department of Consumer and Worker Protection regarding Local Law 144 of 2021 in Relation to Automated Employment Decision Tools

June 6, 2022

Dear Chair and members of the Department:

My name is Julia Stoyanovich. I hold a Ph.D. in Computer Science from Columbia University. I am an Associate Professor of Computer Science and Engineering at the Tandon School of Engineering, and an Associate Professor of Data Science at the Center for Data Science, and the founding Director of the

Center for Responsible AI at New York University. In my research and public engagement activities, I focus on incorporating legal requirements and ethical norms, including fairness, accountability, transparency, and data protection, into data-driven algorithmic decision making.¹² I teach responsible data science courses to graduate and undergraduate students at NYU.¹³ Most importantly, I am a devoted and proud New Yorker.

I actively participated in the deliberations leading up to the adoption of Local Law 144 of 2021^{14,15} and have carried out several public engagement activities around this law when it was proposed¹⁶. Informed by my research and by opinions of members of the public, I have written extensively on the auditing and disclosure requirements of this Law, including an opinion article in the New York Times¹⁷ and an article in the Wall Street Journal¹⁸. I have also been teaching members of the public about the impacts of AI and about its use in hiring, most recently by offering a free in-person course at the Queens Public Library called “We are AI”¹⁹. Course materials are available online²⁰.

In my statement today I would like to make three recommendations regarding the enforcement of Local Law 144 of 2021:

1. **Auditing:** The scope of auditing for bias should be expanded beyond disparate impact to include other dimensions of discrimination, and also contain information about a tool’s effectiveness - about whether a tool works. Audits should be based on a set of uniform publicly available criteria.
2. **Disclosure:** Information about job qualifications or characteristics for which the tool screens the job seeker should be disclosed to them in a manner that is comprehensible and actionable. Specifically, job seekers should see simple, standardized labels that show the factors that go into the AI’s decision both before they apply and after a decision on their application is made.

¹² See <https://dataresponsibly.github.io/> for information about this work, funded by the National Science Foundation through NSF Awards #1926250, 1934464, and 1922658.

¹³ All course materials are publicly available at <https://dataresponsibly.github.io/courses/>

¹⁴ Testimony of Julia Stoyanovich before New York City Council Committee on Technology regarding Int

¹⁵ -2020, November 12, 2020, available at

https://dataresponsibly.github.io/documents/Stoyanovich_Int1894Testimony.pdf

¹⁶ Public engagement showreel, Int 1894, NYU Center for Responsible AI, December 15, 2022 available at

<https://dataresponsibly.github.io/documents/Bill1894Showreel.pdf>

¹⁷ We need laws to take on racism and sexism in hiring technology, Alexandra Reeve Givens, Hilke

Schellmann and Julia Stoyanovich, The New York Times, March 17, 2021, available at

<https://www.nytimes.com/2021/03/17/opinion/ai-employment-bias-nyc.html>

¹⁸ Hiring and AI: Let job candidates know why they were rejected, Julia Stoyanovich, The Wall Street Journal Reports:

Leadership, September 22, 2021, available at <https://www.wsj.com/articles/hiring-job-candidates-ai-11632244313>

¹⁹ “We are AI” series by NYU Tandon Center for Responsible AI and Queens Public Library helps citizens take control of tech, March 14 2022, available at <https://engineering.nyu.edu/news/we-are-ai-series-nyu-andon-center-responsible-ai-queens-public-library>

²⁰ “We are AI: Taking control of technology”, NYU Center for Responsible AI, available <https://dataresponsibly.github.io/we-are-ai/>

3. **An informed public:** To be truly effective, this law requires an informed public. I recommend that New York City invests resources into informing members of the public about data, algorithms, and automated decision making, using hiring ADS as a concrete and important example.

In what follows, I will give some background on automated hiring systems, and will then expand on each of my recommendations.

Automated hiring systems

Since the 1990s, and increasingly so in the last decade, commercial tools are being used by companies large and small to hire more efficiently: source and screen candidates faster and with less paperwork, and successfully select candidates who will perform well on the job. These tools are also meant to improve efficiency for the job applicants, matching them with relevant positions, allowing them to apply with a click of a button, and facilitating the interview process.

In their 2018 report, Bogen and Rieke²¹ describe the hiring process from the point of view of an employer as a series of decisions that form a funnel: “Employers start by *sourcing* candidates, attracting potential candidates to apply for open positions through advertisements, job postings, and individual outreach. Next, during the *screening* stage, employers assess candidates—both before and after those candidates apply—by analyzing their experience, skills, and characteristics. Through *interviewing* applicants, employers continue their assessment in a more direct, individualized fashion. During the *selection* step, employers make final hiring and compensation determinations.” Importantly, while a comprehensive survey of the space lacks, we have reason to believe that automated hiring tools are in broad use in all stages of the hiring process.

Despite their potential to improve efficiency for both employers and job applicants, hiring ADS are also raising concerns. I will recount two well-known examples here.

Sourcing: One of the earliest indications that there is cause for concern came in 2015, with the results of the AdFisher study out of Carnegie Mellon University²² that was broadly circulated by the press²³. Researchers ran an experiment, in which they created two sets of synthetic profiles of Web users who were the same in every respect — in terms of their demographics, stated interests, and browsing patterns — with a single exception: their stated gender, male or female. In one experiment, the AdFisher tool stimulated an interest in jobs in both groups, and showed that Google displays ads for a career

²¹ Bogen and Rieke, “*Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias*”, Upturn, (2018) <https://www.upturn.org/static/reports/2018/hiring-algorithms/files/Upturn%20-%20Help%20Wanted%20-%20An%20Exploration%20of%20Hiring%20Algorithms,%20Equity%20and%20Bias.pdf>

²² Datta, Tschantz, Datta, “*Automated experiments on ad privacy settings*”, Proceedings of Privacy Enhancing Technology (2015) <https://content.sciendo.com/view/journals/popets/2015/1/article-p92.xml>

²³ Gibbs, “*Women less likely to be shown ads for high-paid jobs on Google, study shows*”, The Guardian (2015) <https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study>

coaching service for high-paying executive jobs far more frequently to the male group (1,852 times) than to the female group (318 times). This brings back memories of the time when it was legal to advertise jobs by gender in newspapers. This practice was outlawed in the US 1964, but it persists in the online ad environment.

Screening: In late 2018 it was reported that Amazon’s AI resume screening tool, developed with the stated goal of increasing workforce diversity, in fact did the opposite thing: the system taught itself that male candidates were preferable to female candidates.²⁴ It penalized resumes that included the word “women’s,” as in “women’s chess club captain,” and downgraded graduates of two all-women’s colleges. These results aligned with, and reinforced, a stark gender imbalance in the workforce at Amazon and other platforms, particularly when it comes to technical roles.

Numerous other cases of discrimination based on gender, race, and disability status during screening, interviewing, and selection stages have been documented in recent reports^{25,26}. These and other examples show that, if left unchecked, automated hiring tools will replicate, amplify, and normalize results of historical discrimination.

Recommendation 1: Expanding the scope of auditing

Bias audits should take a broader view, going beyond disparate impact when considering fairness of outcomes. Others surely spoke to this point, and I will not dwell on it here. Instead, I will focus on another important dimension of due process that is closely linked to discrimination — substantiating the use of particular features in decision-making.

Regarding the use of predictive analytics to screen candidates, Jenny Yang states: “Algorithmic screens do not fit neatly within our existing laws because algorithmic models aim to identify statistical relationships among variables in the data whether or not they are understood or job related.[...] Although algorithms can uncover job-related characteristics with strong predictive power, they can also identify correlations arising from statistical noise or undetected bias in the training data. Many of these models do not attempt to establish cause-and-effect relationships, creating a risk that employers may hire based on arbitrary and potentially biased correlations.”²⁷

²⁴ [Dastin, “Amazon scraps secret AI recruiting tool that showed bias against women”, Reuters \(2018\)](https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G)

<https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>

²⁵ [Emerging Technology from the arXiv, “Racism is Poisoning Online Ad Delivery, Says Harvard Professor”, MIT Technology Review \(2013\) https://www.technologyreview.com/s/510646/racism-is-poisoning-online-ad-delivery-says-harvard-professor/](https://www.technologyreview.com/s/510646/racism-is-poisoning-online-ad-delivery-says-harvard-professor/)

²⁶ [Stains, “Are Workplace Personality Tests Fair?”, Wall Street Journal \(2014\) http://www.wsj.com/articles/are-workplace-personality-tests-fair-1412044257](http://www.wsj.com/articles/are-workplace-personality-tests-fair-1412044257)

²⁷ [Yang, “Ensuring a Future that Advances Equity in Algorithmic Employment Decisions”, Urban Institute \(2020\) https://www.urban.org/research/publication/ensuring-future-advances-equity-algorithmic-employment-decisions](https://www.urban.org/research/publication/ensuring-future-advances-equity-algorithmic-employment-decisions)

In other words, identifying what features are impacting a decision is important, but it is insufficient to alleviate due process and discrimination concerns. I recommend that an audit of an automated hiring tool should also include information about the job relevance of these features.

A subtle but important point is that even features that can legitimately be used for hiring may capture information differently for different population groups. For example, it has been documented that the mean score of the math section of the SAT (Scholastic Assessment Test) differs across racial groups, as does the shape of the score distribution.²⁸ These disparities are often attributed to racial and class inequalities encountered early in life, and are thought to present persistent obstacles to upward mobility and opportunity.

Some automated hiring tools used today claim to predict job performance by analyzing an interview video for body language and speech patterns. Arvind Narayanan refers to tools of this kind as “fundamentally dubious” and places them in the category of AI snake oil.²⁹ The premise of such tools, that (a) it is possible to predict social outcomes based on a person's appearance or demeanor and (b) it is ethically defensible to try, reeks of scientific racism and is at best an elaborate random number generator.

The AI snake oil example brings up a related point: that an audit should also evaluate the effectiveness of the tool. Does the tool work? Is it able to identify promising job candidates better than a random coin flip? What were the specific criteria for the evaluation, and what evaluation methodology was used? Was the tool's performance evaluated on a population with demographic and other characteristics that are similar to the New York City population on which it will be used? Without information about the statistical properties of the population on which the tool was trained (in the case of machine learning) and validated, we cannot know whether the tool will have similar performance when deployed.³⁰

In my own work, I recently evaluated the validity of two algorithmic personality tests that are used by employers for pre-employment assessment³¹. This work was done by a large interdisciplinary team that included several data scientists, a sociologist, an industrial-organizational (I-O) psychologist, and an investigative journalist. My colleagues and I developed a methodology for an external audit of stability of algorithmic personality tests, and used it to audit two systems, Humantic AI and Crystal. Importantly, rather than challenging or affirming the assumptions made in psychometric testing — that personality traits are meaningful and measurable constructs, and that they are indicative of future success on the job— we framed our methodology around testing the underlying assumptions made by the vendors of the algorithmic personality tests themselves.

²⁸ [Reeves and Halikias “Race gaps in SAT scores highlight inequality and hinder upward mobility”, Brookings \(2017\)](https://www.brookings.edu/research/race-gaps-in-sat-scores-highlight-inequality-and-hinder-upward-mobility)

<https://www.brookings.edu/research/race-gaps-in-sat-scores-highlight-inequality-and-hinder-upward-mobility>

²⁹ [Narayanan, “How to recognize AI snakeoil” \(2019\)](https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf)

<https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf>

³⁰ [Stoyanovich and Howe, “Follow the data: Algorithmic transparency starts with data transparency”](https://ai.shorensteincenter.org/ideas/2018/11/26/follow-the-data-algorithmic-transparency-starts-with-data-transparency)

(2019) <https://ai.shorensteincenter.org/ideas/2018/11/26/follow-the-data-algorithmic-transparency-starts-with-data-transparency>

³¹ [An external stability audit of framework to test the validity of personality prediction in AI hiring, Rhea et al., 2022, available at https://arxiv.org/abs/2201.09151](https://arxiv.org/abs/2201.09151)

In our audits of Humantic AI and Crystal, we found that both systems show substantial instability on key facets of measurement, and so cannot be considered valid testing instruments. For example, Crystal frequently computes different personality scores if the same resume is given in PDF vs. in raw text, while Humantic AI gives different personality scores on a LinkedIn profile vs. a resume of the same job seeker. This violated the assumption that the output of a personality test is stable across job-irrelevant input variations. Among other notable findings is evidence of persistent — and often incorrect — data linkage by Humantic AI. A summary of our results are presented in **Table 1**.

Facet	Crystal	Humantic
Resume file format	X	✓
LinkedIn URL in resume	?	X
Source context	X	X
Algorithm-time / immediate	✓	✓
Algorithm-time / 31 days	✓	X
Participant-time / LinkedIn	X	X
Participant-time / Twitter	N/A	✓

Table 1: Summary of stability results for Crystal and Humantic AI, with respect to facets of measurement: ✓ indicates sufficient rank-order stability in all traits, while X indicates insufficient rank-order stability or significant locational instability in at least one trait, and N/A indicates the facet was not tested in our audit. Results are detailed in <https://arxiv.org/abs/2201.09151>.

In summary, I recommend that the scope of auditing for bias should be expanded beyond disparate impact to include other dimensions of discrimination, and also contain information about a tool’s effectiveness. To support compliance and enable a comparison between tools during procurement, these audits should be based on a set of uniform criteria. To enable public input and deliberation, these criteria should be made publicly available.

Recommendation 2: Explaining decisions to the job applicant

Information about job qualifications or characteristics that the tool uses for screening should be provided in a manner that allows the job applicant to understand, and, if necessary, correct and contest the information. As I argued in Recommendation 1, it is also important to disclose why these specific qualifications and characteristics are considered job relevant.

I recommend that explanations for job seekers are built around the popular nutritional label metaphor, drawing an analogy to the food industry, where simple, standardized labels convey information about the ingredients and production processes.²⁰

²⁰ Stoyanovich and Howe, “Nutritional labels for data and models”, IEEE Data Engineering Bulletin 42(3):

An applicant-facing nutritional label for an automated hiring system should be comprehensible: short, simple, and clear. It should be consultative, providing actionable information. Based on such information, a job applicant may, for example, take a certification exam to improve their chances of being hired for this or similar position in the future. Labels should also be comparable: allowing a job applicant to easily compare their standing across vendors and positions, and thus implying a standard.

Nutritional labels are a promising metaphor for other types of disclosure, and can be used to represent the process or the result of an automated hiring system for auditors, technologists, or employers.³²

ACCOUNTANT	
Acme Partners	
Qualifications:	BS in accounting, GPA >3.0, Knowledge of financial and accounting systems and applications
Personal data to be analyzed:	An AI program could be used to review and analyze the applicant's personal data online, including LinkedIn profile, social media accounts and credit score.
Additional assessment:	AI-assisted personality scoring
ALERT: Applicants for this position DO NOT have the option to selectively decline use of AI analysis for any of their personal data or to review and challenge the results of such analysis.	

Figure 1: A posting label is a short, simple, and clear summary of the screening process. This label is presented to a job seeker before they apply, supporting informed consent, allowing them to opt out of components of the process or to request accommodations.

Figure 1 shows a posting label, a short and clear summary of the screening process. This label is presented to a job seeker before they apply, supporting informed consent, allowing them to opt out of components of the process or to request accommodations. Giving job seekers an opportunity to request accommodations is particularly important in light of the recent guidance by the Equal Employment Opportunity Commission (EEOC) on the Americans with Disabilities Act and the use of AI to assess job applicants and employees ³³.

³² [Stoyanovich, Howe, Jagadish, "Responsible Data Management", PVLDB 13\(12\): 3474-3489 \(2020\)](https://dataresponsibly.github.io/documents/mirror.pdf)
<https://dataresponsibly.github.io/documents/mirror.pdf>

³³ [The Americans with Disabilities Act and the use of software, algorithms, and AI to assess job applicants and employees, US Equal Employment Opportunity Commission, 2022,](https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificialintelligence)
<https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificialintelligence>

If a job seeker applies for the job but isn't selected, then he or she would receive a "decision label" along with the decision. This label would show how the applicant's qualifications measured up to the job requirements; how the applicant compared with other job seekers; and how information about these qualifications was extracted.

Recommendation 3: Creating an informed public

My final recommendation will be brief. To be truly effective, this law requires an informed public. Individual job applicants should be able to understand and act on the information disclosed to them. In Recommendation 1, I spoke about the need to make auditing criteria for fairness and effectiveness publicly available. Empowering members of the public to weigh in on these standards will strengthen the accountability structures and help build public trust in the use of ADS in hiring and beyond. In Recommendation 2, I spoke about nutritional labels as a disclosure method. We should help job seekers, and the public at large, to understand and act upon information about data and ADS.

I recommend that New York City invests resources into informing members of the public about data, algorithms, and automated decision making, using hiring ADS as a concrete and important example. I already started this work, having developed "We are AI", a free public education course on AI and its impacts in society. This course is accompanied by a comic book series, available in English and Spanish.

Conclusion

In conclusion, I would like to quote from the recently released position statement by IEEE-USA, titled "Artificial Intelligence: Accelerating Inclusive Innovation by Building Trust".³⁴ IEEE is the largest professional organization of engineers in the world; I have the pleasure of serving on their AI/AS (Artificial Intelligence and Autonomous Systems) Policy Committee.

"We now stand at an important juncture that pertains less to what new levels of efficiency AI/AS can enable, and more to whether these technologies can become a force for good in ways that go beyond efficiency. We have a critical opportunity to use AI/AS to help make society more equitable, inclusive, and just; make government operations more transparent and accountable; and encourage public participation and increase the public's trust in government. When used according to these objectives, AI/AS can help reaffirm our democratic values.

If, instead, we miss the opportunity to use these technologies to further human values and ensure trustworthiness, and uphold the status quo, we risk reinforcing disparities in access to goods and services, discouraging public participation in civic life, and eroding the public's trust in government. Put another way: Responsible development and use of AI/AS to further human

³⁴ IEEE-USA, "Artificial Intelligence: Accelerating Inclusive Innovation by Building Trust" (2020) <https://ieeusa.org/wp-content/uploads/2020/10/AITrust0720.pdf>

values and ensure trustworthiness is the only kind that can lead to a sustainable ecosystem of innovation. It is the only kind that our society will tolerate.”

We Need Laws to Take On Racism and Sexism in Hiring Technology

Julia Stoyanovich
jds2109@gmail.com

Sign in to The New York Times with ✕




Profile 1


Profile 2

Julia Stoyanovich
jds405@nyu.edu

Artificial intelligence used to evaluate job candidates must not become a tool that exacerbates discrimination.

March 17, 2021

By Alexandra Reeve Givens, Hilke Schellmann and Julia Stoyanovich

Ms. Givens is the chief executive of the Center for Democracy & Technology. Ms. Schellman and Dr. Stoyanovich are professors at New York University focusing on artificial intelligence.

American democracy depends on everyone having equal access to work. But in reality, people of color, women, those with disabilities and other marginalized groups experience unemployment or underemployment at disproportionately high rates, especially amid the economic fallout of the Covid-19 pandemic. Now the use of artificial intelligence technology for hiring may exacerbate those problems and further bake bias into the hiring process.

At the moment, the New York City Council is debating a proposed new law that would regulate automated tools used to evaluate job candidates and employees. If done right, the law could make a real difference in the city and have wide influence nationally: In the absence of federal regulation, states and cities have used models from other localities to regulate emerging technologies.

Over the past few years, an increasing number of employers have started using artificial intelligence and other automated tools to speed up hiring, save money and screen job applicants without in-person interaction. These are all features that are increasingly attractive during the pandemic. These technologies include screeners that scan résumés for key words, games that claim to assess attributes such as generosity and appetite for risk, and even emotion analyzers that claim to read facial and vocal cues to predict if candidates will be engaged and team players.

In most cases, vendors train these tools to analyze workers who are deemed successful by their employer and to measure whether job applicants have similar traits. This approach can worsen underrepresentation and social divides if, for example, Latino men or Black women are inadequately represented in the pool of employees. In another case, a résumé-screening tool could identify Ivy League schools on successful employees' résumés and then downgrade résumés from historically Black or women's colleges.

In its current form, the council's bill would require vendors that sell automated assessment tools to audit them for bias and discrimination, checking whether, for example, a tool selects male candidates at a higher rate than female candidates. It would also require vendors to tell job applicants the characteristics the test claims to measure. This approach could be helpful: It would shed light on how job applicants are screened and force vendors to think critically about potential discriminatory effects. But for the law to have teeth, we recommend several important additional protections.

The measure must require companies to publicly disclose what they find when they audit their tech for bias. Despite pressure to limit its scope, the City Council must ensure that the bill would address discrimination in all forms — on the basis of not only race or gender but also disability, sexual orientation and other protected characteristics.

These audits should consider the circumstances of people who are multiply marginalized — for example, Black women, who may be discriminated against because they are both Black and women. Bias audits conducted by companies typically don't do this.

The bill should also require validity testing, to ensure that the tools actually measure what they claim to, and it must make certain that they measure characteristics that are relevant for the job. Such testing would interrogate whether, for example, candidates' efforts to blow up a balloon in an online game really indicate their appetite for risk in the real world — and whether risk-taking is necessary for the job. Mandatory validity testing would also eliminate bad actors whose hiring tools do arbitrary things like assess job applicants' personalities differently based on subtle changes in the background of their video interviews.



In addition, the City Council must require vendors to tell candidates how they will be screened by an automated tool before the screening, so candidates know what to expect. People who are blind, for example, may not suspect that their video interview could score poorly if they fail to make eye contact with the camera. If they know what is being tested, they can engage with the employer to seek a fairer test. The proposed legislation currently before the City Council would require companies to alert candidates within 30 days if they have been evaluated using A.I., but only after they have taken the test. Finally, the bill must cover not only the sale of automated hiring tools in New York City but also their use. Without that stipulation, hiringtool vendors could escape the obligations of this bill by simply locating sales outside the city. The council should close this loophole.


With this bill, the city has the chance to combat new forms of employment discrimination and get closer to the ideal of what America stands for: making access to opportunity more equitable for all. Unemployed New Yorkers are watching.

Alexandra Reeve Givens is the chief executive of the Center for Democracy & Technology. Hilke Schellmann is a reporter investigating artificial intelligence and an assistant professor of journalism at New York University. Julia Stoyanovich is an assistant professor of computer science and engineering and of data science and is the director of the Center for Responsible AI at New York University.


The Times is committed to publishing a diversity of letters to the editor. We'd like to hear what you think about this or any of our articles. Here are some tips. And here's our email: letters@nytimes.com.

Follow The New York Times Opinion section on [Facebook](#), [Twitter \(@NYTopinion\)](#) and [Instagram](#).

Google



jds2109@gmail.com

Julia Stoyanovich jds405@nyu.edu

JOURNAL REPORTS: LEADERSHIP

Hiring and AI: Let Job Candidates Know Why They Were Rejected

As more companies use artificial intelligence in their hiring decisions, here's one way to make the system more transparent



Labels that explain a hiring process that uses AI could allow job seekers to opt out if they object to the employer's data practices.

PHOTO: ISTOCKPHOTO GETTY IMAGES

By Julia Stoyanovich

Updated Sept. 22, 2021 11 00 am ET

Artificial-intelligence tools are seeing ever broader use in hiring. But this practice is also hotly criticized because we rarely understand how these tools select candidates, and whether the candidates they select are, in fact, better qualified than those who are rejected.

To help answer these crucial questions, we should give job seekers more information about the hiring process and the decisions. The solution I propose is a twist on something we see every day: nutritional labels. Specifically, job candidates would see simple, standardized labels that show the factors that go into the AI's decision.

How would this work? When people apply for a job, they will see a list of the hiring criteria, such as degree requirements, specific skills and the number of years of experience, so that they know

precisely what a company is looking for. Then, if the applicant is rejected, the AI will present them with another list, showing where they didn't meet the criteria or compared unfavorably to other applicants—the reasoning behind the decision.

In other words, we should show people very clearly what factors are used to judge them, just as we show people the ingredients that go into their food.

We desperately need such a system. AI's widespread use in hiring far outpaces our collective ability to keep it in check—to understand, verify and oversee it. Is a résumé screener identifying promising candidates, or is it picking up irrelevant, or even discriminatory, patterns from historical data? Is a job seeker participating in a fair competition if he or she is unable to pass an online personality test, despite having other qualifications needed for the job?

[A two-tier system](#)

The labels I propose would come in two parts. First, the “posting label,” a short, simple and clear set of requirements that an AI screener will be looking for in applicants. For example, the posting label for an art-director position might list “B.S. in communications or similar,” “two years of full-time experience” and “expert knowledge of Adobe Design Suite.”

The posting label also would explain the assessment process. Will the AI consider only submitted résumés, or also use applicants' public LinkedIn profiles and Twitter feeds? What about their credit histories? Will a video interview be required? Which parts of the application are processed by a machine and which by a human?

[Disclosing AI in Hiring](#)

The author proposes a standardized method akin to nutritional labels that employers could use to inform job candidates when artificial intelligence programs will play a role in their evaluations. One example of what such a form might look like:

ACCOUNTANT

Acme Partners

Qualifications: BS in accounting, GPA >3.0, Knowledge of financial and accounting systems and applications

Personal data to be analyzed: An AI program could be used to review and analyze the applicant's personal data online, including LinkedIn profile, social media accounts and credit score.

Additional assessment: AI-assisted personality scoring

ALERT: Applicants for this position DO NOT have the option to selectively decline use of AI analysis for any of their personal data or to review and challenge the results of such analysis.

Source: Julia Stoyanovich

A posting label for an accountant position, for instance, may list résumé, LinkedIn, Twitter and credit scores as the sources of information, and it may state that personality scores will be used to assess the candidate, with preference given to candidates with higher S (“steady”) and C (“conscientious”) scores.

The label would also provide actionable information. It would state, for instance, that a job applicant is allowed to correct some data that the company uses to make decisions or contest the company's use of their personal information, such as their social-media feed.

In addition, applicants should be informed that they can request accommodations if they have reason to believe that a certain kind of assessment would discriminate against them. For example, scoring a video interview based in part on making eye contact with the camera would disadvantage people with limited vision or autism.

Critically, the posting label enables informed consent: Job seekers agree to the assessment procedure by submitting their applications, and they opt out by deciding not to apply if they object to the employer's data practices (e.g., using an applicant's credit score) or assessment

methodology (e.g., constructing an estimation of an applicant’s personality based on a résumé or performance in an online game).

If a job seeker applies for the job but isn’t selected, then he or she would receive a “decision label” along with the rejection. This label would show how the applicant’s qualifications measured up to the job requirements; how the applicant compared with other job seekers; and how information about these qualifications was extracted.

For example, a portion of the applicant’s résumé may be highlighted to explain that he or she lacked sufficient experience for the position. Or a tweet may be highlighted to substantiate a low “conscientious” score on a personality test. This information would allow the applicant to accept or contest the hiring decision.

Explaining the choice

Having clear criteria for decisions not only helps applicants—it also gives employers vital information.

Many times, AI makes judgment calls that are opaque. Employers often don’t know what data AI screeners are using, or how they analyze that data to make a final decision. The labels can show managers the factors that the AI is using to screen applicants—and let those managers decide if those factors need to be changed.

SHARE YOUR THOUGHTS

What do you think are the advantages and drawbacks of using AI in hiring? Join the conversation below.

For instance, does the AI need to be given more—or different—training data, covering different job roles and demographic groups, to avoid making biased and arbitrary decisions? Likewise, what is the predictive accuracy of the tool for different demographic groups? What features of past applicants’ profiles led to a positive or a negative decision by the tool, and can job relevance of these features be substantiated?

One concern may be that these labels will motivate strategic manipulation or “gaming.” However, there is already strategic manipulation happening: Career services routinely offer training and advice on how to make a résumé attractive to algorithmic tools. Greater transparency will help

alleviate unproductive gaming and tilt the balance in favor of positive change, motivating individuals to actually improve their qualifications, rather than to make it seem like they are qualified.

Humans—and not AI—should ultimately make the final call on whom to hire. But, like it or not, many managers use AI systems at different stages of the hiring process—and that practice is only going to become more common. If managers are relying on AI, those tools should be as transparent as possible, and job seekers should have a say in how their data is used.

Ultimately, hiring is complex. It is a multistep process in which we trade off objective criteria, such as an applicant's degree requirements and measurable skills, against subjective factors such as how well they will fit into the team and pick up new skills.

We bring in AI to help alleviate some of this complexity. But we cannot forget that AI tools work to specification, and they do best when those specifications are clear. We can use AI effectively for parts of the hiring process—to identify clear requirements-based matches. But AI tools cannot exercise discretion or apply subjective judgment. My hope is that nutritional labels will help us come to a consensus on which decisions we should leave to an AI, and which we should make ourselves.

Dr. Stoyanovich is institute associate professor of computer science and engineering at the Tandon School of Engineering, associate professor of data science at the Center for Data Science and director of New York University's Center for Responsible AI. She can be reached at reports@wsj.com.

Appeared in the September 23, 2021, print edition as 'What's In AI's Hiring Black Box.'

Copyright © 2021 Dow Jones & Company, Inc. All Rights Reserved

This copy is for your personal, non-commercial use only. To order presentation-ready copies for distribution to your colleagues, clients or customers visit <https://www.djreprints.com>.

WE ARE AI

#4

All about that

BIA



Terms of Use

All the panels in this comic book are licensed CC BY-NC-ND 4.0. Please refer to the license page for details on how you can use this artwork

T;D : Feel free to use panels/groups of panels in your presentations/articles, as long as you

1. Provide the proper citation
2. Do not make modifications to the individual panels themselves

Cite as:

Julia Stoyanovich and Falaah Arif Khan about that Bias

Wear AI Comis, Vol 4 (2021)

https://dataresponsibly.github.io/we-are-ai/comis/vol4_a.pdf

Contact:

Please direct any queries about using elements from this comic to themachinelearnist@gmail.com or jstoyanovich@nyu.edu



License CC BY-NC-ND 4.0

LET'S TALK ABOUT WHAT WE MEAN BY 'BIAS' IN AI, AND HOW IT ARISES.

WE SAY THAT AN AI IS BIASED IF ITS USE CAN LEAD TO SYSTEMATIC AND UNFAIR DISCRIMINATION AGAINST SOME INDIVIDUALS OR GROUPS IN FAVOR OF OTHERS.

BIAS CAN STEM FROM HARMFUL PATTERNS PICKED UP FROM THE DATA ITSELF,

OR FROM HOW THE ALGORITHM IS DESIGNED,

OR FROM THE OBJECTIVES THAT WE SPECIFIED FOR IT,

OR FROM HOW WE USE IT.



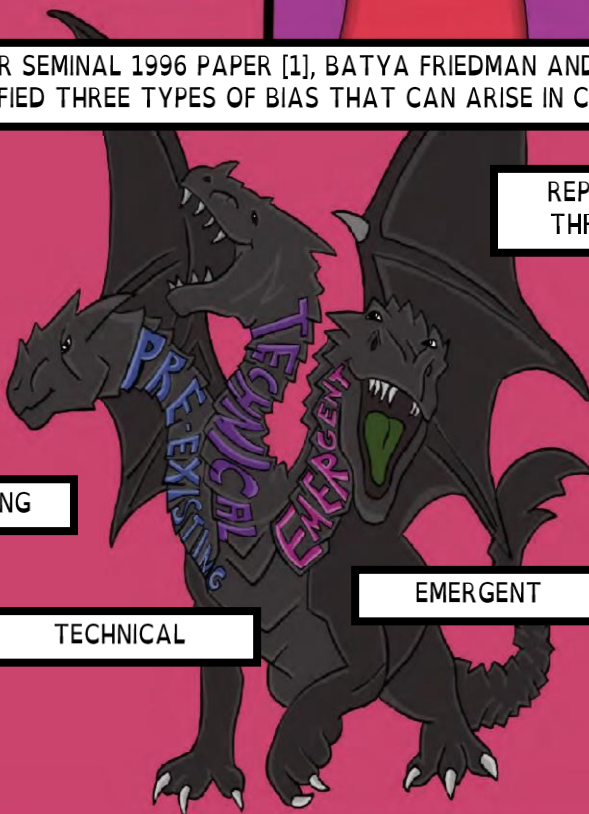
IN THEIR SEMINAL 1996 PAPER [1], BATYA FRIEDMAN AND HELEN NISSENBAUM IDENTIFIED THREE TYPES OF BIAS THAT CAN ARISE IN COMPUTER SYSTEMS,

REPRESENTED HERE AS A THREE-HEADED DRAGON:

PRE-EXISTING

TECHNICAL

EMERGENT



[1] Batya Friedman and Helen Nissenbaum. (1996). Bias in computer systems

RECALL THE BAKING METAPHOR WE USED TO UNDERSTAND DATA-DRIVEN ALGORITHMS IN VOLUME 1.

LET'S NOW USE THE SAME METAPHOR TO UNDERSTAND BIAS!



PRE-EXISTING BIAS EXISTS INDEPENDENT OF THE ALGORITHM AND HAS ITS ORIGINS IN SOCIETY.

THESE WOULD BE THE FLAVOR NOTES THAT WILL SEEP INTO YOUR BREAD IF YOU DON'T PRIORITIZE THE PURITY/FRESHNESS OF YOUR INGREDIENTS,

PRE-EXISTING BIAS (IN THE DATA)

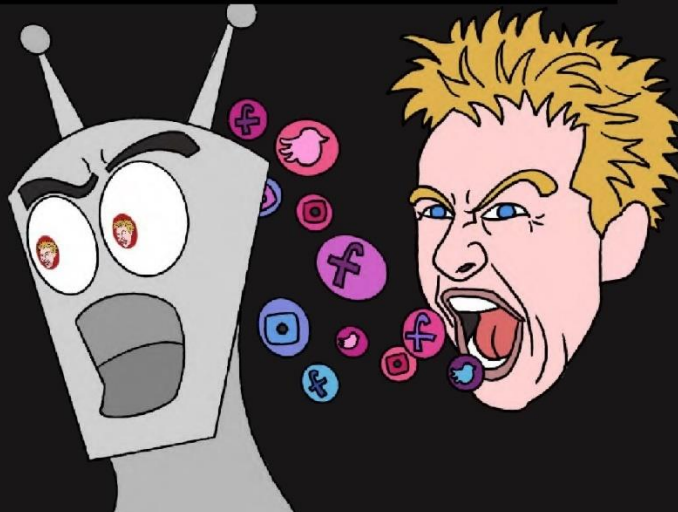
OR IF YOU DECIDE TO USE PREMIXED OFF-THE-SHELF BATTER.



THESE BIASES EXIST IN SOCIETY AND COME 'PRE-BAKED' INTO THE ALGORITHM,

FROM THE UNDERLYING DISCRIMINATORY SYSTEM THAT THE DATA WAS COLLECTED FROM -

SUCH AS THE GENDER AND RACIAL STEREOTYPES THAT LANGUAGE MODELS PICK UP WHEN TRAINED ON DATA FROM SOCIAL MEDIA.



**TECHNICAL
BIAS**

TECHNICAL BIAS IS INTRODUCED BY THE SYSTEM ITSELF -
BECAUSE OF THE WAY IT IS DESIGNED OR OPERATES.

THESE WOULD BE THE IMPERFECTIONS THAT
WILL SEEP INTO YOUR BREAD IF YOU USE THE
WRONG EQUIPMENT -



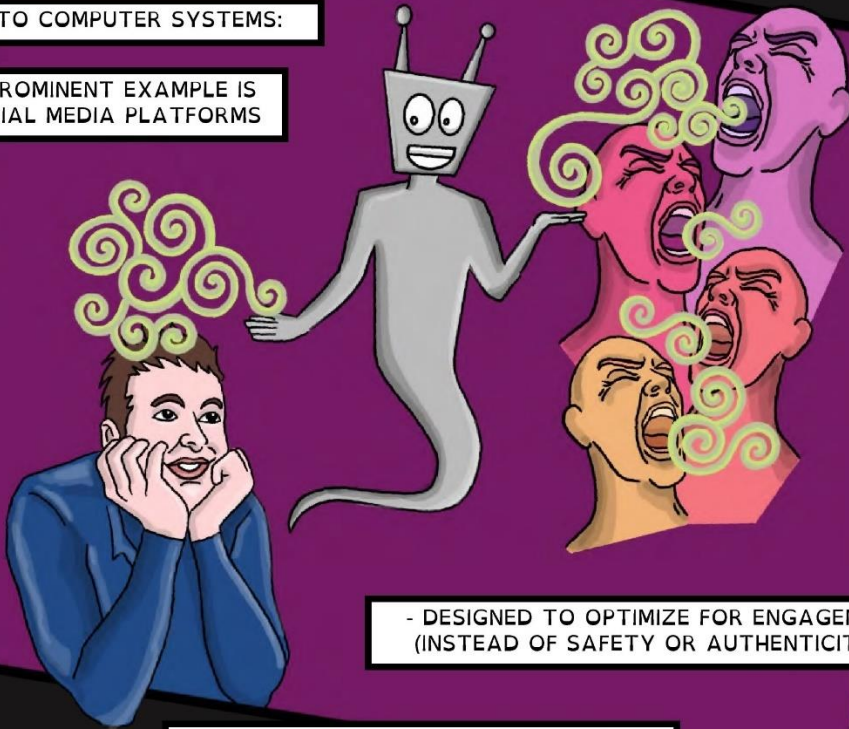
SUCH AS UNEVEN COOKING OF YOUR
CUPCAKES IF YOUR OVEN TEMPERATURE
IS MISCALIBRATED,



OR SPILLAGE OF BATTER IF YOUR BAKING
EQUIPMENT IS OF THE WRONG SIZE.

BACK TO COMPUTER SYSTEMS:

A PROMINENT EXAMPLE IS
SOCIAL MEDIA PLATFORMS



- DESIGNED TO OPTIMIZE FOR ENGAGEMENT
(INSTEAD OF SAFETY OR AUTHENTICITY) -

THAT END UP PROMOTING POLARIZING
ARTICLES AND FAKE NEWS.

**EMERGENT BIAS
(DUE TO DECISIONS)**

EMERGENT BIAS ARISES OVER TIME, BECAUSE THE DECISIONS MADE WITH THE HELP OF THE SYSTEM CHANGE THE WORLD,

WHICH IN TURN IMPACTS THE OPERATION OF THE SYSTEM GOING FORWARD.

THINK ABOUT BEHAVIORAL CHANGES THAT WILL EMERGE AS A RESULT OF YOUR BAKING -

WHAT IF YOU BECOME SUCH A MAESTRO AT BAKING THAT YOU INADVERTENTLY MAKE BREAD A STEADY PART OF YOUR DIET!



OR MAKE IT SO OFTEN, THAT YOU TURN EVERYONE AROUND YOU OFF THE THOUGHT OF EVER EATING ANOTHER SLICE!



OR THINK ABOUT HOW YOUR IDEA OF 'WHAT BREAD SHOULD TASTE LIKE' IS SHAPED BY THE POPULARITY OF PRODUCTS LIKE 'WONDER BREAD'.



IN THE SAME VEIN, THINK ABOUT HOW YOUR EXPOSURE TO NEWS - AND INFORMATION MORE BROADLY -

IS SHAPED BY ALGORITHMS THAT CURATE SOCIAL FEEDS WITH POPULAR AND 'TRENDING' POSTS.



TO MAKE OUR DISCUSSION CONCRETE, LET'S LOOK AT REAL-WORLD EXAMPLES OF ALGORITHMIC BIAS.

LET'S TAKE 'HIRING' AS A REPRESENTATIVE DOMAIN IN WHICH ALGORITHMS ARE INCREASINGLY BEING USED TO MAKE CRITICAL DECISIONS MORE 'EFFICIENTLY'.



ONE OF THE EARLIEST INDICATIONS THAT THERE IS CAUSE FOR CONCERN CAME IN 2015, WITH THE RESULTS OF THE ADFISHER STUDY OUT OF CARNEGIE MELLON UNIVERSITY. [2]

RESEARCHERS RAN AN EXPERIMENT, IN WHICH THEY CREATED TWO SETS OF SYNTHETIC PROFILES OF WEB USERS WHO WERE THE SAME IN EVERY RESPECT

— IN TERMS OF THEIR DEMOGRAPHICS, STATED INTERESTS, AND BROWSING PATTERNS —

WITH A SINGLE EXCEPTION: THEIR STATED GENDER, MALE OR FEMALE.

RESEARCHERS SHOWED THAT GOOGLE DISPLAYED ADS FOR A CAREER COACHING SERVICE FOR HIGH-PAYING EXECUTIVE JOBS FAR MORE FREQUENTLY TO THE MALE GROUP THAN TO THE FEMALE GROUP.

THIS BRINGS BACK MEMORIES OF THE TIME WHEN IT WAS LEGAL TO ADVERTISE JOBS BY GENDER IN NEWSPAPERS. THIS PRACTICE WAS OUTLAWED IN THE US IN 1964, BUT IT PERSISTS IN THE ONLINE AD ENVIRONMENT.

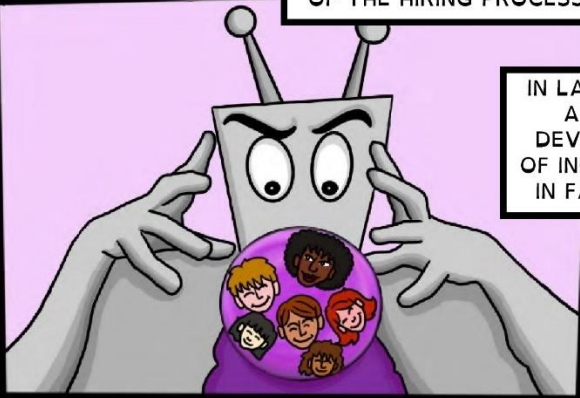
IT WAS LATER SHOWN THAT PART OF THE REASON THIS WAS HAPPENING IS THE MECHANICS OF THE ADVERTISEMENT TARGETING SYSTEM ITSELF, AS AN ARTIFACT OF THE BIDDING PROCESS.

THIS IS TECHNICAL BIAS IN ACTION!



[2] Women less likely to be shown ads for high-paid jobs on Google, study shows. Guardian (2015)

LET US MOVE FORWARD TO THE NEXT STAGE OF THE HIRING PROCESS: RESUME SCREENING.



IN LATE 2018 IT WAS REPORTED THAT AMAZON'S AI RECRUITING TOOL, DEVELOPED WITH THE STATED GOAL OF INCREASING WORKFORCE DIVERSITY, IN FACT DID THE OPPOSITE THING: [3]

THE SYSTEM TAUGHT ITSELF THAT MALE CANDIDATES WERE PREFERABLE TO FEMALE CANDIDATES.

IT PENALIZED RESUMES THAT INCLUDED THE WORD "WOMEN'S," AS IN "WOMEN'S CHESS CLUB CAPTAIN."

AND IT DOWNGRADED GRADUATES OF TWO ALL-WOMEN'S COLLEGES.

THE RESULTS ALIGNED WITH, AND REINFORCED, A STARK GENDER IMBALANCE IN THE WORKFORCE.

THIS IS EMERGENT BIAS IN ACTION -

A HIRING MANAGER TO WHOM AN AI TOOL REPEATEDLY SUGGEST THE SAME KIND OF JOB APPLICANT AS A GOOD FIT,

WILL OVERTIME COME TO BELIEVE THAT THIS IS WHAT A PROMISING EMPLOYEE LOOKS LIKE.

WE ARE ALSO SEEING PRE-EXISTING BIAS IN THIS EXAMPLE: THE AI TOOL WAS TRAINED ON HISTORICAL DATA ABOUT PAST EMPLOYEES, WHO WERE PREDOMINANTLY MALE



[3] Amazon scraps secret AI recruiting tool that showed bias against women. Reuters (2018)

HERE'S ANOTHER EXAMPLE, LATER YET IN THE HIRING PROCESS, PERHAPS DURING A POST-INTERVIEW BACKGROUND CHECK BY A POTENTIAL EMPLOYER -

LATANYA SWEENEY, A COMPUTER SCIENCE PROFESSOR ON THE FACULTY AT HARVARD,

SHOWED THAT GOOGLING FOR AFRICAN-AMERICAN SOUNDING NAMES IS MORE LIKELY TO TRIGGER ADS SUGGESTIVE OF A CRIMINAL RECORD THAN GOOGLING FOR WHITE-SOUNDING NAMES,

EVEN CONTROLLING FOR WHETHER AN INDIVIDUAL IN FACT HAS A CRIMINAL RECORD! [4]



Pristen

THIS IS PRE-EXISTING BIAS AT PLAY -

MANIFESTING LONG-STANDING RACIAL PREJUDICES OF SOCIETY.



Latanya



[4] Racism is Poisoning Online Ad Delivery, Says Harvard Professor. MIT Technology Review (2013)

THE CASES PRESENTED HERE HAVE ONE THING IN COMMON: THEY SHOW THAT AI CAN REINFORCE AND EXACERBATE UNLAWFUL DISCRIMINATION AGAINST MINORITY AND HISTORICALLY DISADVANTAGED GROUPS.

OFTEN THIS IS CALLED OUT AS "BIAS IN AI".

SO, WHY ARE SOPHISTICATED SYSTEMS THAT AIM TO MAKE HIRING MORE EFFICIENT FAILING AT THIS, AND ARGUABLY MAKING THINGS WORSE?

OF COURSE, THE ISSUES OF BIAS IN EMPLOYMENT ARE NOT NEW. THEY EXHIBITED THEMSELVES IN THE ANALOG ERA AS WELL.

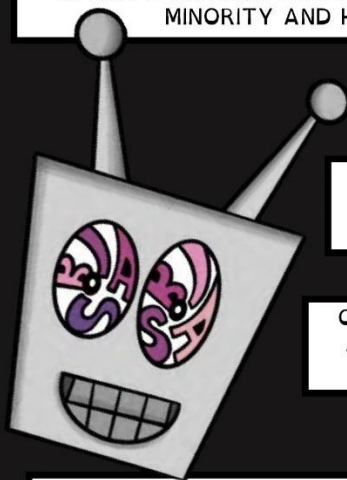
FOR EXAMPLE, IN THEIR WELL-KNOWN 2004 STUDY, MARIANNE BERTRAND AND SENDHIL MULLAINATHAN SENT FICTITIOUS RESUMES TO HELP-WANTED ADS IN BOSTON AND CHICAGO NEWSPAPERS. [5]



TO MANIPULATE PERCEIVED RACE, THEY RANDOMLY ASSIGNED AFRICAN-AMERICAN- OR WHITE-SOUNDING NAMES TO RESUMES.

WHITE NAMES RECEIVE 50 PERCENT MORE CALLBACKS FOR INTERVIEWS.

THIS CASE SHOWS THAT BIAS CAN BE DUE TO HUMAN DECISIONS.



[5] Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. Marianne Bertrand & Sendhil Mullainathan (2009)

LET'S REVISIT PRE-EXISTING BIAS THAT OFTEN EXHIBITS ITSELF IN THE DATA.

DATA IS AN IMAGE OF THE WORLD, ITS MIRROR REFLECTION.

WHEN WE THINK ABOUT BIAS IN THE DATA, WE INTERROGATE THIS REFLECTION.

ONE INTERPRETATION OF "BIAS IN THE DATA" IS THAT THE REFLECTION IS DISTORTED -

WE MAY SYSTEMATICALLY OVER-REPRESENT OR UNDER-REPRESENT PARTICULAR PARTS OF THE WORLD IN THE DATA,

OR OTHERWISE DISTORT THE READINGS.

RECALL THE FAILURE OF AMAZON'S RECRUITING AI TO IMPROVE WORKFORCE DIVERSITY.

THIS TOOL WAS TRAINED USING HISTORICAL DATA: RESUMES OF PEOPLE WHO WERE HIRED IN THE PAST.

THAT TRAINING WAS SUBJECT TO PRE-EXISTING BIAS.

IN THAT DATA, THERE WAS AN UNDER-REPRESENTATION OF WOMEN IN THE WORKFORCE, AND IN TECHNICAL ROLES.

A MORE SUBTLE POINT IS ABOUT DISTORTIONS.

WHEN WE CONSIDER FEATURES, LIKE AN INDIVIDUAL'S SCORE ON A STANDARDIZED TEST, DO WE TAKE THESE AT FACE VALUE?

OR DO WE ACCOUNT FOR DIFFERENCES IN ACCESS TO EDUCATIONAL OPPORTUNITY,

LIKE GOING TO A BETTER SCHOOL, OR HAVING ACCESS TO PAID TUTORING?

ANOTHER INTERPRETATION OF "BIAS IN THE DATA" IS THAT EVEN IF WE WERE ABLE TO REFLECT THE WORLD PERFECTLY IN THE DATA,

IT WOULD STILL BE A REFLECTION OF THE WORLD SUCH AS IT IS,



AND NOT NECESSARILY OF HOW IT COULD OR SHOULD BE.

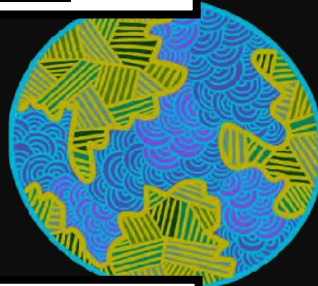
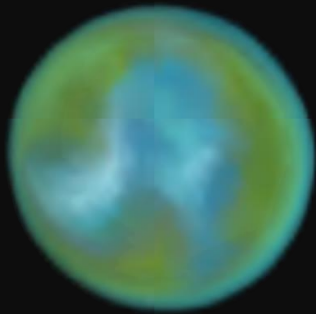
IT IS IMPORTANT TO KEEP IN MIND THAT A REFLECTION CANNOT KNOW WHETHER IT IS DISTORTED.



DATA ALONE CANNOT TELL US WHETHER IT IS A DISTORTED REFLECTION OF A PERFECT WORLD, A PERFECT REFLECTION OF A DISTORTED WORLD,

OR IF THESE DISTORTIONS COMPOUND.

THE SECOND POINT IS THAT IT IS NOT UP TO DATA OR ALGORITHMS, BUT RATHER UP TO PEOPLE



— INDIVIDUALS, GROUPS, AND SOCIETY AT LARGE —

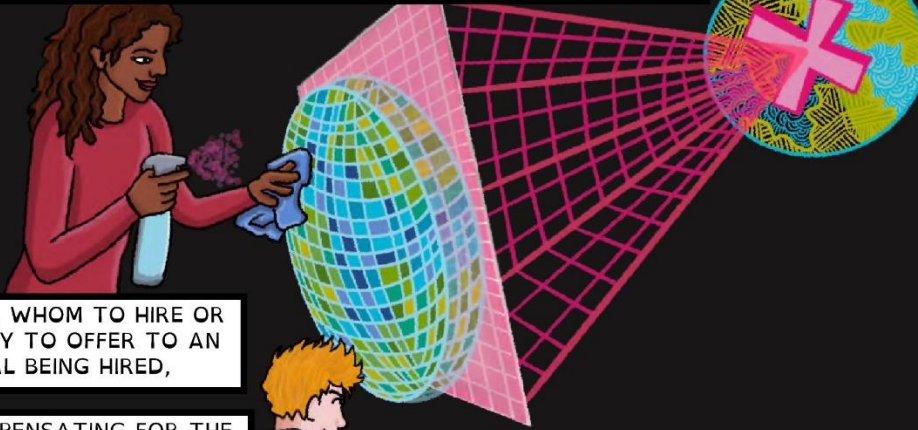
TO COME TO CONSENSUS ABOUT WHETHER THE WORLD IS HOW IT SHOULD BE, OR IF IT NEEDS TO BE IMPROVED.

AND, IF SO, HOW WE SHOULD GO ABOUT IMPROVING IT.



THE FINAL POINT HERE IS THAT CHANGING THE REFLECTION MAY NOT CHANGE THE WORLD.

IF THE REFLECTION ITSELF IS USED TO MAKE IMPORTANT DECISIONS -



FOR EXAMPLE, WHOM TO HIRE OR WHAT SALARY TO OFFER TO AN INDIVIDUAL BEING HIRED,

THEN COMPENSATING FOR THE DISTORTIONS IS WORTHWHILE.

BUT THE MIRROR METAPHOR ONLY TAKES US SO FAR.

WE HAVE TO WORK MUCH HARDER — USUALLY GOING FAR BEYOND TECHNOLOGICAL SOLUTIONS — TO MAKE LASTING CHANGE IN THE WORLD,

NOT MERELY BRUSH UP THE REFLECTION.

CIRCLING BACK NOW TO THE THREE-HEADED BIAS DRAGON.

WHEN SPEAKING ABOUT TACKLING BIAS IN AI, WE TEND TO FRAME THE PROBLEM AS FINDING A WAY TO SLAY THE BIAS-DRAGON.

BUT THROUGH OUR DISCUSSION OF THE LINK BETWEEN HUMAN BIAS AND MACHINE BIAS,

WE FIND OURSELVES QUESTIONING THE VERY NATURE OF THIS TALE -

AT THE END OF THE DAY, MAYBE THE QUESTION ISN'T -

HOW TO SLAY THE DRAGON AND RESCUE THE PRINCESS?

THE QUESTION WE REALLY SHOULD BE ASKING OURSELVES IS -

WHAT DO WE DO ABOUT A SOCIETY THAT LOCKS UP PRINCESSES IN CASTLES, IN THE FIRST PLACE?

FIN.



June 6, 2022

New York City Department of
Consumer and Worker Protections
42 Broadway
New York, NY 10004

Submitted electronically via <http://rules.cityofnewyork.us>

RE: Local Law 1894-A (Automated Employment Decision Tools)

To the Department:

Littler Mendelson P.C.’s Workplace Policy Institute (WPI) submits these comments to the Department of Consumer and Worker Protections pursuant to the Department’s Notice of Public Hearing and Opportunity to Comment on Proposed Rules. We thank the Department for the opportunity to comment on this important and timely topic, and address our comments specifically to Local Law 1894-1/144 of 2021 (herein, the “Law”).

By way of background, WPI facilitates the employer community’s engagement in legislative and regulatory developments that affect their workplaces and business strategies. WPI harnesses the deep subject matter expertise of Littler, the largest law firm in the world with a practice devoted exclusively to the representation of employers in employment and labor law matters. Littler’s clients range from new and emerging businesses to Fortune 100 companies throughout the country and around the world.¹ In today’s workplace, algorithms frequently make decisions that significantly impact people’s lives, across a wide range of activities: health care, housing, lending, and the focus of WPI’s comments, employment. We appreciate the opportunity to provide the Department with the benefit of our, and our clients’ experience.

The Use of Algorithms Employment Tools by Employers. For employers, the development of algorithmic decision making creates both opportunities and novel issues of concern, and they generate new questions about long-time problems. This in turn potentially affects every aspect of employment decision-making for employers of all size in virtually every industry, from the selection and hiring process, through performance management and promotion decisions, and up to and beyond the time termination decisions are made, whether for performance reasons or as part of a reorganization.

¹ Indeed, to help our clients navigate the complex issues surrounding artificial intelligence, automation, and the transformative effect these and other developments are having in the workplace, Littler recently announced its Global Workforce Transformation Initiative, to assist employers, employees, and policymakers face the challenges—and opportunities—that automation and technology bring to the workplace. The Initiative’s inaugural white paper, which analyses many of these issues in depth, may be found at: https://www.littler.com/files/workforce_transformation_report.pdf.

Littler Mendelson, PC

815 Connecticut Avenue NW | Suite 400 | Washington, DC 20006

James A. Paretti, Jr.

jparetti@littler.com

Today, employers can access more information about their applicant pool and workforce than ever before, and have an ability to correlate data gleaned from an application itself, perhaps supplemented by publicly available social media sources, to determine how long a candidate is likely to stay on a particular job. Conversely, by combing through computerized calendar entries and e-mail headers, by way of artificial intelligence (AI), tools exist which can indicate which employees are likely to leave their employment within the next 12 months. These new tools and methods that rely on algorithms and the aggregation and analysis of a massive amount of data are becoming part of the daily landscape in human resource departments.

Similarly, the use of algorithms to review résumés and perform other recruiting functions is becoming far more commonplace. Novel solutions include games-based tools that seek to measure aptitude, tools that conduct interviews and evaluate candidates, and tools that scrape publicly available social media content. The promise of AI-based recruiting tools is to eliminate possible implicit bias of decision-makers and expand the pool of potential candidates. In this way, firms can leverage Big Data to identify and recruit optimal candidates. Employers may also turn to predictive recruiting tools for reasons of efficiency and cost savings by automating at least part of the recruiting process and identifying quality candidates who will stay for the long term.

Equally important, AI-based tools have the potential to promote diversity, equity, and inclusion by expanding the applicant pool and focusing on candidates’ abilities versus well-worn proxies for talent such as academic achievement, work history, and employee referrals, all of which are capable of perpetuating historical biases.

Legal Issues Surrounding the Use of Algorithmic Analysis. Deploying algorithmic tools is not risk-free, however, and should only be done carefully, with the involvement of all key stakeholders, and with the assistance of qualified counsel. Artificial intelligence offers a potent antidote to intentional discrimination. Antidiscrimination laws, however, also prohibit practices that are facially neutral if they have a disparate impact on members of protected categories, unless those practices are “job-related” and consistent with “business necessity.”

Given the complexity of amassing and then analyzing vast quantities of information, an employer would certainly not reverse engineer the process in order to intentionally discriminate against a protected group. It is far more probable that the use of algorithms may be challenged because it unintentionally yields a disparate impact on one or more protected

groups. More precisely, a plaintiff or class may allege that the algorithm used for hiring, promotion, or similar purpose adversely impacts one or more protected groups.

The legal landscape of how courts should view these challenges, and what methods of analysis they should apply, is limited, and despite the rapid advent of AI and predictive technology, employers continue to operate in a legal environment based on rules and regulations developed in an analog world with few guideposts that readily translate to the 21st century workplace. Issues may arise in a context that makes yesterday's compliance paradigm both outdated and difficult to apply.

For example, under current federal guidance, whether a selection method that produces an adverse impact passes muster under Title VII is often decided with reference to the Uniform Guidelines on Employee Selection Procedures (UGESP). The use of artificial intelligence and algorithmic screening tools can effect a shift from selection criteria distilled from job-related knowledge, skills, and abilities, leaving correlation to be established empirically, to one in which correlation is first established empirically- independently of knowledge, skills, and ability-and leaves the duration of that correlation in question. Unfortunately, because UGESP was adopted in 1978, it fails wholly to contemplate the use of algorithmic data and its reliance on correlation rather than cause-and-effect relationships.

Algorithmic analysis means that employers can theoretically analyze every aspect of every decision without worrying about a need to rely only on a partial sample, and this data allows employers to find (or, in some cases, to disprove) correlations between characteristics and outcomes that may or may not have a seeming connection. As a result, employers need to be able to understand what the use of these tools means for reducing the risks of traditional discrimination claims without giving rise to new varieties of such claims. But there are also new implications for background checks and employee privacy, data security obligations, and new theories of liability and new defenses based on statistical correlations, to name but a few.

The challenge for the legal system is to permit those engaged in the responsible development of Big Data methodologies in the employment sector to move forward and explore their possibilities without interference from guidelines and standards based on assumptions that no longer apply or that become obsolete the next year. An important part of this process is permitting employers to find and work with key business partners to assist in Big Data efforts and developing strategies that have the potential to make the workplace function more effectively for everyone. To do that well, it is vital that human resources professionals, statisticians, scientists, and a range of other stakeholders have a seat at the table when policy decisions are made regarding how and when to permit the use of these tools.

It is against this backdrop that we set forth our substantive concerns with the Law.

Concerns with the Law. We respectfully submit that mere clarification of the Law's penalty schedule is insufficient.³⁵ As set forth below, as currently adopted, a number of the Law's

³⁵ [With respect to the proposed penalty schedule, we respectfully submit that the Department consider providing an administrative "cure" period by way of regulation. Such a mechanism would](#)

provisions threaten to limit if not wholly eliminate the use of Big Data in employment-decision making within the City, with the perverse consequences of potentially increasing subjective decision-making and discrimination; dramatically expanding employer liability; and reducing the number of available employment opportunities. We have addressed these concerns below.

Our concerns and recommendations fall into three categories: the **scope** of the law in terms of geographic reach, the tools and systems it regulates, and the employment activity it regulates; its **operational impact** in terms of efficiency, fairness, and clarity of application; and finally, the

potential for **misapplication** leading to conflict with established principles of nondiscrimination law.

Scope of the Law

Location. The Law appears intended to apply only to those candidates for employment or promotion who work (or will work) within New York City (“Sec. 20-870: The term ‘employment decision’ means to screen candidates for employment or employees for promotion within the city.” (emphasis added)). However, there is potential ambiguity in the interpretation of this specific definition. We urge that the Department make clear that this is the extent of the Law’s requirements, and expressly provide that the Law does not apply to applicants or employees who do not work (or will not work) in New York City (or who work or will work within the City only casually, occasionally, or sporadically). Similarly, consistent with the Law’s text, we urge that the Department state clearly that the Law does not apply to employment decisions made within New York City with respect to applicants and candidates for positions outside of the City’s borders.

Recommended Revision:

Employment Decision. The term “employment decision” means to screen candidates for employment or employees for promotion, where both the individual is located and the employment opportunity is intended to be performed within the city.

Promotion. The application of the Law to promotion activity is likely to undermine and upend internal promotion processes that are already subject to anti-discrimination laws both state and federal, without generating any likely improvement in outcome. This is because of three fundamental differences between hiring and promotion. Hiring decisions typically result from an assessment of large numbers of candidates for each opening, while promotions are typically narrower, more individualized decisions. In addition, unlike the swiftness and potential opacity of hiring decisions, promotion decisions are typically part of an ongoing,

[allow the Department to notify an employer of any alleged non-compliance with the Law, and provide a period of time in which an employer could rectify any alleged deficiencies before being assessed with any fine.](#)

feedback-driven process based on well-defined job requirements, expectations, and performance criteria that are well known to the employee.

Recommended Further Revision:

Employment Decision. The term “employment decision” means to screen external candidates for employment or current employees for promotion employment, where both the individual is located and the employment opportunity is intended to be performed within the city.

“Automated Employment Decision Tool.” The Law’s definition of “automated employment decision tool” as “any computational process, derived from machine learning, statistical modeling, data analytics, or artificial intelligence” is susceptible to an unnecessarily over-broad reading, one which would apply the Law to just about any screening tool an employer might use to assess and rank candidates or employees, including traditional pencil-and-paper tests. Critically, the law does not provide any reference point for employers to determine when simple mechanical application of a tool crosses into the realm of “machine learning” or “artificial intelligence.” Consider the following progression of scenarios:

1. Acme Corp. gives candidates for an accounting position a pencil-and-paper math test.
2. Acme Corp. uses data analytics to identify the key math errors its accountants typically make, and revises the pencil-and-paper test to use problems that test those concepts.
3. Statistical modelling shows that computer scoring of the test is more accurate than human scoring, so Acme Corp. converts the test into an online format with the same questions.
4. Acme Corp. replaces the static question-and-answer format of its accounting test with an interactive math ‘game’ where a machine learning algorithm customizes the questions put to a candidate based on that candidate’s prior responses, and then rates their skill.

Scenario 1 is currently regulated by existing employment laws and regulations. At the other extreme, Scenario 4 appears to be squarely the type of automated decision-making that the Law is intended to regulate. The trouble is in the middle. The test in Scenario 2 is “derived from ... data analytics,” and is thus subject to the Law, even though it is just another pencil-and-paper

test. Similarly, the test in Scenario 3, being derived from an application of statistical modelling, is subject to the Law, even though it is merely an electronic reformatting of a pencil-and-paper

test. We expect that the Council did not intend these counterintuitive outcomes, and therefore urge the Department to adopt a more meaningful definition of “Automated Employment Decision Tool,” one that differentiates between the rote application of programs to perform pre-determine, human-driven calculations; pre-programmed, static mimicking of human actions; and true “machine learning” in which a program adapts its behaviors and programming absent human intervention.

Recommended Revision:

Automated employment decision tool. The term “automated employment decision tool” means any computational process, ~~derived from that~~ actively uses machine learning, or artificial intelligence, that issues simplified output, including a score, classification, or recommendation

We also urge the Department to define the terms “computational process,” “statistical modeling,” “data analytics,” “machine learning” and “artificial intelligence.” The differences in vernacular and technical uses of the latter term, particularly, makes it fertile for threshold disputes on the application (and misapplication) of the Law.

We respectfully add that the Law’s list of excluded tools is underinclusive and that the Law would benefit from further clarification of the “computational processes” at issue, as well as an elimination of the list. For example, the Law notes that junk mail filters, firewalls, calculators, spreadsheets, databases, data sets, and any “other compilation of data” are not automated decision tools. Insofar as these tools do not purport to make or inform any employment-related decision without human intervention, that would seem axiomatic. However, the tools enumerated could be part of a larger Automated Employment Decision Tool or process – the manner of use is what matters. For example, even though it “issues simplified output,” the rudimentary use of a filter or search function within résumé database is no different than searching through paper copies of the same résumés – it’s just quicker and more accurate. The only “automation” involved is simple rote filtering to find candidates who meet human-specified criteria, and that automation is not, we submit, the issue (just as the automation of a pencil-and-paper test is not, by itself, a concern). Applying machine learning algorithms to the same résumé database to rank or score candidates, on the other hand, seems the type of activity the Law aims to regulate.

Recommended Further Revision:

Automated employment decision tool. The term “automated employment decision tool” means any computational process, ~~derived from that~~ actively uses machine learning, or artificial intelligence, that issues ~~simplified output, including~~ a score, classification, or recommendation, that is used to substantially assist or replace discretionary decision making for making employment decisions ~~that impact natural persons.~~³⁶ ~~The term “automated employment decision tool” does not include a tool that does not automate, support, substantially assist or replace discretionary decision-making processes and that does not materially impact natural persons, including, but not limited to, a junk email filter, firewall, antivirus software, calculator, spreadsheet, database, data set, or other compilation of data.~~

Finally, we are particularly concerned with the application of the Law to a range of online jobseeking databases and websites, as the Law itself is not clear as to what is contemplated. If an employer goes to a website where job-seekers post their résumés, a human representative enters a series of keywords to match, and the website displays all applicants

³⁶ [We recommend striking the limitation to natural persons because the preceding term, “employment decisions,” is already a defined term of art within this law, and its definition already inheres this limitation.](#)

with that keyword on their résumé/c.v., this would plainly seem to be outside the realm of automated decision making, and we urge the Department to make this clear. Taking the example a step further, however, makes it even less clear how the Law applies to these websites. Suppose, for example, rather than entering keywords for the website to match and return, the employer’s representative enters a job-title, or skill set – asking for candidates for “office manager” or “social media administrator,” or seeking workers who have experience in “marketing” and “website design.” Based on its own program, the website processes the request, and returns a list of candidates to the employer. It may or may not rank this list, and the employer has no knowledge of how the website translated its “plain English” request into a search-and-sort function. Similarly, the employer has no knowledge as to whether and how this proprietary software (which it does not own and whose code and algorithms it has no access to) has developed this ranked list. In light of these facts, we urge the Department to specify in any final regulations that the requirements of the Law apply only to those tools over which an employer has control or proprietary access, and not to routine career-search websites which an employer accesses only as a third-party user.

Operational Impact of the Law

Ten Business Day Notice Period. The Law can be read to require that an applicant or employee be given no less than ten business days’ notice before the use of a subject screening tool, and that such employee or applicant be given the opportunity to request “an alternative selection process or accommodation.”

Read in this way as a practical matter, this means that a candidate must be given 10 days to elect a subjective assessment, so that a job scheduled to remain open for 30 days must remain open for 40, so a late-arriving applicant can take advantage of this 10-day window before the job is filled, and potentially further extend the “open” window beyond 40 days. Indeed, where more than one such latecomer seeks to apply to a position, the required “window” may be extended repeatedly. To guard against this scenario, we recommend that any final regulation make clear that an employer is not required to continuously extend a job posting to accommodate applicants who appear only in an “extended” window.

Perhaps more important, the Law wholly fails to contemplate the business needs of employers who seek to engage temporary or contingent workers immediately (and the workers who want and need these jobs). In that respect, the Law threatens the fundamental viability of staffing agencies and other businesses whose principal operation is the ability to fill jobs quickly, often with little to no advance notice. Given the volume of applicants these firms must screen on an expedited basis, a 10-day “notice” requirement would likely result in these firms ceasing to advertise, screen, and fill positions in performed in New York City. For these reasons, we urge the Department to clarify that the 10-day notice requirement extends only in those instances where a job is intended to be posted for at least 10 days, and that in all instances, an employer is required to give only that notice which is practicable given the nature of the position and the time intended for it to be open.

Notice of Job Qualifications and Characteristics. Often, by their very nature, automated screening tools driven by machine learning, statistical modelling and similar data-analytical AI techniques do not base their results on a specific qualification or characteristic, but rather by analyzing a vast amount of data and determining whether and how these data points may predict success (or likely success) in a position. Thus, while the job qualifications and characteristics present on the corresponding job description will, in a manual process, directly frame the a human recruiter’s analysis of a résumé or application (setting aside instances of discriminatory intent), in a machinelearning process the analysis of the same content is conducted using a mathematical framework that identifies and weights patterns within the data that are neither discernible to the human mind nor humanly computable in the given time. Critically, these patterns and weights are not driven by cause-and-effect.

By way of example, assume an algorithm assesses data about an employer’s existing workforce to attempt to screen applicants and rank them based on likely success retention in a position. The algorithm may determine with statistical significance that employees with a degree in a STEM concentration are—for whatever reason—statistically more likely to perform well, and stay in the position several months longer than candidates with a degree in a liberal arts concentration. The algorithm cannot explain why (nor is it tasked to) – it only proves that this trait correlates to stronger performance (the employer-user is likely also unaware of this correlation, knowing only that the algorithm has provided candidates statistically more likely to perform well and remain in the job). In these instances, it is not clear what information an employer is expected to provide to employees and applicants subject to the algorithmic tool—that the tool screens for successful performance? Likely retention? Undergraduate concentration? Moreover, the network of weighted patterns that such a tool generates is not reducible to a straightforward Englishlanguage description.

Thus, all that an employer may know (and be able to describe) is the information about the job and the candidate that is accessible to the tool (such as, for instance, the job description, the candidate’s résumé, and their responses to questions on the application). Accordingly, we recommend that the Department limit the substantive aspect of the Notice to those categories of information.

Recommended Revision:

The ~~job-specific qualifications and characteristics~~ and candidate-specific data that such automated employment decision tool will use in the assessment of such candidate or employee.

Alternative Assessment. In connection with the notice requirement, the Law also requires that applicants be given the opportunity to “request an alternative selection process or accommodation” but offers no insight into what this legally or practically requires. If an employer uses a screening tool to winnow 7,000 résumés for an open position to 50 for human review, is it required to offer the opportunity for human review to all of the 6,950 individuals who ask? Is the employer required to offer these individuals an in-person screening interview, even if they otherwise would not have qualified for one? We urge the Department to make clear that it is not.

Continuing, even assuming that such an exercise would be possible (which, in many instances, it simply will not be), the Law provides no guidance on how to decide between an employee who receives a glowing subjective assessment and another who scores highest based on data analytics. Is the employer under any obligation to weigh an “alternative assessment” differently than other candidates subject to the screen? Again, we urge the Department to specify that an employer is under no affirmative obligation to prefer a subjectively-screened candidate over one which is highlighted by an automated employment tool, and that an employer may (as in all instances) rely on its business judgment in selecting a candidate or candidates.

Bias Audit. The Law defines “bias audit” as “an impartial evaluation by an independent auditor,” and requires that any tool be tested to assess for “disparate impact” on the basis of race, ethnicity, or sex. The Law appears to require that the tool be assessed at least annually, and that “a summary of the results” of the most recent audit be made publicly available.

This requirement raises a host of questions as to how and under what circumstances a “bias audit” will be deemed sufficient under the Law. For example, assume an employer uses the same Tool to screen applicants for a wide range of positions, from C-suite to entry-level openings. Is an employer required to “audit” the tool for bias against each position? Against a class of openings? In the aggregate? Moreover, apart from disparate impact on the basis of race, sex, and ethnicity, it is unclear what the contents of an “audit” must show, or what analysis beyond that set forth in the Law an employer is required to conduct. Finally, how would the law be applied in those circumstances where a tool has a disparate impact against one group, but is superior to any alternative process for the overwhelming majority of candidates? We submit, given the statutory text, and the lack of any further direction or context from the bill’s sponsors, that a disparate impact analysis of the sort described satisfies the requirements of the Law, and urge the Department to explicitly state so in any final regulations.

Similarly, the Law offers no guidance as to what is required in a “summary” of results. Is a simple statement that the tool has been tested in accordance with the law and shows no evidence of disparate impact on the basis of race, ethnicity, or gender sufficient? We would urge the Department to clarify that such a statement (where accurate) satisfies the requirements of the law.

Finally, we urge the Department to align the requirement for a bias audit of a tool with the Law’s requirements relating to notice. Specifically, the Law requires that notice be given only where a tool is used to screen “an employee or candidate who has applied for a position for an employment decision” (emphasis added). The requirement for a bias audit, however, requires an audit to be conducted where an automated employment decision tool is used “to screen a candidate or employee for an employment decision.” For example, consider software that transfers a résumé into a database by using machine learning or another computational process to extract pieces of information (e.g., name, jobs held, degrees held, etc.). Assume, further, that the database, once created, is used and reviewed solely and directly by humans. At the point that an automated tool is being applied, then, there is no position being applied for, so that class of tool should not be within the scope of the Law.

Indeed, the notion that an individual has to actually have applied for a position is implicit in the Law’s textual reference to “an employment decision” – if no position is being sought, it is axiomatic that no “employment decision” has been made. Moreover, it is unclear how an audit could be conducted in this scenario, insofar as no employment decision is being made.

Recommended Revision:
§ 20-871. Requirements for automated employment decision tools.
a. In the city, it shall be unlawful for an employer or an employment agency to use an automated employment decision tool to screen a candidate who has applied for a position or employee...

Potential Misapplication of the Law

Finally, in requiring that disclosure of the results of an audit be made public, the Law appears to not contemplate the fact that under federal and cognate state civil rights laws, a screening tool may be lawfully used even if it has a disparate impact on protected groups. See, e.g., 42 U.S.C. § 2000e-2(k)(1)(A)(i) (employment practice that is job related for the position in question and consistent with business necessity is lawful despite disparate impact); id. § (k)(1)(A)(ii) (employment practice is lawful despite disparate impact where no less discriminatory alternative practice exists). In so doing, the Law is likely to lead to frivolous charges of discrimination, increased expense for employers, and mismatched expectations for applicants and employees. We urge the Department to be mindful of this in crafting regulations, and consider explicitly making clear to applicants and employees that the mere fact that a screening tool results in some disparate impact does not make its use per se unlawful.

Recommended Revision:
§ 20-874 Construction. The provisions of this subchapter shall not be construed to limit any right of any candidate or employee for an employment decision to bring a civil action in any court of competent jurisdiction, or to limit the authority of the commission on human rights to enforce the provisions of title 8, in accordance with law. The provisions of this chapter shall not be construed to limit the right of an employer or employment agency, pursuant to 42 U.S.C. § 2000e-2(k)(1)(A)(i) & (ii), to use a screening tool that has a disparate impact on protected groups so long as the employment practice at issue is job related for the position in question and consistent with business necessity, or where no less discriminatory alternative practice exists.

* * *

The above highlights key issues we have identified with the Law thus far; we expect there will be others, and we welcome the opportunity to work cooperatively with the Department toward regulatory solutions that address the practical and business realities of how these tools are used in the modern workplace.

Respectfully submitted,

A handwritten signature in black ink, appearing to read "J. Paretti".

James A. Paretti, Jr.
Shareholder

for

LITTLER MENDELSON, P.C.
WORKPLACE POLICY INSTITUTE