

Algorithmic Bias Safeguards *for* Workforce

Overview

DISCLAIMER TO THE ALGORITHMIC BIAS SAFEGUARDS FOR WORKFORCE

These Algorithmic Bias Safeguards for Workforce (“Safeguards”) are made available to you solely for informational purposes. The Safeguards, which include the “Evaluation” “Education and Assessment,” “Scorecard,” and “Implementation Guidance” documents and materials, and all associated documents and materials made available to you in connection therewith, were created by the Data & Trust Alliance to facilitate the collection of information from potential vendors regarding their potential to introduce unfair bias into workforce processes. The Safeguards are intended solely for consideration as a supplement to your existing vendor selection procedures. You remain free not to use the Safeguards or to use only part of the Safeguards. Any conclusion or action taken by you in connection with your use of the Safeguards shall be and is made or taken in your sole discretion. In no event shall any such conclusion or action be construed as reflecting the opinion, endorsement or direction of the Data & Trust Alliance. By using the Safeguards, you hereby agree not to make any statement, directly or indirectly, contrary to the foregoing.

The Safeguards do not constitute legal or other professional advice. The Data & Trust Alliance is not a law firm and is not engaged in the practice of law or providing legal services. All legal questions regarding whether and how to use the Safeguards, including any information collected in connection therewith, should be directed to legal counsel retained by you.

The Data & Trust Alliance does not, and will not, provide advice to you about your business relationships with individual vendors. You are free to decide your own hiring, advancement, data use, AI, procurement, HR and vendor policies. In connection with, and as a condition to, your use of the Safeguards, you agree not to share non-public information with other members of the Data & Trust Alliance about your internal decision making procedures, individual employment candidates or individual vendors (including without limitation completed vendor questionnaires, scorecards, pricing or technical aspects of vendor software) and you shall not collectively with such other members refuse to deal with individual vendors or certain categories of vendors. While the Data & Trust Alliance intends to collect your feedback about the Safeguards, it will not collect feedback about individual vendors and any feedback will only be published by the Data & Trust Alliance on an aggregated basis.

The Data & Trust Alliance does not warrant the accuracy, completeness, or usefulness of the Safeguards or any information collected in connection therewith and undertakes no obligation

to update the Safeguards at any subsequent time or in response to any subsequent developments. You should not and are not authorized to rely on the Safeguards and your use of the Safeguards and any such information is strictly at your own risk. The Data & Trust Alliance disclaims all liability and responsibility arising from any reliance placed on the Safeguards or any information collected in connection therewith, or by anyone who may be informed of any of the Safeguards’ contents or responses thereto.

YOUR USE OF THE SAFEGUARDS, INCLUDING BUT NOT LIMITED TO ANY INFORMATION COLLECTED IN CONNECTION THEREWITH, IS AT YOUR OWN RISK. THE SAFEGUARDS ARE PROVIDED ON AN “AS IS” BASIS, WITHOUT ANY WARRANTIES OF ANY KIND, EITHER EXPRESS OR IMPLIED. TO THE FULLEST EXTENT PROVIDED BY LAW, THE DATA & TRUST ALLIANCE HEREBY DISCLAIMS ALL WARRANTIES OF ANY KIND, WHETHER EXPRESS OR IMPLIED, STATUTORY, OR OTHERWISE, INCLUDING BUT NOT LIMITED TO ANY WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT, AND FITNESS FOR PARTICULAR PURPOSE.

TO THE FULLEST EXTENT PROVIDED BY LAW, IN NO EVENT WILL THE DATA & TRUST ALLIANCE, ITS AFFILIATES OR THEIR LICENSORS, SERVICE PROVIDERS, EMPLOYEES, AGENTS, OFFICERS, MANAGERS, OR DIRECTORS BE LIABLE FOR DAMAGES OF ANY KIND, UNDER ANY LEGAL THEORY, ARISING OUT OF OR IN CONNECTION WITH YOUR USE OF THE SAFEGUARDS INCLUDING BUT NOT LIMITED TO ANY INFORMATION COLLECTED, OR THE FAILURE TO COLLECT INFORMATION, IN CONNECTION THEREWITH, INCLUDING ANY DIRECT, INDIRECT, SPECIAL, INCIDENTAL, CONSEQUENTIAL, OR PUNITIVE DAMAGES, INCLUDING BUT NOT LIMITED TO PERSONAL INJURY, PAIN AND SUFFERING, EMOTIONAL DISTRESS, LOSS OF REVENUE, LOSS OF PROFITS, LOSS OF BUSINESS OR ANTICIPATED SAVINGS, LOSS OF USE, LOSS OF GOODWILL, LOSS OF DATA, AND WHETHER CAUSED BY TORT (INCLUDING NEGLIGENCE), BREACH OF CONTRACT, OR OTHERWISE, EVEN IF FORESEEABLE.

The Data & Trust Alliance name and the Data & Trust Alliance logo are trademarks of the Data & Trust Alliance. All other names, logos, product and services names, designs and slogans are the trademarks of their respective owners.

©2022 Data & Trust Alliance

Thank you for your interest in the Data & Trust Alliance and the Algorithmic Bias Safeguards for Workforce.

The Data & Trust Alliance brings together leading businesses and institutions across multiple industries to learn, develop, and adopt responsible data and AI practices.

As part of the D&TA initiative on Algorithmic Safety, the Algorithmic Bias Safeguards for Workforce are designed for HR teams to evaluate vendors on their ability to detect, mitigate and monitor algorithmic bias in workforce decisions.

As norms around algorithmic systems and algorithmic bias change, these materials will evolve. We anticipate updates to the Safeguards as the use of these criteria matures, and as industry adapts.

We welcome feedback and your engagement. If you are interested in accessing the Safeguards, please contact us at algorithmicbias@dataandtrustalliance.org.

About the Data & Trust Alliance

The Data & Trust Alliance is a not-for-profit consortium established in September 2020. It brings together leading businesses and institutions to learn, develop and adopt responsible data and AI practices. It is co-chaired by Ken Chenault, chairman of General Catalyst and former chairman and CEO of American Express, and Sam Palmisano, former chairman and CEO of IBM.

dataandtrustalliance.org

5M
employed by Alliance
member organizations

\$3.6T+
market capitalization of
Alliance companies

\$1.6T+
revenue of
Alliance companies
in 2020

Member Companies & Institutions (as of January 2022)



About the Algorithmic Bias Safeguards for Workforce

Businesses and institutions are increasingly applying data, algorithms and AI to support their workforce decisions—from hiring and promotion to productivity and compensation.⁰¹

These technologies help identify talent in larger and more diverse pools of candidates, better match the right talent to the right opportunity, personalize employee experiences, and automate routine tasks to free up time for more meaningful work.

However, these technologies also come with risks. Alliance member organizations identified unfair bias as one of the highest risks when using these technologies in workforce.

Most of the algorithmic systems used to support workforce decisions are introduced and maintained by vendors—including software providers, professional networking sites, consultants, and recruiting firms.

This prompted the Alliance to develop the Algorithmic Bias Safeguards for Workforce—criteria and education for HR teams to evaluate vendors on their ability to detect, mitigate and monitor algorithmic bias in workforce decisions.⁰²

01 The Safeguards use the terms algorithmic system, AI, AI system, model, and algorithmic decision-making/decision support system collectively and interchangeably to cover several related but distinct terms, including algorithms, statistics, rules, artificial intelligence, machine learning, deep learning, and neural networks, as applicable. Briefly, an algorithm is a finite series of well-defined, computer-implementable instructions or rules. While all AI uses algorithms, not all algorithms use AI. AI in turn includes machine learning, and machine learning in turn includes deep learning. See for example, Rebecca Kelly Slaughter, *Algorithms and Economic Justice*, 23 *Yale J.L. & Tech. Special Issue 1, 2* (2021).

02 The Safeguards define algorithmic bias as predictions or outputs from algorithmic systems that exhibit unjustified differential treatment between two groups. When these groups are distinguished by legally protected characteristics such as disability, race, age, or sex, algorithmic bias may lead to unlawful discrimination. Addressing the problem of algorithmic bias therefore reduces the risk of engaging in unlawful discrimination.

Safeguards Goals

The Alliance has three principal goals in the development and distribution of these Safeguards:

- 1. Adoption.** Use of these Safeguards across industries will help establish bias mitigation as a key criteria for developing, selecting and safely operating algorithmic systems in HR.
- 2. Learning.** We intend to learn from implementation and emerging best practices, and evolve the Safeguards over time.
- 3. Partnership with vendors.** This work is designed to support both buyers and vendors. This requires collaboration, partnership, and consistent feedback.

Adapting the Safeguards

Organizations are at different stages in their uses of algorithmic systems. Some have robust algorithmic governance processes while others are early in their journey. Not all organizations or vendors are equipped to complete the comprehensive evaluation.

The Safeguards are designed to be used in their complete form, but can be adapted to fit existing systems. An abbreviated Evaluation of 20 questions (from the original 55) is available for foundational compliance.

Components of the Algorithmic Bias Safeguards for Workforce

The Algorithmic Bias Safeguards for Workforce include four components to support organizations that will implement these safeguards into their systems—so they can evaluate potential vendors on their ability to detect, mitigate, and monitor algorithmic bias.

01 Evaluation

55 questions in 13 categories for completion by the HR vendor. Answers are not shared among member companies.

02 Education & Assessment

To enhance algorithmic literacy and to provide detailed guidance for HR teams assessing vendor responses to the Evaluation.

03 Scorecard

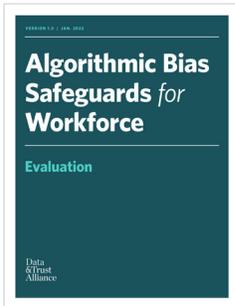
To qualitatively grade and compare vendors and document issues.

04 Implementation Guidance

For integrating the Safeguards into an organization's systems. The Safeguards supplement member companies' vendor selection procedures.

01
Evaluation

The Evaluation is a set of questions to evaluate HR vendors on their ability to detect, mitigate, and monitor algorithmic bias—55 questions across 13 categories, designed for use in the RFI/RFP process.

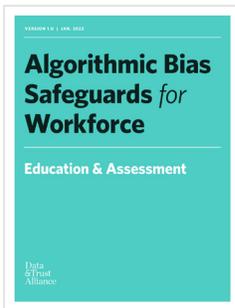


	Category	What the category evaluates
00	Applicability of the Evaluation	Does the offering employ an algorithmic system? Does the offering influence employment decisions? Offerings that do not apply will not need to complete the Evaluation.
Value Proposition		
01	Purpose & Business Value	How is the system meant to be used, and what business value does it provide?
Addressing Bias Across System Lifecycle		
02	Model Design & Training Data	What measures are taken to detect and mitigate bias in (1) the data used to train the model and (2) the design of the model itself?
03	Model Training	How is bias minimized while ensuring maximum performance during the model training stage?
04	Bias Testing	Which legally required and emergent best practice tests are used to detect bias?
05	Bias Remediation & Business Justification	What approaches are used to remediate bias? What is the business justification for any remaining bias?
06	Deployment & Monitoring	What practices are used to mitigate bias during deployment, as well as ongoing practices to monitor bias in the system?
Addressing Bias Through Organizational Practices		
07	Performance	What measures have been taken (and what documentation is available) to demonstrate that the system performs as intended, and as claimed?
08	Governance	What governance procedures are in place to insulate against the legal and ethical risk resulting from bias in the system?
09	Transparency & Accountability	How are transparency, explainability, and override enabled within the system?
10	Compliance, Standards, Insurance, and Certifications	How well are legal liabilities and related compliance practices understood and addressed?
11	Education	How thorough is the education for both your organization, and the buyer organization, to properly use the system and mitigate bias?
12	Ethics & Diversity Commitments	What are commitments to ethical practice—and how have they translated into practice?
13	Accommodations & Alternatives	How does the system account for users with varying needs and disabilities—and how are alternatives and opt-in/-out provided, when needed or requested?

02
Education & Assessment

The Education & Assessment is designed to (1) help an HR buyer or vendor selection teams build a baseline of algorithmic literacy and familiarize themselves with algorithmic bias and (2) assess vendor responses to the Evaluation.

It includes an algorithmic bias primer, an overview of key terms, and deep-dives on each question, providing guidance on how to assess vendor responses.



TOPIC 05 | BIAS REMEDIATION & BUSINESS JUSTIFICATION

5.3 Business Justification

To the extent that any potential bias has not been remediated or fully identified, describe why. Be as specific as possible when describing any potential bias and its impact, including the potential harms created by false negative/false positive outcomes and/or any erroneously high or low numeric predictions. Include any business or performance reasons that are relevant to your determination.

This question seeks to understand the vendor's business rationale for not taking further efforts to mitigate any previously detected bias.

Legal requirements for bias testing generally recognize that bias can never be fully removed from employment decisions.

If bias has been detected (through prior testing), a vendor should show that all reasonable steps have been taken to address that detected bias. If a vendor purposefully leaves detected bias in the system, they should describe why, based on legitimate business purposes (e.g., job required, technical skill)

Note: U.S. anti-discrimination standards in employment decisions require that any remaining bias be mitigated or justified in relation to business necessity.

Answer Guidance

Red	• Vendor offers an answer claiming to have eliminated all bias. This demonstrates a lack of basic understanding around (or attempting to hide) the fact that all probabilistic decision-making systems will exhibit some degree of bias.
Yellow	• Acknowledgment of remaining bias (that has previously been detected), yet inability to describe associated potential harms or business justifications for continuing with bringing the product to market with specificity.
Green	• The vendor explains why previously detected bias is present, based on legitimate business purposes. This is done alongside demonstrating that there is no reasonably better way to achieve similar results, given the model's objectives. • The vendor is able to describe the bias that remains within the model in the form of specifying potential harms. Those harms should be minimal and justified by legitimate business requirements.

Red
This subject implicates higher risk.

Data & Trust Alliance | Algorithmic Bias Safeguards for Workforce | 17

Key takeaway
A high-level description of the question in plain language

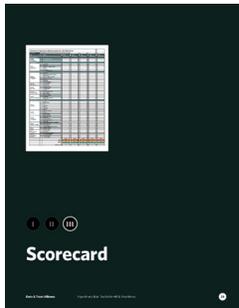
Explanation of question
An overview that explains why the question matters, shares foundational concepts, and outlines risks

Assessment guidance
For each question, guidance is provided around how to interpret a vendor's answers—from highest bias risk (red) to moderate bias-risk (yellow) to lowest-risk (green) answers.

03
Scorecard

The Scorecard is designed to help reviewers qualitatively grade and compare vendor responses to each question in the Evaluation.

The scoring can help a reviewer flag which vendors should be advanced or challenged on the basis of algorithmic bias mitigation practices. The Scorecard is not designed as a stand-alone decision tool, but rather helps inform a wider qualitative assessment of a vendor and their offering.



Scorecard Sheet
A tally sheet to grade and compare vendors

Assessment Guidance
Assessment guidance from Education

Scorecard Instructions

This document is designed to help reviewers assign a qualitative bias risk score to a vendor's offering.

The relative risk and liability level of every vendor relationship is a choice each organization must make individually.

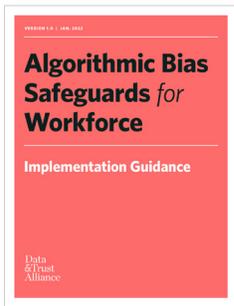
- Appendix 5 highlights an overview.**
- Scorecard instructions to reviewers, for each vendor submission.**
- Scorecard instructions to reviewers, for each vendor submission.**
- Scorecard instructions to reviewers, for each vendor submission.**

Instructions
Guidance for use

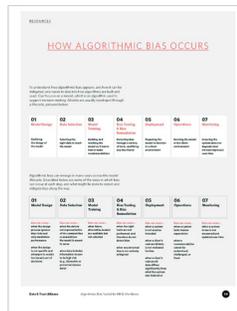
04 Implementation Guidance

The Implementation Guidance supports bringing the Safeguards into use in an organization. Designed for use by HR teams, procurement practitioners, and governance committees, the Implementation Guidance features instructions for each component of the Safeguards and their intended use.

The Implementation Guidance also includes foundational overviews of algorithmic systems and algorithmic bias, as well as support for communicating about these topics within organizations and with vendors.



Instructions for implementation
Guidance around operationalizing the Safeguards



Overview of algorithmic bias
Algorithmic bias and model lifecycle primer



Glossary
Key definitions to ground best practice

Contributors

The Alliance engaged a breadth of background and expertise—from algorithmic accountability to diversity, equity, and inclusion—in the development of the Algorithmic Bias Safeguards for Workforce.

200+ experts

from more than
15 industries

2,000+ hours

of interviews and
co-creation sessions

40%

of contributors from
outside the Alliance

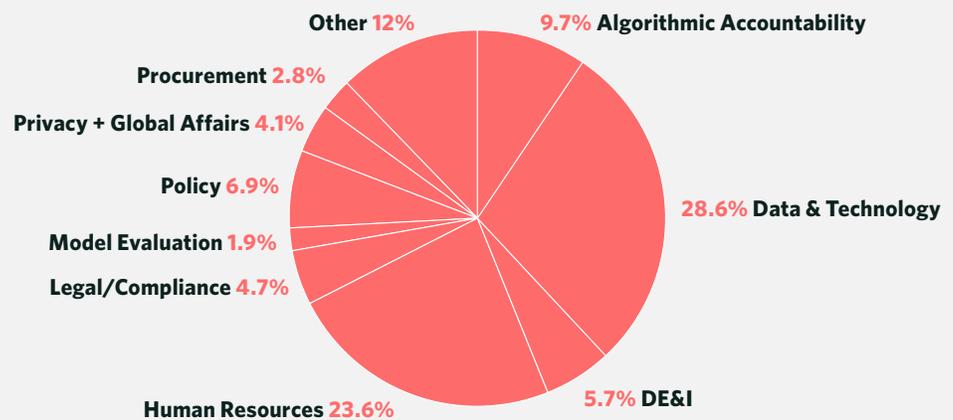
65 contributors

from academia, government,
and civil society—advancing
ethics and technical
accountability

20+ vendors

engaged in the
development process

Key Areas of Expertise



Co-chairs

Leadership Council members of the Data & Trust Alliance who shepherded this work.



Mike Capps, Diveplane,
Chairman & CEO

“For communication about responsible data practices, we must first have a shared language. Then education, especially in such a dynamic field. Next, transparency of vendor practices. And only then can we all collaborate to improve the state of the art.”

Dr. Michael Capps is a well-known technologist and CEO of Diveplane Corporation, a machine learning platform company focused on keeping humanity in AI. Before co-founding Diveplane, Mike had a legendary career in the videogame industry as president of Epic Games, makers of blockbusters Fortnite and Gears of War. For his research in VR, he was featured in SIGGRAPH’s documentary on computer graphics pioneers. He is a regular host of multiple television series on the Discovery and Science channels.



Bob Darin, Healthcare Analytics and Technology Executive, *former Chief Data & Analytics Officer, CVS Health*

“This is an accelerator—it helps achieve commitments to diversity and equity while enabling broader use of AI tools. The importance of this initiative has been recognized across the hundreds of people that have helped develop this, and it is an important component of overall corporate citizenship.”

Bob Darin is a nationally recognized expert in healthcare analytics. Most recently, he served as chief data officer for CVS Health, and has held executive positions at Bupa Healthcare (UK) and Cardinal Health. He has led the development of data science applications across healthcare settings, and is currently working with several healthcare AI growth-stage companies. Bob holds an honors MBA from the University of Chicago and received a degree in economics from Harvard College.



Nuala O'Connor, Walmart Inc.,
SVP & Chief Counsel, Digital Citizenship

“This is not only an anti-bias tool. It is an improved outcomes tool—for a time in which talent recruitment and retention are becoming critical for fairness, opportunity, and future business success.”

Nuala O'Connor oversees the Digital Citizenship team responsible for advising Walmart on issues related to privacy, data use and governance, emerging technologies, cybersecurity, and records management. She is a member of the President’s Inclusion Council focused on efforts to promote inclusive environments. Before Walmart, Nuala served in various privacy and trust leadership roles across the public and private sectors, including as the first chief privacy officer for the U.S. Department of Homeland Security.

Core Working Team

Experts drawn from Data & Trust Alliance member organizations, responsible for originating and validating the Safeguards.

AI Model Evaluation



Chris Kennedy, Regions Bank, SVP, Strategic Initiatives, Technology & Operations (former Deputy Head, AI Model Evaluation)

“Artificial intelligence and machine learning can amplify, often unintentionally, biased or undesirable outcomes. The initiative is one way to shine a light on this risk and make it easier to do the right thing.”

Data & AI



Anshul Sheopuri, IBM, VP & CTO, Data & AI, HR; IBM Distinguished Engineer

“Trustworthy AI is not just a nice to have but a societal imperative to ensure equal access to opportunity to all. The potential of shaping the future of AI deployed at scale is an exciting yet humbling experience.”

DE&I



Jonathan Beane, NFL, SVP, Chief Diversity & Inclusion Officer

“The Evaluation is a critical tool that ensures technological advancements are held to a standard of fairness, equity, and the opportunity for all to be evaluated on the merits. It addresses this for the tools of today and, most importantly, for the technological tools of tomorrow.”

Human Resources



Esther Gallo, Mastercard, SVP, Workforce Analytics and Innovation

“As HR professionals, we have the responsibility to minimize potential bias in our processes and systems. Technology, and especially AI, is going to change how we operate in the HR function, but we need awareness of the associated risks—and to also keep our vendors accountable.”

Legal & Compliance



Kat Robison, Nike, Associate General Counsel, Global Privacy & Security

“The initiative challenged us to proactively consider how to unlock the appropriate and responsible use of AI within our organization. The Evaluation provides a usable framework that helps us feel confident that our use of AI does not unintentionally undermine our broader goals.”

Procurement



Matt Iannetta, CVS Health, Sr. Director, Enterprise Modernization

“This initiative provided our Procurement organization with an effective evaluation framework to ensure we partner with suppliers that can demonstrate the deployment of responsible AI practices within their organization.”

Thank You

Connect with Us

The Data & Trust Alliance will continue to learn how these Safeguards meet the needs of industry and the workforce. Contact us at algorithmicbias@dataandtrustalliance.org.