The Institute for
WORKPLACE EQUALITY

TECHNICAL ADVISORY COMMITTEE REPORT

*EEO AND DEI&A CONSIDERATIONS IN
THE USE OF ARTIFICIAL INTELLIGENCE
IN EMPLOYMENT DECISION MAKING*

December 2022

1920 I Street NW • Washington DC 20006 • 202.293.2220
Theinstitute4workplaceequality.org

PREFACE

The Institute for Workplace Equality ("The Institute") is honored to have established and facilitated the Artificial Intelligence Technical Advisory Committee ("AI TAC") to prepare the *Report on EEO and DEI&A Considerations in the use of Artificial Intelligence in Employment Decision Making.*

In issuing the Report, The Institute provides analysis of and recommendations on how to address key compliance issues currently arising in relation to the use of AI in employment. The Report is available at no charge for all stakeholders and should become a key resource in the broader policy debate on how AI is used in employment decisions and how the agencies that govern the workplace will address AI in their policy and compliance efforts.

Under the leadership of Victoria ("Vicki") Lipnic, former EEOC Acting Chair and Commissioner, and with the robust participation of professional, multi-disciplinary members of the AI TAC, a very detailed and timely Report has been produced, following 18 months of extensive efforts. The Report was prepared by a wide range of subject matter experts, including data scientists, industrial organization psychologists, employment attorneys representing employers and workers, employers using AI tools for employment decisions, vendors who develop and provide AI tools for employment decisions, and former EEO government agency officials.

We want to thank Vicki Lipnic for her tireless efforts in chairing the AI TAC, the individual AI TAC members for their significant contributions and efforts, and the staff members who have donated significant time and resources to enable the Report to be issued.

Respectfully,

David Cohen
Co-Founder, The Institute

David Fortney
Co-Founder, The Institute

TABLE OF CONTENTS

FOREWORD

This report is a very deep dive into some of the most important EEO and anti-discrimination issues as applied to the use of Artificial Intelligence ("AI")-enabled employment tools as they operate within the current regulatory framework.

In the late summer of 2021, the founders of The Institute for Workplace Equality, David Cohen and David Fortney, both of whom I have known for years in the equal employment opportunity legal and compliance world, approached me about an advisory committee they were putting together to study and produce a report about the use of AI in employment decision making. They were organizing a group of multi-disciplinary experts to study the issues and asked me to chair the group. I agreed to do so.[1]

This was not my first time examining the use of AI in employment. In 2016, when I was serving as a Commissioner at the U.S. Equal Employment Opportunity Commission, I helped organize the EEOC's first look into the world of "Big Data" in Employment.[2] At that time, talk of "big data" was everywhere and my fellow EEOC commissioners and I wanted to understand how it was being used in the employment context. Even prior to my tenure at the EEOC, as early as 2004, when I served as Assistant Secretary of Labor and the Office of Federal Contract Compliance Programs ("OFCCP") was part of my enforcement responsibilities, a precursor to the world of "big data" was the sheer volume of "internet applicants" and those concerns led us to promulgate regulations to address that topic.

Six years after that EEOC public hearing the questions about big data are more pointed about the use of Artificial Intelligence. Check most major news outlets on a daily basis and you will likely find some article reporting about some application of AI in some aspect of our lives. As the technology has advanced and concerns about AI have multiplied so, too, have the studies, reports, articles by interest groups, business groups, academics, and other professionals. In just the year-and-a-half that this Technical Advisory Committee has been doing its work, numerous organizations have produced reports suggesting approaches or principles to be followed in the

---

[1] Full disclosure, I serve on the Advisory Board and the ethics council for an AI company. https://www.theinstitute4workplaceequality.org/. Not all Members of the AI TAC are members of The Institute.

[2] U.S. Equal Emp. Opportunity Comm'n, *Meeting of October 13, 2016 - Big Data in the Workplace: Examining Implications for Equal Employment Opportunity Law* (October 13, 2016), https://www.eeoc.gov/meetings/meeting-october-13-2016-big-data-workplace-examining-implications-equal-employment

use of AI broadly and in specific contexts. “Accountability,” “explainability,” “fairness,” “ethics,” “and bias” are among the most important cautionary words used in these studies.

Importantly, the use of AI in the employment realm falls into a particular legal context, namely, our well-developed anti-discrimination laws and the requirements to ensure equal employment opportunity. Enter into the public discourse this Report and this particular group of experts. The members of this group, by design, come from different disciplines – lawyers representing employees, lawyers representing employers; industrial-organizational psychologists; data scientists; psychologists; labor economists; developers of AI-enabled tools; employee advocates; academics; former government officials. It is a deep bench with tremendous credentials and years of experience. But the characteristic that distinguishes the group the most - and of which I am especially proud - is their shared commitment to ensuring that people do not experience discrimination in their desire to be employed and in their employment.

With this goal in mind, the Report provides a 360-degree technical review explaining existing professional standards and established legal precedent about data collection, employee selection procedures, statistics and adverse impact, and the important concepts of transparency and fairness, and cross walks those to the world of AI-enabled tools in employment. This Report should serve as an important reference source for users and developers of these tools and for the regulators contemplating their implications.

My sincere thanks to everyone who participated in this effort.

Victoria A. Lipnic
Washington, DC
December 16, 2022

EXECUTIVE SUMMARY

The increasing use of Artificial Intelligence ("AI")-enabled selection tools and processes for making decisions across the employment life cycle raises pressing questions about the Equal Employment Opportunity ("EEO") and Diversity, Equity, Inclusion, and Accessibility ("DEI&A") issues that arise in relation to the use of such tools. To address those questions, The Institute for Workplace Equality ("The Institute") created an Artificial Intelligence Technical Advisory Committee ("AI TAC") consisting of 40 subject matter experts and tasked them with identifying key issues and providing recommendations for how best to approach them. The Institute is a non-profit employer association that provides training and education to assist companies in understanding affirmative action and equal employment opportunity obligations. The Institute's programming addresses a wide range of human resource management strategies to assist employers in creating and maintaining diverse organizations free from workplace bias.[1]

The AI TAC includes labor economists, data scientists, industrial and organizational psychologists, attorneys representing employers and workers, civil society advocates for technology and democracy, vendors who develop and provide AI tools for employment decisions, employers using AI tools for employment decisions, and former Equal Employment Opportunity Commission ("EEOC") and Office of Federal Contract Compliance Programs ("OFCCP") officials. Not all members of the AI TAC are members of The Institute. The AI TAC is chaired by Victoria A. Lipnic, former Commissioner and former Acting Chair of the EEOC.

The AI TAC process began with creation of an extensive survey designed to identify and measure AI TAC Members' concerns regarding a range of issues related to EEO and DEI&A and the use of AI tools in employment decision making. Once the survey results were received, the AI TAC Members were split into Subcommittees to address specific aspects of AI use in the employment context. Over the following months, each Subcommittee held meetings and discussions during which its Members reviewed the survey results, analyzed the specific EEO and DEI&A issues arising within their focus area, and considered the best ways to address those issues. Each Subcommittee then wrote up an initial draft of a Report Section. The full AI TAC Committee then held a full-day meeting in person during which the Members gave feedback to

[1] https://www.theinstitute4workplaceequality.org/.

each of the Subcommittees. Following that meeting, each Subcommittee finalized the writing of their Section and the Editorial Committee then edited those Sections and Appendices into a full Report. This Report was made publicly available on December 21, 2022 for the particular attention of the EEO and DEI&A community.

Some of the key findings and recommendations discussed in the Report include the following:

- The most prevalent use of AI in the employment context today is in sourcing candidates and making hiring decisions. But employers are increasingly using or seeking to use AI-enabled tools across the full employment lifecycle.
- While there currently are no federal laws imposing specific transparency, notice, consent, or privacy requirements for use of AI in the employment process, several states and localities have implemented such laws, which may require employers to adjust their approach to the use of AI tools.
- Employers who use AI should be transparent about when AI is part of a specific selection procedure or is used in the overall selection process. Specific notice that an AI-enabled tool or process is being used should be provided to anyone who will be assessed using it.
- Just as employers need to provide information to applicants about the use of AI tools, vendors of those AI tools need to provide information about the tools to the employers utilizing them. The level of detail that should be shared may differ depending on the target audience and purpose of the notification.
- The issue of consent for the collection and use of data for employee selection processes is more complex in the context of AI-enabled processes than traditional selection processes because it is often difficult for applicants in an AI-enabled process to discern what data are being collected and the criteria upon which they are being evaluated. While it may be possible to imply consent by an applicant who has received notice and then proceeded to engage in an AI-enabled process, it is a better practice for an employer not only to provide clear notice to applicants that AI is being used and how it is being used, but also to obtain consent from the applicant.
- The use of AI-enabled selection procedures also poses unique privacy challenges. A fundamental characteristic of AI – and especially of machine learning algorithms – is the ability to gather and quickly process a high volume of information. This AI-specific characteristic may result in large amounts of personal data about applicants and employees being obtained, used, and stored. Employers should take specific steps to address this concern and be mindful of applicable data privacy rules.
- As a matter of fairness, employers should establish, and maintain with regular auditing, explicit procedural safeguards to ensure equitable treatment and comparable access while limiting algorithmic bias.
- Use of AI-enabled employment selection tools may have amplified impact on people with disabilities. Employers should therefore familiarize themselves with the recent Technical Guidance Documents addressing the interaction of Artificial Intelligence and the

Americans with Disabilities Act issued by the U.S. Equal Employment Opportunity Commission and U.S. Department of Justice.

- Employers who rely on AI technology in recruiting and hiring need to know the source and quality of the data being used. Importantly, some techniques for establishing reliability of data may or may not apply to newer sources of data used in AI systems.
- The data used in machine learning, as well as the key choices made in collecting, cleaning, training, and maintaining a dataset and building a selection algorithm, should be documented and a rationale provided. Data collection procedures should include sufficient documentation to assess data reliability and validity and allow for computational reproducibility.
- As regards assessing whether an AI-enabled employment selection tool has an adverse impact, the Uniform Guidelines on Employee Selection Procedures ("UGESP") – which explicitly recognize that the science and technology of employee selection continues to evolve – remain the lens through which the federal agencies, employment attorneys, and many courts are likely to analyze the adverse impact and job-relatedness of such tools. Nevertheless, some have concerns regarding the sufficiency of the UGESP validity strategies for certain types of AI applications, such as "black box" algorithms and machine learning.
- The concept of "debiasing" AI models for use in developing employment selection tools is consistent with UGESP's general approach of searching for suitable alternative selection procedures with less adverse impact. However, some debiasing techniques align more closely with UGESP than others, and it is important for compliance experts to carefully inspect and fully understand the details of the debiasing solution before it goes forward.
- There is no one-size-fits-all statistical model that necessarily, unambiguously, and in every case "proves" adverse impact or discrimination. Moreover, understanding the questions that different statistical tests of adverse impact seek to answer, and the assumptions underlying each of them, is increasingly important in a world where AI tools may be deployed at both modest and massive scale.
- "Intelligent" AI – that is, tools that adjust their selection algorithm or scoring criteria over time as they "learn" more about the characteristics of successful applicants or job incumbents, which themselves are changing over time – may strain the traditional framework for assessing adverse impact. It may be advisable for employers to consider developing a policy for version control and associated guardrails around substantive changes to the functioning of the AI-enabled tool in employment contexts.
- Careful attention should be paid when assessments of adverse impact are put forward as a basis to modify an AI-enabled selection tool – for example, where dynamic machine learning algorithms are structured to weigh the data in a manner intended to minimize (or reduce) adverse impact, rather than in a manner that would maximize (or improve) the reliability of measuring job-relevant constructs.
- With analyses of AI-enabled tools, as with more traditional selection tools, finding statistically significant adverse impact does not necessarily mean the tool is

discriminatory or its use legally impermissible. However, such a finding does generally trigger an obligation to understand the drivers behind the analysis and ensure the tool causing the observed disparity is job-related for the position in question and consistent with business necessity. Conversely, a lack of statistically significant adverse impact does not indicate that a tool's use is necessarily appropriate, job-related, or valid in the particular context.

- Vendors and users of AI-enabled employment tools should explicitly address issues of non-statistical fairness. This could take the form of ensuring use of the same type of data, variables/features, procedures, and scoring for all individuals that are being compared to a standard or to each other.

REPORT INTRODUCTION

The use of Artificial Intelligence ("AI") is becoming more prevalent throughout society, including in the workplace. Growing numbers of employers are implementing AI processes across the employment life cycle. The ever-broadening reach of AI in the employment arena raises important questions about the interaction of Equal Employment Opportunity ("EEO") and Diversity, Equity, Inclusion, and Accessibility ("DEI&A") issues and the use of such tools. This interdisciplinary *Report on EEO and DEI&A Considerations in the Use of Artificial Intelligence in Employment Decision Making* (the "Report" or the "AI TAC Report") is intended to address many of these questions.

This Report is the product of an Artificial Intelligence Technical Advisory Committee (the "AI TAC" or the "Committee") that was initiated under the auspices of The Institute for Workplace Equality ("The Institute"),[1] specifically to address EEO and DEI&A considerations in the use of AI in employment decisions, including, *e.g.*, recruitment, hiring, promotions, assignments, performance evaluations, and terminations. In each Section of the Report, the AI TAC Members identify some of the key issues now arising and provide recommendations for navigating this rapidly evolving area in which best practices and regulatory requirements are still being defined.

I. The Background to the AI TAC Committee and the AI TAC Report

This Report is written against the backdrop of a quickly-changing AI landscape. As the use of AI grows across many sectors, governments at many levels, both domestically and abroad, have begun turning their attention to the impact and consequences of that growth. In the United States alone, just in the past two years, there have been significant policy pronouncements and legislation at the federal level to begin to address the real public concerns about privacy, accountability, transparency, fairness, and bias in the use of Artificial Intelligence. Indeed, the White House, Congress,[2] the U.S. Equal Employment Opportunity Commission ("EEOC"), the

[1] The Institute for Workplace Equality is a non-profit employer association that provides training and education to assist companies in understanding affirmative action and equal employment opportunity obligations. The Institute for Workplace Equality's programming addresses a wide range of human resource management strategies to assist employers in creating and maintaining diverse organizations free from workplace bias. https://www.theinstitute4workplaceequality.org/. Not all Members of the AI TAC are members of The Institute.
[2] National Artificial Intelligence Initiative Act of 2020, 15 U.S.C. 9401, *et seq*. (2021).

Federal Trade Commission ("FTC"), and the National Institute of Standards and Technology ("NIST") have each taken steps to address Artificial Intelligence issues.

A. <u>*Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*</u>

In October 2022, the White House Office of Science and Technology Policy ("OSTP") issued a *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People* (the "AI Blueprint").[3] The AI Blueprint establishes a "framework" that

> is intended to support the development of policies and practices that protect civil rights and promote democratic values in the building, deployment, and governance of automated systems.[4]

The AI Blueprint is founded on five principles:

- Safe and Effective Systems
- Algorithmic Discrimination Protections
- Data Privacy
- Notice and Explanation
- Human Alternatives, Considerations and Fallback

Many of the issues raised in the AI Blueprint and its Technical Companion clearly impact the employment sphere. For example, the admonishment that "[d]esigners, developers, and deployers of automated systems should take proactive and continuous measures to protect individuals and communities from algorithmic discrimination and to use and design systems in an equitable way" speaks to, among others, the vendors and employers who design, develop, and use AI tools and processes for employment-related purposes.[5] Similarly, the statement that "[y]ou should know that an automated system is being used and understand how and why it contributes to outcomes that impact you" is applicable to job applicants who participate in a hiring process that uses AI tools and processes.[6] Employers should therefore familiarize themselves with the AI Blueprint and its Technical Companion. However, it should be noted that

[3] Office of Science and Technology Policy, Exec. Office of the President, *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People* ("AI Blueprint") (2022), https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.
[4] *AI Blueprint*, at 2.
[5] *AI Blueprint*, at 5. *See* Report Sections on Data Collection, Uniform Guidelines on Employee Selection Procedures, and Statistics and Adverse Impact, below.
[6] *AI Blueprint*, at 6. *See* Report Sections on Transparency and Fairness and Data Collection, below.

the AI Blueprint and Technical Companion do not constitute binding guidance[7] nor do they mandate that any entity take or refrain from taking any specific actions as regards the use of AI.[8]

B. *The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees*

In May 2022, the EEOC released a technical assistance document entitled *The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job*

---

[7] "The *Blueprint for an AI Bill of Rights* is non-binding and does not constitute U.S. government policy. It does not supersede, modify, or direct an interpretation of any existing statute, regulation, policy, or international instrument. It does not constitute binding guidance for the public or Federal agencies and therefore does not require compliance with the principles described herein." *AI Blueprint*, at 2.

[8] Often, employers can look to court decisions to guide their approach to specific employment issues. But because the use of AI in employment decision making is still relatively new, there have been very few court cases addressing it. The only employment AI cases found in a search of court databases were: *EEOC v. iTutorGroup, Inc., et al.,* Civil Action No. 1:22-cv-02565 (E.D.N.Y., filed May 5, 2022) (Alleging that English-language tutoring company violated the ADEA by using software that solicited birthdates and automatically rejected qualified older applicants); *Deyerler et al. v. HireVue Inc.*, Case No. 2022-CH-00719 (Cir. Ct. Cook Cnty., filed Jan. 27, 2022), *removed as* Civil Action No. 1:22-cv-01284 (N.D.Ill., filed Mar. 10, 2022) (Class action alleging that HireVue failed to obtain written consent for data collection in violation of Illinois's Biometric Information Privacy Act ("BIPA")); *Mobley, et al. v. Facebook*, Case No. 3:16-cv-06440 (N.D. Cal., filed Nov. 3, 2016) (Class action alleging that people of color were discriminated against based on race and national origin in employment, housing, and credit ads; case was settled in 2019); *Riddick v. Facebook*, Case No. 3:18-cv-04529 (N.D. Cal., filed Jul. 26, 2018) (Similar allegations to *Mobley*; also settled in 2019).

Although outside the employment context, employers should take note that this past summer the U.S. Department of Justice ("DOJ") brought its first case challenging use of an allegedly discriminatory algorithm, alleging violations of the Fair Housing Act ("FHA"). *United States of America v. Meta Platforms, Inc.*, Case No. 1:22-cv-05187 (S.D.N.Y., filed June 21, 2022). The complaint in that case – which was settled the same day it was filed – included allegations that Meta "allowed" a machine-learning algorithm "to consider FHA-protected characteristics–including race, religion, and sex – in finding Facebook users who 'look like' the advertiser's source audience and are thus eligible to receive housing ads" and that "Meta's ad delivery system uses machine-learning algorithms that rely in part on FHA-protected characteristics – such as race, national origin and sex – to help determine which subset of an advertiser's targeted audience will actually receive a housing ad." Release, U.S. Dep't of Just., *Justice Department Secures Groundbreaking Settlement Agreement with Meta Platforms, Formerly Known as Facebook, to Resolve Allegations of Discriminatory Advertising* (June 21, 2022), https://www.justice.gov/opa/pr/justice-department-secures-groundbreaking-settlement-agreement-meta-platforms-formerly-known.

Under the terms of the settlement, Meta is required, by December 31, 2022, to stop using its "Special Ad Audience" tool, which the Department alleged relies on a discriminatory algorithm, and to develop a new system for housing ads "to address disparities for race, ethnicity, and sex between advertisers' targeted audiences and the group of Facebook users to whom Facebook's personalization algorithms actually deliver the ads." *Id.* If the Department concludes that the new systems "do not adequately address the discriminatory disparities, the settlement agreement will terminate and the United States will litigate its case against Meta in federal court." *Id.* The terms also provide for an independent reviewer "to investigate and verify on an ongoing basis whether the new system is meeting the compliance standards agreed to by the parties." *Id.* Furthermore, in addition to forbidding Meta from providing targeting options for housing ads "that directly describe or relate to FHA-protected characteristics," the terms also require that Meta "notify the United States if Meta intends to add any targeting options," with the court having the authority to resolve any disputes regarding such new options. *Id.* Finally, Meta is required to pay the maximum penalty allowed under the Fair Housing Act. *Id.*

Applicants and Employees ("the EEOC TAD").[9] This followed the October 2021 launch of the EEOC's *Initiative on Artificial Intelligence and Algorithmic Fairness*, which the agency has used to underscore that "[b]ias in employment arising from the use of algorithms and AI falls squarely within the Commission's priority to address systemic discrimination. While the technology may be evolving, anti-discrimination laws still apply."[10]

The recently-issued EEOC TAD focuses specifically on how the use of AI in the employment context interacts with the Americans with Disabilities Act ("ADA"). The EEOC TAD is very helpful in addressing disability-related questions and we recommend that employers consult it when implementing AI in any aspect of their human resources activities.[11] This Report, however, seeks to provide recommendations on a broader range of EEO and DEI&A issues impacted by the use of AI in employment than is addressed by the EEOC TAD.

C. Federal Trade Commission Guidance

In April 2021, the Federal Trade Commission ("FTC") issued guidance highlighting how it would enforce principles of transparency and fairness with respect to algorithmic decision

---

[9] https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence. The Department of Justice has issued a partner document, *Algorithms, Artificial Intelligence, and Disability Discrimination in Hiring*, https://www.ada.gov/resources/ai-guidance/.

[10] U.S. Equal Emp. Opportunity Comm'n, ("*EEOC") Initiative on Artificial Intelligence and Algorithmic Fairness*, (Oct. 28, 2021), https://www.eeoc.gov/newsroom/eeoc-launches-initiative-artificial-intelligence-and-algorithmic-fairness. The Department of Labor's Office of Federal Contract Compliance Programs ("OFCCP")'s Compliance FAQs state the following regarding employment tools that use AI: "Irrespective of the level of technical sophistication involved, OFCCP analyzes all selection devices for adverse impact. If OFCCP discovers that a contractor's use of an AI-based selection procedure is having an adverse impact at a contractor's establishment, the contractor will be required to validate the selection procedure using an appropriate validation strategy." *Validation of Employee Selection Procedures*, Question 6 (last updated July 23, 2019), https://www.dol.gov/agencies/ofccp/faqs/employee-selection-procedures#Q6.

On September 13, 2022, the EEOC and OFCCP hosted a virtual roundtable with external stakeholders "to discuss the civil rights implications of the use of automated technology systems, including Artificial Intelligence, in the recruitment and hiring of workers." Press Release, EEOC, *READOUT: EEOC and US DOL's OFCCP hosted A.I. and Algorithmic Fairness and HIRE Initiatives roundtable* (Sept. 19, 2022). https://www.eeoc.gov/newsroom/readout-eeoc-and-us-dols-ofccp-hosted-ai-and-algorithmic-fairness-and-hire-initiatives. This event was part of the Hiring Initiative to Reimagine Equity ("HIRE"), a joint initiative of the EEOC and OFCCP, which has as one of its purposes to "[p]romote equity in the use of tech-based hiring systems." EEOC, *Hiring Initiative to Reimagine Equity* (HIRE), https://www.eeoc.gov/hiring-initiative-reimagine-equity-hire.

[11] Employers also should familiarize themselves with the Memorandum issued by the General Counsel of the National Labor Relations Board ("NLRB") on October 31, 2022, entitled *Electronic Monitoring and Algorithmic Management of Employees Interfering with the Exercise of Section 7 Rights.* https://www.nlrb.gov/news-outreach/news-story/nlrb-general-counsel-issues-memo-on-unlawful-electronic-surveillance-and. Among other things, the Memorandum notes that, under existing law, "if employers rely on Artificial Intelligence to screen job applicants or issue discipline, the employer – as well as a third-party software provider – may violate Section 8(a)(3) if the underlying algorithm is making decisions based on employees' protected activities."

making having an impact on consumers, by bringing enforcement actions related to "biased algorithms" under Section 5 of the Fair Trade Commission Act ("FTCA"), the Fair Credit Reporting Act ("FCRA"), and the Equal Credit Opportunity Act ("ECOA").[12] The FTC emphasized the warnings included in that 2021 guidance in a June 16, 2022 report to Congress, cautioning policymakers about the use of AI to combat online problems and urging them to use "great caution" when relying on technological solutions. The June 2022 report also examines concerns regarding bias, discrimination, and surveillance.[13] The chief concerns identified are that AI tools can reflect the biases of developers, lead to unlawful outcomes, and marginalize protected classes in the workplace.

D. *The National Institute of Standards and Technology's Artificial Intelligence Risk Management Framework*

In January 2023, NIST is scheduled to publish an Artificial Intelligence Risk Management Framework ("AI RMF"), as directed by Congress.[14] The NIST AI RMF is intended to improve the ability of organizations to incorporate trustworthiness considerations into the design, development, use, and evaluation of AI products, services, and systems. The framework has been developed over the past year-and-half, in a process involving academia, government, industry, and civil society from the United States and major trading partners from around the world.[15] NIST has also published a companion special publication on managing bias in AI systems.[16]

[12] Federal Trade Commission ("FTC"), Elisa Jillson, *Aiming for truth, fairness and equity in your company's use of AI*, Business Blog (April 19, 2021), https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai.

[13] Press Release, FTC, *FTC Report Warns About Using Artificial Intelligence to Combat Online Problems* (June 16, 2022), https://www.ftc.gov/news-events/news/press-releases/2022/06/ftc-report-warns-about-using-artificial-intelligence-combat-online-problems. Report available at: https://www.ftc.gov/system/files/ftc_gov/pdf/Combatting%20Online%20Harms%20Through%20Innovation%3B%20Federal%20Trade%20Commission%20Report%20to%20Congress.pdf.

[14] H.R. Rep. No. 116-455 (2021); 15 U.S.C. § 278h-1(c).

[15] Two draft versions of the RMF have already been published and NIST has held two workshops to discuss and gather feedback from members of the public. *See* NIST, Dep't of Commerce, *AI Risk Management Framework: AI RMF Development* (last updated Nov. 16, 2022), https://www.nist.gov/itl/ai-risk-management-framework/ai-rmf-development.

[16] *See* Reva Schwartz*, et al.*, *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence, Special Publication 1270*, NIST, at 28 (March 2022) ("NIST Special Publication 1270"), https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf.

The AI RMF outlines seven characteristics of trustworthiness, which organizations put into practice by mapping, measuring, managing, and governing AI risks.[17] These characteristics are:

- Valid and Reliable
- Safe
- Fair – and Bias is Managed
- Secure and Resilient
- Transparent and Accountable
- Explainable and Interpretable
- Privacy-Enhanced

Although economy-wide in scope, NIST's AI RMF will impact how AI is used for employment purposes.[18] Indeed, NIST is encouraging organizations to develop "profiles" to demonstrate best practices in applying the AI RMF to specific use cases and has highlighted hiring as one such example.[19] Although voluntary, previous NIST frameworks on cybersecurity and privacy have influenced policymakers in the U.S. government and technical standards bodies, such as the International Standards Organization.

II. The Creation and Process of the Artificial Intelligence Technical Advisory Committee

A. The AI TAC's Creation

A total of 40 subject matter experts agreed to be part of the AI TAC. Each AI TAC Member has an interest and involvement in the use of AI in employment, whether as practitioners, vendors, academics, or otherwise. The AI TAC includes labor economists, data scientists, industrial and organizational psychologists, researchers, attorneys representing employers and workers, civil society advocates for technology and democracy, vendors who develop and provide AI tools for employment decisions, employers using AI tools for employment decisions, and former Department of Justice, EEOC, and OFCCP officials. A list of AI TAC Members and their biographies can be found in Appendix A.[20]

---

[17] *See* NIST, Dep't of Commerce, *AI Risk Management Framework: Second Draft*, at 10 (Aug. 18, 2022), https://www.nist.gov/system/files/documents/2022/08/18/AI_RMF_2nd_draft.pdf.
[18] NIST, Dep't of Commerce, *Kicking off NIST AI Risk Management Framework*, Workshop #1, Panel 3: AI Risk: Sector Perspectives. https://www.nist.gov/news-events/events/2021/10/kicking-nist-ai-risk-management-framework. *See* NIST, Dep't of Commerce, Summary Analysis of Response to the NIST AI RMF Request for Information. https://www.nist.gov/system/files/documents/2021/10/15/AI%20RMF_RFI%20Summary%20Report.pdf.
[19] NIST, Dep't of Commerce, *AI Risk Management Framework: Second Draft*, at 26 (Aug. 18, 2022), https://www.nist.gov/system/files/documents/2022/08/18/AI_RMF_2nd_draft.pdf.
[20] Over the year-and-a-half of the AI TAC's activities only three Members withdrew from the Committee, due to their concerns about their ability to meet various deadlines.

Once the AI TAC was formed, The Institute asked Victoria A. Lipnic, former Commissioner and former Acting Chair of the Equal Employment Opportunity Commission and partner at Resolution Economics, LLC, to serve as Chair and lead the Committee's efforts. Using their collective years of experience and expertise in the full range of employment-related areas, Chair Lipnic and the AI TAC Members have spent more than a year reviewing, discussing, and analyzing the complicated issues raised by the use of AI-enabled tools in the employment life cycle.

B. The AI TAC's Process

The AI TAC was modeled after a previous Technical Advisory Committee that was brought together with the support of The Institute's predecessor (the Center for Corporate Equality) to address Best Practices in Adverse Impact Analyses. The AI TAC followed a similar process. The AI TAC held its first meeting (virtually) on August 25, 2021. During that meeting, Chair Lipnic discussed the AI TAC's purpose, outlined the requirements and important scoping parameters, explained the AI TAC process, discussed the AI TAC Timeline, and advised on the expectations for an AI TAC Final Report.

1. The AI TAC Survey

Following that first meeting, Chair Lipnic organized a Survey Subcommittee to create an in-depth survey designed to identify and measure AI TAC Members' concerns regarding a range of issues related to EEO and DEI&A and the use of AI tools in employment decision making. The Subcommittee drafted a set of closed-ended questions with content-specific rating scales accompanied by open-ended questions where respondents could provide their input in their own words. In drafting the Survey questions what was apparent was the extensive list of potential AI Survey material. Content for the Survey began to coalesce during November-December 2021 around eight topic areas: AI uses (Recruitment and Selection), AI uses (Other), Data Management, Privacy, Application Development and Documentation, Statistical Issues, Fairness and Equal Employment Opportunity Law, and Communications (among application stakeholders).

While statistics were planned for responses to the closed-ended questions and content analysis for open-ended responses, the Survey responses would not be treated as a final product but rather would be a tool to indicate the direction for further discussion among those knowledgeable for each topic area; it would be a first-cut indication of where there was consensus or controversy on issues. The Survey would also help to identify those who were comfortable with engaging specific topic areas.

In January 2022, a questionnaire was developed by and for Survey Subcommittee members to narrow the range of potential topics. From there, question-writing for the AI TAC Survey proceeded. Besides work on Survey content and the allocation of content to closed- and open-ended questions, the format of rating scales for closed-ended questions was developed. In March 2022, a 91-item Survey (including respondent demographics items) went out to the full AI TAC. The Survey Subcommittee then analyzed the data. Results were tabulated for the closed-ended questions and themes identified for the open-ended questions. Full results were provided to the AI TAC in May, followed by an online briefing.[21]

The results of the Survey were complex. There were many open-ended questions, but generally there were not many questions where multiple respondents seemed focused on a single concern within that question. This indicated that these responses were more for guiding further discussion than for making conclusions. The closed-end questions were amenable to counts of responses, but here also there was ambiguity. A basic concern in interpreting results was that the AI TAC by design has diverse expertise; a number of respondents chose not to answer those questions where they did not consider themselves to be experts. While some questions pointed to a consensus among those who responded, others showed divergence, with the total number of respondents being small and thus making conclusions tenuous regarding that divergence.

Rather than the Survey Subcommittee imposing its interpretation of the results, the better course seemed to be to refer Survey topic areas to those AI TAC Members best fitted to consider them in depth. Once that was done, the Survey was deemed to have made its contribution and the AI TAC began the process of producing a final product.

2. Work by Area-Focused Subcommittees

Chair Lipnic created five Subcommittees based on the topic areas identified in the Survey:

1. Uses and Applications
2. Transparency and Fairness
3. Data Collection

[21] The Survey questions and responses can be found here: https://www.theinstitute4workplaceequality.org/ai-tac-report-release

4. Uniform Guidelines on Employee Selection Procedures ("UGESP")
5. Statistics and Adverse Impact

Chair Lipnic held initial meetings with each Subcommittee in July 2022 to discuss the parameters of the Subcommittee's area of focus. Over the next three months, the Subcommittees carefully examined the points of agreement among Survey respondents, results of deliberations from other organizations[22] relevant to the use of AI in employment, and documents from other sources.[23] Each of the Subcommittees held a series of meetings and discussions during which its Members discussed the specific EEO and DEI&A issues arising within their focus area and explored, analyzed, and debated the best ways to address those issues. Once each Subcommittee had come to general agreement on the key discussion points and recommendations to be made in that regard, Subcommittee Members prepared an initial draft of their Report Section.

3. Final Committee Meeting and Creation of the AI TAC Report

On September 21, 2022, the full AI TAC met in person for a day-long session in Washington, D.C.[24] During that session each Subcommittee shared an initial draft of its Report Section, which was then discussed by each of the other Subcommittees and by the Committee as a whole. Thereafter each of the Subcommittees continued to meet independently and consider the feedback from the wider AI TAC. Once the Members of each Subcommittee reached agreement on the contents of their Subcommittee's Report Section, they finalized the writing of that Section. The Editorial Committee then edited those Sections and Appendices into a full Report.

III. The Final AI TAC Report

This Report follows the topic area structure of the AI TAC and its Subcommittees. In the Sections below the reader will find discussion, analysis, and recommendations addressing the interaction of EEO and DEI&A concerns and the use of AI-enabled processes in employment decision making in the focus areas of Uses and Applications; Transparency and Fairness; Data

---

[22] *E.g.*, Society for Industrial and Organizational Psychologists ("SIOP") *Principles*, SIOP Statement on Artificial Intelligence, American Psychological Association ("APA") Testing Standards, APA Ethical Standards, Association of Test Publishers, Organization for Economic Cooperation and Development ("OECD"), American Statistical Association ("ASA") Ethical Standards, Institute of Electrical and Electronics Engineers ("IEEE") Ethically Aligned Design Glossary. Appendix B contains references for further exploration of these issues.

[23] *E.g.*, Uniform Guidelines on Employee Selection Procedures ("UGESP"), legislative actions, state and local regulations, EEO enforcement agency statements, National Institute of Standards and Technology ("NIST") Plan for Developing AI Technical Standards.

[24] Due to concerns regarding COVID-19 a virtual option was also provided and this option was used by some Members of the AI TAC.

Collection; the Uniform Guidelines for Employee Selection Procedures ("UGESP"); and Statistics and Adverse Impact. Not surprisingly, there is some overlap between the work of the different Subcommittees. For example, issues of data reliability arise not only in the context of Data Collection but also in the contexts of Transparency and Fairness and Statistics and Adverse Impact. Similarly, while UGESP is the primary focus of the Subcommittee that bears its name, it is also relevant to the work of the Statistics and Adverse Impact Subcommittee. There is no way to discuss these issues without some level of interconnectedness.

A. Report Recommendations

This Report brings together experts in a range of disciplines and professions. While all AI TAC Members agree that the use of AI in employment decision making must take into account EEO and DEI&A issues, there are different perspectives among Members as to how that is best done in every instance.

The Members of each AI TAC Subcommittee worked to develop recommendations within that Subcommittee's area of focus and each Subcommittee worked separately to prepare its particular Section of the Report. Thus, the discussion points and recommendations in each Section of this Report generally reflect the views of the Subcommittee Members who prepared that Section. They do not necessarily reflect the views of all AI TAC Members or even of every Subcommittee Member with respect to every particular in a given Section.

The views and recommendations in this Report should not be interpreted as reflecting official positions of the professional organizations, companies, or firms with which AI TAC Members may be affiliated.

Moreover, none of the comments or recommendations included in this Report should be interpreted as asserting legal doctrine, legal advice, or mandated actions by employers using or developers creating AI-enabled selection procedures. Rather, the discussion points and conclusions are intended to provide guidance and inform the continuing public discourse as various stakeholders address the complex and difficult EEO and DEI&A issues that arise with this developing technology.

B. Defining Report Terms

As discussed above, AI TAC Members come from a broad range of disciplines. Each discipline has its own lexicon. Where a Report Section applies a concept or term in a discipline-specific or otherwise context-specific manner, notice of that is made within the Section. However, there are some basic terms used throughout the Report that bear defining from the start:

Artificial Intelligence: Both the trade press and mainstream media label a wide range of employment selection processes as "AI" when many, in fact, are not. This Report adopts the approach taken in the EEOC TAD. The TAD explains that:

> In the employment context, using AI has typically meant that the developer relies partly on the computer's own analysis of data to determine which criteria to use when making employment decisions. AI may include machine learning, computer vision, natural language processing and understanding, intelligent decision support systems, and autonomous systems. [25]

Machine Learning: This Report uses the definition of "machine learning" ("ML") set forth in the National Artificial Intelligence Initiative Act of 2020:

> The term "machine learning" means an application of Artificial Intelligence that is characterized by providing systems the ability to automatically learn and improve on the basis of data or experience, without being explicitly programmed.[26]

Algorithm: This Report uses the EEOC's definition of "algorithm":

> Generally, an 'algorithm' is a set of instructions that can be followed by a computer to accomplish some end. Human resources software and applications use algorithms to allow employers to process data to evaluate, rate, and make other decisions about job applicants and employees. Software or applications that include algorithmic decision-making tools may be used at various stages of employment, including hiring, performance evaluation, promotion, and termination.[27]

Dynamic data: As used in this Report, this means that when compared with traditional methods, the flow of applicant data can be both wider (applicant big data, through applying via LinkedIn,

---

[25] EEOC TAD (definition of "Artificial Intelligence" in "Background" section). The EEOC TAD notes that the National Artificial Intelligence Initiative Act of 2020 defines Artificial Intelligence as "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments."

[26] 15 U.S.C § 9401(3). A broader discussion of AI and machine learning is set forth in the NIST Special Publication 1270.

[27] EEOC TAD ("Background" section).

Indeed, *etc*.) and deeper (more variables, from big data based on automated resume screening, interviews with avatars, interactive games, *etc*.).

Dynamic algorithms: As used in this Report, this means that ML models might be frequently retrained based on dynamic data flowing through an organization (versus a more static application process that changes less frequently).

Applicant: This Report generally adopts the EEOC's approach to the concept of an "applicant":

> The precise definition of the term "applicant" depends upon the user's recruitment and selection procedures. The concept of an applicant is that of a person who has indicated an interest in being considered for hiring, promotion, or other employment opportunities. This interest might be expressed by completing an application form, or might be expressed orally, depending upon the employer's practice.[28]

The OFCCP has a definition of "internet applicant" that may be relevant to some employers. For OFCCP purposes,

> An "Internet Applicant" is an individual who satisfies all four of the following criteria:
>
> - The individual submitted an expression of interest in employment through the internet or related electronic data technologies;
> - The contractor considered the individual for employment in a particular position;
> - The individual's expression of interest indicated that the individual possesses the basic qualifications for the position; and
> - The individual, at no point in the contractor's selection process prior to receiving an offer of employment from the contractor, removed himself or herself from further consideration or otherwise indicated that he/she was no longer interested in the position.[29]

Finally, when reading this Report, it is important to keep in mind that certain terms may have a technical meaning (for example, in the industrial-organizational psychology literature) that can differ from the meaning and use of those same terms in case law applying Title VII of the Civil Rights Act of 1964. "Bias" is a prime example. While the term has particular meaning(s) in the

---

[28] U.S. Equal Emp. Opportunity Comm'n ("EEOC"), *Questions and Answers to Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures* ("Q&As"), No. 15, https://www.eeoc.gov/laws/guidance/questions-and-answers-clarify-and-provide-common-interpretation-uniform-guidelines; 44 Fed. Reg. 11996 (Mar. 2, 1979) and 45 Fed. Reg. 29530 (May 2, 1980).
[29] 41 C.F.R. § 60-1.3.

academic and scientific literature, a finding of "bias" in the technical sense may not translate into a finding of "bias" under Title VII. Differences in pass rates by race, gender, or some other characteristic might be deemed "bias" in the former sense, for example, but might not support a finding of "bias" in the legal sense if the other requirements of Title VII (*e.g.,* job-relatedness, no less discriminatory alternative) are established. "Adverse impact" is another example where the facts that support such a finding in the technical sense may differ from the facts that would support a legal determination of disparate impact under Title VII.

Courts have commented on this challenge. For example, the Second Circuit's 1980 decision in *Guardians Association of New York City Police Department, Inc. v. Civil Service of Commission of the City of New York* contains an extended discussion of the relationship between technical test validation and Title VII:

> The study of employment testing, although it has necessarily been adopted by the law as a result of Title VII and related statutes, is not primarily a legal subject. It is part of the general field of educational and industrial psychology, and possesses its own methodology, its own body of research, its own experts, and its own terminology. ***The translation of a technical study such as this into a set of legal principles requires a clear awareness of the limits of both testing and law.*** It would be entirely inappropriate for the law to ignore what has been learned about employment testing in assessing the validity of these tests. At the same time, the science of testing is not as precise as physics or chemistry, nor its conclusions as provable. While courts should draw upon the findings of experts in the field of testing, they should not hesitate to subject these findings to both the scrutiny of reason and the guidance of Congressional intent.[30]

We encourage readers of this Report to attend to these possible distinctions.

[30] 630 F.2d 79, 89 (2d Cir. 1980) (emphasis added).

USES AND APPLICATIONS

I. Introduction

This Section summarizes the most common current uses of Artificial Intelligence ("AI") in the employment context. It also provides some examples of specific ways different types of AI-enabled tools may be applied in that context, to help the reader better understand the contours of AI use in the workplace.

As attention to AI has increased, there has been an eagerness to apply the label of "Artificial Intelligence" to employment selection processes that do not in fact involve AI. Thus, for example, it is not uncommon to see the AI label affixed to tools that do nothing more than automate some aspect of the selection process. But mere automation of a process does not mean the process is guided by "Artificial Intelligence." Members of the AI TAC think it is important to recognize this distinction between automation and Artificial Intelligence. As can be seen below, however, some inquiries (*e.g.*, surveys) into employers' use of so-called "AI tools" do not do so, instead combining in their questions use of both AI and automation. The information provided in this Section should be understood within that context.

II. How Employers Are Using AI

Employers are increasingly using AI-enabled tools and processes for many employment-related activities. These include everything from initial recruitment and hiring, to identification of skills gaps and responsive training, job assignments, identification of candidates for promotion or other job movements, and identification of employees that are flight risks and the determination of additional rewards or other employment enhancements to retain such talent. Surveys done in the past few years underscore both the increasing use of AI-enabled tools and processes in the employment context and the desire of employers to make ever greater use of them. The Subcommittee did not attempt to ascertain each and every use of AI by any and all employers when making employment decisions. Rather, we identify here the most common uses of AI by employers.

A. Sourcing and Hiring

The most prevalent use of AI in the employment context today appears to be sourcing candidates and making hiring decisions. Employers of all sizes are inundated with expressions of interest from job seekers for open positions; many are also searching available data sources for talented individuals who have not yet expressed any interest in the employer or an open position. They have turned to various AI products, either licensed from vendors or developed internally, to efficiently source and process those candidates, find the best talent, and make speedy offer decisions.

A survey of over 1,500 human resources professionals done earlier this year[31] found:

- 79% use AI and/or automation for recruitment and hiring
- 25% plan to start using or increase their use of AI and/or automation in recruitment and hiring.

In a survey done three years ago (2019), only 56% of human resources professionals said they were then using AI "for the purpose of talent acquisition." [32]

Employers today use AI for a range of purposes within the recruitment and hiring sphere. According to the 2022 SHRM Survey:

- 52 % use AI and/or automation to automate candidate searches
- 64% use AI and/or automation to review or screen applicant resumes
- 25% use AI and/or automation to pre-select applicants for interviews
- 22% use AI and/or automation to administer and score applicant skills assessments

B. Managing and Retaining Current Employees

Beyond the hiring process, employers are increasingly using AI-related tools to manage the performance of employees once they are brought on-board:

- 38% currently use AI and/or automation for performance management[33]
- 20% plan to start using or increase their use of AI and/or automation for that purpose.[34]

---

[31] Society for Human Resources Management, *Automation and AI in HR* (2022) ("2022 SHRM Survey"), https://advocacy.shrm.org/wp-content/uploads/2022/04/SHRM-2022-Automation-AI-Research.pdf.
[32] Oracle Corporation, *2019 State of Artificial Intelligence in Talent Acquisition* (2019), https://www.oracle.com/a/ocom/docs/artificial-intelligence-in-talent-acquisition.pdf.
[33] 2022 SHRM Survey
[34] Id.

Employers are also looking to AI to help them retain employees. According to a recent Mercer study, close to 90% of companies already have or plan to adopt an AI-enabled internal talent marketplace platform as a way to address concerns that reskilled/upskilled talent will otherwise leave.[35]

III. Examples of AI Application in the Employment Context

A. Sourcing Candidates and Making Hiring Decisions

The following examples illustrate the use of AI in the sourcing and hiring processes.

Sourcing/Pre-Screen

A very typical scenario in looking for talent or pre-screening talent for positions is the use of AI (in this case, Machine Learning ("ML")) to evaluate and auto-score resumes for positions. These resume screening tools are typically used as a sourcing tool to identify candidates that might potentially match the requirements of a job opening. Potential candidates are identified by evaluating the current resume database within an organization's applicant database (applicant tracking system) or by scraping social media sources such as LinkedIn. The same ML technology and algorithms can also be used to score resume submissions for a given requisition[36] or job opening. In this scenario the AI technology is usually a type of machine learning like natural language processing ("NLP"). These NLP technologies analyze the text within the resume.

Pre-Hire Evaluation

The use of AI in screening or pre-hire evaluation for larger volume jobs is becoming a normal step in the process when filtering or profiling applicants for a given role. A typical technology used is video interviews and/or novel assessments like game-based tests. Sometimes these newer tools are used independently or together in the selection process and with less frequency they are

[35] Mercer, *The Rise of the Relatable Organization*, 2022 Global Talent Trends Study (2022), https://www.mercer.com/content/dam/mercer/attachments/private/global-talent-trends/2022/gl-2022-global-talent-trends-report-eng.pdf.

[36] A requisition is a request to fill a job – permission to start the hiring process. Lin Grensing-Pophal, *Job Description or Job Requisition: Which Comes First?* (Apr. 24, 2019), https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/job-description-job-requisition-which-comes-first.aspx#:~:text=%5BSHRM%20members%2Donly%20resource%3A,applicants%20will%20need%20to%20possess.

being used in conjunction with more traditional tests (*e.g.,* multiple choice or Likert-style[37] question types). Scoring of video interviews (one-way or live) uses a variety of machine learning technology (*e.g.*, deep learning) from the transcription of the recorded interview answer to a text file, to NLP methods to analyze the text response data. For game-based assessments, the pattern of responses in the game are analyzed with ML to generate scores.

B. <u>Other Uses of AI in the Employment Life Cycle</u>

As noted above, employer use of AI is not limited to the sourcing and hiring processes. It is used in various processes throughout the employment life cycle, such as job assignment, promotion, and pay determination. Most of these processes involve selection decisions, many of which are currently facilitated by AI. Thus, many of the legal, ethical, and scientific considerations associated with use of AI in the sourcing and hiring processes apply as well to the selection decisions made later in the employment life cycle. The following example illustrates the use of AI in the context of later selection decisions.

Internal Mobility:

Employers are using AI-enabled tools to create employee profiles from the wide variety of data on their employee populations. These ML applications are evaluating many employee databases (resumes, skills, job rotations, development courses completed, and job performance reviews) to create detailed employee taxonomies. Organizations are looking to use these taxonomies for various uses, such as to match against open positions, identify skill gaps or development needs, add to promotion benches, and calculate "At Risk for Turnover" scores. For example, employers are using predictive attrition models, based on these taxonomies, to direct rewards or other enhancements to the talent deemed to be a flight risk.

****

In the Sections below, each of the AI TAC Subcommittees has sought to address some of the key EEO and DEI&A issues that arise from the growing use of AI tools and processes in these types of contexts.

---

[37] The Likert-style was developed in 1932 by Rensis Likert to measure attitudes and behaviors. The typical Likert scale is a 5 or 7-point scale used to rate the degree to which responders agree or disagree with a statement. Rensis Likert, *A technique for the measurement of attitudes*. Arch Psychology (1932).

TRANSPARENCY AND FAIRNESS

I. Introduction

This Section reviews and provides recommendations regarding critical issues of transparency and fairness that arise with the use of AI-enabled employment selection tools and processes.

As a starting point, the Subcommittee does not take the position that the use of AI-enabled selection procedures for employment decision making is unethical *per se*.[38] Rather, it is our contention that like any other technological innovation, such selection procedures should be implemented carefully and with consideration of any unfair treatment that may result from their use, either to individual persons or to protected group members. People have inherent rights to be treated with dignity and respect, to have their privacy protected, to have agency over their personal information, to be treated fairly and equitably, and (in the United States and many other countries) to be free from discrimination based on certain group memberships, such as race, ethnicity, gender, religion, age, and disability. Accordingly, we have identified certain aspects of AI-enabled selection procedures that present concerns which must be highlighted and addressed within an ethical framework. Few of these issues are so clear as to allow for unnuanced rules or guidance. However, careful consideration, advance planning, and balancing of the needs and rights of multiple stakeholders can improve the development and deployment of AI-enabled selection procedures to benefit both users and participants - to be both fair to applicants and workable for employers and developers.

II. Transparency, Notice, and Consent Issues for AI Processes

AI techniques allow developers to detect patterns in large unstructured datasets, without specifying a hypothesis or building a theoretical model. This enhanced analytic capability can make use of datasets and data types that were traditionally unusable or non-valuable for employment decision purposes. This process may include the use of data that has been collected from secondary sources, such as data shared by job seekers on social media for purposes other than employment selection, or in a manner that is less obvious, such as response time or cursor location while playing a game. Unlike data intentionally supplied by a job seeker for the purposes of employee selection (*e.g.*, responses to interview questions), the use of secondary or

---

[38] Anna Lena Hunkenschroer and Alexander Kriebitz, "*Is AI recruiting (un)ethical? A human rights perspective on the use of AI for hiring*." AI and Ethics, 1-15. (Jul. 25, 2022) ("Hunkenschroer and Kriebitz").

less obvious data sources, and how those sources impact the AI-enabled selection tool, may not be readily apparent to an applicant or employee. As a result, the issues of transparency, notice, and consent are essential to address.

There are many regulations associated with the use of employment selection procedures; at present, however, there is no federal law in the United States specifically aimed at regulating the use of AI in hiring/selection decisions,[39] nor addressing the nuances associated with transparency, notice, and consent. Emerging state and local laws offer a patchwork of requirements in this area that are evolving and can be complicated. While we cannot and do not attempt to address all of the still-developing legal requirements in this area, in order to encourage the ethical use of AI the Subcommittee strongly recommends that transparency and consent be hallmarks of any selection procedure that harnesses AI.

A. Transparency Considerations

With the introduction of AI in the hiring/selection process – and specifically with the ability of employers to incorporate vast amounts of secondary and less obvious data sources – it may be very difficult for applicants to discern the criteria upon which they are being evaluated and whether the evaluation is done through AI or by a human decision maker. For this reason, as an ethical matter, employers should be transparent about when AI is part of a specific selection procedure or is used in the overall selection process.

While many would likely agree that transparency is important, the level of transparency may be debated:

- Do employers simply need to disclose that Artificial Intelligence is in use?
- Is a general understanding of the data being used and the key criteria being evaluated sufficient?
- Must deep details related to AI be provided (*e.g.,* how algorithms generate scores, how scores are shown to be job-relevant and predict outcomes)?
    - If yes, to whom is this greater level of transparency owed and/or beneficial?
    - Should different levels of transparency be conveyed to applicants, employers, and auditors?

[39] It is important to note that the Office of Federal Contract Compliance Programs ("OFCCP"), an agency within the U.S. Department of Labor, has said it will investigate federal contractor use of AI-enabled employment selection tools in the same way as traditional tools. *See* U.S. Dep't of Labor, Office of Federal Contract Compliance Programs, *Validation of Employee Selection Procedures,* Question 6 (last updated July 23, 2019), https://www.dol.gov/agencies/ofccp/faqs/employee-selection-procedures#Q6.

- How can vendors be transparent while simultaneously protecting their intellectual property and business interests?
- Depending on the specific AI methods used, is it even possible to trace or explain how a specific decision was made based on the most advanced, complex algorithms (*e.g.*, deep learning)?

1. *The Legal Landscape*

To consider these important questions and provide recommendations on the level of transparency required for different stakeholders, it is important to take note of the current legal landscape. There is at present no specific federal requirement for employers to provide notice or information to applicants or to obtain consent prior to the use of AI for hiring/selection. However, the EEOC TAD suggests that employers provide applicants and/or employees "as much information about the tool as possible, including information about which traits or characteristics the tool is designed to measure, the methods by which those traits or characteristics are to be measured, and the disabilities, if any, that might potentially lower the assessment results or cause screen out."[40]

In the absence of binding federal requirements, states and localities have begun filling the void. Recent laws passed in Illinois, Maryland, and New York City create requirements to provide notice to applicants when AI is used. However, the laws vary with respect to the types of selection procedures covered, when the requirements trigger, and the types of notice and/or consent required:

- The Illinois Artificial Intelligence Video Interview Act, effective January 1, 2020, requires an employer to:
  - provide notice to applicants that it will use AI to analyze video recorded interviews;
  - provide information explaining how the AI works and the types of characteristics it uses to evaluate applicants; and
  - obtain consent from the applicant to be evaluated using AI.[41]

- Maryland House Bill 1202, effective October 1, 2020, prohibits the use of facial recognition technology during an applicant's interview for employment without *written* consent.[42]

[40] U.S. Equal Emp. Opportunity Comm'n ("EEOC"), *The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees* (May 12, 2022) (the "EEOC TAD") ("Algorithmic Decision-Making Tools That Screen Out Qualified Individuals with Disabilities" section, Question 12), https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence. The AI Blueprint's "Notice and Explanation" Principle also speaks to these issues.
[41] *See* Section 820 ILCS 42/5.
[42] Md. Code Ann., Lab. & Empl. § 3-717.

- In New York City (effective January 1, 2023),[43] it will be unlawful for employers to use "automated employment decision tools" ("AEDTs") unless certain requirements are met:[44]
    - Before such a tool is used a "bias audit" must be performed "to assess the tool's disparate impact" and a summary of the audit's results must be made public on the employer's website;
    - Audits must be done annually;
    - Applicants[45] must be informed ten days prior to the use of an AEDT that such a tool is being used so that they may request an alternative selection process or accommodation; and
    - Applicants must be provided notice of the job qualifications/characteristics evaluated by the tool as well as information on the type and source of data to be used.

The New York City law does not, however, require explicit consent from applicants.

Employers can also look to the European Union ("EU") for guidance. The EU tackled the issue of transparency in its General Data Protection Regulations ("GDPR") and established a "right to explanation" about data being collected and used. The GDPR provides that entities using automated decision-making systems "ensure fair and transparent processing [by providing] meaningful information about the logic involved."[46] There is much debate on what, specifically, the "right to explanation" requires. However, for the purpose of employment decisions, it has been described as providing an understanding of how an applicant/employee is evaluated using AI.[47]

III. Recommended Approaches to Transparency and Notice

There is consensus among the AI TAC Members that it is critical to ensure transparency around crucial decisions such as employment-related selections. The use of Artificial Intelligence tools in the selection process is no different in this regard. Where secondary data sources (*e.g.,* social media) or less obvious data sources (*e.g.,* click rates) are utilized by AI-enabled selection tools,

---

[43] On December 9, 2022, the NYC Department of Worker and Consumer Protection ("NYC DWCP") announced that "due to the high volume of public comments" on the proposed regulations to implement the new AEDT law (*see* FN 45, below), the NYC DWCP will not enforce the new law until April 15, 2023. 2021 N.Y.C. Local Law No. 144, https://www.nyc.gov/site/dca/about/new-laws-rules.page#2023.

[44] NYC Admin Code § 20-870. As of the date of this Report, the NYC DWCP has issued Proposed Rules to implement the AEDT law. N.Y. Comp. Codes R. & Regs., tit. 6, § 5-300, et seq. (proposed Sept. 23, 2022), https://rules.cityofnewyork.us/wp-content/uploads/2022/09/DCWP-NOH-AEDTs-1.pdf.

[45] The New York City law applies to hiring and promotion and covers candidates and employees.

[46] General Data Protection Regulation, (EU) 2016/679, art. 13, 2 (2016).

[47] Hunkenschroer and Kriebitz, 2022.

enhanced transparency measures may be required. Different levels of transparency should be applied based on the intended audience. Further, the method and timing of the notice must be considered.

A. Transparency from Vendor to Employer

Because employers are potentially liable for the selection procedures used as part of their hiring process, employers require the highest level of transparency from the vendors of AI tools. This does not mean that every person working with the selection procedure requires the same level of detail about the tool. However, there should be a subset of a business's employees who would be considered experts in the selection procedure, and a high level of detail should be provided to these experts as well as to the employer's legal counsel. To this end, vendors should provide employer experts with a very clear understanding of the data collected, the source of the data, how the data are used, trained and/or analyzed, and how ratings, scores, or recommendations are reached. As discussed further in the Data Collection Section of this Report, sufficient detail should be provided to enable the employer to understand and have confidence in the results and to meet any legal requirements.[48]

To be clear, this recommendation is consistent with what vendors of traditional selection procedures have been (or should be) providing to employers and is not intended to represent heightened obligations specific to AI-enabled selection procedures. However, it is recognized that providing this level of transparency is typically much simpler to achieve with traditional selection procedures than may be the case with those using Artificial Intelligence. As a practical matter, employers require sufficiently detailed information and documentation about a selection procedure to determine if it is serving its intended purpose (*e.g.*, predicting successful performance, retention, absenteeism) and to evaluate the risk and legal defensibility of the selection procedure in the context of the employer's hiring process. Employers also must have access to sufficient documentation to comply with any legal requirements, such as the UGESP requirement to include source data in validity studies of any selection procedure.[49]

---

[48] *Id.*; *see also* Data Collection Section of this Report (discussing the need for computational reproducibility).
[49] 29 C.F.R. § 1607.15(B)(11).

B. Transparency from Employer to Applicant

The same level of detail that is provided to employer experts does not need to be provided to applicants and doing so could be considered a breach of information that would undermine the effectiveness of the selection procedure. Rather, applicants should be notified that an AI-enabled selection procedure is being used, and they should be provided information that allows them to understand the general criteria being evaluated by AI. This level of detail should be sufficient to allow applicants to understand how they are being evaluated and decide whether to proceed in the process, withdraw from consideration, request a reasonable accommodation, and so on.

Since individualized advance notice may be impractical for employers, giving all applicants notice by placing information in the job announcement or on the application website may be sufficient unless local regulations require otherwise. Because some regulations have specified timeframes for notice, such as at least 10 days before the AI-enabled selection procedure is used, providing such notice may require an employer to delay processing of applicants who will be evaluated using the procedure. In some situations, such a delay may result in losing qualified applicants to other employers. Yet, providing no notice to applicants is equally problematic. As a practical matter, the Subcommittee recommends an approach that balances applicants' need to receive this information in a timely manner such that they can digest what is involved and make key decisions on how/whether they wish to proceed in the hiring process with the employer's need to fill open roles without undue delay.

C. Transparency Applied

Hypothetical:

> In a pre-hire evaluation, the employer uses a video interview scored using a variety of machine learning technologies ranging from transcription of the recorded interview to Natural Language Processing ("NLP") methods to analyze the interview responses. For the purpose of this hypothetical, we will assume that the algorithm rates applicants on their substantive responses, word choice and articulation. Our hypothetical applicant is an individual whose first language is not English and who speaks with an accent and generally uses simple vocabulary.

In this scenario, the employer should provide on the application website a description of how the interview will be conducted and what characteristics or attributes will be evaluated as part of the video recorded interview; the interview can then be scheduled for a subsequent date and time to allow for the advance notice if it is required. This procedure may allow the applicant to better prepare, to decide to forego the interview, or to request an accommodation.

Other stakeholders, such as personnel involved in the hiring process, should also be provided information about the tool. These front-line decision makers are usually responsible for system implementation, addressing applicants' questions and concerns, and using the system in making hiring decisions. If hiring managers and/or human resources personnel do not understand how a selection procedure works, they will be unable to perform these functions, evaluate requests for accommodations, or feel comfortable relying on the results.

If a selection procedure is legally challenged, yet another level of transparency may be required. Although this Section does not address scenarios involving requests for information during an investigation or audit or discovery during litigation, employers are urged to ensure through contractual agreements that such documentation resides with the employing organization or that vendors (and their successor organizations) will be cooperative in providing the necessary information in these scenarios.

While the appropriate level of transparency may vary based on the intended audience, in all cases the information provided should be sufficient to allow the target stakeholders (employers, human resources personnel, hiring managers, applicants) to make informed choices (*e.g.*, for employers, the choice to adopt the selection procedure for operational use; for applicants, the choice to participate in the selection process).[50] Employers adopting AI-enabled selection procedures should also consider the timing of such notice to applicants/employees so that individuals receive sufficient advance notice to permit requests for reasonable accommodation and to make informed decisions about their participation in the selection process. This approach should allow a vendor to manage the release of information regarding its selection procedure to sufficiently protect any intellectual property or business interests.

[50] Organization for Economic Co-operation and Development ("OECD"), *Recommendations of the Council on Artificial Intelligence*, OECD/LEGAL/0449 (Adopted, May 21, 2019), https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

IV. Recommended Approach to Consent[51]

Despite the current lack of wide-reaching and/or consistent legal requirements related to consent, ethical considerations should be carefully considered by users of AI-enabled selection procedures. Although the American Psychological Association ("APA") ethical standards discuss implied consent to employment testing (predicated on voluntary participation in the hiring process),[52] the use of AI-enabled selection procedures introduces concerns about the viability of implied consent when the selection procedures are less obvious or incorporate data from secondary sources that are not specifically related to the selection process, such as social media. The ethical guidelines of the American Statistical Association ("ASA") acknowledge that a requirement for consent may not apply but nonetheless require members to "[u]se data only as permitted by data subjects' consent, when applicable, or consider their interests and welfare when consent is not required."[53] The standard goes on to recognize and include both primary and secondary uses of data, repurposed data, and linkage of data to other datasets, all of which are frequently incorporated in AI-enabled systems based on machine learning algorithms.[54]

Most applicants who need jobs are unlikely to withdraw their applications because an AI-enabled selection procedure is used by the employer. However, the practice of transparency combined with obtaining consent allows an applicant who does not want this information gathered an opportunity to withdraw from consideration or to appeal the result of selection decisions.[55]

There is disagreement among the Subcommittee Members on whether implied consent is sufficient (*e.g.*, once notice is given, consent is implied if the applicant chooses to move forward in the hiring process) or if explicit consent is necessary. However, the Subcommittee recommends that employers not only provide clear notice to applicants that AI is being used and how it is being used, but also obtain consent from applicants.

---

[51] Further discussion of issues related to consent can be found in the Data Collection Section, below.

[52] *Ethical principles of psychologists and code of conduct* (Am. Psych. Ass'n 2002, amended effective June 1, 2010, and January 1, 2017), http://www.apa.org/ethics/code/index.html.

[53] *Ethical Guidelines for Statistical Practice* § D.5. (Am. Stat. Ass'n 2022), https://www.amstat.org/docs/default-source/amstat-documents/ethicalguidelines.pdf?Status=Master&sfvrsn=bdeeafdd_6/.

[54] It should be noted that obtaining consent does not shield an employer from a legal claim of disparate impact as applicants and employees may not consent to an employer violating the law.

[55] OECD, *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, OECD/LEGAL/0188 (2022) (*See* "Individual Participation Principle," at No. 13), https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0188.

V. Privacy Issues for AI Processes

As a general principle, individuals should be able to maintain their personal privacy except for those things that they choose to make available to others or that are in the public domain. However, as discussed above, some AI-enabled selection procedures allow developers and employers to incorporate secondary and unstructured data sources via the processing of large amounts of information at high levels of power and speed. To the extent applicants are unaware of the data utilized to make consequential decisions about them, AI may magnify the impact of such procedures on individual privacy interests as large amounts of personal information may be rapidly processed through these AI-enabled selection procedures. Privacy concerns also may be magnified where large amounts of data that are non-job relevant or originally intended for a purpose other than employment selection are utilized in an AI-enabled tool, even if the applicant has some awareness of the use of such data.

In the United States, there currently are no specific federal laws that protect individual privacy rights when AI-enabled selection procedures are used.[56] Some states and localities, however, have passed legislation to address privacy concerns in this area. Examples include:

- The Illinois Artificial Intelligence Video Interview Act limits the sharing of video-recorded interviews and requires that video files be destroyed upon an applicant's request.[57]
- As of January 1, 2023, under California's Privacy Rights Act ("CPRA"), employers will be required to inform applicants and employees that personal information is being collected, how the information will be used, and to whom such information will be disclosed. Applicants and employees will be able obtain their personal information, delete or correct it, and opt out of its sale and use across various business platforms.[58]
- Other jurisdictions (*e.g.*, New York City, Maryland, Washington, D.C.) have either passed or are considering similar legal safeguards.

Generally, privacy laws in the United States are based on a model of notice and consent. However, for employers whose businesses require international presence, in particular in the EU, other, generally stricter, regulations may apply.[59]

---

[56] *But see AI Blueprint*, which identifies "Data Privacy" as one of its Five Principles.
[57] 820 Ill. Comp. Stat. Ann. 42/15.
[58] Cal. Civ. Code § 1798.140.
[59] GDPR, art. 13-15.

VI. Recommended Approach to Privacy

The Subcommittee recommends that users of AI-enabled selection procedures take steps to evaluate these privacy considerations and establish protocols to provide appropriate notice of possible uses of collected information and obtain informed consent. Given recent developments, it may be best for users of such selection procedures to provide privacy notices to applicants and employees about the information gathered and explain how such information will be used initially and later.

While notice and consent are important hallmarks of protecting privacy, the very nature of AI also requires a focus on how data are collected, used, stored, and shared.[60] Employers and assessment tool developers/vendors should develop written policies and procedures specifying and documenting how data are handled with particular emphasis on potential personally identifying data and those data that are not obtained directly from applicants (*e.g.*, social media, credit information, educational records). It is recognized that employers have long sought and used information not directly supplied by applicants (*e.g.*, reference requests, background checks). In some cases, such as with credit checks, employers in the United States already must notify an individual when a negative decision is reached on the basis of such reports.[61] The use of AI-enabled assessments may heighten concern regarding obtaining, storing, and using such secondary information, particularly from sources such as social media, and should be prepared to respond to questions and challenges and to ensure appropriate levels of security and confidentiality.

To ensure that the lines of responsibility are clearly delineated, the Subcommittee recommends that contractual arrangements between vendors/developers and users clearly specify:

- which entity "owns" applicant data;
- where those data will reside; and
- responsibilities for management, protection, storage, and deletion of applicant data.

The Subcommittee further recommends guardrails on the collection, use, and maintenance of data used by AI-enabled selection procedures. Users and vendors should implement reasonable security measures to protect any collected information. Such measures should include data

---

[60] *See* Data Collection Section, below.
[61] 15 U.S.C. § 1681.

retention and protection guidelines that limit what has been called "data persistence," *i.e.*, the length of time the data may potentially be used and retained.[62] These measures may be of significant importance given the high level of speed and volume associated with AI data processing. Audits of AI-enabled selection procedures should regularly examine the systems for data spillover (data collected on individuals who are not the targets of data collection) and data leakage (release of personal data or reidentification of anonymized data if security measures are breached). Developers of AI-enabled tools also should ensure that the data collected are limited to personal information that is relevant and necessary to the stated purpose of the selection procedure and that such data are retained for only as long as necessary to achieve the stated purpose. Such guardrails should, at a minimum, include maintaining data and information for periods consistent with existing legal requirements and consideration of whether data relevant to refining or evaluating the AI-enabled selection procedure should be maintained for longer periods.

Finally, users and vendors of AI-enabled selection procedures should consider limits on future uses of collected data when such would be beyond the original purpose, a practice often referred to as "data repurposing." With respect to data repurposing, the Subcommittee recommends that re-sale of data collected through an AI-enabled selection procedure either be forbidden or severely restricted unless explicit consent for re-sale is obtained from applicants.

VII. Fairness Considerations and Recommendations

The Society for Industrial and Organizational Psychologists ("SIOP") and the American Psychological Association ("APA") have adopted a set of principles regarding the use of selection procedures in employment testing.[63] Consistent with the *Standards for Educational and Psychological Testing,*[64] the *Principles* argue that fairness is a social construct with multiple

[62] Employers should be aware of new and conflicting record retention requirements. For example, the Illinois Artificial Intelligence Video Interview Act requires employers to destroy video recordings within 30 days of request from an applicant/employee. This conflicts with, for example, OFCCP's two-year record retention requirement.
[63] *Principles for the Validation and Use of Personnel Selection Procedures* (Soc'y for Indus. Organizational Psych., 5th ed. August 2018) ("the *Principles*"), https://www.apa.org/ed/accreditation/about/policies/personnel-selection-procedures.pdf.
[64] *The Standards for Educational and Psychological Testing* (Am. Psych. Ass'n 2014), https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf.

meanings and offer four possible concepts of "fairness."[65] The first concept may be the most familiar: it turns on whether there are sub-group differences, or alternatively, equal outcomes in, for example, a selection procedure.[66] The second concept is anchored in the equitable treatment of applicants. In other words, to achieve fairness, all participants in a selection process should be subjected to equivalent treatment in terms of all aspects of the process, *e.g.*, assessment conditions, feedback, reconsideration opportunities, and other features of assessment process administration.[67] The third concept is based on comparable access to underlying constructs. Here, fairness requires that applicants be afforded equal opportunities to demonstrate their standing on a construct "without being unduly advantaged or disadvantaged by other individual characteristics."[68] The final concept is statistical in nature and relates to the measurement or analysis of bias.[69] Below we make recommendations on how to incorporate the key concepts of equitable treatment and comparable access into the use of AI in employment selection processes.

---

[65] The *Principles* do not constitute legal requirements or legal guidance. *See* Kelly Trindel et al., *Fairness in Algorithmic Employment Selection: How to Comply with Title VII*, 35 A.B.A. J. LAB. & EMP. L. 241, 241 (2021) for a discussion.

[66] Although this concept subsumes the legal definition of adverse or disparate impact, the *Principles* note that the APA *Standards* (American Educational Research Association ("AERA"), American Psychological Association ("APA"), and National Council on Measurement in Education ("NCME")) reject this definition on the basis that differences in outcomes should result in increased review, but such differences do not necessarily result from bias or unfair treatment. It is important to recognize that some AI-enabled selection procedures conflate adverse impact and fairness, inappropriately concluding that impact ratios greater than .80 imply an assessment is fair. These concepts are not interchangeable, and fairness is a complex concept that requires examining an assessment through multiple lenses.

[67] Particularly relevant to AI-enabled selection systems, the *Principles* note that "[c]onditions related to mode of administration may be particularly important to consider given recent technological advances (*e.g.*, testing via computers, laptops, tablets, and other mobile devices such as smartphones)." *Principles*, at 22.

[68] "Under this view, it may be particularly important to consider whether factors such as age, race, ethnicity, gender, socioeconomic status, cultural background, disability, and language proficiency restrict accessibility and affect measurement of the construct of interest." *Principles*, at 22.

[69] *Id.* at 23. The concept of "algorithmic bias" is especially relevant to the use of AI-enabled employment selection procedures. While that term can have a variety of meanings, in the context of testing bias refers to systematic error in test scores that affect the average performance of different groups of test takers in different ways. When the systematic error (or irrelevant source of variance) causes subgroup mean differences in test scores or criterion measures, this bias is denoted as *measurement bias*. When the systematic error creates subgroup differences in the overall predictor-criterion relationship, it is labeled *predictive bias*. Both types of bias might result from a machine learning algorithm as forms of algorithmic bias – even though it might be difficult to identify them within machine learning algorithms. As general rule, algorithms that are less transparent are less revealing of forms of algorithmic bias, whether it is measurement bias, predictive bias, or some other form of bias.

Measurement bias in employment testing is often difficult to detect. One potential solution is an item sensitivity review. This consists of a careful review of instructions and test items to detect wording that may be differentially understood across cultures or languages or that contain references that are offensive to one group or another. These reviews are conducted by those who have specific expertise and experience in the cultural evaluation of content and may lead to revising or deleting instructions or items on the test. Evaluation of measurement invariance using techniques such as confirmatory factor analysis can also determine if the constructs being assessed by the test are the same for different groups. Regardless of the approaches taken to identify and eliminate

A. Equitable Treatment

Equitable treatment requires all applicants to be treated similarly in the selection process, including things like access to practice materials, performance feedback, and retest opportunities. In the context of AI-enabled selection procedures, this may take the form of using the same type of data, variables/features, procedures, and scoring for all individuals who are being compared to a standard or to each other. Evaluating such concerns goes hand in hand with transparency and explainability, in that to evaluate equitable treatment, one must understand how an algorithm works, what data are used, and how the data are used. To help ensure equitable treatment, employers should establish, and maintain with regular auditing, explicit procedural safeguards to ensure that all applicants are treated in the same manner. Every step in the application process (including AI-enabled selection procedures) should be scrutinized for systemic issues.

The technology landscape has evolved such that many applicants take assessments on cell phones, laptops, and tablets, which makes it difficult to ensure that all applicants are treated similarly. For instance, there may be differences in treatment because of varying access to required equipment, internet availability and speed, quality and speed of computers or other devices, quality of cameras, microphones, and/or speakers, knowledge of process steps and expected participation, availability of distraction-free private environments. Employers should

---

measurement bias, the user should also take steps to ensure that features or variables included in algorithms are not proxies for variables that define a protected class. For example, zip code may indicate race, or participation in some sports may suggest gender. Although many approaches to model development include routines to identify features on which there are large group mean differences – which may indicate that the feature is such a proxy – small sample sizes for the training data may conceal such differences.

Predictive bias in employment testing, including AI-enabled assessments, is of concern because when found, it indicates that a test predicts the criterion or outcome measure in different ways depending on the group to which an individual belongs. Thus, if there is evidence of predictive bias, that indicates that the same test score may predict a significantly higher (or lower) criterion score for the majority group compared to the minority group. Technically, predictive bias occurs when consistent non-zero errors of prediction are made for members of a subgroup. In traditional statistical analyses, predictive bias is often evaluated using moderated multiple regression, in which the regression lines of the majority and minority groups are computed, and the slopes and intercepts are compared. When data across subgroups can be appropriately summarized by the same regression line (*i.e.*, the Cleary Model), then the measure is said to lack predictive bias. Alternative procedures for evaluating predictive bias as a form of algorithmic bias are much less clear and have not been extensively assessed. One reasonable alternative, however, would be to determine whether relationships between predicted and actual scores generated by the algorithm vary by subgroup. It is important not to conflate predictive bias with group mean differences (and adverse impact from such differences). These are two different concepts and one can occur without the other. For example, test scores for different groups may show observed mean differences on employment tests yet show similar prediction of organizational outcomes. Or, the prediction of organizational outcomes may be different, even if mean differences on employment tests are the same. In short, predictive bias should be evaluated even if there is no evidence of group mean differences.

make test takers aware of technological requirements and consider solutions to ensure equitable treatment such as providing applicants the option of taking an assessment using employer-provided technology.

Hypothetical:

> An employer uses a video-based interview that evaluates the content of the applicant's oral response that has been transcribed into text, as well as facial features and voice characteristics. An applicant applies for a position and is screened out based on all three components of the interview. The transcription of the oral text contained numerous errors and uninterpretable passages, and the capture of face and voice material was intermittent because the applicant used an older computer connected to the internet at a slow speed.

The applicant may have met the requirements of the job, but could not be accurately assessed because of the computer issues. This scenario raises concerns of procedural fairness if the employer was not clear about the minimum computer requirements and internet speed or did not offer testing locations where such equipment was available.

It is important for employers to ensure that vendors produce technical reports for selection procedures that should, to the extent possible, document how procedural fairness was considered. Further, evaluations of procedural fairness should reflect the assessment as it is operationally implemented. The goal should be to ensure, to the extent possible, that deployment of AI-enabled systems allows for all applicants to be treated as equitably as possible.

B. Comparable Access

Whether the assessment process consists of automated or semi-automated interviews, online testing, assessment of applications or resumes, games or simulations, or other technologically enhanced and AI-enabled selection procedures, the data generated by such procedures will be evaluated through an algorithmic analytical process to allow some decision about the applicant to be made, either by a human informed by the data analyses or by an automated process. Applicants must be afforded the opportunity to have the data attributed to them be an accurate portrayal of their actual knowledge, skills, abilities, or other characteristics. That is, the collected data about an individual should accurately reflect the person's standing on the attribute being measured, without regard for other, irrelevant, factors. Thus, for example, a measure of a person's level of conscientiousness should not be affected by the level of manual dexterity imposed by a disability. As noted in the *SIOP Principles* quoted above, a range of factors may affect the comparability of measurement, and some may overlap with concerns for equitable treatment.

Comparable access is particularly relevant for individuals with disabilities. The EEOC's TAD states that "the steps taken to avoid . . . Title VII discrimination are typically distinct from the steps needed to address the problem of disability bias."[70] The EEOC encourages users to inquire as to whether the vendor considered issues related to disability in the assessment development process. Specific issues include whether the user interface is accessible, whether materials are offered in accessible alternative formats, whether the vendor determined if the selection procedure may disadvantage individuals with disabilities, and whether any of the traits or variables in the algorithm are proxies for certain disabilities.

These fundamental principles apply to the development and implementation of all selection procedures; however, the impact may be amplified in the case of AI-enabled selection procedures because of the sheer volume of available data and associated features. Further, these principles are not isolated to consideration for individuals with disabilities, but rather can be applied to thinking about any demographic group and ensuring that there are no barriers to comparable access not related to the particular job at issue.

> *Hypothetical*:
>
> For years, an employer used a memory and recall test as part of its hiring process. The test was validated in accordance with professional and legal guidelines, and there is evidence that it is predictive of successful job performance. The vendor, in an effort to modernize the test, now offers a gamified version that purports to measure the same underlying skills and abilities as the original version. An applicant with a vision impairment applies for the job and does poorly on the gamified test even though the applicant has outstanding memory and recall skills. The applicant was not provided notice regarding the use of the gamified assessment or how it worked in advance.

Here, the gamified test did not provide this applicant comparable access and did not fairly measure the targeted skills.

---

[70] EEOC TAD ("ADA Basics" section, Question 10).

Fairness is not a simple concept, nor is there a single approach or metric that allows one to establish that an assessment has been developed and implemented fairly.[71] Beyond the statistical framework for evaluating fairness with respect to the predictor-criterion relationship, fairness requires considering the assessment itself, what data are being used, and how it was designed, as well as how the assessment has been implemented and the processes surrounding the implementation.

Given the complex nature of fairness, and the heightened societal awareness of its importance, employers are strongly encouraged to think carefully about issues of fairness when employing AI-enabled selection procedures. Further, the Subcommittee recommends an open dialogue between employers and vendors or developers of AI-enabled selection procedures on the different aspects of fairness.

[71] While fairness is a complex concept embracing multiple facets, the determination of whether an employment screen has caused discrimination under existing federal law is more established. *See* discussion of UGESP and court cases throughout this Report.

DATA COLLECTION

I. Introduction

This Section identifies key issues and recommended practices in the collection and use of data upon which selection algorithms train.

Most AI-enabled selection algorithms train on predictor data and criterion data to detect patterns present in a data set. Examples of predictor data include item-level responses to a knowledge test or personality test, information from resumes, or interview responses. Examples of criterion data include job outcomes (*e.g.,* overall performance, turnover) or perhaps more indirect indicators of how likely a job applicant is to succeed on the job (*e.g.,* recruiter move-forward decisions, competency ratings, or training performance).

Once trained, these selection algorithms use predictor data from job applicants or employees to make a prediction about a criterion of interest. Because selection procedures may be used for internal employment decisions (*e.g.,* promotions) and external employment decisions (*e.g.,* hiring), different types of data may be available and useful accordingly. For example, algorithms that make predictions regarding promotions may rely on past performance ratings, data that are generally not available for job applicants.

Selection algorithms can use a variety of inputs. For example, asynchronous video interviews ("AVIs") allow candidates to engage in an interview process that does not require a human interviewer: interviewees respond, in a place and time of their choosing, to pre-selected interview questions. Their responses include the content of what they said and the raw audio and video data. Oral responses can be transcribed into text, which itself can be analyzed at a variety of units of language, and the raw audio and video data may also be transformed into other measures such as voice inflection or facial or body geometry action units.[72] Similarly, interactive games can produce information on not only right and wrong answers but also on variables related to the test taker's approach to testing such as cursor location, response time, indecision in response choices, or partially correct choices. Another source of algorithm input is information publicly available on the internet, such as information about what knowledge, skills, or abilities

[72] These are highly contentious data because they are not necessarily job-related and can be viewed as biased (*e.g*., voice inflection as influenced by accent, gender, or disability). Data going into AI-enabled tools should show signs of job relatedness and lack of bias.

job applicants have posted on social media as well as information in social media posts that is not directed at potential or actual employers.

II. Specific Data Collection Issues of Concern with AI Employment Tools

A. <u>Criterion Data</u>

Some algorithms are intended to predict an outcome such as job performance while others indicate the extent to which an applicant is like a group of employees who have been defined as "good" employees. A selection algorithm's predictions about future job performance require a job performance criterion. Other criteria such as turnover, absenteeism, safety behavior, and accidents may also be important to an organization.

Regardless of the nature of the criterion, criterion data raise four main concerns:

- First, criterion data accuracy is often unknown and should be verified. If applicants are evaluated based on their similarity to a core group of "good" employees, then the standards for "good" and the method for determining who was "good" should be apparent; otherwise, evaluating the accuracy of the "good" designation is difficult if not impossible.
- Second, criterion data may be missing in whole or part. When such missing data are not likely to be missing completely at random, the handling of missing data is a critical issue.
- Third, the currency of the criterion data may matter. Job performance information from ten years ago may not reflect the current capabilities of an individual. Excessive absenteeism in the past may be a result of a medical or family issue that has long been resolved. In contrast, educational degrees earned a number of years ago may still be relevant. Developers must determine appropriate time frames for criterion data and document them.
- Fourth, criterion data may not represent the entire domain of job performance. Key performance indicators ("KPIs") often represent important aspects of job performance, are measured regularly, and are stored in accessible databases. However, KPIs do not always represent the entire performance domain. For example, KPIs may focus on aspects of performance that are easily captured electronically (*e.g.*, sales, average call time, production) and neglect other components of the job that are more difficult to measure (*e.g.,* customer service, quality). Formal performance appraisal tools may also not fully cover the performance domain. For example, some tools focus on goal attainment and neglect aspects of quality, timeliness, resource utilization, contributions to team efforts, and the like.

B. Input Data Quality

Selection algorithms are only as good as the data used to develop and train them. Recognition of this provides the foundation for fairness in AI-enabled systems. Decision makers who rely on AI technology in recruiting and hiring need to know where the data come from and what the quality of the data is. Algorithms trained on incomplete data or systematically biased data can lead to biased and inaccurate outcomes that generate discrimination in recruiting and hiring. Using biased data can produce discriminatory outcomes disadvantaging candidates with respect to their sex, race, disability status, or other protected characteristic. Some of the quality issues that may arise with data used in AI systems include:

- Reliance on unrepresentative or incomplete samples to train AI systems
- Data used to train algorithms that reflect previous biased judgments
- Detailed personal data revealing protected characteristics from seemingly neutral information like zip code of residence, social networks, behavioral changes, or other demographic proxies
- Errors contained in datasets disproportionately affecting the records of protected groups.[73]

Personnel selection involves a range of disciplines: industrial and organizational ("I/O") psychologists; other social scientists; data scientists; software engineers; diversity, equity, and inclusion experts; and other stakeholders. Each discipline has its own methods and practices for assessing model performance and accumulating evidence of the validity of the inference made from the test score.[74] Regardless of how the selection procedure is evaluated, however, all

---

[73] *See* Office of Science and Technology Policy, Exec. Office of the President, *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People* (2022) ("AI Blueprint"), at 26, https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf ("Directly using demographic information in the design, development, or deployment of an automated system (for purposes other than evaluating a system for discrimination or using a system to counter discrimination) runs a high risk of leading to algorithmic discrimination and should be avoided. In many cases, attributes that are highly correlated with demographic features, known as proxies, can contribute to algorithmic discrimination.").

[74] The *SIOP Principles for the Validation and Use of Personnel Selection Procedures* ("the *Principles*"), which are intended to "specify established scientific findings and generally accepted professional practice in the field of personnel selection psychology," defines validity as "the degree to which accumulated evidence and theory support specific interpretations of scores from a selection procedure entailed by the proposed uses of that selection procedure." *Principles,* at 1 and 4. In the case of scores on employment tests, this means that assessment scores are job relevant.

Validity is a property of how the assessment is used, not of the assessment itself. Moreover, while different disciplines may use different methods to assess validity, validity itself is a unitary concept. It is therefore important to remember that each method for validation represents a source of evidence and not an alternative approach to validation. The evidence supporting the use of an employment test must be aligned with the interpretation to be methods require accurate data. Consequently, the quality of both the predictor and criterion data must be established.

Importantly, some techniques for establishing reliability of data may or may not apply to newer sources of data. For example, measures of test-retest reliability lose their usefulness when applied to biometric data captured frequently over a long span of time: random error in measurement approaches zero when the number of measurements increases. On the other hand, inter-rater reliability measures meant for ill-structured measurement designs may become more useful to assess rating algorithm reliability. Measures of internal consistency may not make sense when thousands of features are included and the constructs measured are not specified.

The nature of some data sources may suggest unique ways to assess quality. For example, word error rate ("WER") is an important metric in transcription – the underlying technology that allows one to apply natural language processing ("NLP") to raw audio data. The lower the WER, the fewer mis-transcribed words exist in a body of text, which improves data quality and the accuracy of the NLP. Regardless of the nature of the input data, the developer should take steps to assess the quality and remedy any problems found and document both the process and the results.

Another component of data quality is the accuracy of the data input. There is usually little doubt about the accuracy of responses that are entered electronically by the applicant during the selection process. If the test taker responds "B" to item 22, the intended response of the test taker is usually captured accurately by the selection tool, and test takers are fully aware their responses are being captured. In contrast, data collected through secondary sources may present concerns about the extent to which the response accurately reflects the construct being measured. For example, although an applicant may have intentionally entered data into a social media platform, the information may not be true. In addition, some forms of data are assumed, without supporting evidence, to represent a relevant construct. For example, the number of times a response is changed might be assumed to represent indecisiveness, but no evidence of construct validity for the change of response measure is provided.

---

made. For example, a prediction regarding job performance needs to be supported by criterion-related evidence that statistically links test scores and criteria or by content-oriented evidence that relates the test to job requirements through the judgment of subject matter experts.

C. Handling Missing Data

Another data quality concern is missing data. For example, applicants sometimes skip over questions they cannot or do not want to answer.

Missing data matter because they can bias inferences or predictions. To handle missingness, it is important to know or clearly justify assumptions about why those data are missing.[75] The long-standing approach is to classify missing data into three types.[76]

- Missing completely at random ("MCAR"). This means that the probability of missing data is unrelated to (independent of) the distributions of both observed and unobserved variables. For example, if item nonresponse on a survey occurs solely because each survey respondent is by sheer accident equally likely to press the "forward" button twice instead of once, then the resulting missing data are missing completely at random.
- Missing at random ("MAR"). This means that the probability of missing data depends on the observed variables but is independent of the unobserved variables. For example, if sex, race, education, and age are recorded for all the people in the survey, then "earnings" is missing at random if the probability of nonresponse to this question depends *only* on these other observed variables. When a variable is missing at random, it is acceptable to exclude the missing cases as long as the algorithm controls for all the observed variables that affect the probability of missing data.
- Missing not at random ("MNAR"). This means that the probability of missingness depends on the unobserved variables, *i.e.*, missingness occurs because of a process about which the data contain no information. For example, suppose that "surly" people are less likely to answer the earnings question, surliness is associated with earnings, and "surliness" is unobserved. If data are MNAR, the missingness must be explicitly modeled, and the missing values can be imputed, or else some bias in the inferences must be accepted.

Depending upon whether the missingness is MCAR, MAR, or MNAR, proper methods for handling missing data include deleting observations with missing data (case control analysis) as well as techniques for imputing missing data that apply to some machine-learning approaches.[77]

---

[75] Karthika Mohan and Judea Pearl, *Graphical Models for Processing Missing Data*., 116 J. of the Am. S tat. Ass'n 1023–37 (2021), https://doi.org/10.1080/01621459.2021.1874961.
[76] Donald B. Rubin, *Inference and Missing Data,* 63 Biometrika 581–92 (1976), https://doi.org/10.1093/biomet/63.3.581.
[77] Tlamelo Emmanuel, *et al*, *A Survey on Missing Data in Machine Learning*, 8 J. of Big Data 1-7 (2021), https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00516-9

D. Data Sampling

1. Training and Evaluation Samples

AI-enabled selection requires data samples to develop and evaluate an AI-enabled tool. Selection algorithms use two broad samples: a Training sample to develop the algorithm(s) and an Evaluation sample to test generalizability, validity, and fairness of the score produced by the algorithm. Both samples require representative predictor data (*e.g.*, applicant data) and matched-cases on criteria of interest (*e.g.*, incumbent data with job performance metrics).

The Training data set can be further subset (using appropriate stratified random sampling procedures) into Train and Development sets. The Train set is used to develop the model using AI-based techniques (*e.g.*, Machine Learning). Model exploration methods are used on the Train set to generate multiple model iterations that predict the criteria of interest. The Development set is then used to identify the best performing model that will be the basis of an AI-enabled tool. Development data are used to test for overfitting of the various trained models, providing the first indication that the final model selected from the multiple iterations does not capitalize on chance results (and exhibits predictor-criterion relationships). We recommend that procedures be applied to ensure both sample sets have representation on key variables while providing the sample sizes needed (*e.g.*, 80%/20% split; k-fold cross validation techniques).

The Evaluation sample is considered pure hold-out or validation data. Developers use this sample only after selecting a final AI model from the Training data. This sample allows researchers to test the generalizability of the model to an independent data set and evaluate validity relationships among predictor, criterion, and demographic variables. More specifically, developers use applicant data in the Evaluation sample to independently examine the psychometric properties of predictor scores from the AI-based model and to test for generalizability in how different subgroups (*e.g.*, race, gender) perform. Criterion data (*e.g.*, performance data collected after the applicant has been hired and performed in the job for some period of time) in the Evaluation sample allow for data not yet seen by the modeling process to test and verify that the expected predictor-criterion relationships persist. Whereas Training data can be used to explore performance of multiple competing models, Evaluation data are withheld from the training process and used just once to explore the validity of the final selected AI-enabled tool.

2. *Recommended Sampling Considerations*

To avoid common pitfalls, we recommend considering the following when identifying Training and Evaluation samples:

- Representation across the target population (*e.g.*, applicants). Developers should identify the critical characteristics of the applicant population prior to establishing samples, and when feasible (*i.e.*, where race, ethnicity, gender, age, and other relevant characteristics are collected from applicants), use appropriate sampling techniques to ensure all groups are represented in appropriate proportions in both the Training and Evaluation samples. These characteristics of the population can influence how well scores on AI-enabled assessments generalize from the sample to the broader applicant population, and likely impact the tool's validity and fairness.
- Representation across the incumbent population. Developers should ensure both the Training and Evaluation samples contain proportional representation. Considerations include demographic representation as well as business areas in target organizations as well as other relevant variations such as different equipment or work processes.
- Alignment between predictor and criterion populations. Developers should identify in advance systematic differences between the predictor (*e.g.*, applicant) and criteria (*e.g.*, incumbent) populations. For example, applicants may have stronger or weaker skill sets than incumbents. These differences may limit the generalizability of a tool's results. We recommend using appropriate methods to account for such differences in model training procedures.
- Sample Size. Developers should identify the target number of matched predictor-criterion cases required for analyses before generating samples. Samples that do not meet the target sample sizes should motivate developers to examine the reasons for and impacts of any missing data. We also recommend that developers prioritize retaining cases to meet the Evaluation sample needs. Once those needs are met, developers can employ appropriate methods to fill out the Training sets. These may include cross-validation sampling techniques (*e.g.*, k-fold) between the Training and Development sets or data imputation techniques for missing data. Data imputation is not recommended for the Evaluation sample, only for the Training sets. All sampling and imputation methods should be properly documented.
- Point-in-Time. It may be appropriate to split Training and Evaluation samples by a point-in-time to further ensure generalizability of results. For example, all predictor and criterion cases in the Training sets may occur in a date range prior to cases in the Evaluation set.

E. Notice

As noted in the Transparency and Fairness Section of this Report, job applicants and employees should receive notice about the collection and use of data in a selection algorithm. The level of information in the notice should be detailed enough to:

- obtain the applicant's or employee's consent to collect and use the data;
- provide enough information about the selection process for the typical applicant or employee to ask for disability or other accommodations; and
- identify the data sources used, especially secondary sources such as social media that may include information not intended to be shared for employment selection decisions, so that the applicant or employee has the opportunity to evaluate (and perhaps correct or update) such information.

Some AI-enabled selection tools utilize only data provided intentionally by applicants for the purpose of employee selection; indeed, AI methods can be used to build prediction algorithms from an array of traditional item types (*e.g.*, Likert-type self report measures, multiple choice questions with right/wrong answers). Other AI-enabled selection tools utilize data shared by the test taker for purposes other than employment selection (*e.g.*, social media) or in a manner unbeknownst to the applicant (*e.g.,* response time or cursor location while responding to the item or playing a game). When applicants respond to test items, interview questions, and game challenges during a selection process, they usually know their responses will affect their odds of selection. Applicants are less likely to know how less obvious or secondary data affect their test scores. For example, applicants may know that, in an AVI, the content of what they say matters for selection. But they may not know specifically about other data that an AVI produces about them that are used by the algorithm for selection (*e.g.*, facial expressions, voice characteristics, and word frequency). Similarly, applicants may not understand that social media posts or data from other secondary sources are being utilized in AI-enabled tools without notice. Applicants also likely do not know that data about them (*e.g.*, financial information and shopping habits) that can be purchased from third parties (*e.g.*, data brokers) may be used as inputs for algorithms.

To determine whether notice to applicants is sufficient, one should consider (a) what the applicable law requires and (b) whether such notice comports with an organization's business practices and its extra-legal (ethical or supererogatory) commitments. In turn, a notice's purpose(s) will affect the type and degree of information to be communicated, including:

- the fact of use or intended use of applicant data as training data for processing
- categories of applicant data collected (*e.g.*, job knowledge, skills, abilities, other personal characteristics, demographic information)
- sources of applicant data (*e.g.*, the job-applicant, a third-party source, the selection process itself)
- the purpose(s) of the processing

- the legal basis for the authority to process such data
- the manner in which the data will be used and/or disclosed before and after processing
- procedures for correcting errors/omissions when data from secondary sources such as social media are used.

While organizations must provide notice to applicants, they must also be cautious about what they disclose in order to protect their trade secrets or other proprietary information.[78]

1. Notice for Consent

Legal definitions of "consent" sometimes expressly require that individuals' indication of agreement to the processing of their personal data must be "informed."[79] Even when consent requirements leave "consent" undefined,[80] courts and agencies are likely to treat the individual's knowledge as relevant to whether the individual has consented as required by law.

For notice to suffice for "informed" consent, one should ask (1) whether the information in the notice accurately describes the nature, scope, and specific purposes of the intended and expected data collection and use; and (2) how likely it is that the notice will cause the typical applicant to understand what they are being asked to accept. These key questions, in turn, affect how likely it is that a court or agency will find any particular notice to be "clear" and "concise" enough to qualify as "informed" consent. These same questions also affect whether what a firm later does with the data falls within the scope of the consent.

In practice, many employers assume if a job applicant provides them with data or posts that data on a publicly accessible platform (*e.g.*, LinkedIn, Facebook, TikTok), the applicant has consented to let them use such data for any selection process and in any other way they wish. However, that is not necessarily the case.[81]

---

[78] See, *e.g.*, 18 U.S.C. § 1836.

[79] *E.g.*, GDPR, 2016/679, art. 4(11), 2016 O.J. (L 119) (EU), https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679; California Consumer Privacy Act, Cal. Civ. Code § 1798.140(h). The current exemption applicable to data collection for AI hiring activities in the California Consumer Privacy Act, Cal. Civ. Code § 1798.145(m)(1), will "become inoperative on January 1, 2023," *id*. § 1798.145(m)(4).

[80] See, *e.g.*, 820 Ill. Comp. Stat. Ann. 42/5; Md. Code Ann., Lab. & Empl. § 3-717(b).

[81] Compare Cal. Civ. Code § 1798.140(o)(1)(I) (defining "personal information" to include "[p]rofessional or employment-related information") with *id*. § 1798.140(o)(2) (only excluding from definition of "personal information" any information "that is lawfully made available from federal, state, or local government records").

A few jurisdictions afford individuals the right to request businesses to delete some kinds of data obtained about them from a job-applicant selection process.[82] When notice of such right is required as well,[83] a court or agency may take as relevant to "informed" consent whether, how, and when job-applicants were informed of their right to ask for deletion of any personal information that their consent was otherwise supposed to have covered.

Finally, when it comes to consent, the ethical guidelines of a profession or institution may recommend doing more than what the law requires. Thus, for example, the Code of Ethics of the Association for Computing Machinery provides that "[c]omputing professionals should establish transparent policies and procedures that allow individuals to understand what data is being collected and how it is being used, to give informed consent for automatic data collection, and to review, obtain, correct inaccuracies in, and delete their personal data,"[84] while the guidelines issued by the American Psychological Association require that "[p]sychologists obtain informed consent for assessments, evaluations, or diagnostic services . . ., except when . . . informed consent is implied because testing is conducted as a routine educational, institutional, or organizational activity (*e.g.*, when participants voluntarily agree to assessment when applying for a job. . . ."[85]

2. Notice for Accommodation

Although employers generally are required to make reasonable accommodations to a job-selection process for an applicant's disability, religion, or other legally-mandated basis, the applicant typically must ask for the accommodation before any legal obligation triggers. For example, the Americans with Disabilities Act ("ADA") bars discrimination against employees and job applicants based on "disability," while also barring "inquiries of a job applicant as to whether such applicant is an individual with a disability or as to the nature or severity of such

---

[82] *See, e.g.*, 820 Ill. Comp. Stat. Ann. 42/15 (applicant video interviews); Cal. Civ. Code § 1798.105(c) ("any personal information about [them] that the business has collected"); GDPR, art. 17.
[83] *See, e.g.*, Cal. Civ. Code § 1798.105(b).
[84] *ACM Code of Ethics and Professional Conduct* (Ass'n for Computing Mach. 2018), at § 1.6, https://ethics.acm.org/. *See also*, *Ethical Guidelines for Statistical Practice* (Am. Stat. Ass'n 2022), at 5 ("The ethical statistical practitioner . . . [u]ses data only as permitted by data subjects' consent, when applicable, or considers their interests and welfare when consent is not required. This includes primary and secondary uses, use of repurposed data, sharing data, and linking data with additional data sets.").
[85] *Ethical principles of psychologists and code of conduct* § 9.03(a) (Am. Psych. Ass'n 2002, amended effective June 1, 2010, and January 1, 2017), http://www.apa.org/ethics/code/index.html.

disability."[86] Thus, a job applicant with impairments to, for example, visual acuity or manual dexterity may receive an accommodation to an employment test affected by the impairments (if the test does not aim to measure such skills). To obtain that accommodation, however, the applicant has to be informed about that test in advance to know that, given their impairments, an accommodation may be worth requesting.

To provide enough notice for this purpose, an employer should ask whether:

- the information therein accurately describes the nature of the selection process so that typical job applicants understand how much their impairments might affect their performance in the selection process;
- the notice reminds applicants how they may, if they wish, request reasonable accommodations;
- the notice does not ask applicants to disclose the nature or severity of any disability they may have; and
- the notice itself does not reduce the accuracy of a test or another selection-process measure of a job-related skill of the applicant.

<u>*Hypothetical*</u>

Suppose a selection process includes an AVI. The employer or software vendor may provide notice to applicants that the selection process involves an AVI and inform them of a way to ask for reasonable accommodation. Applicants, however, may not know that selection-relevant data from an AVI includes not just the content of what they say, but also other data, such as word counts, facial expressions, and voice characteristics.[87] As a result, applicants with speech impairments or uncontrollable facial spasms may not ask for an alternative to AVI, unless the notice informs them otherwise.[88] If, however, a notice about AVI data would systematically change job-applicants' words or nonverbal behavior during an AVI, the notice may thereby increase measurement error.

If a business does or must provide such notice, that notice's timing matters also. While some jurisdictions may set fixed times for notice by law,[89] a notice's proper timing depends largely on (1) how quickly an employer or vendor can satisfy a reasonable accommodation request made in response to a notice, and (2) the expected duration between when a selection process begins and

[86] 42 U.S.C § § 12112(a) and 12112(d)(2)(A).
[87] *See* FN 72, *supra*.
[88] *See* U.S. Equal Emp. Opportunity Comm'n ("EEOC"), *The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees* (May 12, 2022) (the "EEOC TAD"), ("Promising Practices for Employers" section, Question 14) (identifying as a "promising practice" for ADA compliance "[d]escribing, in plain language and in accessible formats, the traits that the algorithm is designed to assess, the method by which those traits are assessed, and the variables or factors that may affect the rating.").
[89] *E.g.*, N.Y.C. Admin. Code § 20-871(b)(1) ("no less than ten business days before such use").

when one or more applicants are actually selected. Because these conditions vary by industry and workplace setting (*e.g.*, immediate placement of a substitute teacher), we caution against setting any fixed time periods for notice that ignore such variation.

3. Notice for Data Accuracy

To date, while some jurisdictions require notice of an individual's right to request corrections of inaccurate personal information, organizations appear not to be legally obligated to provide notice to past or current job applicants that they might in fact have incorrect or incomplete data about them.[90]

Nonetheless, to improve data quality, some employers or vendors may wish to provide enough notice to job applicants so that they can correct or update inaccurate or incomplete personal data about themselves. For example, where firms rely on sources of data over which the applicant has control (*e.g.*, LinkedIn, Facebook), they may include, as part of their notice to applicants, a recommendation that the applicant correct or update personal information that is found on those sites before applying. Any such notice should be provided in such a way that the notice itself does not reduce the accuracy of a test or other measure of a job-related skill of the applicant.

F. Retraining Algorithms

Sometimes there is a change in the job requirements, the applicant population, or the process that generated the data upon which a selection algorithm was trained. This may call for retraining an algorithm or optimizing how the algorithm functions after it has been put into use. Two potential reasons for retraining an algorithm are model drift and technological advancements.

In either situation, the following practices should be adopted in preparation for retraining:

- First, if the data are available, test how susceptible an algorithm is to performance degradation. This involves training an algorithm on older data and testing it on newer data with set time intervals (*e.g.*, six-month old data, one-year old data). This process will allow one to estimate how much the algorithm's performance changes over time and plan when to check the live algorithm for degradation.
- Second, once an algorithm is finalized but before it is implemented, store the model's predictions, the outcome data used to train the model, any necessary linking variables, as well as the original predictor and criterion distributions for future analyses that will let one detect underlying changes to the distribution of

[90] *See* Cal. Civ. Code § 1798.106(b); GDPR, art. 16.

the predictor and criterion data or changes to how the variables or features are related to one another.

- Third, anonymize training data before storing them. Data storage can be impacted by the data retention policies of vendors and organizations, and regulatory requirements. Anonymizing training data before storage helps prevent data loss that might otherwise occur as a result of such policies.
- Additional pre-retraining practices can include detecting model degradation (*i.e.*, if a certain threshold is crossed in model performance) which can trigger either a flag for manual review or, when appropriate, automatic retraining of the algorithm on new data.

1. Model Drift

Model drift occurs when the underlying data processed by an algorithm over time are different from the data the algorithm was trained upon. Model drift often happens when a selection practitioner conducts a concurrent validation study to validate a new selection procedure. In a concurrent design, the practitioner administers a new selection procedure to a group of job incumbents and simultaneously collects criterion data. An algorithm is derived from the incumbents' responses to the procedure and their existing criterion data. This algorithm is then applied to the responses collected from applicants to the job to identify those who will most likely do well on the job (or meet another criterion). The applicant population often differs from the original incumbent population due to selection effects, and the algorithm should be adjusted when enough applicant data are collected and the applicants are hired so that the criterion data are available.

The fundamental issue of model drift is that some underlying assumption about the data used to train an algorithm has changed. Applicants differ from incumbents; applicant characteristics shift over time; or the job requirements themselves change, leading to different response patterns, demographic compositions, or performance standards. Professional standards such as the *Principles for the Validation and Use of Personnel Selection Procedures* advise the test developer to be aware of organizational changes that affect the job requirements, conduct additional research as needed, and update the associated documentation.[91]

Given the risk of model drift, when to retrain the algorithm depends in part on the realities of collecting new data in an operational setting. Organizational hiring may ebb and flow seasonally;

---

[91] *See Principles*, at 38-39.

performance ratings may be collected only once a year; and even customer-vendor or labor contract terms can affect when an algorithm can and should be retrained. Accordingly, we recommend that developers work closely with organizational partners to create a plan for regular inspection and potential updating of any algorithm that takes into account the particular organization's needs.

2. Technology Updates

Technology updates are also a reason to retrain an algorithm, absent any change to predictors, the criterion, or the circumstances surrounding the algorithm's use. An underlying data collection mechanism may now produce more accurate raw data. For example, transcription services have decreasing error rates in identifying words uttered, especially among speakers with accents. In the case of open-ended text data, an interview-based algorithm may have been trained with word or context embeddings produced by a natural language processing ("NLP") model to predict job performance. Although neither the predictors (the embeddings), the criterion, nor the circumstances surrounding the use of the algorithm have changed, this algorithm may still benefit from retraining because the transcription technology now generates more accurate data.

Whether to retrain because of technology updates depends on the organization's data and use of the algorithm. We again recommend working closely with the organizational partner to determine whether, when, and how to incorporate new technological updates to their algorithms.

G. Documentation

We recommend sufficient documentation to assess data reliability and validity, including techniques used to handle missing data. This is consistent with professional guidance on employment testing, which requires documentation of test development and validation regardless of the nature of the test:

> Reports of validation efforts should include enough detail to enable a testing professional competent in personnel selection to know what was done, draw independent conclusions in evaluating the research, replicate the study, and make recommendations regarding the use of the selection procedure.[92]

Sufficient documentation should include:

- stating the assumptions that are made;

[92] *See Principles*, at 33-34.

- clarifying the methods used to assess reliability and validity;
- explaining why those methods are the best methods for the given system and data sources; and
- describing how to interpret the output of those measures (especially when audiences are diverse in expertise).

This documentation also includes cases in which machine-learning itself is used to improve the quality of data used by selection algorithms. Machine learning can identify patterns (*e.g.*, missingness), associations (*e.g.*, auto-correlated features due to duplication), and unusual values or outliers. We recommend determining what an AI-enabled data cleaning technique should achieve and measuring its accuracy and reliability with metrics that are suitable for the particular data and use case.

As is required for more traditional testing, documentation in the AI context should be sufficient for computational reproducibility, *i.e.*, others should be able to reproduce the results of a model, given documented details around the data, code, and conditions.[93] Enough documentation should be produced to make transparent all of the different assumptions and choices that developers adopt to apply to a selection algorithm. We understand that complete computational reproducibility may be limited by concerns about maintaining data privacy as well as preserving trade secret and proprietary information.[94] To this end, some have proposed templates for documenting the data used in machine learning, as well as the key choices made in collecting, cleaning, training, and maintaining a dataset and building a selection algorithm.[95] By extending such templates to selection algorithms, developers and users of those algorithms could better maintain an in-depth understanding of the data, assumptions, and methodologies implemented. Such documentation supports not only computational reproducibility but also accords with analogous requirements for documenting selection procedures under *SIOP Principles* and under

[93] *See, e.g.*, 29 C.F.R. §1607.15(B)(11) (Reports of criterion-related validity studies should include "*Source data.* Each user should maintain records showing all pertinent information about individual sample members and raters where they are used, in studies involving the validation of selection procedures. These records should be made available upon request of a compliance agency. In the case of individual sample members these data should include scores on the selection procedure(s), scores on criterion measures, age, sex, race, or ethnic group status, and experience on the specific job on which the validation study was conducted, and may also include such things as education, training, and prior job experience, but should not include names and social security numbers. Records should be maintained which show the ratings given to each sample member by each rater"); *see also* 29 C.F.R. §1607.15(D)(11) (discussing source data for construct-related validity studies).
[94] *See* discussion of data privacy and deletion in the Transparency and Fairness Section, *supra*.
[95] Timnit Gebru, *et al*., *Datasheets for Datasets*, 64 Communications of the ACM 86-92 (2021), https://doi.org/10.1145/3458723.

the Uniform Guidelines for Employee Selection Procedures, discussed in the following Section of this Report.

UNIFORM GUIDELINES ON EMPLOYEE SELECTION PROCEDURES

I. Introduction

The Uniform Guidelines on Employee Selection Procedures (the "Guidelines" or "UGESP"), promulgated in 1978, remain the lens through which federal agencies, employment attorneys, and many courts are likely to analyze the job-relatedness and potential disparate impact of employment selection tools, including those based on Artificial Intelligence. The Guidelines pertain to tests or other selection procedures used as the basis for an "employment decision." They have their roots in the Supreme Court's 1971 decision in *Griggs v. Duke Power Co.,*[96] in which the Supreme Court articulated the "disparate impact theory of discrimination." Under *Griggs*, an employee selection process that disproportionately excludes a protected subgroup will run afoul of Title VII, unless the employer can demonstrate that its use of the assessment is "job-related and consistent with business necessity."[97] In 1978, the Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and Department of Justice jointly promulgated the Guidelines to establish a single, uniform set of principles for determining the proper use of pre-employment tests and other selection procedures under Title VII.[98]

In general, the Guidelines are not legally binding,[99] but some courts have recognized that they are entitled to "great deference."[100] At the same time, the field of industrial and organizational psychology has continued to develop over the more than four decades since the Guidelines were adopted, and in recent years commentators have criticized the Guidelines as being out of touch with the professional standards of that field.[101] Notably, however, the Guidelines explicitly

[96] 401 U.S.§ 424 (1971).

[97] Congress has codified the disparate impact theory at 42 U.S.C. §2000e-2(k).

[98] The Guideline apply to Title VII and Executive Order 11246, but do not apply to the ADA or ADEA. *See* 29 C.F.R. § 1607.2(D).

[99] For federal contractors under the oversight of the OFCCP, however, the Guidelines are binding regulations. 41 C.F.R. § 60-3.1, *et seq.*

[100] *Compare Clady v. Los Angeles County*, 770 F.2d 1421, 1428 (9th Cir. 1985) (explaining that the Guidelines "have not been promulgated as regulations and do not have the force of law"), *with Allen v. Entergy Corp*., 181 F.3d 902, 905 (8th Cir. 1999) (explaining that the Guidelines are "entitled to great deference").

[101] *See e.g.*, Matthew U. Scherer, Allan G. King & Marko J. Mrkonich, *Applying Old Rules to New Tools: Employment Discrimination Law in the Age of Algorithms*, 71 S.C. L. Rev. 449, 482 (2019) ("[T]he Guidelines' forty-year-old standards are overdue for revamping or replacement to bring them in line with the modern social science of test validity, which has evolved considerably in the decades since the Guidelines first appeared. . . . [t]he Guidelines' validation standards have, in fact, remained unchanged in the four decades since their promulgation. In the interim, the American Psychological Association (APA) has issued revised versions of the [Society for Industrial and Organizational Psychology] *Standards* . . .").

recognize that the science and technology of employee selection continues to evolve, and they are written broadly enough to provide a framework that can apply to today's Artificial Intelligence tools.[102]

Under UGESP, covered employment decisions include, but are not limited to, "hiring, promotion, demotion, membership (for example, in a labor organization), referral, [and] retention[.]" Other actions, "such as selection for training or transfer," may also be considered covered employment decisions if they lead to any of the decisions listed above.[103] The Guidelines do not apply directly to sourcing, recruiting, or other non-selection processes. This means that some commonly used AI-enabled recruiting tools – including "passive recruiting" tools – fall outside the Guidelines' purview.[104]

Under the Guidelines, the phrase "test and other selection procedure" is defined as any "measure, combination of measures, or procedure used as a basis for any employment decision." This includes "the full range of assessment techniques from traditional paper and pencil tests, performance tests, training programs, or probationary periods and physical, educational, and work experience requirements through informal or casual interviews and unscored application forms."[105] By these broad terms, the Guidelines' reach clearly encompasses AI-enabled selection tools used in the context of an employment decision.

A. <u>UGESP Principles</u>

The Guidelines establish detailed principles for determining the proper use of pre-employment tests and other selection procedures under Title VII. Under the Guidelines, employers have a threshold obligation to maintain and have available, *for each job*, information on whether the

---

[102] *See* U.S. Equal Emp. Opportunity Comm'n ("EEOC"), Questions and Answers to Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures ("Q&As"), https://www.eeoc.gov/laws/guidance/questions-and-answers-clarify-and-provide-common-interpretation-uniform-guidelines; 44 Fed. Reg. 11996 (Mar. 2, 1979) and 45 Fed. Reg. 29530 (May 2, 1980). *See also*, U.S. Dep't of Labor, Office of Federal Contract Compliance Programs, *Validation of Employee Selection Procedures* (last updated July 23, 2019), at Question 6, *https://www.dol.gov/agencies/ofccp/faqs/employee-selection-procedures#Q6*.
[103] 29 C.F.R. §1607.2(B).
[104] While the Guidelines do not apply directly to recruiting or sourcing tools, it is interesting to consider the extent to which the employer restricts the applicant pool to only those who have been recruited via the tool, and whether that may be considered a *de facto* hiring tool and thus within the purview of the Guidelines.
[105] 29 C.F.R.§1607.16(G); *see also* Q&A No. 5 (explaining that covered selection procedures include "interviews, review of experience or education from application forms, work samples, physical requirements, and evaluations of performance") and Q&A Nos. 6, 16.

"overall selection process" for that job has an adverse impact based on protected characteristics; that is, whether the selection process disproportionately excludes individuals of one race/ethnic or gender group more than others. If an overall selection process has an adverse impact, the employer must then assess the component selection procedures that make up the overall selection process to determine whether each component has adverse impact.[106] The Guidelines' adverse impact tracking and analysis requirements are addressed in more detail below in section II.

If an overall selection process has an adverse impact, this also triggers additional data-collection requirements. The employer must maintain and have available for agency review, for each component of the process that is determined to have adverse impact, documentation regarding the validity of that component.[107] As discussed in the Data Collection section of this Report, validity is the demonstration of the job-relatedness of a selection procedure. Colloquially, this means that the assessment scores are meaningful for their intended use. Validity is a property of how the assessment is used, not of the assessment itself.[108]

B. Validity Under UGESP

The Guidelines define three types of acceptable evidence of validity:[109]

---

[106] 29 C.F.R.§ §1607.4(A) and (C); 1607.15A(2)(a); Q&A No. 13. The overall selection process is the combined effect of all selection procedures leading to a final employment decision. The Guidelines call for records to be maintained by sex, and the following races and ethnic groups: Blacks, American Indians (including Alaskan Natives), Asians (including Pacific Islanders), Hispanics (including persons of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish origin or culture regardless of race), and whites (Caucasians) other than Hispanics. *See* 29 C.F.R.§1607.4(B).

[107] 29 C.F.R.§§1607.15(A)(2)(a) and (A)(3).

[108] This concept is perhaps most clearly described in the *Principles for Validation and Use of Personnel Selection Procedures* (2018) (the "Principles"), which are published by the Society for Industrial and Organizational Psychology ("SIOP"). The *SIOP Principles* are intended to "specify established scientific findings and generally accepted professional practice in the field of personnel selection psychology" and define validity as "the degree to which accumulated evidence and theory support specific interpretations of scores from a selection procedure entailed by the proposed uses of that selection procedure." *Principles* at 1 and 50.

We view the *SIOP Principles* as a useful resource, particularly given that UGESP notes that "[t]he provisions of these guidelines relating to validation of selection procedures are intended to be consistent with generally accepted professional standards for evaluating standardized tests and other selection procedures, such as those described in the Standards for Educational and Psychological Tests" developed by the American Psychological Association ("APA") and educational research organizations, and referenced in UGESP as the A.P.A. Standards. 29 C.F.R. §1607.5(C). The *Principles* are the current elaboration of the Standards applied to personnel selection and are an official policy statement of the APA. The *Principles* are additionally relevant because UGESP states "[t]hese guidelines are based upon and supersede previously issued guidelines on employee selection procedures. These guidelines have been built upon court decisions, the previously issued guidelines of the agencies, and the practical experience of the agencies, *as well as the standards of the psychological profession*." *Id.* at §1607.1(C) (emphasis added).

[109] 29 C.F.R. §1607.5(B).

- *Criterion-related validity* is demonstrated by identifying the criteria that indicate successful job performance and then correlating test scores with those criteria.
- *Content validity* is demonstrated when the content of a test closely approximates tasks to be performed on the job by the applicant.
- *Construct validity* is demonstrated when examinations are structured to measure the degree to which job applicants have identifiable characteristics that have been determined to be important in successful job performance.

These validation strategies are addressed in more detail in section III.

The Guidelines also establish highly detailed and technical standards for conducting a validity study and documenting the results, listing specific categories of information that must be included in, or provided with, such studies. Key among these is the requirement for job analysis, or a review of job information, to determine measures of work behavior or performance that are relevant to that job. However, certain criteria, including production rate, error rate, tardiness or absenteeism, or length of service may be used without a full job analysis if the employer can show the importance of the criteria to the particular employment context. A standardized rating of overall work performance may also be used where the study of the job shows that it is an appropriate criterion.[110]

The Guidelines further require that, in determining whether a selection procedure is appropriate for operational use, the degree of adverse impact of the procedure and the availability of other selection procedures of greater or substantially equal validity should also be taken into account.[111] A validation study should, therefore, explain (1) what alternatives were considered and how they were evaluated, and (2) why the version used was ultimately adopted. This requirement is addressed in more detail in section IV.

II. Evaluating AI Tools for Adverse Impact

The Guidelines establish a "rule of thumb" for determining adverse impact: Where the selection rate for one group is less than four-fifths of the selection rate for the group with the highest rate, then the selection process or procedure is presumed to be discriminatory.[112] The resulting

[110] *Id.* §§ 1607.14(A), 1607.14(B)(3).
[111] *Id.* §§ 1607.14(B)(6), (8).
[112] *See* 29 C.F.R. § 1607.4(D). For example, if the selection rate for white applicants is 80 percent and the selection rate for Asian applicants is 60 percent, the selection rate for Asian applicants would be 75 percent of the selection rate for white applicants. Because 75 percent is less than 80 percent (*i.e.*, four-fifths), the assessment would be presumed to have adverse impact.

calculation is called the "impact ratio."[113] The four-fifths rule is not, however, intended as a legal definition.[114] Rather, it is meant to establish a numerical basis for drawing an initial inference of discrimination. In other words, a process or procedure found to have adverse impact under the four-fifths rule will be deemed discriminatory unless it has been validated in accordance with the Guidelines.

The Guidelines acknowledge that the four-fifths rule is not necessarily reliable or appropriate in all circumstances. For example, while a selection ratio of greater than or equal to four-fifths (.8 or higher) is generally not regarded by Federal enforcement agencies as evidence of adverse impact, smaller differences in selection rates may nevertheless be viewed as adverse impact "where they are significant in both statistical and practical terms" or "where a user's actions have discouraged applicants disproportionately on grounds of race, sex, or ethnic group." Likewise, greater differences in selection rates may not constitute adverse impact, such as "where the differences are based on small numbers and are not statistically significant," or where "special recruiting or other programs cause the pool of minority or female candidates to be atypical" of the normal applicant pool.[115]

While courts often rely on the four-fifths rule to assess whether a plaintiff has established a *prima facie* case of disparate impact discrimination,[116] the rule has come under criticism, and courts routinely accept measures of statistical significance – the probability of an observed disparity occurring by chance – as being more probative than the four-fifths rule.[117]

In the context of disparate impact litigation, there are two related concepts of statistical significance that have been used most commonly: (1) standard deviations on a probability distribution and (2) probability levels. A standard deviation is a unit of measurement that allows

---

[113] Q&A No. 12.
[114] Q&A No. 11.
[115] 29 C.F.R. § 1607.4(D); *see also* Q&A Nos. 20, 21.
[116] *See, e.g., Stout v. Potter*, 276 F.3d 1118, 1124 (9th Cir. 2002) (applying "four-fifths rule of thumb" in considering adverse impact of selection procedures); *Boston Police Superior Officers Fed'n v. City of Boston*, 147 F.3d 13, 21 (1st Cir. 1998) (affirming district court's use of four-fifths rule in context of consent decree, holding that, although "violation of the four-fifths rule, standing alone, is not conclusive evidence of discrimination," it nonetheless serves as an "appropriate benchmark").
[117] Further discussion of the use of statistics for assessing adverse impact is in the Statistics and Adverse Impact Section of this Report.

analysis of disparities in common terms.[118] The greater the number of standard deviations from the mean, the greater the likelihood that an observed result is not due to chance. The Supreme Court has stated that "[a]s a general rule for ... large samples, if the difference between the expected value and the observed number is greater than two or three standard deviations, then the hypothesis that the [result] was random would be suspect to a social scientist."[119] Probability levels (also called "p-values") are the probability that the observed disparity is random. A 0.05 probability level means that one would expect to see the observed disparity occur by chance only one time in twenty cases; that is, there is only a five percent chance that the disparity is random (due to chance alone). Courts have accepted p-values of 0.05 or below as sufficient to rule out the possibility that the disparity occurred at random and is therefore not meaningful.[120]

It is important to note the criticality of conducting an analysis that mirrors the reality of the employment decision making context.[121] The analyst should think about unit of analysis considerations including dimensions such as time period, location, job, sub-applicant pool, and any other strata of interest. This nuance is not always captured in the Guidelines. For example, the Guidelines require employers to make determinations as to whether their overall selection process has adverse impact – and, if so, whether their component selection procedures have adverse impact as well – "at least annually" for each group that "constitutes at least 2 percent of the labor force in the relevant labor area or 2 percent of the applicable workforce."[122] Annual adverse impact determinations may not, however, account for selection procedures that – like some AI-enabled tools – change the characteristics used to evaluate applicants, scoring

[118] Technically, a standard deviation is defined as "a measure of spread, dispersion, or variability of a group of numbers equal to the square root of the variance of that group of numbers." David Baldus & James Cole, *Statistical Proof of Discrimination* 359 (McGraw-Hill Book Co., 1980).
[119] *Castaneda v. Partida*, 430 U.S. 482, 496 n.17 (1977). In *Castaneda*, the court defined standard deviation as "the measure of predicted fluctuations from the expected value," *i.e.*, how much a group is underrepresented; *see also Waisome v. Port Auth.*, 948 F.2d 1370, 1376 (2d Cir. 1991) ("Social scientists consider a finding of two standard deviations significant, meaning there is about one chance in 20 that the explanation for a deviation could be random and the deviation must be accounted for by some factor other than chance").
[120] *Palmer v. Shultz*, 815 F.2d 84, 92-96 (D.C. Cir. 1987) (noting that "statistical evidence meeting the .05 level of significance ... [is] certainly sufficient to support an inference of discrimination").
[121] *See* Center for Corporate Equality, *Technical Advisory Committee Report on Best Practices in Adverse Impact* (September 2010), https://irp.cdn-website.com/b44ff977/files/uploaded/TAC-Adverse-Impact-FINAL.pdf.
[122] 29 C.F.R. §§ 1607.4(C); 1607.15(A)(2)(a). The Guidelines recognize that, in some circumstances, annual analyses may be unreliable because an insufficient number of selections occurred for a job in a year, *id.* § 1607.4C; 1607.15A(2)(c), in which case the employer should continue to collect, maintain, and have available "information on the impact of the selection process and the component procedures until it can determine that adverse impact does not exist for the overall process or until the job has changed substantially." Q&A No. 85.

algorithms, and even assessment content as they learn or evolve. At the same time, a requirement of more-frequent determinations of adverse impact, such as each time a tool evolves, may be impractical and may well be based on so few observations that meaningful analyses are not feasible. This begs the question of when a tool has evolved enough to constitute a new selection procedure, warranting a refreshed analysis. As such, it may be advisable for employers to consider developing a policy for version control and associated guard rails around substantive changes to the functioning of the AI-enabled tool in employment selection contexts.

The appropriateness of any adverse impact measurement method is in part a function of how employment decisions are made and the research questions being asked. The decision to analyze data within or across time periods, locations, jobs, requisitions, or any other strata should be made after considering the selection procedure of interest as well as the theoretical and empirical similarity of the strata of interest. In some situations, more complicated statistical techniques may be necessary to appropriately analyze more complex data structures. *See* further discussion in the Statistics and Adverse Impact Section of this Report, below.[123]

III. Validation of AI Assessments Under UGESP

UGESP remains relevant despite advances in science and professional practice because it provides the process for demonstrating and documenting selection procedure validity in the equal employment opportunity context. It is important to note that a given selection procedure may be appropriate – that is, valid – for some uses but not others. To give an example, if an AI-enabled tool is used to screen applicants out of a selection process, then the underlying basis for the screen-out must be essential to the job in question. There must be a job-related rationale for why those above the cut score are selected and those below are not. To give another example, if an AI-enabled tool is used for ranking applicants, the tool must cover enough of the success factors of the job (or enough of a specific critical factor if that is the purpose) so that the ranking is meaningful (*i.e.*, there is meaningful differentiation among scores). In other words, to say only that an AI process predicts something related to the job does not include the specificity needed to establish how scores are to be used and so is not sufficient on its own to establish validity.

[123] Readers may also want to consult the *Technical Advisory Committee Report on Best Practices in Adverse Impact Analyses*, as some of that report's recommendations, although made in the context of traditional employment selection processes, may also be applicable to AI-enabled ones.

A. Potential Complexity Stemming From AI-Enabled Tools

In theory, AI-enabled tools can be evaluated using validation research like any other selection procedure. For tools that involve the analysis of existing data as part of the development process, criterion-based validation strategies may be feasible and can be frequently performed. However, some commentators have raised concerns that UGESP is not suited to determining the validity of certain types of AI applications. Among these concerns are:

- "Black box" algorithms that make it difficult to specify what the selection procedure is, the rationale for how it works, and the theory for why it works. This can make demonstration of validity difficult, the more so if each component of the final model is a discrete procedure. For example, a given AI application could generate unknown derivatives that ultimately make a determination that a non-job relevant social or athletic activity is correlated with successful job performance without providing any explanation as to why that is the case.
- With some machine learning, the algorithms can develop their own final parameters for assessment after more experience with data; they are not limited to initial attributes provided at the start of their training. UGESP does not address selection procedures that modify themselves over time. For that reason, the Guidelines do not indicate how often such a procedure should be validated.
- The line between legitimate variables included in a selection procedure that correlate with protected class (measurement of a relevant predictor and outcome is confounded with protected class) and variables that are proxies for protected class is thin.

It should be noted that most of these concerns implicate complex machine learning applications, which sometimes operate with little to no rationale for linking predictors and criteria to an actual job. These types of applications may be challenged under UGESP's provisions addressing job analysis and avoidance of capitalization on chance. Notably, UGESP warns that studies claiming validity featuring many subjects and low correlation between predictor and criterion "will be subject to close review" by federal agencies if there is also large adverse impact.[124] However, this caveat does not set a minimum for the statistical relationship.

B. What Validation Strategies Are Feasible?

More recent professional guidance (*e.g.*, *SIOP Principles*) describe UGESP's three versions of validity (criterion, content, and construct) as *strategies* for demonstrating validity, not different

[124] 29 C.F.R § 1607.14(B)(6).

types of validity[125] Regardless of the wording, the primary means of validity demonstration in employment testing remain assessing (1) the statistical relationship between predictor and criterion (UGESP "criterion validity") and (2) representation of the content of the job within the assessment (UGESP "content validity"). UGESP's construct validity is commonly considered a combination of both criterion and content validity, and current professional discussion views construct validity as the totality of validity evidence, including how an assessment relates to other assessments of conceptually similar or different underlying constructs. While the Guidelines do not indicate a requirement that the selection procedure be demonstrated as causally related to performance on the job, the inclination of contemporary experts to consider the totality of validation evidence suggests that a causal inference may be at the heart of more recent professional guidance.

The nature of many AI-enabled tools and common development processes tends to align most closely with a criterion strategy, because in many situations, an existing database is used as the starting point for model development and prediction. Depending on the inputs to the system and their overlap with job requirements, however, a content validation approach may be similarly appropriate. Construct validation may be relevant where an examination of the interrelationships among a large number of predictors and outcomes could inform the nature of the constructs and their interrelationships with outcomes.[126]

1. Criterion-Related Validation Research

Assuming that data are available and that predictor and criterion conditions are met, criterion research assessing the relationship between a selection procedure and some measure of work behavior may be feasible for AI applications. However, there may be obstacles to using this approach in the context of some AI-enabled applications, including the following:

- AI research may focus primarily on establishing a statistical relationship between a set of predictors and an outcome, sometimes without considering the nature of either predictor or outcome. Thus, whether the relationship is meaningful using theory or logic is not considered. The fact that there is a relationship alone is taken as

[125] *See e.g.*, *Principles*, at 28-29 ("Once testing professionals have worked with the organization to define its objectives for developing a selection procedure, understand the requirements of the work, and reach agreement on the type of selection procedure(s), testing professionals must decide what validation strategy or strategies will be pursued to accumulate evidence to support the intended inference(s).").
[126] *See* Statistics and Adverse Impacts Section, below, for further discussion of the application of different validation strategies.

establishing validity. While the Guidelines do not require theoretical analysis, they do stress the importance of job relevance and state that "[a]ny validity study should be based upon a review of information about the job for which the selection procedure is to be used."[127]

- Relationships are generally tested for statistical significance, *i.e.*, the probability that the relationship is due primarily to chance is low. The logic of statistical significance testing means that, as the number of observations in the analysis becomes large, even a relationship of trivial magnitude could be statistically significant. While the Guidelines state that "[g]enerally, a selection procedure is considered related to the criterion, for the purposes of these guidelines, when the relationship between performance on the procedure and performance on the criterion measure is statistically significant at the 0.05 level of significance…" they also state that "Users should avoid reliance upon techniques which tend to overestimate validity findings as a result of capitalization on chance…."[128]
- In traditional criterion validation, having some idea as to the nature of the underlying construct being measured can provide some insight into how well it is being measured. This consideration may be more nuanced for more complicated AI tools. Low correlations between predictors and criteria may indicate inefficient prediction, particularly when prediction is lower than what is expected based on the literature.[129] Perhaps not enough of the criterion construct is being captured in either predictor or criterion measurement (assuming that there is a coherent criterion construct to measure), or construct contamination obscures the relationship under examination. A statistically significant predictor alone only indicates a relationship other than zero between predictor and criterion in the sample population from which the data came.

2. Analysis of Work in the Context of Criterion-Related Validation Research

UGESP has a strong focus on job behaviors and work products in the context of validation.

Obstacles this may pose in the context of AI tools include the following:

- Identification of critical or important duties, work behaviors, or work outcomes to be used as criteria are to be identified, usually through a "full"[130] job analysis. In the criterion validation context, job analysis appears geared toward the development of criterion measures. The rigor of job analysis used to develop AI applications may vary depending on context.

---

[127] 29 C.F.R. § 1607.14(A).
[128] *Id.*, at § 1607.14(B)(5) and (6).
[129] Evaluating the results of a criterion validation study in light of literature summarizing similar variable relationships can be particularly informative. That is to say, are the results of a local study similar to what we would expect based on cumulative scientific knowledge? This comparison can be difficult if it is unclear exactly what the AI application is assessing.
[130] The *SIOP Principles* define an "analysis of work" as "[a]ny method used to gain an understanding of the work behaviors and activities required, or the worker requirements (*e.g.,* knowledge, skills, abilities, and other personal characteristics), and the context or environment in which an organization and individual may operate."). *Principles*, at 46.

- Certain criteria may be used without full job analysis (*e.g.,* error rate, tardiness, absenteeism, length of service) if the importance of the criteria to the particular employment context can be shown. This may preclude AI applications that match profiles of applicant characteristics to those of "top performers" without predicting top performance itself.
- Job analysis is seen as essential when "transporting" criterion validity evidence from one situation to another. How much similarity between situations is necessary is a matter of debate, particularly in light of post-UGESP discussions regarding the appropriateness of generalization of validity results; but in those circumstances there is customarily a supporting theory that may not be relevant to AI applications.
- Some AI-enabled employment selection tools are designed in a bespoke manner and thus designated for use specifically in the context in which they were created. While some experts may interpret these tools as limited by their local validation strategies and lacking reliability, the Guidelines stress the importance of validation evidence for the job in its specific context over general validation evidence across contexts.

UGESP's technical standards for criterion validity also contain a lengthy discussion of "unfairness," much of it dealing with the lack of feasibility in evaluating fairness.[131] From a statistical perspective, this infeasibility of fairness analyses is often due to insufficient numbers of subjects in the demographic groups of interest. The science in this regard has matured and recent research has identified some limitations with localized fairness analyses.[132] Some constructs and their tests (*e.g.*, cognitive ability) have been extensively researched. But there is no comparable research for novel constructs assessed within some AI applications, which can vary in how they are constructed, what predictors are used, and what the predicted outcomes are.

3. Overstating Validity

UGESP notes that overstatement of validity findings due to "capitalization on chance" must be avoided "unless an appropriate safeguard is taken."[133] One such instance illustrating capitalization on chance is reliance on a select number of predictors and criteria where many were studied. Another is "use of optimal statistical weights for selection procedures computed in one sample."[134] These provisions of the Guidelines might call into question the selection of a final AI model from several models produced during application development, unless it can be

[131] 29 C.F.R § 1607.14(B)(8).
[132] *See, e.g.*, Christopher M. Berry & Peng Zhao, *Addressing criticisms of existing predictive bias research: Cognitive ability test scores still overpredict African Americans' job performance*, J. of Applied Psych., 100, 162–179 (2015).
[133] 29 C.F.R. § 1607.14(B)(7).
[134] *Id*.

shown that statistical principles were not violated. The Guidelines indicate that "use of a large sample is one safeguard, cross-validation is another."[135]

C. What About Methods Not Specifically Mentioned in UGESP But Generally Considered Acceptable?

There is a distinction between validation strategy and the specific research tactics used to implement that strategy. UGESP does not discuss nuanced research tactics but does state that the Guidelines are intended to be consistent with generally accepted professional standards found in the APA *Standards*, textbooks, and journals.[136] Whatever methods are used must be directed at demonstrating validity. UGESP also does not involve itself in the details of how reliability and validity might be addressed for specific kinds of applications. For establishing a statistical relationship between predictors and criterion, UGESP advises "using professionally accepted statistical procedures" without further specification that might limit methods used with AI applications.[137] The conventional statistical significance standard, with probability of a false conclusion of validity at 0.05 or lower, would continue to apply, with safeguards as described above. It is possible that new statistical approaches, both for AI-enabled and conventional assessments, could use some other definition and benchmark for statistical significance. Practical significance will also be a consideration, and is commonly assessed by a variety of effect sizes that quantify the magnitude of relationship between predictor and criterion.

IV. Suitable Alternatives and "Debiasing"

A. Suitable Alternatives in UGESP

While the idea of "debiasing" models for use in developing employment selection tools may seem like a new or novel concept in the age of AI, the general approach of searching for suitable alternative selection procedures with less adverse impact goes back at least as far as the Guidelines. There are several references to suitable alternatives in UGESP, but the concept is discussed most directly in "Discrimination Defined: Relationship between use of selection

[135] *Id.* For machine learning applications, cross-validation performance (*i.e.*, a form of concurrent criterion validation) is often evaluated using a metric known as the receiver operating characteristic area under the curve ("ROC AUC", or simply "AUC"). The AUC represents the probability that a randomly drawn true positive (*i.e.*, "good candidate") will be rated higher than a randomly drawn true negative (*i.e.,* "bad candidate"). Therefore, an AUC of .80 means that a model correctly classified two users as a "fit" and "not a fit" 8 times out of 10. The model with the highest AUC is considered to be the best fitting model.
[136] *Id.* at § 1607.5(C).
[137] *Id.* at § 1607.14(B)(5).

procedures and discrimination."[138] After first specifying the core tenet that a selection procedure having adverse impact constitutes discrimination unless justified as valid, the section goes on to provide direction regarding alternative selection procedures: "Where two or more selection procedures are available which serve the user's legitimate interest in efficient and trustworthy workmanship, and which are substantially equally valid for a given purpose, the user should use the procedure which has been demonstrated to have the lesser adverse impact.… Whenever the user is shown an alternative selection procedure with evidence of less adverse impact and substantial evidence of validity for the same job in similar circumstances, the user should investigate it to determine the appropriateness of using or validating it in accord with these guidelines."[139]

B. Debiasing of AI Tools

If the process of debiasing an AI-enabled tool means selecting the version of the tool that has the least evidence of adverse impact and has substantially similar validity evidence to other versions, then it aligns closely to the consideration of suitable alternatives as described in the Guidelines. When applying this logic to the development or validation of an AI-enabled tool, the following questions arise:

- *Can a model or algorithm be considered a selection procedure?* UGESP defines a selection procedure as "any measure, combination of measures, or procedure used as a basis for any employment decision. Selection procedures include the full range of assessment techniques . . . ."[140] Based on this definition, a model or algorithm can be considered a selection procedure, in the context of an AI-enabled tool, if it is used as a "basis for any employment decision." As such, choosing the model or algorithm that has been demonstrated to have the lesser adverse impact would constitute an alternative selection procedure, so long as the models/algorithms are substantially equally valid for a given purpose.

- *What if the AI-enabled tool utilizes many individual models or algorithms to reach a score or recommendation?* The Guidelines direct the user to analyze the effect of tools and screens as they are used in the particular employment selection context. Where multiple equations, points of analysis, or AI-endabled models are combined into one recommendation or decision point stemming from the AI tool of interest, it would be appropriate for the user to analyze and document impact, validation, and consideration of suitable alternatives at that recommendation or decision point. This could be considered similar to the more manual practice of combining aspects or

[138] *Id.* at § 1607.3.
[139] *Id.* at § 1607.3(B).
[140] *Id.* at § 1607.16(R).

components of traditional assessments or screens in order to reduce adverse impact and maximize validity (such as combining aspects or components of cognitive ability or personality assessments into one screen or assessment).

- *Can an AI-enabled tool be debiased without regard to validation evidence?* The Guidelines suggest that users should investigate suitable alternatives with lesser adverse impact that are "substantially equally valid for a given purpose."[141] While the Guidelines do not define what is meant by "substantially equally valid" they do indicate that in the consideration of suitable alternatives validity must be taken into account and evidence compared across different tools or screens. This is important for users to keep in mind in the context of "debiased" AI-enabled selection tools.

C. Debiasing Approaches Discussed in Contemporary Guidance

As general guidance and best practice recommendations continue to arise from various government and non-governmental sources, it will be important for users of employee selection procedures to keep in mind that legal requirements by sector remain in place. At the time of this writing, contemporary guidance for the consideration, management, and governance of AI risks across sectors is emerging from the National Institute of Standards and Technology ("NIST") within the U.S. Department of Commerce. In the context of developing its AI Risk Management Framework, NIST has recently released a special publication that discusses debiasing techniques in three distinct categories:

- *Pre-processing:* transforming the data so that the underlying bias is mitigated. This method can be used if a modeling pipeline is allowed to modify the training data.
- *In-processing:* techniques that modify the algorithms in order to mitigate bias during model training. Model training processes could incorporate changes to the objective (cost) function or impose a new optimization constraint.
- *Post-processing:* typically performed with the help of a holdout dataset (data not used in the training of the model). Here, the learned model is treated as a black box and its predictions are altered by a function during the post-processing phase. The function is deduced from the performance of the black box model on the holdout dataset. This technique may be useful in adapting a pre-trained large language model to a dataset and task of interest.[142]

Some of these techniques may be more appropriate from a Guidelines perspective than others:

- *Pre-processing* debiasing techniques in the context of employee selection tools might involve modifying training data pre-build. As such, these techniques seem most

[141] *Id.* at § 1607.3(B).
[142] Reva Schwartz, et al., *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence, Special Publication 1270*, National Institute of Standards and Technology (2022), at 28, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf.

aligned with the Guidelines. For example, a system that is trained on balanced data (representative of people who come from different demographic backgrounds), as opposed to unbalanced data (raw), might evidence less adverse impact but similar validity in testing. The comparison of these outcomes and choice of the system trained on balanced data would seem reasonable in the context of developing employee selection tools under the Guidelines.

- *In-Processing* techniques involve modifying algorithms or models during their build, pre-deployment. In-Processing techniques could be aligned with the Guidelines; however, this is where intentional discrimination or disparate treatment could occur if the system is not carefully designed and appropriately transparent to ensure UGESP alignment. It is not appropriate to simply include a demographic identifier as an input to the algorithm, even if the purpose is to mitigate adverse impact in the system's recommendations. However, if the developer imposes an optimization constraint to minimize adverse impact in recommendations that does not include demographic information in the algorithm, such a constraint could be in alignment with the Guidelines. For example, a developer might test hundreds of variants of an employment selection algorithm designed for maximum validity against a dataset that includes demographic identifiers (*e.g.*, race and gender) to select the version with the least evidence of disparate impact and substantially similar validity. So long as this process occurs pre-deployment, the approach would seem aligned with the Guidelines.
- *Post-processing* techniques involve modifying the outputs of algorithms post-deployment. These techniques seem least aligned with the Guidelines, and most likely to be challenged on the grounds of intentional discrimination or disparate treatment, to the extent that the modification affects the recommendations flowing from the selection system after those recommendations are made. This type of approach would involve altering the outputs of the selection process post-deployment, which has been expressly considered disparate treatment at least since the passage of the Civil Rights Act of 1991.[143]

A fulsome discussion of algorithmic debiasing techniques is beyond the scope of this Report. Interested readers are encouraged to consider the developing work under NIST, but are reminded to check these approaches against requirements in the employment law space. The takeaway from this brief discussion is that certain debiasing techniques should be considered aligned with the Guidelines under the suitable alternatives directive, but that it is important for compliance experts to carefully inspect and fully understand the details of the debiasing solution before it goes forward. What may be considered a reasonable debiasing approach for an AI-enabled tool outside of the employment selection context may or may not generalize as reasonable for consequential employment decisions that are subject to UGESP's requirements.

[143] 42 U.S.C. § 2000e-2(l).

STATISTICS AND ADVERSE IMPACT

I. Introduction

The advent and adoption of AI-enabled tools to inform employment decisions is an occasion to revisit the ways we identify whether such tools have an impermissible "adverse impact." While the previous Section of this Report focused on the analysis of adverse impact and validation specifically in the context of the Uniform Guidelines for Employment Selection Procedures ("UGESP"), this Section takes a broader view. Below we will reference UGESP, but also important developments in case law as we consider adverse impact analysis in the context of AI-enabled employment selection tools.

The typical means of assessing adverse impact is through one or more statistical methods. But the Supreme Court has made clear that "statistics are not irrefutable; they come in infinite variety and, like any other kind of evidence, they may be rebutted" and "their usefulness depends on all of the surrounding facts and circumstances."[144] In other words, there is no one-size-fits-all statistical model that necessarily, unambiguously, and in every case "proves" adverse impact or discrimination.

As AI-enabled assessments and employment decision making tools become more common and are deployed at both modest and massive scales, it becomes increasingly important to understand how different methods of assessing adverse impact might suggest opposing verdicts on whether a given tool can or cannot be used or, at least, whether its use must be justified by proof of job-relatedness, and enjoined if there is a less discriminatory alternative that can be shown to serve the employer's purpose equally well. This Report Section surveys some of the key decision points that need to be considered in determining how to assess the adverse impact of a given tool, along with some of the unique challenges posed by "intelligent" AI (that is, tools that adjust their selection algorithm or scoring criteria over time as they "learn" more about the characteristics of successful applicants or job incumbents, which themselves are changing over time).

One guiding principle is that mere bottom-line comparisons of, *e.g.*, applicants to hires is unlikely to prove meaningful as a way to assess "adverse impact" in compliance with legal standards. The absence of bottom-line impact could mask impermissible impact at intervening steps; conversely, the presence of aggregate, bottom-line impact need not entail that any

---

[144] *Int'l Bhd. of Teamsters v. United States*, 431 U.S. 324, 340 (1977).

particular step in the process has legally impermissible adverse impact. In other words, looking only at the bottom-line could yield both "false negatives" and "false positives."[145]

In this vein, the Supreme Court has cautioned that "the inevitable focus on statistics in disparate impact cases could put undue pressure on employers to adopt inappropriate prophylactic measures."[146] Accordingly, the Court underscored, proving a case of disparate impact "goes beyond the need to show that there are statistical disparities in the employer's workforce."[147] Although identifying adverse impact is only the first step in establishing a disparate impact claim, the Court has long highlighted the importance of job-relatedness, holding that "good intent or absence of discriminatory intent does not redeem employment procedures or testing mechanisms that operate as 'built-in headwinds' for minority groups and are unrelated to measuring job capability."[148] Again, though, the Court has eschewed any appeal to bright-line rules or a single numeric standard: "Our formulations, which have never been framed in terms of any rigid mathematical formula, have consistently stressed that statistical disparities must be sufficiently substantial that they raise such an inference of causation."[149]

Among the issues surveyed in this Section is the need for employers, litigants, and courts to understand what different ways of measuring "adverse impact" do and do not actually entail, and why different models actually test different hypotheses and answer different questions. This is key: "[The] failure to identify the specific practice being challenged is the sort of omission that could result in employers being potentially liable for the myriad of innocent causes that may lead to statistical imbalances."[150] Understanding what it means to select different dependent or independent variables for study, or different benchmarks of "neutrality," can aid in the

---

[145] *See, e.g., Connecticut v. Teal*, 457 U.S. 440 (1982) (importance of examining each step in selection system).
[146] *Watson v. Fort Worth Bank and Trust*, 487 U.S. 977, 997 (1988) (O'Connor, J., plurality opinion).
[147] *Id. accord E.E.O.C. v. Joe's Stone Crab, Inc*., 220 F.3d 1263, 1276 (11th Cir. 2000) ("[H]olding employers liable for statistical imbalances per se is inconsistent with Title VII's plain language and statutory purpose."). It is worth noting, however, that in *Joe's Stone Crab* the problem identified by the Court was not lack of statistical evidence of disparate outcomes, but that the EEOC challenged a subjective selection practice; as a result, there was no direct causal link shown, and a causal link between hiring practices and adverse impact is a requirement for disparate impact claims. The Eleventh Circuit actually found the evidence potentially probative of intentional discrimination and remanded the case to address that claim. *Watson* also involved a challenge to subjective decision-making, where causation was again a fuzzier question. With AI-enabled selection devices, in contrast, just as with more traditional screening tests, the question of causation is generally straightforward.
[148] *Griggs v. Duke Power Co.*, 401 U.S. 424, 432 (1971).
[149] *Watson*, 487 U.S. at 997; *accord Segar v. Smith*, 738 F.2d 1249, 1273 (D.C. Cir. 1984) (citation and internal quotation marks omitted) ("Because the facts necessarily will vary in Title VII cases, a specific test for the sufficiency of a plaintiff's initial proof is not possible.").
[150] *Smith v. City of Jackson*, 544 U.S. 228, 241 (2005) (citation and internal quotation marks omitted).

interpretation of what the results of a given statistical test mean for whether and which underlying practices need to be modified. This Section aims to further that discussion by advocating for a deeper appreciation of the statistical tools most commonly brought to bear and the key modeling assumptions and choices that can, in many cases, make the difference between whether a given AI-enabled employment tool "passes" or "fails" the adverse impact test.

II. Legal Background

Title VII of the Civil Rights Act of 1964 ("Title VII") prohibits discrimination in employment because of an individual's race, color, religion, sex, or national origin.[151] Specifically, Title VII makes it "an unlawful employment practice for an employer" to "fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual … because of such individual's race, color, religion, sex, or national origin."[152] Title VII also makes it unlawful for an employer "to limit, segregate, or classify his employees or applicants for employment in any way which would deprive or tend to deprive any individual of employment opportunities or otherwise adversely affect his status as an employee, because of such individual's race, color, religion, sex, or national origin."[153]

Title VII violations generally are of two kinds: disparate treatment and disparate impact discrimination. Disparate treatment involves intentional discrimination. Disparate impact discrimination, by contrast, occurs when an employer uses a facially neutral employment practice that "causes a disparate impact on the basis of race, color, religion, sex, or national origin and the [employer] fails to demonstrate that the challenged practice is job-related for the position in question and consistent with business necessity."[154] Disparate impact discrimination can occur when a selection device causes an adverse impact on individuals of a particular protected group (defined by race, gender, etc.) and is not otherwise justified by Title VII defenses, such as job-relatedness and business necessity.[155]

[151] *See* 42 U.S.C. § 2000e-2(a).
[152] *Id*. § 2000e-2(a)(1).
[153] *Id*. § 2000e-2(a)(2).
[154] 42 U.S.C. § 2000e-2(k)(1)(A)(i).
[155] *See, e.g., Albemarle Paper Co. v. Moody*, 422 U.S. 405 (1975) (striking down pre-employment test under disparate impact analysis); *Griggs*, 401 U.S. at 425 (Title VII forbids the use of employment tests that are discriminatory in effect unless the employer meets "the burden of showing that any given requirement [has] …. a manifest relationship to the employment in question").

A. Title VII Disparate Impact Standards

Courts apply a three-step test when they analyze alleged disparate impact under Title VII. First, a plaintiff must make a *prima facie* case that an employer uses a particular employment practice that causes a disparate impact on the basis of a prohibited factor, such as race or gender.[156] If a plaintiff fails to make this showing, the case will be dismissed.[157]

If, however, a plaintiff establishes a *prima facie* case, the case proceeds to the second stage of the analysis: whether the employment practice that was shown to cause the adverse impact "is job-related for the position in question and consistent with business necessity."[158] If the employer carries that burden, the analysis addresses a third and final question: whether the employer refused to adopt an available alternative employment practice that has less disparate impact yet would still serve the employer's legitimate needs.[159]

When analyzing whether the use of an employment test has violated Title VII, the focus of the inquiry is on whether the test as applied to a specific set of individuals seeking a specific job has produced a disparate impact that is not job-related. Although a user of a test – *i.e*., an employer – may rely on validity studies or manuals created by test publishers, the users themselves are responsible for compliance with Title VII and UGESP.[160] Accordingly, the questions of whether a test caused a disparate impact and whether use of that test was job-related, are specific to the employer and not to the test itself. (*See* further discussion of job-relatedness and alternatives, below.)

---

[156] *See* 42 U.S.C. § 2000e-2(k)(1)(A)(i); *Lewis v. City of Chicago, Ill.*, 560 U.S. 205 (2010); *Ricci v. DeStefano*, 557 U.S. 557 (2009).
[157] *See, e.g., Grant v. Metropolitan Gov't of Nashville*, 446 F. App'x 737, 742 (6th Cir. 2011); *Aliotta v. Bair*, 614 F.3d 556, 561 n.4 (D.C. Cir. 2010).
[158] 42 U.S.C. § 2000e-2(k)(1)(A)(i).
[159] 42 U.S.C. § 2000e-2(k)(1)(A)(ii) & (C); *see generally Ricci*, 557 U.S. at 578; *Watson*, 487 U.S. at 997-98.
[160] 29 C.F.R. § 1607.7(A). The Guidelines state that "Users may, under certain circumstances, support the use of selection procedures by validity studies conducted by other users or conducted by test publishers or distributors and described in test manuals. *While publishers of selection procedures have a professional obligation to provide evidence of validity which meets generally accepted professional standards …, users are cautioned that they are responsible for compliance with these guidelines.* Accordingly, users seeking to obtain selection procedures from publishers and distributors should be careful to determine that, in the event the user becomes subject to the validity requirements of these guidelines, the necessary information to support validity has been determined and will be made available to the user." (Emphasis added.)

B. Whether An Employment Practice Has A "Disparate Impact"

Plaintiffs seeking to prove that a "particular employment practice" has "cause[d] a disparate impact on the basis of race, color, religion, sex, or national origin"[161] most frequently rely on the use of statistics. Courts apply a case-by-case approach for using statistics to determine whether any numerical disparities allegedly caused by discrimination are significant or substantial, and (where at issue) whether the statistics prove that the challenged practice "cause[d]" disparate outcomes by race, gender, or another protected characteristic.[162]

As noted in the previous Section of this Report, two statistical devices are most frequently used to prove disparate impact: (1) the EEOC's four-fifths or 80% rule; and (2) standard deviations, which measure the extent to which outcomes for different groups deviate from the expected distribution. The four-fifths rule has been adopted by the EEOC through UGESP.[163] Under that rule, federal enforcement agencies may infer adverse impact, and the plaintiff may argue that they have achieved a key step in proving disparate impact, if the minority group is selected at a rate less than 80% of the rate for the most-selected group on the test or other selection device or screen.[164]

The second test frequently used by courts in assessing disparate impact is a standard deviation analysis. This method begins with the hypothesis that group membership does not influence hiring rates. Experts then conduct a statistical analysis to measure the probability that the result

---

[161] 42 U.S.C. § 2000e-2(k)(1)(A)(i).

[162] *Id. See Isabel v. City of Memphis,* 404 F.3d 404, 412 (6th Cir. 2005) (noting that nothing forbids reliance on varying statistical analyses and that courts "require only that the statistical analyses be 'relevant'") (internal citation omitted). *See also Ricci,* 557 U.S. at 587 (describing EEOC's 80% rule as a "rule of thumb for the courts"); *but see Watson,* 478 U.S. at 995-96, n. 3 (noting that the 80% rule "has been criticized on technical grounds.").

[163] 29 C.F.R. § 1607.4(D) ("Adverse impact and the four-fifths rule. A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact. Smaller differences in selection rate may nevertheless constitute adverse impact, where they are significant in both statistical and practical terms or where a user's actions have discouraged applicants disproportionately on grounds of race, sex, or ethnic group. Greater differences in selection rate may not constitute adverse impact where the differences are based on small numbers and are not statistically significant, or where special recruiting or other programs cause the pool of minority or female candidates to be atypical of the normal pool of applicants from the group.")

Although UGESP is codified in the regulations of the EEOC, they are Guidelines, not formally adopted EEOC regulations. In contrast, the Department of Labor's Office of Federal Contract Compliance Programs ("OFCCP") has formally adopted the UGESP as regulations.

[164] *See, e.g., Smith v. Xerox Corp.*, 196 F.3d 358, 365-70 (2d Cir. 1999), *overruled on other grounds by Meacham v. Knolls Atomic Power Lab.*, 461 F.3d 134 (2d Cir. 2006); *Vulcan Society, Inc. v. City of New York*, 637 F. Supp. 2d 77 (E.D.N.Y. 2009).

that actually occurred could have been a random deviation from the starting hypothesis of equal hiring rates. As the number of standard deviations between the observed hiring rates and the hypothesis of equal hiring rates increases, courts are more likely to infer that the result runs afoul of Title VII.[165] Although the Supreme Court has not set a bright line test for when the standard deviations test conclusively establishes discrimination,[166] both it and other courts will generally infer disparate impact at two or three standard deviations.[167]

III. Adverse Impact Statistics

With that legal framework in mind, we turn to an overview of how different statistical measures may bear on the use of AI tools. Consider comparing pass rates between Black and White job applicants on a pre-employment test. If their pass rates are exactly equal, then the case for test fairness may seem clear, absent other considerations. However, often there are observed pass rate disparities – such as when 15% of Black applicants but 20% of White applicants pass the test – that may or may not lead to a legal case concerning adverse impact. Based on two pass rates such as these, one might choose any of a wide range of adverse impact statistics that, while equally reasonable, might reach different conclusions about the presence or absence of adverse impact.[168] For a review of the specific statistical tests utilized by the Office of Federal Contract Compliance Programs ("OFCCP") when auditing federal contractors, see the agency's FAQ on investigating disparities.[169]

A. Considerations of Statistical and Practical Adverse Impact Analyses

Generally speaking, adverse impact statistics (like any statistic) help one address conclusions regarding *statistical significance* and/or *practical significance*. *Statistical significance* refers to whether or not the data are consistent with an initial (null) hypothesis, as noted above with the

---

[165] *See, e.g., Vulcan Society, Inc.*, 637 F. Supp. 2d at 87 (discussing *Waisome v. Port Auth. Of New York & New Jersey*, 948 F.2d 1370, 1376 (2d Cir. 1991)).
[166] *Watson*, 487 U.S. at 995 n. 3.
[167] *See, e.g., Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 308 n. 14 ("'[A]s a general rule for such large samples, if the difference between the expected value and the observed number is greater than two or three standard deviations,' then the hypothesis that teachers were hired without regard to race would be suspect.") (*quoting Castaneda v. Partida*, 430 U.S. 482, 497, n. 17 (1977)); *Stagi*, 391 F. App'x at 137; *Malave v. Potter*, 320 F.3d 321, 327 (2d Cir. 2003).
[168] *See* Leo Alexander, III. & Fred L. Oswald, *Free Adverse Impact Resource* (2019), https://orgtools.shinyapps.io/FAIR.
[169] *See* U.S. Dep't of Labor, Office of Federal Contract Compliance Programs ("OFCCP"), *Validation of Employee Selection Procedures* (last updated July 23, 2019), at Question 4, https://www.dol.gov/agencies/ofccp/faqs/employee-selection-procedures#Q4.

standard deviation rules. For example, if the null hypothesis is that pass rates are equal among all members of a population, then by statistical convention, a *p*-value less than .05 (roughly two standard deviations) would signify that under this hypothesis, one would observe a difference in pass rates at least as extreme as what was observed less than 5% of the time. This relatively low probability might lead one to the inference that pass rates between groups are indeed unequal. However, note that one of the key features of AI tools is their ability to handle and analyze large amounts of data, which in turn can lead to differences in observed selection or "pass" rates that will almost always be statistically significant, simply because of large sample sizes.

This may inform and perhaps underscore the extent to which *practical significance* has a role to play. *Practical significance* addresses an important follow-up question: If findings are statistically significant and the null hypothesis is rejected (per the above), does the observed discrepancy in hiring rates actually matter and make practical difference? The answer here is not as cut-and-dried as observing whether a *p*-value is or is not less than .05. Practical significance is an expert judgment call about whether the *magnitude* of the effect matters. Even the most reasonable experts can disagree in making this judgment call. Going back to the previous example, if the pass rate on a pre-employment screen is 15% for Black applicants and 20% for White applicants, does this discrepancy in pass rates matter? By the four-fifths rule, on which many experts rely, it does matter, because 15%/20% = 3/4ths or .75, which is less than 4/5ths or .80. But if this is instead framed as a 5% difference, whether this difference matters or not could be a subject of debate. Again, the answer to questions like this are often not cut-and-dried. Moreover, not all courts agree that practical significance is required.[170]

All statistics for determining adverse impact can be evaluated through the lens of the statistical and practical significance considerations outlined above. Note that the four-fifths rule is often used with an exclusive focus on practical significance in a very narrow manner: if the observed ratio of hiring rates is 4/5ths or higher, the employer is in the clear; otherwise, there is a *prima*

---

[170] *Compare Jones v. City of Bos*., 752 F.3d 38, 51–52 (1st Cir. 2014) (Title VII did not require plaintiffs to prove challenged practice was "practically significant" under four-fifths rule as part of their *prima facie* case where they showed statistical significance) *with Waisome*, 948 F.2d at 1373 (disregarding finding of 2.68 standard deviations where pass rate of blacks to whites was 87.2%; court noted that if two additional black candidates had passed the test the statistical disparity would have disappeared). For federal contractors and other employers, a review of the OFCCP FAQs on practical significance is recommended to understand the Agency's position on the issue. *See* OFCCP, *Practical Significance in EEO Analysis Frequently Asked Questions* (last updated January 15, 2021), https://www.dol.gov/agencies/ofccp/faqs/practical-significance.

facie case for adverse impact. However, the ratio alone does not consider whether underlying selection rates are very liberal or restrictive, which may be relevant.[171] Also, the impact ratio alone does not consider how precise (or imprecise) the underlying rates are (though it could do so),[172] and the impact ratio does not consider whether the absolute difference in rates is large or small.

Table A below considers the questions about statistical significance and practical significance together:

Table A

JOINT CONSIDERATIONS FOR ADVERSE IMPACT	
Statistical Significance:	
	● Test the null hypothesis that two subgroup hiring rates are exactly the same
	● Was this null hypothesis rejected (*e.g.*, $p < .05$)?
	● Statistical significance means only that the data are not consistent with the hypothesis that the two subgroup hiring rates are exactly equal. It does not necessarily mean that observed hiring rate differences matter in terms of expert judgments about practical significance (see below).
Statistical Estimation:	
	● Estimate and consider the magnitude of the two subgroup hiring rates, their precision, and the subgroup sample sizes
	● Estimate and consider the ways in which differences in subgroup hiring rates can be usefully portrayed: *e.g.*, the impact ratio (four-fifths rule), raw percent difference, odds ratio
	● Estimate the precision of the above estimates (*e.g.*, 95% confidence intervals)
JOINT CONSIDERATIONS FOR ADVERSE IMPACT (cont.)	
Practical Significance:	
	● Are observed hiring rate differences (a) large enough and (b) precise enough to conclude that they make a practical difference?

[171] *See, e.g., Frazier v. Garrison I.S.D.*, 980 F.2d 1514, 1524 (5th Cir. 1993) (95% of applicants passed test and thus a 4.5% difference in pass rates was considered trivial by the court).
[172] Scott B. Morris & Russell E. Lobsenz, *Significance tests and confidence intervals for the adverse impact ratio*, 53 Personnel Psychology 89-111 (2000), https://doi.org/10.1111/j.1744-6570.2000.tb00195.x.

	● Both of these matter: ○ A large effect might be so imprecise that no solid conclusions about practical significance can be drawn (even if statistically significant) ○ A small effect can be precise (and statistically significant) even if practically negligible in the judgment of an expert

As noted, a large effect might indicate adverse impact at first glance yet be intolerably imprecise because of small sample sizes. For example, you might observe that 20% of Black applicants vs. 50% of White applicants were selected by or "passed" a given selection test. However, say that these observed pass rates were based on extremely small sample sizes of 2 out of 10 Black applicants, and 5 out of 10 White applicants. In this case, the difference is not statistically significant, and no statistical support for practical significance can be made.

Conversely, a small effect, such as 20.1% of Black applicants vs. 20.2% of White applicants being selected, could be precise and statistically significant, if based on very large sample sizes. Whether .1% is practically significant depends critically on expert judgment. For example, some experts may say that .1% amounts to thousands of applicants and that rectifying even this discrepancy is possible. Other experts may find .1% to be a level that would be found in most any sample. In short, the determination of adverse impact is more clear-cut when selection rates are not only statistically significant, but also large, and precise by practical standards that most experts can agree upon. It is less clear cut otherwise, yet courts generally focus on statistical significance, and only some courts consider the magnitude of the disparity for practical significance.[173]

It should be noted that *observation*, *estimation*, and *inference* differ in important ways that can help clarify statistical discussions of adverse impact, whether in the AI context or otherwise. For example, say that you observe an impact ratio of .65 (suggesting adverse impact) in a circumstance where this estimate is based on small subgroup samples and is highly inaccurate (*e.g.*, 95% confidence interval from .40 to .90). Because of this inaccuracy, the hypothesis that the four-fifths rule holds cannot be rejected (despite .65 being observed), where the *p*-value is >

[173] 29 C.F.R. § 1607.4(D); *See also,* EEOC, *Questions and Answers to Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures* ("Q&As"), Nos. 11 and 19, https://www.eeoc.gov/laws/guidance/questions-and-answers-clarify-and-provide-common-interpretation-uniform-guidelines; 44 Fed. Reg. 11996 (Mar. 2, 1979) and 45 Fed. Reg. 29530 (May 2, 1980).

.05. Conversely, you might observe the same impact ratio of .65 based on very large subgroup samples and that is very accurate (*e.g*., 95% confidence interval from .63 to .67). In this latter example, the estimate is highly accurate, and the null hypothesis that the four-fifths rule holds is rejected, where the p-value is $< .05$. All of this assumes that observed samples are representative of the larger applicant population. Any misalignment between the nature of observed samples and the nature of the population of interest is often a substantive issue, not a statistical one.

B. Issues Raised by AI-Enabled Tools and Processes

In contrast to the traditional disparate impact cases, which deal solely with the adverse impact statistics discussed above, cases that involve AI machine learning ("ML") models may have to deal with a wide range of new issues involving both *dynamic data* and *dynamic algorithms*.[174] The extent to which the currently existing legal and analytic frameworks remain applicable is worth consideration.

For example, an AI-enabled tool that uses ML to "improve" the applicant selection algorithm over time may in practice apply a different algorithm (*i.e.*, different scoring and/or selection criteria) on Day 2 than it did on Day 1. Courts and practitioners will need to consider the extent to which statistical analyses may need to be modified in light of these changes – for example, whether it is appropriate to pool and evaluate together (at least at the adverse impact stage of a disparate impact analysis) applicants from Day 1 and Day 2, if they were in fact subject to different algorithms.

These issues regarding ML and adverse impact raise a number of unique questions and challenges, including:

- How is the affected class defined under conditions where different versions of a dynamic ML algorithm have different implications for adverse impact?
- What if dynamic ML algorithms primarily weigh the data in a manner intended to minimize adverse impact, rather than in a manner that would improve the reliability of measuring job-relevant constructs? A random-number generator is an extreme example of the former.

[174] By *dynamic data*, we mean that compared with traditional methods, the flow of applicant data can be both wider (applicant big data, through applying via LinkedIn, Indeed, etc.) and deeper (more variables, from big data based on automated resume screening, interviews with avatars, interactive games, etc.). By *dynamic algorithms*, we mean that ML models might be frequently retrained based on dynamic data flowing through an organization (versus a more static application process that changes less frequently).

- If the algorithm is designed to minimize subgroup differences by race or gender, without coding race or gender directly in the algorithm, does this amount to violating the Civil Rights Acts of 1991, which explicitly prohibits norming by race? Or can this be considered a high-tech version of considering less discriminatory alternatives?
- Would it be fair to have a different ML algorithm apply to rejected job applicants who reapply?

Before applying traditional adverse impact frameworks to these new AI and machine learning tools – and in particular those that utilize dynamic data and dynamic algorithms – it would behoove attorneys and analysts alike to consider if and how those frameworks can accommodate these new challenges.

IV. Comparisons and Benchmarks

> "In order to identify the effects of an employer's practices on favored and disfavored groups, two preliminary inquiries are necessary. First, the particular employment practice or selection device believed to create the disparate impact must be identified. Second, the populations representing the groups to be compared must be ascertained."[175]

The particular employment practice provided as an example in Paetzold and Willborn's important treatise on the use of statistics in employment discrimination was the high school diploma requirement in *Griggs v. Duke Power Co*., but a modern AI-enabled tool might also be evaluated through this lens. A game-based aptitude testing algorithm generating a series of scores that measure candidates' relevant abilities and skills, for example, could also be considered an employment practice subject to an adverse impact analysis. Whether an employer deploys a traditional method or an automated decision tool to identify the best candidates by predicting an individual's future productivity, an adverse impact analysis requires clearly identifying each decision point that could favor one group over another throughout the entire selection process.

Each stage of a structured applicant selection process may represent a single or a series of employment decisions.[176] For a complete evaluation of the process, it is necessary to understand each employment decision thoroughly. This involves clearly defining the individuals competing against each other and identifying the specific qualifications or characteristics considered in the

[175] Ramona L. Paetzold & Steven L. Willborn, *The Statistics of Discrimination: Using Statistical Evidence in Discrimination Cases* 204 (Thompson Reuters, 2013-2014 ed. October 2013) ("Paetzold and Willborn").
[176] Typical stages of a structured applicant selection process can be as follows: (1) Applicant Pool; (2) Applicant Assessment Test; (3) Interview; (4) Background Check; (5) Offer; (6) Hire.

decision to select or not select at each stage. Algorithmic tools can be used in one or more of these stages. For example, an automated natural language processing ("NLP") tool that analyzes text and uses scoring algorithms can be used to identify resumes from applicants who possess relevant skills. The scoring system of the NLP tool must be evaluated to ensure that it not only categorizes applicants into qualified versus not qualified but also identifies the detail in each resume upon which the decision to select or not select was made. For another example, a hiring manager may rely on the scores generated by a gaming algorithm that measures candidates' job-related abilities and skills when deciding which candidates to interview. Hiring managers may also rely on text transcripts generated by machine learning technology from a recorded video interview. An adverse impact analysis tests the extent to which each stage or decision point alone or in combination contributes to observed differences in outcomes between favored and disfavored groups with similar qualifications.

A. Applicant Pools

One of the most contentious issues when conducting a bottom-line, applicant-to-hire adverse impact analysis concerns the attempt to identify the relevant applicant pool that would include all and only those individuals who are ready, able, and willing to work for the employer in a natural selection process. This is true regardless of whether the selection process relies solely on traditional methods or involves ML techniques. Even when the applicant tracking data are available, there is not always a straightforward answer as to whether that data include all of the individuals who are interested in a particular job at a particular location during a particular time period.

When the applicant tracking data are not available, two commonly used alternatives to approximate the relevant applicant pool and represent availability are (1) general population and (2) relevant labor market.[177] Such availability measures usually come from a dataset provided by the Census Bureau (*e.g.*, American Community Survey) and are typically constructed in an effort to represent availability within a geographic area, which usually covers where the employer operates and its surrounding areas and will often need to be adjusted for commuting distance and times as well as availability of transportation means. "A potential issue is the quality of the availability data, particularly if it is derived from a complex survey process [such as Census

[177] Paetzold and Willborn, at 117-127.

Data]. Standard errors for the availability estimates in small geographic areas or uncommon occupations may not support all the comparisons that might be desired."[178] Attempts to expand the geographic areas to reduce standard errors and increase the accuracy of the availability measures usually result in inclusion of individuals who are not interested in the job in question. In *Wards Cove*, "the Supreme Court indicated that a population is too broad as a proxy for qualified job applicants if it contains persons who would not be applicants for the at-issue job and too narrow if it excludes qualified persons in the labor force who would be interested in the at-issue job."[179]

Even when conducting an adverse impact analysis focused on a specific stage in the selection process, identifying the proper comparison groups can be challenging. Although detailed applicant flow data may be available for each step of the entire process, determining which individuals to pool together for purposes of evaluating the process can be complicated. Examples of potential complications include:

- Applicants may withdraw or a hiring manager or an ML algorithm may decide an applicant is not qualified at different stages of the process.
- Not all applicants take the game-based assessment tests at once so some test scores become available to the hiring manager or an ML algorithm earlier than others. As a result, the hiring manager or the ML algorithm might assess applicants in batches, comparing applicants in each batch separately.
- Whether the disposition reasons are clearly and accurately recorded will have an impact on how precise the adverse impact comparisons made at each decision point are.

B. Possible Benchmarks

A side-by-side comparison of the results of an adverse impact analysis and several other benchmarks – while not directly relevant to assessing whether adverse impact exists in a particular employment practice – can help highlight the potential areas of legal or other risk for a company. These benchmarks may come from internal or external sources, and when combined with adverse impact analysis, may provide reassurance or alternatively underscore the

[178] Richard Tonowski, *Thoughts from an EEO Agency Perspective* (Taylor & Francis Group, 1st ed. 2016); Scott B. Morris, Eric M. Dunleavy, *Adverse impact analysis: Understanding data, statistics and risk* 287 (Routledge, 1st Ed. 2016).
[179] Paetzold and Willborn, at 211.

importance of probing possible sources of and solutions to adverse impact. Some potential benchmarks to consider may include:

- The employer's Diversity, Equity, Inclusion, and Accessibility reports
- The employer's own Affirmative Action Programs
- The employer's own EEO-1 reports
- Industry/Peer EEO-1 reports
- Representation rates from the publicly available data by industry and occupation
- Adverse impact analyses from earlier years, other jobs, or other locations
- Adverse impact analyses conducted by similar employers[180]
- Some algorithmic decision tool vendors provide bias audits to their clients. These could be one-time only or periodic analyses conducted by the vendor.

Care should be taken when comparing these benchmarks with the results of an adverse impact analysis, and drawing any conclusions based on them. For example, the job groupings in an EEO-1 report are not as detailed as the job titles or job codes that employers typically use in their daily operations, and thus may not provide relevant apples-to-apples comparisons.

C. Further Issues

There are several other considerations that can lead to modification of comparison groups or use of different benchmarks. Any or all of them may warrant consideration in a particular case:

- Representativeness of the sample
    - Are the data used for training the AI-enabled tool and the data used by the tool in the selection process representative samples of the population subject to the selection process?
    - Do these data capture all the individuals who are available and willing to apply or work?
    - Are these data large enough samples of the population from which the candidates or applicants are drawn?

- Changes in the employment selection process
  Changes affecting the employment selection process almost always require updating the construction of the adverse impact analysis, especially the identification of comparison groups. A couple of examples:
    - A typical scenario in looking for talent or pre-screening talent for positions is the use of AI (Machine Learning) techniques to evaluate and auto-score resumes for positions. These resume screening tools are typically used as a sourcing tool to

[180] As noted previously, New York City will require employers to conduct and publish yearly "bias audits" of automated employment decision tools. NYC Admin Code § 20-870. It is not yet clear how much detail these audit reports are required to provide, but they may have utility as benchmarks at least for employers operating in New York City. (Although the New York City law is effective January 1, 2023, the NYC Department of Worker and Consumer Protection has announced that it will not enforce the new law until April 15, 2023.)

identify candidates that might potentially match the requirements of a job opening. Potential candidates are identified by evaluating the current resume database within an organization's applicant database (applicant tracking system). But consider an employer that changes its policy about the duration for which the applicant resumes are kept in consideration by its algorithmic resume screening tool for a particular "evergreen" posting. As a concrete example, let's say the employer reduces the duration from two months to one month to reduce the proportion of the resumes that are dormant. In the former policy, the applicant pool contains applications submitted within the past two months, whereas in the latter policy, it contains applications submitted only within the last month.

  - Consider another employer that changes the minimum requirements for a job. For example, some technology firms have been relaxing the college degree requirement for coding jobs.

  The size and the demographic composition of the applicant pools in both examples change substantially with these policy or process changes, which in turn, can have a dramatic effect on the outcome of the adverse impact analysis. The same is true in the context of AI-emabled tools, particularly if they are dynamic (see discussion above and below).

- Geography Changes
  - Similar to the effect of policy changes, any changes in the geographical area where the employer operates, such as opening or closing a location, will generally require modifying comparison groups and benchmarks used in the adverse impact analysis. Usually, the geographic area the employer draws its applicants from is not homogenous or symmetric. Adding or removing a location affects this area as individuals consider distance, access to, and cost of public transportation when deciding whether to apply for a particular position. A new location may require including additional individuals with a wide variety of occupations and other characteristics, some of which may make them less likely or unlikely to be interested in a job. This becomes a particularly thorny issue when a job becomes virtual.

- Changes in Race and Gender Information
  - The number of individuals who do not disclose their race or gender varies greatly from employer to employer. Whether these individuals are included in the protected group or reference group can change the outcome of an adverse impact analysis.
  - The ability of applicants to report their gender as non-binary may require an analysis of adverse impact to account for multiple gender categories rather than the traditional male-female binary.
  - How race and ethnicity information are recorded in an applicant tracking system or HR information systems can vary from employer to employer as well as over time for the same employer. The comparison groups and the benchmarks should recognize and reflect such changes.

- Reason for Analyzing Adverse Impact

  - There are various reasons one might conduct an adverse impact analysis of an AI-enabled tool, including legal compliance, litigation, internal audits to assess legal risk, external audits, or public disclosure (*e.g.*, DEI&A assessments in Environmental, Social, and Governance ("ESG") reporting). The structure and the level of detail in an adverse impact analysis depends on the reason for which it is conducted.

- Changes to the algorithm
  - Automated employment decision tools often update their algorithms depending on the volume of the applicants processed through the algorithm. (This could be as often as every year, every quarter, every month, etc.) Every time the algorithm changes, comparison groups for each decision point may change as well. For example, if an algorithm is updated to remove the college degree requirement from some coding jobs, the definition of comparison groups or the external availability benchmarks for the purposes of the adverse impact analysis need to be updated to include individuals without college degrees as well. For another example, if the algorithm increases the weight given to assessment scores and decreases the weight given to prior experience in the selection decision, the threshold for the minimum years of prior experience may need to be modified, which in turn will change the comparison groups.

V. Next Steps After Finding of Adverse Impact

Finding statistically significant adverse impact does not necessarily mean the tool is discriminatory or its use legally impermissible. However, such a finding (assuming it is the result of a proper analysis) does trigger an obligation to understand the drivers behind the analysis and ensure the tool causing the disparity is "job-related for the position in question and consistent with business necessity." Tests are "job-related" when they measure traits or characteristics of the individual being considered for the position against the essential skills required for the position.[181] As with the adverse impact analysis, the "job-related" analysis is specifically focused on the test's (or other selection tool's) relation to the particular position(s) for which particular applicants are applying.

A. Validation as a Way of Showing Job-Relatedness

"[D]iscriminatory tests are impermissible unless shown, by professionally acceptable methods, to be predictive of or significantly correlated with important elements of work behavior which

---

[181] *See, e.g., Gulino v. New York State Educ. Dept*., 460 F.3d 361 (2d Cir. 2006); *Gillespie v. State of Wis*., 771 F.2d 1035, 1040 (7th Cir. 1985).

comprise or are relevant to the job or jobs for which candidates are being evaluated." [182] Validation studies are an often-used way of assessing whether a selection device that causes adverse impact is nonetheless job-related and thus permissible under Title VII. The Uniform Guidelines on Employee Selection Procedures speak directly to this issue. The previous Report Section provides in-depth analysis of UGESP. The aspect of UGESP that the following discussion focuses on[183] is the admonition that "[t]he use of any selection procedure which has an adverse impact … will be considered to be discriminatory … unless the procedure has been validated in accordance with these guidelines, or the provisions of [29 C.F.R. § 1607.6] are satisfied."[184] UGESP describes standards for so-called "validity studies" that will enable an employment test or other selection device to comply with Title VII even though use of the test or device results in an adverse impact.[185] However, "[v]alidation studies 'are by their nature difficult, expensive, time consuming and rarely, if ever, free of error.'"[186] At a minimum, validation studies must actually relate to the particular workforce at issue.[187]

In order to determine whether the application of a particular test is job-related, courts still often turn to UGESP's guidelines on validation studies where possible. As noted in the previous Report Section, UGESP identifies three different types of validation studies that courts apply when analyzing whether employment tests are "job-related": content validity, criterion-related validity, and construct validity.[188] Some types of validation study may be more or less

---

[182] *Gulino*, 460 F.3d at 361 (*quoting Albemarle Paper*, 422 U.S. at 431). As noted above, this is a requirement that applies to the specific user (*i.e.*, the employer).
[183] The discussion that follows focuses on UGESP while leaving to the side other types of evidence of "job-relatedness" that may be appropriate, particularly if the "employment practice" alleged to have caused adverse impact is not a traditional test or other clear-cut screening device like a high school diploma requirement.
[184] 29 C.F.R. § 1607.3A. Whether compliance with UGESP is necessary, sufficient, or both to establish "job relatedness" within the meaning of Title VII in all cases, and for all types of practices that can influence employment selections, is not a settled matter. *See, e.g., Guardians Ass'n of New York City Police Dep't, Inc. v. Civil Service of Comm'n of the City of New York*, 630 F.2d 79, 90, 93 (2d Cir. 1980) (cautioning that "[t]he danger of too rigid an application of technical testing principles [like UGESP] is that tests for all but the most mundane tasks would lack sufficient validity to permit their use," and concluding that "[t]o the extent that the Guidelines reflect expert, but non-judicial opinion, they must be applied by courts with the same combination of deference and wariness that characterizes the proper use of expert opinion in general.").
[185] 29 C.F.R. § 1607.5.
[186] *AMAE*, 231 F.3d at 587 (*quoting Cleghorn v. Herrington*, 813 F.2d 992, 996 (9th Cir. 1987)).
[187] *See* 29 C.F.R. § 1607.7(A) (advising users who obtain selection procedures from publishers to "be careful to determine that, in the event the user becomes subject to the validity requirements of these guidelines [i.e., there is adverse impact], the necessary information to support validity has been determined and will be made available to the user."); *see also Lanning v. SEPTA*, 181 F.3d 478, 491 n. 18 (3d Cir. 1999) (noting that district court's reliance on an expert's study of the work of Anne Arundel County, Maryland, police officers was irrelevant to SEPTA transit officer work absent a showing that the actual function of the two positions is comparable).
[188] 29 C.F.R. §§ 1607.5; 1607.14.

appropriate to use when evaluating a given test. For example, in determining whether to assess content validity as opposed to another type of validity, courts may evaluate the employment test at issue for: "(1) the degree to which the nature of the examination procedure approximates job conditions; (2) whether the test measures abstract or concrete qualities; and (3) the combination of these factors, i.e., whether the test attempts to measure an abstract trait with a test that fails to closely approximate the working situation."[189]

Tests that can be validated by *content validity* most frequently involve characteristics or traits that can actually be measured, for instance typing tests or other physical testing.[190] Under the Guidelines however, content validation is not appropriate to demonstrate "the validity of selection procedures which purport to measure traits or constructs, such as intelligence, aptitude, personality, commonsense, judgment, leadership, and spatial ability."[191]

Tests that can be assessed via *criterion-related validity* frequently include those addressing "traits or constructs, such as intelligence, aptitude," and other factors that are not proper for content validation including, for example IQ tests.[192] Criterion-related validity generally requires empirical data demonstrating that the selection procedure is predictive of or significantly correlated with important elements of job performance.[193] Specifically, UGESP requires criterion-related validity studies to first review the job to determine measures of work behavior that are critical or important job duties to the job in question.[194] When the employer is determining what criteria will be used to measure or predict success in a position, including selecting items for inclusion in a selection device, those criteria must be "important or critical work behavior(s) or work outcomes."[195]

Once a test user has successfully established a selection procedure and the criterion against which the candidates will be measured, UGESP analyzes the comparison between the selection

---

[189] *Gillespie*, 771 F.2d at 1043. *See also Guardians Ass'n of New York City Police Dep't, Inc. v. Civ. Serv. Comm'n of City of New York*, 630 F.2d 79, 93 (2d Cir. 1980) ("[A] validation technique for purposes of determining Title VII compliance can best be selected by a functional approach that focuses on the nature of the job."); 29 C.F.R. § 1607.14 (offering guidelines for appropriate use of each test).
[190] *See* Charles A. Sullivan & Lauren Walter, *Employment Discrimination Law & Practice* 303 (Aspen Publishers, 4th ed. 2009) ("Sullivan & Walter").
[191] 29 C.F.R. § 1607.14(C)(1); *see also Gulino*, 460 F.3d at 384 (describing content validation as appropriate for "observable abilities" but potentially not others).
[192] *See Banos v. City of Chicago*, 398 F.3d 889, 893 (7th Cir. 2005).
[193] *Id*.
[194] 29 C.F.R. § 1607.14(B)(2).
[195] *Id*. § 1607.14(B)(3).

procedures and the measuring criterion to determine whether the procedure is "related to the criterion," or "job-related."[196] UGESP explains:

> [A] selection procedure is considered related to the criterion … when the relationship between performance on the procedure and performance on the criterion measure is statistically significant at the 0.05 level of significance, which means that it is sufficiently high as to have a probability of no more than one (1) in twenty (20) to have occurred by chance. Absence of a statistically significant relationship between a selection procedure and job performance should not necessarily discourage other investigations of the validity of that selection procedure.[197]

Courts applying the criterion-related validity test have explained that a test will be considered relevant when the "correlation coefficient" shows a significant positive correlation between success on the test and success in the job.[198] If the test is scored such that high scores are desirable (e.g., measures of correct responses), positive correlation coefficients indicate that the test perfectly relate to higher job performance, whereas negative correlation coefficients indicate that higher test scores relate to lower job performance.[199] If the test is scored such that low scores are desirable (e.g., measures of reaction time), negative correlation coefficients indicate job success and positive correlation coefficients indicate failure.

Beyond merely showing that there is a correlation coefficient between test performance and job performance, employers are also required to demonstrate that the correlation is statistically significant at conventional levels of significance such as 0.01 and 0.05.[200] Note that correlations for criterion-related validity studies will often underestimate the predictive power of an employment test, if incumbents were selected on the test itself (or any other measure correlated with it). This phenomenon of *range restriction* can be addressed mathematically to estimate the validity that would be obtained in the full job applicant sample.

UGESP notes that the degree of adverse impact and the availability of other tests for use are factors that should be considered when determining whether to implement a test with known disparate impact.[201] However, "if other factors rema[i]n the same, the greater the magnitude of

---

[196] *Id*. § 1607.14(B)(5).
[197] *Id*.
[198] *See, e.g., Williams v. Ford Motor Co*., 187 F.3d 533, 540 (6th Cir. 1999).
[199] *Id*.
[200] 29 C.F.R. § 1607.14(B)(5); *see also Hamer v. City of Atlanta*, 872 F.2d 1521 (11th Cir. 1989).
[201] 29 C.F.R. § 1607.14(B)(6).

the relationship (*e.g.*, correlation coefficient) between performance on a selection procedure and one or more criteria of performance on the job, and the greater the importance and number of aspects of job performance covered by the criteria, the more likely it is that the procedure will be appropriate for use."[202]

UGESP's third validity test is *construct validity*.[203] The purpose of a construct validity analysis is to document accumulated evidence that the assessment actually measures what the employer asserts that it measures. The focus is typically on latent constructs or traits including, among others, scholastic aptitude, mechanical comprehension, neuroticism, and anxiety.[204] At the time UGESP was promulgated in the late 1970s, construct validity was a developing field with little literature applying it in the employment context.[205] Sullivan and Walter note that, in the years since their promulgation, the Guidelines have not been amended to be more receptive to construct validated tests.[206] Further, because UGESP requires that construct validity satisfy the same job analysis and correlation coefficient requirements applied to criterion-related validity tests, the courts seem to have treated the two tests somewhat similarly.[207]

B. Failure To Consider Alternative Employment Practices

Employers who prove that their test was job-related may nonetheless be liable under Title VII if a plaintiff establishes that the employer failed to adopt a less discriminatory alternative employment practice that would equally meet the employer's legitimate needs.[208] Plaintiffs attempting to establish that an employer failed to adopt a less discriminatory alternative employment practice must first identify a specific alternative employment practice that the employer could and (the plaintiffs contend) should have used.[209]

---

[202] *Id*.
[203] *See* 29 C.F.R. § 1607.14(D).
[204] *See* Sullivan and Walter, at 318 (quotation and citation omitted).
[205] 29 C.F.R. § 1607.14(D)(1).
[206] Sullivan & Walter, at 318-19.
[207] *See Gulino*, 460 F.3d at 384 (distinguishing between "content" and "construct" without mention of criterion-related validity tests); *Williams*, 187 F.3d at 540 (similarly describing the construct and criterion-related tests).
[208] *See* 42 U.S.C. § 2000e-2(k)(1)(A)(ii), (C); *Ricci*, 129 S. Ct. at 2673; *see also* 29 C.F.R. § 1607.3(B) ("Where two or more selection procedures are available which serve the user's legitimate interest in efficient and trustworthy workmanship, and which are substantially equally valid for a given purpose, the user should use the procedure which has been demonstrated to have the lesser adverse impact.").
[209] *See Allen v. City of Chicago*, 351 F.3d 309 (7th Cir. 2003) ("A vague or fluctuating proposed alternative ordinarily would frustrate th[e] statutory scheme," though such proposals may suffice if employer refused to consider any alternative).

Second, plaintiffs must demonstrate that the alternative employment practice "'would also serve the employer's legitimate interest.'"[210] To establish that the alternative employment practice would also serve the employer's interests, the plaintiffs must establish that the alternative would lead to a substantially equally qualified workforce as the challenged practice.[211] As part of this analysis, "'[f]actors such as the cost or other burdens of proposed alternative selection devices are relevant in determining whether they would be equally as effective as the challenged practice in serving the employer's legitimate business goals.'"[212]

Third, plaintiffs must prove that the alternative employment practice would produce less adverse impact.[213] Plaintiffs must actually demonstrate – not simply speculate or surmise – that the alternative would have a less discriminatory impact.[214]

Finally, plaintiffs must prove that the alternative they have proven would have less adverse impact was available and that the employer refused to adopt that alternative.[215] Although the statute requires a plaintiff to establish that the employer "refuse[d] to adopt" the alternative practice,[216] it is unclear whether an employer can escape liability by showing that it was never given the opportunity to adopt the less discriminatory alternative.[217]

Once an employer establishes a process for making employment decisions based on a determination that the process is job-related for the position in question and consistent with business necessity, the employer may not ignore or overlook the results of the process merely because the employer has concerns about adverse impact. "Doing so, absent a strong basis in evidence of an impermissible disparate impact, amounts to the sort of racial preference that

---

[210] *Id*. at 312 (*quoting Albemarle*, 422 U.S. at 425 (internal quotation omitted)); *accord Adams v. City of Chicago*, 469 F.3d 609, 613 (7th Cir. 2006).
[211] *Id*. at 314.
[212] *Id*. at 314-15 (*quoting Watson*, 487 U.S. at 998).
[213] *See* 42 U.S.C. § 2000e-2(k)(1)(A)(ii); *Albemarle*, 422 U.S. at 425; *Allen*, 351 F.3d at 315.
[214] *See, e.g., Allen*, 351 F.3d at 315 n. 11 (noting that plaintiffs' burden required offering more than "a prediction" that the alternative would be less discriminatory); *Brown v. City of Chicago*, 8 F. Supp. 2d 1095 (N.D. Ill. 1998) (finding city refused to adopt less discriminatory alternative that it knew to be less discriminatory based on statistical analysis).
[215] *See Bryant*, 200 F.3d at 1094; *see also Adams*, 469 F.3d at 613 (finding that test was "unavailable" because there was no record evidence that the city could have feasibly developed and applied a valid merit selection method for promotions in the month between the time the alternative was proposed and the promotions occurred).
[216] *See* 42 U.S.C. § 2000e-2(k)(1)(A)(ii).
[217] *See* Barbara Lindemann & Paul Grossman, *Employment Discrimination Law* 156 and n. 183 (The Bureau of National Affairs, Inc., 1976).

Congress has disclaimed, § 2000e-2(j), and is antithetical to the notion of a workplace where individuals are guaranteed equal opportunity regardless of race."[218]

As the Second Circuit has explained:

> [A] strong basis in evidence of disparate-impact liability is an objectively reasonable basis to fear such liability. It is evaluated at the time an employer takes a race-conscious action. It relies on real evidence, not just subjective fear or speculation. Because it focuses on liability rather than mere litigation, it requires both objectively strong evidence of a *prima facie* case (or perhaps actual proof of a *prima facie* case) of disparate impact, and objectively strong evidence of non-job-relatedness or a less discriminatory alternative.[219]

The court further noted that "the employer must show a strong basis in evidence that, at the time the race- or gender-conscious action was taken, the employer was faced with disparate-impact liability and that the race- or gender-conscious action was necessary to avoid or remedy that liability."[220] A better practice, however, may be to consider the potential for adverse impact before adopting such a process, and if possible, pilot and evaluate its use for a subset of employment decisions before launching its use on a larger scale.

Until there is a body of caselaw that directly addresses claims of disparate impact discrimination by AI-enabled employment selection processes, the well-developed caselaw assessing adverse impact, including the use of statistics, provides reliable guardrails. In light of the scale by which AI-enabled tools can operate, it is essential that employers develop at least a basic familiarity with the statistical issues and validation techniques (and how courts deal with them) in order to assess such tools.

[218] *Ricci*, 557 U.S. at 585.
[219] *United States v. Brennan,* 650 F.3d 65, 113 (2d Cir. 2011).
[220] *Id.* at 72.

CONCLUSION

The analysis and recommendations provided in this Report are intended to contribute to and inform the on-going public discourse on the role of Artificial Intelligence in employment decision making. As the use of AI in the employment arena continues to develop and change, it is imperative that all stakeholders understand and take into account the particular EEO and DEI&A issues implicated by such use.

The AI TAC submits its recommendations and welcomes the opportunity to collaborate with the many constituencies who are focused on these issues, including enforcement agencies, employee advocates, employer organizations, professional organizations, and academic institutions, to ensure that stakeholders can address the complex and difficult EEO and DEI&A issues that arise with developing AI technology.

ACKNOWLEDGEMENTS

I am deeply grateful to everyone who participated on this Technical Advisory Committee. Professional efforts like these take many players playing different positions at various times. Everyone gave their talent and, perhaps even more important from busy professionals, their time in this significant volunteer endeavor.

I have to offer special thanks to a number of people who participated in this process. First, Richard Tonowski, Ph.D., former chief I/O psychologist at the EEOC. Without his initial scoping of the issues and chairing the initial Survey Committee this project would never have gotten off the ground. Dave Schmidt, Ph.D., at DCI Consulting, brought the survey to life and Gurkan Ay, Ph.D. at Resolution Economics, gave us the survey results. I thank Eric Dunleavy, Ph.D., at DCI Consulting and Nancy Tippins, Ph.D., for many heavy lifts from the beginning and for taking all of my calls (and those of others) and answering my emails and giving steady advice on many perplexing issues.

Many thanks to the team at Fortney Scott, especially Valerie Higgs, Alyson Cina and Deirdre Bell, for organizing and providing our meeting space for our in-person meeting in Washington, DC and doing so with their usual gracious hospitality.

Enormous thanks to Margo Pave, Esq., Director at Resolution Economics, without whose precision analysis, incredible writing and extraordinary editing there would not be a final report. High fives to Adriana Joens, soon-to-be-Esquire, of Fortney Scott, who came into this project at a crucial time and researched, cite-checked, organized, proof-read, and tracked down absolutely everything we threw at her. She never shied away from giving us her comments and criticism, a skill that will serve her well in her chosen career. And finally, incredible thanks to Jonathon Kaehn, at Resolution Economics, who served as the project manager and utility player from the beginning and without whose organization, technical skills, patience, and steady, cooperative spirit this project would never have gotten out the door.

Thank you to all.

Victoria A. Lipnic
AI-TAC Chair

The Institute for
WORKPLACE EQUALITY

Appendix A Accompanying Technical Advisory Committee Report on EEO and DEI&A Considerations in the Use of Artificial Intelligence in Employment Decision Making

December 2022

1920 I Street NW • Washington DC 20006 • 202.293.2220
Theinstitute4workplaceequality.org

The Institute for WORKPLACE EQUALITY

CO-FOUNDERS OF THE INSTITUTE

David Cohen DCI Consulting Group, Inc.	David Fortney Fortney & Scott, LLC

The Technical Advisory Committee was a volunteer effort by the following individuals:

CHAIR OF THE AI TAC:

Victoria A. Lipnic
Resolution Economics, LLC

LEAD EDITOR:	PROJECT MANAGER:
Margo Pave Resolution Economics, LLC	Jonathon Kaehn Resolution Economics, LLC

MEMBERS OF THE AI TAC:

NAME	ORGANIZATION	TAC SUBCOMMITTEE
Alexander Alonso	Society for Human Resource Management	UGESP
Gurkan Ay	Resolution Economics, LLC	Statistics and Adverse Impact
Stephanie Beers	Microsoft Corporation	UGESP
Tara Behrend	Purdue University	Uses and Applications
Lynn Clements	Berkshire Associates	Transparency and Fairness
Eric Dreiband	Jones Day	Statistics and Adverse Impact
Michelle Duncan	Jackson Lewis P.C.	Transparency and Fairness
Eric Dunleavy	DCI Consulting Group, Inc.	UGESP
Chai Feldblum	EEO and DEI Consultant – Self Employed	Transparency and Fairness
Mark Girouard	Nilan Johnson Lewis PA	UGESP

Chris Gokturk	Littler Mendelson P.C.	Statistics and Adverse Impact
Valerie Hoffman	Seyfarth Shaw LLP	Uses and Applications
Richard Landers	University of Minnesota	Transparency and Fairness
Craig Leen	K&L Gates LLP	Uses and Applications
Kathryn Mantoan	Orrick, Herrington & Sutcliffe LLP	Statistics and Adverse Impact
S. Morton McPhail	Valtera Corporation	Transparency and Fairness
Kris Meade	Crowell & Moring LLP	Uses and Applications
Frederick Melkey	Emerson Electric Co.	UGESP
Nathan Mondragon	HireVue	Data Collection
Fred Oswald	Rice University	Statistics and Adverse Impact
Sachin Pandya	University of Connecticut	Data Collection
Margo Pave	Resolution Economics, LLC	Uses and Applications
Frida Polli	Harver and pymetrics	UGESP
Eric Reicin	BBB National Programs	Data Collection
Matthew Scherer	Center for Democracy & Technology	Uses and Applications
David Schmidt	DCI Consulting Group, Inc.	Transparency and Fairness
Savanna Shuntich	Fortney & Scott, LLC	UGESP
Robert Stewart	Amazon	Data Collection
Eric Sydell	Modern Hire	Uses and Applications
Nancy Tippins	The Nancy T. Tippins Group, LLC	Data Collection
Richard Tonowski	Formerly EEOC	UGESP
Kelly Trindel	Workday	UGESP
Rae Vann	WayFair	Data Collection
Christine Webber	Cohen Milstein Sellers & Toll PLLC	Statistics and Adverse Impact
Colin Willis	HireVue	Data Collection
Ye Zhang	Resolution Economics, LLC	Data Collection

TAC MEMBER BIOGRAPHIES

Victoria A. Lipnic is a Partner at Resolution Economics. She leads the Company's Human Capital Strategy Group. The Human Capital Strategy Group combines the Company's expertise in data analytics and deep knowledge of regulatory requirements with an interdisciplinary approach to advise organizations on the full range of their human capital needs and reporting requirements including recruitment, selection, promotions, DE&I, pay equity, and ESG, especially as to equal employment opportunity and anti-discrimination issues.

Ms. Lipnic joined Resolution Economics in 2021. She has broad experience in the full range of human capital, labor and employment issues, especially from the regulatory enforcement perspective. Prior to joining the Company she served as Commissioner of the U.S. Equal Employment Opportunity Commission ("EEOC") from 2010 to 2020 and Acting Chair from 2017 to 2019. She was appointed to the EEOC by President Barack Obama and confirmed by unanimous consent by the U.S. Senate. At the EEOC she worked on policy, cases, and regulations falling under all of the statutes enforced by the Commission including Title VII of the Civil Rights Act of 1964, the Age Discrimination in Employment Act (ADEA), the Americans with Disabilities Act (ADA), the Equal Pay Act (EPA), the Pregnancy Discrimination Act (PDA) and the Genetic Information Nondiscrimination Act (GINA). While at the EEOC she participated in numerous agency regulatory initiatives including the final GINA regulations, the ADA, as amended, regulations, and the revisions to the EEO-1 form to include pay data reporting. She organized the agency's first public meeting on Big Data in Employment, created its Chief Data Officer position, oversaw development of the Office of Enterprise Data and Analytics and published a significant report on age discrimination. She co-chaired the EEOC's Select Task Force on the Study of Harassment in the Workplace, and co-authored its seminal report, issued in 2016, before the #MeToo movement. Prior to the EEOC, she practiced law with Seyfarth Shaw. She also served as Assistant Secretary of Labor for Employment Standards from 2002-2008, appointed by President George W. Bush, where, among other regulatory enforcement agencies, she oversaw the Office of Federal Contract Compliance Programs.

Alexander Alonso, Ph.D., SHRM-SCP is the Society for Human Resource Management's (SHRM's) Chief Knowledge Officer leading operations for SHRM's Certified Professional and

Senior Certified Professional certifications, research functions, and the SHRM Knowledge Advisor service. He is responsible for all research activities, including the development of the SHRM Competency Model and SHRM credentials.

Throughout his career, Dr. Alonso has published works in peer-reviewed journals such as Industrial and Organizational Psychology: Perspectives on Science and Practice, Journal of Applied Psychology, International Journal of Selection and Assessment, People and Strategy, Personality and Individual Differences, Quality and Safety in Health Care, and Human Resources Management Review. He has also authored several chapters on community-based change initiatives in workforce readiness, as well as co-authoring Defining HR Success: A Guide to the SHRM Competency Model in Practice.

Dr. Alonso received his doctorate in Industrial-Organizational Psychology from Florida International University in 2003. His works have been recognized for their contribution to real-world issues. They include being recognized by the Society for Industrial Organizational Psychology (Division 14 of the APA; SIOP) with the 2007 M. Scott Myers Award for Applied Research in the Workplace for the development of the federal standard for medical team training, TeamSTEPPS; being awarded a 2009 Presidential Citation for Innovative Practice by the American Psychological Association for supporting the development of competency model for team triage in emergency medicine; and receiving the 2013 SIOP Distinguished Early Career Contributions for Practice Award.

Gurkan Ay, Ph.D., is a Director at Resolution Economics, LLC. He conducts economic and statistical analysis of labor and employment issues (including analysis of issues related to compensation as well as hiring, promotion, and termination practices), wage and hour claims, and environmental justice related matters. He has significant experience in proactive monitoring of compensation with respect to pay equity. He regularly assists clients with DEI&A reporting needs for public disclosure and ESG reporting purposes, and with the proactive analysis and compliance needs related to algorithmic bias.

Gurkan has a Ph.D. in Labor Economics from the George Washington University and an M.S. in Finance from University of Wyoming, as well as a B.S. in Industrial Engineering from Istanbul Technical University. Prior to joining Resolution Economics, he was a consultant at ERS Group.

Gurkan also served as an adjunct member of the McCourt School of Public Policy faculty at Georgetown University.

His prior experience includes analyses of market timing and late trading in mutual funds, computation of costs and benefits of IRA products, state taxation of businesses, antitrust cases in telecommunications sector, and the FCC's spectrum auctions.

Stephanie Beers has practiced employment law for over 15 years and is currently an Assistant General Counsel at Microsoft leading the Workforce Regulations and Policy team within Microsoft's Corporate, External, and Legal Affairs department. In her role, Stephanie manages OFCCP compliance and provides strategic advice to business and HR professionals regarding federal contractor compliance, diversity and talent acquisition policies and programs, and legal and compliance training. Stephanie also advises on workforce public policy, including diversity and inclusion, economic security, and the impact of automation and AI on skills and employability. She holds B.A. degrees in Psychology and Planning, Public Policy & Management from the University of Oregon and a J.D. from the University of Washington School of Law.

Tara Behrend, Ph.D., joined the Purdue faculty as an Associate Professor of Psychological Sciences in Fall 2020. She was previously Assistant Professor of I-O Psychology (2009-2015) and Associate Professor of I-O Psychology and Management (by courtesy) (2015-2020) at the George Washington University. She is an internationally recognized expert in workplace technology use, having published extensively on topics related to big data, surveillance, privacy, and learning. Her work in the area of STEM careers and education has been widely influential as well.

Lynn A. Clements is Director of Audit and HR Services for Berkshire Associates, Inc. She oversees Berkshire's audit defense, pay equity and advanced people analytics practice. Throughout her career, Lynn has provided clients with strategic insight and practical compliance and risk management guidance, including in hundreds of government investigations before OFCCP and EEOC.

With 25 years of experience in the EEO and affirmative action space, including nearly eight years working as a former senior official at the EEOC and OFFCP, Lynn brings a unique blend of regulatory knowledge and compliance experience to every project she oversees. Lynn is a regular presenter of management and employee training at all levels of organizations and has testified before Congress and federal enforcement agencies on EEO and pay equity issues. A regular panelist at national and local industry conferences, Lynn is a member of the NILG Advisory Board, a faculty member at The Institute for Workplace Equality and co-authored the U.S. Chamber of Commerce's Fall 2017 Report on Recommendations for Reform of the OFCCP.

Eric S. Dreiband is a Partner at Jones Day and represents clients in investigations, litigation, and counseling in civil rights, employment discrimination, whistleblower, wage and hour, and other matters. Prior to rejoining Jones Day in 2021, Eric served as the 18th Assistant Attorney General for the Civil Rights Division at the U.S. Department of Justice (DOJ), and he also served as the 12th General Counsel of the U.S. Equal Employment Opportunity Commission (EEOC).

Under Eric's leadership, DOJ's Civil Rights Division set enforcement records for prosecutions of law enforcement officers and sexual harassment, religious liberty, and servicemember cases; charged the highest number of hate crimes cases in decades; significantly expanded resources for human trafficking prosecutions; prosecuted race and other forms of illegal discrimination in education, employment, housing, lending, and voting. In recent years, he has reached historic disability rights settlements with several states; opposed unlawful COVID-19-related civil liberty restrictions; and successfully litigated to protect the Constitutional and civil rights of all people in the United States.

As EEOC general counsel, Eric led the Commission's litigation of the federal employment antidiscrimination laws, and he issued the Regional Attorneys' Manual, which established the policies of EEOC's litigation program. Eric also served at the Department of Labor (DOL) as deputy wage and hour administrator and directed DOL's enforcement of the Fair Labor Standards Act (FLSA), the Family and Medical Leave Act (FMLA), and other laws.

Michelle L. Duncan is a principal in the Denver, Colorado, office of Jackson Lewis P.C. Her practice is focused on representing employers in affirmative action and employment discrimination matters before OFCCP.

Michelle joined the firm after working for nearly fourteen years as a trial attorney with the U.S. Department of Labor, Office of the Solicitor. She served in the National Office in Washington, D.C., as well as in the Dallas and Denver Regional Offices. As a senior trial attorney, she litigated some of the Labor Department's most complex cases.

During her tenure with the U.S. Department of Labor, Michelle was widely regarded as a leading authority on OFCCP litigation. She litigated numerous OFCCP cases and provided ongoing counsel to high-level OFCCP officials. In addition, Michelle provided legal and enforcement training on a national level to both OFCCP enforcement personnel and other attorneys in the Office of the Solicitor. This unique experience enables her to provide both strategic and practical advice to Jackson Lewis clients with government contracts.

Eric M. Dunleavy, Ph.D., is an Industrial/Organizational Psychologist and Director of the Employment & Litigation Services (ELS) Division at DCI Consulting, where he leads a division of Psychologists and Labor Economists involved in a wide variety of personnel selection and litigation support services. Eric has particular expertise in the areas of EEO analytics, job analysis, psychological measurement, and selection procedure development, evaluation and validation.

Eric received his M.A. (2002) and Ph.D. (2004) in I/O Psychology from the University of Houston. Since then he has conducted high stakes applied research on employment outcomes such as hiring, promotion, and pay in a wide variety of contexts and for a wide variety of clients. He has also published articles in various journals and recently co-edited (with Scott B. Morris) the book "Adverse Impact Analysis: Understanding Data, Statistics and Risk" (Taylor & Francis, 2017). He has been adjunct faculty and taught graduate level courses at both George Mason University (GMU) and the University of Maryland at Baltimore Country (UMBC). He is also a faculty member of The Institute for Workplace Equality. Before joining DCI, Dr. Dunleavy worked for the American Institutes for Research (AIR), where he was involved in large-scale social science and employee selection research for medical, educational, and federal agency clients.

Chai R. Feldblum is a long-time civil rights advocate and scholar. Chai played a leading role in drafting and negotiating the Americans with Disabilities Act of 1990 and later the ADA Amendments Act of 2008. As a law professor at Georgetown Law in Washington, DC for 18 years, Chai created a Federal Legislation Clinic where she and her students helped non-profit organizations advance their legislative social justice goals.

From 2010 to 2019, Chai served as a Commissioner of the Employment Opportunity Commission (EEOC) where she helped advance employment civil rights, including for LGBTQ people (establishing their protection under sex discrimination laws), people with disabilities, and women. She also led a proactive effort to prevent harassment in the workplace. For two years, Chai was a partner at the law firm of Morgan Lewis, where she helped employers work proactively to create safe, respectful, diverse and inclusive workplaces.

In 2021, Chai became a free-lance civil rights advocate and scholar. She assists with legislative and regulatory work regarding civil rights, particularly for LGBTQ+ people, people with disabilities and women. She also serves as Vice Chair of the AbilityOne Commission, a federal agency devoted to the employment of people with significant disabilities.

Mark J. Girouard is a Shareholder and Chair of the Labor an Employment practice at Nilan Johnson Lewis PA. Mark represents employers in single-plaintiff cases, private class actions, and litigation against the EEOC and other government agencies, including defending numerous nationwide discrimination and wage-and-hour class and collective actions.

Mark also advises employers regarding a wide range of state and federal employment law issues, helping them proactively manage the risks associated with pay equity, pre-employment assessment, background check and drug screening, military leaves, family and medical leaves, disability accommodations, and FLSA and state wage-and-hour compliance, among others. In addition to advising and defending businesses ranging from startups to Fortune 50 corporations, he has nearly two decades of experience representing municipal entities, including serving as labor and employment counsel to several Minnesota cities.

Chris Gokturk is currently a non-attorney Principal with Littler Mendelson P.C. where she assists companies in understanding and mitigating their affirmative action and systemic discrimination risks, particularly in the areas of talent acquisition, compensation, performance management, as well as workforce planning and restructuring.

Chris has more than 20 years of professional experience in compliance, enterprise risk management, and statistics. She has helped clients across all major industry groups develop, implement, and defend compliant affirmative action programs.

Chris provides a broad array of services, including, preparing fully compliant, data-driven affirmative action programs, managing all phases of the Labor Department's Office of Federal Contract Compliance Programs (OFCCP) compliance evaluation process, conducting rigorous self-critical analyses of compensation data, designing and executing statistical analyses of employment transactions data, developing legally defensible diversity and inclusion metrics, assessing organizational compliance with talent acquisition requirements, assisting in implementation of applicant tracking and HRIS systems, and conducting risk assessments in all areas related to human resources and compliance

She is a frequent speaker and trainer on pay equity and the requirements enforced by the OFCCP. In addition to her practice, Chris is also an active participant in Littler's Diversity Group.

Valerie J. Hoffman is a Partner at Seyfarth Shaw LLP. As counsel to chief executives and management, Valerie advises employers of all varieties and sizes, including Fortune 100 technology, financial services, professional services, hospitality, and manufacturing companies. While Valerie has a broad-based employment law advice and counsel practice, she has a particular focus on strategies to prevent and resolve issues relating to employment discrimination, including pay equity and glass ceiling issues, OFCCP defense and affirmative action compliance, and advice about legally defensible diversity practices. She is a core leader of the firm's Pay Equity Group and founded and is co-chair of the firm's nationally prominent OFCCP, Affirmative Action, and Diversity practices.

In addition, Valerie co-founded and co-chairs Seyfarth's People Analytics group. This forward-facing people analytics practice works with employers to design and implement metrics for tracking, assessment and achievement of workforce objectives including effective talent acquisition and management, workforce planning, leadership development, pay equity, diversity & inclusion, reductions in force, and other objectives where data can be leveraged to improve human capital management and work life.

Valerie also provides counsel and advice on diversity and inclusion best practices, metrics, progress management and change management. Employers across the country value her broad knowledge of legally defensible and practical solutions for achieving diversity and inclusion objectives. She has extensive experience counseling senior leaders about these issues and their intersection with employment discrimination law.

Richard N. Landers, Ph.D., is the John P. Campbell Distinguished Professor of Industrial-Organizational Psychology at the University of Minnesota and Principal Investigator of TNTLAB (Testing New Technologies in Learning, Assessment and Behavior). His research concerns the use of innovative technologies like games, gamification, machine learning, artificial intelligence, and virtual reality, to improve psychometric assessment, employee selection, adult learning, and research methods. He is a Fellow of the Society for Industrial and Organizational Psychology, American Psychological Association, and Association for Psychological Science. His work appears primarily in psychology and interdisciplinary human-computer interaction journals. He currently serves as associate or consulting editor or on the editorial board of four academic journals. He is author of a textbook and has developed two edited scholarly volumes. He is featured frequently in the popular press, such as Forbes, Business Insider, and Popular Science and regularly consults with industry as president of Landers Workforce Science LLC, primarily by auditing employee hiring systems incorporating artificial intelligence and machine learning.

Craig E. Leen is a Partner at K&L Gates in the Labor, Employment & Workplace Safety practice group, where he co-leads the firm's OFCCP and Affirmative Action Compliance focus area. He is the former Director of the Office of Federal Contract Compliance Programs at the U.S. Department of Labor, with the agency publishing extensive guidance on equal employment opportunity, pay equity, and employee selection during his tenure, including specific guidance in the area of Artificial Intelligence.

Craig is also a Professorial Lecturer in Law at The George Washington Law School, a member of the faculty for The Institute for Workplace Equality, and a member of the advisory board of the NYU Center for Labor & Employment Law. Craig also serves as Vice Chair of the District of Columbia Advisory Board to the U.S. Commission on Civil Rights, as Chair of the Civil & Human Rights Committee of the Bar Association of the District of Columbia, and as a member of the Rules of Professional Conduct Review Committee of the DC Bar. Craig serves on the boards of Circa, RespectAbility, and Friends of Kenilworth Aquatic Gardens, and the advisory boards of Eightfold.ai (where he serves on the AI Ethics Council) and Disability:IN DC Metro.

Kathryn G. Mantoan is Of Counsel in Orrick's global Employment Law & Litigation group, working across the San Francisco and Portland offices. She focuses on high-stakes employment litigation, compliance counseling, and litigation avoidance measures and co-chairs Orrick's Pay Equity Task Force.

Katie regularly writes on issues in EEO law, including co-authoring "Mind the Gap: Pay Audits, Pay Transparency, and the Public Disclosure of Pay Data," 33 ABA J. Lab. & Emp. L. 1 (2018) and "Job-Relatedness & Validity Evidence Under Title VII," presented at the 2021 ABA Annual Labor and Employment Law Conference.

Katie received California Lawyer's Attorney of the Year Award in 2016, and was part of trial teams recognized with American Lawyer "Litigators of the Week" accolades in September 2020 and June 2022. She received her A.B. in Philosophy from Princeton University, magna cum laude, and her J.D. from Harvard Law School, cum laude; and she is a Ph.D. candidate in Philosophy at the University of California, Berkeley.

S. Morton McPhail, Ph.D., Dr. McPhail practiced Industrial/Organizational Psychology for more than 35 years. He received his BA in Psychology from Trinity University in San Antonio and masters and doctoral degrees in I-O from Colorado State University. In 1981, he co-founded the consulting firm of Jeanneret & Associates and became a Principal. After merger, he became a Senior Vice President with Valtera and subsequently with CEB until his retirement. His consulting work included serving as an expert in litigation involving such diverse issues as job analysis, selection practices, equal employment opportunity, compensation, and statistical analyses, development, and validation of an array of scientifically based assessment tools, creation of performance management systems, designing training and development programs, and executive assessments for selection and development.

A Fellow of SIOP, Dr. McPhail served as its Secretary-Financial Officer from 2009-2011, and as President (2016-2017). He has published and presented on numerous topics, including editing a volume in SIOP's Professional Practice Series on validation strategies, and presented frequently at professional meetings. He is adjunct faculty for the University of Houston and Rice University and has served on the editorial boards of SIOP's Professional Practice series and Industrial and Organizational Psychology. Until his retirement, he was a licensed psychologist and served for 15 years on a Texas Psychology Board committee responsible for the State's Jurisprudence and Ethics Examination for licensure. He previously served APA on its Continuing Education Committee and on the Advisory Panel for National Standards for High School Curriculum in Psychology.

Kris D. Meade is co-chair of Crowell & Moring's Labor & Employment Group. He is also a member of the firm's Management Board and Executive Committee. He counsels and represents employers in the full range of employment and traditional labor law matters, including individual and class action lawsuits filed under Title VII of the Civil Rights Act of 1964, the Age Discrimination in Employment Act, ERISA, and companion state statutes. Kris represents employers in connection with union organizing campaigns, collective bargaining, labor arbitrations, and unfair labor practice litigation. In 2020, Chambers USA recognized Kris as a leading labor and employment lawyer.

Kris also counsels and represents employers in connection with pay equity and affirmative action compliance matters, including the Office of Federal Contract Compliance Programs' corporate management (or "glass ceiling") reviews and affirmative action compliance audits. Kris has particular experience directing sophisticated statistical analyses in connection with employment class action litigation, compensation audits conducted by the Office of Federal Contract Compliance Programs, and pay equity self-audits conducted by employers to ensure compliance with federal regulations, Title VII, and the Equal Pay Act. He frequently teams with labor economists who serve as consulting or testifying experts in private litigation or government enforcement actions. In addition, Kris has substantial experience in managing and coordinating large-scale discovery in nationwide, serial litigation.

Kris likewise represents and counsels' employers in business tort cases, Sarbanes-Oxley whistleblower retaliation matters, matters involving the protection of trade secrets, and litigation over non-competition agreements and other restrictive covenants. He also counsels' employers with regard to their use of predictive analytics and AI in the employment and personnel decision-making context.

Frederick J. Melkey has over 16 years of experience related to Affirmative Action compliance. In his current role as the Corporate Director of EEO and Affirmative Action for Emerson Electric Co., he has responsibility for over 100 AAPs. Fred began his career as an Engineer. For his "mid-life crisis" he attended law school in the evenings. He has held EEO/AA related roles in a wide variety of industries including working for Intel, The Hartford Insurance, The Massachusetts Institute of Technology (MIT), and Emerson. Fred earned a BS in Industrial Engineering from Purdue University, an MBA from the University of New Mexico, and a JD with Honors from the University of Connecticut School of Law.

Nathan Mondragon, Ph.D., is Chief IO Psychologist at HireVue and responsible for building, researching, and maintaining the AI-driven assessment product. In this role the IO Psychology team at HireVue work side-by-side with the Data Science team to create and deliver AI-based Talent solutions. Nathan has over 30 years of extensive experience in the Talent Management space and is a recognized expert in the blending of IO Psychology tools with technology to deliver seamless integrated recruitment, hiring, and talent development solutions. In 1996, Nathan helped lead the creation and delivery of the first ever online selection assessment, in 2004 built from the ground up the first integrated assessment solution within an enterprise-wide ATS (Taleo), and in 2015 as part of HireVue began work to deliver an industry leading AI-driven pre-hire assessment solution. He has over 50 manuscripts, presentations, and workshops on IO Psychology and technology solutions, has been quoted or featured in over 30 popular press articles, and has been interviewed on NPR, BBC, WSJ, Bloomberg, USA Today, and HBO about the future of Recruiting and Assessment. He has held leadership positions with Aon, SHL, DDI, Dell, Taleo, and Oracle.

Nathan received his Ph.D. in Industrial and Organizational Psychology from Colorado State University.

Fred Oswald, Ph.D., is Professor and Herbert S. Autrey Chair in Social Sciences at Rice University. As an industrial-organizational psychologist, his research, grants, and 150+ publications (peer-reviewed publications, edited books, chapters, and technical reports) focus on developing, implementing and evaluating the wide range of tests that everyone encounters in their life within employment and educational settings (e.g., personality, knowledge, performance tests). His recent work addresses the scientific, ethical, and legal implications of AI-based employment tests. Dr. Oswald's recent leadership positions include being the current Chair of the Board on Human-Systems Integration (BOHSI) at the National Academies; current member of the National Artificial Intelligence Advisory Committee (NAIAC); current Chair of the Board of Scientific Affairs (BSA) within the American Psychological Association; and past president of the Society for Industrial and Organizational Psychology (SIOP, 2017-18). He is a Fellow of the American Psychological Association (Div. 5, 8, 14), the Association for Psychological Science, and the Society for Industrial and Organizational Psychology. Dr. Oswald received his Ph.D. in industrial-organizational psychology from the University of Minnesota in 1999.

Sachin S. Pandya Sachin S. Pandya is a Professor of Law at the University of Connecticut. He researches the law of work and anti-discrimination in the United States. For publications and more, see https://orcid.org/0000-0001-7387-1307. Before law teaching, he worked as an appellate and civil rights lawyer for the New York State Attorney General and as a judicial clerk for Judge Jon O. Newman, United States Court of Appeals, Second Circuit.

Margo Pave, Esq. is Director, Human Capital Strategy Group, at Resolution Economics, LLC. She has extensive expertise in employment law and discrimination issues. At Resolution Economics, she oversees projects that combine the Company's skill in data analytics and deep knowledge of regulatory requirements with an interdisciplinary approach to advise organizations on the full range of their human capital needs and reporting requirements including recruitment, selection, promotions, DE&I, pay equity, and overall talent allocation.

Previously, she served as Assistant General Counsel for Appellate Services at the U.S. Equal Employment Opportunity Commission. In that role, she managed a team of attorneys litigating cases under all the statutes enforced by the EEOC, including Title VII of the Civil Rights Act of 1964, the Americans with Disabilities Act, the Equal Pay Act, and the Age Discrimination in Employment Act.

Prior to her time at the EEOC, she was a partner in private law practice, where she litigated a broad range of labor and employment matters on behalf of both employees and employers. Ms. Pave is a former co-chair of the Equal Employment Opportunity Committee of the American Bar Association’s Labor and Employment Law Section.

Frida Polli, Ph.D., is a former Harvard and MIT neuroscientist. She is the Chief Data Science Officer at Harver, the industry leading hiring solution helping organizations optimize their talent decisions. Previously, she was the CEO and founder of pymetrics, the leader in unbiased soft skills assessments, pymetrics. Under Frida’s leadership, pymetrics was a World Economic Forum’s Technology Pioneer, Global Innovator, and Global Future Council member, an Inc 5000’s Fastest Growing company, and Forbes AI 50 company.

Additionally, Frida served as a cognitive neuroscientist at Harvard and MIT. Her work focused on multimodal imaging of health and disease. Her work won the MIT 100K, a BBRF Young Investigator Award, numerous NIH grants and an HBS Life Science Award. She belongs to Institute for Workplace Equality Technical Advisory Committee, the World Economic Forum's Future Skills Alliance, the IEEE’s Organizational Governance of AI, WEF’s Human-Centered AI in HR, and the MIT Work of the Future Task Force. Her research has been presented and published in the Proceedings of the National Academy of Sciences, the Journal of Neuroscience, Brain, the President’s Circle of the National Academy of Sciences, ACM FaccT, and academic forums at Harvard Kennedy and Business Schools, Stanford University, Columbia University, Insead, and others.

Eric D. Reicin joined BBB National Programs as President and Chief Executive Officer on November 4, 2019. Most recently, Eric served as Vice President, General Counsel, and Corporate Secretary for MorganFranklin Consulting, LLC and MorganFranklin, LLC, a global management consulting firm and government contractor (DOD and civilian). He also led MorganFranklin's Corporate Investigations & Dispute Solutions consulting practice. Eric's strategy and execution work in MorganFranklin's transformation from founder owned to an LLC with minority management ownership to an ESOP led to his recognition as the Association of Corporate Counsel – NCR Outstanding Chief Legal Officer in 2016. Vaco (backed by Olympus Partners) acquired MorganFranklin on July 31, 2019.

Eric previously served as Senior Vice President and Deputy General Counsel at Sallie Mae, then a Fortune 500 diversified financial services company (NASDAQ: SLM). He spent 14 years working closely with the Board of Directors and senior executive management during times of transformational change, senior executive turnover, negative press, government investigations and class action litigation, financial uncertainly, and regulatory scrutiny. He also successfully led a large team of attorneys, compliance personnel, and professionals based in six cities.

Eric served a six-year term on the global board of the Association of Corporate Counsel, which has a presence in 85 countries. Eric previously served as president of the Association of Corporate Counsel - National Capital Region, the largest regional in-house bar association. Eric served as the 2016-2019 co-chair of the D.C. General Counsels Club. He currently serves on the advisory board of the Georgetown University Law Center CCI, the Board of Directors of the American Employment Law Council (AELC) and is a Fellow of The College of Labor and Employment Lawyers. In 2019, Legal 500 named Eric to the General Counsel Powerlist – United States.

Matthew U. Scherer is Senior Policy Counsel for Workers' Rights and Technology Policy at the Center for Democracy and Technology (CDT), where he studies how emerging technologies affect workers in the workplace and labor market. He works to advocate for both governments and private organizations to adopt policies that protect workers' digital rights and ensure that new technologies enhance social justice and equality.

Matt came to CDT from Littler Mendelson, a global labor and employment law firm, where he advised employers and tech companies on algorithmic bias, HR tools' compliance with antidiscrimination laws, and related privacy and ethical issues. He also worked as an analytics project manager, conducting and overseeing complex data science projects. Before joining Littler, Matt practiced traditional employment law at Buchanan Angeli Altschul and Sullivan. Before entering private practice, he completed judicial clerkships with Judge Gregory M. Sleet at the U.S. District Court for the District of Delaware, Judge Deborah Cook at the U.S. Court of Appeals for the Sixth Circuit, and Justice Charles Wiggins at the Washington Supreme Court.

Matt is also a noted writer and commentator on the legal and policy issues surrounding Artificial Intelligence. His articles include Regulating Artificial Intelligence Systems and Applying Old Rules to New Tools: Employment Discrimination Law in the Age of Algorithms.

Matt received his J.D. from Georgetown University Law Center, where he served as editor-in-chief of The Georgetown Journal of Legal Ethics, and holds an M.S. in Educational Policy from the University of Pennsylvania's Graduate School of Education.

David Schmidt, Ph.D., is an Industrial and Organizational Psychologist who joined DCI Consulting's Employment & Litigation Services Division in 2019. Dave consults with organizations on topics such as selection, validation strategies, legal defensibility, and the appropriate use of technology and data. Dave has over 25 years of experience and expertise in test and assessment development, data quality and integrity, and research methods. This is coupled with a strong technical foundation in analytics, statistics, and psychometrics. Dave has extensive experience conducting independent expert reviews and audits of a wide range of selection tools (ranging from traditional methods to those enabled by Artificial Intelligence/machine learning) to evaluate adherence to legal and professional standards, as well as relevant local laws.

Prior to DCI, Dave spent about 11 years at Development Dimensions International as their resident expert on legal issues, managing activities supporting clients defending against legal challenges. Dave helped shape DDI's standard practices related to legal defensibility (e.g., job analysis, content validation, criterion validation, validity transportability, adverse impact analyses, disability accommodations). Dave helped build DDI's analytics capabilities, emphasizing repeatable/scalable analytics, custom client analytics, and data quality/integrity issues. Prior to DDI, Dave spent over 11 years in consulting and product development at Aon Consulting, working on the development and implementation of testing and high-fidelity assessment solutions. Dave helped shape Aon's standard practices in selection, validation, legal defensibility, statistical and psychometric analyses, test and assessment development, web-based product design, quality assurance, and client technical support.

Dave received his B.S. in Psychology (minor in Mathematical Sciences) from Oregon State University and his M.S. and Ph.D. in Psychology (minor in Statistics) from Iowa State University.

Savanna L. Shuntich is an associate attorney with Fortney & Scott, LLC in Washington, D.C. She advises businesses on employee policy matters, OFCCP compliance, and the use of Artificial Intelligence in hiring. Prior to joining Fortney & Scott, Ms. Shuntich focused her practice on employment litigation and appeared before federal and state courts and administrative agencies. She was named a Super Lawyers Rising Star in the field of Employment Litigation. Ms. Shuntich received her bachelor's degree from Davidson College and her law degree from the American University Washington College of Law.

Robert Stewart, Ph.D., is a Senior Research Scientist in Amazon's Global Hiring Science team, where he designs custom personnel assessment solutions to objectively identify top talent, match candidates to roles, provide a great candidate experience, and raise Amazon's hiring bar. Prior to joining Amazon, Dr. Stewart spent six years at PDRI, an SHL company, where he built a variety of human capital solutions for public and private sector clients. His professional experience includes developing alternate validation strategies to provide content, construct, and criterion-related validation evidence for Artificial Intelligence/machine learning-based (AI/ML) tools and fairness research in the AI/ML space. His expertise includes job analysis, assessment development and validation, employee selection, competency modeling, and program evaluation. He holds a Ph.D. and MA in Industrial/Organizational Psychology from the University of Houston and a BA in Psychology from the University of Delaware. Dr. Stewart's work has appeared in peer-reviewed journals, book chapters, professional conferences, and technical reports. He currently resides in Tacoma, Washington.

Eric Sydell, Ph.D., is an industrial-organizational psychologist, seasoned entrepreneur, and skilled consultant with more than two decades of experience working in the recruiting technology and staffing industries. An expert in Artificial Intelligence, machine and deep learning, psychometrics, and their practical application in hiring, Eric regularly writes and speaks on these topics, both in the media and at academic and industry conferences, worldwide, and is the co-author of the book Decoding Talent: How AI and Big Data Can Solve Your Company's People Puzzle, from the Fast Company Press. He has been quoted in Fast Company, USA Today, Yahoo Finance, VentureBeat, Silicon Republic, and many other outlets.

Eric was one of the founding scientists of Shaker International, an innovative hiring consulting and solutions firm, where he directed research and innovation. He currently serves as Executive Vice President of Innovation at Modern Hire, where he oversees all research and product innovation initiatives, including the data science-focused Labs team of PhD-level I-O psychologists and data scientists.

Nancy T. Tippins, Ph.D., is a Principal of the Nancy T. Tippins Group, LLC, where she brings more than 40 years of experience to the company. Her firm creates strategies related to work force planning, sourcing and recruiting, job analysis, employee selection, succession planning, executive assessment, and employee and leadership development. Much of her current work in the area of tests and assessments focuses on evaluating programs for legal risks, including concerns regarding validity, adverse impact, record keeping, consistency in administration, and uses of test information.

Throughout her career, Nancy has participated in the creation and revision of professional standards for tests and assessments, serving on the committees that revised the Principles for the Validation and Use of Personnel Selection Procedures for the Society for Industrial and Organizational Psychology (SIOP) in 1999 and co-chairing the committee that revised the Principles in 2018. She was also a member of the committee that revised the Standards for Educational and Psychological Tests for the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) in 2014 and was a U.S. representative on the committee that revised the ISO 10667 International Assessment standards in 2011. She is currently involved in a SIOP effort to create an addendum to the Principles that addresses tests and assessments that incorporate Artificial Intelligence.

Nancy has made many presentations and authored a number of articles on tests and assessments. Her most recent paper explored the use of Artificial Intelligence in employee selection. Recently, she co-authored Designing and Implementing Global Selection Systems, co-edited the Handbook of Employee Selection, and another edited volume, Technology Enhanced Assessments. She has served as the associate editor for the Scientist-Practitioner Forum of Personnel Psychology and editor of SIOP's Professional Practice Series. She is currently on the editorial boards of the Personnel Psychology, Industrial and Organizational Psychology: Perspectives on Science and Practice, Journal of Psychology and Business, Personnel Assessment and Decisions, and the International Journal of Selection and Assessment.

Active in professional affairs, Nancy has a long-standing involvement with the Society for Industrial and Organizational Psychology where she served as President (2000-2001). She is a fellow of SIOP (Division 14 of the APA), Quantitative and Qualitative Methods (Division 5 of the APA), the American Psychological Association (APA), and the American Psychological Society (APS) and is an active participant in several private industry research groups. Nancy received her M.S. and Ph.D. in Industrial and Organizational Psychology from the Georgia Institute of Technology.

Richard Tonowski, Ph.D, joined the U.S. Equal Employment Opportunity Commission ("EEOC") in 2001 as a Psychologist, worked as the assistant HR director for strategic policy and planning from 2003 to 2006, and then served as Chief Psychologist from 2006 to 2020. In that role he was an in-house expert witness for selection procedures and statistical analyses. He wrote and edited the "On the Legal Front" column in the Society for Industrial and Organizational Psychology's (SIOP's) The Industrial-Organizational Psychologist from 2014 to 2020 and has participated in several sessions at annual SIOP conferences involving legal issues, Big Data, and Artificial Intelligence. During that time, he has also taught a graduate course regarding law, ethics, and I-O psychology issues; previously he had taught a graduate course in human resources management. Prior to his time with EEOC, he had over 20 years of experience involving public sector test development and validation, performance appraisal, employee surveys, diversity management, and labor relations. He has a Ph.D. degree in psychology from Rutgers University and has held senior professional certification in human resources from the Society for Human Resource Management and the Human Resources Certification Institute.

Kelly Trindel, Ph.D., is Head of Machine Learning Trust at Workday. Kelly earned a Ph.D.. in Experimental Psychology at the University of Texas, Arlington. She went on to serve as an Assistant Professor of Psychology at Wingate University, teaching psychology, research design, and statistical analysis courses to undergraduate students. Kelly served for 7.5 years as a Social Science Research Analyst, and later as Chief Analyst, at the Equal Employment Opportunity Commission, where she assisted investigators and attorneys with systemic discrimination case development and analysis, among other duties. Kelly's interest in AI for employment selection began at the Commission where she testified at the EEOC's 2016 hearing on 'Big Data' and organized and led an internal task force on the issue thereafter. Kelly has since worked for technology companies in New York City and Silicon Valley on issues related to developing governance programs for responsible and trustworthy AI systems. She is published in psychology, computer science, and legal journals on relevant issues.

Rae T. Vann is Director of North America Employment Law for Wayfair, a global e-commerce retailer of all things Home. Prior to joining Wayfair in January 2022, Rae spent over twenty years in private practice, where she represented and advised corporate clients on a variety of workplace compliance matters, with an emphasis on EEO, nondiscrimination, and affirmative action. She also previously served as senior vice president and general counsel of the Center for Workplace Compliance, the national employer association formerly known as the Equal Employment Advisory Council (EEAC). Rae's thought leadership has been featured in media including the Wall Street Journal, Bloomberg Law, Law360, National Law Journal, National Public Radio, PBS NewsHour, Politico, and Reuters.

Christine E. Webber is Co-Chair of the Civil Rights & Employment practice group. In this role, Ms. Webber represents victims of discrimination and wage and hour violations in class and collective actions.

Ms. Webber has had the honor of representing clients in some of the largest, groundbreaking discrimination and Fair Labor Standards Act (FLSA) class and collective actions in the United States, including Keepseagle v. Vilsack (D.D.C.), a historic nationwide race-based discrimination class action brought by Native American ranchers and farmers against the United States Department of Agriculture (USDA). The landmark $760 million settlement required the USDA to pay $680 million in damages to thousands of Native Americans, to forgive up to $80 million in outstanding farm loan debt and to improve the farm loan services the USDA provides to Native Americans. Ms. Webber was lead counsel in In re Tyson Foods FLSA MDL (M.D. Ga.), a collective action involving FLSA claims at over 40 Tyson chicken processing plants, which ultimately resolved the claims of 17,000 chicken processing workers who had been denied compensation for donning and doffing required safety and sanitary equipment; and Hnot v. Willis Group Insurance (S.D.N.Y.), where she represented a class of women vice presidents in Willis' Northeast region, who complained of discrimination with respect to their salary and bonuses, as well as promotions. This "glass ceiling" case settled for an average payment of $50,000 per woman, a record-breaking settlement in 2007 for a sex discrimination class action. Ms. Webber continues the fight in Dukes v. Wal-Mart – a nationwide pay and promotion sex discrimination class action that went to the U.S. Supreme Court in 2011 and addressed standards for class certification in employment discrimination matters.

Colin Willis, Ph.D., is an Industrial Organizational Psychologist at HireVue, Inc. and leads the science team's legal strategy and research into neurodiversity, Artificial Intelligence, and hiring in addition to consulting with customers on assessment use. Colin has worked regularly with external partners to develop award-winning products and research and engage with policymakers on Artificial Intelligence and employment. Colin has presented internationally at the Society for Industrial and Organizational Psychology, the European Association for Work and Organizational Psychology, the National Association of Colleges and Employers, the Autism at Work Research

Workshop, and the Age in the Workplace Meeting. Colin received his Ph.D. in Psychology from Colorado State University (Fort Collins, CO).

Ye Zhang, Ph.D., is a Director at Resolution Economics, LLC. He holds Ph.D. and M.A. degrees in Economics from University of Maryland and a B.A. in Economics from Fudan University in Shanghai, China. Prior to joining Resolution Economics, Dr. Zhang directed many large and complex labor and employment policy research studies for Federal, State, and local governments. Prior to his consulting career, Dr. Zhang served as an Assistant Professor of Economics at Indiana University.

Dr. Zhang is a labor economist with significant expertise in economic modelling and statistical analysis in all areas involving labor and employment issues. His practice areas include analysis related to allegations of employment discrimination in compensation, hiring, promotion, and termination. Dr. Zhang also specializes in data analysis related to FLSA wage and hour cases, FMLA actions, state law issues of minimum wage and employee misclassification, EEOC investigations, and OFCCP investigations of federal contractors. Dr. Zhang is also an expert in designing and implementing rigorous policy research and program evaluations. He has developed and applied econometric evaluation methods and statistical analyses to a variety of projects funded by the Wage and Hour Division (WHD) and the Chief Evaluation Office (CEO) of Department of Labor (DOL), State, and local government agencies.

Dr. Zhang's research has been published in the Journal of Population Economics and the Review of Economics of the Household and he has served as referee for many journals such as the Journal of Labor Economics, Journal of Human Resources, Journal of Public Economics and American Economic Journal: Applied Economics. He often makes presentations on the economic analysis of labor and employment issues and has presented research and served on panels for the government and various professional associations.

THE INSTITUTE FOR WORKPLACE EQUALITY FACULTY MEMBERS

David Cohen is the founder and President of DCI Consulting. He provides consulting services to employers and management law firms on a wide range of human resource risk management strategies, particularly in the areas of EEO/affirmative action program development, systemic compensation statistical analyses, comprehensive human resources self-audits, and employee selection and test validation.

In addition, Mr. Cohen is the co-founder of The Institute for Workplace Equality, a national nonprofit employer association that trains and educates federal contractors in understanding and complying with their affirmative action and equal employment obligations.

Recognized as a national EEO and affirmative action compliance expert, Mr. Cohen speaks frequently before corporate leaders from Fortune 500 companies, and at regional and national ILG conferences and OFCCP events. In 2006, he co-authored a book entitled "Understanding Statistics: A Guide for I/O Psychologists and Human Resource Professionals," which was published by Wadsworth. Mr. Cohen is also the Associate Editor of the Applied HRM Research.

He also created DCI's HR Equator™ salary equity software, which enables companies to conduct systemic and other compensation analyses to comply with OFCCP and EEOC requirements.

Mr. Cohen has a M.S. degree in Industrial and Organizational Psychology from Radford University and B.A. degree in Psychology from West Virginia University. He is also an adjunct faculty member at the University of Maryland Baltimore County at Shady Grove.

David S. Fortney is a co-founder of FortneyScott. His practice focuses on a wide array of compliance initiatives and litigation matters involving wage and hour, compensation, Equal Employment Opportunity, labor matters, federal contractor's affirmative action and non-discrimination obligations, collective bargaining, and workplace health and safety. He advises clients in developing and implementing strategic plans involving federal agencies, Congress, and the press that enhance a client's reputation and take into account the interests of various stakeholders, including stockholders, the workforce, and customers. David brings experience from both the public and private sectors and frequently represents clients before federal and state agencies, including the U.S. Department of Labor's agencies, the Equal Employment Opportunity Commission, and the National Labor Relations Board.

David is a Fellow in the College of Labor and Employment Lawyers. For the past 10 years, he has been named one of the leading employment lawyers in Washington, D.C. by the CHAMBERS USA survey of America’s Leading Lawyers for Business and has been selected for inclusion in the 2009 through 2016 editions of The Best Lawyers in America, Washington D.C.’s Best Lawyers, and Super Lawyers. He was also awarded an AV rating (the highest level) by Martindale-Hubbell.

Before co-founding the firm, David served as the Acting Solicitor of Labor and held other senior policy positions in the U.S. Department of Labor. He is a frequent lecturer and writer on employment-related topics.

The Institute for
WORKPLACE EQUALITY

Appendix B Accompanying Technical Advisory Committee Report on EEO and DEI&A Considerations in the Use of Artificial Intelligence in Employment Decision Making

December 2022

1920 I Street NW • Washington DC 20006 • 202.293.2220
Theinstitute4workplaceequality.org

REFERENCES

ACM Code of Ethics and Professional Conduct (Ass'n for Computing Mach. 2018) at § 1.6, https://ethics.acm.org/.

Anna Lena Hunkenschroer and Alexander Kriebitz, *Is AI recruiting (un)ethical? A human rights perspective on the use of AI for hiring*, AI and Ethics (Jul. 25, 2022), https://doi.org/10.1007/s43681-022-00166-4.

Barbara Lindemann & Paul Grossman, *Employment Discrimination Law* 156 and n. 183 (The Bureau of National Affairs, Inc., 1976).

Center for Corporate Equality, *Technical Advisory Committee Report on Best Practices in Adverse Impact* (September 2010), https://irp.cdn-website.com/b44ff977/files/uploaded/TAC-Adverse-Impact-FINAL.pdf.

Center for Democracy & Technology, *Civil Rights Standards for 21st Century Employment Selection Procedures* (December 2022), https://cdt.org/wp-content/uploads/2022/12/updated-2022-12-05-Civil-Rights-Standards-for-21st-Century-Employment-Selection-Procedures.pdf.

Charles A. Sullivan & Lauren Walter, *Employment Discrimination Law & Practice* 303 (Aspen Publishers, 4th ed. 2009).

Christopher M. Berry & Peng Zhao, *Addressing criticisms of existing predictive bias research: Cognitive ability test scores still overpredict African Americans' job performance*, 100 J. of Applied Psych. 162–179 (2015).

Colin Willis, *et al.*, *Examining the Use of Game-Based Assessments for Hiring Autistic Job Seekers*, 9 J. of Intel. 53 (2021), https://doi.org/10.3390/jintelligence9040053.

Donald B. Rubin, *Inference and Missing Data,* 63 Biometrika 581–92 (1976), https://doi.org/10.1093/biomet/63.3.581.

Ethical Guidelines for Statistical Practice (Am. Stat. Ass'n 2022), https://www.amstat.org/docs/default-source/amstat-documents/ethicalguidelines.pdf?Status=Master&sfvrsn=bdeeafdd_6/.

Ethical principles of psychologists and code of conduct (Am. Psych. Ass'n 2002, amended effective June 1, 2010, and January 1, 2017), http://www.apa.org/ethics/code/index.html.

Frank J. Landy, *Stamp Collecting Versus Science, Validation as Hypothesis Testing* (Am. Psych. Ass'n 1986), https://doi.org/10.1037/0003-066X.41.11.1183.

Franziska Leutner, *et al., The Future of Recruitment* (Emerald Publishing Limited 2022), https://doi.org/10.1108/978-1-83867-559-220221009

Franziska Leutner, *et al., The potential of game- and video-based assessments for social attributes: examples from practice*, 36 J. of Managerial Psych. 533 (2021), https://doi.org/10.1108/JMP-01-2020-0023.

International Privacy Subcommittee of the ATP Security Committee, *Artificial Intelligence and the Testing Industry: A Primer, A Special Publication from ATP* (Ass'n of Test Publishers Jul. 6., 2021), https://www.testpublishers.org/assets/ATP%20White%20Paper_AI%20and%20Testing_A%20Primer_6July2021_Final%20R1%20.pdf.

International Test Commission and Association of Test Publishers, *Guidelines for Technology-Based Assessment* (2022), https://www.intestcom.org/upload/media-library/guidelines-for-technology-based-assessment-v20221108-16684036687NAG8.pdf.

Jason R. Bent, *Is Algorithmic Affirmative Action Legal?*, 108 Geo. L. J. 803-853 (2020), https://www.law.georgetown.edu/georgetown-law-journal/wp-content/uploads/sites/26/2020/04/Is-Algorithmic-Affirmative-Action-Legal.pdf.

Joshua A. Kroll, *et al.*, *Accountable Algorithms*, 165 U. Pa. L. Rev. 633 (2017), https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3.

Julie M. McCarthy, *et al.*, *Distressed and distracted by COVID-19 during high-stakes virtual interviews: The role of job interview anxiety on performance and reactions*, 106(8) J. of Applied Psych. 1103–1117 (2021), https://doi.org/10.1037/apl0000943.

Karthika Mohan and Judea Pearl, *Graphical Models for Processing Missing Data*., 116 J. of the Am. Stat. Ass'n 1023–37 (2021), https://doi.org/10.1080/01621459.2021.1874961.

Kaylynn R. Griswold *et al.*, *Global differences in applicant reactions to virtual interview synchronicity*, The Int'l J. of Hum. Res. Mgmt. (2021), https://doi.org/10.1080/09585192.2021.1917641.

Keith E. Sonderling, Bradford J. Kelley & Lance Casimir, *The Promise and The Peril: Artificial Intelligence and Employment Discrimination*, 77 U. MIA L. Rev. 1 (2022), https://repository.law.miami.edu/umlr/vol77/iss1/3.

Kelly Trindel et al., *Fairness in Algorithmic Employment Selection: How to Comply with Title VII*, 35 A.B.A. J. Lab. & Emp. L. 241, 241 (2021), https://www.americanbar.org/content/dam/aba/publications/aba_journal_of_labor_employment_law/v35/no-2/fairness-algorithmic-employment-selection.pdf.

Leo Alexander, III. & Fred L. Oswald, *Free Adverse Impact Resource* (2019), https://orgtools.shinyapps.io/FAIR.

Matthew U. Scherer, *et al.*, *Applying Old Rules to New Tools: Employment Discrimination Law in the Age of Algorithms*, 71 S.C. L. Rev. 449 (2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3472805.

Nancy T. Tippins, Frederick L. Oswald & S. Morton McPhail, *Scientific, Legal, and Ethical Concerns about AI-Based Personnel Selection Tools: A Call to Action*, 7 Pers. Assessment and Decisions 1-22 (2021), https://scholarworks.bgsu.edu/cgi/viewcontent.cgi?article=1170&context=pad.

Organisation for Economic Co-operation and Development ("OECD"), *Recommendations of the Council on Artificial Intelligence*, OECD/LEGAL/0449 (Adopted, May 21, 2019), https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

Pauline T. Kim, *Auditing Algorithms for Discrimination*, 166 U. Pa. L. Rev. Online (2017), https://scholarship.law.upenn.edu/penn_law_review_online/vol166/iss1/10.

Pauline T. Kim, *Race-Aware Algorithms: Fairness, Nondiscrimination and Affirmative Action*, 110 Cal. L. Rev. (Oct. 2022), https://californialawreview.org/print/race-aware-algorithms-fairness-nondiscrimination-and-affirmative-action/.

Principles for the Validation and Use of Personnel Selection Procedures (Soc'y for Indus. Organizational Psych., 5th ed., August 2018).

Ramona L. Paetzold & Steven L. Willborn, *The Statistics of Discrimination: Using Statistical Evidence in Discrimination Cases* (Thompson Reuters, 2013-2014 ed. October 2013).

Reva Schwartz, *et al.*, *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence, Special Publication 1270*, National Institute of Standards and Technology (2022), https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf.

Richard Tonowski, *Thoughts from an EEO Agency Perspective* (Taylor & Francis Group, 1st ed. 2016).

Robert M. Guion, *Assessment, Measurement, and Prediction for Personnel Decisions* (Routledge, 2nd ed., 2011), https://doi.org/10.4324/9780203836767.

Sara Kassir, *et al.*, *AI for hiring in context: a perspective on overcoming the unique challenges of employment research to mitigate disparate impact*, AI and Ethics (Sept. 22, 2022), https://doi.org/10.1007/s43681-022-00208-x.

Sara Mattingly-Jordan, *et al*., *Ethically Aligned Design, First Edition Glossary* (Inst. of Elec. and Electronics Eng'r 2019), https://ethicsinaction.ieee.org/wp-content/uploads/ead1e_glossary.pdf.

Sayash Kapoor & Arvind Narayanan, *Leakage and the Reproducibility Crisis in ML-based Science* (2022), https://arxiv.org/pdf/2207.07048.pdf.

Scott B. Morris & Eric M. Dunleavy, *Adverse impact analysis: Understanding data, statistics and risk* 287 (Routledge, 1st Ed. 2016).

Scott B. Morris & Russell E. Lobsenz, *Significance tests and confidence intervals for the adverse impact ratio*, 53 Personnel Psychology 89-111(2000), https://doi.org/10.1111/j.1744-6570.2000.tb00195.x.

SWJ Nijman, *et al.*, *Missing data is poorly handled and reported in prediction model studies using machine learning: a literature review*, 142 J. of Clinical Epidemiology 218-229 (2022), https://www.jclinepi.com/action/showPdf?pii=S0895-4356%2821%2900375-9.

Timnit Gebru*, et al*., *Datasheets for Datasets*, 64 Communications of the ACM 86-92 (2021), https://doi.org/10.1145/3458723.

Tlamelo Emmanuel, *et al*, *A Survey on Missing Data in Machine Learning*, 8 J. of Big Data 1-7 (2021), https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00516-9.

Tressy Thomas & Enayat Rajabi, *A systemic review of machine learning-based missing value imputation techniques*, 55 Data Tech. and Application 558-585 (2021), https://doi.org/10.1108/DTA-12-2020-0298.

U.S. Equal Emp. Opportunity Comm'n ("EEOC"), *The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees* (May 12, 2022).

World Economic Forum, *A Blueprint for Equity and Inclusion in Artificial Intelligence White Paper* (June 2022), https://www3.weforum.org/docs/WEF_A_Blueprint_for_Equity_and_Inclusion_in_Artificial_Intelligence_2022.pdf.

World Economic Forum, *Human-Centered Artificial Intelligence for Human Resources: A Toolkit for Human Resources Professionals* (Dec. 2021), https://www3.weforum.org/docs/WEF_Human_Centred_Artificial_Intelligence_for_Human_Resources_2021.pdf.