Phenom Fit Score

2024 Report

phenom

Table of Contents

Executive Summary

Phenom Fit Score helps recruiters decide whether or not to invite a candidate to a job interview.

Using an industry-standard and legally defensible analysis, Phenom has examined the potential for Fit Score usage to lead to adverse impact.

Statistical tests show that the use of Fit Score is unlikely to introduce or worsen adverse impact during the interview decision stage of an employee selection process.

- **Fit Score recommended moving candidates forward to the interview phase at rates that were comparable across gender groups.**
- **Fit Score recommended moving candidates forward to the interview phase at rates that were comparable across racial/ethnic groups.**
- **Fit Score recommended moving candidates forward to the interview phase at rates that were comparable across gender and racial/ethnic intersectional groups.**

These findings show that Fit Score contributes to customers' fair hiring processes with respect to gender, race/ethnicity, and the intersection of the two.

This report is intended to give Phenom customers insight into the Fit Score model. The information provided herein does not, and is not intended to, constitute legal advice.

About the Report

Introduction to Phenom Fit Score

What is Phenom Fit Score?

Fit Score is an AI model that powers the Phenom Intelligent Talent Experience platform. This model helps recruiters identify candidates to interview, bringing speed and objectivity to talent acquisition. Fit Score on its own does not make hiring decisions. Fit Score is designed to be part of a hiring process, where it can be overseen by humans.

Why did Phenom develop Fit Score?

Phenom developed Fit Score to help recruiters handle large flows of job seeker data and empower recruiters with objective and fair decision support. The reality of modern recruiting is that the volume of job seekers can be overwhelming, such as when a Phenom career site attracts many more applicants than a traditional system. Further, recruiters may be unfamiliar with some applicants' highly specialized skills. These challenges of large volumes of data and resume complexity can be exacerbated by time pressures from hiring managers and job seeker expectations. Under pressure, human decision-making can introduce bias, or make bias harder to detect.

Fit Score rises to these challenges by streamlining the hiring process and saves time and resources by prioritizing candidate job fit in terms of objective job-related characteristics. This allows recruiters to maintain oversight over a large number of job applications that need to be reviewed manually, ultimately leading to better outcomes for both the organization and job candidates.

How does Phenom Fit Score work?

Fit Score works with objective job-related characteristics, such as relevant career paths and skills-based hiring, to compute a value that aims to predict how a given candidate profile fits the job match criteria. To a recruiter, this value is represented as a letter grade: A, B, C, and No Fit.

In brief, the Fit Score model computes the letter grade by comparing the resume to the job description. The AI behind Fit Score includes AI language modeling of similarities between candidate information and job-related criteria, such as the match between a candidate's skills and the skills listed in a job description. Fit Score is applied to all candidates equally, regardless of their demographic or other protected characteristics. The only inputs to Fit Score are job-related; there are no inputs of demographic or biometric information.[1]

What are appropriate ways of using Fit Score?

Fit Score helps recruiters decide whether or not to invite a candidate to a job interview. In general, the higher the letter grade, the stronger the match between a resume and the job match criteria.

One simple, intuitive way to think about Fit Score is that it is an approximate, at-a-glance summary of how well the candidate matches a job. Without AI, a recruiter could sort and filter candidates only by structured fields or Boolean keyword queries, such as the name of a specific prior employer or an exact matching job title. Without AI, a search for a "computer programmer" would miss all those who had the title "software engineer," and vice versa. With Fit Score, it is possible to sort and filter candidates in terms of their attributes relative to job requirements, such as whether the resume lists a skill that is a close semantic match to the job criteria. The recruiter can manage the inputs to Fit Score for both the candidate and the job, the latter affecting all candidates. In this way, Fit Score can represent whatever is important to the recruiter.

[1]Additional background information on Fit Score can be found in the Phenom white paper *Empowering Recruiters with Phenom AI*, available on request from Phenom Product Knowledge Management.

Recruiters and employers have established workflows prior to becoming Phenom customers, and these workflows can incorporate Fit Score in different ways. Consider these examples:

A recruiter posts a position and receives a number of applications, each of which is processed by Fit Score. The recruiter reviews several applications that have Fit Score grades of “A” and “No Fit” to confirm that Fit Score is calibrated in accordance with the recruiter’s expectations. The recruiter can then use Fit Score to have an at-a-glance summary of the application strengths in terms of the job matching criteria.

A recruiter posts a position and receives a large number of applications, each of which is processed by Fit Score. In the past, to save time and resources, the recruiter’s go-to strategy without Fit Score was to only closely examine the applicants who were employed in a similar role by prominent competitors. This approach screens candidates and imparts a condition that may not necessarily be job-relevant. There may be a risk of overlooking diverse talent. With Fit Score, the recruiter has a better alternative. Fit Score makes decision-making based on job-relevant attributes easy, eliminating the need for screening with potentially irrelevant criteria. As Fit Score is computed for all applications, all candidates can be considered fairly. The recruiter can now prioritize top applicants. For example, applications that receive grades “A” or “B” can be reviewed manually for additional factors, before scheduling initial interviews.

A recruiter posts a position and receives a number of applications, each of which is processed by Fit Score. On paper, the position has two hiring managers, but the recruiter knows which of the two will make the ultimate decision. The recruiter adjusts the job matching criteria to express that manager’s priorities especially closely.

A recruiter posts a position and receives a number of applications, each of which is processed by Fit Score. Fit Score gives these applications grades “C” and “No Fit”; no applications receive grades “A” or “B”. The recruiter uses this distribution of grades in the applicant pool to determine that the job posting is not reaching its desired audience, and revises the posting and its marketing.

By contrast, it is inappropriate to use Fit Score as the sole basis or most important factor in choosing how to act on applications, or to use Fit Score to overrule factors such as human decision-making.

Is Fit Score an Automated Employment Decision Tool under New York City Local Law 144?

As designed, Phenom Fit Score is not intended to be an Automated Employment Decision Tool ("AEDT") under **New York City Local Law 144** (LL 144).

New York City Local Law 144 (LL 144) "prohibits employers and employment agencies from using an automated employment decision tool unless" certain conditions are met to audit the tool and notify the public. The New York City Department of Consumer and Worker Protection adopted rules to implement LL 144. The rules define what an Automated Employment Decision Tool (AEDT) is and what is required to audit an AEDT. The rules were adopted in April 2023 and LL 144 took effect on July 5, 2023.

According to the adopted rules (**6 RCNY § 5-300**), an algorithm is an AEDT if it is used "to substantially assist or replace discretionary decision making", defined as follows:

1. to rely solely on a simplified output (score, tag, classification, ranking, etc.), with no other factors considered; or
2. to use a simplified output as one of a set of criteria where the simplified output is weighted more than any other criterion in the set; or
3. to use a simplified output to overrule conclusions derived from other factors including human decision-making.

It is inappropriate to use Fit Score as the sole basis (i) or most important factor (ii) in choosing how to act on applications, or to use Fit Score to overrule factors such as human decision-making (iii). None of these three criteria apply to Fit Score when Fit Score is used appropriately. If Fit Score is misused or relied too heavily upon by an employer, it could be considered an AEDT.

Phenom developed Fit Score for other purposes, and recommends that customers use Fit Score only in ways that match its intended design. Phenom communicates appropriate practices to customers in trainings and other customer communications. (For examples of how to use Fit Score appropriately, see **What are appropriate ways of using Fit Score?** above.)

Has an AI risk assessment been conducted for Fit Score?

Yes. Phenom conducts risk assessments following its Governance Policy for AI Technologies. This policy is based on the **Model Artificial Intelligence Governance Framework** originally developed by the government of Singapore for the World Economic Forum, and it is similar to other AI governance approaches, such as the **Google AI Principles**.

The Governance Policy states that a risk assessment should consider the probability that the AI may lead to harm, the potential severity of harm, and strategies for mitigation.

According to the risk assessment, Fit Score, when used according to best practices, has a low probability of harm and a low severity of harm. The risk assessment suggests monitoring for adverse impact as one approach to harm mitigation. This audit report implements this suggested harm mitigation approach.

What are bias and adverse impact?

Adverse impact refers to the tendency of certain policies, practices, or procedures to disproportionately and negatively affect certain groups of people, based on their protected class. The legal definition of adverse impact is a "substantially different rate of selection in hiring, promotion, or other employment decision which works to the disadvantage of members of a race, sex, or ethnic group." Adverse impact reports are commonly used in legal inquiries on employee selection.

Adverse impact can occur even when the policies or practices in question are not intended to discriminate. Adverse impact examines the consequences of policies and practices, not their intent.

The term "bias" has **many definitions**, including in cognitive psychology, statistics, computer science, and industrial/organizational psychology. For example, the American Psychological Association (APA) Principles for the Validation and Use of Personnel Selection Procedures defines bias as "systematic error in a test score that differentially affects the performance of different groups of test takers". In this definition, bias relates to the tool and score itself. From the perspective of computer science and software engineering, Phenom Fit Score development addresses bias by only using job-relevant variables and developing our model on broadly representative data.

Scientifically, bias and adverse impact are distinct concepts, but adverse impact analyses provide a practical way to think about bias. For example, selection rates and impact ratios are the foundation of the "4/5ths rule", which is one way to measure adverse impact. Selection rates, scoring rates, and impact ratios are also how New York City Local Law 144 defines its "bias audit" formula.

What is a bias audit, and why is it necessary?

A **bias audit** is a process of examining and evaluating data, algorithms, or systems for potential biases that could negatively impact certain groups or individuals. The goal of a bias audit is to identify unfair or discriminatory practices in a particular system or process.

Bias audits are necessary because they help ensure that systems and processes are fair for all users, regardless of their demographic characteristics. Without a bias audit, these biases may go unnoticed. After an audit, corrective actions may be taken to mitigate bias.

If Fit Score does not use protected class information in its design, why is an audit desired?

By conducting a bias audit, we hope to demonstrate Phenom's commitment to fairness and to help build trust and confidence in our systems. Given Fit Score's design, the probability of bias is very slim. Still, an adverse impact analysis provides concrete evidence.

In addition, by conducting a bias audit, we aim to help our customers comply with regulatory requirements. Many hiring practices in U.S. organizations are subject to regulation at the federal level (i.e., EEOC) or from state or local governments. These regulations call for identifying and mitigating bias in hiring decisions.

Audit Scope and Process

How was the bias audit conducted?

A Fit Score bias audit was conducted through an adverse impact analysis by a third party. For this analysis, Fit Score was considered to be an assessment that gives a score to each candidate who applied to a job, and a rule that recommends whether a candidate should or should not be invited to an interview based on that score. Statistical tests were then conducted to determine if Fit Score resulted in any significant differences in the rates at which members of demographic subgroups were recommended to advance in the hiring process.

What protected classes were covered in this audit?

Results from a series of statistical tests of group differences in interview recommendations based on Fit Score were examined to identify any patterns of adverse impact.

This audit included three demographic categories: gender, race/ethnicity, and the intersection of gender and race/ethnicity.

Gender was defined as a binary variable, Male or Female. Race/ethnicity was defined as a categorical variable, including Native American or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian or Pacific Islander, White, and Two or More Races. The definitions of these variables were a “common denominator” across candidate self-identification questionnaires on job application forms.

In any statistical analysis such as this one, it is common for some group comparisons to not be included for various reasons such as:

Sample Size

The adverse impact report only includes groups with a large enough sample size to ensure statistical reliability. If a particular group has a small sample size, it was not included in the report as the results would not be statistically significant.

Not a protected group

Adverse impact reports typically focus on protected groups such as race, gender, age, and disability.

Lack of data

Not all protected group classifications were collected. For example, sexual orientation is not typically collected on candidate self-identification forms, and therefore could not be included in the report.

Does this report comply with standards and best practices?

Yes. The EEOC's Uniform Guidelines for Employee Selection Procedures (UGESP) is the cornerstone of U.S. federal regulations on identifying adverse impact in hiring practices. The UGESP has given rise to statistical methodologies and best practices. Our audit of Fit Score follows these methodologies and practices.

Does this report satisfy the requirements for an independent auditor?

Yes. Conductor AI, the third-party auditor that conducted this audit, is independent of Phenom. A link to Conductor AI's test results is provided in the appendix to this report.

Does this audit satisfy the "bias audit" requirement of New York City Local Law 144 on Automated Employment Decision Tools?

Yes, this report satisfies the requirements of Local Law 144 (LL 144). However, **Phenom Fit Score was not designed to be an Automated Employment Decision Tool, so New York City Local Law 144 may not apply to Fit Score when Fit Score is used as intended**. In the spirit of LL 144 specifically and supporting regulation and transparency in general, Phenom sought a third-party audit of Fit Score as if LL 144 did apply to Fit Score.

What data sets were used in preparing this audit report?

The report includes a subset of anonymized and de-identified data from some Phenom customers who used the Phenom platform. It does not include data from all Phenom customers, or all jobs or applicants from any one customer. The sample represents Fit Score use cases appropriately, as the data covers multiple United States geographies, jobs, job families, employers, and talent pools. All data included in this report pertain only to applications in the United States between January 1st and December 31st of 2023.

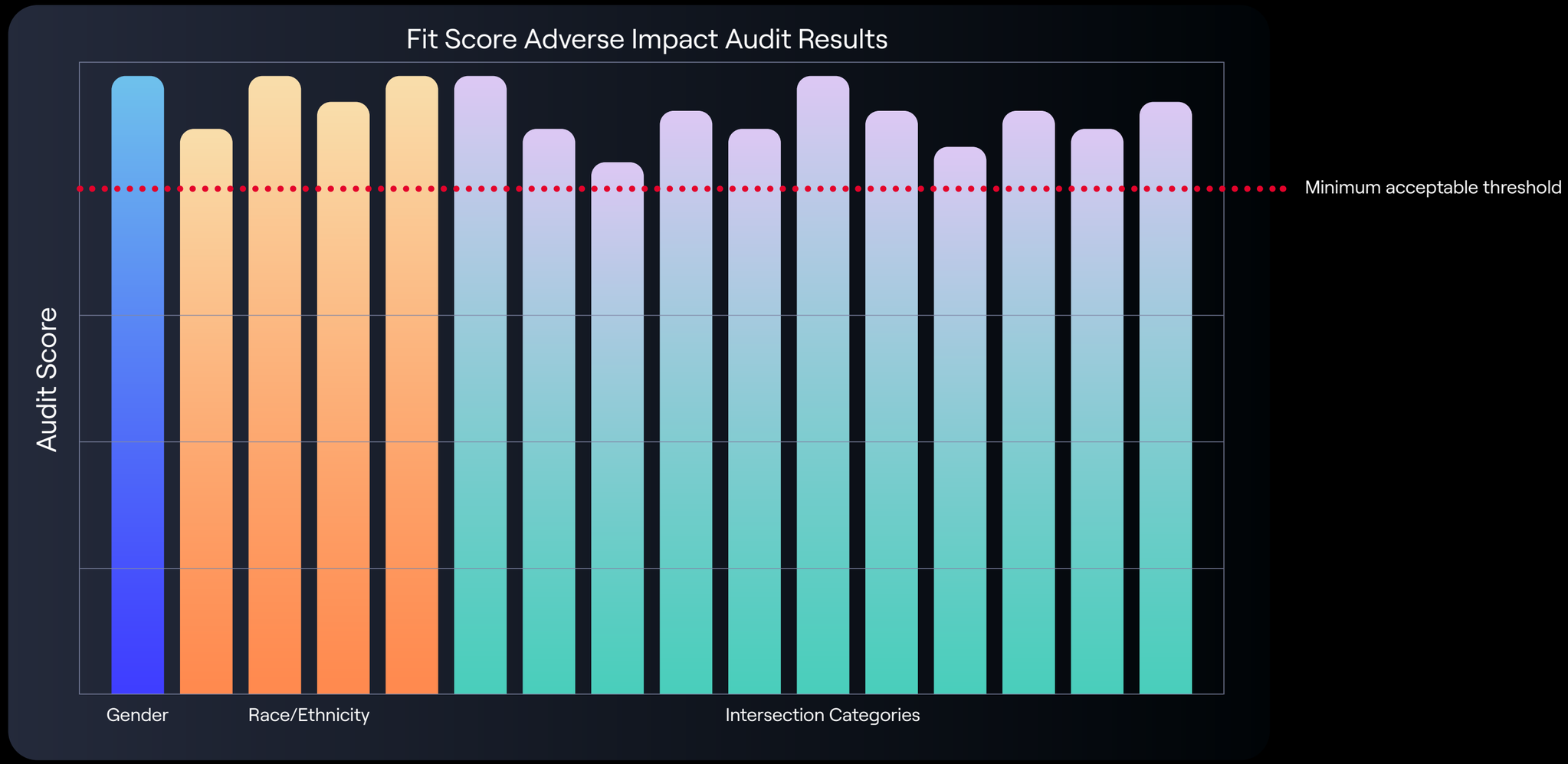
How were customer data and candidate data protected?

Confidentiality, data privacy, and integrity were central to this process. Customer information was never shared with third parties and all analyses and findings were determined based on aggregated non-identifiable applicant data from a broad pool of candidates. Our approach complies with Phenom policies on information security and data privacy.[2]

[2]Relevant Phenom policies are available on request from Phenom Account Management, Customer Value and Global Customer Care teams.

Audit Findings and Impact

The following discussion interpreting the audit findings is solely the responsibility of Phenom and not the third-party auditor. As is typical for LL 144 audits, the auditor's scope of work only included computing the relevant impact ratios, not interpreting them.



What did the audit find?

The results from a series of statistical tests comparing group scoring rates based on Fit Score of multiple protected classes of over three million applicants showed no evidence of adverse impact. The audit found that Fit Score recommendations have no adverse impact for gender, no adverse impact for race/ethnicity, and no adverse impact for gender and race/ethnicity intersectional demographic categories.

This dataset included Phenom customer hiring data across 21 job families. In total, the dataset included over 60,000 jobs and over 3,000,000 applicants.

Families with at least 20K jobs	Families with fewer than 20K jobs
Community and Social Service	Farming, Fishing, and Forestry
Life, Physical, and Social Science	Construction and Extraction
Transportation and Material Moving	Building and Grounds Cleaning and Maintenance
Educational Instruction and Library	Personal Care and Service
Production	Installation, Maintenance, and Repair
Architecture and Engineering	Protective Service
Healthcare Support	Legal
Arts, Design, Entertainment, Sports, and Media	Food Preparation and Serving Related
Healthcare Practitioners and Technical	
Office and Administrative Support	
Sales and Related	
Business and Financial Operations	
Management	
Computer and Mathematical	

Is Fit Score a fair tool to use in hiring procedures?

Based on this adverse impact analysis, we confirm that Fit Score is a fair selection tool for prioritizing candidates for interviews.

If I use Fit Score as part of my selection procedure, what is the risk of treating protected groups unfairly?

As Fit Score is based on only job-related characteristics, Phenom does not expect a direct influence of demographics on Fit Score. By design, all candidates, regardless of their demographic characteristics, are treated equally when their data is input into the Fit Score computation process. This adverse impact analysis confirms that the risk of treating protected groups unfairly by using Fit Score is low.

Can Fit Score help to reduce adverse impact in hiring?

Yes. Many conventional hiring practices can subject protected groups to unintentional unfair treatment. Fit Score promotes fair treatment by providing an objective evaluation of candidates based on job-related attributes such as skills and standardizing hiring criteria for a job for all candidates, which minimizes conscious and unconscious biases.

My company is a Phenom customer and uses Fit Score. Can my employer rely on this bias audit for LL 144 compliance?

Under LL 144, an employer "may rely on a bias audit of an AEDT" under one of two conditions (**§ 5-302 Data Requirements**):

- **the employer provided their historical data to the auditor conducting the bias audit, OR**
- **the employer has never used the vendor's AEDT before**

Therefore, employers that use Phenom Fit Score in their selection process and that need to comply with LL 144 should contribute their data to Phenom's bias audit. Employers who do not need to comply with LL 144 are welcome to also contribute their data, as the generalizability and reliability of the overall audit will improve with a larger pool of employers and jobs.

How do I contribute data to Phenom's bias audit so that I can rely on the audit for LL 144 compliance?

Please reach out to your Phenom point of contact, typically the Customer Value Manager or Account Manager.

Do the findings from this audit generalize beyond LL 144?

Because different employers have their own recruiting and hiring workflows, Phenom customers differ in how they use Fit Score. A comprehensive adverse impact audit for an employer ought to be specific to the employer's jobs, candidates, hiring process, and role of Fit Score in that process.

That said, the findings in this report can be generalized. These findings provide useful insights as a starting point in understanding the use of Fit Score in many organizations and hiring processes. The applicability of these findings depend on various factors such as the size of an organization, types of jobs and job family classifications, the number of applicants and interviews in each job, the diversity of the applicant pool and workforce, and the selection processes aside from Fit Score.

Fit Score is the only AI element in my company's hiring process. Is Fit Score the only element that I need to audit?

It is not sufficient to audit only AI elements of a hiring process.

Customers that are subject to federal and/or local laws on non-discrimination audit and reporting, such as Title VII of the Civil Rights Act of 1964, the Age Discrimination in Employment Act, or the Americans with Disabilities Act, should not solely rely on Fit Score to demonstrate compliance with the respective obligations.

Laws require employers to demonstrate non-discrimination and/or affirmative action practices throughout the entire hiring process (and potentially also after finding an employee). Specifically, if an audit is conducted, the focus of the audit and the legal issue will be whether (1) the selection process as a whole results in adverse impact against protected groups and (2) whether the selection criteria used are job-related and consistent with business necessity. Fit Score is a tool to aid recruiters and does not make decisions by itself. Use of Fit Score may be only one element of a hiring workflow, and companies differ in their workflows that include Fit Score. Although we conducted an audit of Fit Score, selection processes that involve Fit Score should not automatically be considered free of bias. Sole reliance on our Fit Score audit cannot fully alleviate any legal burden.

Employers should conduct regular audits of their overall selection processes to identify bias against protected groups.

What best practices should my company follow in using Fit Score to manage the risk of adverse impact in our hiring process?

Fit Score alone does not make hiring decisions. Phenom recommends that Fit Score should be used as one part of a hiring process, subject to human-in-the-loop control by a recruiter. (See **What are appropriate and inappropriate ways of using Fit Score?**) Here are additional best practices that an organization can adopt to further mitigate the risk of adverse impact.

Conduct a job analysis

Identify the essential functions of the job and the knowledge, skills and abilities required to perform these functions. Use the Edit Matching Criteria function in the Phenom CRM to ensure that the matching criteria used by Fit Score are appropriate to the job.

Train hiring process personnel

Train recruiters and hiring managers on the importance of using fair hiring practices, identifying and addressing potential bias, conducting interviews, and evaluating candidates.

Use multiple selection criteria

Properly designed, multiple selection criteria and evaluation methods will converge on a holistic evaluation of a candidate. Depending on the type of occupation, helpful evaluation approaches may include general ability and personality tests, work sample tests, and structured interviews.

Review job descriptions

Ensure they accurately reflect the essential functions of a job and its requirements. Avoid language that may discourage qualified individuals from applying. Scientific research suggests that the language of job descriptions can affect men and women differently, especially in majority-male occupations. Use the Phenom X+ job description generator to create tailored job descriptions and avoid biased language.

Standardize selection processes

Evaluate candidates in a consistent way. Use the same interview procedures, questions, and scoring systems for all candidates.

Monitor your hiring practices for bias

Review your hiring practices to manage the risk of bias. This includes, but is not limited to, measuring adverse impact, reviewing selection criteria, and analyzing your interview questions.

Will Phenom produce audit reports in the future?

We aim to release a large-scale audit report annually.
We previously released a bias audit report in 2023.

APPENDIX - THIRD-PARTY BIAS AUDIT RESULTS