

HIDDEN BIAS IN HIRING

Examining Applicant
Screening Tools



**Learning
Collider**



”

Algorithms have been shown to outperform human decision-makers across an expanding range of settings...

Yet the rise of algorithms is not without its critics who caution that automated approaches may codify existing human biases and allocate fewer resources to those from underrepresented groups.

Danielle Li, Lindsey Raymond, and Peter Bergman. *Hiring as Exploration*

ACKNOWLEDGMENTS

This white paper summarizes actionable findings from collaborative research conducted by economists from NBER, MIT, and Learning Collider. For in-depth technical explanations and results, refer to the working paper:

*Hiring as Exploration
Danielle Li, Lindsey Raymond, and Peter Bergman (Learning Collider's Founder)
NBER Working Paper: 27736*

Learning Collider extends gratitude to our research partner and anonymous data provider. Access to real-world data leads to relevant, impactful research, especially in the development of responsible technology as a mechanism for equity and economic mobility.

We appreciate our readers for exploring the research findings and sharing this resource widely.



ABSTRACT

Hiring algorithms are not new. These gatekeepers are here to stay. What's needed now is a deeper, collective understanding of how they work - from development to decision-making.

Hidden Bias in Hiring examines machine-learning algorithms used to screen applicants in the early phases of the hiring funnel. We begin with a foundational understanding of hiring algorithms, how they function, how they are trained, and why it matters. We then delve into opportunities to improve algorithmic hiring technology.

This white paper draws from collaborative academic research leveraging real-world data from an anonymous Fortune 500 company.

A team of researchers - including Learning Collider's founder - initiated their study by building and testing a typical screening tool. What they found wasn't surprising: **standard machine-learning algorithms** are problematic in their design, training, and outputs. Instead of their perceived objectivity, this commonplace tool codifies the bias and blindspots of antecedent human decision-makers.

By dissecting the inner workings of the standard model, researchers identified opportunities to improve the algorithmic design, training, and subsequent decision-making. Their next phase was developing and testing an **inclusive machine-learning algorithm**.

Through simulations with data from roughly 89,000 job applications, the inclusive model outperformed the standard model and human recruiters, selecting more diverse and high-quality applicants for interviews. The estimated savings, given boosted hiring rates, is more than 2/3 of the total interviewing cost (from \$1.1m down to \$337,000 for 445 offers).

Inclusive machine-learning algorithms have the potential to transform applicant screening by retaining the cost-efficiency benefits of technology while systematically improving hiring decisions and advancing economic mobility for underrepresented applicants.

03 INTRODUCTION

04 A PRIMER ON ALGORITHMS IN HIRING

08 HOW WE SCREEN MATTERS

12 HIRING BY SCREENING TYPE + CONCLUSION



INTRODUCTION

Who is Learning Collider? We are social scientists and data scientists who envision technology as a force for good. We conduct research at the intersection of business benefit and social impact, where evidence-based innovations are positioned for scale.

Recently, our research collided with the beating heart of business: Human Resources.

Despite labor shortages, high turnover, and pandemic pivots, the heart of business continues to pump. Human capital flows in all directions as recruiting teams work to fill every artery and capillary, supplying the entire business with the people needed to survive. The work of this muscle is integral - yet stressful - and filled with opportunity.

That opportunity is the upside of present-day challenges in HR. Labor shortages can push employers to reconsider and revise job requirements, like dropping education or experience minimums. This opens up opportunities for minority candidates who are less likely to have four-year college degrees, attend elite universities, or have post-graduate degrees. Dropping educational requirements can immediately diversify hiring pools. And by doing so, we often learn these credentials rarely correlate with high performance.

Another opportunity is rethinking hiring and HRTech stacks. In particular, a deeper look at hiring algorithms reveals the problematic nature of their designs, training, and outputs. **But. We can make hiring algorithms better - for employers, recruiters, and applicants.**

Learning Collider studied the ins and outs of applicant screening algorithms using real-world data from an anonymous Fortune 500 firm. We uncovered deficiencies and biases in standard screening algorithms, an opportunity for more learning.

Our research then led to the development, iteration, and testing of inclusive machine-learning algorithms that yield **better hiring potential** while **increasing the diversity of applicants selected** for interviews. Before we get to those details, we'll lay some groundwork, explaining how the standard screening algorithm works and where the trouble lies.



A PRIMER ON ALGORITHMIC TECHNOLOGY IN HIRING

In each stage of the hiring funnel, algorithms make decisions for recruiting teams.

Websites like Indeed and ZipRecruiter use algorithms to determine who sees which postings. These same platforms - and thousands of others - make it possible for applicants to find and apply for jobs in abundance, the precursor to our algorithms of interest.

Opportunities for employers and prospective applicants abound from these market-and-match algorithms.

For employers, new hires can be sourced from broad and diverse labor markets. For prospective applicants, jobs with upward mobility - across geographic borders and sectors - are seemingly accessible with a few clicks...until the next step.

FOCUSING ON APPLICANT SCREENING

Applicant screening, candidate reviewing, resume screening, or some mashup of these names - we'll call them screening algorithms - are tasked with the initial screening of applicant pools. Increasingly, recruiters lean entirely on efficient screening algorithms to select applicants for interviews.

In the case of Learning Collider's data provider, a Fortune 500 firm hiring for high-paying professional service jobs with upward mobility, an average of 200 applications are received for each hire made. Their screening process cuts this pool by 95%, leaving 5% to be interviewed. Of those interviewed, 10% are hired.

Screening algorithms are packaged in many forms. They are as simple as the felon box, a checkbox acknowledging an applicant's penal history that has been used to screen and discard applicants quickly. While now banned in most states, the felon box perpetuated a legacy of racial discrimination in hiring.

More and more, companies rely on more complex machine-learning algorithms to incorporate past hiring decisions into future hiring decisions. These algorithms may be built into a company's talent management software and trained by internal data from historical candidate pools. Alternatively, third-party recruiting vendors train algorithms by scraping data from platforms like LinkedIn to learn and formulate patterns.

In terms of speed, these tools outpace human recruiters, translating to less personnel time, higher productivity, and lower hiring costs. These immediate benefits - not necessarily hiring results - drive their demand.

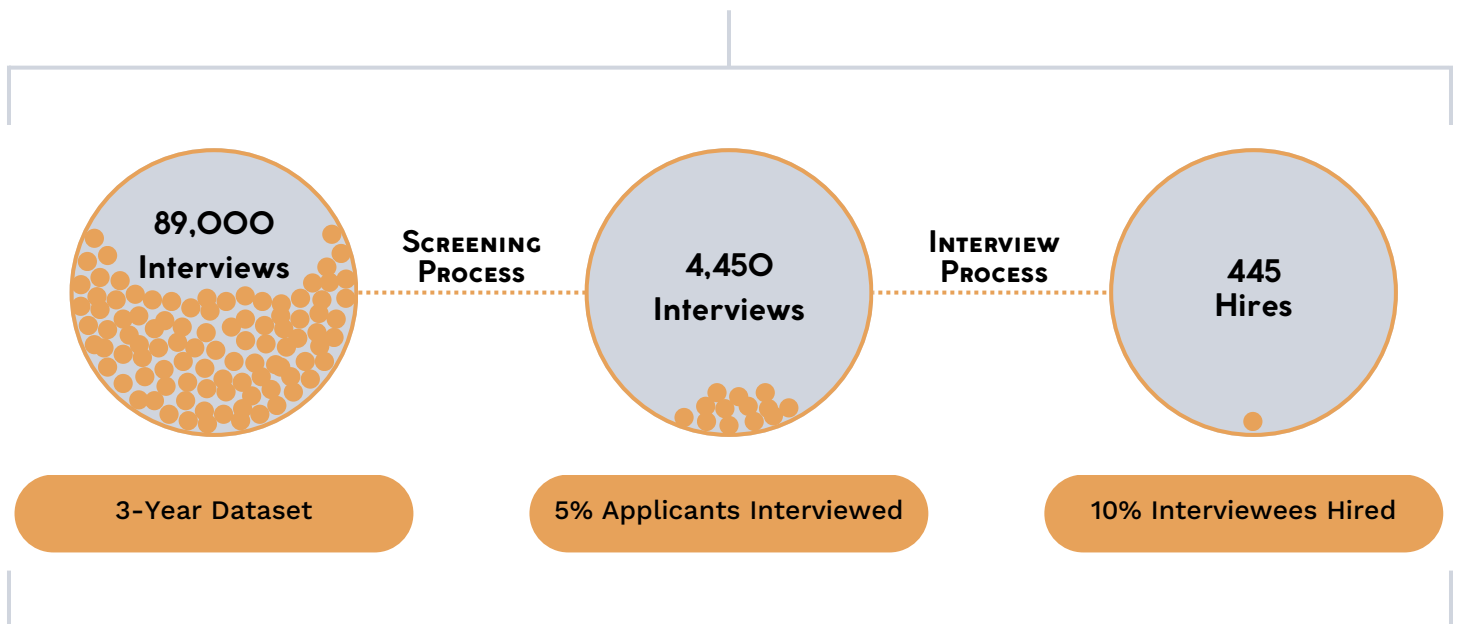
A deeper look into their design, more specifically their training, reveals the problematic nature of standard exploitative machine learning. First, let's peek at where our research started: data from nearly 89,000 job applications received by a Fortune 500 firm from 2016 to 2019.

Where We Started

The groundwork for *Hiring as Exploration*




One (Anonymous) Fortune 500 Firm



\$1,100,000 Interview Cost

Assuming a per-interview cost of \$250, this scenario expends \$1.1m for just one step of the complex hiring process.

 How can this firm - and others - leverage the screening process to improve diversity and overall hiring? How can these improvements be made without increasing costs or compromising hiring quality?



“TRAINED” EXPLAINED

All machine-learning (ML) algorithms - no matter their application or intent - are “trained” to make decisions using some input data and some definition of success.

Standard ML screening algorithms are designed to replicate recruiters' decisions on whom to interview or hire. For example, training could incorporate the following:

- > **Input Data: historical applicant data, typically from resumes.**
- > **Success: the average likelihood of being hired.**

From historical data, a standard screening algorithm will learn the likelihood that applicants with specific characteristics should be hired. If four candidates studied chemistry, and two were hired, the algorithm will conclude that chemistry graduates have a 50% chance of success.

Once trained, algorithms produce outputs; in this case, applicants to interview. The algorithm applies its training to new data - job applications - to find and match patterns, producing applicants who share traits with those hired in the past.

Some standard ML algorithms remain constant or **static** in their training. Others are designed with a feedback loop to learn from subsequent outcomes, known as **updating** or dynamic. Whether static or updating, the training baseline is embedded in the past. The perpetuation of past hiring cycles should raise a red flag when hiring for the future.

On the surface, the training may seem logical. “Remember Jared? He was great! Let’s hire another Jared!” Standard ML algorithms can select more “Jareds” to interview. As we dig further, we remember this: people, skills, and workplaces constantly evolve.

WHERE THE TROUBLE LIES

Arguably the most infamous example of ML screening algorithms comes from Amazon. The company invested significant resources to design and train a “holy grail” automated hiring tool. Instead, Amazon made headlines for codifying a preference for male applicants and discriminating against women. The four-year project was shelved in 2018.

Similarly, another HRTech company developed a black-box tool with odd predictive traits. Applicants named Jared who once played high school lacrosse correlated with strong job performance.

You may have inferred where the trouble lies. It’s not the bot. It’s us.

Humans make the decisions that design and train these tools. Then, once in play, standard ML screening algorithms replicate - or regurgitate - decisions made by human recruiters. But whom have recruiters historically interviewed and hired?



As we know, humans are consistently inconsistent, noisy, and unconsciously (or consciously) biased. Any human blindspot or bias can be learned and codified by algorithmic gatekeepers.

Let's review a more relatable blindspot than the outrageous Jared who played high school lacrosse: a college education. Recruiters typically know elite schools, neighboring colleges, and large state schools; past recruiting methods may have included on-campus events or job fairs. Smaller colleges, especially those serving specific populations, are less likely to be recognized.

This blindspot leads to bias, putting graduates of Historically Black Colleges & Universities (HBCUs), Tribal Colleges, and Women's Colleges at a disadvantage. Additionally, if a firm recruits from a school with low minority enrollment, the algorithm will replicate decisions that mirror similar demographics.

Put bluntly, applicants from underrepresented communities and diverse socioeconomic backgrounds are discarded by algorithmic design. In our research, this is especially true for Black and Hispanic applicants.

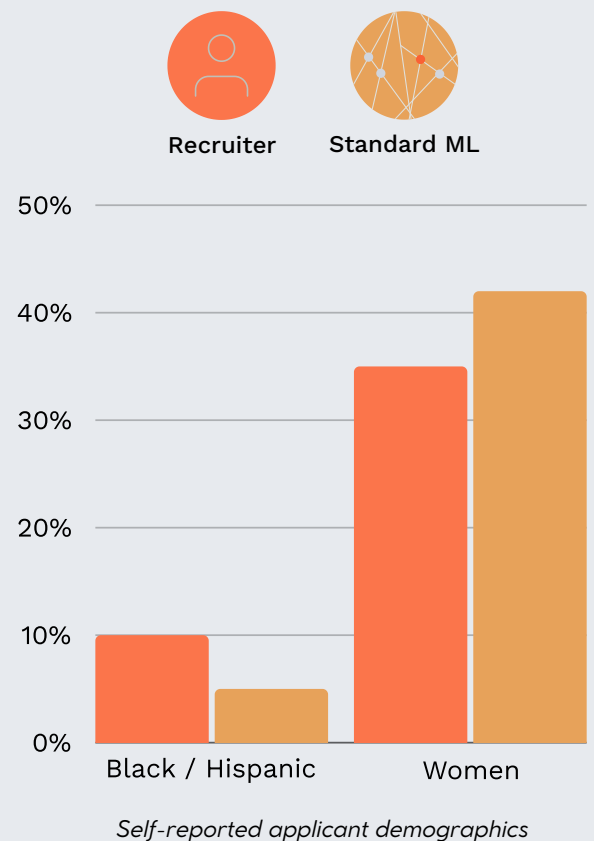
In terms of DEI, an employer might be leaning on HR and recruiters to question the lack of diverse candidates. The common conclusion disregards the impact of screening algorithms, "We just didn't have any qualified applicants from that demographic."

Here's how this plays out in our research:

Recruiters select more Black and Hispanic applicants for interviews than the standard ML algorithm. The reverse is true for female applicants.

Same data pool. Different outputs.

Underrepresented Applicants Selected for Interviews by Recruiters & Standard ML





How We Screen Matters

In the case of our Fortune 500 firm, qualified candidates do exist across demographics, but the way we screen matters. [Spoiler alert: our research also identifies a better way to screen using inclusive algorithmic design.]

Another angle to explore algorithmic shortfalls is more nuanced. Economists, business leaders, and HR professionals often discuss the “Future of Work.” This phrase captures an array of workforce trends, prompted by a shared recognition that how we work, where we work, and who we work with are changing rapidly. Hard skills are changing. Soft skills are increasingly valued. Remote work is expanding. Entire sectors are shifting; along with them, workplaces and workforces evolve.

For hiring to accommodate such innate dynamism, why are we relying on algorithms that propel past hiring cycles into our future?

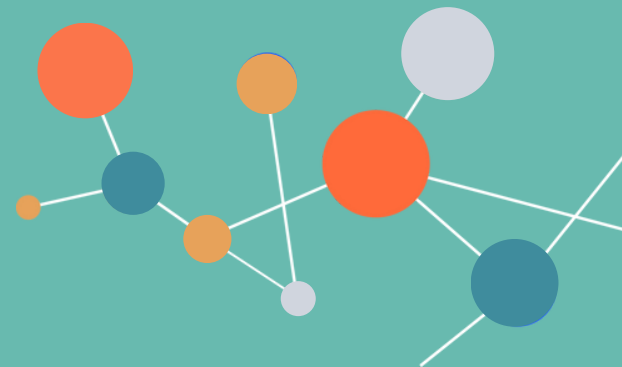
Learning Collider’s research builds on a catalog of studies examining the flaws of standard ML algorithms. The primary differentiator is how our study took the examination a step further. Instead of simply diagnosing, we prescribe a treatment for the dependency on exploiting past decisions.



Not all algorithms are
created equal.

The design of an algorithm
crucially impacts the
consequences of these
decision tools.

Learning Collider





ALGORITHMIC INNOVATION

Learning Collider's innovative treatment leverages exploration, essentially built-in and real-time experimentation. While standard ML exploits what we know from past data and decisions, inclusive ML explores what we don't know by valuing those traits in the model. The resulting innovation is inclusive "exploratory" ML.

> Input Data

- Historical applicant data, e.g., from resumes, serves as the baseline for:
 - What we know - applicants with these characteristics were interviewed and hired in the past; and
 - What we don't know - applicants with these other characteristics have never been interviewed (or hired) so we don't know how they will perform in interviews or on the job.

> Definition of Success

- The highest likelihood to be hired; in this case, our researchers defined success as those most likely to be hired using the Upper Confidence Bound. In other words, the basis of success is past hires who were above the bar.

For end-users, this design may seem very similar to standard "exploitative" ML. But the balance of both exploitation (what we know) and exploration (what we don't know) is critical. Comparing standard ML and inclusive ML, their algorithmic formulae and results vary in important ways.

INCLUSIVE ML & HIRING RATES

Inclusive ML encourages experimentation and risk. When the algorithms don't recognize something in the data, the model tests it out and learns from its performance. An applicant from a small HBCU receives bonus points. Recruiters then learn how that applicant performs in interviews and potentially further down the hiring funnel.

Through the updating feedback loop, inclusive ML learns how each new trait advances - or doesn't. The algorithm also learns how past successful traits perform moving forward. Every cycle, it learns and adjusts its predictions, getting smarter over time while lessening its reliance on the initial baseline training.

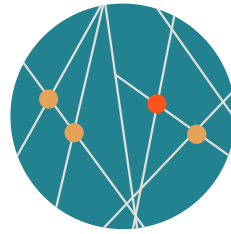
Inclusive ML outperforms recruiters and is arguably just as efficient as standard ML.

To compare outcomes - hiring rates - resulting from each screening method, the research team modeled interview decisions made by recruiters, simulating how each screening method performs in selecting applicants who would ultimately receive an offer. Recall that the firm's hiring rate is 10%, screening by human recruiters.

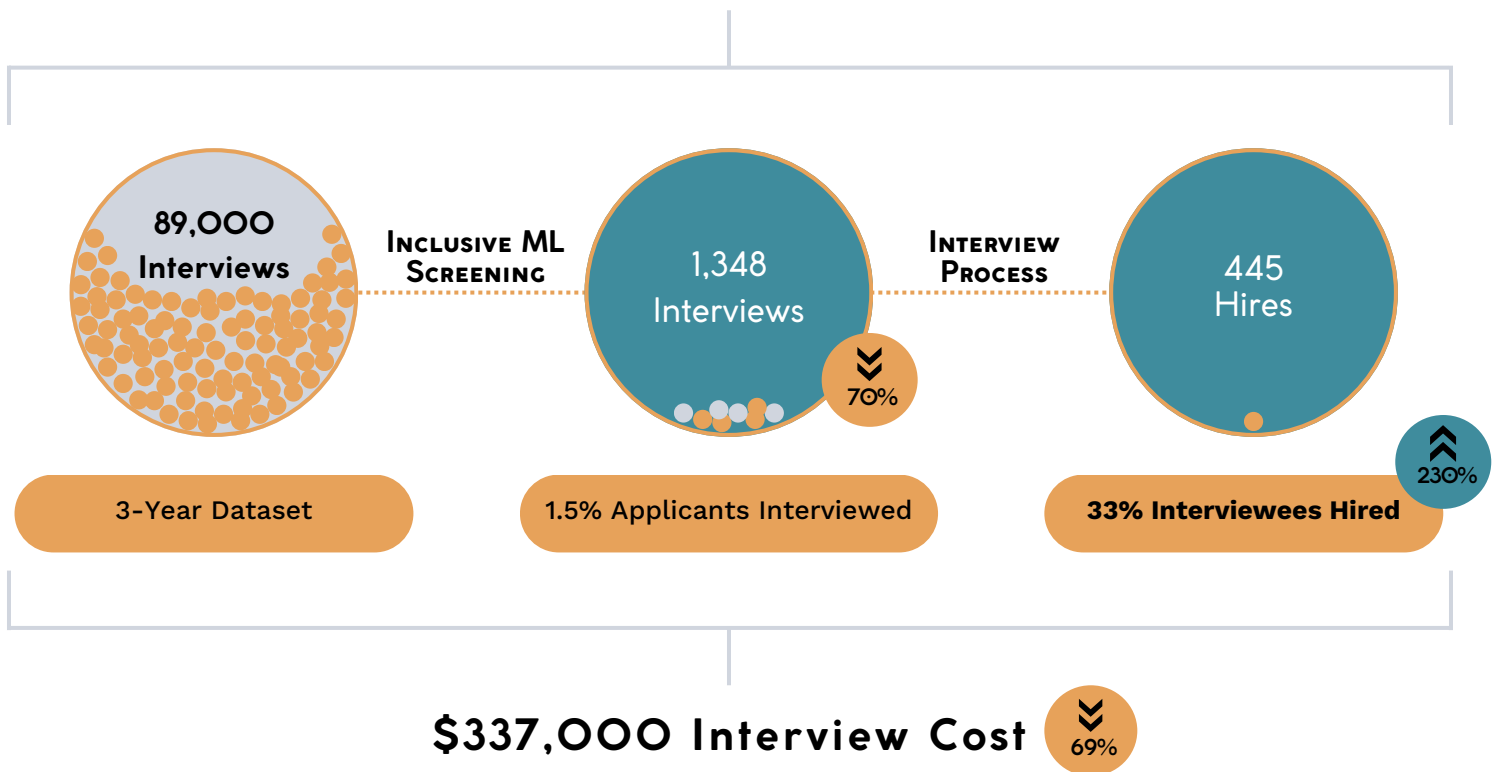
- **Standard ML results in a 24% to 35% hiring rate depending on their feedback loops (static or updating).**
- **Inclusive ML results in a 33% hiring rate. When compared to the mean of the standard ML range (29.5%), inclusive ML takes the lead.**

The Future of Hiring

Inclusive ML developed & tested in *Hiring as Exploration*



Inclusive ML



Assuming a per-interview cost of \$250, Inclusive ML saves \$763,000 in the interviewing process compared to the firm's typical process.



Through exploration and success predicted by past hires who are above the bar. Inclusive ML selects high-quality applicants, leading to fewer interviews per hire.



INCLUSIVE ML & DEI

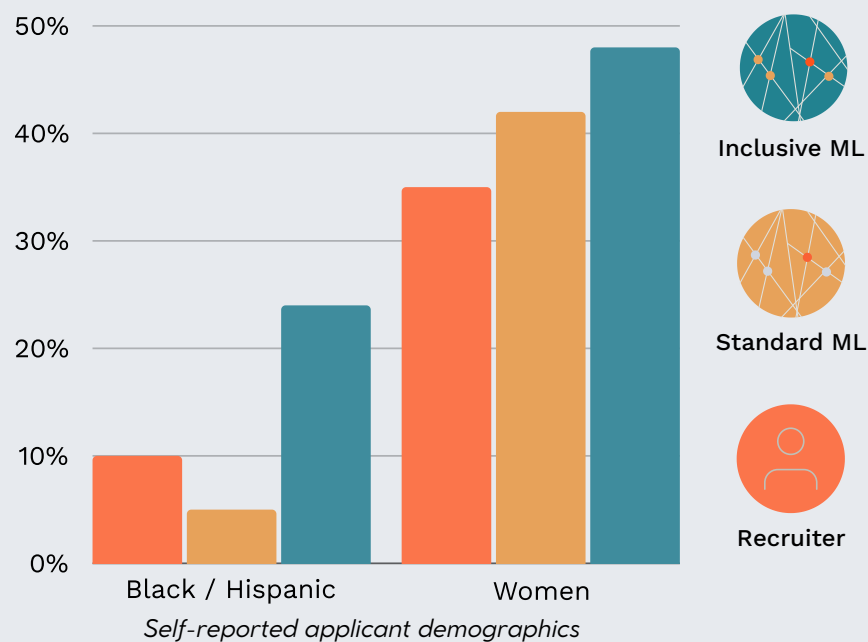
Next, let's dig into inclusive ML's impact on DEI, as measured by the percentage of applicants interviewed from backgrounds with historically low representation in the sector.

Inclusive ML outperforms both standard ML and human recruiters significantly. Particularly for Black and Hispanic applicants, inclusive ML more than doubles access to interviews when compared to human recruiters. Against standard ML, inclusive ML quintuples access to interviews.

How? Why? Inclusive ML is like our research. It is designed to learn. To do so, it values what is new or unknown. It does not explicitly value underrepresented demographics; rather, it explores and learns from all applicant variables.

On all counts, inclusive ML helps humans make smarter decisions as we commit to building dynamic, diverse, and productive workforces. And it does so without compromising efficiency or increasing costs. What it does require from humans is a new way to embrace algorithms - with a willingness to experiment, learn, and diverge from the past.

Underrepresented Applicants Selected for Interviews by All Screening Methods

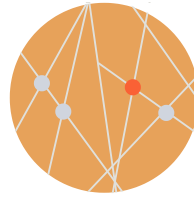


Hiring by Screening Type

Comparisons of Typical and Automated Screening Methods



Recruiter



Standard ML



Inclusive ML

How it Works

Noise
Judgments
Beliefs

Exploitation
Average
predicted success

Exploration
Exploitation
Upper bound
predicted success

Hiring Rate*

10%

24% - 35%

33%

DEI Impact on Interviews*

10% Black/Hispanic
35% Women

5% Black/Hispanic
42% Women

24% Black/Hispanic
48% Women

Estimated Interview Cost

\$1.1m

\$138k - \$464k

\$337k

Estimated interview cost calculated using the Hiring As Exploration results. Assumes a per interview cost of \$250.

**Hiring Rates and DEI Impact on Interviews pertain to the Hiring as Exploration research data.*



CONCLUSION

Demands on HR - that beating heart of business - are constant. As we assess HRTech and test the decisions tasked to algorithms, we observe immediate benefits: efficiency, productivity, and cost-savings.

Technology can indeed help the heart pump. It can also constrict and stress the critical muscle, limiting performance in its core function: hiring.

When we uncover the design and training of applicant screening algorithms, we find they are not inherently better at selecting quality applicants to interview, nor are they inherently unbiased. In reality, standard ML is trained to replicate decisions made by human recruiters in the near or distant past. The result is codifying - and perpetuating - human bias and noisy decision-making.

Within these problems, we find opportunities. **First, how we screen applicants matters. Second, how we design and train our applicant screening technology matters.** On both counts, we can make progress.

Learning Collider's research is a significant step forward in diagnosing problematic hiring algorithms and prescribing a transparent, accessible treatment. The results of the treatment are two-fold: **better hiring outcomes while advancing DEI** in interviews.

The advantages of the prescribed inclusive ML are clear for both employers and applicants. Dynamic, diverse, and productive workforces require better, more transparent HRTech.

So, where do we go from here? Now, we recognize where bias is hidden in hiring. We have the understanding and tools* to innovate our tech stacks, improving hiring outcomes for employers and career opportunities for applicants. Let's make it happen.

***Technical explanations of Learning Collider's inclusive ML are detailed in [Hiring as Exploration](#).**

Website

WWW.LEARNINGCOLLIDER.ORG

Email

INFO@LEARNINGCOLLIDER.ORG

