



BY EUGENE YIGA

The Algorithmic Gatekeepers

Navigating the frontiers of recruitment bias requires humane leadership.

Once upon a time, getting a job was a lot simpler. Candidates would send a CV to a recruiter or the HR department of a prospective company, an actual human being would read it, and the candidate would either be shortlisted for an interview or turned down. Nowadays, with hundreds of people sometimes applying for a single job, it's not feasible to process the applications manually. Unfortunately, despite promises to streamline the process in a way that would be fair to everyone, recruitment systems seem to have become worse.

“As a computer scientist with a focus on data science and a deep involvement in algorithmic discrimination research, I’ve come to lead a significant project on algorithmic hiring,” says Carlos Castillo, a researcher on algorithmic fairness and crisis informatics at the Pompeu Fabra University in Barcelona. “In the rapidly changing landscape of employment, where AI-driven recruitment tools are increasingly prevalent, we’re witnessing a shift where algorithms, not humans, are the first to review job applications. This transformation, driven by the sheer volume of applications – a jump from around 100 per job in 2010 to over 250 today – highlights the urgency and complexity of addressing biases within these systems.”

Documenting discrimination across AI

In exploring algorithmic discrimination, it’s clear that these biases are not isolated incidents but are pervasive across various sectors where AI is applied. Notably, the notorious instance of an Amazon recruitment tool that downgraded female applicants is a testament to the widespread nature of these biases. This tool, like many others, wasn’t designed with malice but ended up reproducing societal biases in a highly consequential setting.

“Through our research, including interviews with migrants, a recurring theme is the scepticism around the neutrality of AI hiring tools,” Castillo says. “One poignant testimony shared during our research underscores this sentiment, noting that if most hires share similar backgrounds, it’s indicative of inherent biases in the system, contradicting the supposed neutrality of these tools. This revelation often comes to light in subtle ways, such as receiving rejection notices at times when no human HR professional would be working, leading applicants to rightly suspect they were assessed by an algorithm.”

Regulatory landscape and challenges

The new EU AI Act categorises recruitment tools as high-risk applications, thus necessitating stricter oversight compared to less impactful AI applications. Castillo believes that this regulatory approach is essential in addressing the nuanced challenges of ensuring fairness in AI-driven employment, which was a sector already under scrutiny for discrimination long before the advent of AI.

Algorithms, not humans, are the first to review job applications.

“Addressing biases in AI is not merely a technological problem but a socio-technical challenge, where understanding and intervention must transcend computational solutions,” Castillo says. “My colleagues and I often grapple with operationalising fairness in a way that balances mathematical precision with the complex realities of societal biases. Moreover, HR departments, overwhelmed by the volume of applications, might unwittingly exacerbate the issue by adopting tools that indiscriminately filter out candidates, reducing their workload but potentially sidelining perfectly qualified individuals.”

Synthetic data and quality monitoring

One of the innovative approaches Castillo is pursuing is the creation of synthetic CV databases to better understand and mitigate biases. This initiative involves collecting CVs to analyse the statistical relationships between demographic profiles and the phrasing within CVs. By donating their CVs, individuals can contribute to a dataset that helps refine the understanding of how demographic factors influence job application success, thus informing the development of fairer AI systems.

A recurring theme is the scepticism around the neutrality of AI hiring tools.

“While the path forward is full of challenges, it is the collaborative effort between computer scientists, policymakers, and the public that will pave the way for more equitable and transparent AI-driven recruitment processes,” Castillo says. “I look forward to discussing these issues further and exploring potential solutions to ensure that AI serves as a tool for inclusivity rather than exclusion in the job market.”

For more information: <https://findhr.eu/>

RACE BY PROXY

In their comprehensive analysis, Thao Phan from Monash University and Scott Wark from the University of Kent explore the complexities of how race and racialisation operate within the realm of AI and data-driven systems. Their research, titled *Race by Proxy*, explores the intricate role of proxies in algorithmic culture.

“Proxies in AI and data science serve as substitutes or surrogates for variables that are challenging or impossible to measure directly,” Phan says. “For example, recommender systems use observable behaviours, such as clicks and views, as proxies for consumer desires, which are inherently complex and elusive. This substitution allows for computational actions but often fails to capture the true nature of human wants, leading to potential misrepresentations and biases.”

Manifestations of racial discrimination

Phan and Wark argue that while modern systems aim to neutralise visible markers of difference, they inadvertently perpetuate existing racial legacies. This occurs through proxy discrimination, where indirect discrimination materialises via correlated datasets rather than direct discriminatory actions. For instance, using zip codes as a proxy for race or income as a proxy for gender can lead to biased outcomes in decision-making processes, such as credit eligibility or employment.

“Our study critically examines the misguided assumption that increasing transparency and accountability in AI

systems can straightforwardly ‘fix’ inherent biases,” Wark says. “Instead, we point out that these technologies, by their predictive and classificatory nature, use proxies that embed and amplify societal inequalities. This, we argue, calls for a deeper understanding of how AI practices, under the guise of neutrality, continue to foster environments where racism can thrive in new forms.”

Technical and ethical implications

Phan and Wark highlight the irony that systems designed to mitigate harms such as racism actually extend or entrench these issues through their operational mechanisms. They discuss the limitations of current methods aimed at addressing proxy discrimination, such as the development of synthetic data. These techniques themselves rely on proxy logics, indicating a recursive problem where the tools intended to observe and measure biases are also entrenched in them.

Ultimately, *Race by Proxy* serves as a call to critically examine the deeper layers of how race is enacted and transformed within digital and algorithmic systems. It challenges the prevailing views that technological solutions can simply erase deep-seated social issues such as racism. Instead, the researchers advocate for a nuanced approach that recognises the complexities and inherent limitations of proxies in AI, urging us to re-evaluate how these systems are designed and implemented so that we can genuinely address racial injustices.

FAIR AI DECISION-MAKING UNVEILED

As a researcher in fair and responsible artificial intelligence immersed in his doctoral studies at the European Laboratory for Learning and Intelligent Systems (ELLIS) Alicante Foundation, Adrián Arnaiz offers insight into a key question: how exactly does bias and discrimination occur?

Through his research and public discourse, Arnaiz strives to provoke thoughtful discussion on refining AI processes to foster fairness, thus shaping a more equitable digital and social infrastructure. In a recent presentation, he explained the complexity of bias in artificial intelligence, particularly in recruitment systems. His insights explore the 'classical pipeline' of machine learning to shed light on the critical stages where biases may infiltrate and perpetuate discrimination.

Data:

At the data-collection stage, biases can occur when the datasets used to train AI systems are not representative of the target population, or when they inadvertently contain discriminatory biases. Arnaiz discusses the potential of modifying data to mitigate these biases, mentioning techniques such as synthetic data generation, undersampling, and oversampling. He also highlights the importance of 'data valuation', a method to assess how much each data point contributes to the model's behaviour, all with the aim of enhancing fairness in algorithmic decision-making.

Model:

The model-building phase is essential for embedding fairness. That's why Arnaiz emphasises the need for mathematical restrictions that not only optimise the model's accuracy but also integrate fairness directly into the algorithm's design. This approach helps detect and prevent bias during the model's training phase. He also points out the relevance of considering relational data in social network contexts, where decisions about individuals are influenced by their social connections, thus propagating societal biases.

Decision:

Finally, the decision-making process itself can introduce biases based on how model outputs are interpreted and used in real-world scenarios. Arnaiz discusses the role of threshold settings, which determine how model probabilities are translated into actionable decisions. Adjusting these thresholds can help counteract biases in final outcomes. Furthermore, he raises concerns about the long-term social impacts and the dynamics of gamification that arise when AI systems are deployed, emphasising the interaction between technical systems and societal structures.



Khyati Sundaram

BREAKING BIAS

For organisations looking to modernise their hiring practices and promote a more inclusive and fair recruitment environment, Applied is a platform that integrates behavioural science to help reduce bias. It's designed to improve the quality of hires and increase diversity within companies.

"Our platform removes opportunities for bias to creep into hiring decisions through anonymous applications, skills tests, and a 'blind', peer-review scoring system," says CEO Khyati Sundaram. "We're not interested in candidates' backgrounds or what they've done in the past. We're interested in finding out what candidates are truly capable of based on their role-relevant, transferable skills – the most accurate indicators of performance we have."

Applied has found that focusing on candidates' skills – rather than their names, previous experience, or other arbitrary proxies – gives everyone a fair chance to succeed. As a result, the companies it works with see minority ethnic hires increase by up to 300%, and the number of women hired into senior roles increase by up to 70%.

"Improved diversity isn't the only benefit," Sundaram says. "Since this approach ensures that hires have the skills needed to succeed in their new roles, retention rates improve from 83% to 93% on average. And since tokenism is taken out of the equation and hires are based on merit, companies are able to build more inclusive cultures. Plus, candidates have a better experience knowing that they're being given a fair chance."

Book a free demo at
www.beapplied.com GIBS