

Ensuring Equitable Outcomes from Automated Hiring Processes

The [ODD](#) (Optimizing Diversity with Disability) initiative is a part of the IDRC's [WeCount](#) project. A central aim of ODD is to investigate bias in widely deployed hiring algorithms and to suggest more inclusive alternatives.

Automated algorithms supporting the infrastructure of society

These automated hiring and matching algorithms, implemented by major corporations such as LinkedIn, Amazon, and others can be positioned in the wider context of automated processes, that use machine learning/AI algorithms, and support the infrastructure of society. These systems inevitably result in inequitable outcomes, as largely unintended consequences both of:

- the algorithmic structure of their optimisation processes, which naturally steer towards central regions of parameter spaces, and
- their embedding in traditional inference workflows inherited from the scientific tradition, which assume that beneficial future outcomes must necessarily share the same structure as those it is inferred have been seen in the past.

It's also arguable there is a further source of bias since the system is deployed in the context of capitalistic utility functions that tend to favour accruing benefits privately, socialising losses. This results in bias since the community facing the losses (the public at large) is inevitably more diverse and less legible than the community accruing the benefits (the corporation deploying the system). There are numerous other sources of bias, both conscious, and unconscious, which are widely described elsewhere.

Risks of Automated Hiring Systems

The above considerations are common to the deployment of any kind of machine learning systems as part of the infrastructure of society. Turning to hiring systems more specifically, it is becoming widely recognised that, as recorded in a [recent Harvard Business School report](#), the architecture of these hiring systems is leading to inequitable outcomes, some of which can be accounted for by the generic considerations above. Workers who do not fit the algorithm's self-reinforcing profile of those who are appropriate for work are being excluded from participation. Inevitably, these are a more diverse, heterogeneous group than those who fit the profile. Let us imagine that those impacted will want to mount an investigation of one of these systems, so that they can better understand who is being marginalised and how the system can be improved to be more inclusive.

These are some responses arguing against such an investigation, and offering an evaluation of their risks and costs:

- The corporation deploying the hiring system might argue that since it employs those with arguably the highest expertise in the operation of such systems, it should be permitted to self-regulate and assure the fairness of these systems internally. This argument is weak since the fairness of the system would be much better assured by the involvement of independent, acknowledged experts in the field who have no vested interests entailed with the corporation's profitability. These are the professional standards we take for granted in all other parts of society's infrastructure — would you accept an accountant who refused to be audited, or a civil engineer who refused to accept safety inspections from an independent body?
- The corporation may argue that since there is no precedent for the system, it can only be assessed in terms of its operation in the field operating using real data. This is a shaky argument on many grounds, including that large-scale production software installations are never deployed without at least some form of validation and monitoring workflows, at least for internal use. The question is for: For which audience do these workflows operate, which aspects of the system's function do they assess, and who is entitled to inspect, interpret and design their outputs? Given that, once the system is running, it is having a large societal impact (thousands of people are gaining and losing their livelihoods) it is arguably too late to avoid these harms if we have no means of assessing them up front. The [precautionary principle](#), increasingly enshrined in law in many jurisdictions, implies that we do our best to foresee, plan for and prevent irreversible damage resulting from systems that we deploy.
- A further risk of only permitting evaluation of the system as it runs is forcing the disclosure of citizens' private data in order to operate the evaluation. For example, in the context of an automated hiring system, the corporation may propose that system users are invited to "self-identify" — for example, to declare what disabilities they have — as data for the private use of the corporation to run its internal audit of unconscious/unintended hiring bias. The harmful privacy implications of this disclosure are obvious. There are numerous instances where contract for use of such data was deemed unclear, misinterpreted or the data was subject to an unintended breach. It is far preferable, and precautionary, that the system operate by publicly understandable considerations and algorithms which can be checked by audit in advance of harms occurring.
- Responding to the problem of a lack of inclusion by requiring greater disclosure from the community (even with promises that this disclosure will remain private to the corporation) represents a harm in itself. Data's use grows and changes in unexpected ways over time, and once it has been used to "schematise" a population at risk, harm has already done — what happens to identity when it is so deeply schematized as it must be to be operated on algorithmically? And further, how much more harm can result from the use of this disclosure of schematized information when it is formalised in an automated system whose effects can't be accounted for, except in use?

All of these considerations tend towards requiring massively greater transparency in such systems — in accordance with WeCount's aims towards creating inclusive data ecosystems. Systems should not only be much more accountable for the data which they require from citizens, but also for the use they make of such data. It should be possible to trace the use of one's disclosed data, the purposes to which it has been put and, if necessary, to revoke its use by the corporation. In the particular context of hiring systems, it should be possible to discover how personal data has been schematized by the system, as well as to trace how differently schematized individuals will be treated differently by the system. Without this, it is extremely likely that inequitable and harmful outcomes will result, often in ways that are opaque to all of those involved.

The Lifecycle Leading to Society's Acceptance of Technology

We accept the loose standards applied to the deployment of machine learning (and more widely, software engineering) systems because they are part of an immature industry whose nature and impacts on society are still poorly understood. It is clear from the history of many previous such immature industries that it is inevitable that society gains the expertise, the will, and the capability to effectively oversee and regulate these industries. The open questions are: What is the trajectory of society establishing this oversight, how quickly does it proceed, and who acts to accelerate or impede the process?

A example that is very familiar from the history of technology is the evolution of safety regulations governing the [operation of steamboats in the 19th century in the US](#). In 1817, the boilers of these steamboats were a cutting-edge technology, with strong incentives to compromise on safety standards by fabricating them with dangerously thin steel, running them overpressure and above safe speeds. The steamboat owners insisted that it would be impossible to comply with such onerous safety standards and still run their businesses at a profit. They effectively lobbied for the prevention of regulations, and when regulations were first enacted in 1838, supervised their operation internally rather than participating in external audits by disinterested inspectors. This argumentation continued for decades, resulting in thousands of deaths, until effective legislation was finally introduced to regulate the industry in 1852 — from then onwards it was found perfectly possible to run the industry both profitably and safely.

For those involved in these processes, the question to be asked is: What is their part in it to be? Should they be those acting to bring in effective oversight of such innovative technologies by competent, independent authorities in order to ensure fair outcomes for society as quickly as possible, or should they be those acting to ensure that profits can be extracted from the unsafe, inequitable practices as long as possible, perhaps stalling them by the promise of self-regulation?

Activities for the ODD Project

As we begin conversations with ODD project stakeholders, we intend to gather answers to all these questions, and gain insights into further important qualities for those who deploy automated hiring processes, in their aim to avoid damaging consequences to their clients. We will assemble datasets from both real and synthetic profiles of postings and individuals in order to suggest improvements to these hiring algorithms in line with "[fair machine learning](#)" metrics, as well as suggesting wholesale alternatives to these algorithms based on recent research on exploratory, rather than optimizing approaches.