Trend Topic: **AI + Skilling**

What can businesses learn from the use of AI in the criminal justice system

By

Dr. Kameel Khan

Founder, Project ReMake,
Visiting Scholar & DCI
Fellow, Stanford University
(2017-2019), Judge (Ret.), UK

Imagine that you've spent the last five years in prison, and now you're up for parole.

The parole board will use two sources of information to decide if you should be released. One is your lawyer. The other is a detailed dossier of information, including reports and records, to assess your behaviour and suitability for release. This dossier is compiled by HMPPS (prison & probation) and includes reports from various staff, including prison officer and governors. Sometimes it's accurate, and sometimes it's not. Sometimes it's filled with hearsay and gossip inserted by third parties.

Now, do you think the parole board would make a better decision if they had an algorithm to help them out?

That algorithm would be designed to predict the future — specifically, your future behavior, and your chances of committing a parole violation if released.

Since 2001, the UK probation services have used a largely opaque algorithmic system, the Offender Assessment System (OASYS), to help predict the likelihood of re-offending and guide decisions of sentencing, parole and rehabilitation. Normally these tools increase the accuracy of decisions. But this system has been criticized as having discrepancies in accuracy based on gender, age and ethnicity.   Traditionally, parole boards make the right decision 50% - 62% of the time. The experience in the US is that algorithms reach the right decision over 80% of the time. While not all algorithms are fair, equal or accurate (it depends on the quality of the data), good algorithms can have a big impact on incarcerated people, on society, and on government budgets and efficiency. It can enable more people to be safely released without impacting public safety. It can also enable governments to better allocate staff and budget. Instead of monitoring people who don't need to be monitored, they can do more to help low-risk people reintegrate into society and re-think their approach to high-risk people.

The use of AI by parole boards gives us a window into how AI can help improve decision-making more widely, and nudge us in arriving at better outcomes than we could achieve on our own. If we can build an algorithm that provides valuable input to a life-changing decision such as parole, we should be able to leverage this same technology to also help businesses make better data-informed decisions. By "better" I mean more accurate, less biased, and administered more quickly and efficiently. In particular, I see three potential benefits for businesses based on how the criminal justice system has leveraged AI and they are:

> Traditionally, parole boards make the right decision 50% - 62% of the time. The experience in the US is that algorithms reach the right decision over 80% of the time.

- Becoming more proactive and less reactive. We can use AI to look for things that could go wrong, to flag anomalies and create early warning systems. That can help us be more agile and more aware, and can give organizations valuable breathing room to plan, strategize, and respond thoughtfully before problems balloon.

- Reducing institutional bias. Researchers from the University of Cambridge have found that algorithms that are designed to promote a more diverse workforce are akin to pseudoscience. But this over simplifies a complex and evolving field. While it is true that some algorithmic approaches may be poorly designed or lack rigorous validity, this does not invalidate the entire domain of algorithmic fairness or diversity enhancing technologies. I still believe it is possible to build a hiring system that is robust in detecting bias, even if it is yet to be accomplished.

- Achieving better governance. If leaders integrate AI into risk management, and perhaps make it available as a tool for their risk committees, we could encourage more responsible decision making.

At critical mass, better decision-making should build organizational resilience, improve efficiency, and support more successful innovation. In an age of heightened geopolitical conflict, climate change, and great economic uncertainty, organizations of all persuasions should be considering AI as a powerful tool to improve our cognitive abilities and the potential for our futures. Of course, no tool can do this on its own – we need the skills to use the tool properly, interpret its findings, and implement them wisely.

So what will it take to develop and trust AI to look into the future for us? There are actions we must build into our algorithms, and others that we must foster in our workforces.

We must develop new skills, individually and organizationally

A forward-looking AI has the potential to provide us with information, that, organizationally, we're not prepared for. If you know that there is a chance of failure in one specific part of your supply chain, how do you use that information to produce

better organizational resilience? Do you have the right people, with the right skills, using the right processes to make data-driven decisions?

Questions such as these go well beyond the technical skills that dominate so many debates on the skills gap. Technical skills are relatively straight-forward to identify, if difficult to hire for: facility with general-purpose programming languages such as python, for example, strong mathematical and statistical foundations, machine learning, data analysis and visualization.

But we also need people who are experts at interpreting data, at critical thinking, and at solving problems and communicating. Institutions and organizations must prioritize the development of these skills, even when they don't neatly fit in the category of technical skills that organizations so often seek out. Without these interpretative, problem-solving, and communication skills, we can't design and implement efficient and ethical AI systems that can serve as accurate guides to the real world – or the almost-real future.

We must learn to identify and remove embedded bias

Embedded bias can be pernicious and difficult to detect. If a hiring algorithm is looking to match a profile of successful candidates within an organization, and the leadership of that organization is overwhelmingly male, the algorithm will filter out female candidates. Unless specifically steered otherwise, the algorithm will be looking backward in an attempt to replicate the status quo. For algorithms to be trusted they must be fair and rely on accurate information.

Algorithms used in parole board decision-making are considered successful and trustworthy in part because the developers have worked to eliminate embedded bias. These tools process a range of inputs - sometimes just a few variables, other times more than 100 - to assign a risk score based on things such as a likelihood of arrest or failure to appear in court. This is the way accuracy is created – however some data isn't considered because it's viewed as discriminatory: a person's race and gender, for example. Data about the number of times someone has been stopped by the police is also off-limits because that information may reflect police behavior more than the behavior of the incarcerated person. Zip codes don't factor in, either, because they may also include racial bias. The more fairness you require, the less accurate the prediction.

> For algorithms to be trusted they must be fair and rely on accurate information.
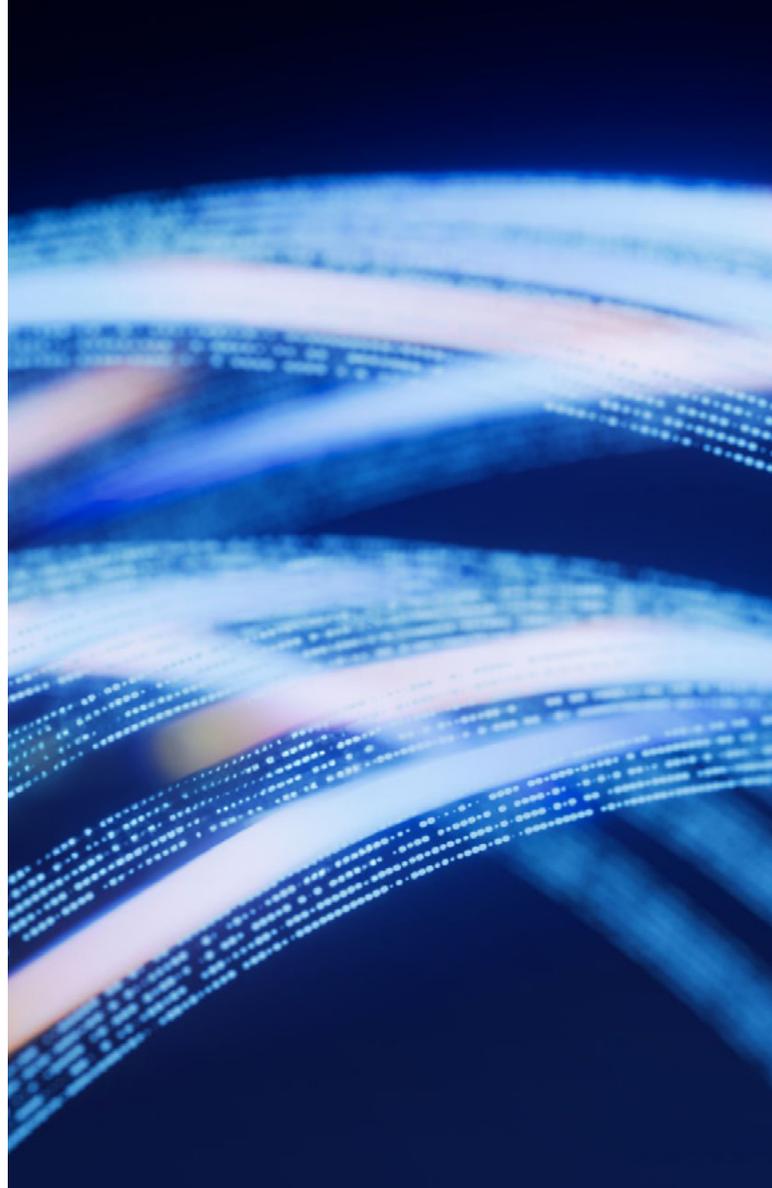
We must interrogate the model, its inputs, and outputs

In many jurisdictions, we still face troubling ethical issues around the use of algorithms that may not have sufficiently mitigated embedded bias and do not provide transparency around their inputs or their decision making. In the well known 2013 case, Eric Loomis was sentenced to six years in jail, based in part on a risk assessment performed by an algorithm. He sued to find out why the algorithm determined that he deserved that sentence. He lost, on the grounds that the algorithm was the intellectual property of the company that had produced it. This is exactly the type of outcome we need to guard against if we want to use AI to make decisions that are both efficient and fair.

The alternative is to become skilled at interrogating these models on a regular basis, rather than resigning ourselves to becoming passive users. Is the model asking the right questions? Does it have the right data? Have we excluded the right data?

We must advocate for regulation

Many business leaders are allergic to regulation, but reasonable regulation provides the guardrails that businesses need to be credible to their customers and other constituents. In the U.S., there is currently no version of the U.S. Food and Drug Administration to determine the safety or efficacy of an algorithm. While algorithmic decisions are frequently referred to as black boxes, we should insist on some level of transparency. That means the ability to understand how models make decisions, how specific results are produced, and what data is used as inputs and for training. We should also be able to audit the outputs to ensure they are fair and accurate.

Not everything can be regulated. Algorithms predict outcomes and the predicted accuracy of an algorithm is reduced if you eliminate gender, race and other differentials in your results. There will always be bias but perhaps less than if a human made the decision. There would certainly be more transparency and accountability.

...we need to be relentless in ensuring the intelligence we build is capable of steering us toward a better world.

To reap the benefits of a future-looking AI, we need to be relentless in ensuring the intelligence we build is capable of steering us toward a better world. Removing embedded bias is a critical step in this direction, as is insisting on transparency in our algorithms, their inputs, and their training.

To properly vet and manage these increasingly powerful algorithms, we need to advocate for responsible regulation and make sure that the problem-solving and critical thinking skills are just as valued as the more easily defined technical skills. By working together to ensure that AI is as trustworthy and accurate as possible, we can gain a powerful new tool in our attempts to make sense of, and respond to, a fast-changing and uncertain world.