MARION OSWALD

'GIVE ME A PING, VASILI. ONE PING ONLY'

WHY THE SUCCESS OF MACHINE LEARNING DEPENDS ON EMPOWERED PEOPLE

While we're seeing some promising developments in the introduction of machine learning and data science methods to support law enforcement risk assessments, we shouldn't assume our technology is answering the question we need to answer.

The quote in the title is from a 1990 Cold War film 'The Hunt for Red October.' Sean Connery plays Captain Marko Ramius, commander of the Soviet Union's newest submarine, which is fitted with an innovative propulsion system undetectable to passive sonar. As Captain Ramius and his officers want to defect to the United States, the story features a race between American and Soviet submarines to detect the Red October. The Americans need to make contact with it before the Russians find and sink it. Captain Ramius's famous '*Give me a ping, Vasili*' comes as the talented sonar officer Jonesy attempts to track the Red October using his underwater acoustics software.

Jonesy does not take the output of the software at face value. He knew they did not originally build it for tracking nuclear submarines but for detecting seismic anomalies. He used this knowledge to interpret the result in the complex situation and was supported by his commander because he was able to explain and justify his findings.

One moral of this story, which applies to today's preoccupation with data analytics, machine learning, and AI, is: **Don't assume your technology is answering the question you need to answer.**

To uphold this moral, we need to understand what AI tools are doing and the immediate and longer-term consequences of using them within our decision-making processes. Epstein argues that:

> 'In a truly open-world problem devoid of rigid rules and reams of perfect historical data, AI has been disastrous… When narrow specialization is combined with an unkind domain, the human tendency to rely on experience of familiar patterns can backfire horribly.'
>
> (Epstein, 2019)

Perhaps this might seem a touch harsh when set against claims made for some predictive techniques and diagnostic tools that appear almost daily. Here, I unpack what such predictive use cases are really doing.

WHAT ARE 'PREDICTIVE' USE CASES REALLY DOING?

Much is written about predictive risk assessments using machine learning methods, often based around random forest decision trees. But are they really predicting or risk assessing anything? They use group data from the past to make a prediction about an individual's future. It's more accurate to say they are categorising or comparing by comparison with certain characteristics of a specified group in the past and these characteristics will only be those that can be translated into a datapoint or a numeric scale. The question these tools are really answering is how do the characteristics of an individual (which can be translated into a datapoint or a numeric scale) compare to the characteristics (as translated into datapoints) of a specified group of people in the past.

If we want to understand and evaluate a tool, we need to know details like: what input data is being used and how has it been translated into datapoints? Are these data relevant to the question I need to answer? What is the analysis doing with these datapoints? And what uncertainties and provisos are attached to the analysis?

We know that in many public sector contexts, recorded data can be partial, entered in different formats, out of date or missing. For example, a BBC report on Greater Manchester Police's Integrated Operational Policing System quoted one serving officer's concerns that 'there's a black hole where the recent intelligence should be.' If machine learning methods are implemented without a deep understanding of the underlying data, the impact of errors and missing information could be both amplified and hidden from the user.

PREDICTIVE USE CASES — DO THEY 'WORK'?

Law enforcement is increasingly expected to adopt a preventative, rather than reactive posture, with greater emphasis on anticipating potential harm before it occurs, identifying vulnerable individuals in need of safeguarding, and targeting interventions towards the highest-risk persistent and prolific offenders. Actuarial risk assessments have been used for many years to support such a preventative approach; what's new is the introduction of machine learning and data science methods to produce the algorithm and ever-increasing types and volumes of datasets.

> **High accuracy rates at the group level can often conceal very low accuracy rates for specific individuals or groups of individuals within that larger group.**

High accuracy rates at the group level can often conceal very low accuracy rates for specific individuals or groups of individuals within that larger group. All individual predictions are associated with a confidence interval (a margin of error), which is often not taken into account when reporting the overall predictive accuracy of the tool (Babuta and Oswald, 2019).

To quote one of my favourite fictional detectives:

> 'While the individual man is an insoluble puzzle, in the aggregate he becomes a mathematical certainty. You can, for example, never foretell what any one man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant.'
>
> (Arthur Conan Doyle, 1890)

This article is a version of a keynote address given to the RUSI-Trustworthy Autonomous Systems conference 'Trusting Machines? Cross-sector lessons from healthcare and security' on 1 July 2021.

> What question can the tool contribute to answering and is this the question we need to answer?

Examples in health also illustrate the importance of validation and contextualisation of the AI output by an expert human. While AI-supported breast cancer risk prediction has produced promising results, researchers have highlighted the need for improvements based on additional clinical risk factors, closer consideration of strategic screening aims (early detection, reduction of false positives and so on) and validation on diverse patient populations and clinical environments. What question can the tool contribute to answering, and is this the question we need to answer?

An evaluation of a sepsis detection algorithm by academics at the University of Michigan claims that the particular tool has poor predictive value despite its widespread adoption in clinical settings. The research suggests that the tool does not catch patients at an earlier stage of sepsis (which is when you would want to catch them from a clinical point of view) and therefore does not do what its manufacturers state that it does (Wong et al., 2021).

DECISION-MAKING AUTHORITY

There's another reason why we should ask **whether the technology is answering the question we need to answer.**

Decisions in national security, policing and health are subject to important legal tests, including those set out in the human rights framework and in specific laws governing coercive or intrusive powers. There is a risk of relinquishing decision-making authority if we conflate algorithmic outputs with the answer to a legal test (Oswald, 2018).

Let's take the requirement for 'reasonable grounds for suspicion' to justify the exercise of police powers. According to Code A pursuant to the Police and Criminal Evidence Act 1984, 'generalisations or stereotypical images that certain groups or categories of people are more likely to be involved in criminal activity' cannot support reasonable suspicion. Probabilistic outputs based on reference class may not satisfy the requirement for reasonable grounds, as they fall within the exclusions of generalisations, category-based suspicion, and suspicion based on general association.

We've seen that algorithmic predictions effectively compare an individual against datapoints from a group in the past, and so are likely to be seen as equivalent to suspicion based on general association, as set out in code A of PACE.

All this is not to say that data analytics have no place in national security, policing, and healthcare — far from it (Oswald, 2020). We're seeing some very promising methods being developed to join the dots between different pieces of information to suggest connections between those involved in organised crime or previously unidentified crimes of modern slavery. In national security and policing terms, such analysis is a form of intelligence and therefore should be assessed and handled as such, with its potential uncertainties appreciated.

THE NAMING OF ALGORITHMS

As noted above, an algorithm might predict an average behaviour, but for an individual (especially when the algorithm's output could be used to 'do something' in the real world that might affect that individual's rights), badging something as predictive is potentially misleading and risks creating over-reliance. We should name these algorithms in a way that accurately describes what they do in a more specific and circumspect way, e.g., as an 'Organised Crime Group Association Suggester' or 'Public Order Deployment Suggester'.

RECOMMENDATIONS

I conclude by returning to Jonesy and Commander Mancuso and expanding on the recommendations that flow from their stories. We should:

- Ask what the tool was built to do.
- Ask what the tool is really telling us — question the headline.
- Ask what the tool is NOT telling us and what is missing or uncertain.
- Ask whether the output of the tool is relevant to the decision that needs to be taken.

Mancuso and his reaction to Jonesy is equally important in this story as it tells us the following about AI and empowered people:

- Operators and managers need appropriate training, knowledge and skills to understand AI tools.
- Skilled operators need discretion to decide how, if at all, to use the output, provided that they can justify their decision, and management should be supportive of the exercise of skilled discretion.
- Management should take a critical approach both to how AI works and the purposes for which it is proposed to be used.

Dr Marion Oswald works at Northumbria University and the Alan Turing Institute. Her research focuses on law, ethics, technology, policing and national security. She sits on the Advisory Board of the Centre for Data Ethics and Innovation.