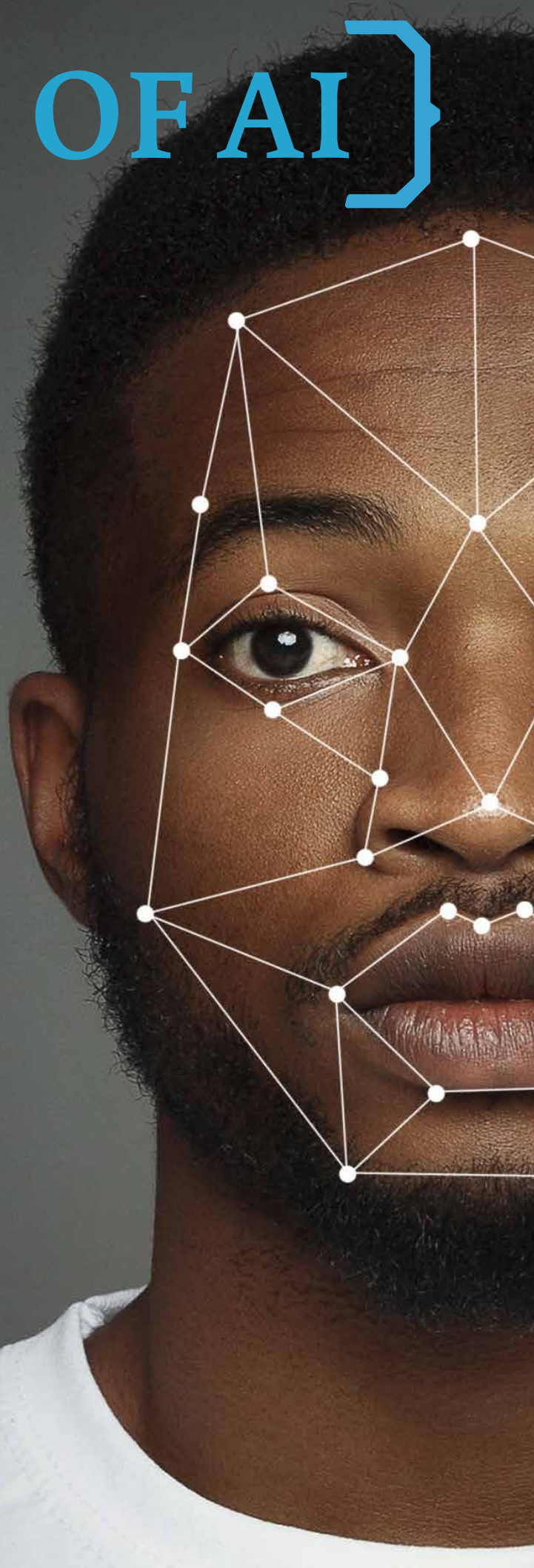
ETHICS OF AI

AI will raise some of the toughest ethical dilemmas companies have ever faced.

How can CDAOs work with boards and senior executives to reduce risk?



Corinium
connected thinking

Why did we make that decision?

We're just at the starting line of AI and already CDAOs are grappling with ethical dilemmas that create risk at all organisational levels. Poor AI governance can lead to significant reputational, market and financial risk. CDAOs will need to educate and work with boards and senior executives to develop new governance methods in a rapidly expanding risk universe.

Who do we trust?

Despite overwhelming evidence that AIs can make faster and more accurate decisions than humans, people still trust humans more than machines. According to AI expert, Guarav Ahuja, who currently leads Digital Analytics and Data Science for a large Australian discount department store retailer, "When an AI makes a decision, those impacted don't just want to know the 'what' – they want to know the 'why." And this is opening up corporate decision-making to a whole new level of scrutiny.

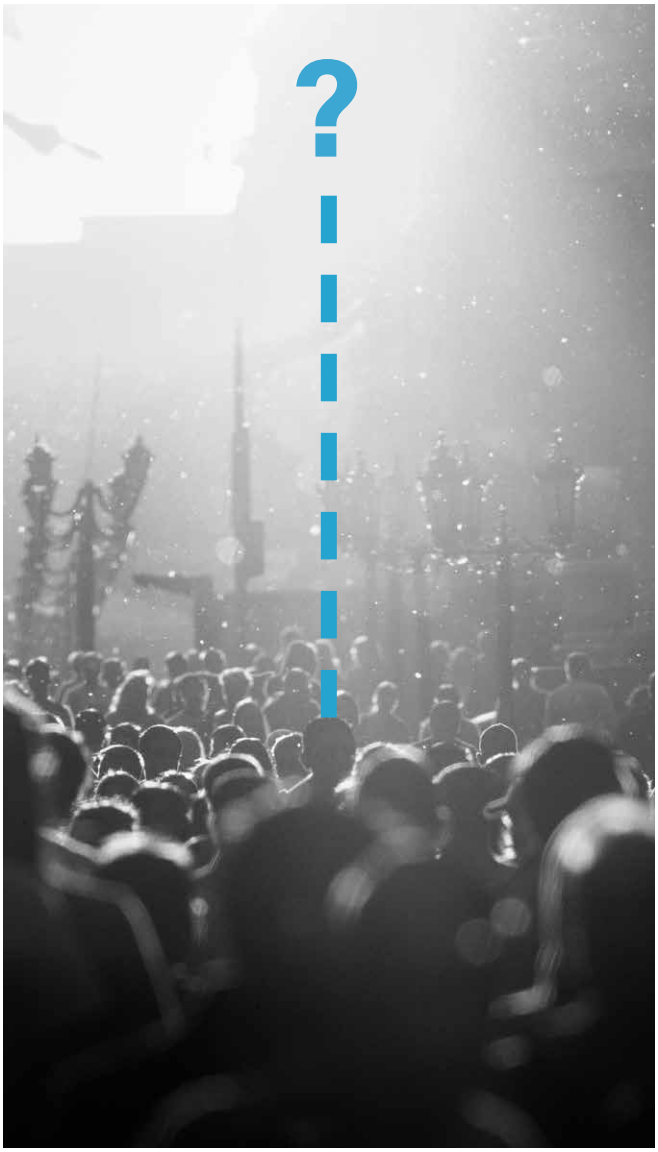
For more than a century, business people have been making decisions that affect human lives: deciding on safety features or evacuation procedures; approving (or rejecting) credit; or choosing job candidates. But, unless something went drastically wrong, these decisions were seldom questioned – even though they were made by inherently biased humans and based on incomplete or inaccurate data, gut instinct and 'finger-in-the-air' guesses.

Outside the stressful conditions of a Royal Commission, we rarely asked businesses: "Why did you make that decision?"

But now we have AIs making decisions, it's a different story.

Any AI decision with an even minor negative impact on human beings results in a public outcry. The media is full of stories of algorithms discriminating against women – or recommending African Americans get longer sentences than their white peers.

Felipe Flores, the founder and host of the DataFuturology podcast reflects on what happened when Uber started testing autonomous vehicles: "The project had the potential to cut costs by 50%, but then a test car killed a bike messenger. Uber discovered that, even though the AI knew it would hit the cyclist six seconds ahead of time – it did it anyway. Within a week, the program was shut down."



Why are we blaming AIs?

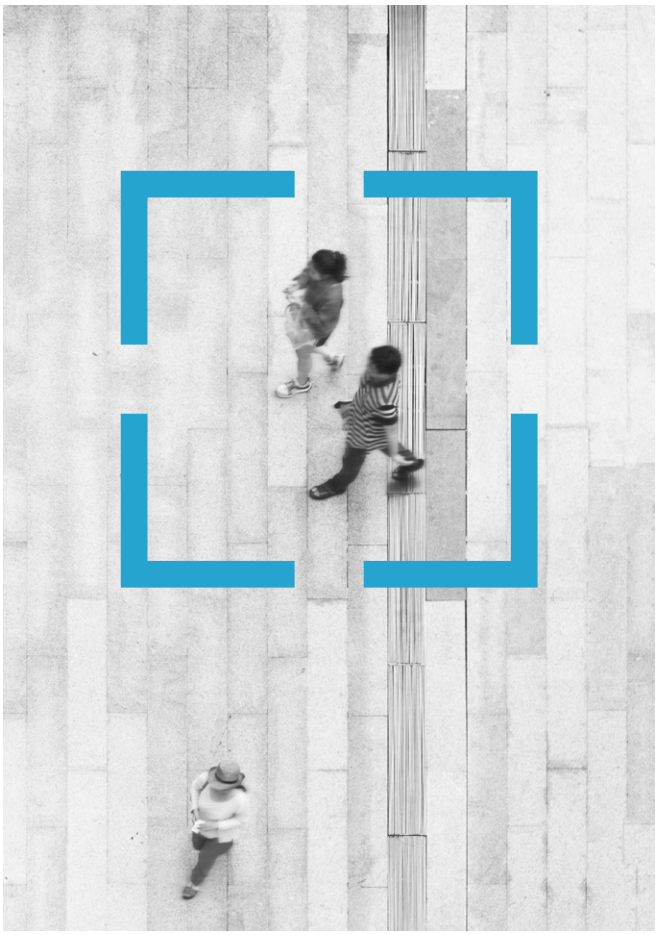
Beyond the Uber use case, NAB's Data Science Manager, Richard Balson, questions the knee-jerk reaction of shutting something down because it fails. He's says AI is a new capability for many organisations and the expectations of what value it will provide needs to be managed carefully as the capability is built out and evolves.

"It's one of our biggest challenges. It's our role to ensure business leaders understand that AI models can still make mistakes because they're based on averages. There will always be outliers. There's no such thing as a perfect prediction. We need to educate leaders about the limitations of AI, to help them understand that if an algorithm fails once, we must continue to model it. If we don't, then we'll never get AI to realise its potential.

What will happen when AI decisions come under the microscope?

Customers, regulators and shareholders are demanding 'explainable AI' – so they can understand why an algorithm made a decision. And this is almost certainly the future for many AIs. "I think companies that develop algorithms they can't explain will struggle to survive," says Flores.

Ahuja agrees: "Some AI tasks will remain black boxes because understanding the decision process is irrelevant or unnecessary. But, in many cases, people will demand to know why a decision has been made – and especially why their request has been approved or denied."



When corporate decisions are laid bare, will the board be comfortable with what this will reveal?

Flores says corporate leaders are often scared by what they find when they unpick decisions. "One company had automated the way they hired call centre agents based on hundreds of thousands of hours of human interviews. Once they'd stripped out the obvious biases – race, sex, age – something was still wrong. On a hunch, a data analyst discovered their hiring practices were biased towards people who wore glasses! This is not something anyone wanted made public."

Will explainable AI reveal some ugly truths?

Flores offers a great example from the early years of AI, when someone fed the passenger list from the Titanic into an algorithm to predict who would survive.

"As you can probably guess, women, children and first-class passengers had the greatest chance of surviving the night. Third class passengers effectively had a death sentence. Now imagine that scenario as an airline or a cruise ship operator. Do we accept that, in the event of an accident, people sitting at the back of the plane or in the lower cabins are more likely to die?"

He acknowledges the sensitive nature of these issues but, like Balson, questions whether walking away is the answer.

"As a society, we're having this conversation for the first time. We need to think more about the trade-offs we're making. When something goes wrong, will we give up and walk away – or have the courage to set a precedent?"

Balson sees a societal upside in holding up a mirror to past behaviours. "It's exciting that we can now identify biases that may exist in society through machine learning techniques. I think it will prove to be a powerful means to prompt evidence-based conversations that will help us resolve these issues."

"If, for example, an organisation were to use AI to predict salaries of potential candidates based on current and historical data, there would be a bias in the salaries offered to male and female candidates due to the known salary gap. Being able to identify this bias when building the AI system could allow the organisation to identify what drove this bias and alter its decision-making process. Using data, they could also then test the impact this has on the pay-gap."

Flores believes this will be essential if companies are to avoid coming under fire in new ways from customers and regulators. "We can now track what happens to people who do or don't get a loan or a university place. The former may build wealth or have a greater life expectancy. The latter may fall into deeper poverty. In 15 years' time, will we see the children of people who didn't get loans bringing class actions against the institutions that knocked them back?"

How can we govern in an automated environment?

The ethical issues attached to AI are so challenging that Ahuja expects an international regulator will eventually emerge. *"We need a regulator focused on critical AI implementations across geographies. It will have to be an entity with the authority – and enough teeth – to act on complaints and proactively drive ethical behaviour. At that point, governance may be onerous, but at least it will come with certainty. Companies will design controls to make sure they don't fall foul of the regulator."*

Until then, Ahuja says it's up to CDAOs to help their organisations put in place the principles of AI governance at two junctures:

- **Building and deployment** – "To avoid conflicts of interest, those building and implementing AI software and systems should not be the ones who are testing and certifying it."
- **Business-as-usual** – "Once running, just like any other process or system, AIs will need continuous monitoring and governance controls frameworks in place – right up to board level. This will include forming an Ethical AI Committee."

"In the absence of a regulator, it's an organisation's responsibility to make sure it's accountable for the decisions made by AIs – not to mention the imperative to ensure data privacy and protect against reputational risk."

Balson says the starting point is for boards to think about why their organisation is automating decision-making. "Generally, it's about making better and faster decisions. Boards should be considering what it means to make better decisions for their customers. To do this, boards need to think about what they will prioritise when there are competing demands and articulate the trade-offs they are willing to make to ensure the best possible customer outcome."

Who should sit on an Ethical AI Committee?

The controversial issues that can arise when AIs make decisions means choosing candidates for an ethical AI committee is extremely challenging. Witness Google shutting down its External Advisory Board for AI just a week after forming it.

Ahuja believes it's vital to find a balance of people with "the authority, interest and ability to judge the accuracy or quality of an AI implementation."

Flores agrees: "The main problem is that a lot of people don't understand how AI works and how to pull the ethics out of the data. You need a high level of technical competency, otherwise the governance process will slow progress to a crawl."

He's worried that the AI governance could go the same route as data governance where, in the early days, non-technicians oversaw the work of data scientists to the detriment of speed to market. "I worked with a bank whose snail-like data governance process, overseen by a non-technical team, meant it could take six months to get from finding a business problem to working on it. By the time they made a move, competitors were already there."

Balson is more concerned that the conversation about AI Ethics is elevated to a sufficiently high level.

"It's a philosophical discussion that needs support from the most senior levels. It's really hard to get consensus, so you need strong leaders who can understand what you're trying to solve and reach agreement to drive outcomes."

What assurances do consumers need to trust machines to be fair?

Balson believes consumers want transparency around how decisions are made, but questions how much detail people will require. "Organisations and society are trying to solve for what level of explain-ability customers expect for AI systems. Is it enough for the consumer to know an organisation's principles for building AI systems, or do they need to understand each decision made by the system?

He points out that it won't always be possible or desirable to offer granular explanations. "When people understand how decisions are made, they change their behaviour, making the AI system less accurate." Instead of providing granular explanations, he sees a regulatory trend coming to Australia that will allow consumers to challenge or opt out of auto-decisioning. "If you think you're being unfairly treated by the AI system, you may have the opportunity to put a human in the loop."

Balson is emphatic that consent communications shouldn't end up like terms and conditions, where consumers are bombarded by so much information they don't read any of it. "The good news is, digital channels now give organisations the opportunity to have tailored conversations with customers based on their level of data literacy."

He believes that, as people become more data literate, they will get more comfortable to both challenge organisations who are doing the wrong thing and give consent to those who are using data to deliver customer benefits.

"There's a misconception that data is being used for things that don't benefit consumers. In fact, most organisations are using customer data for good: to improve offerings, to get better at interacting with customers or to combat fraud."

"Once we have more conversations about why we're using data and how customers will benefit, perceptions will change."

5 steps to building an AI Privacy and Ethics framework

Ahuja says that, rather than having a 'code of AI ethics', organisations should adopt a privacy and ethics framework. "Codes can become pedantic. You need to be able to adjust based on outcomes."

He suggests a practical five-step process for developing that framework:

1. **Determine and document the purpose of AI implementations** – "Getting agreement on the purpose is critical. What's the vision and mission of why you're investing in AI? If it's purely about profitability, I would question that. In the current environment, no business can survive unless it lifts its goals above basic financial performance. You have to consider how the customer will benefit – how you can align your AI development with what your customers value."

2. **Consider what could go right and wrong** – "Scope out the best and worst case impact of the AI system on individuals, families, organisations – even countries. Only then will you understand the intensity of potential consequences. Classify risks with different severity levels."

3. **Decide what needs to be done to manage these risks** – "How do you mitigate against the worst case scenarios? Do you need a dispute or issue resolution process?"

4. **Plan how to rectify issues** – "What steps will be needed in each case? Do you need a recovery plan?"

5. **Plan how to manage legal liability** – "This is a critical part of the framework. How will you manage legal liabilities and reputational risk? Develop a communications plan encompassing shareholders, board members, customers and the media."months to get from finding a business problem to working on it. By the time they made a move, competitors were already there."

Balson is more concerned that the conversation about AI Ethics is elevated to a sufficiently high level. "It's a philosophical discussion that needs to be had at the most senior levels. It's really hard to get consensus with these issues, so you need strong leaders who can reach agreement and then go out into the business and get everyone comfortable with where the conversation has landed."

Is this the most important conversation you'll ever have?

Ethical AI is most likely to develop in organisations with a mature, fail-fast culture, where leaders understand both the bigger picture of what's possible and the limitations of algorithms. Ahuja says the task of educating the board and executives on these issues will fall to CDAOs, requiring them to be:

- **Educators** – "CDAOs should be presiding over regular forums where builders and owners present the effects of the AI system to the board. To grasp ethical issues, boards need to understand how an AI system works and then be updated regularly on: what it's learning, how it's continually improving and the implications of changes or new functionalities."
- **Storytellers** – "That means, CDAOs must be able to take the board through high-level stories to articulate underlying functionality and different outcomes. As data analysis and data gurus, this is not just our opportunity but our responsibility. We are the only people who can explain to a board-level audience the business meaning and ethical implications behind AI functionality. CDAOs need to partner with colleagues in marketing, communications and data visualisation to help put together stories and refine messaging."
- **Translators** – "Importantly, these stories need to be told in the board's language and underpinned by meaningful metrics. If an AI is conducting an automated marketing campaign, the board has no interest in programming issues. They want to know: how we're getting permission before sending communications, how many customers were contacted over the last quarter and what's happening to the conversion and unsubscribe rates. To make a point hit home, use verbatim quotes from customer feedback and complaints."

But, to have these conversations, CDAOs must be prepared for complete transparency.

"The most important thing is to operate in a culture of honesty and trust," says Balson. "The advances of AI and machine learning are changing every day. AI experts, and team using these systems have to be open to having conversations about what they don't know, what can go wrong and what's no longer working."

As organisations infuse AI into their operations, leaders need to understand the ethical implications of harnessing this powerful technology. As well as being the technical architects of AI, CDAO must also become educators, storytellers and translators to ensure strong, relevant governance that protects organisations without hampering AI-based growth strategies.

Corinium would like to thank our expert contributors:

Felipe Flores
Gaurav Ahuja
Richard Balson