DATA LEADERS WHO'S WHO

DRIVING INNOVATION WITH DATA

Featured in this week's interview

Marek Rucinski
Deputy Commissioner | Smarter Data
Australian Taxation Office (ATO)

Data Futurology

Talent insights.

Data Futurology

Talent insights.

TECHNOLOGY

Editors note

"We are excited to bring you the Data Leaders Who's Who, 2021. This publication is a collection of stories from the frontline - thought leadership from data chiefs who are driving change and making an impact with data. We extend our sincere thanks to the leaders featured for contributing to this initiative and sharing their insights with our audiences in support of lifting the data capability of the community."

James Lecoutre, Partner, Talent Insights I Felipe Flores, Founder, Data Futurology

AI AT THE ATO

Marek Rucinski
Deputy Commissioner
Smarter Data
Australian Taxation Office (ATO)

The ATO has established a careful approach to where and when AI projects should be implemented in order to deliver quantifiable value. Marek shares AI/ML use cases at the ATO and strategic advice on how to scale projects in order to avoid what he calls "proof of concept hell!" This rich interview also features insights on structuring teams, as well as really practical steps to make sure your AI ethics principles aren't just a poster on the wall, but actually implemented in day-to-day operations.

AI presents huge opportunities for organisations to tailor customer experience and present more convenient offerings to your customer base. How do you determine which projects to use AI for and prioritise your projects?

We consider the use of supervised and unsupervised AI techniques to derive actionable insights for business where the solution to the problem requires predictive insights, consumption of large and multiple data sets, or data that is semi-structured or unstructured. We apply our data ethics principles to guide us on how to apply these techniques.

AI is not always the most suitable answer to a data and analytics project. Basic slice and dice analysis and visual reporting of high-quality data can also deliver significant value. Visual data representations can be a simple, but powerful tool to create new insights and consume complex data patterns.

Whether you use AI or not, you can realise additional value by integrating the insights or data into human and/or robotic workflows. But human oversight will always be needed!

We use several criteria to prioritise our data and analytics projects, including:

- The extent to which it is required to deliver legislative changes and government programs

- Alignment to the ATO's strategic objectives

- Our overall data and analytics capability to deliver it.

Where the project is not related to legislative changes or government programs, we assess the value potential, strategic fit and feasibility of execution.

It is often difficult to quantify the ROI with AI and machine learning projects, is there anything you do specifically to measure the impact of an analytics product or project?

How to measure ROI for an AI or machine learning project may not always be obvious so you need to have an experimental mindset to ensure you understand the impact and benefits derived.

Benefits can be realised in different ways. They may focus on:

- Business efficiency, for example, automation projects with AI decision nodes that deliver high throughput saving labour costs.

- Generating new, better, faster insights from data. This can translate to additional revenue or increased efficiency as the models provide more relevant actionable insights compared to blind data sampling.

- Improving the client experience, for example, by pre-filling information to reduce the time and effort required to meet tax and super obligations

- Legal compliance where the project relates to policy or legislative implementation.

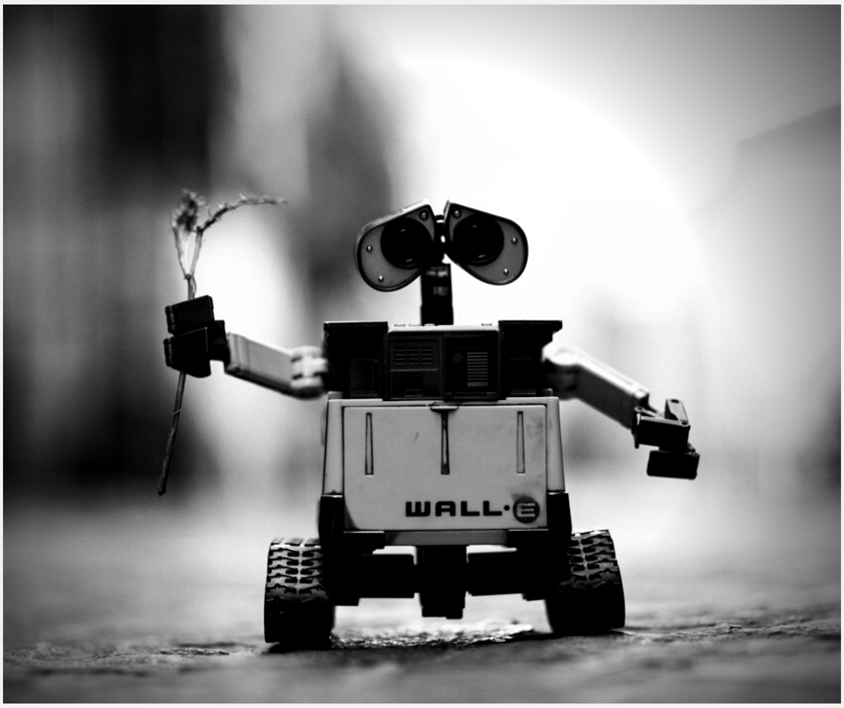
Evaluating value realisation is essential to ensuring we achieve what we set out to do. It is also informs the achievements that can report to government and the community. For example, the government funded Tax Avoidance Taskforce, which has formal reporting requirements, relies heavily on the delivery of data and analytics solutions to achieve its outcomes and deliver a return on investment for government.

Can you share any examples of AI/ ML solutions deployment or use cases where this has worked well?

In the ATO, we recently used machine learning to help our Tax Avoidance Taskforce ensure multinational enterprises, large public and private businesses (and associated individuals) pay the right amount of tax in Australia. We adopted a hybrid approach to augment a previous business-rule method that identified taxpayers of interest based on certain criteria. The hybrid approach leverages business expertise and machine learning, which provides coverage for a wider range of known and emerging profit shifting risks and improves the case selection outcomes.

We also delivered CbC Interact, a tool that makes unstructured data from country-by-country (CbC) reporting statements discoverable, accessible and usable for staff – this has expanded the data available for analytics and reduced manual effort in analysing the information.

For GST compliance activities, we currently adopt a hybrid approach where we use a combination of high-level business rule and complex analytical models. The business rules are based on clients' registration details, maturity of business, lodgement patterns, unusual behaviour, certain ratios and industry benchmarks.



"THESE MACHINE LEARNING MODELS IDENTIFY HIGH RISK GST REFUNDS AND RANK THEM BASED ON THE LIKELIHOOD AND CONSEQUENCE OF THE RISK."

These high-level rules are supplemented by advanced analytical models. These machine learning models identify high risk GST refunds and rank them based on the likelihood and consequence of the risk. Similar machine learning models are deployed to identify fraudulent behaviour. These models result in increased accuracy and improved client experience.

For large enterprises with legacy technology, how can we overcome the data hurdles in deploying AI?

One can argue there is a natural trade-off between scale and agility. Though the scale of an organisation can create greater opportunities to generate value in AI and data and analytics generally. The value generation must be relevant and aligned to the strategic objectives of the organisation.

Large organisations acquire and generate lots of data. A wise person once said, “quantity has its own quality”, which rings true in this context. Large data sets, when connected and curated are the corner stones of AI projects. These projects require training data, linear data series, networks of connected data sources to fuel the models, and generation of breakthrough insights. More often than not, this will be found in large organisations. These organisations have a foundational competitive advantage in AI and data driven projects. That is why so many of them are feverishly trying to acquire even more data.

What are the biggest blockers to AI deployment?

The trick is in how we create organisational agility in larger organisations to leverage the data, identify value rich use cases, and create coherent energy in pursuing the use cases to generate momentum for the whole data driven culture to emerge and flourish. Larger organisations are data dependent and must be driven by use cases.

We can look at this in terms of strategy, technology, process and people.

"INTEGRATION OF SMALL-SCALE PROOF OF CONCEPT TO AN INDUSTRIAL PROCESS IS A KEY POINT OF FAILURE FOR MANY ANALYTICAL PROCESSES."

STRATEGY

AI and analytics use cases need to align to the strategic priorities of the organisation, otherwise they will only exist as back-room curios. They need to have distinct potential to generate value for the organisation to garner the investment required to pursue them. If use cases do not have that, they will be blocked and rejected.

TECHNOLOGY

This is becoming less of a barrier, as its components are becoming increasingly commoditised. However, the data must be available to fuel the analytical process. The technology components also must be integrated and of relevant scale to support the process. Without this, projects will not move beyond pure concept. Or worse, they will get stuck in “proof of concept hell”, where they showed promise to deliver value, but failed to scale.

PROCESS

Integration of small-scale proof of concept to an industrial process is a key point of failure for many analytical processes. It is often simple to show a one-off result. It is much harder to scale and integrate them into a workflow. Sophisticated and experienced practitioners clearly understand and value this ability to take the business on a journey, to tune and change the processes to take advantage of the new capabilities and insights. This is where value gets generated, but more importantly, it gets multiplied.

PEOPLE

Analytics and AI is a team sport. A lone data scientist may have a brilliant idea, however they will not be able to make it real without data managers and analysts feeding the right data to the model, data engineers curating the right data sets, and business stakeholders actually using the insights to generate value. They all play a role and need careful orchestration.

What are the limitations to scaling AI/ ML solutions in your opinion?

Firstly, as outlined before, ML/DL solutions are not the only path to generate value from data and analytics. I view them on the high end of the spectrum. They have a place in helping us cope with specific use-cases that depend on consumption of large data sets, or ones requiring predictive and prescriptive types of insights.

These considerations create natural boundaries for ML / DL applications. Nevertheless, this still leaves a vast landscape of potential value for their application.

Overall, we are not even close to realising the limits yet, and it is very difficult to actually speculate on the potential limits. We are still limited by computing power, although much less than in pre-cloud days.

We also need to adhere to ethical boundaries, just because we can does not mean we should. We can use DL to generate new insights and decisions, but if we cannot explain them, their value to the business is diminished. At the same time, we can use them to advise the business on patterns and this leverages them.

Another limitation is the fact that even in DL guise, we are still dealing with narrow AI. It follows through on the instructions we give it, and maybe that is a good thing. The opinion on the outcomes of singularity and broad / general AI is rather split right now and explored widely in SF realm.

In May 2021, only 22% of Data Futurology respondents polled in our Advancing AI series said they have automated monitoring of ML models in production. What would be your advice on why models need monitoring?

I have mentioned the ethical constructs needed to guide the implementation of the ML or DL models in production. This is foundational. We are very careful to ensure they augment, not replace human judgement. I think this is key to their adoption and success.

ML models are easier to implement as they offer higher levels of transparency and communicability. DL models are of value as they offer breakthrough insights, that the human brain would have difficulty replicating individually in context of connecting data points to drive insight or at scale. ML and DL models also need tuning and re-parametrisation to ensure they stay within tolerances of the expected performance and efficacy. I believe they can be deployed with success in the context of the above considerations and by combining them with an industrialised process integration framework that manages the model lifecycles, linking them to data management and software development concepts. In the ATO, we are working on this dev / data / model ops mash-up.

How do you integrate AI ethics into the fabric of the organisation?

The ATO has published six data ethics principles that form the backbone of all that the whole organisation contemplates and executes with data and analytics. It took us over a year to develop these principles, from concept, through global benchmarking, to executive level discussions and endorsement.

Of course, the key here is not just the generation of the statements but the day-to-day application of them across the fabric of the organisation. Only then, are cultural norms influenced and new values added. To achieve this, data ethics is part of the ATO's corporate mandatory training for all staff, including contractors and secondees to the ATO. In addition, we are developing an operational framework and guidance material that unpacks the data life cycle and illustrates how the principles apply. We then work with the business stewards to ensure these are real across the data sets and analytics projects.

What processes do you have in place to measure bias in decision making?

This is a complex issue and there are several factors to consider. Firstly, the above-mentioned ethical principles serve as an important reference point for model development and application, and thereby organisational bias management for data and analytics overall. Secondly, as a government agency, the ATO is guided by the laws it administers. The laws guide us in how we administer the tax and super system and the decisions we make.

Lastly, we can delve deeper into how we construct our models – especially in the ML/DL space. This focuses on constructing the right training sets to minimise bias yet preserve the population patterns we are trying to leverage for predictive reasons.

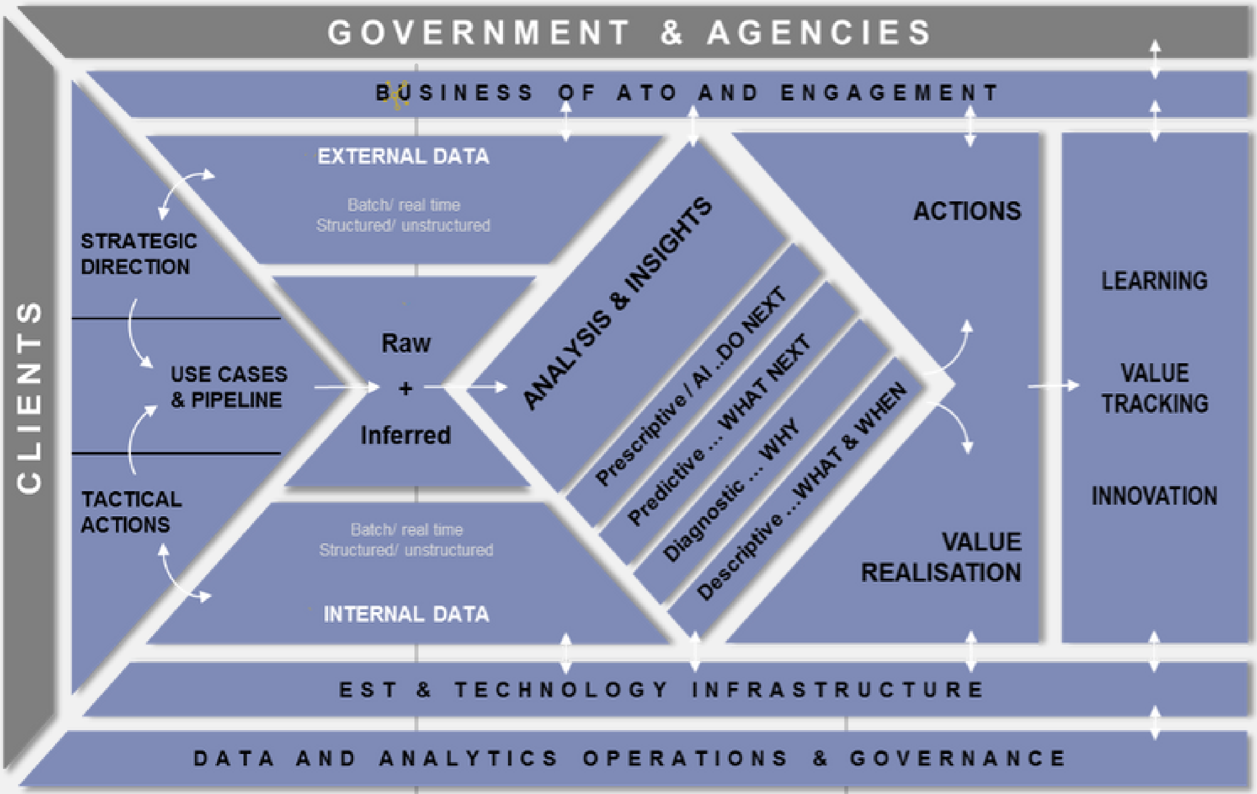
That is, we apply due diligence, review AI outputs on test sets and monitor results while in production. We review input data to explore and understand possible bias due to collection methodology or sampling. These are established practices in Smarter Data. We're also looking towards how we can augment these human / practice driven processes with open source tools that are designed to quantify model bias. These tools will help automate a mix of statistical and ML techniques to support our team in executing good judgment on the presence and mitigation of bias.

What have you done at your company to build organisation-wide trust in your AI?

The ATO is very structured in how we apply AI to decision making. We are focused on augmenting our human capital, not replacing it.

As such, our efforts focus on how we can create confidence in our models for their users. This then focuses on how we make them explainable and transparent to create confidence in both the data used, as well as the model's logic. Our data scientists spend a lot of time with business users to test models to ensure insights are actionable and relevant. We pilot them in small scale and implement progressively. Once implemented, we monitor the model performance and tune as required to ensure that efficacy is maintained. We also try to share our approaches, case studies and successes broadly to educate and excite the business community about the potential present in directed application of analytics to drive value.

"THE KEY HERE IS NOT JUST THE GENERATION OF THE STATEMENTS BUT THE DAY-TO-DAY APPLICATION OF THEM ACROSS THE FABRIC OF THE ORGANISATION."



What is the best way to structure your data and analytics teams? What processes and methodologies are key to underpinning analytics project success?

I will address the where and how of this.

WHERE

- The analytics team needs to be close to its clients and consumers of data. The aim is for the team to become trusted advisors, not pure providers of services.

- The team needs to engage in constant two-way dialogue with the business to refine the problem statements, iterate the solutions, hone the delivery and track the results.

- Teams that fail usually are further removed from the people that they serve, they are buried in IT or finance teams, and as such do not possess the intimacy of contact, or clear identity that this capability requires in the organisation.

- I usually talk of the triangular-relationship that is essential to the success of an analytics team, one between business, IT and analytics. This shape, if in balance, is very strong. This trick is to retain the balance.

HOW

- Here the considerations are fluid again, as the capability and professional patterns of data and analytics organisations are still in flux, because the whole capability is still maturing. It is where IT or marketing was 30 years ago. Things will still evolve.

- At the same time, we chose to demystify things as much as we can and relate our structure to the data and analytics value chain. To show it is a process, one that can be mechanised, scaled, pulled apart to show how it works.

We are seeing a real demand across the industry for ML engineers, do you see that changing in the long term? Will ML just become a fundamental component to the data scientist's role?

This is true not only for ML engineers but also data engineers. We are going through a peak now as organisations are stepping up their efforts to harness the data stores, create the right pipes and scale the modelling capability to generate value. This reflects the high growth phase of the current capability. It will plateau as demand gets fulfilled. I also firmly believe as we move into the low / no code stage of programming, these skills will become more democratised in regard to pure specialisation. The individuals that will retain high value are the ones that can connect the components of the analytics value chain and thus architect the sections of the whole or whole solutions. Those skills will always be in demand.

For those who are hiring on potential, what skillsets are most important to be able to build on?

As a starting point, good technical skills whether in maths, AI or statistics combined with coding in open source are somewhat baseline now in the data science space.

Beyond that, the below skills are also important:

COMMUNICATION SKILLS

There is less and less room for pure “back room people”. We need individuals that can connect with the users of their models and insights, to elevate the solutions through dialogue and to collaborate with rest of data and analytics team on design, build and delivery. That is, those who can communicate complex and/or technical issues clearly and simply and have the ability to tell the story.

STRATEGY SKILLS

These become more relevant the more senior the role. This is not just about looking at the “now” with the client and solution. It’s about the ability to develop a clear future vision that links with the organisation’s big picture and deliver outcomes that align with the long-term perspectives. This allows for the development of more enduring and continually evolving solutions.

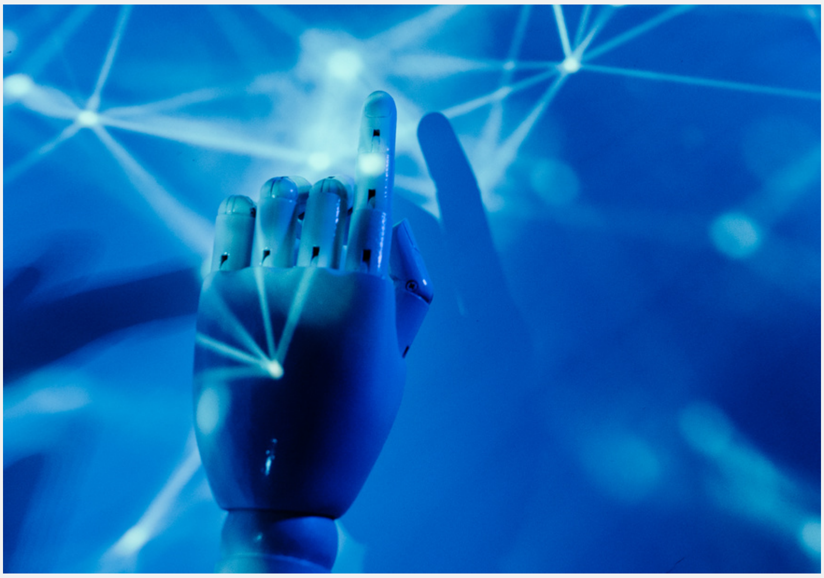
What would be your recommendations to leaders looking to attract diverse talent to their teams?

- Tell stories and project your identity.

- It is important to use social channels and industry forums to create an image, tell stories around how data and analytics is done, or what and how things are done at your organisation. In many interviews I have conducted, good applicants do this research. They check the footprint of you and your organisation.

-- Beyond that, having a clearly articulated operating model and career path, and community of practice is important to ensure that the talent lifespan is extended.

Last but not the least – interesting and meaningful work. We can easily point to the impacts we have on the community – that helps!



Lots of companies have fairly restrictive data governance policies, some are still formulating theirs. How do you find the balance between data governance and allowing analytics and data science teams the freedom to build models and provide insights to their business stakeholders?

Data, especially private data needs to be safeguarded. The ATO houses lots of such data and have clear processes to protect it – it is what the community expects of us. Therefore, I cannot be dismissive of the restrictive mindset. It creates confidence in the actual ethical usage of data in the analytical models.

The trust and confidence is hard to achieve, and is easily lost. As such, the balance is essential. We have secure sand boxes for data scientists to play in with business to prototype. Many of our analysts and data scientists have certain security clearances that allow them to work hand in hand with business stakeholders to ensure what we create is relevant.

Are there any data governance tools or frameworks you have been impressed with?

This is a bit of a moving target, as things change all the time. We constantly benchmark with consulting companies across the industry and our peers in Australia and overseas – we try to take 'best of breed' approach across components and then modify it to suit our needs.

What new technology and innovations do you see as being the most critical to your industry over the next 18 months?

- Cloud computing is still evolving in context of penetration at scale. I think it's already proved itself to be transformative regarding compute on demand – more will follow.

- Open source will continue to grow. The exchange and legalisation of algorithms are very exciting as it accelerates the development cycles.

- Low code / no-code data science products are key to near future democratisation of this field.

- Automation will be more pervasive and coupled with AI. We see huge potential in this to increase efficiency and effectiveness of our interactions and experiences.

The data landscape will continue to flex and expand so there will be more signals to integrate, but also to leverage.

Finally, what work are you most proud of at the ATO?

There is a lot to be proud of as to what our ATO teams do and deliver. To name a few – the:

- Immediate impact we made in supporting the Australian community during COVID-19 pandemic – our quick implementation of data and analytics solutions to support the administration some of the Australian government's economic stimulus measures, including JobKeeper, early release of super and the cash flow boost, will be remembered by all who were involved.

- Work around various taskforces – where the analytics behind the decisions helps the collection of additional millions of dollars of revenue under the laws the ATO administers.

- Constant innovation across the board – implementing robotic process automation, new visualisation tools and re-engineering the whole data stack.

For me personally it is important to reflect on HOW we do things and how this has changed in an enduring fashion we:

- Evolved our function from a side / support team, to one which is viewed as integral to the strategic priorities for the ATO

- Created and evolved our operating model to scale our ability to deliver and industrialise our operations

- Are transforming the way the broader ATO accesses, uses and leverages data themselves by uplifting data literacy and enabling self-service tools. This will last for a long time.