DATA LEADERS WHO'S WHO

DRIVING INNOVATION WITH DATA

Featured in this week's interview

Felipe Flores, Founder, Data Futurology

Data Futurology

Talent insights.

TECHNOLOGY

Editors note

"We are excited to bring you the Data Leaders Who's Who, 2022. 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 | Felipe Flores, Founder, Data Futurology

THE FUTURE OF DATA SCIENCE

Felipe Flores
Founder
Data Futurology

Felipe shares on why prioritising change management is key to getting AI products into production, and the importance of stakeholder engagement in lifting data capability across the organisation. Here are his broad insights to help in uplifting the strategy, execution, and future of data science.

How do you devise a data strategy? What sets apart the good from the bad?

One of the simple definitions I like of a data strategy is that firstly, it gets people aligned on where the organisation is today, in terms of both data and analytics maturity and secondly where they want to be in the future. The data strategy creates an execution path to get from point A to point B. Several use cases will emerge from that execution path, and these will need to be prioritised.

In terms of what sets apart the good from the bad, a lot of it is the internal alignment, and whether the strategy has been able to create the desire for the organisation to change. That desire starts with awareness and an understanding of where the organisation is now. Why it needs to change and where we need to get to. Then the goal is to adopt an aspirational outlook. “Aspirational” here doesn't necessarily mean it has to be very exciting, futuristic, or difficult to achieve. That's not necessarily aspirational from a data strategy perspective.

Rather, aspirational could still be realistic, it’s just important that there’s the goal to step up from where the organisation is today.

In essence, a data strategy is broader than analytics and tech strategies and must encompass everything including culture, technology and aspirational goals. The shared understanding of the organisation in terms of where the organisation is today and the aspirational goals is a key component that sets apart a good data strategy. Also, in order to lift your data strategy you must to ensure that data leaders have a shared consensus on the project prioritisation for future use cases and a preliminary road map with logical groupings of use cases according to business impact or customer desire.

Felipe, project prioritisation is a really important point. How do you determine which projects to use AI for and how do you go about prioritising your projects?

AI can be used for operational decisions and for strategic decisions. They both need to be handled differently.

Operational decisions happen thousands, if not hundreds of thousands of times per day. Hence there is a lot of work that can be done at the MLOps and ML engineering area that can help us improve the scale at which these services are delivered.

On the other side, there are relatively few strategic decisions. If you think about an acquisition, for example, it is an important strategic decision for an organisation, but it would happen a couple of times a year at most. But there's a lot of analysis and insights that needs to go into large strategic decisions and AI can be used for insights to assist there.

In terms of which projects to use AI for – I’m usually thinking about operational. We need well prepared, high quality data presented to the model creation part which will create better models over time. With strategic AI it is much slower, in small groups, using AI/ ML to create insights and consult or advise on decisions.

There are four components that help with project prioritisation: 1. Feasibility – can it be done with the data, technology, people and resources we have now? 2. Desirability – does the customer want this? Can we quantify that market desire and how important that feature is for the market? 3. Viability – is there a financial return on this? Is there a path to profitability? 4. Organisational readiness – is the organisation ready to bring the product to life for the customer or ready to consume that product?

This readiness encompasses two sides – technical readiness and also the people side – are people aware that this is a problem or that customers want this and do we have the momentum to bring this product to life? Timing here is critical.

How have you found success in raising data literacy in your organisation? How do you get involved in educating peers and execs?

It has become industry standard to have data literacy programs within the organisation that have multiple levels, based on the existing competency of the team and also the business function they are in e.g. finance, sales etc. As an organisation we want to lift all individuals' data literacy levels, although everyone will have differing levels of data literacy already. Ideally you want to manage this as a group, based on your data literacy, so groups within the organisation can be trained simultaneously. Then it's important to have the accountability, so that managers have the desire and KPIs set. Where there are KPIs set from a senior leadership level, this helps managers to build in time for their teams to raise data literacy. Managers can experiment with the format that will work for their team, whether it's concentrated sessions over a day or two, a couple of hours per week spread over a longer period, or a combination of the two. 101 levels tend to focus on data visualisation, SQL – a great start to unlock the power of data in the organisation. 201 levels may cover a gooey tool to do ML. 301 is on the programming side, extending to cloud skills, hopefully also data engineering and extending the 201 skills.

In terms of engaging peers and execs – this is a great question because it is so challenging! In my experience in one-to-one sessions and later small groups. It takes a long time to create the awareness and the desire.

Execs need an awareness of what data literacy is, how it can align with your plans and support your goals. That is not achievable in a one-off conversation or presentation. It takes time to bring the group along and multiple touch points to build the trust and grow their desire to invest. That's how I've seen success and even seen bosses to do formal studies to gain a better understanding of the data field. This signals to the rest of the organisation that data literacy will support business goals.

Success involves starting with awareness and desire, moving to knowledge, and additionally allowing people to find and follow the curiosity, while connecting them with others in the organisation for the shared knowledge journey.

In terms of execs – the key is small groups, spend time on this and you will see results!

"EXECS NEED AN AWARENESS OF WHAT DATA LITERACY IS, HOW IT CAN ALIGN WITH YOUR PLANS AND SUPPORT YOUR GOALS. THAT IS NOT ACHIEVABLE IN A ONE-OFF CONVERSATION."

"THERE ARE CRITICAL COMPONENTS THAT INCREASE THE PRODUCTIVITY OF THOSE DATA SCIENTISTS THAT ARE NOT SEEN BY EXECUTIVES. DATA WAREHOUSE WORK IS A COMMON EXAMPLE."

What do you wish senior leadership knew or understood?

The thing that I find particularly challenging to convey myself is the value of the parts of the data end-to-end ecosystem that the executives don't interact with directly.

Traditionally, executives are happy to invest in data scientists and get the tools for them. If the use of those tools is generating more value, then you're going to continue to get support. But then there are critical components that increase the productivity of those data scientists that are not seen by executives. Data warehouse work is a common example.

This comes back to raising the data literacy with peers and executives. I've worked with executives and bosses that have started to spend time in the data warehouse building queries, and then they start to get an appreciation of the complexities that live in there. As a result then we've been able to get more support to continue to improve the warehouse. This is also the case for information security, access controls, governance and so many other unseen areas of the data ecosystem.

What execs don’t realise is that the benefits of the areas they don’t interact with make the areas they do interact with so much faster, so much more powerful and so much more accurate. It really supercharges the capability in a way that is difficult to convey because of the complexities of the data ecosystem. I wish that execs understood that area better – I think it is improving a little. I know that data leaders are putting a lot of focus on this in their own organisations and I am grateful for that.

Operationalising AI. How do you deliver AI at scale and get more models into production?

The first step is having an easy way of creating large quantities of models on an ongoing basis. When we think about the creation of the models, it starts with a with a good feature store, with a curated data set that can be used as an input for multiple machine learning and AI projects. In a lot of cases, having vendor solutions gives you that first benchmark, and from there, the organisation can invest on creating more handcrafted models where it makes sense to.

From what I've seen, a lot of a lot of the work on improvements can be done on the data side. You can work on the quality of the data itself, or bring in new datasets and new features, and spending time on that will give you a more useful more valuable AI model, reaping greater rewards for the business.

"IF YOU DON'T DO THAT CHANGE MANAGEMENT PIECE UPFRONT AND EARLY IN YOUR JOURNEY, IT TAKES ABOUT FOUR TIMES LONGER LATER ON."

Sometimes handcrafting models is necessary, but I think I think we can get away with doing auto ML for a large proportion of the models that we need.

What are some of the lessons learned you've encountered when getting AI products into production?

One of the key lessons I've learned here is to work on the organisational components as early as possible. These organisational components are usually people, so the change management piece is crucial and the earlier you begin that journey the better. I've mentioned already that aligning our projects to the organisational strategy is imperative. We need to make sure AI is going to make a difference where it counts. For that we need to have conversations early on to make sure that people are aware of the difference AI can make and there is a real desire for it to happen as part of the business strategies. Often the technical solution is the easy part but getting the organisational buy-in and backing for that solution to be operationalised and adopted throughout the organisation is the challenging part. If you don't do that change management piece upfront and early in your journey, it takes about four times longer later on.

For this reason we are starting to see change or relationship managers being hired into analytics teams so that they can act as a business liaison, raising awareness of AI success and growing the desire to operationalise AI. On-going communication to continually remind the business of organisational wins and success with AI beyond the technical project or product duration will support you in productionising your AI.

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

The best way to structure teams depends on the organisation and its maturity. The way that I've currently structured my team is in a matrix or in a hybrid structure, where the formal reporting lines are by capability. We have data scientists that report to data science manager, who reports to me. However, then the data scientists don't work with each other in the same project unnecessarily, because then we have cross-functional project teams across the organisation. To manage that, we have our people almost feel like they have two teams. There's a project team where they work day-in, day-out and they might be the only data scientist in those teams. We then have the capability area, which is responsible for upskilling people, leveraging lessons learned and sharing models, platform, and code to reuse across the business. This where we have the on-going connection for the team in the capability section. We have several ways to bring the capability team together including a book club and meetings where we showcase projects.

In terms of methodologies underpinning success, over the last decade agile has been something that a lot of organisations have adopted. However, traditional agile processes like agile and scrum can be a little slow for analytics. Traditionally you have a two-week cadence with a two-week sprint, where the work is planned for that sprint and then it gets delivered. However, you're able to get a much clearer picture of what you're shooting for in data science and it's much more about discovery and iteration. As a result, I've typically favoured using Kanban as the approach within the agile umbrella.

In Kanban, the backlog is being continually reprioritised. When someone on the team has capacity to take on new work, they pick up the work that's at the top of the priority list. What Kanban does is limit the amount of work in progress by having a small number of items in the in progress items under 'to-do' and 'in-progress.' That's been beneficial in terms of getting focus and allowing teams to reprioritise approaches to getting the answers they are looking for.

Finally the other component around stakeholder engagement that impressed me lately came from my colleague Ram Kumar, CDO at Cigna. He has created monthly surveys where the business stakeholders and analytics team rate each other on how easy they are to work with, how much support they are getting from each other. They then have monthly meetings to review that and I loved that concept. They also have KPIs that are structured around the team's survey results as well as individual ones so that they become team players and support each other across the team for success!

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?

Perhaps things will change in the long term, but not in the short to medium term.

"DATA ENGINEERING SKILLS ARE HUGELY IN DEMAND AND WILL CONTINUE TO BE SO FOR A LONG TIME SO INVESTING IN THESE SKILLS IS IMPORTANT."

As to the question of whether ML will become a fundamental component to a data scientist role, I would change that a little bit to say that engineering will become a fundamental component of the data scientist's role.

Once we have a version of a model that we want to start to get value from, we move into the world of engineering, and particularly of machine learning engineering and data engineering. I like to think about this in terms that for every data scientist that an organisation has, they will need between two to four data engineers or machine learning engineers. That really shows why we have such a shortage of engineers today, and the demand for engineers in this space will continue for some time. There's a big opportunity for upskilling and cross skilling, where data scientists are learning more about data engineering and machine learning engineering.

Over the long term, I do think that data engineering will become a part of the data scientist's role. Data engineering skills are hugely in demand and will continue to be so for a long time so investing in these skills is important. I envisage that will have more generalists in most small to medium sized companies. Larger companies may still have very niche roles but for smaller companies, generalists with both data science and engineering skills will have a greater value within the organisation.

How do you ensure you are leveraging new tech for innovation, rather than tech for tech's sake?

There are so many organisations that get caught up in chasing the new shiny object. For technical delivery teams like analytics, AI, and technology teams, it's easy to come up with areas where we see the problems or we're excited by the possibilities of what could be done.

What I found is that when the ideas and the brainstorming is generated within the technical delivery team, the business impact will be either limited or misaligned to where the organisational priorities are.

Instead, what I encourage is still brainstorming, but rather than then going and building the technology, or a proof of concept, and then trying to sell it internally. Instead, what you do is you start talking to people about the idea itself. The aim is to get the right business stakeholder to be excited by this, and then to request it from the technical delivery teams, the data science and the data engineering teams. When the business is asking you for your idea, that's huge.

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

There are a few frustrations and areas where, despite the huge advances we have made in the tech space, some unsolved problems remain. Five areas in particular come to mind. The first one is around making it easier to do data preparation, and particularly with regards to the versioning of data, which is something that I think is still a largely unsolved problem in the ML and AI space that might have some simple solutions. Understanding what version of the data was used for a particular model and being able to create that kind of end-to-end link is still largely an unsolved problem.

As part of that data preparation umbrella, the ability to create feature stores at speed, at scale in a larger way, with monitoring and alerting is still a high bar. This is the whole movement of dataops and the space that plays in. I'm really excited that this is a movement as it will help us to make improvements in this space.

The second one is democratising AI, sometimes referred as citizen data scientist. We have to move into that world where more people in the organisation need to have these tools at their fingertips and get them to a point where there's value being created from this wonderful technology. The barriers to entry are currently too high, but that's changing to make the creation of models more accessible through the organisation. There needs to be an easier way for operationalising of those models. Making AI more accessible is the only way to transform our organisations with this technology and for that it needs to be held at the standards we hold our data science & ML teams to in terms of the processes, reliability, and security. Bridging that gap for non-technical people will be critical.

The third is better auto ML. We've had a lot of progress in the creation of models in that space, but the efforts are still a little bit disjointed and consequently there are still a lot of gaps in terms of the smarts that can go into auto ML. Currently there are a lot of repetitive things that data scientists need to check for and adapt and still many logical decisions to be made by hand. This needs to change and there's still a lot of room for automation and optimisation that can happen in the creation of models.

This is definitely a space I would expect to see improvements going forwards.

The fourth huge opportunity is for better ops. Dataops, MLOps – this movement has made huge progress over the last 18 months but still has a long way to go to take us to that place where every organisation has thousands of models in production. It is super important to be able to reliably deploy, monitor, track and alert the quality of predictions that are being created by models in production.

Finally, and possibly the most important is responsible and ethical AI. This has to be an area where we will see technology advancements. We must have the confidence to create models with the impact we want in the world and that means that we need improvements in responsible and ethical AI. Most organisations today just don't have the data required to make fair, unbiased models because the data wasn't captured with the purpose it is used for. We need to improve the way we are either capturing data or sharing the data. We need to be exploring how to create the data to reflect the world we want and not the existing biases that exist in the world today. How can we use AI to build the world we want? That is a space that I hope is going to evolve a lot over the coming months.