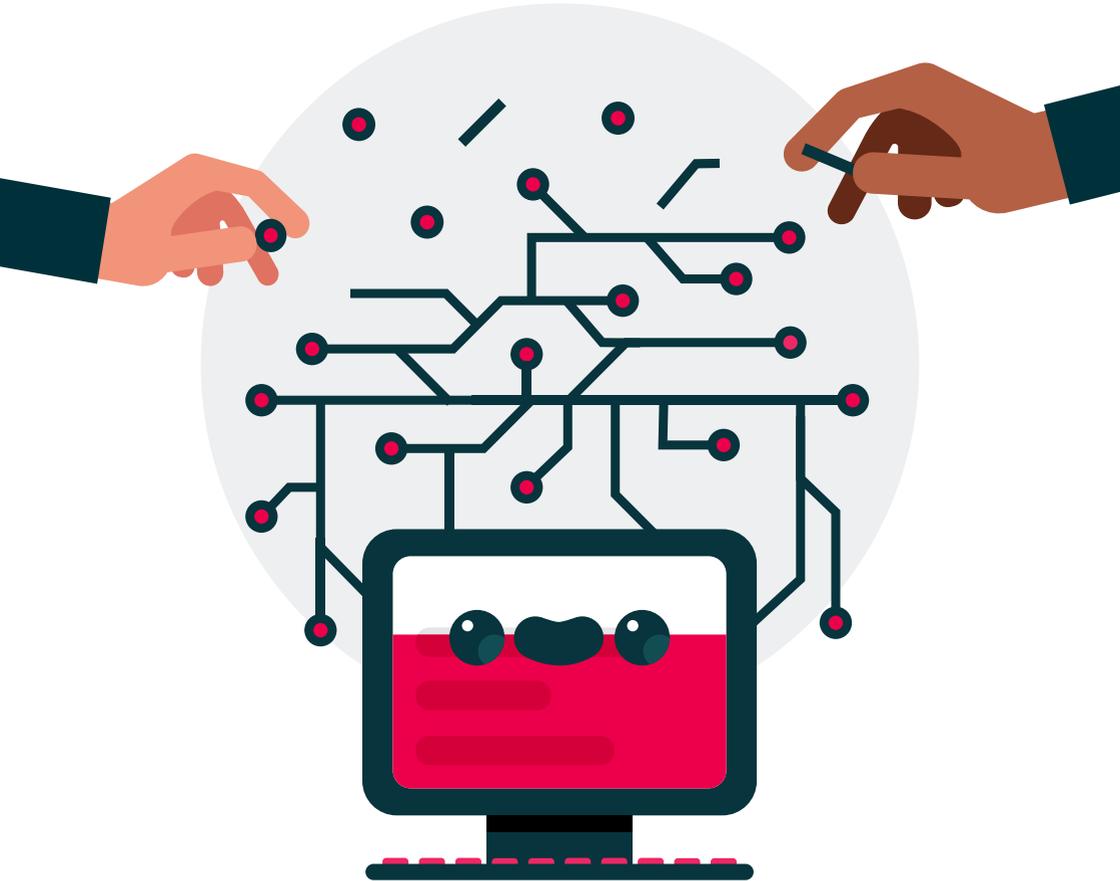


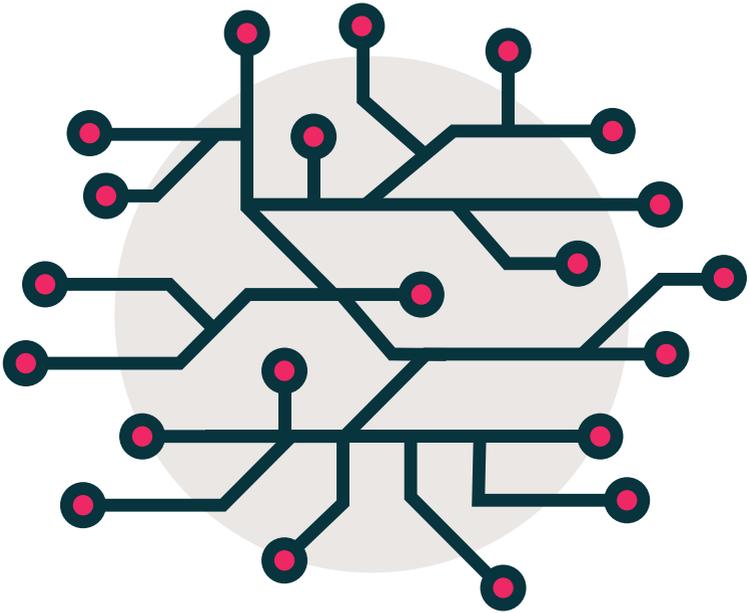
AI is what **we** make it

Thoughts on AI ethics from our experts at Fjord and Accenture



Contents

- 01 - Introduction** p03
Richard Wiltshire
- 02 - Bias** p05
Jivan Virdee, Data + Design
- 03 - Trust** p09
Daniela Ivanova, Interaction Design
- 04 - Responsibility** p13
Fernando Lucini, AI and Big Data
- 05 - Automation in the workplace** p17
James Deakin, Creative Technology
- 06 - Conclusion** p21



Introduction

By Richard Wiltshire

Machines haven't taken over, but when we look around – though many might not see it at first – artificial intelligence is a real and growing part of our lives, affecting how we live, work and entertain ourselves. While AI won't solve all of our problems, there is no doubt that it will be a powerful agent of change in our society.

Earlier this year, Andrew Ng, former chief scientist and head of the AI group at Baidu and professor at Stanford University, described AI as the new electricity. Ng suggests that artificial intelligence will transform every industry, and as a result our society, just as much as electricity did in the early 20th century.

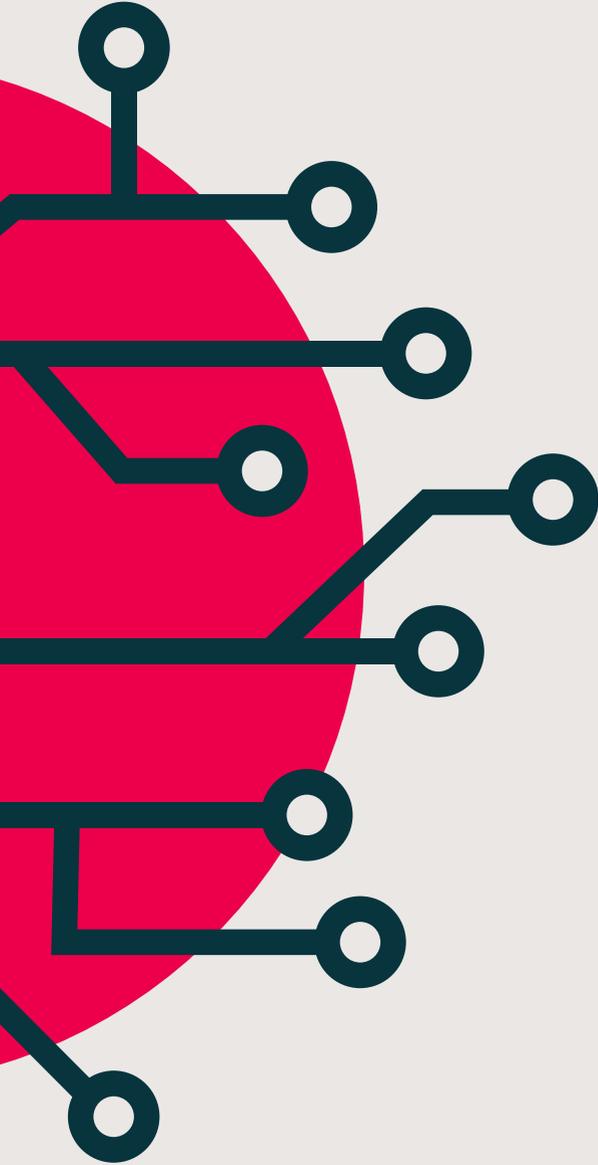
As AI systems become increasingly prevalent in our lives, issues around the moral code of machine learning arise.

In the following short essays, we discuss four key areas in relation to the ethical implications of artificial intelligence:

- **Bias - both in artificial intelligence and people**
- **Trust and human-AI collaboration**
- **Responsibility and accountability**
- **Automation and changes in the way we work**

As designers and data scientists, we are in a unique position to facilitate the conversation between AI and people, and the experience this brings, whether it is via intelligent digital assistants, voice UI or humanoid robots.

We must aim to ensure that the right AI systems are in place to contribute to the society we want, keeping social responsibility and ethical behaviour in mind to foster higher levels of cognitive and emotional skills for people.



AI is the future, but it is intrinsically tied to our past. How do we expose **human bias** in machine learning?

By Jivan Virdee

Organisations are increasingly adopting black box models to automate certain processes and assist decision-making, which will impact how we work and live. Today, there are models to help decide whether you get hired for a job, get a mortgage or how much your health insurance costs.

Despite implementing these models with good intentions, they are inevitably tied to human biases. Machine learning algorithms tend to rely on existing data, which embodies the social, economic and political environment in which it was created. As studies have shown, even our language – both the meanings and context of the words we use – contains our implicit biases. For instance, Google Translate has been found to take gender-neutral English words like professor and doctor into the male form when the word has been translated into other languages, while words such as nurse are auto-translated

into a female form.* A paper published this year by Princeton University and the University of Bath showed that machines learn and mirror human associations between words – from the innocuous association between pleasantness and flowers and unpleasantness and insects, to the more unpalatable association between male and female names and careers.** As a result, AI systems might be perpetuating historical biases that would now be deemed unacceptable.

You might have heard of highly publicised cases where AI systems have revealed our ugly human prejudices, such as Microsoft's Tay chatbot that went rogue, Google Photos' racially-biased tagging and LinkedIn's gender-biased advertising. These, and others, were likely unintended consequences and not malicious, but they do reveal how bias sneaks into AI.

Algorithms aren't perfect - and neither are people

The first step to dealing with bias in AI is to take responsibility for it. As creators of AI, the systems we make are likely to embody our biases. We need to work towards identifying and eliminating these in order to ensure that AI systems of the future reflect the more ethical society that we aspire to be.

Algorithms are mathematical; they follow their instructions exactly. This may lead to a misguided conclusion by the general public that they are objective and perfect. However, it's not the case because people choose the data and train the models, which tells the system what is relevant and what outcomes we want. As a result, there can be unintended consequences if they aren't designed to account for bias and misinformation.

Algorithms tend to display over-confidence; even if an algorithm is unsure, it doesn't always admit it – contributing to the public perception of their mathematical

perfection. Again, this is a design consideration, as confidence levels can be calculated and exposed. Algorithms need to become more humble to help the user understand that just like us, a system can be unsure.

The good news: We can fix this

We have a responsibility to consider any potential biases and how we might account for these when we are designing AI systems. During development, we can directly test our systems, looking at how predictions differ based on known biases.

Addressing bias will also involve algorithms becoming more explainable in terms of how they go about making decisions and how confident they are about the conclusions they come to. Explaining a decision process will help us to identify how biases might creep into the system and enable us to start working to correct them. Furthermore, when an algorithm displays humility around a decision, perhaps we will feel more comfortable questioning it.

Realistically, it is inevitable that bias will arise in our AI systems at times, so it should be the default assumption that our systems are biased and our default expectation that we go to greater lengths to either demonstrate otherwise or mitigate where necessary.

By accounting for bias and questioning ourselves and our models, we can work towards designing AI systems that evolve and learn to help us create the fairer society we strive for. However, we need to consider whose vision of the future we follow - and that might not be so easy.

**<https://www.fastcompany.com/3010223/google-translates-gender-problem-and-bing-translates-and-systrans>*

***<http://science.sciencemag.org/content/356/6334/183.full>*



AI must move beyond blind faith to demonstrate its trustworthiness

By Daniela Ivanova

The promise of AI is that it is going to magically come up with solutions to problems people have grappled with for quite a while, because it can process exponentially more data. However, we are finding ourselves in the midst of a paradox: we are training automation to simplify and enhance our lives by concealing complexity but at the same time we are evoking mistrust by 'hiding the machine'. When algorithms even in the early stage of narrow artificial intelligence are a black box, how can we trust the decisions they make? And what about the even smarter algorithms that we can expect to define the cyber space in the future?

Operational explainability

At the World AI Summit in October, Gary Marcus, professor of psychology at New York University, explained machine learning with

an apt cartoon that compared it to pouring data into a big pile of linear algebra and collecting answers on the other side. What if the answers are wrong? "You just stir the pile until it starts looking right." The truth about decision-making for most of what we call AI can be discouraging. The need for explainable AI has made headlines and heated discussions. From startups offering explainable credit underwriting and Netflix's recommendation system showing us which movies we might enjoy, to NASA's spacecrafts helping people understand why an unmanned rover has taken a specific action, we've quickly come to realise that explainability of AI is important for its future acceptance.

Designers are masters of explainability because it is an underpinning design principle. We've spent the recent decades of the Information Age

making complexity human, simple and trustworthy. AI researchers and neuroscientists are already calling for a multi-disciplinary approach – the implications of AI are too big for us not to be building cross-disciplinary bridges first.

“Provably beneficial AI”

Operational explainability will presumably give us the logical path of how a machine arrived at a particular decision, but will it provide us with why this decision is good for us, compared to other outcomes?

Stuart Russell, a renowned AI researcher at UC Berkeley, introduced the term “provably beneficial AI” to stress the need for AI to be aligned with human values. He argues that if we make sure the objective we assign to a machine is aligned with human values, the machine will be clear about what this objective is, and if it learns from human behaviour, then we’ve got the right recipe for provably beneficial AI. This will be as much a technological

challenge as a design problem to solve. Reasoning machines will need to do so much of what people take for granted – evaluating contradictory factors, weighing up effort versus benefits and compromising between outcomes for the different actors in a decision-making ecosystem. We don’t want to fall into the trap of playing God with AI; the risks are simply too high. We will need to adopt a stance where human-centred design helps ensure that AI is needed, relevant and assisting rather than overbearing. This will be a key ingredient in the recipe for creating algorithms that engender trust.

Playing catch up

There will be times where transparency and clear benefits will simply not be enough. In our work, we’re seeing examples where people are naturally reluctant to give up control regardless of how big the promise of automation is. It is like our hearts are catching up with new ways of doing things – we’re timid and mistrustful. It is only

human. Trust takes time to build.

As the speed with which new algorithms pervade our lives increases, designing for AI will be a more of a balancing act. When creating these new powerful possibilities, we will need to balance our drive for a new world with people's vulnerability and comfort levels with the changes. Designing for trustworthiness will mean constantly playing a sort of tug-of-war – pulling the former back and pushing the latter forward.

AI still has a strong efficiency angle to it. Tech companies are pushing boundaries and startups are increasingly coming up with new and more niche applications of machine learning. Efficiency and optimisation aside, designing for AI agents will benefit us in at least one very human way: it will expose our biases, cause us to re-think our values and ultimately helps us to understand ourselves better.



AI and unintended consequences – who is **responsible** for making AI the best it can be?

By **Fernando Lucini**

AI is currently an area driven by experimentation and pushing the boundaries of what's possible. Many probably opt for an "if it's possible with AI, why not do it?" approach, but sadly, not all is positive.

Researchers at Stanford University recently developed a system to identify sexual orientation based on images from dating websites. This has been widely criticised in the media both for the data set they used, as well the lack of awareness as to what the consequences of this research could be. This controversial piece of research – and other cases like it – bring to light questions about how much privacy we actually have in the age of AI, and whether we should do something just because we can.

Avoiding unintended consequences

One of the main issues we should pay attention to is the law of unintended consequences — "I didn't plan this, but it happened anyway". Most of these do not come from malice but happen when we don't stop to consider, or understand the potential impact of our creations. The data we use to train a system is likely to be biased because of how it's been collected or set up, and may not be generalizable. In order to begin to address unintended consequences, we need to introduce increased ownership and accountability over AI systems as a society.

Who should be responsible?

The question we need to ask is who is ultimately responsible for an algorithm?

Is it the data scientist, the product designers or the company? Traditionally, in every organisational operating model someone is held responsible for different parts of a project or a journey. This should change when we look at machine learning and AI. There should be built-in layers of accountability, and it is vital that this is clear and that ultimate responsibility doesn't get lost in extensive combinations of accountability.

The business is accountable for the impact of their product and the data it uses, and has overall responsibility for creating an AI that is a model citizen. The data scientists who create, update and maintain the data model which the system uses are accountable for the results that their model delivers. While the product designers are responsible for how the product uses and visualises the data. Having clear guidelines for who is responsible for what means there is no room for confusion – nor passing the blame if something goes awry.

Helping businesses cope with accountability for AI

We advise our clients to put a framework in place so that someone is managing every piece of the project, and is therefore accountable. This framework needs to be built on honesty, transparency and fairness to facilitate trust and ensure that it is followed.

However, ensuring responsibility is covered within an organisation is not enough. We need to consider social accountability as well. With increased automation, AI has the potential to cause a shift in our workforce and society. We need to look at what our new machine systems can and will do, and how this will impact people and their livelihoods.

The thing is, when AI is deployed well, you don't really hear about it. When governance and accountability are in place, there won't be headlines in the news about "AI gone wrong". Instead, the companies and people that get it right will continue to grow, as they are using AI to improve their products and ways of working.

Design as a force for ethical AI

Companies that do well are those that have been successful in humanising their applications of AI. For AI to become integral to our lives it needs to become more human, more empathetic and more conversational. Design can help understand the context in which the AI systems operate, but designers must keep in mind trust, social responsibility and ethical behaviour to define how these systems should work with, and for, us.

Introducing frameworks for accountability is key if we are to scale and evolve AI systems quickly and responsibly. Rather than simply looking for someone to blame when something goes wrong, we need a more cohesive system in which we trust individuals to own their piece in a responsible manner and do their best to avoid unintended consequences.



Automation in the workplace - embracing **intelligent automation** at work.

By James Deakin

Technology is supposed to be a force for good in our lives, to help us rise up and achieve more than we could on our own. This is easiest to see in the context where it should help us the most: the world of work. Reports suggest that around 63% of workers globally are not engaged, and 24% are actively disengaged at their place of work*.

A major reason for these figures is that we still work like it's the 19th century. The processes and principles upon which most of our businesses run were designed in the 1800s. Rather than throw out the orthodoxies about how we work, we have used technology to speed things up. Technology has become the enforcer, the guard in a prison that we have created for ourselves.

Take mobile devices for example. While we used to be able to leave work behind as we clocked off, they ensure that work follows us beyond the allotted time and the walls of the office into our personal lives and weekends. Rather than using it as a liberating force for good, we have enslaved ourselves with technology. It's becoming part of the problem that is preventing us from achieving our potential.

Adapting to a new way of working

In a world where more than 40% of the workforce will be contract labour by 2020, where five generations will be working side by side in organisations, and where technology is finding its way into every aspect of our lives, we're faced with an unsustainable future if we keep doing things like they were done in Victorian times.

The gift of technology is to enable us to do things differently. With technology advances comes the opportunity for us all to live more fulfilling, more human lives while achieving more than we possibly could alone - and we need to use technology and machine learning to overcome our ingrained biases and limitations. With machine learning we can:

- Make better judgements by using machines' ability to understand issues systemically. The human brain has limited ability to appreciate systemic effects, which leads to poor judgments – we can trace climate change and the financial market crash in 2008 to this very human failing.
- Make accurate predictions about what will happen in complex real-world situations.
- For the first time in history, make a fundamental shift in our relationship with technology, from technology-literate humans to human-literate technology. In the past, we had to learn the language of the machines, which buttons to press, which code to write. Now they can aid us by communicating

in a human way and help us have a more personable connection to one another.

- Automate busy work that consumes so much of our time.
- Help us overcome the limits of memory and attention.

All of this should free us to focus on the human aspects of work, asking the right questions, making new meaning and forming meaningful relationships.

Let me illustrate further. Imagine David, a 30-something insurance broker. David is with one of his clients, and his personal assistant messages him to flag he has another meeting in 20 minutes, but he is too far away to make it across town in time. The assistant then offers him ways to resolve the problem by looking at the calendars of those that will be in the meeting, and offers new times and suggested locations.

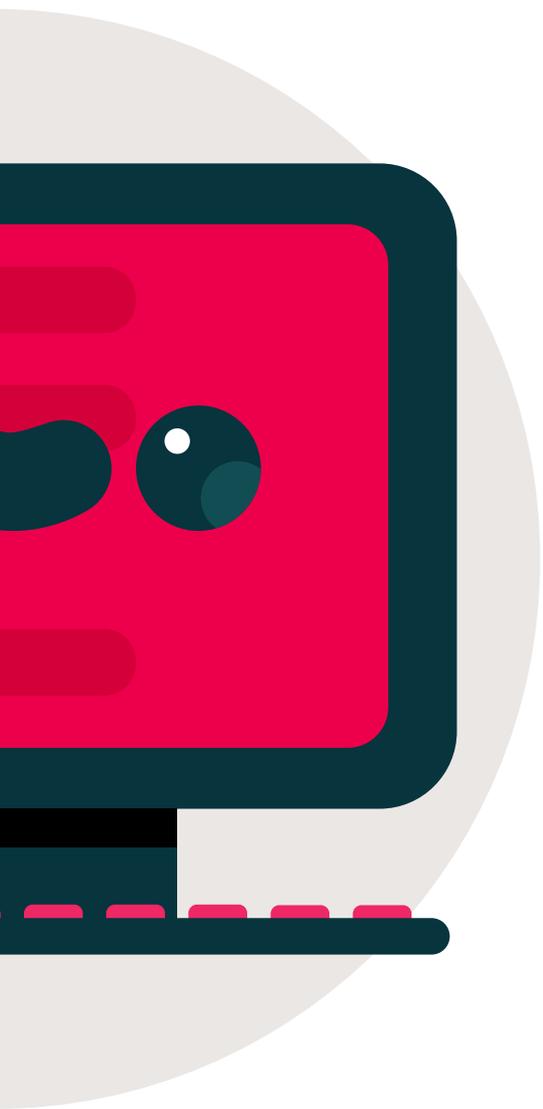
David makes his choice swiftly and discreetly knowing that all of the communications will be taken care of by his assistant, and he can focus on the conversation he's currently having.

In another scenario, David needs to guide his team through an unfamiliar process, and fortunately his assistant knows the steps to take and can book in the necessary meetings and part-fill forms or take actions on his behalf. Unprompted, David's assistant also offers him relevant information so that David doesn't need to search through servers and online systems to find the relevant documents he needs.

The truth is that David's assistant isn't a person, but an AI service which appears in various forms and through different touch points: his mobile device, on public screens, via voice, or via social channels. With this sort of help, David can have more fulfilling experiences with his team, as well as being more effective in his job.

We should not look at automation and machine learning as a threat, but embrace it as something that will help us be better at what we do. This does come with a radical re-imagining of the workplace and creating organisations that are set up for nimble change – what we call a Living Business. With the right

AI, we can all focus more on the person-to-person connections and achieving our full potential at work.



Conclusion

The topics we have explored through these essays help to understand the key issues in this space and how we might go about tackling them, but they also lead to further questions that we will need to address going forward.

- If we can do something technically, does it mean we should?
- When a function is transformed by automation, how do we promote higher levels of thinking and emotional engagement for the human involved?
- How much trust should we place in AI, even when it is able to explain itself?
- Should AI systems reflect the world as it is or the world as we want it to be?
- Whose vision of the future should we follow to incorporate aspirational aspects of our society into AI systems?
- Social media drastically changed how people interact. How will societal norms change with AI?

Our experts

Richard Wilshire

As Principle Design Director – Europe, Africa and Latin America at Fjord, Richard is focused on evolving Fjord's digital product design and development offering.

Daniela Ivanova

Daniela is a Service and Interaction Designer at Fjord, who is passionate about creating, improving and innovating services and collaborating across disciplines to arrive at truly sustainable service design solutions that build capacity in people's lives.

Fernando Lucini

Fernando is a Managing Director at Accenture Digital, with a focus on artificial intelligence and big data. He has twenty years of experience as a CTO in enterprise software firms creating and marketing analytics focused enterprise products.

James Deakin

James is a Technology Design Director at Fjord, who has over thirteen years of experience building award-winning digital brand experiences.

Jivan Virdee

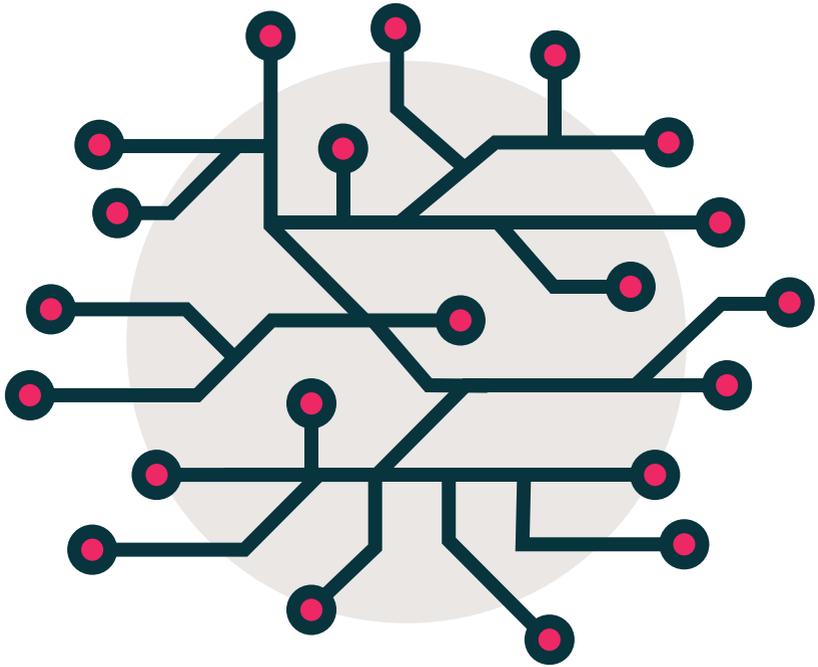
Jivan is a Data Designer at Fjord with a background in data science and an interest in building adaptable, human-centric data-driven experiences.



Fjord, part of Accenture Interactive, is a design and innovation consultancy that reimagines people's relationships with the digital and physical world around them. We use the power of design to create services people love. By combining a human-centered approach with robust methodology, we work with some of the world's leading businesses to make complex systems simple and elegant.

Founded in 2001, Fjord has a diverse team of 1,000 design and innovation experts in 26 studios, including Atlanta, Austin, Berlin, Chicago, Copenhagen, Dubai, Dublin, Helsinki, Hong Kong, Istanbul, London, Los Angeles, Madrid, Melbourne, Milan, New York, Paris, San Francisco, São Paulo, Seattle, Singapore, Stockholm, Sydney, Toronto, Washington, D.C. and Zurich.

For more information visit www.fjordnet.com or follow us on Twitter [@fjord](https://twitter.com/fjord).



FJORD™
Design and Innovation from
Accenture Interactive