McKinsey&Company

The ethics of artificial intelligence

January 2019

Executives and companies can enjoy the benefits of artificial intelligence while also being aware of potential drawbacks and taking careful steps to mitigate their effects.

In this episode of the *McKinsey Podcast*, Simon London speaks with McKinsey Global Institute partner Michael Chui and partner Chris Wigley about the ethical implications of artificial intelligence (AI) and how executives can engage more thoughtfully on the use of AI and its potential repercussions.

Podcast transcript

Simon London: Hello, and welcome to this edition of the *McKinsey Podcast*, with me, Simon London. Today we're going to be talking about the ethics of artificial intelligence. At the highest level, is it ethical to use AI to enable, say, mass surveillance or autonomous weapons? On the flip side, how can AI be used for good, to tackle pressing societal challenges? And in day-to-day business, how can companies deploy AI in ways that ensure fairness, transparency, and safety? To discuss these issues, I sat down with Michael Chui and Chris Wigley. Michael is a partner with the McKinsey Global Institute and has led multiple research projects on the impact of AI on business and society. Chris is both a McKinsey partner and chief operating officer at QuantumBlack, a London-based analytics company that uses AI extensively in his work with clients. Chris and Michael, welcome to the podcast.

Chris Wigley: Great to be here.

Michael Chui: Terrific to join you.

Simon London: This is a big, hairy topic. Why don't we start with the broadest of broad brush questions which is, "Are we right to be concerned?" Is the ethics of AI something—whether you're a general manager or a member of the public—that we should be concerned about?

Chris Wigley: Yes, I think the simple answer to this is that the concerns are justified. We are right to worry about the ethical implications of AI. Equally, I think we need to celebrate some of the benefits of AI. The high-level question is, "How do we get the balance right between those benefits and the risks that go along with them?"

On the benefit side, we can already see hundreds of millions, even billions of people using and benefiting from AI today. It's important we don't forget that. Across all of their daily use in search and things like maps, health technology, assistants like Siri and Alexa, we're all benefiting a lot from the convenience and the enhanced decision-making powers that AI brings us.

But on the flip side, there are justifiable concerns around jobs that arise from automation of roles that AI enables, from topics like autonomous weapons, the impact that some AI-enabled spaces and forums can have on the democratic process, and even things emerging like deep fakes, which is video created via AI which looks and sounds like your president or a presidential candidate or a prime minister or some kind of public figure saying things that they have never said. All of those are risks we need to manage. But at the same time we need to think about how we can enable those benefits to come through.

Michael Chui: To add to what Chris was saying, you can think about ethics in two ways. One is this is an incredibly powerful tool. It's a general-purpose technology—people have called it—and one question is, "For what purposes do you want to use it?" Do you want to use it for good or for ill?

There's a question about what the ethics of that are. But again, you can use this tool for doing good things, for improving people's health. You can also use it to hurt people in various ways. That's one level of questions.

I think there's a separate level of questions which are equally important. Once you've decided perhaps I'm going to use it for a good purpose, I'm going to try to improve people's health, the other ethical question is, "In the execution of trying to use it for good, are you also doing the right ethical things?"

Sometimes you could have unintended consequences. You can inadvertently introduce bias in various ways despite your intention to use it for good. You need to think about both levels of ethical questions.

Simon London: Michael, I know you just completed some research into the use of AI for good. Give us an overview. What did you find when you looked at that?

Michael Chui: One of the things that we were looking at was how could you direct this incredibly powerful set of tools to improving social good. We looked at 160 different individual potential cases of AI to improve social good, everything from improving healthcare and public health around the world to improving disaster recovery [Exhibit 1]. Looking at the ability to improve financial inclusion, all of these things.

For pretty much every one of the UN's Sustainable Development Goals, there are a set of use cases where AI can actually help improve some of our progress towards reaching those Sustainable Development Goals.

Simon London: Give us some examples. What are a couple of things? Bring it to life.

Exhibit 1

Artificial intelligence (AI) has broad potential across a range of social domains.

AI use cases per domain, number

Note: Our library of about 160 use cases is not comprehensive and will continue to evolve. This listing of the number of cases per domain should thus not be read as exhaustive.

McKinsey&Company | **Source:** McKinsey Global Institute analysis

Michael Chui: Some of the things that AI is particularly good at—or the new generations of AI are particularly good at—are analyzing images, for instance. That has broad applicability. Take, for example, diagnosing skin cancer. One thing you could imagine doing is taking a mobile phone and uploading an image and training an AI system to say, "Is this likely to be skin cancer or not?"

There aren't dermatologists everywhere in the world where you might want to diagnose skin cancer. So being able to do that, and again, the technology is not perfect yet, but can we just improve our accessibility to healthcare through this technology?

On a very different scale, we have huge amounts of satellite imagery. The entire world's land mass is imaged in some cases several times a day. In a disaster situation, it can be very difficult in the search for humans, to be able to identify which buildings are still there, which healthcare facilities are still intact, where are there passable roads, where aren't there passable roads.

We've seen the ability to use artificial-intelligence technology, particularly deep learning, be able to very quickly, much more quickly than a smaller set of human beings, identify these features

on satellite imagery, and then be able to divert or allocate resources, emergency resources, whether it's healthcare workers, whether it's infrastructure construction workers, to better allocate those resources more quickly in a disaster situation.

Simon London: So disaster response, broadly speaking—there's a whole set of cases around that.

Michael Chui: Absolutely. It's a place where speed is of the essence. When these automated machines using AI are able to accelerate our ability to deploy resources, it can be incredibly impactful.

Chris Wigley: One of the things that I find most exciting about this is linking that to our day-to-day work as well. So we've had a QuantumBlack team, for example, working with a city over the last few months recovering from a major gas explosion on the outskirts of that city. That's really helped to accelerate the recovery of that infrastructure for the city, helped the families who are affected by that, helped the infrastructure like schools and so on, using a mix of the kinds of imagery techniques that Michael's spoken about.

Also there's the commuting patterns—the communications data that you can aggregate to look at how people travel around the city and so on to optimize the work of those teams who are doing the disaster recovery.

We've also deployed these kinds of machine-learning techniques to look at things like, "What are the root causes of people getting addicted to opioids? And what might be some of the most effective treatments?" to things like the spread of disease in epidemiology, looking at the spread of diseases like measles in Croatia.

Those are all things that we've been a part of in the last 12 months, often on a pro bono basis, bringing these technologies to life to really solve concrete societal problems.

Simon London: The other thing that strikes me in the research is that very often you are dealing with more vulnerable populations when you're dealing with some of these societal-good issues. So yes, there are many ways in which you can point AI at these societal issues, but the risks in implementation are potentially higher because the people involved are in some sense vulnerable.

Michael Chui: I think we find that to be the case. Sometimes AI can improve social good by identifying vulnerable populations. But in some cases that might hurt the people that you're trying to help the most. Because when you're identifying vulnerable populations, then sometimes bad things can happen to them, whether it's discrimination or acts of malicious intent.

To that second level that we talked about before, how you actually implement AI within a specific use case also brings to mind a set of ethical questions about how that should be done. That's as true in for-profit cases as it for not-profit cases. That's as true in commercial cases as it is in AI for social good.

Simon London: Let's dive deeper on those risks then, whether you're in a for-profit or a not-for-profit environment. What are the main risks and ethical issues related to the deployment, AI in action?

Chris Wigley: One of the first we should touch on is around bias and fairness. We find it helpful to think about this in three levels, the first being bias itself. We might think about this where a data set that we're drawing on to build a model doesn't reflect the population that the model will be applied to or used for.

There have been various controversies around facial-recognition software not working as well for women, for people of color, because it's been trained on a biased data set which has too many white guys in it. There are various projects afoot to try and address that kind of issue. That's the first level, which is bias. Does the data set reflect the population that you're trying to model?

You then get into fairness which is a second level. Saying, "Look, even if the data set that we're drawing on to build this model accurately reflects history, what if that history was by its nature unfair?" An example domain here is around predictive policing. Even if the data set accurately reflects a historical reality or a population, are the decisions that we make on top of that fair?

Then the final one is [about whether the use of data is] unethical. Are there data sets and models that we could build and deploy which could just be turned to not just unfair but unethical ends? We've seen debates on this between often the very switched on employees of some of the big tech firms and some of the work that those tech firms are looking at doing.

Different groups' definitions of unethical will be different. But thinking about it at those three levels of, one: bias. Does the data reflect the population? Two: fairness. Even if it does, does that mean that we should continue that in perpetuity? And three: unethical. "Are there things that these technologies can do which we should just never do?" is a helpful of way of separating some of those issues.

Michael Chui: I think Chris brings up a really important point. We often hear about this term algorithmic bias. That suggests that the software engineer embeds their latent biases or blatant biases into the rules of the computer program. While that is something to guard against, the more insidious and perhaps more common for this type of technology is the biases that might be latent within the data sets as Chris was mentioning.

Some of that comes about sometimes because it's the behavior of people who are biased and therefore you see it. Arrest records being biased against certain racial groups would be an example. Sometimes it just comes about because of the way that we've collected the data.

That type of subtlety is really important. It's not just about making sure that the software engineer isn't biased. You really need to understand the data deeply if you're going to understand whether there's bias there.

Simon London: Yes, I think there's that famous example of potholes in Boston I think it was using the accelerometers in smart phones to identify when people are driving, do they go over potholes. The problem with that at the time that this data was collected is that a lot of the more disadvantaged populations didn't have smart phones. So there was more data on potholes in rich neighborhoods.[1]

Chris Wigley: There's a bunch of other risks that we also need to take into account. If the bias and fairness gives us an ethical basis for thinking about this, we also face very practical challenges and risks in this technology. So, for example, at QuantumBlack, we do a lot of work in the pharmaceutical industry. We've worked on topics like patient safety in clinical trials. Once we're building these technologies into the workflows of people who are making decisions in clinical trials about patient safety, we have to be really, really thoughtful about the resilience of those models in operation, how those models inform the decision making of human beings but don't replace it, so we keep a human in the loop, how we ensure that the data sources that feed into that model continue to reflect the reality on the ground, and that those models get retrained over time and so on.

In those kinds of safety critical or security critical applications, this becomes absolutely essential. We might add to this areas like critical infrastructure, like electricity networks and smart grids, airplanes. There are all sorts of areas where there is a vital need to ensure the operational resilience of these kinds of technologies as well.

Michael Chui: This topic of the safety of AI is a very hot one right now, particularly as you're starting to see it applied in places like self-driving cars. You're seeing it in healthcare, where the potential impact on a person's safety is very large.

In some cases we have a history of understanding how to try to ensure higher levels of safety in those fields. Now we need to apply them to these AI technologies because many of the engineers in these fields don't understand that technology yet, although they're growing in that area. That's an important place to look in terms of the intersection of safety and AI.

Chris Wigley: And the way that some people have phrased that, which I like is, "What is the building code equivalent for AI?" I was renovating an apartment last year. The guy comes around from the local council and says, "Well, if you want to put a glass pane in here, because it's next to a kitchen, it has to be 45-minutes fire resistant." That's evolved through 150, 200 years of various governments trying to do the right thing and ensure that people are building buildings which are safe for human beings to inhabit and minimize things like fire risk.

We're still right at the beginning of that learning curve with AI. But it's really important that we start to shape out some of those building code equivalents for bias, for fairness, for explainability, for some of the other topics that we'll touch on.

Simon London: Chris, you just mentioned explainability. Just riff on that a little bit more. What's the set of issues there?

[1] The Street Bump program is not in active use by the city of Boston.

Chris Wigley: Historically some of the most advanced machine learning and deep-learning models have been what we might call a black box. We know what the inputs into them are. We know that they usefully solve an output question like a classification question. Here's an image of a banana or of a tree.

But we don't know what is happening on the inside of those models. When you get into highly regulated environments like the pharmaceutical industry and also the banking industry and others, understanding how those models are making those decisions, which features are most important, becomes very important.

To take an example from the banking industry, in the UK the banks have recently been fined over 30 billion pounds, and that's billion with a B for mis-selling of [payment] protection insurance. When we're talking to some of the banking leaders here, they say, "Well, you know, as far as we understand it, AI is very good at responding to incentives." We know that some of the historic problems were around sales teams that were given overly aggressive incentives. What if we incentivize the AI in the wrong way? How do we know what the AI is doing? How can we have that conversation with the regulator?

We've been doing a lot of work recently around, "How can we use AI to explain what AI is doing?" The way that that works in practice we've just done a test of this with a big bank in Europe in a safe area. This is how the relationship managers talk to their corporate clients. What are they talking to them about?

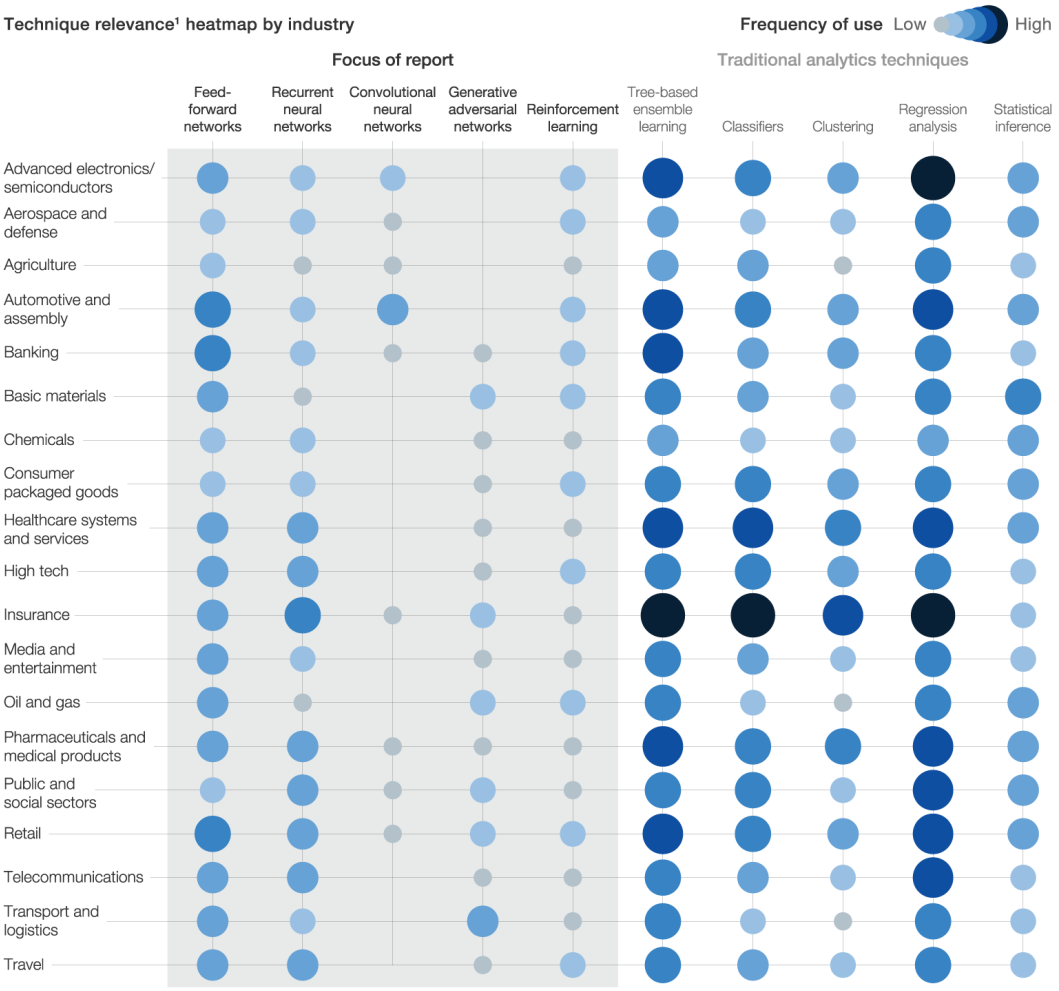
The first model is a deep-learning model, which we call a propensity model. What is the propensity of a customer to do something, to buy a product, to stop using the service? We then have a second machine-learning model, which is querying the first model millions of times to try and unearth why it's made that decision. [See Exhibit 2 for more on deep-learning models.]

It's deriving what the features are that are most important. Is it because of the size of the company? Is it because of the products they already hold? Is it because of any of hundreds of other features? We then have a third machine-learning model, which is then translating the insights of the second model back into plain English for human beings to understand. If I'm the relationship manager in that situation, I don't need to understand all of that complexity. But suddenly I get three or four bullet points written in plain English that say, "Not just here is the recommendation of what to do, but also here's why." It's likely because of the size of that company, of the length of the relationship we've had with that customer, whatever it is, that actually A) explains what's going on in the model and B) allows them to have a much richer conversation with their customer.

Just to close that loop, the relationship manager can then feed back into the model, "Yes, this was right. This was a useful conversation, or no, it wasn't." So we continue to learn. Using AI to explain AI starts to help us to deal with some of these issues around the lack of transparency that we've had historically.

Exhibit 2

Advanced deep learning artificial intelligence techniques can be applied across industries, alongside more traditional analytics.



[1]Relevance refers to frequency of use in our use case library, with the most frequently found cases marked as high relevance and the least frequently found as low relevance. Absence of circles indicates no or statistically insignificant number of use cases.
Note: List of techniques is not exhaustive.

McKinsey&Company | **Source:** McKinsey Global Institute analysis

Michael Chui: You could think about the ethical problem being, "What if we have a system that seems to work better than another one, but it's so complex that we can't explain why it works?" These deep-learning systems have millions of simulated neurons. Again trying to explain how that works is really, really difficult.

In some cases, as Chris was saying, the regulator requires you to explain what happened. Take, for example, the intersection with safety. If a self-driving car makes a left turn instead of hitting the brakes and it causes property damage or hurts somebody, a regulator might say, "Well, why did it do that?"

And it does call into question, "How do you provide a license?" In some cases what you want to do is examine the system and be able to understand and somehow guarantee that the technical

system is working well. Others have said, “You should just give a self-driving car a driving test and then figure out.” Some of these questions are very real ones as we try to understand how to use and regulate these systems.

Chris Wigley: And there’s a very interesting trade-off often between performance and transparency. Maybe at some point in the future there won’t be a trade-off, but at the moment there is. So we might say for a bank that’s thinking about giving someone a consumer loan, we could have a black-box model, which gets us a certain level of accuracy, let’s say 96, 97 percent accuracy of prediction whether this person will repay. But we don’t know why. And so therefore we struggle to explain either to that person or to a regulator why we have or haven’t given that person a loan.

But there’s maybe a different type of model which is more explainable which gets us to 92, 93 percent level of accuracy. We’re prepared to trade off that performance in order to have the transparency.

If we put that in human terms, let’s say we’re going in for treatment. And there is a model that can accurately predict whether either a tumor is cancerous or another medical condition is right or wrong. To some extent, as a human being, if we’re reassured that this model is right and has been proven to be right in thousands of cases, we actually don’t care why it knows as long as it’s making a good prediction that a surgeon can act on that will improve our health.

We’re constantly trying to make these trade-offs between the situations where explainability is important and the situations where performance and accuracy are more important.

Michael Chui: Then for explainability it’s partly an ethical question. Sometimes it has to do with just achieving the benefits. We’ve looked at some companies where they’ve made the trade-off that Chris suggested, where they’ve gone to a slightly less performant system because they knew the explainability was important in order for people to accept the system and therefore actually start to use it.

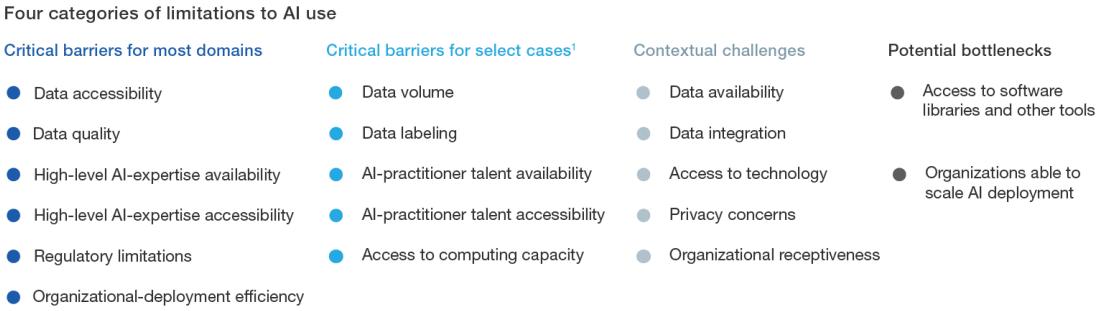
Change management is one of the biggest problems in AI and other technologies to achieve benefits [Exhibit 3]. And so explainability can make a difference. But as Chris also said, “That can change over time.” For instance, I use a car with [anti-lock] braking systems. And the truth is I don’t know how that works. And maybe earlier on in that history people were worried: “You’re going to let the car brake for itself.”

But now we’ve achieved a level of comfort because we’ve discovered this stuff works almost all the time. If we start to see that comfort change in an individual basis as well.

Simon London: I’m going to ask an almost embarrassingly nerdy management question now. Stepping away from the technology, what’s our advice to clients about how to address some of these issues? Because some of this feels like it’s around risk management. As you think about deploying AI, how do you manage these ethical risk, compliant risks, you could phrase it any number of different ways. What’s the generalizable advice?

Exhibit 3

Eighteen bottlenecks could limit artificial intelligence's (AI's) benefit to society.



Note: This list of bottlenecks was derived from interviews with social-sector experts and AI researchers and tested against our use cases.
[1]Bottlenecks that are critical for some domains as a whole or for individual use cases within those domains.

McKinsey&Company | **Source:** McKinsey Global Institute analysis

Michael Chui: Let me start with one piece of advice, which is as much as we expect executives to start to learn about every part of their business and maybe you're going to be a general manager, you're going to need to know something about supply chain, HR strategy, operations, sales and marketing. It is becoming incumbent on every executive to learn more about technology now.

To the extent to which they need to learn about AI, they're going to need to learn more about what it means to deploy AI in an effective way. We can bring some of the historical practices—you mentioned risk management. Understanding risk is something that we've learned how to do in other fields.

We can bring some of those tools to bear here when we couple that with the technical knowledge as well. One thing we know about risk management: understand what all the risks are. I think bringing that framework to the idea of AI and its ethics carries over pretty well.

Simon London: Right. So it's not just understanding the technology, but it's also at a certain level understanding the ethics of the technology. At least get in your head what are the ethical or the regulatory or the risk implications of deploying the technology.

Michael Chui: That's exactly right. Take, for example, bias. In many legal traditions around the world, understanding that there are a set of protected classes or a set of characteristics around which we don't want to actually use technology or other systems in order to discriminate.

That understanding allows you to say, "Okay, we need to test our AI system to make sure it's not creating disparate impact for these populations of people." That's a concept that we can take over. We might need to use other techniques in order to test our systems. But that's something we can bring over from our management practices previously.

Chris Wigley: As a leader thinking about how to manage the risks in this area, dedicating a bit of head space to thinking about it is a really important first step. The second element of this is bring someone in who really understands it. In 2015, so three years ago now, we hired someone into QuantumBlack who is our chief trust officer.

No one at the time really knew what that title meant. But we knew that we had to have someone who was thinking about this full time as their job because trust is existential to us. What is the equivalent if you're a leader leading an organization? What are the big questions for you in this area? How can you bring people into the organization or dedicate someone in the organization who has that kind of mind-set or capabilities to really think about this full time?

Michael Chui: To build on that, I think you need to have the right leaders in place. As a leadership team, you need to understand this. But the other important thing is to cascade this through the rest of the organization, understanding that change management is important as well.

Take the initiatives people had to do in order to comply with GDPR. That's something that again I'm not saying that if you're GDPR compliant, you're ethical, but think about all the processes that you had to cascade not only for the leaders to understand but all of your people and your processes to make sure that they incorporate an understanding of GDPR.

I think the same thing is true in terms of AI and ethics as well. You think about everyone needs to understand a little bit about AI, and they have to understand, "How can we deploy this technology in a way that's ethical, in a way that's compliant with regulations?" That's true for the entire organization. It might start at the top, but it needs to cascade through the rest of the organization.

Chris Wigley: We also have to factor in the risk of not innovating in this space, the risk of not embracing these technologies, which is huge. I think there's this relationship between risk and innovation that is really important and a relationship between ethics and innovation. We need an ethical framework and an ethical set of practices that can enable innovation. If we get that relationship right, it should become a flywheel of positive impact where we have an ethical framework which enables us to innovate, which enables us to keep informing our ethical framework, which enables us to keep innovating. That positive momentum is the flip side of this. There's a risk of not doing this as much as there are many risks in how we do it.

Simon London: Let's talk a little bit more about this issue of algorithmic bias, whether it's in the data set or actually in the system design. Again very practically, how do you guard against it?

Chris Wigley: We really see the answer to the bias question as being one of diversity. We can think about that in four areas. One is diversity of background of the people on a team. There's this whole phenomenon around group think that people have blamed for all sorts of disasters. We see that as being very real. We have 61 different nationalities across QuantumBlack. We have as many or more academic backgrounds. Our youngest person is in their early 20s. Our oldest person in the company is in their late 60s. All of those elements of diversity of background come through very strongly.

We were at one point over 50 percent women in our technical roles. We've dropped a bit below that as we've scaled. But we're keen to get back. Diversity of people is one big area.

The second is diversity of data. We touched on this topic of bias in the data sets not reflecting the populations that the model is looking at. We can start to understand and address those

issues of data bias through diversity of data sets, triangulating one data set against another, augmenting one data set with another, continuing to add more and more different data perspectives onto the question that we're addressing.

The third element of diversity is diversity of modeling. We very rarely just build a single model to address a question or to capture an opportunity. We're almost always developing what we call ensemble models that might be a combination of different modeling techniques that complement each other and get us to an aggregate answer that is better than any of the individual models.

The final element of diversity we think about is diversity of mind-set. That can be diversity along dimensions like the Myers-Briggs Type Indicator or all of these other types of personality tests. But we also, as a leadership team, challenge ourselves in much simpler terms around diversity. We sometimes nominate who's going to play the Eeyore role and who's going to play the Tigger role when we're discussing a decision. Framing it even in those simple Winnie the Pooh terms can help us to bring that diversity into the conversation. Diversity of background, diversity of data, diversity of modeling techniques, and diversity of mind-sets. We find all of those massively important to counter bias.

Michael Chui: So adding to the diversity points that Chris made, there are some process things that are important to do as well. One thing you can do as you start to validate the models that you've created is have them externally validated. Have someone else who has a different set of incentives check to make sure that in fact you've understood whether there's bias there and understood whether there's unintended bias there.

Some of the other things that you want to do is test the model either yourself or externally for specific types of bias. Depending on where you are, there might be classes of individuals or populations that you are not permitted to have disparate impact on. One of the important things to understand there is not only is race or sex or one of these protected characteristics—

Simon London: And a protected characteristic is a very specific legal category, right? And it will vary by jurisdiction?

Michael Chui: I'm not a lawyer. But, yes, depending on which jurisdiction you're in, in some cases, the law states, "You may not discriminate or have disparate impact against certain people with a certain characteristic." In order to ensure that you're not discriminating or having disparate impact is not only that you don't have gender as one of the fields in your database.

Because sometimes what happens is you have these, to get geeky, these co-correlates, these other things which are highly correlated with an indicator of a protected class. And so understanding that and being able to test for disparate impact is a core competency to make sure that you're managing for biases.

Chris Wigley: One of the big issues, once the model is up and running, is, "How can we ensure that while we've tested it as it's being developed, that it maintains in operation both

accuracy and not being biased." We're in the reasonably early stages of this as an industry on ensuring resilience and ethical performance in production.

But some simple steps like, for example, having a process check to say, "When was the last time that this model was validated?" It sounds super simple. If you don't do that, people have very busy lives, and they can just get overlooked. Building in those simple process steps all the way through to the more complicated technology-driven elements of this.

We can actually have a second model checking the first model to see if it's suffering from model drift, for example. And then translate that into a very simple kind of red, amber, green dashboard of a model in performance. But a lot of this still relies on having switched-on human beings who maybe get alerted or helped by technology, but who engage their brain on the topic of, "Are these models, once they're up and running, actually still performant?"

All sorts of things can trip them up. A data source gets combined upstream and suddenly the data feed that's coming into the model is different from how it used to be. The underlying population in a given area may change as people move around. The technologies themselves change very rapidly. And so that question of how do we create resilient AI, which is stable and robust in production, is absolutely critical, particularly as we introduce AI into more and more critical safety and security and infrastructure systems.

Michael Chui: And the need to update models is a more general problem than just making sure that you don't have bias. It's made even more interesting when there are adversarial cases. When in fact just to say, for instance, you have a system that's designed to detect fraud. People who are fraudulent obviously, don't want to get detected. So they might change their behavior understanding that the model is starting to detect certain things.

And so again, you really need to understand when you need to update the model whether it's to make sure that you're not introducing bias or just in general to make sure that it's performing.

Chris Wigley: There's an interesting situation in the UK where the UK government has set up a new independent body called the Centre for Data Ethics and Innovation that is really working on balancing these things out. How can you maximize the benefits of AI to society within an ethical framework?

And the Centre for Data Ethics and Innovation, or CDEI, is not itself a regulatory body but is advising the various regulatory bodies in the UK like the FCA, which regulates the financial industry and so on. I suspect we'll start to see more and more thinking at a government and inter-government level on these topics. It'll be a very interesting area over the next couple of years.

Simon London: So AI policy broadly speaking is coming into focus and coming to the fore and becoming much more important over time.

Michael Chui: It is indeed becoming more important. But I also think that it's interesting within individual regulatory jurisdictions, whether it's in healthcare or in aviation, whether it's what happens on roads, the degree to which our existing practices can be brought to bear.

So again as I said, are driving tests the way that we'll be able to tell whether autonomous vehicles should be allowed on the roads? There are things around medical licensure and how is that implicated in terms of the AI systems that we might want to bring to bear. Understanding that tradition and seeing what can be applied to AI already is really important.

Simon London: So what is the standard to which we hold AI? And how does that compare to the standard to which we hold humans?

Michael Chui: Indeed.

Chris Wigley: Absolutely. In the context of something like autonomous vehicles, that's a really interesting question. Because we know that a human population of a certain size that drives a certain amount is likely to have a certain number of accidents a year. Is the right level for allowing autonomous vehicles when it's better than that level or when it's better than that level by a factor of ten?

Or do we only allow it when we get to a perfect level? And is that ever possible? I don't think that anyone knows the answer to that question at the moment. But I think that as we start to flesh out these kinds of ethics frameworks around machine learning and AI and so on, we need to deploy them to answer questions like that in a way which various stakeholders in society really buy into.

A lot of the answers to fleshing out these ethical questions have to come from engaging with stakeholder groups and engaging with society more broadly, which is in and of itself an entire process and entire skill set that we need more of as we do more AI policy making.

Simon London: Well, thank you, Chris. And thank you, Michael, for a fascinating discussion.

Michael Chui: Thank you.

Chris Wigley: It's been great. □

Michael Chui is a partner of the McKinsey Global Institute and is based in McKinsey's San Francisco office. **Chris Wigley** is a partner in the London office. **Simon London**, a member of McKinsey Publishing, is based in McKinsey's Silicon Valley office.