Actuaries Institute Podcast
Transcript – Ethics Beyond Human Behaviour

Date: 17 August 2018
Interviewer: Anthony Tockar
Guest: Tiberio Caetano
Duration: 23:00min

Anthony: Hello and welcome to your Actuaries Institute podcast. I'm Anthony Tockar, Director at Verge Labs and Curator of the Data Digital Track at the upcoming Actuaries Institute General Insurance Seminar. The rise in data and artificial intelligence as a tool in business and our personal lives raises questions around trust and public interest.

Humans act under the law and their own moral code but what is the moral philosophy beyond human behaviour? What is the agency of AI's, cyber systems and networks and markets of cyber agents?

Today I'm lucky enough to have someone who is published extensively at top machine learning academies and thinks deeply on these issues, Tiberio Caetano.

Welcome Tiberio.

Tiberio: Thank you for having me Anthony

Anthony: Apart from being the Co-Founder and Chief Scientist at Ambiata, one of the most respected data science companies in Australia, Tiberio is a keynote speaker at the General Insurance Seminar in November. We're very excited to hear you speak, Tiberio, especially about the program you're currently working on which has the practical goal of equipping AI and cyber systems with the capacity of ethical conduct towards humans.

Now that raises all sorts of questions that sci-fi movies have tried to capture for decades but we as actuaries, understanding the building blocks of machine learning, can perhaps step away from the hype and perceive this concept differently. So, in this episode we're talking about the capacity of machines to behave ethically.

Why this is important, for businesses in general insurance, but also more broadly, and what we know and are striving towards in this field. So, to start us off Tiberio, can you tell us what comes to mind when you hear the term AI?

Tiberio: I think many things. One is I'd say fear, the other is excitement, but probably the most significant one is confusion. I think there's way too much confusion around that expression and hopefully today we'll be able to dive into that a little bit and try to take care of a little bit of that confusion.

Anthony: Yeah I know what you mean, so I guess turning it on to ethics a little bit, so Google of course, famously announced recently, seven guiding ethical objectives for their AI products and research. Why do you think ethical AI has suddenly risen to prominence?

Tiberio: I think perhaps one of the reasons is that those companies have started to actually pay the price of not being ethical, but not in a way that they are obtaining consequences of

deliberate unethical actions, quite contrary, they have systems and engineering that inadvertently are creating unethical outcomes for society and what you notice is that the reason why we notice that those outcomes are unethical is because the decisions that those systems actually end up making are not aligned with what a human would consider ethical and we think we are ethical, so that's probably one of the reasons why the companies have become more and more aware of this issue.

Anthony: So, on the topic of ethics in particular, we've spoken before about this topic and I know that you've identified really two problems within ethics, so the valuation problem and the causation problem, so I'll let you elaborate more but can you tell us a bit more about this?

Tiberio: Yeah this is a way that I like to think about the problem. The problem of ethics, it's a very complex problem that dates back to at least Aristotelic times. Many, very intelligent people have reflected on this and have pondered on the ethics problem, not only from a philosophical perspective but in practice. Businesses considered ethical questions on a daily basis and there are ethical trade-offs that any business has to face on a daily basis when they are dealing with major decisions. Now, I think it's very helpful to make progress in this conversation to create a demarcation between two subproblems. One I would call the valuation problem, as you said and the other, the causation problem.

One way to understand the valuation problem, it's really about what's the desirability of a given world? a world in which everyone is suffering, everyone is in hunger is, oh well that's probably objectively worse than a world in which everyone is flourishing right? so you can conceive of many different worlds and each of those worlds you can conceive, has some inherent ethical value in so far as its desire to a different extent right? some worlds are objectively better than others, it's not really a matter of opinion whether you think that everyone going through the experience that most Kenyans' are going at the moment versus everyone going through the experience that most Norwegians' are going through at the moment. If you think that the Kenyans are better off, well I'm sorry but I'm not going to take you seriously, so ethics in that sense isn't an objective question and morality is very much an objective, much more objective than people think it is. Now, so valuation is about attributing a value, a desirability score so to speak in language of actuaries to each possible world you can imagine.

Now causation is a very different problem. Causation is the problem okay. Let's assume that we knew the desirability score of every possible world okay, given the world we are in now what do we actually have to do in order to migrate to a different world with a better desirability score? so that's the causation question and why causation? because causation is about what do I have to do in order to obtain the result that I want? so the causation is more like a navigation problem. Once you have a map, you have to figure out well what do I do now to get to my destination? I turn right and turn left and so on. It's really useful to separate those two problems because a lot of the confusion that happens in ethical debates, in my view, they arise from the fact that people have different ideas about what different ways to achieve the same final objective as opposed to what they actually believe is the best final objective.

Anthony: Sure, well that certainly all makes sense. One thing I was discussing recently was I guess around how ethics does to an extent, change over time not to I guess can't tell what you're saying about the valuation problem but certainly over time things like privacy for example even within the same time in certain communities, it's seen very differently certainly you know in modern communities and cities and all that we all live in our little worlds and we don't want people to maybe find out about more intimate details but in other places is different. So how do you think in terms of that navigation, we can address that in a useful way?

Tiberio: Anthony you raise a really interesting point. You're talking about what philosophers would say, descriptive ethics, the two ways for (us) to conceptualize ethics in a different way, in a different dimension as to valuation and causation, so one is called descriptive ethics and the other is normative ethics. Descriptive ethics is more about what do people think is right and wrong? and if you ask this question here in Australia or in Sydney, you get one answer on average. If you ask this question in Afghanistan ever you get a different answer or Norway different answer. Different cultures and different times, they will give you different answers to what is right and what's wrong, that's fine. That's pure result of the fact that we are evolved creatures that evolved it independently in different parts of the world after the departure from Africa.

Now that doesn't tell us in fact anything about what intrinsically is good feelings or good sensations. For instance what I mean is over time we have noticed that even though these cultures are very different in what they consider right and wrong. There has been a convergence to a large extent of that notion, for instance today, most people believe that it's not acceptable to shoot someone who is walking on the other side of the street. In fact we have institutions that highly discourage that action right whereas not long ago, that was very acceptable in many societies right and just a few decades ago or even today in many cultures, there are certain actions towards women that were tolerated. Today, it's very different and there is more convergence about the idea that of human rights or the fact that women and men should be treated on the same ground, so there seems to be overall an evolution of morality towards what I would say some... I don't want to deliberately use the word universal morality but certainly a more consistent way to identify what's actually good and bad and that is what we would call more normative ethics. Normative ethics would look into the question what should we actually do in order to improve the world in an objective way?

Anthony: Yes absolutely. I totally agree and I think you know coming into this from the idea of writing code, algorithms etc to help improve this problem, we can use kind of some of these maybe we're gonna call them axioms but guidelines for fundamental human rights and things like that, so to turn the conversation a bit towards algorithms and I guess machine learning and data science etc within that world how do you see some of these ethical problems manifest?

Tiberio: Wow, they are manifested in many different ways so let me perhaps give you a couple of examples. Let's consider if you are if you are a machine learning practitioner or if you are an actuary, if you're trying to understand risk. You have a scenario where you have (a) history of events like incidents that happened, for example let's talk about insurance, you have instances of claims of certain instances you've had in the past and you want to use that

data in addition to demographic data, in addition to other kinds of behavioural data to estimate the likelihood that a certain person is going to claim again in the future, what's the risk of that person in the future? Fundamental questions associated with insurance. Now there's a purely technical question which is how do you estimate that risk as well as possible in the future? that's a purely technical question and so far it's not really an ethical consideration.

Now that's fine, so let's say we do that and we use our machine learning algorithms, the best machine learning algorithms you can have. You collect your data, you develop your models and you predict certain risk scores that are different for different customers of an insurance company for instance. Now the real important question now is what are you going to do with that knowledge? That is the ethical question because ethics is the philosophy of conduct. It's normative ethics, how do you norm your behavior? what do you do with what you know? It's not about knowledge per se, it's more like what you do with what you know? I mean okay now I know that this is the risk.

This is my best prediction for the risk of this of claims or future claims for this particular customer, now what should I do about it? well the most naive version would say well you should price your insurance according to that risk, you should look at that risk, you should put some profit margin on top of it and take care of some costs and price your insurance policy according to that risk within regulatory constraints yeah? We know that there are regulatory constraints even though the insurance industry is exempt from many regulatory and discrimination acts, they're still constrained. That's fine up until now, it's all good but here you start really looking to certain complex questions right? because what determines your risk?

Many things determine your risk. Certain things like your driving behaviour determine your risk for car insurance. Where you live is a good predictor of your risk but look where you live is also correlated with your race and should you take your race into account when actually pricing your risk? Then you start to look into those questions and you start to become confused very quickly (as I said confusion right) because you look at well is it fair to price the risk of someone based on what people like them did in the past? it seems like fair but maybe the reason why your risk is high. It's not because you're genuine autonomous behavior is suggesting that it should be high but because you ended up born in scenario in the contingency, in a society, in a particular neighborhood that would necessarily push you towards a kind of behavior that would lead to high risk to begin with and that has nothing to do with your agency. Now immediately you see that maybe it's not that simple and those questions start to permeate the thought of people who are reflecting hard about these questions, so this is a particular serious example.

Anthony: Yeah, look I really like the idea of this normative ethics in which you have the knowledge at hand. I've certainly heard arguments and one can easily think of areas within actuarial work where you build a model and there's an argument sometimes to go for weaker models because you're less likely to make those kinds of errors so you're almost limiting the knowledge that you get so that the decision you make is less likely to be an adverse one and I've known that's never really sat that well with me.

I feel like you should use the data at hand, you should use the modeling and then it's what you do with it that is essentially where the ethics and the fairness comes in.

Tiberio: I think that's a great point. I think it's really important that people separate knowledge from action because if you start from the assumption that you want less knowledge well part of that knowledge that you've given up is possibly knowledge that will help you to take better actions, so it's not necessarily a great strategy.

Anthony: Sure, so I wanted to ask I guess pragmatically, if you're an analyst, an actuary, a data scientist today, what actions can you take to improve your ethical decision-making?

Tiberio: Right, quite a few but I think we need to focus on very few. I think number one you need to understand that ethics is the norm is how you normally conduct your behavior. Every action that you take needs in principle some ethical governance or some ethical guideline or some ethical justification. Of course, we can't be thinking about this all the time but we need to just permeate our minds with this idea that ethics is not a discipline that you go and go to University and you sit there for the semester and you learn a few things and you never look at it again okay. Ethics is a way of life.

Ethics is is just how you should behave and and that's not really arbitrary. It's not really up to you (or) up to everyone else to choose if you are going to be killing everyone on the street. If you're going to be exploiting minorities, it doesn't work that way. We need to actually learn what actually is good for people and that's about knowledge so I would say;

1. Start respecting that word, ethics and don't interpret it as mushy do Buddhism as many people do. Respect that's discipline in philosophy that has been around for thousands of years

2. Do not separate ethical behavior from knowledge, that's crucial. Many people out there call themselves ethical, claim that are behaving ethically, whereas they have a great valuation function but they don't actually know what they're doing.

The causation components are all sorted out. I mean as I've said this numerous times I'd love to give this example of witch-hunt you know during witch hunts yeah people were you know committing those witches to the flames and in fact the evaluation function was perfectly reasonable.

They wanted to they prefer to kill for a woman that were dancing counterclockwise then letting thousands of people died because of plague because they thought it was because of those women, that the plague was coming right so they were just conducting perfect utilitarian calculus under the wrong assumption that the cause of the massive death was for women dancing in a particular way, so it's important on the side of ethics is about intent and about competence. It's about understanding the world as it is not as we would like it to be, as it is. So that's where science comes into ethics. So I'd say for science for data scientists, for practitioners, actuaries, respect the word ethics. Go back to philosophy as the tradition but associate that with knowledge. It is about knowledge. Get to know what actually is good for people and start behaving accordingly.

Actuaries Institute Podcast
Transcript – Ethics Beyond Human Behaviour

Anthony: Thank you, that's really fascinating and I think any of the listeners can certainly take away a few messages and a few actions that can apply to our careers so that's about all we have time for, for this episode, so thank you very much for joining us Tiberio. It's really great to hear from someone with your experience and someone who's kind of looked at this topic so closely and it's kind of so close to the latest thinking as well so thank you.

Tiberio: It's been my great pleasure Anthony, thank you very much.

Anthony: So we hope you enjoyed the discussion. Look out for the registration for the General Insurance Seminar. You can find out more details on the Actuaries website of course and Tiberio will be speaking along with a range of other thought leaders across General Insurance, so please get in touch with us with any comments about today's conversation and let us know what topics you'd like covered in your Actuaries Podcast. I'm Anthony Tockar, bye for now.