COMMUNICATING THE SOCIAL IMPACTS OF AI

IN CONVERSATION WITH NAT KENDALL-TAYLOR



We interviewed Dr. Nat Kendall-Taylor, CEO, FrameWorks, for the Response-ability.tech podcast. Nat received his PhD in Anthropology at UCLA and he joined the FrameWorks Institute in Washington, D.C. in 2008. During our conversation Nat shared the findings from a project, commissioned by the MacArthur Foundation, on how to frame and communicate the social impacts of artificial intelligence.

FrameWorks is a non-profit research organisation in Washington, D.C. that uses rigorous social science methods to study how people understand complex social issues such as climate change, justice reform, and the impact of poverty on early childhood development. FrameWorks is centrally concerned with how people use heuristics, mental models, and patterns of reasoning to make sense of social issues, and then developing evidence-based techniques that help researchers, advocates, and practitioners explain them more effectively.

Their project on how to frame and communicate the social impacts of artificial intelligence has shown, that to bring the general public along with us on these issues, rather than simply repeating the words and phrases "structural racism" or "artificial intelligence advances existing systems of structural racism", we should "unpack what that means, how that works, and give examples of issues where you can see that come into play".

As Nat says, much of this "is about examples and explanation", and he says their work "suggests that it is a responsibility, it's an obligation, of those who understand the process of how these things work to bring the public along, and to deepen people's understanding" of the social impacts of artificial intelligence.

What follows is an edited version of the part of our interview with Nat where we discuss the framing project. To listen to the whole interview, go to Apple Podcasts or Spotify or wherever you listen.

HOW FRAMEWORKS CAME TO WORK ON THE PROJECT

The project on how to frame and communicate the social impacts of artificial intelligence came to us through a ongoing long-term collaboration with the MacArthur Foundation, which is one of the larger foundations in the United States, based in Chicago.

The Foundation has an area of interest called technology and society. And so, artificial intelligence and particularly the social implications of that, particularly the ways in which artificial intelligence propagates and advances existing inequities across

Some of their Grantees working on issues such as algorithmic justice and facial recognition had been having a lot of difficulty in advancing their ideas in the public realm and running up against some walls in terms of getting their messages through.

social groups in the United States, is a commitment they have, a larger commitment to social and racial justice. They have a set of Grantees, really incredibly smart innovative folks, who are working on issues such as algorithmic justice and facial recognition. Those grantees — I think understandably to you and me and probably folks who are listening to this — had been having a lot of difficulty in advancing their ideas into the public square, had been running up

against some walls in terms of their ability to really get their messages through.

And so that tends to be the point at which folks come to us — when they've been hitting their head against a wall on communications for some time. They know they've got important things to say, they know they've got great data and research and findings, but they just can't get people to get it, to think about it, to consider it in supporting different solutions.

HOW PEOPLE UNDERSTAND AI AND ITS SOCIAL IMPLICATIONS

And so the MacArthur Foundation came to us and asked if we'd be interested in taking this on, from a framing perspective: How do people understand artificial intelligence and the social implications thereof. And then prescriptively, from a reframing perspective, how can those people who are working on these issues make informed choices about how they position their information to allow people to more openly and readily appreciate these social kind of equity components of the issue.

We started a project about a year ago, which has gone through (and is going through) the typical parts of the process that we use to do the research. So we started off with interviewing members of the sector, both Grantees and folks outside of their Grantee circle, people who are working on the social implications of artificial intelligence, to figure out: what were the core ideas, the information they wanted to more effectively put out there and get to land, cut through and be understood by people who

We had to figure out how normal people — people who don't live, eat, sleep, and breathe artificial intelligence, and the social implications thereof — think about these issues.

are not in the field of artificial intelligence. And so that was the first step.

The second step was then to figure out what they're up against, how normal people — people who don't live and eat and sleep and breathe artificial intelligence, and the social implications thereof — think about these issues. Again, from a cultural models perspective, what are the deep patterns of reasoning that either make it hard to appreciate this perspective on the issue or, in some cases, allow people to engage with it in helpful and meaningful ways.

That's the kind of point that we're at right now. There is a report that will shortly be published that documents those first two steps in our research process.

We're currently working with members of the field in the US — and this is one of the most fun parts of the process — to brainstorm a bunch of different framing hypotheses. For example, we think



New York City. Photo by Patrick Ho on Unsplash.

the value of justice is going to be really effective, and we think this particular metaphor is going to be important in helping people understand some of the problems with algorithms and the source of information upon which they draw to make predictions and decisions.

We'll then empirically test those [framing hypotheses] using a series of qualitative and quantitative methods, classically designed framing experiments, where you get large representative samples of Americans and randomly assign them to different treatment groups where they hear different frames. Then you can compare the effect that those frames have to discern what kind of 'frame effects', as they are called in our field, these different decisions are having.

And then we get to the the hardest but the most important part of the work, which is where you take all of that work that we will have done for a year

and a half, and try to put it in forms and products that members of the sector, given the fact that they are so unbelievably busy, can use in their work to inform the decisions they are making when it comes to framing.

The term artificial intelligence has come to encapsulate a tremendous range and breadth of technologies, some of which are not even, by definition, artificial intelligence. In the first stage of the project, when we were doing stakeholder interviews and a grounded theory consensus analysis, the decision was made to really focus on algorithms and their effect on three issues in particular: health and health care; child protection; and predictive policing.

The decision to focus on these areas was made in collaboration with members of the sector who thought that, wow, if we could really move some of our ideas on on that relatively narrow slice of the field that would be really valuable and powerful.

If the general public doesn't understand what the thing is that you are claiming has pernicious impacts on certain groups of people, then it becomes very hard to have a meaningful conversation about what those impacts are, who is affected how, and so forth. That's a major obstacle.

LACK OF UNDERSTANDING IS IMPEDING MEANINGFUL CONVERSATIONS

In the second stage, unfortunately, when it comes to artificial intelligence, a lot of those findings, in terms of how people think about these issues, reveal and draw into relief the obstacles the field faces and how they communicate. Some of them are fairly obvious, such as the public does not have a very good grasp of what artificial intelligence is and what it isn't.

And some of that is a product of the sector and the technology companies, how they communicate, and the kind of allure that they draw from the tag of artificial intelligence and the relatively loose application thereof on a whole bunch of things that are innovative or cutting-edge, but probably not definitionally artificial intelligence.

But if you've got an audience, the general public in this case, who doesn't understand what the thing is that you are claiming has pernicious impacts on certain groups of people, then it becomes very hard to have a meaningful conversation about what those impacts are, who is affected how, and so forth. That's a major obstacle.

More specifically, when you get into predictive algorithms, there's even less of an understanding about what predictive algorithms are. And there's a sense that artificial intelligence has things that you put into it and things that come out of it. But not a lot of sense in terms of the dirty data issue: that if you feed algorithms and they learn on data that is biased, then you create algorithms and predictions that are similarly biased. Which is obviously a major issue in the United States when it comes to predictive policing practices, that they are fed data that is not unbiased data about crime. It's data that's deeply biased by structural racism which then produces an algorithm that is structurally racist and then produces predictions which are equally racist.

So this lack of understanding about how that particular part of it works, assumptions about the fact that these algorithms are always correct, the inability to see that they are fallible in the same way, and that they do have components in which human error enters into the algorithms. And so, if you understand algorithms as just being neutral, then it becomes also very difficult to see how they could be advancing inequities and how they would need to be changed, altered, and used in different ways to avoid some of those exposures.

PEOPLE DON'T UNDERSTAND WHAT STRUCTURAL OR SYSTEMIC RACISM MEANS

There are things that go beyond this issue, but are certainly in place and a lot of what I'm about to say comes from work in the US so I don't know how internationally applicable it is. Over the last 18 months in the United States, since the murder of George Floyd, the words "structural racism" and "systemic racism" have become ubiquitous.

We've got another really interesting project called the culture change project, which is looking at how cultural models have or have not changed over the last 18 months in this country at a really deep level. So there's a correlation that people are drawing that they were previously not drawing to the same degree

between those words and terms and the idea of racism.

But what our research shows really clearly is it doesn't extend deeper than those words and terms. People don't really have a sense what structural or systemic racism means, people really don't understand, outside of a few issues, how that might work and what the outcomes of that might be. One of the exceptions we're seeing is policing, where there has been so much focus over the last 18 months, in driving home the idea that, when we see people being killed by police, that it's not just that individual police officer and their racial animus that is the problem. It is the systems of training, of recruitment, of incentives, of accountability that are at play. And so you see on a couple of issues this surface-level linguistic relationship actually deepening into some understanding of how it works.

There is a tremendous amount of work ahead of us in explanation. Not just repeating those words and phrases "structural racism" or "artificial intelligence practices advance existing systems of structural racism", but actually unpacking what that means,



Justice for George Floyd. Photo by Nathan Dumlao on Unsplash.

how that works, examples of issues where you can see that come into play.

But I think this — and this is one of our main recommendations from this culture change project which pertains to the artificial intelligence project — is that there is a tremendous amount of work ahead of us in explanation. Not just repeating those words and phrases "structural racism" or "artificial intelligence practices advance existing systems of structural racism", but actually unpacking what that means, how that works, examples of issues where you can see that come into play.

This is the work we have ahead of us in that project and why we selected those three systems in particular —health, child protection, and policing — because it's eminently more possible to explain artificial intelligence and its impacts on those particular issues than it is just to try to explain the social implications of artificial intelligence.

A lot of this is about examples and explanation.

It is a responsibility, it's an obligation, of those who understand the process of how these things work to bring the public along, and deepen people's understanding of how using algorithms can be problematic.

Which is good news because people who work in these sectors know the explanation, and I think, to be a little provocative, they've been under the illusion that they can't go deeper on the explanation because they think people don't and won't understand. I think our work throws that into question and actually suggests that it is a responsibility, it's an obligation, of those who understand the process of how these things work to bring the public along, and deepen people's understanding of how using algorithms to, for example, make resourcing decisions in police departments can be seriously problematic.

Follow **Nat Kendall-Taylor** on Twitter at @natkendallt and connect with him on LinkedIn. Visit the FrameWorks website for more information and follow FrameWorks on Twitter @FrameWorksInst.

Postscript, 21 October 2021: Read the FrameWorks Strategic Brief for this research project, published on October 19.

The Response-ability Summit, formerly the Anthropology + Technology Conference, champions the social sciences within the technology/artificial intelligence space. Sign up to our monthly newsletter and follow us on LinkedIn and Twitter. Subscribe to the Response-ability.tech podcast on Apple Podcasts or Spotify or wherever you listen. Watch the talks from our events on Vimeo.