AI ETHICS

Indigenous Data Sovereignty

By Maui Hudson, Te Kotahi Research Institute, University of Waikato

00:02
My name is Maui Hudson, I'm a member of the Factoria Nation and I am an associate professor at the University of Waikato. I'm really pleased to be a part of this program on AI Ethics: Providing a Global Perspective. And in particular, this lecture will be talking about indigenous data sovereignty. Just give me a moment to get my presentation up, and we will get underway. So, as I mentioned, this talk is going to be about indigenous data sovereignty. I'm based at the Te Kotahi, a Research Institute at the University of Waikato. And I've been involved with a number of organizations which I will talk about shortly. My affiliations to Waikato, which is a Iwi or a tribal nation out here in New Zealand. I'm a member of the tribal council, the Factoria Māori Trust Board and then negotiated for their tribal Treaty claims. In a professional capacity, I've been involved with co-authoring guidelines on Māori research ethics, genomic research with Māori and biobank with Māori. I'm a founding member of Te Mana Raraunga, which is the Māori data sovereignty network, as well as the Global Indigenous Data Alliance. And I've been a co-author of the key principles for Indigenous Data Governance, and a co-director of Local Contexts, which is one of the developers behind the bi-cultural labels, and a co-director of Enrich, which is an international network based around equity for indigenous research and innovation as a coordinating hub. So just to start off, I did want to talk about the indigenous data sovereignty in the context of an international environment. And this is a movement that has crossed as crossing the globe. There are a number of nation state based indigenous led communities of practice to manage the donor being one of those and in New Zealand. But there's also the US indigenous data sovereignty network. And Maiam Nayri Wingara, indigenous data sovereignty, Clicked Event Australia, as well as nascent movements in Canada and places like Sweden and Mexico. The Global Indigenous Data Alliance also draws in people from broader, broader set of countries. But it's primarily an international network and the Research Data Alliance hosted international indigenous data sovereignty interest group as well. So what are indigenous data - data information and knowledge in any format that impacts indigenous peoples, nations, and communities at collective and individual levels. This is a slide that was put together by the US Indigenous Data Sovereignty Network. And you'll see there's really these three particular categories. This data about indigenous resources and environments; there is data about indigenous people as individuals, and then there's data about the collectives, as nations and as peoples that spans, you know, a wide variety of data sets from traditional knowledge, right through to scientific data about geology, or plants, through to administrative data collected about tribal peoples. So digital status sovereignty is a new term. Data sovereignty comes from cloud computing. It's a term that's used to refer to data being subject to the laws of the nation within which it's stored. That indigenous data sovereignty flips that and talks about data being subject to the nation from which it's collected, and clearly includes tribal nations. So it's really a discourse about rights and interests in relation to data.

04:31
And draws on thinking that emerges from you know, indigenous rights, treaty rights, cultural intellectual property rights, as well as indigenous research ethics and data ethics. And there are two main themes. One is data for governance. So just thinking about access to data, and how that data can be used to support governance decisions in indigenous contexts. And then as data is available and accessible to everyone. And other people access

information about our nations, what's our role in terms of governance of data? So managing access to data so that it's used as relevant and responsive. And just on the right side here is the cover of really the first book that put together talking about indigenous data sovereignty towards an agenda which was released in 2016. And probably the strongest expression in terms of principles as it relates to indigenous data sovereignty, as reflected in the OCAP principles, and emerges from the First Nations Indigenous Governance Center in Canada, predates the discussions around indigenous data sovereignty but still reflects, you know, ideas of indigenous control of indigenous data by focusing on issues of ownership, control, access and position. And in the New Zealand context, we have created, to monitor and diagnose, created the Māori data sovereignty principles, taking six common valued values and concepts and giving them an expression in the context of data. So while Donnatella Tanner said relative autonomy relates to authority, but it's not necessary in the fullnes, the full meaning of the Maori word, but in this context, that's how it's being applied. And so, as an example, you can see here that annotator tongue authority control might have an inherent right to exercise control over Māori data and Māori data ecosystems. This right includes but is not limited to the creation, collection, access and analysis, interpretation, management, security, dissemination, use and reuse of Māori data. Jurisdiction, decisions about the physical and virtual storage of modern data should enhance control for current and future generations, wherever possible, Māori data should be stored in Altidore New Zealand. This is particularly relevant at the moment as people are talking about offshoring of data, or making use of cloud data storage options where the servers are based overseas. And then self-determination of the Māori right to data that is relevant and empower sustainable self-determination and effective governance. And if we think about the sort of data that's collected, and then what you're able to do with that, and then in an AI space, having the right sort of data collected becomes an important component. Similarly, if we think about Whakapapa, and in this context, how that relates to relationships and relationships with and through data. To put in one context, all data has a Whakapapa or genealogy. Accurate metadata should at minimum provide information about the provenance of the data, the purposes for its collection, the context of its collection, and the parties involved. Data disaggregation - the ability to disaggregate Māori data increases its relevance for Maori communities and Iwi. Māori data should be collected and coded using categories that prioritize modern needs and aspirations. Once again, thinking about how data has been classified, and future uses, current decision making over data can have long term consequences good and bad for future generations of Maori, that governance Maori data should be reasonable as to enable long term decision making in order to minimize future harm. So you can see here that each of these principles is thinking about different dimensions of data as it relates to a core indigenous concept. And this is important, in part because even amongst our own community, there's this challenge of competing interests. Because we both have support for open data and open science, things that push data into those spaces. And that comes about in part because of our increasing participation within STEM subjects and cutting-edge science and technology, as well as our aspirations for indigenous data sovereignty or greater indigenous control of indigenous data. So this goal around or a sort of a goal sits there around indigenous communities benefiting from innovation and development, that they're coming about via greater control of indigenous knowledge and indigenous data. And you see on the right-hand side here, the most recent book being released around indigenous data and policy, this came out in late 2020.

09:59

And discussions around indigenous data sovereignty have become more pressing and more relevant as we increasingly move into this era of open data, and big data and open science, and the use of machine learning and artificial intelligence to deal with, with the volume of data that's being digitized. And sitting behind that are these challenges where every indigenous community has, you know, large, enormous collections and data held in archives, museums, libraries, and repositories, then significant information about those collections, including

community names, and proper provenance is often missing. That those communities are largely not the legal rights holders and they're often excluded from discussions about how they should be managed. And issues of responsibility and ownership, as well as the mistakes that exist within the metadata are continuing from the physical environment into the digital environment as that material gets digitized. At the same time, we're not just dealing with historical collections, but the new material being generated through the researchers that are, you know, working on collecting data and samples from indigenous communities, and this is only increasing as that sort of activity increases. So taking a little bit of shift here, and really thinking about when that information, and you know, most indigenous data is sitting in institutions or agencies or other places outside the control, and often the purview of indigenous communities. So what sorts of indigenous frameworks can be put together, that give some influence or some enhance the ability for indigenous people to make decisions about what's happening with that data. And so, what we have here are the key principles for indigenous data governance, that were released in 2019. You can see at the bottom, the reference for the journal article, which talks about the development. The photo here is the group of people that were involved in the conceptualizing of the key principles. This was at a conference in Botswana. And the principles themselves promote collective benefit authority of control, responsibility and ethics as the primary principles for this framework, but that it has these sub-principles as well. So collective benefit, as in the context of inclusive development and innovation, for improved governance and citizen engagement and for equitable outcomes. And each of the principles has a similar set of sub-principles as well. And so the key principles for indigenous data governance in part arose from our evaluation of existing principles. And when we look at the fear principles, here fear principles of scientific data management, findable, accessible, interoperable and reusable, really talking about the characteristics of the data themselves. And so they operate as a data centric set of principles. And when we did an assessment of Māori data sovereignty principles, indigenous principles that are emerging in other contexts, and realized that they tend to talk more about the people and about the purpose. And so the fear principles on the key principles become complimentary. And that's how they've been discussed in international forum, particularly the Research Data Alliance, which has been responsible for operationalizing or implementing the principles and is now becoming a strong advocate for the development of criteria for the key principles, and that they be applied alongside fear. And we're also starting to see the key principles included in the policy documents. This is the Australian Institute for Aboriginal and Torres Strait Islander Studies (AIATSIS).

14:38

They have a code of ethics for Aboriginal Torres Strait Islander research, which has recently been updated. You'll see on the right-hand side, the areas which they think are particularly important, and in the context of this kind of research, and the responsibilities associated with it. And in the upper right-hand quadrant around indigenous leadership. This specific reference to indigenous knowledge and data. And if we move into that part of the document itself, you'll see a point there that says researchers must be aware of and apply the international data principles of fear and care.

15:22

And so thinking about this assertion of indigenous rights and interest, which is being made through indigenous data sovereignty, how indigenous peoples become involved in the governance of data, and indigenous data governance applies to, you know, this information that's sitting within other institutions or contexts, and then how people get to access that. So some spaces that will be a sort of open access protocol, anyone can go and access it, and other places that will be controlled, and it will be accessed on request. And so in what ways can indigenous knowledge inform those data access protocols. And the example which I'm going to talk to here comes out of

New Zealand and is based around Ngā Tikanga Paihere, which guides culturally appropriate and ethical data use in relation to the integrated data infrastructure or the IDI which is a database which combines administrative data from a variety of government institutions, which is linked, and then de-anonymized, or anonymized de-identified. And that is available to researchers to apply for access. And the frameworks started from some work that we did, which was looking at what Māori concepts that informed the sharing of traditional knowledge, and what we were interested in is to what degree might those concepts inform ways we might think about data access. And so the concepts on the left hand side here, you can see there's 10 of them that emerge through this work really relate to three particular areas. Some of them related to the assessment of the data, or the knowledge that was, you know, being requested or people were thinking about sharing part of it related to an assessment of what the purpose of the use of that data was, and part of the assessment related to the data users. And so on this diagram, the black and blue spirals relate to the assessment of the data. The green axis relate to the data use and the red axis relate to the data users. And we use these 10 concepts to create a set of questions that showed how they could be relevant to data. And so we gave each of the concepts a characteristic and we see them they're. Tapu, which normally means sort of something that's sacred or something that's special that that needs to be managed or restricted, related to the level of sensitivity. And alongside the characteristics that we created an assessment question. And the purpose of thinking about this in this way is so that we can look at a data set. And on balance, what level of management is required for the status here? And if most of the answers came up, and they were highlighted in the green spaces, then it was probably a data set that could sit in an open data environment. If it was orange, you'd be thinking about, well, we probably need some, some rules around it, but they might be sort of policy or practice related expectations or roles. And ones that were primarily in the red space could still be accessible, but would need to be accessible on request. And that framework or that set of questions was used as the basis for developing Ngā Tikanga Paihere. What we realized is that the 10 questions aligned quite nicely with the Five Safes framework. That was the framework which Statistics New Zealand was using to inform access to the data from the IDI. The five stages framework was also used by other national statistics organizations around the world, including in Australia and the UK. And that provided the basis for thinking about these areas which you can see through the middle part there and the orange, and the Māori concepts that span both sides of it. So anyone accessing Māori data from the IDI now currently has to answer questions related to each of these areas to get up before approval can be granted.

20:28

And so that Ngā Tikanga Paihere framework can see as something that works in a sort of essentially in a controlled and a cold controlled access sort of situation. And also thinking about how to maintain indigenous control or indigenous relationships with data, which might be sitting in more open spaces as well. And then a part comes down to thinking about what kind of meta data sets and what sort of cultural information is available as people are accessing data sets from different places. And brings us to thinking a little bit more about the provenance of indigenous data. And what I'm going to talk about here is the traditional knowledge labels, services initiative that, you can find at localcontext.org that has emerged over the last 10 years or so, developed by Professor Janie Anderson from NYU, in collaboration and partnership with indigenous communities primarily in North America. And these 18 traditional knowledge labels are machine readable digital tags that establish provenance for indigenous data in part by creating a jury system cultural metadata, and that enhance community control and definition in part because they can only be delivered by the communities themselves, while the icon remains stable. The cultural metadata associated with the icon is customizable by the community. And this is then either in use or been discussed with a large number of users, large number of indigenous user institutions across the globe in a variety of countries. And just to give you an example of what it looks like in practice, this was the

online catalog record serving within the Library of Congress. For the first ever recordings of songs that were made on tribal lands, and the tribe in question is the Passamaquoddy, you can see it highlighted there. But outside of the name, there's not really anything that connects the people to this material. And after the tribe engaged with the Library of Congress around this project, you now see what the record looks like. And so there's a much richer description of cultural traditional knowledge associated with the songs, creates a much richer record for the public, given that they can still access this. And you can see sitting there, really visible on the record, the traditional knowledge labels, which the Passamaquoddy tribe has placed on this material. So when the first labeling attribution talks about the connection back to the person calling themselves. Outreach refers to the fact that they're happy for this material to be used by the public for outreach purposes. But it should also only be used for non-commercial purposes. And this becomes the outward facing record, an aspect, you know, within the Library of Congress, but there's also things that need to happen on the background to allow this to happen. And one of those things is changes within the digital infrastructure. So within the mark record, there are now space for the traditional large labels to be recognized. And similarly in the JSON file, there is now a place in the database built in there for traditional knowledge labels. So this means that any item within the Library of Congress could now have a traditional knowledge label attached to it. Alongside this change, there was also the addition of the traditional knowledge labels to the rights advisory and a recognition of them as the first rights so they sit there alongside the intellectual property right, which is held by the Peabody Museum, at Harvard University.

25:11

And so the traditional knowledge labels, obviously associated with cultural heritage information, and traditional knowledge. And I've been working with Jane Anderson over the last two years to apply this model to genomic data generated from genetic resources. And so we've been working through a project here in altaira, to develop and apply them. And we have six initial bicultural labels, provenance label, tops, the same things as attribution, are considered verified; labels that indicate desire by communities that might be open to future collaboration, or being open to commercialization activities; a label that recognizes multiple communities where there might be a range of different communities that have an interest in a particular genetic resource; and also a research use, research use only then. And so the labels themselves and the other section was around provenance, and that's one of the things which the labels are doing is creating a sort of a provenance story that connects back to these traditional communities. But it also does a couple of other things in the sense that there are protocols, traditional protocols associated with material, which ought to be respected by public accessing the data. And that's reflected in some of the labels down at the bottom that might be restricted to men or women, or apply to particular seasons of the year. And there's also labels that refer to what purposes are being permissioned in the future. And, you know, as we talked about what the outreach label or the non-commercialization mean. And this work has led to the need to start thinking about where these labels or where this provenance information was set within, you know, the whole variety of databases and systems across the world. And so that's led us to initiating a project with the IEEE, to create a standard or recommended practice for the provenance of indigenous peoples data. And this activity has got underway in mid-2020, and will probably take a few years before we can generate the information that can then be applied as a standard across the globe. And so the last thing I wanted to cover in this session was a bit that then moves the thinking a little bit along from the questions of indigenous data, and a little bit of thinking about algorithms and artificial intelligence. And so this next session is based on a bit of work we've just completed around Māori perspectives on automated decision making. And that's been part of a project being done by the digital council thinking about trust and AI. And so there were two key activities that we did for this project: one was a literature review; and the other was, we had an expert workshop. And the similar key points that emerged from those two activities. One was that the literature of Māori data sovereignty, or indigenous data

sovereignty more generally, primarily focuses on data access and governance. And there are some comments made on appropriate use and analysis, but limited discussion about algorithms or automated decision making. And this generally reflects all the things that I said generally reflect concerns that are expressed by minorities and other jurisdictions. A couple of key points is that data itself can't really be separated from algorithms. And you know, that they're connected to each other in terms of what they create. The algorithms themselves can't be separated from the processes or the systems within which they operate. And so when bias exists within historic data sets, that's going to be reflected in the outputs of an algorithm. And if there's bias within the system that's using the algorithm that's also going to be reinforced and reflected through the use of that algorithm. And so the, as a part of the project, we looked at the Māori values that have been used in the context of Māori data sovereignty, Māori principles for Māori data sovereignty, which are promoted by Te Mana Raraunga but also used by the data Iwi leaders groups, as the Tribal Chiefs Forum has a working group focused on issues of data. And we've used that to frame some of the thinking that emerged from here too. So when thinking about financial autonomy obligations, it's important to increase transparency about the use or proposed uses of automated decision making. Effective partnership and participation means conversations about the ethical aspects of the technology happen at the same pace as the use of the technology or the development of the technology itself. It's important to create audit processes or design checks so that we can work out which automated decision making systems are good or effective and less effective. And that should be said or done using criteria of the Māori and now this kind of goes down to that need to calibrate the likely impact of any kind of automated decision making systems. In terms of monetary authority, we need a strong regulatory body to support the ongoing governance and use of automated decision making systems and software in the context of application across government or government departments. Māori governance across ADM is necessary to mitigate harm, stigmatization and inequitable outcomes. So how did it as we're creating a range of control mechanisms that can apply across the different kind of situational contexts, which we'll find ourselves so whether those are jurisdictional ones, or whether those are temporal ones. In terms of court a tangle collective benefit, building Māori capability and automated decision making machine learning and AI will likely improve outcomes and potential benefits. The opportunities to ensure Tamati that Māori worldview and Māori values frame problem definition and the way that decision making is applied, and need to be thinking about or recognizing that bias exists. And how do you work with that. And if we can create more transparency about the bias, and be selective, and the way it's used to further collaborative approaches, then that might address some of the problem issues associated with it. In terms of the Whakapapa relationships, what is the basis of a decision, what are the different kinds of decision or points where information has been drawn together, or applied or logic supplied, to come up with decision which comes on the back end of the algorithm, just you know, recognizing that historical data is often being generated from it and the critical process and that can create unequal outcomes. And that there needs to be more diverse range of characteristics and relationships within the data sets applied with an automated decision making to improve outcomes and trust in those systems. In terms of Manaakitanga, Māori need to understand the data that informs their decision making those made by automated decision making processes. There's limited examples of really good trust relationships, often between agencies of Māori communities, with most of the agencies expecting Māori to trust them, but not having much trust in Māori themselves. And, you know, this can only be addressed by meaningful participation and partnership. And in terms of kaitiaki, tongue or guardianship, data with an automated decision making systems as from people, and therefore has this sort of life, isn't it at all to be automated decision making algorithms or be viewed as but living requiring guardianship and relational, they're connecting a network of obligations and relationships and that needs to be appropriately maintained. And the analogy being used here as you don't leave a cabin alone in the rain. So all the effort that goes into place in terms of creating that, that town will create an artifact, you then just don't let it waste away. So, you know, just thinking about in the context of

algorithms how chi chiacchio guardians could be responsible for ensuring the algorithms themselves can withstand changes and context over time to ensure that they continue to be fit for purpose. And so out of that project, there were a range of recommendations that we put forward. Based on Māori perspectives on automated decision making, is the need to build capability and capacity around Māori data and digital capacity. Well, then both Māori communities and across networks of Māori practitioners, we need to develop robust equity assessment protocols for algorithms to assess algorithms to ensure Māori participation and institutional algorithms, self assessment processes, so often the internal processes being used as people are developing and implementing algorithms don't involve the people on which the algorithms are going to be applied to. And you know, that needs to be thought about particularly during self assessment processes. To support collaborative partnership and project governance and development and use of algorithms, it'd be useful to create a Māori values framework and guidelines based on cultural protocols that support the design, development, use, and maintenance of automated decision making tools. And also having the chance to explore Māori use cases. So, you know, much of the critique of AI and automated decision making is the negative aspects, which can be generated from them, which contribute to you know, continued colonization of indigenous communities. But there's also this element where the tools themselves, if you have the right sort of data, if you have the right sort of, kind of question, if you have the right sort of design could be applied to enhance language and enhance the use of, of traditional knowledge in contemporary environments. And that becomes one of the things to, you know, really consider and think about going forward as is not just about decolonizing data or decolonizing algorithms. It's also what is the potential there to create indigenous artificial intelligence? And what might that look like? And so, you know, through, there are people around the world, including, you know, this website here, which is bringing together indigenous researchers and communities to discuss what kinds of protocols might inform artificial intelligence. And as well, we're working on a new project here in New Zealand called Tikanga in Technology, which also has a component that's thinking about, what does it mean to indigenize artificial intelligence. So, just in closing, really highlighting, you know, indigenous data sovereignty is about indigenous control of indigenous data. But that's not the end in itself. You know, the control of data allows the indigenous peoples to control the sorts of narratives that are generated about them, and by them, and that's about those communities being self-determining, and that they're able to contribute to the realization of their own aspirations, and those things that behind, you know, really the activities of the Global Indigenous Data Alliance, and the discussions about decolonizing data about indigenous data sovereignty, and that these things need to be thought about in the context of effects as it relates to data and effects as it relates to AI. So norina tena, koutou Tena koutou Tena koutou katoa. I hope you can take the thoughts and ideas from this and have them add to to what you learned during this course.