

Show Notes:

Plurality: A Vision of the Future of Democracy and Society

Taiwan's inaugural Digital Minister, Audrey Tang, and author and founder of RadicalxChange Glen Weyl, join Johanna for an in-depth conversation on utilising technology to reduce division, improve governance, and regulate and the speed of innovation

Audrey and Glen share their inspiring stories of grass-roots political action turned tech revolution and speak about Australia's unique democratic inventiveness, their experiences unifying groups of people through new types of social media, opposing foreign state actors, prebunking misinformation, building bridges between diverse political positions, and making democracy quicker and more representative - all while keeping pace with rapid technological development.

They share their thoughts on the future of AI, open source vs closed source, accountability, steerability, and explainability, and the value of developing technology as public infrastructure.

Read Plurality: The Future of Collaborative Technology and Democracy here: <https://www.plurality.net/>

Connect with RadicalxChange: <https://www.radicalxchange.org/>

Transcript:

[MUSIC]

[TOPIC: Introduction to guests and topics.]

JOHANNA WEAVER: Hi, I'm Johanna Weaver, and you're listening to Tech Mirror, a podcast that talks about how humans shape technology and how technology is shaping our society.

Welcome to Tech Mirror. Today we have a real treat for you. We have two special guests to inspire: Audrey Tang, Taiwan's digital minister, and Glen Weyl, who's the author and founder of RadicalxChange at Microsoft. What is extraordinary about this story, actually, is not so much the fact that Audrey is Taiwan's first digital minister, but rather that Audrey is a public sector technologist and a minister, and during her term, she has built and rolled out many technologies that enhanced Taiwanese democracy and built consensus in a way that many of us dream about, but Audrey has made a reality. Much of the participatory democracy technology that we talk about with Audrey and Glen on this pod has been built and deployed, and it's increasingly being rolled out around the world by people just like you, dear listener, who were interested in building a better and a different future.

Back in 2016, when Audrey was first appointed Taiwan's digital minister, she was given the task of writing her own job description, and with typical Audrey flair, they responded to this request, not by creating a bureaucratic, boring document, but by writing a poem. This poem so powerfully encapsulates Audrey's vision for a different relationship between people and technology than we currently have, and so I asked her to read it to us, to frame the conversation that will follow. So let's get into the good stuff. This is Audrey Tang, Taiwan's inaugural digital minister, reading her job description.

AUDREY TANG: When we see the Internet of Things, let's make it an Internet of Beings. When we see virtual reality, let's make it a shared reality. When we see machine learning, let's make it collaborative learning. When we see user experience, let's make it about human experience. And whenever we hear that a Singularity is near, let us always remember the plurality is here.

Now, I think that's the way we use the word "plurality" in 2016. It is in sharp contrast to the inevitability-ish feeling of singularity of technology, which is progress so fast that humanity will no longer be able to make sense of it, the idea of plurality, then was that as long as we, the technologists and policymakers, make an

intentional choice for shared reality, human experience, internet of beings and so on, the capability of human coordination will grow in the same cadence, in the same tempo, as the technologies that needs our democratic steering, and so our hands stay on the steering wheel the whole time. So instead of saying that we hit the brake or the pedal toward a cliff, or was that a cliff, or is it Flying Spaghetti Monster at the end, instead of having that debate of AGI, we can just steer to a different direction.

[TOPIC: Audrey Tang's early life and involvement in the Sunflower Movement]

JOHANNA: Audrey, I just love this analogy, the idea that instead of being scared of technology or by the spectre of artificial general intelligence, AGI and slamming on the brakes, that we can instead take hold of the steering wheel and steer in a different direction. One of the things that we talk a lot— talk about a lot on Tech Mirror, is how we can awaken a sense of agency in people, so that they understand that we have the power to shape technology differently. We're going to delve into specific examples of how you have actually built these types of technologies in a moment. But before we do that, perhaps, can you share a little bit about your journey? Where did your belief that you had the power to harness technology for better outcomes for people. Where did that journey start for you?

AUDREY: Well, I think personally, it started when I dropped out of junior high school when I was 14 years old. I won the national science fair competition and was very interested in this new website called arXiv, or archive that has a lot of preprint papers that people just post before they even go to peer review, or certainly before going to journals, and started writing the researchers related to my science fair subject, which is machine learning and reasoning. And then they didn't know I was just 14 years old, and started doing research together. And then for a while, I thought, yeah, maybe I'll continue to go to school, but then research sounds much more fun.

So I printed out those emails, brought it to my head of the school, principal Du Weiping, and explained that I would like to spend more time doing research. And she was like, well, but how about getting a degree? How about getting into a senior high, University, into a PhD, into a lab doing research with those important professors? I'm like, you know principal, these professors are working with me now. They have no idea I was just 14 years old. So after reviewing those emails, she said, Okay, from tomorrow on, you don't have to go to my school anymore. I help convincing your parents. And I'm like, Yeah, but it's compulsory education. If I don't show up, my family gets fined. And she was like, Okay, I'll just fake the records for you. So that is— [JOHANNA: the principle said that?] That is the point where I started to believe in bureaucratic innovation, to see that people are willing to bend the rules when the rules no longer fit in reality.

JOHANNA: I just love that. It's such an incredible story. Let's fast forward to that rule breaking or rule bending principal, from there to 2014 and the streets of Taipei are full of protesters, mainly young people and students, and they're protesting a trade deal that's being negotiated between Taipei and Beijing and very famously wearing yellow headbands or holding sunflowers, those protesters then went on to occupy Taiwanese parliament in a movement that was known as the Sunflower Movement, and Audrey, perhaps unintentionally, you were right there in the middle of it. How did that happen?

AUDREY: So I dropped out of junior high, and I started a few companies, started when I was 15. So by the time that it was 31/32 it was 2012 or so. I'm already semi retired, started a few companies, just investing my time into civic technologies. At the time there's this G0v, or gov zero, movement that downloads the government websites, scrapes it from the web, put it into much more mobile friendly, much more interactive open source versions, uploads them at g0v.tw and shame the government into merging it back in. So I was having a lot of fun doing that: dictionaries and things like that. And so, interestingly, that brought us very close to civil society organisations at the time, and many of them were

secretly plotting to occupy the government in 2014 with a bunch of students to protest against the Cross-Strait Service and Trade Act that would have opened up, for example, our forging infrastructure to the likes of Huawei among other things.

Now, the thing happens to us, is that we're asked to support the network connectivity of the protests outside, but during that protest, when I brought my phone to offer, you know, a high-speed connection, livestreaming support and so on, people just started breaking into the parliament and starting to occupy there. Now, what's very important then to keep a non-stop livestream, because the very next day the media, and some of those anti-social social media corners, were portraying the protestors as violent people that breaks the stuff or have a fight with police while none of this has happened. There's certainly no violence. To have a non-stop livestream also means that people discover the situation when police surrounded the occupied parliament and people came from all over Taiwan and counter surrounded the police to ensure the safety of the people inside. And almost immediately the people inside began deliberating the trade agreement and turn it from a protest, which is against something, into a demonstration, which is for something.

In this case, for public participation in deliberation. So for the next three weeks we took up the movement, helped to provide broadband access, livestreaming to the facility data, daily summaries, you name it, logistic support, and, along with the 20 or something CSOs managed to converge upon a series of very coherent demands at the end of three weeks which was then adopted by the speaker of parliament. So it was one of the very rare successful occupiers that changed how people see because after this peak experience, you cannot say anymore that people cannot come to coherent ideas. They just did that for one of the most complex issues. And therefore, at the end of that year, I was then tapped to join the cabinet as a youth advisor and also to help reform the curriculum.

[TOPIC: Glen Weyl and the Plurality community.]

JOHANNA: And you then progressed from being the youth advisor to being appointed Taiwan's inaugural digital minister. I mean, that's quite a step up. And while you were in the— that position, you rolled out many extraordinary technology programs in the service of democracy and in the service of citizens, which we're going to talk about in a moment. But before I do that, Glenn, I'm really sorry I've been neglecting you, and you also have an extraordinary story. Can you tell us how you first came across Audrey's work and what drew you into it and led you to having taking such an active role in the Plurality community, including by co-authoring with Audrey and that community, the book *Plurality*, which— which has just been released. So where did this start for you?

GLEN WEYL: So I grew up in Silicon Valley. Both of my parents were entrepreneurs, and I came of age in an era where there was tremendous optimism about the effects of technology. Technology helped swipe President Obama into power and was transforming his White House. It was igniting movements for freedom, the Arab Spring, in the Occupy movements, et cetera. But by 2016 when I was really deciding the direction of my career, that optimism was largely gone. People saw technology igniting partisan divisions, spreading disinformation, leading to broad social conflict, and it was at that point that I decided that, you know, my career trajectory had to take another look, that I needed to engage with the public to try to figure out how to bridge some of these divides. And I wrote a popular book that caught Audrey's attention.

We ended up working together thanks to an introduction from Vitalik Buterin, the founder of Ethereum, on trying to bring some of those tools into participation in Taiwan. And as I got to know her, I really started to see that Taiwan was the counter example. It was the one case where these movements had really led to something meaningful, to building consensus, to strengthening democracy, to lifting up people and reducing social and economic inequalities. And it was because of that that I really devoted my life. I turned my life

towards trying to tell her story, making films and writing this book together, and now we've been on a global tour for three months for my 10th year anniversary sabbatical at Microsoft.

[TOPIC: Building bridges and reimagining 'anti-social networks' to be 'social networks'.]

JOHANNA: And one of the core premises of the book is that technology and democracy can be natural allies. I think that at least in the Australian landscape, there's a lot of scepticism about that. When we talk about technology and democracy, it's usually technology undermining democracy. So can you talk a little bit about the— both the tension and the opportunity that you see around technology and democracy?

AUDREY: Certainly. In Taiwan in 2012 to 2014 we have observed, as did everybody else, does those so called social media really manufacture counter power? That is to say, they're much more anti-social than social. So, while it is easy to start social movements of outrage and so on there, it is actually very difficult to heal relationships, to build the bridges across people of different ideas and so on, on those platforms. And, so like any good technologists, we decided to build our own open source, pro-social social media. Now, Taiwan has a long tradition of doing so. The National Academy supported the open foundry. Even before GitHub came around, we have something like GitHub that is supported as infrastructure by the government, the National Taiwan University supported PTT for almost 30 years now, a open source equivalent of Reddit, except there's no shareholders or advertisers. It's just a student club. So it is part of the norm in Taiwan for the charities and academics and public sector to continue to invest in the public square online and gov zero, because we do bi-monthly hackathons in the National Academy did a lot of those tools.

So during the Occupy, you can enter your company name or the serial number of your company in order to find out exactly how the trade deal affects you. On the street, deliberations were captured

and synthesised with the online platforms such as Lumio and so on. And right after the Occupy there's a wave of experimentation using open source discussion tools like Discourse, as well as tools that highlights the common ground across differences, like Polis into this system called vTaiwan that starting 2014 worked with the cabinet, but not for the cabinet, to resolve some of the most difficult emerging issues such as Uber, Airbnb and the like. And the great thing about these systems is that they're modular. It's not a kind of off the shelf solution, but every time that we decide to try something new, we communicate with stakeholders in the weekly meetings and things like that, and just adjust the technology to fit the reality of the particular issue at hand. And so I think the most important lesson here is that there can be algorithms that identify the differences and highlight the bridges, and if the government can then commit to use those bridging statements as agenda for multi stakeholder conversation, then people are actually very willing to build those bridges. If you give this, what we call bridging bonuses. So I think the traditional social media nowadays, it's been 10 years, has now been learning from those algorithms. So Twitter, community notes, community notes in Google, many others are now starting to learn this bridge making systems, but when it was prototyped in Taiwan, it was entirely the work of the Civic and public sector.

[TOPIC: Using technology to identify public consensus and shape government policy.]

JOHANNA: I mean, that's just so extraordinary, isn't it? Just think about— stop and think about that for a moment for us, like social media that builds bridges rather than amplifies divisions. Can you imagine that? And here's the thing, we don't actually have to imagine it, because Audrey, you and so many others in Taiwan have actually built it, and not just built it, but you've used it to effectively inform and shape government policy. And I think Glenn, when, when many people hear what Audrey has managed to do in Taiwan, there can be, well, I think most people have a sense of awe, but I've noticed that this is sometimes followed by a sense of

powerlessness, or perhaps this feeling that it isn't translatable, that people admire what Audrey has managed to do, but is sceptical as to whether or not it can be repeated in their country. With plurality, you're looking to build a global movement. So how do you respond to this type of scepticism? I mean, it's well meaning scepticism, but it is scepticism nonetheless.

GLEN: You know, the UK, I believe developed industrialism because they had a shortage of workers and a surplus of coal. And Taiwan, I believe developed this model because technology was such a core part of the economy, they export 50% of their GDP as digital exports, and democracy was the core of their identity. That's how they differentiate themselves from folks who might want to change their system of government. And so it was absolutely critical that they had to fuse these things together and develop what they did. But industrialism was adopted many other places in the UK who really had to transform their societies in order to do that. And you know, I really notice that these days, there's a real obsession with this sort of abstract, floating, hanging technology. Ideas like, AI is going to do it for you, you know? And it's like, okay, so where's an example of AI actually making a difference to anyone? And it's like, well, it's kind of hard to figure out, you know, but it's just thought, it's this universal thing that's going to transform everyone's life, or blockchain, it's going to make this amazing difference. And there's no real examples of making a difference for people. So to me, yeah, is porting something between two cultures challenging? It can be. But if you have a real example of a society, not just like a flashy computer demo that's been transformed by the integration of technology with society, to me, that holds much more hope than something that is like a free floating spaghetti monster in the air, like AI or blockchain, and maybe is worth trying to figure out that pattern.

[TOPIC: Technology's vital role in the Sunflower movement.]

JOHANNA: A free floating spaghetti monster in the air. I just, I love that description. Maybe let's delve from there into some of the

transformative ways that the technology you're talking about has actually been integrated into Taiwanese society. Audrey, can you take us back to the Sunflower Movement? So we're back to 2014 people on the streets. Parliament is occupied. How many people were involved in this? And what— How did you use technology in order to support those protesters?

AUDREY: Yeah, so it's half a million at a height on the streets, but within the occupied parliament, maybe 100 or so. And the thing that we set up is a real time live stream, so that people from the outside, using their phones or just looking at the projectors on the walls, can see in real time what's going on and little deliberations at a time. There's like stenographers, core reporters that just types everything everybody says across all the sides of deliberation so that we can just manage them and massage them and summarise them at the end of the day to find out the common ground. And so later on, of course, all these functions so AI is good for something for transcription and translation and summarization.

And so because they model language, so that is what we are now using, and to your point of information integrity nowadays, in Taiwan, there is a very strong ecosystem of co-facts or collaborative fact checking that lets people see what the trending rumours are when people long tap and report it from its end to end encrypted channels to a common register that's open source, and then a language model compares each incoming message to the previous messages and its templates and so on, and come up with like rapid response. Like the first take of what could be a counter narrative to that, before opening to everybody to add more context and so on. And it is also part of our civics education to encourage students to participate in such end of use. And also pre bunking, because after a while, you tend to anticipate what the foreign information manipulators will say around particular topics, whether it's about masks, vaccine or election. And so you can actually prebunk all these messages by saying to people that these are coming, they will look like this, and then relying on people's ability to

communicate lightheartedly their counter narratives to these kind of assaults on information integrity.

[TOPIC: Adding social context to views and shaping legislation with Polis.]

JOHANNA: Yeah, it's extraordinary. The entire ecosystem that you have built from, you know, the pre bunking to misinformation/disinformation responses by collaborative fact checking, right from those first days of summarising the positions, using technology to do pretty basic things that allowed people, almost manually to identify their common ground to the systems that you're now using in Taiwan that help to actually build the bridges and to reach common positions. So can you talk us through some of that specific technology. Maybe, if you can have an example of— of one of those technologies that you have built, and then the impact that that has then had on shaping policy.

AUDREY: Sure. I mean, the original Uber example of 2015 you would go to a website, it's called Polis, and you're going to see one statement from your fellow citizen. We always ask, what do you feel about the Uber X situation? And maybe some feel that it's convenient, maybe some feel that it's extractive. It's all fine. And as you click, I agree. I don't agree. I agree. I don't agree to each statement, you see your avatar moving among the clusters. So initially there's clusters of Uber drivers, Uber passengers, taxi drivers, people who care about the rural places, so on, so forth. And so each cluster is visualised not according to how many people are in it, but rather the plurality of the experiences and the ideas contained within each cluster. And we also show via the Polis system, what are the unifying statements in each cluster that distinguish them from other clusters, as well as what are the kind of global statements, the breakthroughs that communicate across those clusters.

So for example, one earlier breakthrough is that surge pricing is fine when the demand is high, but undercutting existing metres is not

fine because it destroys the kind of dignity, right, the basic livelihood of drivers. Now that is something that can unite two islands, two clusters together. So over the course of three weeks, you literally see those clusters moving into the middle. And then people kind of compete to post statements that gain us even more and more bridging bonuses. And at the end of the day, the top 10 or so of those statements were then taken as agenda in a live streamed multi stakeholder consultation with all the stakeholders, and then they were like, oh, yeah, we can do such changes. And Uber is now a legal taxi fleet that Q taxi in Taiwan, but the law has been changed to admit search pricing and many other flexibilities, as well as allowing the rural places to co ops and so on, to join the app based dispatch. So it's win, win, win for everyone. But the reason why we can get that is that the system was very open ended and only asked, What do you feel and what's your feeling toward other people's feelings?

GLEN: And I just wanted to build on that to say I think this notion of bringing people together is a very good starting point, but I actually think it goes beyond that. I think the real problem is not about division or unity, but about people not seeing themselves in social context. There's no problem with there being differences of opinion if we have awareness of them. The problem is when people are constantly fed information as if it was universally relevant, when it's really the narrow opinion of a community. I mean, we have all kinds of diversities in communities. We know that, for example, religious content for various religion is not a consensual thing that everybody agrees on. That's not a problem. That's— that's diversity. That's wonderful. It's— the problem is that if you think that that content is what everyone believes and thinks, but actually it's just representative of a narrow community, and that's not— if that's hidden, that's the problem. So actually, what we really want is technologies that give us an accurate and transparent view of what the social communities are and which ideas are shared and which ones are narrower, so that we can place ourselves within the social context and not be sort of individualised to think that whatever we're served is the global view.

[TOPIC: Combatting foreign misinformation and anti-democratic influence campaigns with pre-bunking.]

JOHANNA: I really want to emphasise something that you've said there, Glenn, at the start of your response, that the problem is not so much that we have differences of opinion, is that we have no awarenesses— no awareness of the differences of opinions, and we need to, as you just said, see ourselves in that social context. And you know, the reality is that most of the social media that we have and use today reinforces the messages of people that think like us, that speak like us, that have the same views as us, and they serve that content up that is so similar that we don't get that diversity of views, and therefore we don't have the opportunity to build bridges where it might be possible. I mean, it's not always going to be possible. Again, this is what I find so extraordinary about what— what you and Audrey are building with the Plurality community. There's so many people out there that talk about reimagining social media, but in Taiwan, you've actually done it and it— it is proof that it is possible. So I think another really interesting example of the different approach that Taiwan has taken is in regards to mis and disinformation. This is a topic that's very topical in Australia, and in particular, Audrey, the way that you make a point in Taiwan to distinguish between the policy and societal responses that are required for foreign and domestic mis and disinformation. So Audrey, can you talk us through that approach and why it's so important?

AUDREY: Sure. I mean, during the elections, of course, each candidate is going to offer some platforms, campaigns and so on, but we do not allow donations from foreign actors to the political campaigns for the very simple reason that democracy is within our politics, right? And so we do not allow, when we cast the votes, people who do not have citizenship to suddenly, million, you know, hundreds of 1000s of mail-in votes. That's not the case, right? So, I mean, in a polity, it really helps to distinguish between the layers of the actor, the behaviour, and the content, exactly the same content,

if it is posted by a domestic actor, that may be just political speech. But the same content, if it is foreign, sponsored and masquerading as local actor, that is inauthentic behaviour. So in this point, there really should not be a content layer. Only regulation if you all you have delivered is at the content layer, then you like mix everything into this very ambiguous idea of a fake news or something, which does not help the conversation.

So what we did instead, in Taiwan is that when there's foreign interference, the national security, the think tanks related to them and so on, do like full attribution, just show the people that there is foreign interference going on. But on the other hand, if it is just domestic misinformation, then the ecosystem of like immunity makers, pre bunkers and so on, the co-facts ecosystem, the Trend Micro and Google. Look all those private sector antivirus companies and so on. They all offer products and services that you can install, maybe in your chat group, maybe as an app that manages your notifications and so on that compares these incoming messages against the collaborative fact database. So in a sense, people then learn that instead of taking anything down you— we surround everything that is the misinformation with contextual information so that people can then use these material to make prebunking, so that whenever something like that happens again, people see that and say, oh, yeah, that's that's the prebunked material. And that's how you learn about, you know, journalistic perspective. That's how you learn about media competence. That's how you build the general immunity against outrage and polarisation attacks, if everything is taken down by the administration that actually fuels conspiracy theories and more than anything else.

[TOPIC: Motivating citizens to participate more actively in democracy.]

JOHANNA: And how do you get people motivated to do that, because I think we would say in Australia, yes, we want, we want to have that kind of an environment, but it feels so far away from where we are now.

AUDREY: Yeah, I think one key trick is to reduce unnecessary work. So if to report a trending disinformation, all I have to do is long tap on my messenger. That's easy. By flagging spam. Anybody can do that, if to receive the clarification, the balance report and so on, that is just a built in feature of my favourite antivirus tool or things like that. Then again, everybody can participate in that. And if you see a bit of context that you would like to join, you can just say, I'd like to add to that, and then talk to your phone. Again, many people can do that as well. So the point is that everybody have maybe two seconds of kindness, and after two seconds, maybe two minutes of kindness, but nothing beyond that. So the point is, as people are young and have a lot of time, just make sure that when they were 12 or 13 or 14, they learned the art of this kind of community collaboration, and then it becomes a habit of sorts, and once they learn about that, they can also teach their parents and grandparents

[TOPIC: The Plurality movement's goals and methods.]

JOHANNA: And part of making this translatable at scale is the Plurality movement. So Glenn, can you tell us a little bit about what the movement is and the motivation behind writing the book as well?

GLEN: We believe that these kinds of examples can inspire people around the world, empower people around the world to address the problems that they face within each of their contexts, and that their examples are exactly what we need to tell the story, and that therefore we needed to write an open book in which people around the world can contribute to it, so that while we light the flame of a new possibility that the things that catch fire around the world are things we can't possibly predict, which has been very much what's turned out To be the case and now maybe almost a majority of the things we tell are things that got inspired around the world, but that have gone forward in ways we couldn't have imagined. People in Croatia are using quadratic funding to support local businesses.

People in Japan are crowdsourcing platforms for very popular political candidates. People in the UK are writing novels that imagine a world where this plays out. People in the United States are making films.

So the Plurality movement is really just the opening up of our stories to become anyone who wants to contribute stories by making things free of copyright and participative, and going around the world and just asking people if there's a way we can be a platform using our stories or their stories to gain more momentum and have more impact. And hope we can do that for you all here in Australia and the wonderful work you do, Joanna.

[TOPIC: Enacting changes in Australia and bringing new tech driven solutions to policy making.]

JOHANNA: Well, there's certainly a lot of appetite for it. I think you've been in Canberra, and there's a lot of excitement, both in the meetings with you, but then trailing behind you, which is wonderful to see. Let's— let's stay on that theme. If you were an Australian politician who wanted to get involved and have more direct citizen participation in the way that they represented their constituents, what do we need to do? What are the steps that we need to take?

AUDREY: Well, if you're a politician and you need to run for your next re election, you can look at what Takahiro Ano-san has done in Japan, which is toll free numbers or hashtags and so on, to crowdsource your platform, which is taking your existing platform document and training a voice clone of yourself or a V tuber of yourself. That explains this in a way that you probably don't have the time to individually to each and every people who care to dial that toll free number.

It is also to make sure that you do this: tell us that answers the top most reasonable policy change proposals to your existing platform, and also this, there is this talk to the city, the open source summarisation module that can then do broad listening and get all

the feedbacks to the ideas that you proposed, and to tell you, in very newest way what the constituents think, in a sense that they can suggest even more nuanced versions of how to narrate, how to tell the platforms to the various different parts of your constituents. And so all this is like off the shelf, open source, meaning that you can tailor make it to your potential applications and so on, but there's nothing preventing you from using these technologies at scale. And so I think the most important thing to remember here is that this is not a fixed set of technologies. This is a lot of toolkit that enables broad listening, and this commitment to listening broadly that really distinguishes people who engage the plurality— stack the plurality work, versus people who were kind of stuck right in the ideas of broadcasting only politics.

GLEN: And my understanding is that there's a coalition of folks coming together here in Australia: funders, technical experts, people of backgrounds in public opinion, who really want to support leaders, who want to do this. And so hopefully you'll have the opportunity to connect with some of them and to build momentum around this.

[TOPIC: Decision making power vs agenda setting power.]

JOHANNA: Yeah, I couldn't agree more, and I can feel that momentum building. Now, both of you have spoken about a suite of tools that increases participation in democracy, things like, you know, crowdsourcing or refining political platforms, town halls, tools like Talk to the city that that listen and facilitate a two way conversation between politicians and their constituents, but what do you say to the sceptics, or maybe actually the political die hards who say that by increasing this level of participation, you're handing too much power to the people or or people that don't understand how Parliament really works. And I just want to be clear here, I'm not advocating for autocracy, but I do think some of these concerns that people have, or reservations that they have about how these technologies might distort or change power are— are real concerns. So how do you— how do you respond to that?

AUDREY: Well, I mean, there's decision making power and there's agenda setting power. Every tool that we just described shares agenda setting power. It is literally just making sure that when there's emerging issues that affects the society, the people can discover it and also discover the various different system dynamics that causes it, but it is in no way a replacement for, say, referenda, right? So this is small d democracy. The idea is that these tools can be used within a organisation, within a school, within a soccer club, to quadratically vote for their favourite coach, which is something that actually happened in Brazil, and many other things. And so think of them not as sharing your final decision making power. Think of them as a kind of better, better radar that you can dance this out what is coming to your constituencies, hearts and minds, and that helps you to win those hearts and minds.

GLEN: To take a analogy from the business world, I used to work in the office of the CTO at Microsoft, and certainly, you know, the base level, programmers, product managers, don't know how the office of the CTO works, but neither does the office of the CTO know what's going on in all of those product lines. And so the question is not, how do we eliminate expertise? The question is, how do we have more expertise? Because the expertise of the office of the CTO is very, very narrow. Everyone's expertise is very narrow. And so we want to be able to ensure that we can listen to all the expertise we need to get the right decision.

JOHANNA: What a novel concept listening to expertise and diverse views to get to the right decision. Look, as you know, design thinking sits behind much of the work that we do here at the tech policy design centre. And Audrey, I've heard you talk about design thinking and describe what we've just been talking about as the double diamond. So how you use policy as a design tool that is enhanced by technology. Can you explain what you mean by the double diamond model to us?

AUDREY: Sure. So agenda, selling power is in the first diamond in terms of the double diamond model, it begins with a very wide mapping of discovery: discovering people's feelings around any particular subject. There's no right or wrong when you're discovering. It is just that mapping process. And then after you mapped out the stakeholders, their values, their feelings and so on, you begin the first converging process, which is called define. And the thing that's defined is this common set of “how might we” questions: how might we make sense of all this? How might we balance these different values? How might we live together and so on. But those “how might we” questions, again, just outlines agenda for things to do. It is not a set of policies that is waiting to be read by the parliament.

So these are much more value based, much more abstract, but also based in people's lived in experiences. And they are not concrete in a sense of like rules and regulations, but they are much more closer to the ground, in a sense that you can point to the discovery process and say, This is where the pain point comes from. So very interestingly, it serves as kind of the goals. It doesn't prescribe that specific implementation rules to those goals. So this looks a little bit like farther away from now, but also because it is crowdsourced from the collective discovery process, communicating those “how may we” questions is much easier than the nitty gritty details of the rules and regulations. So the first diamond maps out the stakeholders, their values, their ideas, and it defines the common things that we can all live with. And then it connects to the second diamond, which is to develop various different kind of solutions that could fit the spec of the “how might we” questions and finally, delivering, getting the products and services to people's hands. And so when we're talking about, say, the legislative process, the implementation process, the procurement process. We're talking about the second diamond, which is about development and delivery.

JOHANNA: And I think sometimes when people hear the concept of participant— participatory democracy, they think that they were out

to replace parliament with one kind of super app where every citizen gets to vote on every issue. And I'm sure there are some people out there that advocate for that, but what you're talking about is actually much more subtle. It's about using technology to map stakeholders, their views, their ideas and to define the areas of commonality where they exist, and then to make real policy progress in those areas. And then, more importantly, I guess, through all of this, is also to be using the technology to improve the efficiency and the effectiveness of the implementation of the policy.

AUDREY: Yeah, in Taiwan, even when we do random stratified sampling, maybe mini public and so on, the goal is always just to make sure that there's a wide variety of people to define the common values. It is never the idea that, because these people are statistically representative, we can dismiss the legislature. It is never about that.

[TOPIC: Bridging and consensus building technology using deliberative polling.]

JOHANNA: Yeah, exactly. And Glenn, can you talk a little bit about this idea of bridging and consensus building technology, and in particular, how it treats people who have views at the edges? You know, these are people whose views might be minority views, but they're very vocal, and often their polarising views are amplified by traditional social media. So how do the technologies that the Plurality movement is building and championing, how do how do your technologies, or the suite of technologies, treat those types of voices

GLEN: Well, as I mentioned before, I don't think that the goal is just to repress the extreme voices. In fact, often we really want to hear the extreme voices. We just want to understand what they are. We want everything to be heard in the way that it is: in its context, so that two things, first, we can build bridges across those contexts, because if you don't understand a context, you can't build a bridge across it. I'll give an example of that in a minute, but second so that

those people see each other, and if they want to work together and campaign and press that they can unite. So we don't actually want to eliminate those voices just because they're different from other voices. We actually want them to be able to see each other, to find each other, to work together, but also to understand the context that to have an action if that's what they want for the society at large, they need to build bridges to other people.

So let me give you one example of this. I come from a very secular family from Silicon Valley, but I've come to believe that actually the strongest ally that people who want tech reform, and a better way of doing tech in the world that I want have is deeply religious people. And a lot of what tech has done in the past is actually suppress religious expression. So if you try to search for various types of religious things on Google, Bing, things like this, you'll— you'll usually get punnies, you know, and Christmas trees, not actual authentic expressions of that religious tradition. Now some people would say, Well, that's good, because then we don't get into these debates over religion. But actually, if you want to build bridges, if you want to work with religious people, you have to understand the terms in which they understand the issues, not suppress them because they could be too divisive. It's actually a source of energy for working together to understand our differences, not something that by suppressing it— all we do by suppressing it is we lose the opportunity to use that energy and store it up so it'll explode sometime in the future.

JOHANNA: And Audrey, can you build on what Glenn has just explained here? You know, I know some people will be listening to this and saying, look, it sounds great, but maybe a little bit utopian, that it isn't something that is in— within reach. But I think it's really important that we remember that what Glenn is describing isn't fiction. It's real. You personally, with a team of eager supporters have built it. So to make this more tangible, can you give us an example of this type of deliberative technology, things like the alignment assemblies, or perhaps Polis that you have actually used to build bridges and solve real world policy challenges in Taiwan.

AUDREY: Sure, I'll use a recent example, because it also relates to the Australian recent conversation around social media. In Taiwan, the NCC, the National Communication Commission, has been trying for a while to tackle the issue of social media, but the intermediary laws that is modelled after the Digital Service Act was not very popular when it was brought for consultation in 2022 and so when the ministry of digital affairs started, we're like, yeah, content moderation, that is the purview of NCC, but we can actually see those social media companies also as advertisement platform which is within our preview. And so we did a alignment assembly around information integrity in those large platforms using these consultative technologies. Specifically, we worked with, first the Polis technology to ask people: what do you think about generative AI in general, and what do you think are priorities that we should talk about?

And then when information integrity came up as the thing to talk about, we moved to the defined stage and then worked with Stanford using online deliberative polling. The way it works is that we sent 200,000 SMS to random numbers in Taiwan coming from 111, the trusted government number. One half of the SMS started out and people filling some surveys. We asked people, How do you feel again about information integrity? We collected, I think, less than 10,000 responses, so we sent out the other half saying that we're going to pay you 70 US dollars for almost a day of participation. And much more people responded to the survey. And the idea here is that across age, ethnic, region, gender, location and so on, metrics, we did a random stratified sampling, choosing the people who responded randomly so that 500 people gets chosen as a statistical representation, the mini public, of the Taiwanese society. Now, 450 of them showed up, and in 45 rooms of 10 people each, they deliberated around the information integrity issue.

Now, two important things: first, those rooms are auto facilitated. There's no human facilitator, so we can facilitate as many rooms as

possible. And if people are too quiet the room, push them, not just them, to speak up. You can distract people, but just for five seconds each, there's a real time transcripts going on all the time. There's a group selfie showing that whether the group is ready to move to the next topic, so on and so forth. That's the first thing. And second, there's real time summarization across all the groups. And so when all the groups are talking about things. There's other engines like talk to the CD and so on that can go back in time and analyse all the transcripts at once and give us this overview effect of what people are talking about. And even pick up some of those nuanced statements together, so that, for example, people really don't like content moderation as an angle, but people really like anti fraud as an angle.

For example, people would think that social media companies should secure public key signature from celebrities if they want to authenticate their likeness, their deep fakes, their digital twins for advertisement, but if the social media don't do that, and somebody could count for 1 million, then people think those social media companies should be liable for that 1 million, so and so forth. So all these fragments of ideas, when pieced together, actually form very coherent lines that could have been, you know, written into law, and so we read them out at plenary, experts talk about the repercussions and so on. And then people break out again into 45 rooms and so on. And so the idea here is that at the end of the day, we collect these consensus that after deliberation, has more than 85% of people across all different ideological camps who support them. And then in the following multi stakeholder meeting, we read them out to meta, to Google and also Microsoft, open AI, those the sorts, and say, you know, these are coming. This is what our people want. Is this technically feasible? And they're like, of course, yes, it just cost a lot of money, but anyway, and so and so now, as we record this, it is already law for a while.

Now in Taiwan, it's only three months between the alignment assembly, online deliberation and the parliament actually fast tracking this into an actual law. And so it shows that democracy can

move at a speed of technology. It doesn't have to be always trailing behind technology. The fact that most of the laws trails behind technology, almost like doing it after it's too late, is because we spend a lot of time doing this back and forth, because our listening skills are not as broad and not as deep as the current generation of technologies are.

[TOPIC: The place of experts within the system.]

JOHANNA: Yeah, I mean, for me, this is music to my ears, because I've said before that we have to bust this myth that regulation has to be slow. You know, I really do believe that if we're going to be effective in our mission to use technology. To shape a better world, that we have to be able to regulate at the pace of innovation, and I really love how you're framing this, that we need to stop wasting time with poor listening schools and substandard consultation, and that we can harness the technologies to deepen and broaden the way that we listen or that we consult so that we can develop technology at speed and at scale. And I guess to all of those policy makers that are listening to this and getting really excited about the opportunity of using the technology like this, but maybe getting just a little bit worried that it will do them out of a job, can you provide some reassurance in these types of deliberative forums that we've just spoken about. What is the role of experts? Do we still need experts, policymakers and practitioners.

AUDREY: The plenary reading is by the experts, and we also include in the 450 people, 150 actual practitioners on media. And so it is both in those mixed settings, where we do have rooms of experts talking among themselves and also with people, and also the plenary, which puts the experts on tap, not on top, responding to people's proposals and ideas and so on. And I think that is the role of experts, because everyone is an expert in their own feelings and resonance and lived experiences, but people who are trained in academic and in professional languages and have the experiences can evaluate the possibility space much quicker. But it shouldn't be them to dictate the possibilities. It is more like the possibility space

mapped out by the discovery process then, as we're defining it, the experts then put much sharper contours on how to translate those general idea, those general values and so on into something that has a chance of working.

GLEN: I mean, I think the role of experts is in many cases, very similar to the role that like Als play in these things, which is, you know, people define the different clusters of opinion, and then someone needs to propose bridging statements. And maybe the participants can propose those. Maybe the experts can say, "Oh, if those are the things that matter, this thing might appeal to everyone," but then it's ultimately everyone's job to figure out whether that did do the job of bridging. And it creates a different type of an expert. It creates an expert who's an expert in listening and an expert in perceiving the synergies, rather than expert in saying that they know what the issues are that need to be bridged.

[TOPIC: The strength of Australian democracy, and the place for quadratic voting.]

JOHANNA: It sounds to me like the skills of a very good diplomat....

So Glen, you've been in Australia for a few days now, and I think it's fair to say that Australians pride themselves on the fact that we have a strong democracy and that we are electorally inventive. Shall we say so? For example, I mean things like the secret ballot was invented in Australia. Compulsory voting comes from Australia. What's your sense of the appetite in Australia for this type of transformation and innovation?

GLEN: My perception is that there are, there's like a very rich tradition at the centre of Australian politics that's always trying to reach one step further for something that might work better, that might keep society tied together. And I've met a number of people like that, perhaps unsurprisingly, from our global experience, overwhelmingly women. And that opportunity, I think, to be on that cutting edge and to show a version of tech leadership that is about

bringing people together, and often that is something that is disproportionately gendered as female. I think that's an incredible opportunity that Australia has right now. Taiwan obviously has had to innovate a number of those things, but Australia maybe has a chance to institutionalise them. Taiwan still has a first past the post system for President. Australia is past that in many ways, and maybe you can go even further. I wouldn't be at all surprised if Australia ends up as the first quadratic voting jurisdiction, given all the ways that you use an inventive mix of different voting methods to elect people.

JOHANNA: Maybe, can you tell it— speak to what what quadratic voting is?

GLEN: So quadratic voting is a voting system that allows people to express how strongly they feel, and not just the direction of their preference. You get a bank of credits and you can vote in favour of or against. Candidates or proposals or initiatives or infrastructure and say how strongly you favour it or strongly you oppose it. And that just goes even further than preferential voting, instant runoff, things like this do in ensuring that you don't just get locked into the choice of two lesser evils, because you can make clear in a very nuanced way. This is really important to me. I'm kind of indifferent between these. Therefore it allows a much more diverse situation, rather than just having to have like one political spectrum, talking to a politician last night who was saying that issues are complex and highly multi dimensional here. And when you have a chance to say yes, strongly, no, weakly, et cetera, you have the chance to map out that rich multidimensional space, rather than putting everything into like, you know, left, right, or something like that.

JOHANNA: And Audrey, can you speak to an example of how that— because, again, I think as an example that people hear, but then go, how does that actually work in real life?

AUDREY: Like every year, Taiwan runs the Presidential hackathon, and for the past five, six years, we've used quadratic voting. And

the experience is that you get a link, maybe from your friends, saying, go to here and vote and support my project. And the project usually a local project or small scale pilot project, once they make the top five, will get a presidential promise to get the infrastructure behind it to be national infrastructure. So this is one of the highest honour credit a public servant can get when it comes to innovation. And so the idea here is to get 99 points. And as you vote your friends project, you'll discover very quickly that one vote costs 1.2 votes, 4.3 votes, 9 points. And with 99 points, you can only vote nine votes, which costs you 81 points. Now do the math, and you have 18 left. Nobody want to squander those 18.

So they look at another project and cast four votes, which cost 16, and you do still have two points left. And so you're motivated to look at two other projects. And so after you look at four projects that usually have some synergy, maybe you feel that, you know, 81 points really too expensive. I want to take some of that nine back. Maybe I want to do a seven and seven. Maybe I want to do a three and three and two and so on. And so at the end of the day, many people vote for like, 5/6/7, projects that have some synergy. And so it has two effects. One is that when the top 20 is announced, pretty much everybody feel they have won, because they have some stake in the ones that made the cut. And the second is that the other 180 or so projects that did not make the cut, because there is now a group selfie, they can see how synergistic they are to the top 20 that did make the cut, and therefore can actually maybe join their efforts, right?

So it creates a decidedly positive sum benefit to the people both voting and participating in the projects. It creates a collaborative spirit, because it is much easier if two or three projects advertise together and mobilise together rather than first past the post a single vote per person, which makes it, strictly speaking, zero sum, right? And so quadratic has many properties, but the main property is that the chance of the each vote of getting somebody some project winning is exactly proportional to the marginal cost that you're spending on that one extra vote cast. And so it tends to get

people into the mood of actually evaluating the actual synergy between those projects, instead of just, you know, go vote for your friend, close the window.

[TOPIC: Open source vs closed source vs public code. Code as public infrastructure instead of commercial product.]

JOHANNA: And again, what I love about what you're describing here with this system of quadratic voting is that it's gently, just gently nudging us, not only to look at that shiny thing that our friend has asked us to vote for, but it's also sort of incentivising us to want to look at these other proposals and these other things that are out there, and in doing so, expanding our view of the world rather than narrowing it. And this is all due to the design choices that are being made by the people that are making the technology. And I think another choice that you made as digital minister in Taiwan was to build your public technology, open source. And I know there's a lot of debate about the security of open source tech.

And I think again, here Taiwan is such a fascinating example. You're a country that is famous for your technological prowess. Much of your technology is public technology. You built it open source. While at the same time, withstanding a daily digital onslaught from— from China. So to me, Taiwan's experience, your lived experience, Audrey, really busts the myth out of the water, the idea that open source can't be secure. But how do you respond to people who still have reservations about public or open source technology?

AUDREY: In Taiwan, we say public code, so it's well, roughly speaking, describing the same software. But when you say open source is a kind of private sector practitioner, saving costs, sharing maintenance, kind of worldview, very practical aspects. When you say free software, it is about the moral right of the people using it to learn, to change and so on, and but if you say public code, it means that you're treating it as public infrastructure. In Taiwan, the Ministry of digital affairs is in the committee of transportation, meaning that

when it comes to public infrastructure budgets, we're thinking of it in terms of like airports or train stations. It doesn't quite make sense to have the airport only fly one private airline, right, and then forcing each airline to build their own airports. At some point, you will have no competition.

But rather the form of interoperability that is taken in the infrastructural layer is coupled with the code of regulation, of penetration testing, of red teaming of software, bill of material, of zero trust, network architecture, of defensive depths of real time indicator, sharing all this code that is needed for the code that is free software, open source, to run in the public sector. And so if you think of it as a resilient public infrastructure, it is not about saving costs. It is not about just going from one vendor to the other. It is also about something that is stable, something is jointly maintained and stewarded by multiple departments that can use it as an exchange and can be trusted over time. So this is how we branded public code. And people all understand that it doesn't automatically cost less. It actually initially would cost more because you would need to build a community of practitioners that understand the in and out, as well as how to ethically roll it out within the operational environment in the government, and also making sure that the same procurement in two adjacent layers doesn't go to the same vendor. So there's the need here to speak interoperability using open protocols, which is as important as open source.

So we're not fundamentalists. If the public infrastructure supports Open Document Format, and if one department decides to use Google doc to be a word processor of that open document, they can still procure it, it is just that both the input layer and output layer cannot be in the proprietary format.

GLEN: Yeah, and I mean to follow up with that infrastructure analogy, you know, when we have public bridges, of course, those can be sites of attack. I mean, we have to think about defending them, but the best way to defend them is not necessarily to, like, have them be enclosed and private and only a certain type of car

goes through them, you know. And it also doesn't mean that, like, you never charge a toll anywhere, and we absolutely believe that, but it does mean that, like, the default understanding is that it's for the public interest that charges are to maintain the infrastructure and to avoid congestion, and not just to hide it and control it and cut it off. And then that actually makes it more resilient, because more people are seeing the bridge they're have a sense of public spirit. They if they see something, they say something in a way that you wouldn't do if you're in a private club,.

AUDREY: Yeah, if it's public code then there's a budding industry of red teamers in Taiwan that would very much like to, you know, join and earn bounties and things like that. But if you close off the access to those public code, then what happens, really is that it's security through obscurity. You're not just defending the system using passkeys or other cryptography, you're not just defending that particular attack surface, you're essentially including the entire non free, proprietary source code into your attack surface, so that when people get an excess of copy of that and it was not as thoroughly tested before, they can find loopholes. No matter how good your passkey or whatever authentication system is, they can find loopholes in that source code better than to just use the public code that has been used or tested by other governments or by other international organisations and build upon such build of material components that are tried, tested and true. And also then invite your own red teamers, your own cybersecurity startups and apparatus to then try to attack it and before, well before, the deployment. So it doesn't work if it's suddenly open sourced after being proprietary for like, a decade. It does work., if from the very planning stage, you open up to the public in the design process, then also start those bug bounties and things like that, and very early on.

[TOPIC: The dangers of closed source AI models, and making technological innovation safer via collaboration and open sourcing.]

JOHANNA: yeah, I couldn't agree more. I'm a really big proponent of the Australian Government taking a much more active approach

to bug bounties and open source, especially when it comes to technology that government is building, or public digital infrastructure. And I guess to pivot slightly, one of the areas where there is, I know, a lot of discussion around open source is with respect to artificial intelligence. I'm conscious we've been talking for a while, and this is a huge topic, but just that's at a high level. Audrey, what are your views on the ways that the different types of AI models are being developed and regulated globally at the moment?

AUDREY: Yeah, I think if they are deployed in a way that is obscure, that is opaque to the downstream, then whomever deploying that owe the downstream users a duty of explainability, of steer-ability and things like that. On the other hand, if it is open source, meaning that all the way from the preparation of the data and training parameters and so on, and the resulting model is open for any purpose, and anyone can then take it to fine tune it more, to prescribe a different character, or whatever using that, then I do think that the downstream is, in this sense, not really a downstream, but rather a peer, a co-producer. And it makes less sense then to assign liability to the upstream provider. And so yeah, whether it is a service product or whether it is just in the comments for people to build like Lego blocks. I think these are very different ways to look at AI models.

JOHANNA: Yeah, I agree. And I think the different Lego blocks and this concept of an AI developer versus an AI deployer or co-producer, to use the words that you just used there is— is going to be one that we increasingly see policy makers grappling with. What about you, Glenn? I mean, in addition to working alongside Audrey on the— with the Plurality community, you also work at Microsoft. So how does that shape your views and your thinking about the regulation of artificial intelligence, and this question about the merits or the potential pitfalls of open sources— open source, sorry, in the specific context of artificial intelligence.

GLEN: Well, look, I think that both in the open source world and in the closed world, there's a desire to bridge the gap between those two extremes. So you know, the collective intelligence project that we work closely with has been working with a lot of the closed source developers to help them engage democratic participation to steer their models. And on the other hand, a lot of the open source models need to think about want to think about how they govern various elements of access and security, and they want to use methods like we used in the book to ensure the integrity of the open source, canonical models, even though other people can fork and do whatever with those, the thing that is the big global public resource needs to be something that is treated as such, and that requires not just anyone can add or take but that what is contributed is steered by the community collectively. So I think we're going to move away from these extreme, all or nothing frames that are about, you know, having a lot of debate and two things that really allow us to draw from each model the benefits: to draw the public spirit from the open source and to draw the steer-ability and, you know, governance, that is a feature often of some of these more closed approaches.

JOHANNA: And when it comes specifically to frontier AI models and the question of emergent capabilities of that very specific type of artificial intelligence. There are those that suggest that in the— in that instance, it would be inappropriate to have an open source model because we can't guarantee its safety. This is kind of the argument that in a democracy, there are still some things that we agree that are just too powerful, that that shouldn't be made available to the general public. Where do you fall on that debate?

AUDREY: I mean, if there are clear indicators that a proliferation of capabilities lead to, for example, bio or cyber or some other capabilities. They need to be evaluated in a, you know, API access only versus if you can fine tune it. And you need to compare those baselines to see how much risk it actually increases. Currently, only in the domain of synthetic media do the open source one clearly poses a higher risk, because if you have an API, you can have

provenance and watermark that sort of works. But with the open source models, anyone can just do a remix attack, a rephrasing attack, and boom, none of these methods work anymore, if you just remix it with the open source model. And so in terms of synthetic media, for sure, but whether synthetic media by itself, can be solved conclusively by a restricted API only watermarking model, well, certainly in the text domain, that doesn't look like the case.

Whereas, of course, in video or image domain, there are some successes. But maybe one simpler fix is just to change the norm, to simply assume that nothing is real by content alone, and you need to have the source itself, like if it comes from one on one, or if it has verifiable credential, or it belongs to one of the disciplines with this kind of rigour, like academic, journalistic and things like that. I mean, it was just like, you know, before the invention of photography, all we had is paintings. And no matter how realist is that painting, everybody know, it's fiction, right? And so if you trust a illustration, that is not because it looks like reality, but rather because you trust the newspaper or something like that. And so yeah, a norm change is probably needed. And once we— you do that, then the synthetic media harm of open source model again, becomes lower, and we can do society scale mitigation. But I don't think there is one size fits all solution for all of the cyber, bio, synthetic, media, nuclear, or whatever other domains, and I totally support more open and collaborative red teaming between all those different models, and if the threat is really high in one of those training runs and so on, then maybe the training run itself needs to be clearly monitored, let alone releasing as open source.

GLEN: Yeah. I'd also say that I think this notion that emerging capabilities mean that open source is worse is a little bit misleading, because the thing is that what's really going to get us towards potentially emerging capabilities is the notion that some entity, by massively expending capital and accelerating development can gain some enduring advantage. And if things are going to be open source, it's kind of hard to imagine exactly how that works, right? Whereas, if things are closed source, even if you have some

controls, those controls better be really, really, really good. Because what you're basically saying is: number one principle that we're putting on the table is that if you spend a massive amount of capital, you get power, and then we can talk about how your power is used. But if we start from the standpoint of expending massive amounts of capital, doesn't give you power, doesn't give you dictatorship. Which one is going to lead to emergent capabilities first? I don't think it's— it's clear. I'm not saying that means I'm an absolutist or anything like that. But I think that this assumption that the closed source model is going to reduce that race to dangerous capabilities, I think, is pretty misleading.

AUDREY: And also I mean safety research, although somewhat helped by other research, is its own thing, and at the moment, there's just not enough safety researchers, proportionally to capability researchers, and a lot of safety research is done with open source models, at least this generation of models that helps explain ability, interpretability, durability and so on. So just as a practical matter, to have some open source models to work with already helps safety research, which then helps to mitigate against some of those emerging capabilities.

[TOPIC: Closing statements and how to get involved.]

JOHANNA: Yeah, I think that's a really important point, and I want to double down on the fact that we really need to support and foster and incentivize AI safety research as its own field, as distinct from the development of artificial intelligence. Look, I'm conscious this has just been such an incredible conversation. Glenn, do you have any parting words, or perhaps a call to action to Australian policy makers or politicians listening to this podcast?

GLEN: Connect with Johanna and with all of the amazing people here in Australia who are building an incredible Plurality movement in the public service, in the— in politicians, technologists, civil society. There's an incredible community growing here, and you can add wherever you are, whether it be in the cultural field, in the— in

business, in government, all of these have tremendous amount to add, and we're really looking forward to all the things that you're going to give us to write about in the next edition of the book.

JOHANNA: Thanks, Glenn, that's wonderful. And what about you, Audrey, do you have words of advice for our listeners? I think it's so easy for those of us who work in this field to look at the mountain that needs to be climbed and to feel like it's too high or too far, but what you've shown us, you know, in this conversation, but also in your years of work, is that it is possible to get to the top of that mountain, but there will— I'm sure there were times for you as well, where people were saying it was too much, or it was too transformative, or it can't— couldn't be done. How do you respond to that? Or what words of encouragement do you have for those of us here in Australia working on these issues or listening more globally as well?

AUDREY: Okay, and now I'm going to quote from Dalai Lama instead, who said, and I quote: "In order to carry out a positive action, we must develop here, a positive vision. It is under the greatest adversity that there exists the greatest potential for doing good, both for oneself and for others." So the greatest adversity, which may look like a wall of impossibility for some, for us in Taiwan, really, that is our existential opportunity. If we do not overcome polarisation, if our democracy falls prey to the sewer of discord, if we do not defend against the cyber attacks well and so on. Well, there's no another turn. And, so in Taiwan, we really have no other option but to anticipate what's coming and then just make sure that we're resilient against whatever that's coming. Taiwan is caught between the Eurasian plate and the Pacific plates, and they too bump into each other. There's like three felt earthquake every day somewhere in Taiwan. But the result is the tip of Taiwan, the Yu Shan grows for half a centimetre every year, and so the building code, the resilience, the instant reactivation of internet connection or the investment into Universal Service, is not because that these are easy, that we do it. It is that it is under the greatest adversity that we have this great potential for doing good.

JOHANNA: And Glen, what do you say to people when they say your vision isn't possible?

GLEN: Don't ask why nobody's doing it. You are the nobody.

JOHANNA: I love it. That is so perfect, what a treat of a conversation. We don't do tech policy because it's easy. We do it because it has the greatest potential to have positive impact. And listeners, or whatever you do, don't be the nobody. Audrey and Glen, thank you so much for being with us today. I've really enjoyed this conversation.

AUDREY: Thank you so much. Live long and prosper.

[MUSIC]

JOHANNA: Tech Mirror is a podcast of the Tech Policy Design Centre at the Australian National University. This episode was recorded on Ngunnawal lands with sound engineering by ANU studio. Amy Denmede provided invaluable research support. Post production is by Martin Franklin from East Coast studio. Thanks for listening. Get in touch and get involved.