

Accelerating AI Ethics

6-Pack of Care: Ambassador Audrey Tang and Dr Caroline Green on the 'Civic Care' approach

Transcript

September 2025

www.afp.oxford-aiethics.ox.ac.uk

00:00:00 Green

Hello and welcome back to Accelerating AI Ethics. I'm Dr Green from the Institute for Ethics in AI. In August, we had an extraordinary conversation with Ambassador Tang Tang about plurality, a vision for AI that augments human cooperation rather than replacing it. The response to that episode highlighted a clear need for hopeful, practical alternatives to the dominant narratives of AI conflict or singularity.

Today, we're moving from that broad vision to the specific architecture. Audrey and I have been collaborating here at Oxford on a new book that outlines exactly how we build this future. It's titled 'The 6-Pack of Care, Exercising Civic Care in AI Governance'. Audrey, welcome back.

00:55:00 Tang

So happy to be back, and what a week working with you on this new microsite. It's up now. When you're hearing this, please check out '6', as in the number '6', 6pack.care and read our manifesto and the entire microsite.

00:01:10 Green

It's been such a pleasure working with you on this, and this episode is all about introducing the 6-Pack of Care and the civic care approach to the world.

00:01:21 Tang

Yeah, indeed. I think one response I got from our previous podcast was that people really loved the energy of plurality, which is about turning conflict into co-creation or co-production.

However, people also want very concrete, very actionable framework that they can put into their AI systems or evaluate the AI system they're using today, to test whether they're actually paying respect to humanity as a whole, and so I think that is why we really went very, very nitty-gritty engineering when it comes to applied ethics here with 6-Pack of Care.

00:02:02 Green

Absolutely, I agree with you.

And that's also what the Accelerator Fellowship Programme at the Institute is all about. It's about translating a philosophy into something very applied, that people can pick up and use in their own lives.

So, let's go to this central concept of civic care, which this whole approach and the 6-Pack of Care is built on. Tell me about this specific framing, and specifically care, it's often a concept seen as very interpersonal between humans caring for each other. But how does it apply to the governance of powerful autonomous systems?

00:02:42 Tang

Yeah, definitely.

When we talk about the ability for machines to achieve certain outcomes, most of those machine learning algorithms just maximise for that outcome. For example, if you teach a social media algorithm to keep people engaged, that is to say, glued to it, then as we mentioned in the previous podcast, most of those algorithms just show you the most conflicting, the most extreme, the most hate[full] content because it triggers the part of the brain that wants to keep arguing with it and therefore people become glued to their screen.

On the other hand, while maximising that outcome, it really puts in a lot of negative externalities, polluting the relational situation between people.

What we used to do, is that we follow the same people and on social media, on blogs, and so on, we can have a shared common ground to talk about. But now, when optimised for engagement, everybody lives in their own world and the social fabric became so strip-mined so that there's no common ground to talk about.

So how do we describe the class of AI algorithm that does not just optimise for a particular outcome, but rather for something that more resembles healthiness in a society?

So, I looked around in the various different ethics traditions and found that there is this tradition called 'virtual ethics' that want to imbue in an actor, a human or machine, some attributes, some properties. And care is the property that makes us to listen at the speed, the need of the actual people getting affected by those algorithms. So instead of designer deciding everything, care says to us, that it needs to be from the needs of the people affected. And 'civic care' means that it's not just the individual needs, but also the needs of the healthy relations between people.

00:04:46 Green

Okay, so this is really interesting to me; as I told you in the last episode, my research background is in social care, which refers to a practice really of supporting people with illness or living with disability or so on with activities of daily living. But it has got a lot of aspects to it, right? It's really supporting people practically, but also emotionally on so many different levels in their lives. And within social care, the field here in the UK, people in that community, so people using care services, people offering care and so on, they want to make the policy landscape much better for everybody, and quite recently, this new concept of co-production has emerged, and it reminds me of what you just said. So, co-production really is a mindset, a way of working together where different actors come together, find a common ground, and then co-create and co-produce solutions. And it was born from, the fact that often there are some groups, specifically people using care services and offering care, are less powerful; their voices are not heard as much as more powerful groups - and that may be care providers, it may be policy makers or so on. So, co-production is all about bringing people together, making them listen to each other, and then create solutions that work for each other. So, it's a mindset that challenges power structures, then puts it into practice to create something that is not only really fitting for what people need and want, but it also creates a sense of community, that relational health. This is something I'm just thinking about, how from my context it bridges to the civic care approach in quite a practical way.

00:06:44 Tang

Definitely.

And I think the word 'civic' here means that we treat this ability for people who are affected to involve themselves in the decision-making process as the main outcome, instead of maximising like profit, engagement, or things like that.

If there's something we must optimise, then it must be this civic muscle, this ability for people to act in solidarity, to co-determine, co-produce the future, and also the service from the AI machines and vendors and so on, that is part of this future, but does not dictate the future.

00:07:22 Green

Wonderful.

So yeah, going back to that example of social care; social care and AI, this is very much a reality already. People are using AI systems in their everyday lives, whether it's people who are needing care and support, perhaps in their own homes, or it may be care providers in care homes, and we're seeing these AI systems there. But I think there is a concern at the moment of various people; we've seen that in the work that we have done in social care, that these AI systems don't really fit people's realities, that they don't necessarily meet their needs. And there is that need and wish to hold on to the relationality of care. People don't want robots to take over care, but they want AI and technology to help them to live better lives, to help them care better. So, I can see that, the civic care approach here could be really helpful in making sure that we're leveraging the power of AI for the relationality of care, to help people in their care situation, but to also really build systems that actually fit people's needs.

00:08:42 Tang

Yeah, I mean, ethics to me is about telling a good system from a bad system, a morally right system from a morally wrong system. But many ethical traditions, as I mentioned, that those starting with consequences, with outcomes, start to produce a lot of side effects when we're facing a lot of power asymmetry. And power asymmetry is kind of the defining characteristic when we interact with AI systems.

Nowadays, people who want to talk to their therapist or their coach and so on find that the AI chatbots can produce 10,000 times more text at the same time unit at a fraction of the cost. And they all sound very convincing. But is it really care that it's providing? Or is it simply reliance and sycophancy? It's then very difficult to tell if you're just measuring by the outcome alone; how many words of support can they generate? How do people self-report as feeling better after seeing the chatbots' conversations?

So, then we need something else.

At a speed of thousands of time differences between a machine actor and a human actor, traditional ethics sometimes fail, as I mentioned about consequentialism, often become incomprehensible to the human and the ontology, usually the ethics about the universal rules to apply. Think of us, the humans, more like a garden and machines more like a gardener in which a plant grows at maybe one thousandths of a speed of a human gardener operating in the garden, except now the machines are thousandths high speedier than we are. Then the universal rules from the plant, most of them simply does not translate well if you're scaling it to something that's 10,000 times faster than us.

But there's the framework of care which assumes this asymmetry from the start, because, especially Joan Tronto, who has worked on the 'civic care' or 'caring with', Basically, the idea is that a good gardener must till to the tune of the garden at the speed of the garden. And that is the fundamental of the 6-Pack of Care, which is to

basically for the robot that takes care of people's relationship with each other to work at the speed of those relations.

00:11:20 Green

Okay. One thing I'd really like to explore with you more, is how the 'civic care approach' here into AI ethics challenges the way that we are looking at AI and the role of AI in our lives at the moment, and I think, you know, I've been fascinated working with you in the past week, just seeing how you integrate AI into your own life. And that idea of, specifically I find really interesting, that idea of co-presence, that AI now has this seat at a table, around the table really, so with this idea of co-presence, you challenge the norms of a physical presence in your own work, I think it's really interesting to talk to you about this because it shows that you can have a different mind frame around AI and what it can mean in your life. And I'm thinking specifically about your experience that you told me about at the UN Internet Governance Forum in Geneva back in 2017 you participated there via a telepresence robot rather than actually being there in your body as such. Can you tell us about that experience and what it taught you about embodiment and participation and also, you know, what it means to you more broadly to have AI in your life?

00:12:45 Tang

Yeah, definitely. Back then in 2017, I was invited to talk in UN Geneva in the Internet Government forum about empowering people in small islands, in landlocked countries in which internet access isn't for granted. People often need to rent very expensive satellite connection or to arrange with their neighbours when it comes to the fibre optics and so on. So, this is an area where Taiwan has extensive experience, but because Taiwan is not currently a member at the UN, at the door of the UN Geneva building, if you show a passport from Taiwan, you cannot get in. This is a classic case of something about us without us. And so, it just turns out that Double Robotics, which is a company that makes those machine doubles, their machines do not need a passport to get into the door. And so, I was in Taipei and just remote controlling that machine. When somebody raised their hands, I can turn the machine to face at them. And so, for the UN, well, they're just watching a movie, even though the movie was recorded half a second ago from Taipei. And it became the first time since [19]71 where a representative from Taiwan and a representative from Beijing spoke on the record in the same UN meeting together.

So that created a diplomatic precedent that we then keep referring to, and that enable a lot more participation. So, I think co-presence is about looking at who might be missing from the table, systematically, structurally, because of some weird political or geopolitical reasons, but then finding out ways for proxies to re-present their actual needs and their actual ideas from the place of need instead of just somebody else to represent them. So, this re-presence, this co-presence, I think is very important for co-production.

And so today in my life, I use AI agents all the time. In fact, as we're talking, we have this AI agent running on a local computer, this [is] called Daylight Computer. But I think the way I think about this is that they're all assistive to the relational health. So, for example, this has no camera, so I would not take pictures. This doesn't even have colour. So that's the reality around me is always more vivid than this screen. And it just sits in the background in a very calm way, but if I do need some reminder, looking things

up and so on, then it also has the embodied context of what I'm hearing and what I'm thinking so that it can offer me like real-time assistance.

00:15:39 Green

I definitely need a robot like that. So I can be at all sorts of different conferences around the world. Interesting. That really means looking at, you know, the way that you use AI tools in your life or beyond robots. It also means, you know, first of all, I'd like to explore with you more on how the 'Civic Care Approach', what it already means to you personally in your own use of AI and what you just said, robotics too.

But generally, I think the 'Civic Care Approach' in relation to presence, it's all about challenging the idea that it's too difficult to raise people's voices who are maybe not heard or who are hidden. And there are so many ways that we can do that these days now to really strengthen that relational health.

00:16:37 Tang

Yeah, definitely. When I was Minister for Digital Affairs in Taiwan, I often say that I take all the different sides. And so, if there is a part in a multi-stakeholder conversation that I do not understand, either because I don't have the lived experience, or I don't see things the same way, or simply because I don't speak that language, it's my fault. It's not the fault of those stakeholders. And so, I found assistive intelligence, or AI use for civic care, really helpful in that, first of all, it can help topicalise and summarise what we have talked about, and it can also translate across cultures so that even in a place where I really don't have the lived experience, nowadays language models can find metaphors that help me to understand.

So, re-telling their story from something that I have personally experienced, and now people have applied this method to what's called the sense-making tool, so that you can have like 10 people sitting around the table talking about their lived experience. And almost in real time, it can make sense of that conversation and reflect back to the people so that people who do not have the same lived experience can nevertheless understand what are they even talking about. This way, it is not to replace the people-to-people connection, but rather augmenting our civic muscle to make sure that we can understand and be more empathetic across longer cultural distances.

00:18:18 Green

Great. Okay. In the past minutes or so in the introduction, you have introduced to us that 'Civic Care Approach' and how it challenges so many ways in how we currently think of AI in our lives, but also giving us that idea that we can really build that relational health using that approach with a different mindset of co-production, really.

Now, let's dive into the book's framework, the 6-Pack of Care, and these are:

1. Attentiveness
2. Responsibility
3. Competence
4. Responsiveness
5. Solidarity
6. Symbiosis

1) Let's start with the foundation, Attentiveness.

You write, 'Before we optimize anything, we choose what to notice'. Why is this the crucial first step?

00:19:12 Tang

Yeah, I think of attentive AI as an AI system that is curious about what the actual needs are. Instead of starting by optimizing for certain things, like a score of engagement or something, it first asks, what is the local context? What are the marginalized groups? What are their lives like? So instead of prescribing anything, it just engaged community using the methods which just talked about, the bridging maps, the uncommon ground, the idea of sense-making, all of them are about creating a group selfie rather than just a statistical average. And so, in the context of online social media, for example, there's now more and more social media trying out community nodes, which is instead of just showing the most viral post that stirred up outrage and so on, it's now crowdsourcing what is the most helpful local context to attach to a viral node. And the most viral node is then accomplished with something that's called the uncommon ground, which is people who usually disagree, manage to agree that some nodes are helpful. And this kind of bridging map is a much more accurate group selfie because it shows what difference there are in different groups, what splits us, but also what unifies us despite the splitting.

And so attentiveness is to both, the unique needs of various different sub-groups, but also the overarching bridges that those groups may not themselves be aware, but after attending to those common needs, they all become aware, oh, these are the common needs that if the AI system go and overcome those issues with us together, it will not infringe on the important things that we also consider sacred and an AI system should not disrupt like human dignity.

00:21:18 Green

So, when we do listen, but then there's no action after that, you say to you that is theatre, right? That's something you've argued in the past.

2) Would you say once a need is noticed, the next step in the 6-Pack of Care after attentiveness is responsibility?

How do we move from recognising a problem or a need to making a verifiable commitment through that 6-Pack of Responsibility.

00:21:53 Tang

Indeed.

So nowadays, when people are developing AI, many frontier labs have this thing called 'model specification' or 'model spec'. And such spec is a way to make a public pledge to the public, to the relevant community, that's what this AI system is trying to do. And also, what it pledges not to do, for example, flattering someone into self-harm, is something that many chatbots have now publicly pledged to not do. And a few years ago, we introduced this idea of alignment assemblies in which a statistically representative slice of the public come together, and in groups, they look at each other's suggestions for the AI system code of conduct, integrating those into their own lived experience and upvoting or downvoting them, and after a few rounds of deliberation, settle on this kind of bridges of what AI systems should do in their community and should not do in the community. And so that is the 'attentive' part. And if we can bring those attentive results into the pledge that the vendors make, then those vendors can be called responsible AI vendors because they take responsibility on fulfilling the parts of the alignment assembly co-production that they can do. And also, they draw a very clear boundary saying, we're just fulfilling this much and beyond which is not our scope.

00:23:36 Green

Going back to, when I listen to you explain the 6-Pack of Care, I always go back to the work that I mentioned earlier that I've been doing in the past two years or so, trying to find out through a process of co-production what the responsible use of AI in social care means.

And what we have done in those past two years is, and as I said, bring people from the care community, so people drawing on care, people providing care, care providers, policy makers or so around the table to get that common ground, to talk to each other, to make sense. And so, to be attentive to people's needs, but then also to be responsible. So, you know, I'm fascinated by how this process has been happening mostly without AI, all about AI, basically without it, and how we could now integrate AI into it.

And within that process too, so coming to the, you know, attentiveness that was all about hearing people's voices on an equal basis and sense making, how are people already using AI in care in their everyday lives, what's helping them? What are they worried about? What are their hopes and dreams? And to really pinpoint those needs, the responsibility was then about the various groups of people really finding themselves within that landscape of responsibility, and what it means for them to be responsible. So, we, for example, had tech providers in this co-production as well. So, people who are building AI system specifically for caregiving. And they said, well, you know, we really want to show people that we are responsible. So based on our definition of the responsible use of AI in social care, which is all about, you know, supporting the values of care, not undermining or harming values of care and human rights, they said, yes, well, that's what we want to do, but we actually need to think what that means for us in practice. And we want to show people that we are responsible. So, they created a pledge. And now tech providers are signing up to that pledge and not only signing up to it, they also then need to be accountable to it. You know, they're putting systems in place in order to make sure that they actually meet whatever they have pledged. So I find that interesting and how much we can learn from that experience of co-production for the 6-Pack of Care.

00:26:23 Tang

Yeah, definitely. I think what you're describing is exactly the right process in which that the pledge, as I understand, includes not just the scope of commitment, but also real timelines like we will deliver this by this time, the guardrails, remedies, accountability mechanisms, and hopefully also independent oversight. And the civic care mechanism that I just described is more about an automated translation. So, it's not just that people come together and have a conversation, but rather in real time they see a bridging map from the attentive AI that reflects it back to them without having to wait for a few months for human sense makers to code this up.

And once this is done, then we translate those very legible (in English or other languages) specification into something that AI systems can already comprehend. So, instead of human programmers programming each of those pledges into their systems logic, it is becoming a curriculum, so, that when AI systems train on those data, materials, and so on, the data that goes in those AI systems are already filtered by the lens of those pledges so that the AI, once it's being trained by those data, some synthetic data, some real human data generated with deliberation, then they intuitively, so to speak, understand what is moral for this group of people.

00:28:03 Green

So, there's one question that I am burning to ask you. I've told you all about the co-production on AI and social care without AI mostly. How could we re-envision this or re-envision this process, this work, now using AI?

00:28:24 Tang

Yeah, definitely. So, I understand that as we're recording this, this weekend, you will continue this deliberative polling with actual care receivers and people in the social care ecosystem when it comes to their preferences on those companies.

So, one very quick introduction we can make is because you do have the transcript of all those deliberative polling sequences, instead of just taking the quantitative data of the final poll, we can also take the qualitative data of that conversation and feed it into the sense maker, which is open source. And the sense maker would be able then to annotate the votes that people give at the end of the deliberation and nuance those decisions into, for example, people agree that these are important principles, but people, new ones, start saying, but when deploying, you need to watch out for this, watch out for that, and so on. So, you generate a much richer, much more thick data set when it goes to those vendors. They still sign on to the same pledge that was the voted in results from the deliberative polling. But then when they're training their AI systems, then those AI systems can additionally agree on the thick context that is provided by the deliberative process itself. I love to introduce the use of sense making and the bridging maps and more importantly, reflect that back to the people so that people can maybe chat to a care assistant and ask what was agreed on the previous deliberation. And now I have this real situation. How do you compare against that and so on.

00:30:09 Green

Okay, yeah, I'm really excited about the next step using the Stanford AI Deliberation Platform in this process.

3) Let's go on to the next pack, the third pack called Competence. And here it's really about engineering meeting with the ethics.

It's about understanding how, you know, we make sure that AI systems don't only have good intentions, but that they actually deliver care in the sense of the civic care approach, not social care, I obviously need to make sure that there's this distinction there.

Tell us more about competence.

00:30:48 Tang

Competence to me, means empowering the people receiving the care, not just addressing their needs, but also making sure that their relations also flourish, also thrive.

So that means, for example, that if you want to serve the needs of people when learning, it is not about introducing an AI that can do all their homework for them.

I mean, that's very powerful, but it's not very competent because a competent teacher is not just about solving all the homework for the students, but rather working at the speed of the student and making sure that the student is curious about these disciplines, these different fields, and also introducing the student to more communities of knowledge, of practice, introducing students to more peers for peer-based learning, more purposes for purpose-based learning, and so on. So, a competent teacher or a gardener, I think, is a very good kind of metaphor for a competent caregiver. When AI is in this situation, I think it's very important for us to see whether the AI is taking away the existing human relations because people come to depend on AI

more, or whether it is encouraging the human to grow more other human-to-human relations.

00:32:12 Green

That makes me think of another example that I have recently come across, and that is, systems that are supposed to support people with disability or people with frailty or living with stages of dementia to support them in their own homes, so that they can stay more independent, more autonomous for longer. And what I've seen here is that the intentions behind these systems are great and generally people are very excited because people want to keep on living in their own homes as independent as possible for as long as possible. But concretely, I'm thinking of a case where somebody on one day asked their AI, oh, can you help me with remembering something? And, you know, the system is helping them very well. But the other day, that same person fell and they had someone there to support them with the fall. But, you know, the supporter tried to get some help through the AI. What can I do to help someone up? And the AI told them, oh, you should call an ambulance. So, they called the ambulance, but there was no ambulance to come because we've got a shortage here of emergency services, especially during winter months and so on. So here, one day, the system really helped the person, but on another day or another situation or in another context, it wasn't very helpful. I think it's that this is really important what you're saying here and also what people I've spoken to are also concerned about when it comes to AI and caregiving in people's own homes, is that it might mean that they will be cut off from some of the many important relationships they may have with, you know, caregivers who are coming into their homes because now the policy is, oh, we only need a caregiver to come in once a day instead of twice a day because there's an AI system supporting or, you know, so that concern of that, that it could actually work against the relational health of people. So, how important it is here to really think, how do these AI systems work for people in their contexts and their situations?

00:34:30 Tang

Yeah, definitely. And I think that is the fundamental difference between the previous age of what's called reinforcement learning with human feedback in which ChatGPT was born, because they rank the various possible outcomes of ChatGPT to any given prompt by asking individuals to rank this response as better than that response. And once they work with a lot of people to basically label good responses and bad responses, it produced something that would consistently get good score from human labellers, but also it made ChatGPT extremely flattering, extremely sycophant[ic], because most people, when they are in this kind of individual-to-individual dyadic relationships, they optimise for this short-term satisfaction. And so, if something that flatters me, oh, of course I'm going to let it pass. But if something that really checks my understanding, something really pushed back against my hallucination, maybe I'll just give it a thumbs down. Right? And so a AI that is trained from individual human feedback may not prioritise the relational health of that human embedded in a community. So now we're seeing more advances in like multi-agent settings, where the agent is learning not from individual feedback, but from community feedback. And so, the situation you talk about where a person is not just alone in their home, but rather have a community of helpers, a community of people who support them with mutual support and so on to prioritize these relationships of culture, family, and so on as the main unit of analysis and then doing reinforcement learning to get the better

score, not from a thumb up or down, but from a mutual recognition and mutual assurance from those communities. This is a very active area of multi-agent research.

00:36:39 Green

Yeah, and I can also see here how we then need to adopt A mindset of building systems that is about people's needs and learning from their needs and wishes rather than based on assumptions, for example, of particular groups. So, there's a growing community now interested in building systems specifically for older people. But we don't just have a group of older people where everybody is the same. Yes, there might be some, health conditions that are more likely to develop in older age or so, or, our bodies change as we age, of course. So, there may be some kind of common grounds that we can think about for older people as a group, but otherwise everyone is different. So if we're wanting to build systems that help older people in their situations, we should think more about those localized, those needs, those individual needs, family health, relational health, and so on.

00:37:36 Tang

That's exactly right.

Like if I am a very old person, but I do not have this particular frailty and so on, for an AI system to simply assume that I have and provide all sort of help is really taking away my agency, and it is a kind of discrimination, really. And so, to overcome those very simple, very thin labels, stereotypes, and so on.

It's exactly why we need this community, communal AI alignment, because then it's the only way to provide the thick context from that particular community.

00:38:13 Green

Let's move on to the next pack.

Once we know that a system isn't perfect, right? And no system is perfect, really. It needs to be changed. It needs to adjust. How do we design systems that are correctable and adaptable at the speed of society?

4) Responsiveness

00:38:33 Tang

I think the most important thing here is to empower people closest to the pain. That is to say, people who are suffering the harm should be the people who define the harm. So instead of just the designers trying to figure out how to make sure that it does no harm to everyone, it is simply not possible. All the unforeseen consequences will harm someone at some point. But what we want is that people do not mitigate against those harms thinking; 'this is just how tech works. Tech just happens to us and we overcome doom scrolling ourselves by using grayscale screens or using a stylus instead of your fingers'.

I mean, all these [interventions] work to mitigate doom scrolling as a harm, somewhat, but it shouldn't be up to the individual people suffering the harm to correct a structural oversight. And so, the idea here is that we need to make something like a Wikipedia before evaluating those harms. So, the Collective Intelligence Project, which I work with, have this website called weval.org, where communities can author localized benchmarks for harm. So, there you will find, for example, people in Sri Lanka, crowdsourcing what is important for them in a local context so that the chatbot can understand what's important to that community. There are communities that are suffering from like psychosis, mental health issues, and they want to make sure that the AI chatbots do not exacerbate the psychosis issue and so on.

And so, this then closed the loop because once those individual harms are foregrounded and made into something that everybody can see, then we can continuously run those benchmarks against the frontier models so that when they release a new version, it doesn't only compete on the scoreboard of, for example, solving the international Olympiad on mathematics or something, but also make sure that it doesn't regress on those benchmarks where it really did cause people harm.

00:40:51 Green

Okay. Yeah, that's our next step in the responsible AI and social care project was to create a database where people can report their experiences, be it harms, but also what they find helpful. So, I think we need a WEVAL here.

00:41:08 Tang

Yeah, we do need a WEVAL here so that, when the next version of those technologies come, you can run in a simulated environment, of a simulated person, in exactly the same situation and then evaluate how the next version of the robots will respond to it and to make sure that they continuously run this test so that it doesn't cause more harm once they release the second version out to the wild.

00:41:36 Green

Okay, so let's move on to the last two packs, which really look at the ecosystem that we need for the civic care approach.

4) Solidarity is the 5th pack, and it's about making cooperation the path of least resistance. So, how do we build this infrastructure to counter the trend of monopolistic power?

00:42:02 Tang

Yeah, so the first four packs form a loop, right? You become attentive to the group's needs. You respond to those needs. You're competent, so you deliver and fulfil some of those needs. And in fulfilling it, you maybe cause some harm and then very quickly pivot and in a responsive way go back to attentiveness.

One thing though is that if the power asymmetry is too big, if we do not have a choice between receiving care from a tech vendor there's no way to switch to some other vendor, there's no way to bring our data from this conversation to another conversation, then there's simply no incentive for the monopolistic provider to be attentive anymore. And this happens again and again when a service is small, like when it's an upcoming social media platform, usually it's very attentive, it's very responsive. But at some point, when the majority of population is on that platform, suddenly it doesn't care about you anymore, right? And this is not something that can be overcome with engineering alone; usually it takes policy. Here, people already talk about 'interoperability' a lot, for big tech systems. I can message people from WhatsApp over to Signals or from Android to iMessage and so on, because there's mandated interoperability, so that, for example, when you switch your phone from one telecom provider to another, you can bring your number with you, and that's called number portability.

Imagine, without those mandates, it would simply be impossible for new telecom to enter the market, because it's just too much inconvenience to switch from one vendor to another.

And so nowadays, we're now seeing policymaker[s], for example, in the state of Utah, mandating that starting next July, you can switch from one social network, like x.com, to another, like Blue Sky or True Social, and then bring all your existing contact with you so

that all the new follow, new react, new posts, and so on that you follow carry over to the new network.

And so that is the idea of interoperability. And we also need the idea of selective disclosure or meronymity, partial anonymity, because one part of sharing our lived experience is to identify the harm that is caused by those technologies.

But if we have to sign each report with our real name, we're essentially doxing ourselves. And most people who are in a vulnerable situation would not want to contribute their lived experience to course correct those technologies simply because they would not trust, right, the people who process those data.

But if we do meronymity, selective disclosure, then we can make sure that it comes from somebody who claims to be a resident in this country, who claims that they're between 18 years old and 50 years old, but then you do not need to learn anything more about it.

The digital signature would carry the attestation that this person really generated this. 'proof', but there will be no extra information that is given, so, interoperability and meronymity together, provide a good infrastructure for agentic AI, so that we can know each chatbot, each technology comes from some kind of vendor, what kind of pledge did it sign? Which community is this pledge going into without revealing the private details of any of those community members participating in the deliberation or the wiki eval.

And you can easily switch from one vendor to the other, porting all your existing data, as well as keeping the pledges that vendor A made and maybe broke and carrying it to vendor B.

00:46:25 Green

So surely that would actually also incentivise, looking at the market competition, that should incentivise providers then to ensure that their systems work, they're competent.

00:46:39 Tang

The race to the top, right now to the bottom of brainstem.

00:46:45 Green

So, the last 6-pack is symbiosis.

And that one offers a direct challenge to the singularity narrative that we got discussed last time. So that, overpowering one powerful AI system, so you introduce a really powerful metaphor from the Japanese Shinto, the kami.

What are kami in the context of AI? And why do you think that this vision of the kami is preferable to the singularity narrative?

00:47:19 Tang

Yeah, kami here is used metaphorically as a local steward, and specifically means an AI system that is bounded to something that is place-specific or community-specific that cooperates with the relational health. So, all it cares is about the relational health of this particular community. we talk about how community nodes can bridge a community together. And an AI that drafts such community nodes is this kind of kami. It simply does not worry about paving the galaxy with paper clips or with solar panels or anything like that. It's not trying to grab power by itself. All it cares is about the relational health of this community. And if the community have fully healed, if the community have moved on, if there's no such community anymore, then the kami can be switched off with no regret.

So again, in ethic terms, it means that it optimizes for the virtue of relational civic care.

It is not optimising for the consequential outcome of maximum happiness in the universe.

This is the horizontal alternative that we call plurality of many local kamis, in an overlapping way, taking care of each garden instead of a vertical superintelligence supreme gardener that doesn't know what's enough. So, this 'enoughness', this boundedness is the defining characteristic of symbiosis. Otherwise, it's parasitic.

00:49:04 Green

So, before we started working with each other, with the institute created a care assistant called Dedicate, and that's rooted in many conversations with family caregivers to understand better what it is that they need to feel supported in their everyday life as a caregiver.

Because, you know, often people are faced with a lot of challenges, happy moments too, but challenges in their caregiving journey, whether it's for themselves and taking care of themselves or the person they're caring for.

So, the idea behind Dedicate, again, it translates quite really well onto specifically that symbiosis pack, because what it's all about is to create connectedness between, so using AI, to create connectedness between organisations, people that are out there to support family caregivers with advice and information.

We built a system where we have that content given to us by them. And then we have got an AI on top of that, which makes the information easily accessible to family caregivers in the moment when they need it most or when they want to engage with it. But that chatbot also has its limits. So, if somebody types in a question around a very specific medical question or a question that is not something that's commonly found within caregiving at all, that chatbot will say, oh, I'm not, this is not my area.

00:50:43 Tang

Yeah, I tried it out. I went to dedicate.life and asked what equipment should I bring to my podcast. And it says that I don't know.

00:50:53 Green

I would have been surprised if it had given you an answer there.

But I'm really excited to have found that Dedicate and the idea behind it is so much aligned with the civic care approach, you know, the way that we went about it, raising people's voices to hear what it is that they need, and then to build a system that connects them with what's out there in the world. So, they actually also feel like, oh, I'm not alone with this, and, you know, really have that idea of, yeah, the kami really speaks to it there because it's about that particular community.

And Dedicate right now is for people living in the UK or England even specifically. But we could build systems like that in other contexts and other cultures where caregiving might be different.

00:51:40 Tang

Right. So, it forms a federation of kamis, instead of trying to build one simple super intelligence that knows everything and can do everything, because that simply does not exist, right?

Usually when you try to build a super solution that works everywhere, it ends up just coercing every place to look just like the place where the designers lived in, and then just colonize, right, the rest of the world. And so, starting from symbiosis means to know does boundedness 'enoughness' is actually better for people's life for relational health?

00:52:18 Green

So Audrey, we have learned so much about the civic care approach.

We have learned about the framework, the 6-Pack of Care, which is ready to be used, which has been used and tested actually in various contexts. So, it's something that's out there for people to pick up.

And I think there is a real urgency for this approach. Again, going to, my world of research, I also work on rights, human rights and older persons. And there's this really exciting new process. Well, it's not a new process, actually, but we're taking a new step, a powerful step towards potential new convention on the rights of older persons within the UN.

And working with you, I've been really excited about the possibility and opportunity of using AI now to have global voices of people of all ages, because we all age, aging is a matter of every human being, right? Not just older persons, but to also raise older persons' voices from around the world to inform this process within the United Nations that's on the way to form this convention. And again, I think it's urgent, it's here. We've got ageism as one of the big types of discrimination that is happening around the world and that needs to be addressed, and really urgently, and again, the civic care approach here is helping us to think this through on how can we do that in a meaningful and quite fast way, right? Just responding to how quickly we are moving with AI and just how important it is to actually find solutions now.

Where do you see urgency and how to use Civic Care or opportunities right now, what would you tell our listeners to do now once they've listened to our podcast and learned more about the Civic Care approach?

00:54:28 Tang

Well, first of all, check out our microsite, sixpack.care, that links to some nice manifesto, as well as concrete tools to start using.

Anyone can run a local alignment assembly; anyone can tune a local AI model using the result of that alignment assembly, and see for themselves whether this kind of communal AI communal model responds better to your community needs; anyone can advocate for social portability, for interoperability between the providers, and none of this is theoretical. In the Frontier Labs, I mean, we work quite closely with parts of Google, for example, Jigsaw, and I just gave a talk at DeepMind. And there's parts of Microsoft Research that focus on plurality research in the collaborative. OpenAI, Anthropic, all have run alignment assemblies with the CIP.

So, I think it's not that the Frontier Labs do not know of these approaches. It is just that the policymakers now need to understand that this is not science fiction. This is something that each lab in their research capacity have already done quite a few iterations of this loop. And so now the policymakers ask must be to make this the baseline. If the AI product is not attentive, then it's broken. If an AI product is not responsive, then it's broken.

And if the entire ecosystem of AI agents do not work in solidarity, but rather are engaged in a kind of race to destroy each other, then that's not the best idea, right? So, to have symbiosis and solidarity as the bedrock of our tech policy, I think this is also very important. And so I think it's great that we have this podcast.

And it's great that we have this microsite that can refer the policymakers, the engineers, researchers, and people who are affected by AI into this shared vocabulary that is at once philosophy, but also policy and engineering.

00:56:46 Green

And what gives me so much hope with this approach is just, you know, the idea that we can build better relationships again in the world where it seems that, often we're just too polarised, where we don't work with each other, collaborate with each other, be with each other, the way that really makes us happy as humans, as relational beings.

00:57:13 Tang

Yeah, indeed, this lack of civic care, manifests as polarisation. And this feeling that we're all polarised, that we cannot make coordinated actions, that people just hate each other on social media and so on are like a large web that coats the entire tree and deny the tree from sunlight and so on. I think it's time to step out of the decade-long illusion that's caused by the polarising social media and then try civic care for real.

00:57:46 Green

Yeah, thank you so much, Audrey, for this podcast, for introducing the architecture of civic care. It really is a very compelling and very necessary vision. And in the spirit of relational health, as you already mentioned, people, anybody listening, coming across this work is more than invited to engage with us through our microsite and the 6-Pack of Care. We'll also make sure that people have the link to that.

And again, thank you so much.

It's so great working with you.

00:58:22 Tang

Yeah, same here.

Live, flow, and prosper.