THE BROOKINGS INSTUTUTION

WEBINAR

GLOBALIZING PERSPECTIVES ON AI SAFETY

WEDNESDAY, FEBRUARY 19, 2025

UNCORRECTED TRANSCRIPT

PANEL:

GRACE CHEGE, Junior Research Scholar, ILINA Program

MAIA LEVY DANIEL, Senior Program Manager, Trust and Safety Foundation

BEN KEREOPA-YORKE, Master's Student, The University of New South Wales

JAM KRAPRAYOON, Strategy Manager, Institute for AI Policy and Strategy

CRAIG RAMLAL, Lecturer, The University of the West Indies, St. Augustine Campus

MODERATOR: CHINASA T. OKOLO, Fellow, Center for Technology Innovation, Brookings

*****

OKOLO: So thank you all for being here for this webinar today and good morning from Washington DC. I appreciate you all for joining the Brookings Center for Technology Innovation for our webinar on globalizing perspectives on AI safety. I'm Dr. Chinasa T. Okolo, a fellow at Brookings within the Center for Technology Innovation. Over the past few years, AI safety has emerged as a field dedicated to ensuring that AI systems operate reliably, ethically, and beneficially. However, AI safety narratives often prioritize Western perspectives and technical risk, neglecting broader societal harms and diverse cultural context. This results in AI systems that perform poorly in diverse environments and also reinforce systemic biases. To critically discuss these issues, we're launching the AI safety and the global majority project with this webinar, along with a series of published commentaries on this topic that have begun to roll out over the past week and will continue over this week. To provide context on how the webinar will go today, we'll start off with short opening remarks from our project contributors, transition into a moderated panel discussion, and end with an open Q &A session. We've already received lots of great questions, and I'll do my best to convey as many questions to the panelists as time allows. If you're watching live, you can submit questions via X by tagging @BrookingsGov with hashtag AI safety or email events at Brookings.edu. To get started, please welcome some of our contributors to the AI safety and the global majority project. First up, representing Africa, we have Grace Chege, a junior research scholar within the ILINA program. Next, for the Caribbean, we're pleased to welcome Craig Ramlal, a lecturer within the University of West Indies St. Augustine campus. And next up, for Southeast Asia, we're also welcoming Jam Kraprayoon, a strategy manager within the Institute for AI Policy and Strategy. And for Latin America, we have Maia Levy Daniel, a senior program manager within the Trust and Safety Foundation. And last but not least, we're very pleased to welcome, representing Oceana, Ben Kereopa-Yorke, a master's student at the University of New South Wales. Next up, we'll have our panelists present their opening remarks for the respective webinar. And we will start off with Oceana, move to Latin America, then Southeast Asia, Caribbean, and end with Africa. So Ben, I'm pleased to have you present your welcoming remarks.

KEREOPA-YORKE: Thank you, doctor. It's great to be here. I begin by acknowledging the traditional custodians of the land from which I join you today, the Turrbal and Jagera peoples of Meanjin, Brisbane. I pay my respects to their elders past, present, and emerging, and acknowledge their continuing connection to land, waters, and community. I speak descended from lands where steam rises from the earth, where my people, Te Atua and Ngapuhi, have maintained sophisticated systems of resource management for centuries. I speak as manuhiri on Turrbal and Jagera country where I witness another people's deep understanding of sustainable technological systems. These aren't relics of history. They're living alternatives

to the computational colonialism consuming our Pacific today. While data centers drain our water and power, Indigenous frameworks like te matauranga offer proven systems for managing resources sustainably. The same forces that once claimed our lands now claim our digital spaces. We've seen this pattern before. The AI safety movement has failed, not because its goals aren't important, but because it has become a form of technological theology that many pushed while ignoring present harms and chasing undefined futures. Standing on the shores of country, watching waves inch closer to ancestral lands, I witnessed their silicon theology being built. Techno saviors conjuring fears of digital gods while real waters rise. Across the Pacific region, Oceania, 16 nations face an existential collision between rising seas and computational colonialism. They say AI safety is about controlling future systems. We say AI safety is about controlling present damage. While Tuvalu implements its Future Now project to preserve culture against territorial loss, data centers consume the equivalent of 50,000 homes worth of energy annually. While our nations face water scarcity, their AI facilities drain up to 6% of district supplies. The mathematics is clear, not in theoretical proofs of controlling undefined systems, but in the thermodynamics of resource consumption. They treat our digital spaces as terra nullius, empty territory awaiting salvation through undefined concepts of alignment and control. But communities are already showing another way, from artists protecting cultural expressions to indigenous frameworks managing distributed systems sustainably. They say we must prepare for tomorrow's artificial gods. We say we must address today's artificial costs. Having survived colonization, indigenous peoples understand what it means to be living on one technological plane and then be exposed to a radically different one. Yet our worldview is not requested nor understood. We are viewed as an extraction source rather than colleagues. The colonization of our lands and language was not enough. Now our cognition, what makes us, us, is forcibly taken. The very idea of safety has been colonized. Safety for whom? Safety from what? Safety where? Meanwhile, the smoky air, the last creatures of their kind calling to dead cannon hoping for replies that will never come. Rising seas, climate refugees, yesterday's allies denying the very truth of events we see with our eyes. The real silicon fetishism is not measured in dollars. It's in the birds, the sea creatures and mammals that exist only in yesterday. Some say these models will transform the world. We say they already have. Look at the land. The path forward isn't through speculative technical solutions or elaborate ceremonies of control. It's through converting impact from corporate checkbox to community power, through making consent and contestation fundamental, not optional, through building tools for community self -determination. AI safety is not a technical issue. It's a governance one. It isn't a future crisis. It's a present reality. It isn't about controlling tomorrow's systems. It's about being accountable for today's impacts. The greatest risk isn't AI. It's artificial exclusion. While computational infrastructure consumes our resources, indigenous knowledge systems offering proven frameworks for sustainable technology are

systematically ignored. The waters are rising, both literally and metaphorically. We can either waste our remaining resources on silicon theology or we can build practical frameworks for community technological self -determination. The time for failed experiments is over. Thank you.

OKOLO: Thank you so much, Ben, for your great opening remarks. And so next, we will have remarks from Maia for Latin America.

LEVY DANIEL: Thank you, Chinasa, and thank you for having me today. So I would start by saying that there's no clear definition of what AI safety means for the region. I think there isn't even a perfect translation for the term in Spanish, actually. And in Latin America, as happened in other regions as well, AI governance initiatives have flourished over the past years. There are AI strategies, local regulations, policies, ethical frameworks, as well as recent regional initiatives that aim to achieve responsible AI. And in terms of what we can find in these initiatives, AI safety usually refers to those requirements for the AI systems and obligations for different responsible parties throughout the AI lifecycle to ensure that harms are minimized and human rights are protected. So there are some requirements in some of the current bills, such as risk evaluations, transparency requirements, algorithmic impact evaluations. But in general, they are very generic. They do not provide any details on how to implement them or any actionable steps. And these requirements are not consistent throughout the different bills. So the conversation is currently focused on risks related to biases, to fairness, to lack of transparency, and the importance of privacy and human oversight. But there's not much on what we be asking companies to do to minimize those risks. So there's, for instance, a bill in Brazil that has already been passed by the Senate that includes a specific section of safety and addresses transparency, obligations to generate documentation and implement evaluations. And there's also a section on algorithmic impact evaluations with a detailed methodology for its implementation. There are other bills in Argentina and Costa Rica, for instance, or the Peruvian law that require some impact evaluations, transparency, and the creation of public registries. But except for the case of the Brazilian bill, which was the result of a three -year discussion that included the participation of relevant stakeholders, requirements are very generic and are not consistent throughout the different bills. So the meaning of AI safety or what practices which should require from companies developing and implementing these AI systems that, in general in Latin America, those AI systems are being designed and developed outside, they're being developed abroad. This definition is not entirely clear in the region and there's a lot of work to be done to achieve a comprehensive understanding of what it implies and how to achieve it in the region. Thank you so much,

OKOLO: Mala, and I appreciate your opening remarks. So next up, we'll have remarks from Jam.

KRAPRAYOON: Hi, thanks, Chanasa, and thank you also to the rest of the Brookings team for inviting me to join the session and also obviously to contribute an essay to the topic. Just really excited to be involved. Also, just to note that my co-author, Sean, couldn't be here today, but I'll be speaking on both of our paths. The main point we make in RSA is that transatlantic countries, so the US, UK, EU, which have been leading at the moment on AI safety efforts should be doing more and engaging more with Southeast Asia as a region on AI risks and that this would be mutually beneficial for both sides. We make this case specifically with reference to one risk where we think that frontier AI will be able to very materially affect the world in the near future, which is uplift in the ability of various actors to conduct offensive cyber operations. The idea being that generative AI, frontier AI will be able to act as a force multiplier for various kinds of criminal activities powered by enhanced cyber capabilities. At the moment, we're seeing a little bit of that in terms of the use of JMAI, mostly demos of how, for instance, you can use certain kinds of large language models to assist in things like automated spearfishing or the generation of persuasive emails to use in but as capabilities advance, we would reasonably expect this trend line to go up. There are three sorts of actors that we talk about and we dive into one in particular cyber criminal networks, which are particularly important to Southeast Asia as a region. As some of you may or may not know, criminal networks are engaged in large - scale scams and scam centers being based in places like Myanmar or Cambodia, where there are already hundreds of thousands of people involved as operators. I think it's estimated that even in just the US, these scam centers are costing Americans around billions of dollars per year. We can expect that you can use generative AI tools, deepfakes, voice cloning, automation of phishing to enhance the impact of these operations. There are also other groups like terrorist groups or nation states that are also obviously a concern to Southeast Asia as a region. You can imagine that generative AI could be used to improve recruitment or generate propaganda material or just proliferate offensive cyber capabilities to actors that otherwise be too low resource to use them in the current day. Given this emerging concrete risk, what could be done? We advocate for three main lines of effort. The first is evaluation, the second is localization, and the third is adaptation. With evaluation, again, as many of you already know, many safety evaluations are being run on Frontier AI systems now. For these to continue to be effective in the cybersecurity space, they should also begin to integrate regional expertise. This is important not to just handle linguistic diversity but also to draw on local countries' knowledge of the operating environment in which cyber crimes actually happen. Creating assessments that take into account how cyber criminals actually operate today and also

how likely and in what ways will they embrace AI power tools, which will help us create a more grounded assessment of risk and uplift. With regards to localization, as many people have written about already, it's important to localize safeguards so that people are able to successfully use jail breaks when they're attempted in low resource languages, which we would expect to also be the case in Southeast Asia. In terms of adaptation or assisting adaptation, the risks from AI are not only going to affect the global majority or the global north or western countries, however you might want to call it, and also the efforts to sort of harden our defenses against some of these risks will also be a pressing concern in Southeast Asia and other countries as well. The only issue is that many of these tools are cost prohibitive. By default, we expect them to be not just in terms of inference but also in terms of training and development of these tools. When you couple that with the region's issues with poor digital infrastructure and limited public resources, this is obviously a big problem. So I think there's a very real opportunity here for partnerships between frontier AI companies and western countries to sort of share distilled models, specialize AI tools, and also to provide compute credit and kind of tech transfer or sort of talent sharing schemes between countries in order to promote safety - promoting applications. Thanks.

OKOLO: Thanks so much, Jan. And so now we'll go to Craig.

RAMLAL: Thank you. Thank you for having me and thank you for allowing me to contribute in this initiative. I listened to the other panelists and we talk about some of the definitions of what AI safety is, and I recall that we don't even have an accepted definition for what AI itself is. I remember serving on the UN Secretary General's High Level Advisory Body on AI and that was one of the calls. What is AI? How can we define this? And then how can we develop what AI safety is? And to me, AI safety is a very broad term. It aligns with societal well -being, ethical principles, the protection of fundamental rights. So it extends beyond the technical robustness that we currently see mostly in the world of what we try to standardize AI safety to be. It should include issues such as cultural sensitivity, human -centered design, and socioeconomic impacts. I think that like Ben in the Caribbean, we also have a unique reality that necessitates a broad definition of what AI safety should be. We do face challenges that are not priority to other regions. These include linguistic and cultural misrepresentations. That is because most of the training data is not available for training in the private organizations, it would seem. So things like Creole and Indigenous languages are not there. So we have cultural misrepresentation because of that. We have dependency on foreign AI systems. This leads to a lack of control over our local data and also leads to economic vulnerabilities where automation threatens some of our key jobs like tourism and business process, outsourcing, call centers, so

on. We have that in our region. We have issues with infrastructure and climate challenges, including unstable internet, power outages, disaster risks that make AI implementation very difficult. And of course, we have limited access to AI resources such as the high -performance computing and the advanced tensor processors. These are because of global shortages and I'm very sure we can get into that later on. So I think if we really think about AI safety and we are looking at an inclusive approach to AI safety, we really need to reframe globally how we think what AI safety is and it needs to focus on additional issues such as data sovereignty, economic resilience, local AI capacity building and environmental sustainability as well. Thanks.

OKOLO: Great. Thank you so much, Craig. And then now we'll wind up our intro remarks with Grace.

CHEGE: Thank you, Dr. Chinasa, for the introduction and Brookings for giving me the opportunity to present my work today. I'm also really looking forward to hearing everyone's contributions. So my piece discusses open access AI as a strategy for advancing AI safety in Africa. Open access AI in this instance refers to any form of model sharing spanning across the gradient of model releases and practices. So this means from more restrictive modes of release such as cloud -based access and API access to more liberal or open modes of access. Open access AI proponents laud it as a public good, reason being it presents a means to democratize AI development, use and profits by allowing demographics that do not have the critical mass and resources to develop foundation models from scratch to participate in AI development. And African countries have been beneficiaries of this. They've been able to leverage open access AI to innovate and come up with local AI solutions in response to of mental challenges in domains like health and education. However, what I find more interesting is that open access AI also plays a pivotal role in AI safety. It promotes transparency and inclusivity, which fosters broader oversight, collaboration and accountability of AI systems, which ultimately reduces the risks associated with the unchecked deployment of advanced AI systems. So recognizing the promise of open access AI in AI safety and the pivotal role it plays in AI safety, and also drawing on recommendations in literature advocating for open paradigms of model release, especially in global South regions, including Africa, I saw it fit to discuss the contextual limitations of open access AI as an approach to AI safety in Africa specifically. I therefore discussed two broad constraints of open access AI as an approach to AI safety and followed this up with some proposals for how to mitigate these constraints and maximize the potential of open access AI in African AI safety. So the first limitation I discussed is the dependency dynamics between model sharers who are primarily located in leading AI countries in the global North and African AI safety researchers. In some, I explain how Frontier AI companies usually hold the unilateral decision on what model components they choose to release, and in some instances, who they

choose to give access to. Further, with the rise of AI nationalism among leading countries in AI, African countries stand to lose as regulatory and policy stances tend towards withholding proprietary information about advanced models in the name of national security and global competition. Africa also stands to face repercussions if they align with certain antagonistic AI partners. All these factors put together compromise Africa's autonomy, sovereignty, and ability to leverage open access AI for AI safety. The second limitation I talk about, which has been alluded to by the previous contributors, is the limited resources. Here I discuss the bottleneck of lack of compute resources such as GPUs and cloud computing access, as well as the general state of the continent when it comes to AI utilities, which include electricity access, internet access, and the overall lack of funding for AI safety on the continent. Thank you.

OKOLO: Great. Thanks, Grace. Thank you so much. And so appreciate everyone for the great opening remarks. And I think that really set the stage for the panel discussion that will follow. And particularly, I wanted to just start off, I know this was a bit covered in some of the opening remarks, but just really to provide a bit more context for our audience. I would love to learn how you all define AI safety personally in your research or just in your work in general. And then also for the regions that you covered and also for that many of you live in, what does this look like in basically AI safety approaches or even just like this concept or discourse around this topic? And so just to start off, I'll have Grace take the floor on this.

CHEGE: Okay. Thank you. So as you've said, we have discussed the contentiousness around what exactly AI safety is, but my rough definition of it would refer to actively researching and implementing mitigation measures to prevent societal harm from frontier AI systems, which include advanced AI systems like this. And I think this means this includes being concerned about current AI risks that are scalable into the future. For what it looks like in Africa, I think that it looks like many things, but crucially, I think it means one, addressing the context specific factors that make it difficult for African countries to address AI safety concerns. So one, this means what is the effect, what is the impact of the economic realities on the continent? Things like limited access to capital and investment, high levels of poverty, as well as the socio - political challenges that the continent faces. So things like recurring violent conflict, weak state institutions. I think AI safety also means finding ways for African countries to substantially engage in AI safety discussions in global AI governance in general. Yes.

OKOLO: Thanks so much, Grace. And so next up, we'll go to Craig to talk about the definitions of AI safety and also what this looks like for the Caribbean.

RAMLAL: Okay. So I think that I would have covered it a bit on what the definition would look like. I'll just reiterate that it would look like this alignment with societal wellbeing, ethical principles and protection of fundamental rights. I think that when we talk about what it could look like for a region, we want to develop a definition or have a holistic definition that covers something globally. And then each region should pull apart from what is contextualized for their case. I think that that is quite necessary to preserve autonomy in certain regions. I think that for the Caribbean especially, we need to be empowered and take control of our AI future. That's what safety would really look like to us. We need to strengthen our data sovereignty. We need to support low -compute AI models, preserve our cultural identity. We need to build local talent and develop AI for our regional challenges.

OKOLO: Great. Thanks so much. And next up, we'll go to Ben.

KEREOPA-YORKE: Thank you. Yeah. Look, I mean, as everyone else has described, we definitely face a sort of definitional and ontological crisis in terms of how we define a lot of these concepts. I think, speaking to what Grace said and also what Craig said, it's regional. The lens at which we view these things can't help but be restricted to the nature of the environments we live in. I think if we zoom out a little bit, the AI safety as a concept is really speaking to people, process and technology through the lens of the impacts on organizations, businesses, the people working in those or those that consume goods and services via AI and also the environment. I think here in Australia, where I am, we have advanced quite far. We have a guardrails that have been out for consultation. So there is a real move to move beyond perhaps the definitional crisis we face and start to think about what some of the guardrails are. And I think as all of us here come from the global majority, it sort of behooves us to be the ones that come up with our own definition. We don't need to inherit it from anyone else. Thank you.

OKOLO: That's very important. Thank you, Ben. And then next up, Maia?

LEVY DANIEL: Yeah, I already addressed this a bit in my opening remarks. But as I said, there's no clear definition of what AI safety means for Latin America. But in general, usually it refers to those requirements for the AI systems and obligations for the different responsible parties to ensure that the harms are minimized and human rights are protected. But as Craig mentioned, societal considerations, including cultural values that are specific from the region should definitely be part of this definition. And that's why I

think there's so much work to do. Requirements included in AI governance initiatives should reflect these specific regional considerations. And I will talk a bit more about this later. But there are regional initiatives that could be really instrumental for this.

OKOLO: Thank you so much, Maia. And then we'll round off with Jam.

KRAPRAYOON: Yeah, sure. Thanks. Yeah, it's a really interesting question. Has it sort of observed draft legislation bills or early legislation coming out from various Southeast Asian countries? I think the one thing I can conclude is that countries have not really converged on a sort of shared understanding. And even at a national level, not even to talk about sort of at a regional level, I don't think there's a clear sense of what safety even means. I think there are several themes that are sort of emerging around the region. I think a lot of them are attached mostly for historical reasons to certain other tech policy themes, like things like being concerned about child safety and the production of materials or access to certain websites. So I think it's kind of just the start of a conversation. And it's unclear at the moment what themes will activate policymaker and societal concerns across the region. With respect to kind of my personal definition, my work primarily focuses on the national security implications of AI. But what's been really interesting to observe if you look at the Paris AI Summit and some conversations I've had with policymakers in the region, is that I think a primary anxiety now has been actually not on kind of concerns of what AI could do or could use to do now or in the future, but on kind of the fear of missing out. There's anxiety of, what if I don't, I'm not part of the kind of the gold rush of a frontier AI. What if I don't build out my compute infrastructure now? What if I get left behind? And I think it's probably really important to seriously address that concern, given that it's rooted in kind of the historical sort of resource difference between the global majority and the global north.

OKOLO: Great. Thanks so much, Jam. And now we're actually going to go back to you because all of your pieces and also just in general, we see that there is a really strong need for these localized and also contextualized AI safety approaches. And so just given the motivations of this project in general, I'd love for all of our panelists, excuse me, to think about why is it critical to think beyond Western -centric approaches when addressing and also implementing approaches to AI safety. Jam, specifically, your piece with Sean touches on advances within AI safety frameworks across Southeast Asia. I'd love for you to share some of these developments. I know you've talked about this a lot already. And also, again, like why it's so important that these approaches are localized to the region.

KRAPRAYOON: Sure. Yeah, I think one obvious reason, probably kind of mundane reason why it's important to localize approaches is because when an AI safety framework is developed in some countries, it's developed with those kind of institutional realities in mind. So if you're porting over things like entire sections of like the EU AI Act or something, they're not necessarily fit for purpose, or the countries don't have the kind of implementation capacity to kind of actually, you know, to regulate AI in that way. For instance, I think the for developing sort of like the technical side of AI safety, but they shouldn't be a prerequisite for having a very technical facility for performing AI safety research shouldn't be a prerequisite for joining the global dialog on AI safety, for instance. So I think, for instance, many countries in ASEAN outside of Singapore are now looking into creating a sort of AI safety network, which will not be that that is going to act as kind of a regional secretariat. So they because, you know, individual countries will not be able to sort of staff, they don't have the resourcing and the staff and AI safety Institute nationally, but they can still participate and have a voice on the global stage. So I think developments like that are kind of showcase why you have to sort of think about what works in a local context. Another point might be around when you're thinking about actual, you know, dangers from the misuse of certain kinds of AI systems. It's important to actually think about, you know, operationally, what do those risks actually look like when they're materialized in the real world? I think we've kind of moved from sort of gesturing and vaguer sort of risks about what AI can or cannot do. And we're starting to sort of develop a more concrete sense of like, okay, if you do this kind of uplift study, a current AI system can actually help a person who's like has this level of expertise do X, Y, Z. I think we can kind of go further. And that's why I mentioned sort of looking at the actual tactics and operational details of how cyber criminals today act in Southeast Asia, because I think that will give us a much more grounded assessment of, you know, how AI tools might change that those dynamics moving forward.

OKOLO: Thank you so much, Jam. I'd be also be happy to open this specific question up to the rest of our panelists. And so I think Grace is unmuted, so happy to have her go.

CHEGE: I think that it's important to think beyond Western-centric approaches, because often they might make assumptions about global majority context. So as people have brought up things like access to critical AI resources is a big issue when it comes to global majority context. Do these approaches take into consideration that such contexts might not have the infrastructure, the computing power, the digital infrastructure, the skilled AI workforce? I'd say also another thing is that AI may pose unique AI risks in global majority context. So Western -centric safety frameworks may overlook the distinct risks that AI poses

in this context and fail to fully realize that they could result in disproportional harmful outcomes. So yeah, do these approaches take into consideration the societal, political and economic dynamics in these regions? Do they recognize that these are possibility of AI harms actually scaling? If you have a region that is prevalent with violence, what happens when you introduce AI systems, for example, autonomous weapons of war, the harm that could result is tenfold. So yeah, I think Western approaches might not take into consideration those nuances. And therefore, as people have said, it's, I guess, on us. We have the onus to come up with safety approaches that speak to our Indigenous situation.

OKOLO: Great, Craig?

RAMLAL: I really like the flow of the panel and what everyone is thinking about. And I think, and I completely agree with it. And I think that one of the key challenges in what we should be thinking of now is now that we understand this is happening and what AI is, the safety definitions are and how they are skewed. What we need, more or less, is inclusivity in what we define AI safety to be. So what would be the mechanisms for how we come together to talk about what is affecting each of our regions and how can we put that into a framework that is truly global and works in each of our senses? I think also too, now, in 2025, it's quite, we are quite fortunate to start developing newer AI safety rules and policy dialogs and guidelines because we know, we have a sense, a very good sense, the world has a very good sense of what AI is in terms of its technological capabilities and where it might go to. So it's quite fortunate for those countries to start thinking about what their AI safety guidelines and rules should, just wanted to make that.

OKOLO: Great. Thanks so much, Craig. I appreciate it. And so I think it's good to just move on to the next question. And I think we've already got a lot of great context from all of our panelists so far. And so now we'll talk about the limitations and considerations for AI safety measures, which is again, obviously something that's really important to think about. We're really interested in just understanding what are the limitations of current AI safety frameworks and also how could they unintentionally perpetuate these global inequities? Ben and Craig, I actually just want to focus on you both for this part of the question. Because you've specifically mentioned the challenges of advancing AI safety in Oceania and also the Caribbean respectively. Given the concerns and risk of AI development more broadly within small island developing states, how do you see these issues exacerbating these present-day inequities? And so I'll have you, Ben, start off with this.

KEREOPA-YORKE: I think it's important to establish definitionally that we are generally talking about generative AI. So the gen AI area here, right? So AI is not a public utility, certainly generative AI is not a public utility. You have a very small group of companies that have the access to the means of the production of our foundation models, right? So fundamentally in Oceania, we face a access issue much like what Grace has spoken to for Africa, in that we are holding to other areas of the world to either give us access to compute infrastructure, compute credits, et cetera, or to fine tune our own foundation models for you know, our purposes within the region. So, you know, again, if we say that when we speak about AI, we're generally talking about generative AI, then there is a philosophical and technical difference between AI safety that is spoken about in the pre-gen AI era versus, you know, today's almost post MLDL era. There's a lot of philosophical carryover, though the impacts of gen AI are actually different through the scale, because prior to generative AI, it was kind of something that happened to you or alongside you, AI, you know, it was the algorithm that defined your experience with the world around you or ranked you or, you know, used image recognition and classified you. Now it's something that we engage with through primarily westernized user interface and user experience, you know, life cycles. So there's a lot of need for us in Oceania to understand what is safety to us and who is the safety for, because we have a long history of understanding colonization. And this begins to look like we have been given a technology without fully understanding what it is that it has a value for us. In closing, I just want to acknowledge that there is a significant segment of, you know, of society, people that are viewed as experts within the AI field, who state that generative AI and large language models have huge ethical considerations that prevent them from being viewed as something that should be deployed into our communities. And not only that, but the things like hallucinations and the error rate and the compute infrastructure costs are so heinous that there should be a pause to think about whether this is a technology that should be used at all. I don't necessarily carry that view, but I think it's important that that view is heard and has a seat at the table. Thank you.

OKOLO: Thanks so much, Ben. And then I'll go to Craig.

RAMLAL: So when it comes to the limitations of the current AI safety frameworks, and I think it's a discussion of priorities, and I don't blame certain countries for having the views that they have. They have to protect their nation's interests. So I completely understand what this is. I emphasize what I'm trying to say. It may not be the best for other regions, but for their region, it might work out the best for them. So I think they do focus on issues of transparency and accountability. And then, like we said before, issues that other regions would face, it will not be there. A term I want to kind of, I think I want to push forward or popularize a

bit, would be this term of digital compute security. We heard it a bit. Ben had talked about it. Grace talked about it. So it's about a country's ability to have access to the computing power that they need for things like AI and scientific research. So it's similar to energy security, but you need it for compute instead of the power grid. And you want to make sure that your nation or your region has enough of it so that you're not totally dependent on other countries for it. And I think that it's quite difficult for each region to get that compute security, that digital compute security that they need. And it could come two forms, one in access or another one in actual getting the chips or having access to the silicon foundries to develop the chips that they need for their region. And I've been trying to push that for the Caribbean for some time, because as we talk about AI, we are not seeing the case, as Ben would have rightly pointed out, this huge study in low compute AI systems, low compute capable AI systems. We are not seeing that for generative AI systems. I mean, there have been some cases where we see it happening for the reasoning models, but generally it's not there. And that could exacerbate most of the digital divide. And since the world is moving towards this technological revolution, AI is nearly ubiquitous now, I think it's quite necessary for this discussion of digital security at a sovereign level to be held.

OKOLO: Thank you, Greg. And Ben, would you like to add more points to this?

KEREOPA-YORKE: I think that that idea should become more popular. In short, that's a great way of looking at it, isn't it? It's about it. And I think that also speaks to some of the work that Jan was talking about in terms of national security. We've all watched what happened post-Paris with the rebrand of the UK AI Safety Institute into AI security. I think there's a real move to looking at those core underlying issues that perhaps have not gotten enough attention while we have the AGI beast in the background. So I don't have anything new to add other than affirm that Oceania would be highly favorable of those concepts granting more traction.

OKOLO: Thank you so much. And Jam?

KRAPRAYOON: Yeah, I'd love to add something here, because this is something that I've been trying to wrestle with, this idea of sovereign AI, sovereign compute. On one hand, I think the arguments for it make sense at some level, as Craig has articulated. But at a practical level, I sometimes wonder, is it the ideal world where every country under the sun is trying to build a national AI champion or spending millions or billions of dollars on infrastructure to do so? Because for every major data center project, it will take X

amount of years to require certain access to certain kinds of energy sources. And it's very possible for a lot infrastructure to become kind of like zombie infrastructure. As chips improve rapidly and access to chips is uncertain. It's hard to know if they'll stay fit for purpose at the point after they're actually completed in construction. And I sometimes always wonder if it's sort of part of the traveling salesman role that kind of Nvidia is taking. They've at least done a tour of Southeast Asia, certainly, and has said, oh yeah, everyone should build data centers here. Everyone should kind of get into chip manufacturing. It's still something I'm kind of wrestling with. I wonder if the national economic model for countries trying to not fall behind is something like this where everyone needs to build their own. Because I guess the analogy there is it's kind of like manufacturing during the industrial revolution. Countries had to build their own factories and kind of the countries that did really well, Japan, Taiwan, People's Republic of China, eventually managed to sort of leap forward by quite a lot and kind of got to the sort of frontier. I wonder if that's the analogy or if the analogy is something more like with the mRNA vaccine, where you have technology that was developed primarily in the West, obviously with kind of international collaboration, but was distributed eventually all around the world for kind of global benefit. I think obviously, if you're looking at that specific case, there are a bunch of reasons why maybe the vaccines could have been distributed more equitably and faster to a lot of different regions of the world. But I am wondering now if we want to encourage that sort of model or view of what kind of frontier AI systems could be and how they could be distributed. Because if you're kind of taking that analogy further, you don't want countries that are sort of stuck with the Sinovac equivalent in terms of infrastructure or AI models. And they're sort of just for largely like nationalistic or patriotic reasons using sort of inferior things that might not be as good as what they could get. I think another point is that it's sort of unclear to me how realistic it is for a lot of these countries to actually get access to the chips in the first place. There's all sorts of kind of export controls. If you look at the AI diffusion rule in the US, is it actually feasible for countries to kind of like get on the tier list and sort of really get involved in sort of manufacturing and creating a chips foundry? I think that's something that I think is still TBD.

OKOLO: Great. Thanks, Jim. So I see Ben and Craig. So I'll go first with Ben to ask the comments.

KEREOPA-YORKE: Thank you. Look, just to speak to your point, I agree that if we were talking about any other type of technology, that there would be compelling arguments as to why a country shouldn't be going on. However, I think there's two major compelling arguments for why you need to look at this type of AI capability build. The first is, as we've seen very recently, the major and dominant players in the AI field perhaps have geopolitical concerns that make them less attractive, perhaps, or secure for the global majority

to rely on. And I'll leave it at that. Secondly, if you think about generative AI as a technology that mediates your interaction with reality. So all it does is you input things into it, and then it gives you this mediated output via text form or video or whatever it is. That's essentially cognitive infrastructure. So if you as a country are outsourcing your cognitive infrastructure or Your population's ability to experience mediated reality to another geopolitical entity. You are facing an existential crisis that is probably not one that you can effectively recover from. Certainly you will find it difficult to later on mediate the effects and ameliorate the effects on your population. So I think there's that recent geopolitical events and there's the nature of the technology itself that needs to be considered.

RAMLAL: Yeah, I completely agree with Ben. And to further add to that, I think that we are living in a world where, yes, it's significantly costly to purchase these tensor processing units as a sovereign mechanism to enable that development. But we are also talking about low -compute capable AI systems. We don't know what that compute will look like in the near future. But you have to understand this is our sovereign data that is leaving our countries. Most governments will not agree for another country to process that kind of data. That's a national security risk. So completely agree with Ben.

OKOLO: And Jam, would you like to add any more comments?

KRAPRAYOON: Yeah, sure. I'm just wondering if there are kind of alternative models that ought to be explored other than each country should sort of do the full build out, for instance, for national security, specific applications. I think there probably are some mix of kind of like practical security measures you can do, like data siloing, secure enclaves, mixed with sort of maybe like more exploratory privacy, the use of privacy preserving technology. So like, you know, you can kind of separate out national security, specific applications of AI models, even if you're not using like a model that was like, look, we trained or developed and maybe there's something around like the kind of gradations of access that other people have already talked about with regards to like, you know, the gradation from like fully open source to fully closed source. I think there's something there. I think there are incentives for frontier AI companies to want to sell access to models. And I think with kind of diplomatic, with a diplomatic push, also, it could maybe you could make the case that like compute access to compute, even if it's hosted outside of a country's borders is sort of like a right that even developing countries should have access to, especially for purely socially beneficial uses, like promoting enterprise or advancing kind of medical science. It just seems like it's more in line with the actual

resource constraints that countries are facing, even if maybe like an ideal from a security standpoint is that you do develop each country has like independent compute. Maybe that's not actually a viable path forward.

OKOLO: Yeah, really great insights from this section. And thank you, Jam, for your remarks. And so we'll actually go into the next question, because we're so deep within this conversation. And I think it'd be interesting to talk about some opportunities that we see for advancing AI safety throughout this respect of global majority regions. And so the next question is really focused on what is the potential of localized approaches for reshaping the safety landscape? And for my particularly, your piece specifically mentioned AI governance concerns and not in America. And how do you believe that efforts towards improving these frameworks, AI safety frameworks, governance frameworks, sorry, governance frameworks could improve AI safety in this respective region?

LEVY DANIEL: Yeah, so as I already mentioned a bit, AI governance initiatives have been quite flourished in Latin America over the past few years. But in general, these initiatives tend to be pretty vague, or do not provide actionable measures, or even a clear definition of AI, as we were discussing before. And many of these initiatives have been clearly influenced by particularly the EU AI Act. And we can see this in the risk -based approach frameworks these initiatives propose, and also in the safety requirements for high -risk AI systems in particular. So on top of these local initiatives, there have been recent developments in terms of regional initiatives. For instance, there have been two ministerial and high -level authority summits on the ethics of AI in Latin America and the Caribbean that were co-organized by UNESCO, which resulted in two declarations and a roadmap with specific actions that should be prioritized for this year. But these initiatives do not mention any safety considerations. So what I argue in my piece is that Latin America can definitely benefit from AI, but it must establish specific guardrails to prevent harms and guarantee human rights, which are usually mentioned in passing and without any further operational steps. So these regional initiatives could be instrumental in raising the awareness of the importance of incorporating robust AI safety measures in policies and regulations, taking into account the specific risks and harms the region is already experiencing. And they could also play a crucial role in facilitating a shared understanding of AI safety across countries, leading to common definitions and approaches that reflect the region's specific challenges, as we were already discussing, along with actionable steps to consider. But I think this thing about actionable steps is really important. And these regional initiatives should meaningfully engage with the AI safety dialogs that are being held at the international level, as these discussions are still defining essential concepts and standards that will directly impact countries in the global majority, including Latin America. So we'll see what

will happen with these forums and institutes recently created, and what role safety will continue having at the international level. But Latin America should also participate in these discussions. So given that the field is still nascent, I think there's a unique opportunity for the region to engage in these discussions, and also address technical AI safety considerations that focus on risks, risks and harms that are specific for the region, and seriously ensure the protection of human rights in Latin America.

OKOLO: Great, thank you so much, Maia. And so next we'll go to Grace.

CHEGE: Thank you, Dr. Chinasa. So specifically, in regards to open access AI, and how localized approaches can improve the safety landscape, I think open access AI presents an opportunity for African led AI safety research that takes into consideration local context, diversity, inclusivity. So because open access AI lowers the barrier of entry for researchers, research institutions, developers, I think this allows them to actively shape safety, ethics, and AI governance in general, in ways that can best serve the continent's unique needs, as well as factor in its unique vulnerabilities. I also think that open access AI presents an additional opportunity for skill development and capacity building when it comes to AI safety specifically on the continent. So it can be instrumental in building local AI safety expertise. So with the availability of AI models, tools, African researchers, students, professionals have the opportunity to learn, experiment, and collaborate with each other in ways that weren't previously available to them due to high costs and lack of resources. I also think it allows, therefore mentioned, to get an opportunity to keep up with state of the art AI safety techniques that were once again, once out of reach due to things like financial constraints. So connected to building skill and capacity, my piece also speaks of the opportunity for African AI safety researchers to identify unmet needs in things like model evaluation and develop niche expertise around these things. And I think that this way, African safety researchers could not only address gaps in model testing and other safety issues, but they also position themselves as possible key contributors to the global AI ecosystem. And I think this makes them more likely, it makes it more likely for external stakeholders to engage with them in exchange for expertise and valuable insights. Further, still connected to that, in the event that African countries choose to get into international agreements with global North partners, I think John had spoken of diplomatic pushes, right? So in the event that African countries choose to do that, developing niche expertise and specialized expertise could also be a lever under which they increase their negotiating and bargaining power because they have something to offer. Lastly, I also think that open access AI could possibly result in better informed regulatory frameworks for AI safety and governance on the continent. So by having open access to models, tools, research, I think that regulators, policymakers, and

other interested stakeholders can stay sort of updated on the latest development in AI technology, and AI safety specifically, and ensure that their regulatory efforts are relevant and informed by real world application of AI.

OKOLO: Thank you, Grace. And I just wanted to see before we move on to our last question for the panel, if Jam, did you want to add anything?

KEREOPA-YORKE: I'm sorry, I think I've just left mine.

OKOLO: Okay, no worries. All right. So yeah, so thank you everyone for that great conversation. And so as we round up the panel, I'm actually want to, you know, ask everyone, just in general, what interesting developments in AI safety do you currently see in your respective regions? And also, what future opportunities exist? And so we can start off with Craig, please.

RAMLAL: Okay, sure. So in our region, we have AI research hubs. That's why we're on the center called the Intelligent Systems Lab at the university. And we do develop the low compute capable AI systems because we do believe that that is quite necessary for our region. And we do it in such a way where we have the stakeholders involved with the adoption of the technology. So the stakeholders who would be involved in the adoption of the technology on board at the start to help develop the technology as it goes forward. We have a couple initiatives as it relates to policies, we have CARICOM's Caribbean Telecommunications Union started an AI task force, they are looking at harmonization policies, ethics safety is under that policy. The Caribbean Examinations Council is also implementing their regional AI policy in terms of education sector, which is a big one in terms of AI safety. And I'm on those two panels. And we have also started developing our well release our masters in AI, MPhil and AI, PhD in AI to serve the 22 countries in the Caribbean. And I think we what we need to have, especially for the future is that we need to have this baseline socio technical understanding of what AI is, what AI safety is, as a participatory approaches in the AI governments.

OKOLO: Great. Thank you so much, Craig. And so next we'll go to Jam.

KRAPRAYOON: Sure. I mentioned one, I think really interesting development in AI safety in the Southeast Asian region, which is the sort of formative potential creation of an AI safety network as opposed to just national individual AI safety institutes, as a means for these countries to sort of participate in global dialogs

like the Ai summits that have happened in the past few years. I think it presents a really incredibly important opportunity for the region to have a voice in global affairs, I think, especially as other panelists have alluded to, like Ben, the conversation around AI is becoming increasingly securitized. And I think my prediction is that, you know, in the next couple of years, it will be increasingly tied into the geopolitical conversation and the increasing tension between the US and China. I think currently, many Southeast Asian countries, just in terms of their positioning, they're sort of hedging between these two giants, essentially. And I think there's an opportunity to have more of a positive voice in these kinds of affairs. I think if we look into sort of the past with the non -aligned movement, I think there's an important role that these countries could potentially play using vehicles like the AI Safety Network, using participation in kind of global AI summits to sort of advocate for their interests with respect to sort of non -proliferation of dangerous AI capabilities, and also for the sort of the sharing of the potential benefits of AI.

OKOLO: Great. Thank you so much. And next, let's go to Grace, please.

CHEGE: I think that a really significant development that's happened lately is Kenya joining the International Network of AI Safety Institute. I think this holds potential to benefit the entire continent because it positions Kenya as part of a global network of countries that are working to establish common standards for AI safety. I hope that this collaboration will allow Kenya to learn from international best practices when it comes to AI safety. And once again, hopefully this trickles downstream to the rest of Africa. I also see prospects of Kenya being able to share insights and strategies to do with AI safety that are specific to the African context. Similar to that, I think there's a lot of opportunity for African countries to develop AI safety institutes, as someone has said, whether it's nationally, regionally or on continental level. Another positive development is that there are more initiatives on the continent that are working on AI safety. So the ILINA program, for instance, where I work, shameless plug, is building AI safety expertise in Africa. And it does this by organizing an annual junior research fellowship, as well as running a seminar that's targeted at young Africans specifically. I think this is commendable. It also produces high quality in -house research and engages with African AI policy makers in coming up with AI that is, AI regulations that are mindful of AI safety issues. I also think that another positive development is the continent coming together through the African Union to come up with the continental AI strategy. Even though it's soft law, I think that this is a good development and it's good that the African stance has been articulated in unity. Hopefully this trickles down to individual states.

OKOLO: Great, thanks so much. And then next, let's go to Maia, please.

LEVY DANIEL: Yeah, so in terms of opportunities for the region, I think that in the next few weeks, in Latin America, will be a way to start discussing seriously what an AI safety framework means, what it should include and how to implement it locally in the region. So as I mentioned, discussions around risks and harms are already happening and some countries already have some AI safety requirements, but they need to provide more details about the procedures and the actionable steps about how they should be implemented. And when providing those details, I think there are a few opportunities. There's an opportunity for the region to seriously focus on human rights and avoid replicating frameworks that are unsuitable for the region. For instance, the EU AI Act framework. There's also an opportunity to invite all the relevant actors to participate, including actors from technical sectors that would provide insights on the feasibility of certain procedures and requirements. It would also be interesting to focus on the way the public sector is using AI systems in the region and the specific safety requirements for those systems. In particular, since they are implementing systems developed in the global north that are using and are using them to automate processes from which people cannot opt out. So we need to see, again, how the conversation moves forward at the international level, but I think a regional discussion on AI safety would be even more interesting now in order to avoid replicating frameworks from abroad that may not be suitable for the region so Latin America can come up with a specific framework that applies to its specific needs and challenges. Great, thanks so

OKOLO: Great, thanks so much, Maia. And then Ben please.

KEREOPA-YORKE: Thank you. Yeah, look, there's a lot of interesting things happening in Oceania, I think specifically in Australia with the rollout of the voluntary AI safety standard. We're about to see operationalized AI safety in this country and that is the chosen vehicle for actually seeing what happens when the rubber meets the road. So yeah, keep your eyes peeled and we'll see what comes out of Oceania. Thank you.

OKOLO: Great, so thank you all to our panelists for the great conversation today. We're actually going to transition into Q&A. We do have a couple questions that have been asked beforehand and so I'm going to go through the most, I would say, important ones and then ask them to our panelists and then if there are other ones that we've received, I'll be sure to try to get them answered in the time that we have. All right, and so even though our conversation really focused on the global majority and the project itself is focused on these respective regions, we know that the US in particular and also the EU and the UK have had a significant

influence on these global AI safety and also AI governance conversations and we had a really interesting question that asked how does the significant shift in the US presidential administration and also the congressional power change, how does it impact the safety aspects of AI conversations? Either for those who have context or I would say experience in the US, how does this impact the global AI safety conversations? And I have anyone answer this. I think Ben.

KEREOPA-YORKE: Yeah, sure. So I've actually recently had experience with this question where it's being raised among large organizations. What is the impact of the Trump administration withdrawing some of the Biden administration's views on AI safety? And the answer categorically for our region is it doesn't mean anything for us because we are in AI safety because we need to ensure that we roll out AI that serves the people in our region and that we are not swayed by moves that take place outside of our region because the incentives for us to pursue AI safety are perhaps different than some of the political ones that we see in the United States. Thanks.

OKOLO: Great. Thanks, Ben. I just want to see if any of our other panelists wanted to address this. I know it's a bit of a tricky question.

KRAPRAYOON: I think it certainly does mean there's a vibe shift. I think not just obviously nationally in the US in terms of certain sort of issues maybe taking precedence or you're seeing even the UK as a very close partner has, I think, partially due to the influence of the US renamed the AI Safety Institute, the AI Security Institute, which maybe it's more of a nominal change or something, but I think does maybe reflect some substantive changes in terms of their focus. I think regionally the vibe shift has been the most apparent in the Paris AI Summit. There's kind of an increasing conversation about one, the security aspect, and then two, the need for innovation and embracing the innovation side of AI, which I think some proponents, the pro innovation folks argue was a viewpoint that was underrepresented. My guess is that you'll see more global support and obviously supporting the US for these AI infrastructure build-outs. Talking about the various sorts of beneficial applications of AI and then, of course, things that kind of promote successful competition against China. Those are my guesses for kind of the focuses in the US and then how they kind of reverberate globally.

OKOLO: Great, thank you so much, Jam. And so we'll get to the next question. It's kind of a two-part question which asks, what are the likely implications of the US-China AI relations for global majority

countries? And also, what are the promising policy choices to pursue in the era of AI nationalism? And this is open up to anyone.

KEREOPA-YORKE: Yeah, okay, I'll take that one. So in the Pacific, in Oceania, we are very, very far away from a lot of things in the world. And something that has recently taken place in our area is China has started to build a lot of relationships with some of the more, the smaller Pacific Island nations like Cook Islands, etc. So this happens both above board and otherwise. And also in Australia, China is our largest trading partner and responsible for a sizable amount of our GDP. So when we say what are the likely implications of a US - China AI race, heavy, heavy. And something that has to be handled both at the diplomatic level. And we also have to understand what the repercussions are from a security perspective. Recently we saw the release of Deep Seek, right? And that has really changed the game on a lot of different levels. Now the Deep Seek app itself is now completely banned in Australia government departments, in education. But obviously the Deep Seek family of models with the open weights is still in use within both research and in organizations. So I think what we might have is two-tier policymaking where a AI system that is hosted in a country. So hosted in China or only accessible via Chinese controlled infrastructure might be banned. But we might be instead with technology that is easily accessible and available via open model weights, et cetera. But there are genuine security concerns that arise from using AI models that you do not know how and what they're being trained on. So that's a long way to say we don't know, but surely it's not good.

OKOLO: Thank you, Ben. And Craig?

RAMLAL: Yeah. And also on because of this, of the release of Deep Seek and what the technological changes are, I mean, it's trained on a different, it's trained on this group policy framework, reinforcement learning framework. And that is really pushing ahead how capable these systems are. And what we are seeing in a case where China released this, well, Deep Seek released this model, we are seeing that a lot of Western institutions are now starting to train their systems based on this framework as well. And so we are seeing at least a very quick push towards just getting faster and more capable systems coming out without really looking at the safety of the implications of what these things are. So it's really a race. It's China and U.S. are racing each other, but the world is also trying to keep behind and still getting some information of what is available to them from the open technologies that is being released. And we really, we're not sure on the, well, at least I'm not sure on the policy side, like how to even make sense of all of this and what is going

to be the next thing that is going to come out. We are seeing very small models being released, now one billion parameter models being released that are very capable. So it's very, very interesting.

OKOLO: Great. Thank you so much, Craig. Jam, did you want anything I saw you?

KRAPRAYOON: Yeah, sure. I just wanted to make a fairly specific point. It sounds like from what Ben said, Southeast Asian countries are in a very similar position with regard to China. Sometimes they're one of, if not the largest trading partner for a lot of countries and often the largest investor in terms of things like infrastructure investments, but simultaneously also Southeast Asian countries are also a frequent target of espionage tax from China as well. So I think the interesting kind of policy choice that might be open now, at least for some countries in Southeast Asia, is to benefit from this sort of like French shoring development or basically like the offshoring of things like semiconductor manufacturing, computer vision in mainland China and also in Taiwan just because they're unfortunately at a kind of state of heightened risk. I think opportunities for countries like Malaysia, Thailand, Singapore to sort of benefit from onshoring some of this semiconductor supply chain. And I think there was also a secondary opportunity on top of that to sort of leverage that new position in the semiconductor supply chain to sort of negotiate for things internationally that are kind of beneficial for the region.

OKOLO: Great, thank you. And Maia?

LEVY DANIEL: Yeah, just to make a quick point, but I think this question is related to the previous one related to the like the impact of this administration shift. And I think this could have an impact on Latin America, in particular on specific countries, because currently there are like different ideas in different countries in the region, like for instance, Brazil has been willing to regulate AI and understanding that there are harms and risks involved. But we also have, for instance, Argentina that was already willing to deregulate AI to make like to create and that well to make Argentina like an AI hub. And it's been like this administration is very aligned with the new administration in the US. And there's this idea that if you're not aligned with the US, then you're aligned with China. So I think some countries will definitely follow the new US ideas regarding AI deregulation and that will definitely have an impact on AI safety requirements.

OKOLO: Thank you so much, Maia. And Grace, I just wanted to check in and see if you wanted to add anything at all? No pressure? Okay, no worries. All right, so we're going to the next question. It's actually

kind of technical. I know we have some engineers and computer scientists in the room. So it's really focused on the capacity of AI to be used for translation and other tasks. And what needs to be prioritized in training these models to be more culturally sensitive? And also how should we be training individuals to engineer their AI prompts in ways that don't perpetuate systemic biases and or hallucinations?

RAMLAL: Yeah, I guess I can take this one because I might be the biggest nerd in the room. So what needs to be prioritized in training models to be more, okay, so the main thing is that we need to have data that represents the cultures that you are trying to train these systems on, as well as have the guardrails around these AI systems so that their responses do not cross that boundary, that bias. And the way that they normally do it is that they align these systems right after they fine -tune it. So fine-tuning is a method that you can use to align as well, but you want to essentially get your foundational model and have this sensitive, not sensitive data, but you have this guardrail data that you want to make sure that your model aligns with. The idea on prompt engineering is something I kind of, I don't know where I stand with it because prompt engineering how do you elicit a response from the model that you want, but it's close enough, it's written close enough to the training data of the system, right? So I don't really, every AI system is going to be trained differently, right? So I don't really follow prompt engineering as much as one should, but I do look at the ways and how do you fine -tune your models and how do you develop. We are looking at something of a layer of protection around your models before it releases some of its responses, right? And you can look at things as an engineer, some things like meantime between failure and those other metrics to see, make sure that it does not perpetuate the systemic biases and those issues.

OKOLO: Great, thank you so much, Craig. And I'll go to you, Ben.

KEREOPA-YORKE: Yeah, thank you. I just want to make the point that no AI system should be deployed or trained on language when the people who speak this language or who are indigenous to it, etc. do not consent. There is a certain attitude related to AI where it is apparent that nobody is allowed to question whether an AI system should be deployed at all. And I believe that there needs to be an avenue for refusal by communities where they can unequivocally state that they do not consent for their data or their language to be ingested into AI systems. And I think that's really important in the global majority that we have a relatively uniform acceptance that there are going to be specific subgroups in our regions that do not want to sign up for the AI journey. And though we are proponents of AI systems ourselves, we should always endeavor to create pathways for consent and refusal to be honored.

OKOLO: Thank you so much, Ben. I think that's really important. And I think the idea of this autonomy and also refusal is something I see growing particularly in the US context but also worldwide and particularly in Oceania with indigenous populations. So thank you for bringing that up. So for our last question, and again, this is a free floor, what are your predictions for an international AI governance framework? And this has also been a tricky question with a series of AI summits and just we've seen this regional questions arise. So Craig, let's start with you.

RAMLAL: So from the 79th UN General Assembly, they are looking to develop the International Scientific Panel on AI and also the International Governance Dialog on AI for policymakers and so on. So it's likely that the governance framework would start from those discussions if there are to be one. I don't know now with the current climate and what's happening if an agreed international governance AI framework would still be on the table. But if the governance dialog policy does come out, I would be more aligned to say that it most likely would be formed from some of those discussions. And we would see at the UNGA 80 if you know what's supposed to happen.

OKOLO: All right. Thanks so much, Craig. I appreciate it. So I want to see if any of our other panelists had anything they want us to contribute to this. Again, it's a very tricky question. All right. So just to stay on time, we will end the webinar here, but I really want to extend my thanks to our panelists for their great contributions to this project and also participating for this webinar in this webinar today. We also thank the audience for attending this webinar and appreciate your questions. Please be sure to check out our published pieces for the AI safety in the global majority project on Brookings.edu and stay tuned for our pieces on Africa, Oceania, and the Caribbean. Thank you again and have a great day.