

THE BROOKINGS INSTITUTION

FALK AUDITORIUM

AI SAFETY BEYOND THE UK SUMMIT:  
A CONVERSATION WITH SECRETARY MICHELLE DONELAN

Washington, D.C.

Monday, November 13, 2023

UNCORRECTED TRANSCRIPT

WELCOME

CAMILLE BUSETTE

Interim Vice President and Director, Governance Studies  
The Brookings Institution

FIRESIDE CHAT

MICHELLE DONELAN

U.K. Secretary of State for Science, Innovation and Technology

MODERATOR: CHINASA T. OKOLO

Fellow, Center for Technology Innovation  
The Brookings Institution

PANEL

MIRANDA BOGEN

Director, AI Governance Lab  
Center for Democracy and Technology

CHINASA T. OKOLO

Fellow, Center for Technology Innovation  
The Brookings Institution

MODERATOR: VALERIE WIRTSCHAFTER

Fellow, Artificial Intelligence and Emerging Technology Initiative  
The Brookings Institution

\* \* \* \* \*

**BUSETTE:** Good afternoon, everyone. I'm Camille Busette, the interim vice president for Governance Studies here at the Brookings Institution. And I'm delighted that you have joined us this afternoon for an important conversation about the future of artificial intelligence and safety. As many of you who have been following the public policy discourse on generative AI like ChatGPT know, the policy conversation around us is fluid and evolving, and there are a range of very interested stakeholders for whom the outcome of any public policy framework or regime has significant consequences. As a result, the formation of public policy in this space has involved multiple global dialogues, as states, companies, innovators, and advocates try to find the right balance between innovation and responsible evolution of the technologies. The most recent of these important conversations was the U.K. AI Safety Summit in early November, which brought together international governments, leading AI companies, civil society groups, and experts in research. It considered the risks of AI, especially at the frontier of development, and discussed how those risks could be mitigated through internationally coordinated action. The architect of that summit and the secretary heading up the U.K.'s efforts in the AI safety space is our esteemed guest this afternoon, the U.K. Secretary of State for Science, Innovation and Technology Michelle Donelan. Ms. Donelan was first appointed secretary of state for science, innovation and technology in February of 2023, and I think it's fair to say that she's moved at warp speed to position the U.K. government as an important interlocutor on AI safety policy. Previous to her current role, she held a number of ministerial posts in education, having been appointed as secretary of state for education in July of 2022. So, fresh off the very successful U.K. summit, please join me in welcoming Secretary Donelan.

**DONELAN:** Good afternoon. It is a pleasure to be here at the world-renowned Brookings Institute this afternoon and to discuss what I believe is the most important transformational technology in human history. Now, just to manage your expectations, I do have to admit that I come bearing a lot of British bugs. So if I cough, splutter, or dry up my throat, please forgive me and bear with me. So, thank you.

Artificial intelligence is on the cusp of changing the world as we know it. And it is nations like ours who stand together with our shared values, creating innovation and driving artificial intelligence to empower us and to change our lives for the better. In the U.K., we have 1% of the world's population, but we have built the third largest AI sector. We have rocketed ourselves to a 688% increase in AI companies basing themselves there in less than a decade. And U.K. AI scale-ups are raising almost double of that of France and Germany and the rest of Europe combined. The U.S., the U.K., and China now make up the top three AI nations in the world and that has extremely important implications for all of us. Because while some speak of the "special relationship" as a tried cliché, what our two nations are doing together in AI speaks louder than words ever can. Just take, for example, DeepMind, one of the leading A.I. companies in the globe, on the globe, a British company that has recently merged, of course, with Google. The unstoppable force of American business enterprise combined with the powerhouse of British innovation and invention, Google DeepMind are now building models that will potentially save thousands of lives from cruel diseases that can genuinely be improved for our lives, for all of us. These partnerships, whether at a business level or state level, are vital if we are to ensure that the models of the future encompass the values that both our nations hold dear: democracy, freedom, fairness, accountability, and transparency. Our two nations, I believe, have a responsibility to compete with China, to stand up for these very values. And that is why it is so important that we established the first-ever global A.I. Safety Summit at Bletchley Park earlier this month, bringing together more than 150 representatives from government, from tech giants, from universities, from civil society, academics, scientists to come together to form a collective commitment to the safe development and deployment of powerful AI. And through the Bletchley Declaration, signed by 28 member states and the EU, we laid the groundwork for a new, unprecedented international partnership. Against all the odds and despite many saying that it was in fact impossible, that partnership included China, a global coalition working together in support of A.I., which is people centered, trustworthy, responsible, and above all, safe. To ensure this mission success, we also agreed that Turing Prize winner Yoshua Bengio, credited, of course, as being one of the very godfathers of AI,

would lead on a "State of Science" report. And he will be supported by an international expert advisory panel, comprising, of course, of experts from the United States and other key nations.

But we didn't stop there because if we are really going to ensure there is that great catalyst, then we have to make sure that the commitments that we take now will stand up in the decades ahead. A.I. may be moving fast, but humans must, I believe, move faster. The U.K.'s new A.I. Safety Institute will play a central role in this vital endeavor, and it will be complemented by the work of the U.S.'s own Safety Institute, which the vice president announced in London during the summit, and it will work with innovators at international level to rigorously test new AI models both before deployment and after. The world is indeed embarking on a new journey towards greater safety and I think that Bletchley Park represents the first few steps on that journey. And that is why summits will also be held in Korea in just six months' time and then in France in one year's time, ensuring that the extraordinary pace of international action is maintained well into the future. Because I have been clear that back in the U.K., an international approach to this issue is not just practical, it is absolutely essential.

Right now, we stand at the crossroads in human history and if we turn the wrong way, it would be a monumental missed opportunity for mankind. The prizes within our grasp are enormous. AI could turbocharge our public services. It could produce tailor-made education services for every child's needs and talents. It could free people everywhere from tedious work and amplify our creative abilities. It could help our scientists unlock bold new discoveries, opening the door to a world without diseases like cancer and Alzheimer's. But we can only really seize those benefits if we double-down on our shared values that the U.S. and the U.K. align on. Our two nations as world leaders in AI have a unique role to play in giving people in every corner of the globe confidence that this technology will work for humanity and not against it. That it will, echoing the words of President Reagan, realize our best hopes are not our worst fears. That it will become a powerful force for good so that communities here in the U.S. and back in the U.K. and around the world can live healthier, happier and easier lives. Thank you.

So I sit here? Yeah.

**OKOLO:** All right. Hi. So I'm Chinasa T. Okolo. I'm a fellow here at Brookings in the Center for Technology Innovation. And I'll be moderating a fireside chat with Secretary Donelan. And so thank you again for your presence here at Brookings today, and we really appreciate it. And as a trained computer scientist attending the AI Safety Summit was a great opportunity for me to engage with different stakeholders from industry, civil society, and academia, and it also widened in my perspectives on AI safety. And so to get started, I'd love to hear more about your perspectives on the summit and broader views on AI governance. And so to kick us off, what do you think were some of the successes or failures of the UK AI Safety Summit?

**DONELAN:** Yeah, so I think what we did, summit was quite remarkable really, because we did bring together so many experts from civil society and also countries from all over the globe, including China and also the companies themselves. And the atmosphere at the summit was really one of teamwork and one where we actually have to tackle these risks and these problems together despite the differences. And there were some lively discussions in some of the breakout rooms, but there was a sense of actually we have to do this globally and not just act in silos. I think one big success, of course, was getting China to not only attend but agree to the Bletchley Declaration, which was something that a lot of commentators said we wouldn't manage to achieve, but I think is vitally important if we're actually going to be serious about this agenda. And then it is the long-term path that we have set ourselves on now that is the true success of what we achieved at the summit. So when I say that, of course I mean the panel that will be setting up, the "State of Science" report that will really ensure that we're on top of the research and we're delving further into the risks, because the reality is nobody has a full handle on the risks. This is an emerging technology that is moving quicker than any technology we have ever seen before. And we also got that landmark agreement to enable us to test the models pre-deployment, which is especially important given the fact that we're on a short timeline as a new set of models will be coming out within the next six months. So it was a tremendous success. I don't believe that was a failure in

any sense. So I won't engage with that part of the question. But it is very much the beginning of this journey and there is a long way to go.

**OKOLO:** Okay, yeah, thank you so much. And so there's also been many debates about the concerns of existential risk overshadowing the present-day risk of A.I., like algorithmic discrimination or election-related disinformation. What's your view on some of these perspectives? And also, do you feel that that they are compatible or at odds when thinking about AI governance?

**DONELAN:** So when we decided to hold the summit, I wanted to make sure that it was an adding to the conversation and not cutting across the conversation. And that's why we focused exclusively on the frontier - so we obviously mean those highly capable models that we see now and that will be developed in the future. And if you look at a lot of the other international work that's gone on already, so I'm thinking of the Hiroshima Process, GPAI, the OECD, etc., the UN's work, none of that focused exclusively on the frontier. That's the missing piece of the puzzle, if you like, that we wanted to achieve via the process that we established at the summit. But we did look at three categories of risk. One of them was loss of control. So that is much more your your future potential risk, your existential threats. One of them was around misuse -- now misuse can be occurring now as well, so it's not just a future risk -- and also societal harms, which do include things like disinformation, misinformation, bias, potential impact on democracy and elections. That's something that I led on in terms of the conversation personally at the summit, because I think it is very important and it's very topical, not just for the U.K., but of course for America as well. And we together are marching towards our elections with this very live and real threat hanging over us. And what I want to come out of those conversations at the summit now, and I will be having more conversations during my stay this week with my counterparts will be very much focused on on that particular topic, because I think we need to work together. We need to combine our expertise and knowledge to be able to tackle that problem head on. Yeah.

**OKOLO:** Awesome, yes. So thank you again, you know, for your responses. And so in the aftermath of the summit, there's been debates about the potential for regulation to hinder open source software in the future A.I. ecosystem. What's your perspective on these debates and what role may this specific conversation play in ongoing regulatory discussions?

**DONELAN:** So on open source, this is something that was discussed a bit at the summit. There are some some big advantages to open source. It also helps to equalize AI, if you like, but there are also some some disadvantages and some concerns. Now, we weren't going to solve that problem or that topic within two days. And it's something that we do need to delve into more. It's why it's also important that we're doing more research into the risk, to understand it better, to be able to apply the right mitigations and the right guardrails. So things like that State of Science report, and of course the institute that we announced that we're establishing, which is the next evolution of our task force, and your institute here in America, which will be working in lockstep with ours, will really be able to delve into that question, but to your broader point about the interaction with regulation and innovation, what we're trying to do in the U.K. is ensure that our regulation is actually encouraging innovation. And we set out a white paper earlier this year that outlines the principles that we want our existing regulators to work to, things like the values that I talked about before. So things like safety and fairness and transparency and accountability. And that's to ensure that we have consistency and cohesion across the board, which will help businesses to to know the roadmap and and to get the same messages from different regulators. But of course, we want to do that in a way that does foster business, does encourage the jobs of tomorrow, and does encourage this technology, which has the ability to change all our lives for the better to actually be developed.

**OKOLO:** Thank you so much. And so I like to go back to some of your earlier points about China's participation in the summit. There was lots of criticism that was received to the U.K. for inviting China to the summit. And how do you think their inclusion contributed to the broader discussions and objectives at the summit?

**DONELAN:** Yeah, so on this I say to people, you know, okay, we can bury our heads in the sand and pretend that China doesn't exist or pretend that China isn't a leading - global leader along with us and America when it comes to AI. But that would be somewhat naive. And the reality is that if we really want to get a grip on these risks and we really want to utilize these amazing opportunities for the benefit of society, then we have to do that globally. AI doesn't recognize geographical boundaries, and so if we don't at least try to get China to be part of that conversation, then I think that is the wrong tact, definitely. And I was quite taken aback by not just how willing China was to attend and to sign up to the agreement, but also the way that they engaged in the summit as well, and really showed a willingness to work with other nations, including ours and America. And there was one moment where I think I was standing at the podium and in the seats, a bit like this, was the EU, was America, and was China. And I think that spoke volumes about just how we had convened people from around the globe to really have this impact on this agenda. And of course, it wasn't just Chinese ministerial representation, it was also civil society as well, because I think that was vitally important that we were getting all these different voices around the table so that we could tackle this issue.

**OKOLO:** Okay, thank you. And so thinking about global participation on a wider scale, when it comes to AI regulation or just governance, how do you think, or how do you position the U.K. as a body in the conversation to include some of the lesser-heard voices within this space?

**DONELAN:** So I think in terms of including the lesser-heard voices, that's why it's particularly important that the summit wasn't a closed environment. We were really transparent to begin with about what the objectives were. We were transparent about the breakout groups afterwards and the key things that were discussed. But then we also made sure that the attendance involved scientists and academics and those from civil society, from across the globe, that we included people from a range of countries, every continent was represented, and that we had not just the leading frontier AI companies at the summit, we also had smaller AI and technology-based companies. In fact, I think one of them, a few of them only had about five employees. So we're trying to get that breadth of views, but we're acutely aware with our summit that it was only two days and so we did a what we called road to the summit where we engaged with multiple stakeholders, again, not just in the U.K., globally, loads and loads of roundtables, different forms of discussion. We got different partners involved in that as well so that we could hear the voice of so many people and they could feed in to this agenda. But the summit was only the beginning of the journey and we need to work closely with countries like America to pursue this in the long term and make sure that we are getting all of those voices and that expertise into this conversation, because as I said, I really do think this is going to be a game-changer. When we when we look back on human history, we will note this point in time. And that's why it's important that we we leave no stone unturned in trying to make sure that we get this right.

**OKOLO:** Thank you so much. And so my Ph.D. work focused more broadly on how AI affects different populations in the Global South, especially when it comes to Africa and India. And also my work here at Brookings also focuses on governance in the Global South. And so I was recently, I recently saw the announcement for the AI for Development Initiative, and I really was really impressed by that. I'd love to hear more about it. And just like, what are the goals for this initiative and others that may be in the pipeline?

**DONELAN:** So I think it's really important that AI is not something that benefits the West exclusively, it has to benefit the entire globe and it has to do that in a number of ways. Specifically, it has to enable us to tackle some of the big global challenges that we're all facing, like climate change. But it also has to be a power for good, a force for good in developing nations to help with things like food inequality, etc., or to help with predicting droughts earlier and things like that. And then that is therefore incumbent upon us to make sure that we are leaning into this technology in the right way. So if we look at the summit, there were, there were countries from Africa that were represented. Rwanda, for instance, was there, a number of other countries participated. And I've spoken to to them and offline as well from the summit and making sure their voice is really heard on this agenda. And they have a lot to say. And in terms of the announcement that we made, that

was about injecting money not just from the British government, also the Canadian government and the Bill and Melinda Gates Foundation, so that money could be directed specifically to a number of African universities for research to catalyze the use of A.I. in that nation. And we hope that that will be the beginning of that process.

**OKOLO:** Great, thank you so much. And so to get into some broader questions on air governance, oftentimes the idea of innovation can be viewed as at odds with regulation. What is your view on this and how may governance solutions like common standards or evaluation frameworks or regulatory sandboxes be leveraged to further foster innovation?

**DONELAN:** Yeah, so my mantra in our department back in the U.K. is very much about regulate to innovate. And I don't think we need to think that regulation necessarily stifles innovation. In fact, when I speak to businesses, they say to me that they want to have clarity over the regulatory environment. They want to know the roadmap that they're going to be working in so that they can have that confidence to make long-term investments. And we have a multitude of different regulators in the U.K. that often overlap a little bit, especially for businesses. And what we wanted to ensure was that they weren't getting conflicting messages, that there was consistency. And that's why our first approach here with the white paper that we published was around setting out the principles that all regulators needed to act. I'm a massive fan of sandboxes, and we've already done some work in this, and we will be announcing some more work shortly in this area because I think that is an important tool to spur on innovation.

**OKOLO:** Thank you. And so there are still significant gaps in our ability to test, evaluate, and measure safety and fairness in advance AI models. Both the U.S. and U.K. governments, have announced the creation of a policy institutes whose mandates include the development of assessment, of evaluation capabilities. Do you think these initiatives are sufficient for the scale of this problem?

**DONELAN:** Yes. And the good news on this one is that we already established what we call the Frontier AI Task Force in the U.K. and we got on board skills from across the globe. So people like Yoshua Bengio and some of some of our own expertise in the U.K., but also from different fields, and we wanted those people to help us to really examine this question and be able to lead on this work. And so that's already established; it has the funding, it's been going, it's already been doing some of this evaluatory work. At the summit, in fact, it presented demos on some of the risks. So around misuse, loss of control, etc.. And the next evolution will be morphing into an institute, which can happen seamlessly now because it's already in existence and we want it to work in lockstep with America's institute when that is off the ground. But the good news is that it can do it immediately, in essence, because it's already started doing it and it has the expertise and the talent and the funding to be able to really evaluate these models. And the companies have given the agreement for that to happen pre-deployment, which is is great news. But we need more from the companies, too. So one of the main conversations that we had at the summit as well was what more can the companies do themselves? Before the summit, in fact, we published a document called the Emerging Processes Document. And what that basically is, is it's a catalogue of all of the different initiatives and policies that these frontier AI companies could use as a safety guardrails themselves. And it is designed to encourage a race to the top. We also, for the first time, got the companies to publish their own safety policies so that you can compare and contrast, if you like, between what they could do and what they do do. Again, encouraging and incentivizing a move forward. And I think that that also complements what came out of the White House with the executive order and just shows that we are working together and we each have a role to play in this agenda, to work globally. And it speaks to why this is so important that we do this together.

**OKOLO:** Great, thank you so much. And so that was my last question for now. I'd like to open up to the wider audience - excuse me - and our online audience to ask any questions they have. Thank you.

**AUDIENCE QUESTION:** Hi. Chris Macrae, Norman Macrae Foundation. The Economist sent my father back in 1951 to spend a year with von Neumann and Einstein. And the idea which

emerged from that was that media should always cover very good AI, because really technology is always a race between very good and very bad. Could you give us some examples of very good AI that maybe came out of the summit? I tried to watch from over here and unfortunately the media conference seemed to ask questions about very bad, and I didn't hear much on very good.

**DONELAN:** Yes, that is a general theme that the media likes to, it does sell headlines and sell papers. So basically, there's a multitude of different areas where AI is already, as you will know, transforming our lives. And some of the key things that we drew out at the summit, we made an announcement domestically in the UK for £100 million to go into what we called an AI accelerator mission. So we've already got these missions in life sciences that we'd established around things like dementia, for instance. And what we want to utilize this hundred million for now is to rocket those missions so that we can actually tackle those areas quicker by the use of AI and so that money can go towards research to achieve that. We also made an announcement around the summit for over £2 million to enable the use of AI for our teachers to be able to produce more individualized lesson plans, etc. And we have a goal in the U.K. to get to a position where every teacher would be utilizing AI in that way so that it would reduce the amount of time that they're spending on the the admin and the bureaucracy and the paperwork and give them more time in the classroom. And I think that to me is one of the key advantages of AI, because when I look back at my education and I'm sure when you look back on yours, what you remember that you remember those teachers that had the time to invest in you, to inspire you to work with you. And if we can do more of that, that's incredible. And it's the same with healthcare, isn't it? If doctors and nurses can spend less time filling out forms and etc., and can spend more time with patients, then we will get better overall results. So for me, AI is very much about being an excellent copilot, enabling people to get on more with the job that they actually trained to do, as well as limitless opportunities in the future.

**OKOLO:** Thank you. Over here.

**AUDIENCE QUESTION:** Hello, my name is Kevin Allison. I'm the president of Minerva, which is a company that looks at the geopolitics and policy of AI. I was curious what, if any, ideas emerging from the summit or from the U.S. executive order do you feel might constructively feed into the next steps in the U.K.'s regulatory process on AI from here? So for example, the next steps to follow the white paper. Is there anything that you've learned over the past couple of weeks that that might change the U.K.'s overall approach, or are there specific policies like reporting requirements or know your customer requirements from the U.S. executive order that you would like to see implemented in the U.K.?

**DONELAN:** So I think it very much assists us because I don't see us as the U.S.' work and the U.K.'s work. I see us working together on this agenda. And we have to be realistic. There are some things that you will be able to do better than us because the majority of these companies are based on U.S. soil. And so it makes sense that you lead in in certain ways and that in that we can then lead on other ways, just like we we set up the task force very quickly, just like we convened the summit at Bletchley Park. So I think when we're now thinking and contemplating about those next steps for the white paper, we do that with the comfort of knowing that we're not acting alone, that we're acting globally, that you've gone on a pathway that will help us to achieve the safe deployment of AI. And we, of course, have got that landmark agreement at the summit, too. And all of these things can't be viewed in isolation, but they must be viewed as the overall picture. And that was very much what we wanted to achieve with the summit. We wanted to ensure that everybody was doing the bit that they could do and that we were working collectively.

**OKOLO:** Okay, Thank you. Second row on my right, please.

**AUDIENCE QUESTION:** Thank you. Thank you for your presentation. My name is Ricardo and I am a diplomat and a consultant in international affairs and risk analysis. I have two very brief questions. The first one is, your government and the United States, along with forty others, put out today a political declaration on military use of artificial intelligence. Can you tell us a little bit of how does that look like in layman's terms? Like what is it exactly that you're trying to achieve with with

the declaration? And the second one on the enforcement and regulatory process, looking ahead, how do you, what is the U.K. trying to do in terms of enforcement globally? Are you looking for the UN to to work in the near future in something like that, or are you going it alone? I just don't have it clear in my head how how is it? It seems that everything is moving way faster than than we might want to or many people might want to. So how would you enforce all your efforts?

**DONELAN:** So in terms of your your second point, so we wanted the work that we were doing at the summit and the forward pathway to complement the existing work that's already happening by the likes of the UN and GPAI, etc.. And but we were never advocating for a global regulator. What we're advocating for is a global approach in terms of everybody actually agreeing that we need to get a better grip on the risks, we need to do more research collectively, that we need to start a conversation on this in terms of what are the best guardrails, and that we need to work very closely with the companies at the cutting edge of this technology so that we can actually ensure that we are keeping pace as this technology develops. In terms of the military use of AI, there does have to be very clear rules of the road. So in layman's terms, that is what that's about, is about making sure that we are abiding to a code of conduct in terms of the safe use of AI in our military defenses.

**AUDIENCE QUESTION:** [off-mic].

**DONELAN:** So, so so as I said, we're not we're not advocating for a global regulator on AI. This is very much the beginning of a journey and very much about starting those networks and these conversations. Now we have hit the ground running because not only did we convene the world at Bletchley Park, we also got a landmark agreement on testing pre-deployment. We also got the 28 nations plus the EU, and that included China, to sign up to the Bletchely Declaration. We set a forward path with the State of Science report and panel and of course, our institute working hand-in-hand with the American institute. And then the next summit will be held in six months and then the summit after that in a year. So you talk about this moving so fast. That's why we've got to keep pace with these these summits and these workstreams. It's why I'm here in America this week as well, having these conversations with my counterparts. It's why we're just a few weeks past the summit and I'm here already because this agenda is vitally important and we want to tackle it on a global stage.

**OKOLO:** Great. Thank you. So we have a couple of questions in the back. The gentleman in the gray shirt and then the gentleman in the black turtleneck.

**AUDIENCE QUESTION:** Hi, thank you very much. My name is Adam Dawkins. My background has been working in health care, both in terms delivering care virtually and also using AI.. So the benefits of filling in forms are also the risks about making decisions. So my question really, have you, has work being done about thinking about accountability? And why I say accountability, just kind of make sense of it, is having worked in regulation and law to some extent with risk management in health care, if you're going to have regulation, you need to have some idea behind beforehand of accountability. And given the lack of transparency of AI having these things in place, is, to my mind, really important because otherwise it's, you don't have an audit trail, and how you actually formulate going ahead is that kind of process. So really just thoughts on accountability, if you'd be kind enough.

**DONELAN:** Yeah, absolutely. And I think that's why every decision that we take has to be the right one and we have to do the research behind it. So it speaks to why the money that we've invested in relation to our AI accelerator mission was around investing into the research on this so that we can actually do this in a safe way, one that does ensure that we have accountability and that we're not using AI in a in a reckless way, in a way that will actually lead to, to to worse health systems, because we want this to enable and empower our medics rather than take away from them or risk patients. So I think at every stage we need to be looking at this very question and making sure that we're using AI in the correct way. And that's why our approach is very much investing in the research rather than just jumping to the, to the end solutions before they're finalized and finessed.



**OKOLO:** Okay, thank you. So the gentleman back there, please.

**AUDIENCE QUESTION:** Yes. My name is Roger Cochetti. I am an editorial contributor on technology policy for The Hill newspaper. And you've touched a couple of times on the theme of regulation, which I think is very important. It's where the rubber hits the road, so to speak. And you correctly pointed out that many companies in this field are seeking some type of regulation so they know what the playing field looks like, but you also pointed out that you are not seeking global regulation of AI. However, one of the key reasons a business would support or seek regulation is so that they can be confident their competitor is playing by the same rules as they're bound to play by. So how do you, I'm hoping you can expand a little bit on your perspective towards regulation, because the worst situation for any company would be to be strapped with the set of regulations for themselves that their competitors are not bound to, and yet you're not seeking global regulation. So how does this all fit together, if you can? Thank you.

**DONELAN:** Yeah, so that's an interesting question. But then it takes me back to where we were, say, a month ago. People were saying to me that we couldn't convene all those countries around the table, that we couldn't achieve an agreement at all, and that we wouldn't be able to get anything from the tech companies. And we did all of those things. But we do need to be realistic as well. And that was monumental in terms of what we actually achieved. I don't think we should by any stretch minimize those first steps that we have taken. And absolutely, each country will want to take a slightly different tact. We see that across the board with all things. But we need some kind of common themes and some kind of coherence. And that's what this process will all start to achieve. But also the other thing that we needed was we needed action quickly because we know that those set of models are coming out within the next six months. So we couldn't wait for some protracted process. We couldn't wait for every country to to undertake legislation, for instance. What we needed was, we needed results and outcomes quickly. And that's what the summit delivered by ensuring that agreement on the pre-deployment testing, which will include those new set of models.

**OKOLO:** Thank you. We have a couple of minutes left. I wanted to see if we have any online questions. No. Okay, great. Let's see.

**AUDIENCE QUESTION:** Thank you. Hi. My name is Olga Balagalova, and I'm a professor at Johns Hopkins SAIS and also leading our Emerging Technologies initiative, which we just launched. And one of the things that we're thinking about is how we sort of train the next generation of policy students about how to govern new and emerging technologies. And so one of the things I wanted to know from you and from your experience with the summit is what can sort of this next generation of students, what are some of the open questions that they should really be grappling with as they study how to govern and build policies for new and emerging technologies problems?

**DONELAN:** Well, I think I think the biggest challenge here for them is the unknown, isn't it? This technology, as you will be an expert in, is emerging so quickly, quicker than we have seen at all. You know, when we think about the fact that man went from the horse and cart to the space race in just over a lifetime, and if we look back five years ago, large language models couldn't even string together coherent sentences. And now they can pass the bar, they can pass SATs, etc.. So who knows where we will be in five years time. So I think the biggest challenge for your students, well, of course, being the to stay nimble and agile in a way that policymakers have never had to do before, and that they will be, to a certain extent, no rules of the road. They will have to be agile and and listen and and stay on top of the agenda as it progresses and moves. And and it speaks to what we were talking about before, about not just relying on one set of views on this because it is so important and because nobody has all the answers. So it is about relying on the views of civil society, but also of industry, and really just engaging in the detail as well.

**OKOLO:** Thank you. And I'll take one last question on the back, blazer, the blue blazer, please. Thank you.

**DONELAN:** I thought I was gonna be a fight there between the two.

**AUDIENCE QUESTION:** Thank you very much. Secretary Donovan, Philip Meehan. I work for one of the tech companies that was at the summit, Amazon Web Services. My question is, maybe on a personal level, we talk a lot about the speed of this. How do you personally keep up? It's more than a full time job just to try to stay up to date with papers and what's going on and the advance in the industry. How do you personally stay up to speed and make sure your personally not falling behind?

**DONELAN:** Well, I have to admit I read more academic journals than I ever did at university now, in this office. How I keep up to speed is really staying on top of the detail and doing the due diligence and the reading and making sure that I'm speaking to those that are in the industry that are working on it day and night, and that I'm speaking to my counterparts across the globe as well, because it is moving fast, thinking is moving fast and it certainly won't stand still. So we need to continue to stay on top of the agenda, which is an incredibly exciting one. We can focus heavily on on all of the risks and and the doom and gloom. But the reality is the reason why we've got to grip those risks is because these opportunities could be transformative for humanity, for my child, for your children, for our grandchildren. And I think that we therefore have a duty to get this right.

**OKOLO:** All right. Thank you so much. I'd like everyone to join me in thinking Secretary Donelan, for her presence today. Thank you so much. And please stay seated as we transition to the panel, please.

**WIRTSCHAFTER:** Hi, everyone. Okay, this is on. Thanks so much for joining us today in person, online. My name is Valerie Wirtschafter, I am a fellow here at Brookings in Foreign Policy and in the Artificial Intelligence and Emerging Technology Initiative, which is a bit of a mouthful. I'm really looking forward to moderating this discussion after what has been a really great session with the secretary. Online viewers, please submit your questions either via - I don't know where the camera is, I don't know where to look at you - but please submit your questions via email to events at Brookings dot edu. Or you can also use Twitter and use the hashtag UKAlsafety. We'll be taking questions of course, from the audience as well. I hope this will be a great conversation. First, I want to briefly introduce each panelist. We have Chinasa Okolo, who you've seen on stage here already. Chinasa, do you want to give a little background about yourself?

**OKOLO:** Oh, sure. Happy to do so. Hi everyone. I just finished my Ph.D. in computer science at Cornell University, and I joined Brookings in August as a fellow in the Center for Technology Innovation, where I focus on AI governance in the Global South and with a focus on Africa in particular.

**WIRTSCHAFTER:** And we have here with us Miranda Bogen, who is the director of a newly formed AI Governance Lab at the Center for Democracy and Technology. Amanda - Miranda, would you like to introduce yourself?

**BOGEN:** Sure, so the Center for Democracy and Technology is a 28-year-old civil society organization, one of the earliest digital rights organizations, that over, you know, in recent years has grown from focusing on traditional issues like privacy and free expression to also thinking about equity and larger societal impacts of technology, and especially in recent years, has been quite engaged in the policy conversation. And we realized that as we're getting to the point where actions are happening, you know, governments are putting out executive orders and international agreements, the next phase of the conversation is coming, where we're going to talk about what are the details? What does it look like to hold systems and practitioners accountable? What are the evaluations that we need to develop to make sure that these systems are safe and not instigating harm in society? And so we founded the AI Governance Lab in order to help deepen the public interest participation in these conversations that are very centered on practitioners, you know, the companies that are developing this technology. I came from industry before this. I was working at mETA advising the teams building foundational AI systems and thinking about some of the implications, but also about what is the infrastructure and processes that companies need to

develop in order to actually fulfill and uphold the commitments that they're making and also to prepare for the regulations that are coming and to address the risks that everyone is becoming aware will need to be dealt with collectively.

**WIRTSCHAFTER:** Great. Thank you so much. So Chinasa, I wanted to start with you because it's my understanding you were at the summit. So what do you think some of the biggest successes were?

**OKOLO:** Yeah, I would say for me, just being, you know, successfully able to get so many different stakeholders together in one room. There were different roundtable sessions throughout the day, and we were, I was able to get a better understanding of just like the different perspectives from civil society and government stakeholders and industry stakeholders, for example. And then also just being able to interact directly with some of the ministers, especially from Nigeria. I'm Nigerian and that was really important for me to connect with them, with the Minister Bosun Tijani, just to understand like what the initiatives are being done in terms of the AI space, you know, within Nigeria and also throughout the continent as well. And then also just really to get a broader understanding of what AI safety means, you know, to these different stakeholders. And also, you know, like what what efforts will be needed to actually work towards, you know, safe or responsible AI.

**WIRTSCHAFTER:** Now, Miranda, CDT was there as well. What was your reaction to the summit? What do you think was kind of the the major successes coming out of it?

**BOGEN:** Well, the collection of civil society organizations that were invited to participate had been concerned that the framing of the summit was overly narrow. And we released a coalition statement to that effect early on in the summit. We were pleasantly surprised that the conversation did acknowledge that there were issues of today that needed to be dealt with, and so we were glad that there was a recognition of that imperative. And that and I think the conversation around figuring out what do we need to do next, what infrastructure, through research, through testing needs to be developed, is important to bring people together on and get them to to commit to. But another important thing was just putting something on the calendar that was so high, you know, a central and so covered in the media, I think put the impetus on participating governments and other institutions to really get their act together and put things out. So we saw the White House executive order released Monday, the same week we saw the Office of Management and Budget guidance also be released, which had been a long time coming. And we saw the Hiroshima principles and recommended practices come out as well. It was a very busy week for those of us in the field, but at the same time having landmarks in this process that is moving so quickly, having moments that push people to take action to make commitments is really important. So we were glad of any opportunity to push both governments and practitioners and companies to make those commitments to publish their work that they might not otherwise be pushed to publish. So we're glad to see that. But, you know, that, more to do for sure.

**WIRTSCHAFTER:** Did you want to add something? Oh, great. Well, now so I know that there was a lot of excitement that came out of the summit. Where do I think, you know, maybe going down, what do you think was sort of missing or kind of wasn't, maybe was a missed opportunity from the summit?

**OKOLO:** Yeah. So something that brought up, I was invited to give opening remarks at one of the morning roundtable sessions. And I mentioned just a lack of focus in terms of the actual harms that AI development poses to specific groups of workers, especially data labelers, content moderation moderators, for example. And so a lot of this work is situated or concentrated in the Global South, mean we're kind of seeing, you know, see similarities to colonial level extraction. And so it's really important to understand how governments can enact protections for these workers and also like data data labelers - excuse me - are also present in the United States and others like quote unquote, developed countries. And so, you know, as the need for data increases, you know, the need for these workers will also increase. And so it's really important to think about how regulation can also cover labor aspects. And then also just in general, the representation. It's

good to see that, you know, countries from, you know, the Global South included in these conversations, but I think there is much more work needed to be done, needed to be done to elevate these voices and understand how we can like level the table or democratize access to these conversations on AI governance and regulation.

**BOGEN:** Absolutely, plus one. One trend that we're seeing in the U.K. Summit was symptomatic of this was the over-focus on the technical elements of the risk at hand. And while there was acknowledgment of broader societal risk, the conversations are on what to do about those risks, tend to focus on the technical interventions and technical evaluations. And I think as anyone who's been studying technology and policy for some time knows that the technology on its own is not the operative piece. It's, it fits in with context, it fits in with society. And so we can't just build testing and evaluations for models themselves. We need to also be thinking about the societal interplay. We need to have the expertise on that interplay part of the conversation. And I worry that will, that, as you know, there are technical problems to be solved, but I worry that will become too enamored with the technical problem and forget that these systems will end up embedded in in domains that have a lot of expertise that's needed to deploy AI responsibly in those domains. Medicine, as an example. There's some great research out of Duke University about actually implementing sepsis detection algorithms, for instance. And much, you know, much of their finding was about the extent to which the humans who are interpreting and dealing with these systems need to be brought into the systems and understand at least the contours of them for that, that, the promise of the technology to actually play out. And so thinking about what is the, what are the concentric circles of research that are needed and of deliberation around what the appropriate interventions are.

**WIRTSCHAFTER:** Great. Thank you. Chinasa. I was wondering if maybe you could elaborate on what you'd like to see more in terms of especially, you know, if we think about the national level conversations we've been seeing and then the international level conversations. What would you like to see more in terms of kind of democratizing that access that you mentioned at either or both of these levels? And, you know, is it more of a domain for the domestic level? Is it part of the international conversation? Is there a space for for both conversations at the same time? What would you like to see more of?

**OKOLO:** Yeah, I definitely think there's room for both. I, you know, asked the secretary earlier about just to provide a little bit more context on the AI for development initiative. And I think that's a really good start but as we see, and probably as you all know, a lot of the AI companies are based, are concentrated within the United States and also within the U.K. and Canada as well. And when we think about the concentration of like computing power, GPUs, internet access, these are all controlled by major telecom, telecom companies. And also these big tech companies are increasingly like gaining a controlling interest in developing and building like undersea cables, for example. And so when we think about AI development, I think it's really important to understand how we can, like democratize access and and more importantly, when it comes to like actual who's developing these technologies and who's being able to, like, control them - let's say, not control, but like implement them in ways that serve the needs of communities. And so, for example, places like Microsoft, IBM and Google have established labs in countries like Nigeria, Ghana, South Africa, India, for example. But again, you know, these companies have their own respective interests. And I think it's really important that governments find, local governments within the Global South find ways to support these note native or local ecosystems, along with getting support from the broader, say, like international community to help train the talent needed to maintain these ecosystems as well. So yeah, it's a bit of a bit of both.

**WIRTSCHAFTER:** Great. And then Miranda, kind of building on your comments, what do you think practitioners can do who developed these systems, and where can I think governments and civil society play a role in sort of integrating and building those conversations more?

**BOGEN:** So the U.K. AI Summit, as the secretary said, was very focused on the frontier, certain models above, above a threshold of capability or compute because of the particular risks that they may pose. But what I think we really need to do is think about the common infrastructure that will help us address both the challenges of today and the risks of tomorrow. And in particular, the companies who are building this technology can implement those, those interventions already for the systems even that are not in scope of these broader concerns as confidence-building measures in on one hand to show that they actually do care about the harms that their technologies might cause. I think the focus on specifically high capability models is again, overly narrow. Many models that are extremely simple can wreak havoc on people's lives now. Smaller, but yet still advanced models can pose some of the same risks that people are concerned about for the frontier models.

And so what are the steps that developers can already take today to begin if not fully mitigating those risks, at least creating the infrastructure where as we learn more about how to mitigate those risks, will be able to go back and figure out where those mitigations need to be applied? There are some really basic steps. For instance, my expertise is on fairness and discrimination, and so few developers of AI systems even disaggregate their top-line metrics at all. A few categories of, where might the performance of this system be lower than we expect if we just look at the global average? And I think building on the earlier question of how do we think about what ought to happen at the international level versus the domestic level versus the practitioner level, when when we're talking at the international level, things can become so abstracted as to be disconnected from the people who will ultimately be impacted. And I think we need the domestic to, you know, also keep that top of mind. That doesn't always happen in policy at all, but at least the domestic regulatory space and the domestic industrial conversation can be aware of the way in which the implementation of different regulations or the deployment of different technologies will affect a particular context that might not be transferable everywhere around the world, but the practice of identifying what that looks like in a in a context that touches and, you know, people, touches touches domains is what we're ultimately going to need to address the risks because not everything is going to be able to be done at the infrastructure layer, at the at the foundation model layer, at the international layer. We know that very few institutions, global governance companies, etc., will give up flexibility or power unless there is some kind of dire circumstance that warrants that. And so I think we need to be realistic about where constraints will be able to be applied effectively.

**WIRTSCHAFTER:** Now, we've seen a lot of, I think, domestic conversations both in the U.S., and I'd love to get each of your reactions also to the executive order. How do these play in, I think, with a lot of those other international conversations, are there things that should be potentially in the kind of international realm, specifically things that should be much more on the domestic side? What are your views on this?

**BOGEN:** I think that the executive order, for instance, set a good example and also the vice president's speech in the U.K. set a good example of how to holistically think about the types of impacts that that AI will have and not overly cabin to a particular frame of the problem. It was an extremely broad order that recognized the actions that current, you know, regulators with their existing authority can take, but also recognized where new studies are needed, where new action is needed on the part of the federal government and the industries that they regulate. So I think that we'd like to see more of that and not let the conversation become overly national securitized, for instance, which can tend to happen when you're at the international level.

**OKOLO:** I also agree with many of the points raised. I would say for me it was surprisingly robust and I think it's really important to understand like how the U.S. will leverage that in order to actually make headway or touch towards regulation. You know, the AI, coupled with the AI Bill of Rights, I think it sets a really good precedent or foundation, you know, for thinking about how the U.S. can enact regulation. But I don't want to get to a point where, you know, we're coming out with so many different orders or strategies, etc., and and not thinking about how we can actually start thinking about or enacting regulation. So that's what I think.

**WIRTSCHAFTER:** Yeah. So I think the secretary said it was the beginning of a journey. Right. And I think we're coming up this year in 2024 is 40 elections around the world. So what kinds of things do you think are kind of possible in the here and now for the risks that are in the here and now, beyond the kind of existential risk conversations that I think are quite prominent?

**OKOLO:** Yeah. So some things I've been looking into, especially when it comes to elections within the African continent, are just the ability of AI-fueled disinformation to spread. I did some work, my Ph.D., that focused on mis - excuse me - misinformation spread throughout the COVID pandemic on social networks and African communities. And so just understanding that there is a lack of literacy when it comes to like being able to discern, let's say, like AI-generated content. I think this stands generally for different populations, but within the African continent, particularly when it comes to things like memes or other kinds of content, it's really hard for like older populations to understand like what this, what kind of context is being relayed in the in these kinds of media. And so given that, you know, a lot of African governments are dealing, well not lots, but there are a couple of African governments dealing with, you know, very I'd say like sticky political situations, especially when it comes to coups. And then also, you know, coupled with internet shutdowns that frequently happen throughout these elections, I'm really concerned about how AI-generated media will impact this. And we've also already seen some use cases, especially within the Nigerian elections that happened earlier this year and also Burkina Faso, some instances of AI-generated content is impacting the political scene throughout that throughout the country as well, so.

**BOGEN:** On the broader question of what can practitioners be doing now, agnostic to the particular harm that we might be concerned about, there are some basic machine-learning best practices that are very often not adopted, especially as you know, we want to create a competitive environment so that that, you know, the main players don't sort of lock themselves in as the as the as the leaders for the next, you know, coming decades. But at the same time, smaller companies, start-ups, you know, don't have the same resources as those big companies and don't necessarily know what enterprise risk management processes might look like to address these concerns. And so some basic practices like internal documentation, you know, very, clarity on what data is being used to develop systems, traceability of where certain models are being used in certain products, these are all practices that will at least increase the confidence of the developers of the technology that they know what they're developing. And it will be very difficult for anyone else to do that if that foundational infrastructure doesn't exist. So anything we can do to incentivize more deliberate development of the technology will help foresee some of these risks. They might not solve all of the risks, but at least people building the technology will be able to flag when something they're building might have a particular implication, when certain evaluations or tests that haven't been done need to be done, that the challenge of of ascertaining which tests and mitigations are relevant in which circumstances is is a daunting one. It's not it's not simple, although we wish it would be.

So the question earlier on of what should we teach policy students of how to how to navigate this space, I, it made me think of the challenge of how do we make sure we don't over-rotate on STEM education because even the people developing the systems will need to be thoughtful about where those systems might be deployed so that they can alert people who have that domain expertise that there might be some steps that need to be taken, even if they're not expert in those steps. The recognition of the implications itself is is a skill that I think we need to invest in so that there's a sense of responsibility on the part of people building the technology, and there's at least a pathway to deploying the the approaches that we're developing to address the risks where they're appropriate.

**OKOLO:** And so kind of leading off of that, especially when it comes to machine learning best practices, explainability has been, I would say, like, expressed as a, one of the maintenance of, I guess, increasing transparency in machine-learning systems overall, especially when it comes to like Chuck Schumer's SAFE framework. And so I did some work in explainability during my Ph.D. and we found that like for novice technology users, or just like users that have very little

experience interacting with AI or don't have the domain knowledge, it's actually very hard for them to understand common explainability methods or what these try, what these methods are trying to explain about models. And so I think it's really important to think about how we can do, I would say like more translating of the impacts and actual results or let's see, like outcomes of these systems to the general public in ways that they can understand.

**WIRTSCHAFTER:** So speaking, I think, of the general public, ChatGPT last year kind of dramatically expanded and exploded some of these conversations. And while there had been conversations ongoing about governance of AI, it kind of created a renewed sense of urgency or maybe created a new sense of urgency around these tools. Do you think that has been helpful or harmful for broader AI governance questions? And then I will open it up also to the floor for questions after that. So think of some of, some great questions because we do have some time.

**BOGEN:** On one hand, the increased attention to the challenges is beneficial. There were, there have been communities calling for action in the area of AI, simpler AI, but AI nonetheless in our grand conception of it, for for years now and we are seeing quite a lot of action now. On the other hand, ChatGPT and similar chat bots I think have shaped what people think of as AI, as being generative AI and interactive language-based systems, which is one modality that we should be paying attention to. But when we're talking about these advanced AI systems, I often see people reverting to talking about it in the context of chat bots and interactive systems. And what I anticipate happening, and the developers and the companies are already talking about this, is building the foundation models that power these chat bots into other systems as agents, as systems that can take action on your behalf, which really closes the circle back to automated decision making, which the many concerns stemmed from to begin with. And so I think we can't forget that there are many shapes of AI systems that we're going to need to address and make sure that policy solutions don't over-rotate on a particular a particular type of these systems as they roll out. And we need to be expansive enough that we're capturing the impacts that are most likely to affect people and potentially harm them.

**OKOLO:** Right. I pretty much agree. Yeah.

**WIRTSCHAFTER:** Great. Well, I will open it up to questions. We've got a couple in the back. I know there's microphone runners. We have a lot of questions. That's wonderful. So, yeah, so if you want to pass it along, we can kind of take a couple questions, maybe the two of you together, and then we'll pop back up to the panel. And then I saw other hands kind of floating around here.

**AUDIENCE QUESTION:** Hi, my name is Abi Reddy. I'm currently a student at Georgetown University. I'm in the Master of Science and Foreign Service program and also a disinformation research assistant at CSAT. I was wondering a little bit more about talent development, what steps do you think countries like the U.S. and the U.K. can be taking right now to ensure that talent development in the Global South remains equitable and to prevent exploitation in the future?

**AUDIENCE QUESTION:** Thanks. My name is Steve Perkins. I do some consulting in data privacy and want to ask about data privacy. Certainly here at Brookings Cam Kerry and CDT both been concerned with data privacy for years, but it seems like data privacy is sort of getting ignored in all this. Is data privacy just not sexy anymore and we're just going to move on to AI? We obviously don't have any federal privacy laws compared to GDPR or anything like that. Do we just forget about data privacy or where does that fit into AI governance?

**WIRTSCHAFTER:** Feel free to take one or both.

**OKOLO:** Yeah, I'll take the first question. And so I think it's really still a bit complicated to understand like what the U.K. and the U.S. can do, because I know that the U.S. specifically has like started advocating or like for AI talent to actually migrate to the U.S. And so I think that we're going to kind of see brain drain continue, especially when it comes to like, you know, how it mirrors current day practices in like the medical context increase when when it comes to technical and like

AI talent specifically. And so I'm actually a bit worried that, you know, as talent is trained within Africa or within Southeast Asia and Latin America, that they're going to be encouraged, you know, to migrate to the U.S. and the U.K. and, you know, increasingly like, I guess like build up that tech ecosystem while leaving, you know, other native local tech ecosystems to kind of I'd say like just like diminish a little bit. And so I think the U.K. and U.S. have to, like, take a measured approach. But I think also governments kind of, like I mentioned earlier, have to really think about how they can invest and support their local existing ecosystems, because there are many thriving tech ecosystems within the continent, especially within Egypt, Nigeria, South Africa, Kenya, for example. And I think that governments have to be really intentional in terms of how they develop regulation to support and not hinder startup growth. And then also think about how they can allocate investments towards, you know, ensuring that AI development continues to rise within these regions.

**BOGEN:** And on the question of privacy, I think you're exactly right. It would be great to have federal privacy regulation legislation to set the ground rules for how companies developing systems can use data which is their lifeblood. At the same time, many of these systems were trained on publicly available information, and our conception of what privacy means in that context is less developed. Oftentimes, people don't know their data has been made public. Public data can be, you know, matched in different ways with private data. And at the end and similarly, there are sort of new questions that put privacy in a little bit of tension with other goals around other harms of AI, for example, discrimination and inclusion. There are often, you know, foundational issues where data sets are not inclusive. And so that creates an incentive for companies to go out and collect more data, but often in sort of sloppy ways. And so figuring out what are those baseline rules, but then how do we open the conversation for ways in which AI is changing the types of issues that privacy, privacy goals were trying to address? I think we'll have to do that as well, because it's, you know, privacy issues are no longer about just data disclosure and use. It's more about what might AI systems end up being able to, if not kind of explicitly identify, then at least have, create the experience that they're that they have information that they ought not, especially as we get into AI systems that are dealing with more and more unlabeled data. So they're not classifying people into certain categories, they're simply learning a sort of vague representation of people's behavior and details and then providing information or acting with that information. It will be, it will create the experience for people that they've disclosed something, even if they haven't officially, you know, they haven't explicitly disclosed it. And and it might create the experience of a system acting on information that they ought not act on, even if that information was never explicit in the system. And so there's a lot lots of issues that we're going to need to continue digging in in the area of privacy. But we should definitely not let that drop in the interest of new and shiny AI policy questions.

**WIRTSCHAFTER:** So a few other hands, some in the back, one up front.

**AUDIENCE QUESTION:** Hi, my name is Maria. I'm with the Electronic Privacy Information Center. And my question was a lot of eyes right now are on the EU with the AI Act, the UK with the summit. I was wondering if there are any other countries that are really putting out regulations or really thinking about these issues? I know Argentina and Brazil and South America are putting out privacy laws, and I was wondering if any countries in Africa were putting out any of this or if you were aware of other international areas that this might be good to look into?

**OKOLO:** Yeah, definitely. So going back into the data privacy space, I think a lot of African countries, I think it's, I'd say like around forty or so have active data regulation. But when it comes to AI regulation, no African country has enacted that yet. But there are lots of AI strategies that have been published or in development. I believe Mauritius was the first African country to produce an AI strategy and along with Egypt, Morocco, Rwanda actually released theirs earlier this year. Nigeria has a draft AI strategy, and what the new minister that's in place, they're in active development of that strategy. They actually have a very unique, I would say, process that they're working on and they're very intentional in recruiting expert, experts from the diaspora to contribute to that. And so I'm not sure of the timeline for that, but will expect to see it within the next year. And



so I've also been personally involved in the African Union's continental strategy for AI. AI continental strategy for Africa, excuse me, and so we've been working on that over the past year and a half, and so right now it is under review and the goal is to publish this within late January or early February, around the time that the AU member-wide meeting happens in Addis Ababa. And so that's one framework that's also in development, too.

**WIRTSCHAFTER:** I think we have one here.

**AUDIENCE QUESTION:** I was wondering if you read Fei-Fei Li's book because it gave me quite a shock. It's only about a week old. But what really shocked me was that, you know, she collected 20 million data photos so as to train a computer to see objects the way we do. So it was a very strange dataset, you know, Ph.D. students, two, only two sort of kept with her over eight years because it was took so long. But the point of it was that if you had such a huge dataset, once you put a computer on it, the computer was as happy with the largest dataset than something that might have been a thousand times smaller. So when you look at the new AI of Fei-Fei Li, or of Demis Hassabis, which is different because, you know, basically he took the game of Go, which has billions and billions of pattern recognitions. They were sort of very strange datasets in the in a certain way. So I'm wondering how to sort of square the circle because I'm not quite sure how we then learn from that in terms of what data we should be regulating or collecting.

**BOGEN:** It's a really interesting question because as we're thinking about where do we draw thresholds beyond which there might be additional requirements, if any of you have ever worked with engineers, you know that if you give them a constraint, they'll do everything that they can to work around that constraint. And so I anticipate that will be the case in this context as well. If it suddenly is much more expensive or burdensome to build a model that is, you know, a hundred billion parameters or whatever, we draw the line they'll look for, one that's, you know, 99 billion parameters. And so I think that's something important. The concept that more data is always better is also one that doesn't always hold. There's advances in few shot learning, reinforcement learning that won't end up needing quite so much data. At the same time, there are hypotheses that as, that scale is the way that we're going to continue growing these systems. And so researchers disagree about that. But the most interesting thing is on that original dataset that Fei-Fei Li built, because we see that embedded in all sorts of computer vision systems. And if you go back and look at the dataset, it was manually annotated with a, you know, a variety of different categories of things that are in those data in that dataset. And those categories are wacky and oftentimes things that are other concepts that we would not use today or concepts that are not visually identifiable. And I think that just points to the importance of developers, of technology being thoughtful about the data that they're using and about how they're leveraging that data, as in some cases, labeling that data. And as we move into the phase of not paying so much attention to the data and data sets for these large models, but focusing on evaluations and how do we test these systems, the same care is going to be really necessary that we're actually building evaluation data sets in a thoughtful way, that we're that we're not mis-specifying what we're measuring, that when humans are evaluating the outputs of systems that will be used in these evaluations, that we're taking our learnings from data labeling practices in general, both about labor and about fidelity of how people are labeling the categories that are available to them and just being aware of the limitations that that layer in the supply chain of AI, the role that that plays in the overall assurance that we might have around the behavior of those systems.

**OKOLO:** And also thinking about like investigating data sets themselves. I'd like to highlight the work of my colleague Abeba Birhane. She's based in Dublin and she does a lot of digging into these datasets. And she she has a paper coming out at NeurIPS later this year that focuses on the LAION dataset, and some of her earlier work had has shown that lots of these datasets contain pornographic images and other just like, you know, unsavory things. And so I think I'm her work has been really important in highlighting the importance of, you know, like auditing these datasets.

**WIRTSCHAFTER:** We have one here and one over here.

**AUDIENCE QUESTION:** I was curious for your thoughts on how we can have a more intelligent and grounded discussion about open source risks, because it feels that at the moment there's two camps, one that is mustering kind of 1990s-era arguments in favor of open source operating systems to argue that more open source is better. And then another camp that's very concerned about national security risks of open source AI systems, that's saying, whoa, that's not even the conversation that we feel is the important one. What, if any, steps would either of you recommend to to make that conversation more productive?

**BOGEN:** I think NTIA was tasked with with gathering information about that exact topic. And I think creating structure structures where people can better articulate their perspectives in this is going to be a start, because the conversation did spin up very quickly and very few people were kind of prepared. And the folks who had been deep in this conversation for for some time hadn't done the work to kind of bring other people along in explaining what their concerns were. I personally still don't feel I have a full, full and and coherent view on what the what the concerns are and whether they are in fact, novel and new circumstances that prevent, excuse me, presented by AI or whether they are similar concerns around any sort of technology that can have multiple uses depending on the intention. And so I think more and more just more articulation of each side and and we're trying at CDT to start bringing folks together to really sit down and better understand the contours of the conversation, because it, the the executive order, for instance, had a section on models above a certain certain threshold and having certain reporting requirements, it came a little bit out of nowhere. I think the national security community was having those conversations, but that didn't extend to beyond that community. And so I think pulling more people in and getting specific about concerns and opportunities will be really important.

**OKOLO:** I delegate that to Miranda.

**AUDIENCE QUESTION:** Hi again, how concerned, given your experience in Meta before, and where you are now at the center, are you concerned about the, any impact of AI in next year's elections in the U.S.? A and B, how concerned are you? Can you can you elaborate a little bit of, we're going to see AI produce voting registration cards? I don't know.

**BOGEN:** I think Arvind Narayanan has some great thoughts on this, which is that the bottleneck here is not AI necessarily developing new, like, misinformation. It's how that information is distributed. And so I think the shape of the problem is similar to what we've seen in recent elections. And I thought that, again, the the U.S. has been considering how do we develop technical approaches to identify the provenance of of artificially generated content, but more so, how do we make sure that authoritative information is is marked in such a way that people can trust that that information is solid? So I don't think we'll solve this by this year's election. CDT does have an elections expert in-house, and so that's something we're going to continue working on. But I think the question to ask in many of these things is what is AI actually changing about the circumstances and what is just going to build on what we've already seen happen and what sort of resilience do we need to that?

**WIRTSCHAFTER:** We have time maybe for one more question.

**AUDIENCE QUESTION:** Thank you. Thank you very much. This has been very educational. My name is Michael. I work with a nonprofit called the Open Voice Network. We focus on the human machine voice because we see this happening now that I can talk to a machine and, you know, coming back to some of the points you made, there's no guidance or laws regarding human to machine voice for cybersecurity or regulation. But also number two, voice is such a strong communication modality that we've just, we've written guidelines on, what if I don't speak English, how can I get that device to speak to me in the language I prefer? And if it does, again, what protects it? So I'm coming back to you to wonder have you looked at that human to machine voice component? And you know what we do in health care or what we do in infrastructure? I want to know what time the bus comes. I want to be able to hear that in my language, but at the same time respect my privacy and my security.

**OKOLO:** I haven't necessarily looked too much into it, but it what you just said mention reminds me of Humane. There's this tech company, they released something called the AI pen. Yes. Yeah. And so I think that work is very interesting. And just in general, for a lot of these, like large language models, they only represent a very small portion of the world's languages. And so I've seen just lots of different examinations or analyzes of, you know, people prompting these systems, especially when it comes to like getting information that's relevant in these native languages. And a lot of times like they're just like literally spitting out gibberish. And so it's really important that these models, you know, account for these contexts. And that's why I also think it's really important that local development, especially when it comes to the African continent and the rest of Global South. Yes. You know, be prioritized, you know, developers within this region be given the opportunity to contribute to larger language models. And they already are. There's lots of communities within African continent like Deep Learning Indaba, Masaka NLP - Masakhane, excuse me - and Ghana NLP, for example, that are actively working towards bridging this, you know, like digital language divide. And so I think that that's something that will be really important as we progress.

**BOGEN:** And my colleagues at CDT have a great recent report also on the linguistic challenges in large language systems. But I think developers are thinking about voice to voice models that sort of surpass the written component of language in general, which is an interesting way to potentially address the fact that written data may not exist in many of these languages. In terms of privacy, I think we saw a few years ago the scandals that hit Amazon Alexa and things like that. And so again, this is a question of fundamentals of people developing this technology. Are they aware of sort of human concerns around how their data is used? Are they giving people some kind of control? Are they allowing them to, if not opt into their data being used to train models in the first place, at least opt out of data being retained in a modality you may not be keen on having data stored in perpetuity.

**AUDIENCE QUESTION:** [off-mic] You know, to add to that, we've been looking at the human machine, but you could be in a location where we're having a conversation and that fourth chair is a device. You want it to communicate to a second device via voice so we humans know if these two devices are talking. and [inaudible] privacy regulation, can I trust that device that's responsible or the developers?

**BOGEN:** And many of the imaginations of what AI can accomplish, if you dig in to what the technology would require, it is really ubiquitous data collection for actual interactive capabilities, for sort of knowledge of preferences, all these things. And I think we'll have to see if that's a tradeoff people are willing to make and make sure that that conversation is sort of had out in the open and not one that's foisted upon people.

**WIRTSCHAFTER:** All right. Well, we have hit time. That was, I think, a great note to end on. Thank you all so much online, in person, for joining us for this really timely conversation. I know we have Brookings are going to continue to have more conversations about this. So stay tuned for more of that. And we look forward to seeing you again here soon. And thank you - can we give a round of applause to these panelists? Thank you so much for joining us and we look forward to seeing you all again soon.