THE BROOKINGS INSTITUTION

FALK AUDITORIUM

ON THE ROAD TO THE INDIA AI IMPACT SUMMIT: GLOBAL AI GOVERNANCE AND THE HAIP REPORTING FRAMEWORK

Washington, D.C.

Thursday, January 22, 2026

WELCOME AND INTRODUCTION:

ELHAM TABASSI
Senior Fellow and Director, Artificial Intelligence and Emerging Technology Initiative, The Brookings Institution

KEYNOTE:

H.E. VINAY MOHAN KWATRA
Ambassador of India to the United States of America

REPORT PRESENTATION:

MIRANDA BOGEN
Director, AI Governance Lab, Center for Democracy and Technology

VALERIE WIRTSCHAFTER
Fellow, Artificial Intelligence and Emerging Technology Initiative, The Brookings Institution

PANEL DISCUSSION: WHAT IS NEXT FOR HAIP?

MODERATOR: CAMERON F. KERRY
Ann R. and Andrew H. Tisch Distinguished Visiting Fellow, Center for Technology Innovation (CTI), The Brookings Institution

KEITA AZUMA
Head of International Affairs, Japan AI Safety Institute

AUDREY PLONK
Deputy Director, Directorate for Science, Technology and Innovation, OECD

SHATAKRATU SAHU
Associate Director for Technology, Media, Telecommunications,
US-India Strategic Partnership Forum

PANEL DISCUSSION: INDUSTRY PERSPECTIVES

MODERATOR: MIRANDA BOGEN
Director, AI Governance Lab, Center for Democracy and Technology

AMANDA CRAIG
General Manager, Responsible AI Public Policy, Microsoft

PHIL DAWSON
Head of AI Policy and Distribution Partnerships, Armilla AI

EMILY MCREYNOLDS
Senior Fellow, Future of Privacy Forum

DAVID WELLER
Senior Director, Emerging Tech, Competitiveness and Sustainability Policy, Google

DOMESTIC EFFORTS: NIST'S DOCUMENTATION ZERO DRAFT

MODERATOR: ELHAM TABASSI
Senior Fellow and Director, Artificial Intelligence and Emerging Technology Initiative,
The Brookings Institution

JESSE DUNIETZ
AI Standards, Policy, and International Engagement Lead
National Institute of Standards and Technology

CLOSING REMARKS:

ELHAM TABASSI
Senior Fellow and Director, Artificial Intelligence and Emerging Technology Initiative,
The Brookings Institution

* * * * *

TABASSI: Hello and good morning. Thanks so very much for being here. And hello, good morning, good afternoon, good evening to many people joining us online depend on what time zone you are. So, we are very grateful to have all of you here, and I'm very pleased to welcome you to our today's event, marking the launch of our report on Hiroshima AI Process reporting framework, developed in collaboration and partnership with Center for Democracy and Technology. Transparency and impact reporting frameworks such as HAIPs provide a credible mechanism for demonstrating responsible AI practice. The India AI Impact Summit shows how this accountability can be applied at national scale. Together with the hype reporting and other transparency mechanism, they connect global transparency standards with real world AI deployment and measurable public impact. The discussions that we are having today cannot be more timely at the, as the momentum builds towards one of the most consequential moments in the AI policy.

This year, the India AI Impact Summit as the fourth in a series of international summits that started in UK in 2023 and the first to be held at the global south. The summit will help shape global conversations on AI governance, adoption, and impact at a critical juncture for both geopolitics and technological development. To discuss India's priorities and vision for the summit, we are most honored to be joined today by his Excellency Ambassador Vinay Kwatra. Ambassador Kwatra is a career diplomat and a senior leader of India's Foreign Service. His distinguished career has included posts as India's ambassador to Nepal, as well as its ambassador to France and permanent representative to UNESCO.

Earlier in his career, he served as director and joint secretary of the South Asian Association for Regional Cooperation in Katmandu, the deputy chief of mission in Beijing, and a variety of other postings around the world, including Ashkin, Durban, Karachi, and Geneva. Most recently, prior to being named India's ambassador to the United States, he served as India's foreign secretary leading the nation's Ministry of Foreign Affairs.

In his current role, he has made AI a high priority for the mission and has been deeply involved in discussions on AI policy in the US and abroad. We are very grateful that he could join us today for today's conversation. And without any further ado, I give the floor to the ambassador.

KWATRA: Thank you very much. Distinguished participants, colleagues, and all of us who are joining us today. Very good morning and thank you very much Brookings for convening today's discussion on global AI governance especially in the context of the upcoming India AI Impact Summit, which is to be held in Delhi on February 19th, 20th. If you look at the history of human civilization, it has always had a very fascinating relationship with the tools that shape it throughout. And I think we are presently witnessing a timeframe in our advancement where the tool of technology manifested through artificial intelligence, what we call artificial intelligence, which are basically mathematical models to process the data sets in a manner that gives you something which is very productivity enhancing, changing the manner in which you engage with your ecosystem and surroundings around you in fundamental way.

India AI Impact Summit is being held in this emerging, fast emerging backdrop, I should say. Naturally the technology, per se looks at how, we undertake many of the tasks differently. How we learn how we work. I would say even how we innovate, govern, create, and ultimately, naturally how we carry on with the daily our daily lives. I don't think it's any longer a question of the impact of AI. I think that is very clear. Some say it's gonna be transformative. Some have different points of view, but I think it's a question of how exactly it'll do it. What would that impact be and what sections of the human society would actually end up benefiting from it more, some less.

Those are all the points which would very much be part of the AI summit. There is naturally, as we see particularly in the us strong imp impactors on very large, huge, substantial investments leading to the development of, new frontier models pretty much every six months. There are big ticket announcements, large action plans. There are multiple ideas and proposals for artificial intelligence, both in the space of development, but also in the space of deployment of those developments. We in India approach the artificial intelligence like any other piece of technology from a purpose and intentionality and from a template that focuses on its value creation for the society.

Not necessarily just wealth creation but also looking at it from the value template. We are a developing economy, a fast growing, developing economy. In fact, the fastest growing large economy in the world the government of India to position this in our own developmental context and the context of our economic growth launched the, what we call the India AI mission. Prime Minister Modi

launched that. And the guiding vision of that is making AI in India, making AI work for India, but also building its cooperation possibilities. For the societies outside India, for the countries outside India. You touched upon it that the discussions covered under the past AI summits such as those in the UK CO Paris they laid a very important groundwork for focusing on multiple aspects of ai safety, ethics how we can work together in the space of ai.

The aim of the AI summit in Delhi would be to shift these conversations more from, I would say commitments more towards outcomes, specifically from principles to practice and from national strategies. Each country has national AI strategy to what could be benchmarked and measurable global progress in the field of ai. And from our perspective under the leadership of Prime Minister Modi, this would be anchored essentially in three themes. The summit would be anchored in three themes. The first theme deals with the people itself. I don't know if you can bring that up or, yeah. So, the first theme is for the people.

Now what it essentially would mean in terms of specifics would be that when you train models on data sets whether those data sets are mindful of cultural diversity of the society in which the model is being pre-trained, for example whether it doesn't have any inherent bias in that. So, to remove inherent bias in the training of the models before you deploy them. So that's the people aspect of it. And naturally that is built into the very design part of the of the models that you develop. But equally in its deployment especially when you spread it across multiple user cases of different industry the planet. So, profit yeah, is obviously focusing on innovation, responsibility responsible innovation, and naturally reducing the resource footprint which the AI consumes while producing its outcome.

And the Progress Sutra, which is the which envisages AI as an engine for inclusive growth. That is very important particularly for. Countries like India, which are developing economies, as I said, that it should be inclusive of everybody in the society. And naturally aligning itself with the global development priorities, with equitable access to the opportunity. It emphasizes, I think my prime minister is particularly very focused on this. It emphasizes democratizing the key AI resources. It should not just be available. It should be available rather to each and every individual in the society so that they can innovate on their own, depending upon what they think is good for them.

But the tool is available to all really. And under this overarching frame these three RAs of people, planet and progress. We have then built in the second layer what we call a seven interconnected chakras. Which essentially would take these AI products that are by developer, not just in India, but all over the world. And then in terms of their deployment, adoption, diffusion. To scale, spread them across these seven chakras of human capital. Essentially building a talent pipeline, not just for development, but also in case of user industries, talent development, not just for India, but also to be able to respond to the global needs in this space.

Inclusion for social empowerment naturally safe and trusted ai, that's very important. Resilience, innovation and efficiency science democratizing AI resources, as I mentioned and AI for economic growth and social good. So, the value creation part is very important and is embedded in the social, in social good. Now, when next month government representatives, technology experts, industrial startup entrepreneurs, and the large number of them we have really got an overwhelming response to the impact AI summit of next month from all over the world including in, in particular from the US when they meet next month to chart the course for defining this AI deployment for future.

I thought in that context, I'll just highlight. Some of the aspects of how India is looking at beyond just the ai summit the larger permeation of artificial intelligence into our ecosystem. We think we bring with us. India brings with it intrinsic strengths to the large-scale deployment and adoption of ai. I and I would even add adoption and deployment to scale, which is very important for any technology to eventually scale up at the level of society. And for the industry and enterprises to draw benefit from technology, it has to be diffused at the level of large society and large enterprises.

We are also probably one of the most, if not the most dug holder of most diverse data sets for the training of AI, and I would say both structured data sets and unstructured data sets and across different industry domain and domains of governance. And our society has a characteristic, which is it is it relates to technology very easily. So, it's able to adopt. And diffuse technology at scale relatively easily. We have classic examples of this in India our digital public infrastructure, which was built on the unique identification, unique payment interfaces, and other population scale platforms. And I say this with acute consciousness, their population scale.

So, they are spread for the entire 1.4 billion people. Yet the cost per unit of the usage is probably the marginal cost is zero, if not even negative. And if you count the benefit that they bring to the society perhaps the cost is even negative in a sense that they contribute more to the growth of the society rather than costing that much to the society. So, we think that the way we adopt this way we relate to technology, gives us a good blueprint for a frugal, inclusive, and a globally deployable innovative tools such as ai. We as an economy, we are currently at about 4 trillion GDP. We expect it to touch 7 trillion by the end of this decade.

And our goal is that by 2047, we are aiming of a GDP of approximately 30 trillion with a per capita income of $12,000. The idea is that this economic growth, the continuously, rapidly trend of economic growth will transform not just India, our society, but will also lift every other country in our region and the region. As such and in this growth, we think technology in general and artificial intelligence in particular would have a huge role to play. Not just in bringing more productivity gains to the existing industry, of course, but perhaps even creating new areas of economy altogether. We don't, I can't even begin to visualize what those would be because those would be function of how the society adopts AI and then innovates on it for them to create new business templates, new industrial templates and new growth templates, but are.

Strategy is focused on to harness the gains of AI. It's focused on one, building a robust compute infrastructure, which is very crucial for us to be able to build our own ai naturally foster our own AI models. Not necessarily, a trillion parameter models like the us, but we have our unique needs, and we can build models to respond to those unique needs. Creating a talent pipeline, as I mentioned earlier, not just for our needs, but also for the global needs and do this through education and of course, startup funding, which is very crucial because that's the nature of this industry and in the end. But the goal is to make it open, affordable, and accessible to all one and all ensuring that innovation reaches to every person at the population scale so that everybody's able to innovate.

If you, if I was to draw a derivative differentiation of this this effort then spans into all the five layers of AI architecture application layer at the top, naturally, the AI models layer. The one below it. The compute layer, which sits below the models to be able to build models on and the data centers and the network infrastructure layer, which then powers the compute, and of course, the energy layer,

which starts at the very base of it. I will not even touch upon the research ecosystem of mathematicians, engineers and the people who work in the labs to, to build these mathematical models, which then runs through these five application layers at one point or another. Our in, in, in our country, we are making progress in all the five layers simultaneously, and I would say in, in, in parallel building our own capabilities.

We are trying to build, we believe firmly that we should build our own sovereign ai sovereign LLMs attracting investments in data centers, not just from within the public and the private sector capital in India, but also capital from abroad. We have recently had a very impressive partnerships announced between India, Microsoft, India, Google, and couple of other countries to build a compute infrastructure that is that will power some of these ai applications at the application layer, as I alluded to earlier. We are focused on the kind of impact that the artificial intelligence and its algorithms will have on our society through impact on industry, governance, agriculture, delivery of healthcare and the range of economic endeavor that the society engages in. There are already large number of successful user cases, which are currently running in India of a kind.

I can, if I start listing, I'll, I can take the whole session just listing them, but it gives you it gives, it's an indication of how comfortable we are in picking up a technology and purposing it to the societal needs of governance and bringing, as I said, gains to the to the economy. Oh. At the models layer, which is very important layer for us to be able to build applications on. We know there are large number of models in the world. There are some prop IT models. There are some open-source models, and there are being developed by many countries in the world. These models in many ways, particularly the open-source models, they lower the entry barriers in reducing the cost of the training and deployment.

And it's very easy, at least on the open-source model, to be able to distill them and then build either new models or new applications that run on, on, on those models. But in India, through these models, startups, researchers developers are building on existing work but also doing completely new work from the scratch. But we, as I said, firmly believe that every nation may have its own sovereign models to ensure I think the important aspects of one data security cultural relevance as I said, the data sets parts in training the models. I think if you were to do an independent search on all the models in the world and say.

If I was to ask myself okay, these are 10 models in the world. What is the extent of Indian data sets on those, all the 10 models? And you will find the responses quite varied. It's not one figure in each of those cases, but if I look at the same Indian models, you'll find that percentage to be much higher. So that is a cultural relevance of the data sets, the way they are embedded in the in the models. And this is important because that then it takes you to the bias which is a model. Model is agnostic to, sentiments. This only recognizes the data and our data in India whether structured unstructured tends to be very context heavy data context heavy data is very important when you come to the question of benchmarking the artificial intelligence.

So, when you measure intelligence, you have to benchmark it on something. And that context, heavy data is very important when you start measuring the performance of AI against a certain benchmark of intelligence. So, all these are very important aspect, which we in our society and through the India Impact Summit will be addressing. We are currently we already have and are building several AI models under the India AI mission. Some of them would be launched at the at the summit next month. I can just give one example, which is a well-known example, which is what is called server ai. The name of the model is server.

Server May, it's a full stack India AI model which is trained from scratch. It's not a distal model. It's a model trained from scratch on 22 Indian languages. That's the context it has embedded in it. So, when it produces the outcome token, what we call next predictable token there are x hundred thousand words in the English dictionary, but here you have the dictionary set of 22 Indian languages on which it can produce the next predictable token of the model. So that's the strength of the extent of heaviness of the context that has gone into the model. It also has recognition for six UN languages. In it. Now there is another model being developed, which is a much smaller size. It's called that model is specific to industry cases.

And we are hoping, we don't know, we are hoping that it'll be a journal purpose model so that it can get adopted across different industry sets. These models will have India centric foundations naturally, because data sets are Indian. It'll support research startups, public sector applications. Again, if AIM remains affordable, inclusive, and populations scale solutions, in this case, AI solutions Indian

context. But of course, it'll be available for others. Also, the third area of the AI stack is the compute needed for training and inference. Naturally we don't have, a hundred thousand GPUs data centers.

Not, that's not the kind of compute we have, but what government has done as an initiative is government has aggregated the compute available in the country and optimized the use of that aggregated compute at a cost, which is one third of the cost, which is available globally. So, you can take smaller compute, but optimize it, bring it, make it more efficient by aggregating it. And giving it where it's needed most. And of course, as we grow compute size would grow. I mentioned about two, three international cooperations. This would lead to larger compute presence in India. We are also at the same time building our own customized chips. They may be CPU chips, they may be GPU chips, but customized chips which would meet the specific training and inference requirements of a given industry.

Current models are not necessarily general purpose. What's applicable on auto may not be applicable on health. What's applicable on health may not be applicable on agriculture. So, you, each industry still needs a unique model. So, we are focusing on that very heavily. And this is backed up under it by our growing semiconductor ecosystem. So, we have a large CPU based semiconductor ecosystem on which this AI layer would write very strongly. We currently, just for information, it may not be AI really related directly, but relevant to ai. We have 10 semiconductor projects, which are currently running in India. And they look at basically legacy nodes, but they are, that's legacy nodes are the ones which are used maximum in the industry including some of them are with collaboration with the us companies.

And these projects we AIM will create chip development capabilities in India and eventually support the compute needs which is what is required for the AI mission. Energy sector is the is the fifth layer which has seen remarkable growth in India on year-to-year basis. We are successfully trying to balance the goals of meeting the rising electricity demand but also promoting sustainability at the same time. Our current installed capacity is at about 475 gigawatts. We recently passed Ashanti bill essentially, which supports nuclear energy innovations, including a small modeler, reactors, a small varies 50 to 250 megawatts capacity for onsite positioning of these model of these of these power generation facilities to essentially bolster the sector because it has a very heavy energy heavy in

industry. So, across these all five layers and the domestic focus the public and private they've all fed into this.

And given the impetus to building this, each of the layers of AI infrastructure while. It has strengthened how we are approaching this particular subject. It has also helped us build very strong international partnerships, particularly with the us, both with the government and also with the private sector. We think that our strengths are complementary. The US brings cutting edge AI research in terms of the new mathematical models new technology streams within ai. So, whether it's the existing l LMS or the new branches of it advanced labs you have meta-scale level cloud providers, which is very necessary in terms of access to the compute deep venture capital ecosystem.

This marries very well with the kind of competencies that India brings to the table across the five layers that I mentioned. You can take each layer and drop a combination chart, and you'll find there is a complementary matrix that is available there. There is already, as a result of this, a very promising momentum of specific cooperation-based projects on the ground. And US AI firms are already expanding their engineering their, and their research centers in India, which is what is the cutting edge of the development of ai. As we move towards the. Delhi Summit India's message, my Prime Minister's message on the AI to the world is the future of AI must be inclusive because it has to bring that value to the society, transparent, safe, and govern through a model that is collected.

We don't, we think that the benefits of AI cannot remain limited to a few nations and a small segment of society. It really, as I said, has to be population scale. And as my Prime Minister emphasizes, technology must be democratized in both its use and deployment. Our digital successes, which I alluded to earlier I think will show us at least the way forward in the field of ai. Also, through this comprehensive five-layer architecture wide strip widespread adoption and diffusion and people first value-based impact. We in India are building an ecosystem that strengthens the nation, uplifts our entire society, and gives us template for building externalities of partnership of AI with other countries of the world. We think that the India Impact Summit 26 will be a pivotal platform to turn the vision. Into really a global collaborative action. That's what we are aiming for. I want to thank again, Brookings and each one of you for sharing your time this morning, coming and listening to this. Thank you so much.

BOGEN: Thank you so much, Mr. Ambassador for those really rich and illuminating remarks. And I think as we can all tell, there's just an immense amount to follow in the world of AI. And that's only if we're talking about the AI layer, not at least all of the other layers, but. In order for the globe, really everyone to come together and make sense of the evolution of this technology. There's a need for some alignment of information about what is even being built, what are the values that are driving different developers of the, of these tools? How are deployers incorporating these tools into real world context, automotive medicine, et cetera, as we heard. And so, in light of that need the Hiroshima AI process reporting framework was, has really been a leading effort to create a locus of that conversation to say, what is it that we actually know about this this field?

What's developing and how can we start to make sense of it and figure out what interventions are needed in a voluntary capacity, as a framework for other accountability initiatives. But as a start, so I'm Miranda Bogen. I'm the director of the AI Governance Lab at the Center for Democracy and Technology. We were excited to partner with Brookings to really dig into the hype reporting framework to understand what role has it played in actually helping to make sense of this space and where can it go from here. So just to ground the conversation, we'll get into a little bit of the research and have some great panel conversations as the event goes on.

But what is the hype reporting framework? Hype. It represents a multilateral effort to promote transparency and accountability in the development of advanced AI systems. It was launched in February 2025 as part of the G7 Hiroshima AI process and provides a mechanism for organizations who are developing AI systems to voluntary, voluntarily report on risk mitigation practices, and facilitate cross organizational comparability and to identify and disseminate good practices, obviously to promote transparency and help address the many risks that I think all of us are aware are coming with this quite advanced technology. And finally, again, facilitating the comparability, I think is a key piece here. A lot of actors in this space are disclosing information about their progress, about their research but it's quite disparate. It takes a lot of effort to pull together. And so, some sort of structured framework where information can be at least organized in a similar manner to be able to look across the field can be helpful.

And that was what motivated this effort. So, what does it, what does the framework include? We'll just run through pretty quickly. Risk identification and evaluation. What are the practices companies are adopting risk management? What are the mitigations that they're taking on to, to address those risks? What sort of reporting is included? Some of the reports that we'll talk about link out to other disclosures and transparency artifacts that companies and developers are already using, and, but simply combining them which can have its own use. What is it governance at an organizational level, what does incident management look like?

How does that kind of ladder up to other organizational governance practices? Obviously content authentication is a huge concern when it comes to generative AI and a big risk that folks are trying to manage. Included here what does research look like? Where are we moving to? What's likely to come next? And finally, how are we supporting this broader ecosystem? The human layer that the ambassador talked about. But one of the questions that comes up quite a bit is, how does this approach fit into other transparency documentation, disclosure artifacts? Are they duplicative?

Are they harmonious? This is an, this is a visualization that that helps me that there's a, there's really an ecosystem of disclosure and accountability frameworks here, starting at the very granular artifact level of documenting data sets to documenting characteristics of the models that are built on those data sets. So, model cards, things like that. Increasingly as we got more and more advanced ai, we're seeing system cards, which talk about a variety of components, not just a model, but other safeguards that are involved, other tools that might be connected to a model, really understanding the full picture of a given AI system.

And then bigger picture disclosures. And this is where we see the high reporting framework coming in. What are the organizational priorities? What are the values that are driving an organization's approach to ai? How are they making sense of all of these things rather than just looking at individual technical artifacts, but, pulling those together into a bigger picture. And of course, testing evaluation, validation verification play a role in all of these artifacts both to really add more detail, but also to demonstrate what risk mitigation practices look like and where the progress is going. Like we said, we think these are we, these are com These are complimentary efforts, but they are distinct activities.

Organization level disclosure is a different act, different activity than specific documentation about systems. They're for different audiences, they're for different purposes, but they do relate to one another. And we'll talk a little bit more about how stakeholders in the hype process understood all of these different activities. And so, I'll turn it over to my colleague Valerie, to talk about the research that we did and some of the findings.

WIRTSCHAFTER: Hi, I am Valerie. I'm a fellow here at Brookings, realizing in this slide deck we didn't provide a link or a QR code to the report, but you can find that on the Brookings website, and we can certainly promote that on this page as well for the event. So, getting into our process for analyzing some of these reports there are a couple of these. They're primarily focused on those who decided to submit. So, the 24 organizations that chose to submit, we wanted to get a little bit broader. So, we looked at the published submissions, of course, trying to compare across them to see how each organization approached the reporting framework. We conducted an open-ended survey of the AI community.

This was really designed to look at people who could have submitted, to understand awareness about the framework to see where the benefits and utilization lie in, in this process. And then where we found gaps still. We conducted some supplemental interviews with targeted stakeholders to just get a bit of a fuller picture of the entire ecosystem and where this really fits in. So, for the reports, we found 24 submissions from eight countries, Japan and the US were dominant which makes some sense. Of course, report lengths varied from nine to 60 pages, so there was quite a diversity in how people chose to fill these out. One third looked linked primarily out to external documentation.

We'll talk a little bit about some of the challenges with that, but there was a lot of linking to external references, especially for Frontier Labs that may have a little bit more robust systems in place for the survey. We received 110 responses. This was just a convenient sample, so nothing in here is representative we'll say. But it does give us a sense of what the broader community feels is important about the framework. Where some of the challenges might be where the benefits lie. The organizations that they represented.

So, it was primarily researchers and public interest actors, though we did have some representation from deployers developers and policy makers. Questions were focused on a variety of things related to familiarity with hype. Some of the logistical aspects of submitting reports, incentives for completion perceived impact all different types of things to get a bit of a sense of what people saw the value in and where the benefits were and where some of the challenges arose in the process as well. So, between eight and 20 questions open-ended people responded to based on that, their familiarity and awareness of the hype process. We know, and this was something that came through in the actual hype reporting framework process. Open-ended questions are hard. But we really didn't want to, because we weren't looking for a representative sample.

We didn't wanna lock people into some of our preconceived notions either. So, what did we find? One of the main things about hype is that it's very flexible, right? And so, it encourages participation from different types of organizations because of that built in flexibility. That's an asset, but it's also an obstacle the flexibility limits comparability. Which was one of the main benefits that we heard from a lot of people about the impact of these reporting the reporting framework. Because of that flexibility, there's an increased time commitment. Potentially some people will give up in the process because they're unsure of how to respond to something, how to measure something.

Especially for companies that maybe have less well-developed risk mitigation practices. It can be a bottleneck or a barrier to actually even beginning to fill out the report. There was initial feedback to the OECD about uncertainty from whether to fill it out from organizational practices, despite what Miranda was mentioning previously, or specific AI systems, whether to write from a developer deployer perspective. This uncertainty was something that we saw quite a bit when looking across reports. Eight wrote from a developer perspective, primarily six from a deployer perspective. Seven kind of blended. The two and three wrote from a different approach. Same thing with the organizational perspective.

We found that, oh, which slide am I on? Oh eight. Did I go there? Okay. There we go. We found that 18 took a, oh, I skipped eight. Wrote from a developer perspective ooh, I skipped this one. I'm sorry. Company-wide approaches 13 organizations. 10 chose to do a sort of hybrid fusion between the organizational approach and a model specific approach. And then one was more model specific. So,

another thing that we heard quite a bit as well was that small and medium enterprises service providers had some capacity and information challenges when com co completing report submissions. Most of the reports were submitted by larger organizations with 250 or more employees, but six organizations with fewer than that did submit reports.

They faced different obstacles. We heard this in surveys, in our interviews, and we saw it even in looking at the reports. Capacity constraints, motivation to submit to show that this is a value proposition lack of policy knowledge were all things that came up quite a bit. So, the framework, of course, has this built in flexibility. But we found also that responses, while informative still remained quite high level and difficult to verify. Some of this is of course the nature of the questions. They often ask things like, does your organization do x? Without eliciting more information on what organizations specifically do.

But some of it was actually in some ways part of built into the process organizations with maybe more capacity to respond that faced potential public scrutiny. Chose to provide less granular responses, chose to link out to curated documentation that has some instability risks. Of course, those links could die. They could disappear. They could change. So, it makes it hard also to do mapping over time to compare how things have evolved in this space. Smaller companies on the flip side, use some of these reports to demonstrate capacity to build credibility, to showcase compliance.

And so there, there were different approaches especially based on the size of the organization. The framework does provide significant value. That was quite clear connection to the G seven, G 20, the global reach of it, the multi-stakeholder development process. The openness to, to submissions efforts to complement existing frameworks were all things that we heard were really valuable. But we also found real limitations in knowledge about the framework. Of course, reminding that the surveys not representative, but even amongst the AI community that responded, only half were familiar with hype as a mechanism for this type of transparency effort. So limited knowledge really of the framework.

And thinking about better awareness of processes could both encourage more participation, but also more impact of the framework as a whole. Finally, and really quickly, we found that, of course,

benefits were really different based on who was participating and what their vantage point was. In the AI ecosystem, developers were using it to identify gaps. Align teams internally demonstrate trust, potentially shape the future of governance frameworks externally. Deployers, were thinking about how to clarify risks. SMEs were using it to signal maturity. Enterprise customers were trying to understand risk handling incidents. Researchers in civil society found some value and understanding current practices, methods and gaps. Policy makers maybe look to these as well as a way to potentially shape future governance.

So, building on these findings really quickly, I'll just run through some of our recommendations. We roughly bucketed these into to four groups on quality and consistency in reporting, clarity of purpose the usability of the submissions and incentives for broader participation and awareness. So, on the reports themselves, thinking about especially that balance between flexibility and uniformity, the open-ended flexible reporting is really valuable to be able to get some of these broader insights, but also it poses a lot of challenges and limitations. And so, I think we suggested revisiting some of this balance, keeping some of the open-ended components, but also potentially more structured questions as a way to identify trends over time, emerging consensus more of a return on investment potentially for submitters, where open-ended questions are included.

More explicit guidance about what makes a good response could be quite useful. Encouraging respondents to answer directly in the framework in the, in their responses rather than linking out is also very valuable due to those stability concerns and the challenges with comparing over time. And then thinking about enhancing transparency around what has been audited what findings emerged, especially where organizations may be detailing practices, but what has been implemented and what has been verified by independent auditors and where that shows up in the framework as well would be beneficial too.

As we mentioned, the framework has many different goals depending on the actors involved. So, figuring out some clarity of purpose for the framework as a whole, I think would be really valuable. Though, having that different ecosystem participate is a very good thing. But there is some ambiguity about how this fits in with the broader AI governance ecosystem which can be further clarified. Different audiences for reports may be also interested in different things. So, clarifying potentially the

target audiences, allowing report submitters to select their target audiences because that will also filter into the level of detail that they're providing and the technical sophistication necessary too.

The framework is an. A step toward accountability. But that can be strengthened if the theory of change or thinking about how reporting actually leads to intended impacts is further developed. And companies are de detail are encouraged to detail how reporting the reporting process actually has changed their practices over time. Comparisons across submissions came up regularly in our conversations and in inter and in survey responses as one of the clearest public value, clearest values of these public facing reports both within companies and for potential customers. But the process for comparison, I don't know if anybody has tried to do it is very hard.

You either have to open up individual tabs for every report and kind of click through them or download individual reports. And so, thinking about a way that allows that comparison to be more sophisticated, filtering by actors, by size, all these types of things intended audience would also be useful. Making the framework iterative changing with developments in markets, policies and feedback while keeping some core stable base as well. And then finally, strengthening. Thinking about ways to strengthen incentives for participation, awareness and adoption. Those incentives, 17 organizations participated in the development of the framework.

Less than half of them. Actually, submitted reports. So, ways to build this community to expand and to demonstrate the value add whether this is actually entering into governance conversations and where that's happening would be a way to demonstrate what that return on investment per submission would be. And then of course, continuing to enlist stakeholders to participate through the various groups that are part of this hype ecosystem. Expanding eligibility criteria, potentially building in a hype light type of framework for smaller companies that maybe have less capacity to respond or where their capabilities are limited. And that is all.

KERRY: Welcome everybody. I am Cam Kerry. I am the Ann and Andrew Tisch distinguished visiting fellow at the Center for Technology and Innovation here at Brookings. For the past six years, I've co-led a project that we call the Forum for Cooperation on ai in partnership with Brussels think tank the Center for European Policy Studies, which has brought together experts in ai from all sectors. With

the several governments, seven governments including the us the European Union and Japan, among others. And I am very grateful to work today, both with the government of India and thank the US mission. And with the center for democracy and technology and a longtime leader in tech policy.

I, as an official in the Obama administration as well as in my work at Brookings, CDT has been an indispensable partner. And so great to partner on this event and on this report. Ambassador Kwatra talked about the variety of strategies of initiatives of, multilateral groupings in the world of AI policy. And a question that often comes up, and it's a question that recurred several times in some of the questions that we received for this event in advance is how do all of these things fit together? And is there fragmentation or things headed in different directions? Do we need some kind of unified governance?

And in our project we have articulated the view that look, the AI is built on the internet, a network of networks around the world. And that we need to replicate that architecture for AI governance, networked architecture with many different nodes, different pathways and the hype. Process is an example of that. We've seen a series of AI principles migrate across networks from, the AI principles from the OECD which now plays a role in the Hiroshima process, the work of the G seven in articulating the Hiroshima process in proliferating it around the world.

And the AI India summit is part of that and part of an iterative process in developing and understanding norms around ai. We are delighted today to have an international panel. Because it's international, everybody is coming remotely. One of our participants was supposed to be here in person Keita Azuma from Japan. But his flight was delayed. We are glad that you can join us on route. I'm gonna begin with Keita because coming from Japan, that is where the Hiroshima process originated. And Japan has played an important leadership role in AI for more than 10 years.

The G7 started the process that led to the OECD AI principles under the Japanese presidency. And the Hiroshima principles similarly emerged under a Japanese presidency. Kaita welcome. I am, I'm delighted that you can join us today from, I guess en route or wherever you are. So, I understand that, so Keita is the director of international affairs at the Japanese AI Safety Institute. So, Keita, if you could tell a little bit about the work of the AI Safety Institutes and the network of AI Institutes and how

that relates to the Hiroshima process. And I understand that you recently met with the Hiroshima global Forum and ai safety Institute. Where do you see that leading in terms of both the work of the ACS and of the Russia network?

AZUMA: Thank you very much. Thank you very much for kind introduction and I would like to express my sense appreciation to the Brookings Institution and the Center for Democracy and Technology for convening this timely and important discussion. And Japanese Safety Institute views will AI governance not as abstract regulatory exercise, but as a practical history shaped by various associated challenges as a regulatory Asian society facing severe labor shortage. Japan sees AI as essential to sustaining core economic vitality and social services, while recognizing the need to manage its risks, responsibility, following extensive domestic and international discussion on a governance Japan Safety Institute enact and Canada Office enact as the Japan AI Act in June last year. And just last month adopted a new AI PC plan.

So together these frameworks aim to make Japan one of the world's most innovation friendly environment for AI development and deployment, while ensuring safety, security, and trustworthiness. The core features of the hir process, stakeholders engagement, openness, transparency, and interoperability are closely aligned with this policy directions in AI safety, student cabinet, office of Japanese government, and help from a foundational pillar of Japan's national strategy. Thank you.

KERRY: Thank you. Let me turn now to Audrey Plonk. I've mentioned the OECD and its role in, in AI. Audrey the deputy director of the Science Technology and Innovation Directorate at OECD which has done the work on ai. And the OECD really played a central role in promulgating the hype reporting process. And now in implementing that. So, Audrey, let me put to you the question that's the topic of this panel. What's next for the Hiroshima reporting process? We're getting some, we're getting some echo. Audrey...

PLONK: okay. It is great. I will try next. Okay. Just to say that just wanna thank Brookings and CDT for the work and the partnership and the research and the survey and the community that's building around the hype reporting framework. We're really excited that in less than a year we've seen so much support as well as progress toward the objectives. I guess just in terms of where we are, you

heard already from the report that there was 24. We have now 25 companies that have reported most recently two more Japanese companies, a Danish company and Indian companies. So, we see the community really growing. We also have 17 reports in process. So, we're expecting, another

KERRY: Audrey, we seem to have lost you. We had a blast of static there... no sound

PLONK: Do you have sound?

KERRY: Yeah. Okay. Now I hear you.

PLONK: I'm so sorry. I don't know what you caught but I was just saying that we have. 25 reports now and we have quite a growing a diverse community and we have 17 reports in the pipeline that should be coming. I don't know when, but that's exciting and we'll hope, hopefully enrich in the. The data that's out there for us to look at in terms of what's next, I would group it into two big buckets. The first is maybe three big buckets actually. The first is to really continue to push on the participation and the existing reporting framework. As you see we're pretty happy with where we've gotten in a year, but there's a lot more that we can do, and that's why the work that with Brookings and CDT and others is so important to help get out the word.

You heard that? Stats from the survey that, there's still a lot of the community that doesn't know about it. And we're open for ideas of how we can continue to push it. But that's one, one big objective for the future is to build that community, to grow it beyond where it is today. The second is to continue to update the framework and keep it relevant. And in that regard, we have some two exciting objectives. The first is to link the reporting framework to our database of tools for trustworthy ai. That's a database that includes a set of educational and technical tools that companies use.

They're reported into the database. I think there's now more than 750 different tools. And so, what we're hoping to do is to be able to tie the reports to that database so that we can start to look at what tools are companies using and their risk management processes. And then, we can start to build out a more detailed picture of what's happening. And so that development work is underway with regard to the database itself, but also with regard to the framework. And then the second piece of the tying it to other tools and data sources that we have access to at the OECD is to tie it to our incidents work.

So, we have an AI incidents monitor, which was really looking at press reported hazards and potential incidents. And so, we would like to tie that that database to report so that we can start to see if there are patterns in in activities over time. And so those are two elements of the future for the framework. And then I think it was already mentioned, but we are really. Recognize that the community, when we negotiated the hype itself in the G seven presidency of Japan, it was only a couple years ago, but it was a fairly different ecosystem, or at least the ecosystem was still quite emerging.

And we'd like the framework to be able to target a more diverse set of actors. And I think it was pointed out in the presentation from Brookings that, we have large companies, but we also wanna attract small companies and not just developers, but also Deployers. And so, we're thinking about adjusting and having maybe a lighter version that the working on a lighter version that is maybe more attractive to deployers and smaller companies that are not major, big model developers, but that could use the framework and want to use the framework. But are maybe, maybe the whole original one that was really focused on, developers of advanced AI systems could be adapted further. And then and then finally just in the, a normal course of maintaining things, we obviously wanna keep the framework streamlined and as up to date as possible with the developments.

And that's always, a bit of a chicken and egg problem because the, everything moves so fast. But that's work that we're doing now and really excited to be here and happy to hear thoughts and feedback of how we can how we can make it more useful and improve it. So, thank you so much, cam. And I, sorry, apologies for my audio issues. I. Moved offices yesterday to a new office and apparently not everything's working yet. So, my apologies.

KERRY: Yeah. One quick follow up question. Any timetable for sort of the next iteration of the reporting framework?

PLONK: Yeah, so we, we hope to be able to talk about some of it, at least in in India in a few weeks. I don't I don't wanna put a deadline on when we can absolutely launch it. But I think we might be able to show some of the progress there and so I think we're moving quickly. Obviously there's a few potential moments in the spring where we could we could make it a public. So, there's the team's working really hard heads down to, to move it forward as, as quickly as possible.

KERRY: Thank you. So let me turn to the other screen in this discussion which is the road to the India Summit. And we have participating Shatakatru Sahu, who's associate director at the US India Foundation. And I understand Shatakatru that, to follow up. I don't know whether you heard Ambassador Kwatra's description of the India Mission, the sutras, the chakras but I understand that now issued guidelines as well, both for the public sector and to some extent the private sector as well. Tell us a little bit about some of that process the action items and what's ahead.

SAHU: Thank you, Cam. And thank you to Brooking for organizing this, and it's great to be here with my panelist data. And Audrey yes, I did hear Ambassador Kwatra very nicely laying out what the India AI Impact Summit is going to be about. And a key feature will be the governance approach. So, in November 2025, when they release its AI governance guidelines, which are voluntary and recommending nature, and it is a result of about 12 months of consultation with industry, were within India. And a previous draft was released in January 2025 also.

I'll just briefly lay out what that governance guideline is. Firstly, it's a function of India positioning itself. As a country looking to highlight, accelerate, and prioritize innovation over anything else. We don't wanna miss the bus in terms of AI development or AI adoption and AI deployment. And we do realize very realistically that we are not a US or China in terms of capacities of AI infrastructure, computer chips, et cetera. So, while that is the reality, the approach is to maximize the impact of AI across sectors, especially those related to welfare. So having said that, what the key takeaways of this governance guidelines are that India in no time in the near future, we'll not be introducing any new AI specific law.

It'll rely on existing laws to counter risks and harms. And it's going to support voluntary measures. While it does that, if there is any specific harm, which is not being addressed by the existing laws, then India will plug in targeted amendments and targeted amendments with the existing laws. What it would also do is, in the meantime, try to understand the AI value chain better. It would try to understand harms in the Indian local context better. To address those risks and develop a risk framework, which is unique to India in its own local context. So that's what in a nutshell the governance guidelines are about.

It is pro innovation, voluntary and reliant on existing laws while trying to understand the AI value chain better and the harms in the local context better. And this comes it ties in really well with the AI Impact Summit on February 19 20th, because what the summit is, and I'll be repeating what ambassador I said, it's a, it's about increasing access to AI and in ensuring there's impact of AI on ground. And that's important because we wanna see it on ground in healthcare, in education and other critical sectors. So, when this happens, I think I hope OECD and bookings release the report and the, at the summit, because India is in a position where it wants to adopt voluntary guidelines, but it's not at that advanced stage where the high reporting is.

So, I think India would benefit from the findings of this report, which is released by Brookings. And there are two, three recommendations from the report, which I would like to highlight. One is increasing participation of companies to adopt this framework. I think that will work well in India. You at this point, it'll be critical to increase participation of companies within India, which are basically. International firms operating in India and also domestic ones. If they should take up this voluntary framework and report on it, I think that will be a good starting point. Second will be incentivizing them to do right now there are many existing laws, as I mentioned, and also new targeted laws coming in towards which the industry slightly apprehensive. If not with the principle of the law, then definitely with the operation of the law. Just as an example, India released his report on copyright engineered ai, where it is suggested a hybrid model where it is suggested on a community basis that everything online is available for AI developers to train their models on unless it's behind the pay.

However, once they access it retrospective manner, they need to compensate the copyright holders. So that's something which Jay’s experimenting with experimenting within terms of incidents and AI issues which are not, properly dealt with existing laws. So, it's a recommendation to amen India's copyright lock so that's being done. And I think incentives will be important because otherwise companies won't hop onto the reporting mechanism. Some incentives have worked well in the past for voluntary self-regulation in India is basically, let's say, winning government contracts. If you wanna win government commercial contracts, then if you can showcase voluntary commitment, that helps your case better.

And I think another suggestion from the hype report, which I read, is being, it, I think for a country like India where the understanding of the AI value chain isn't so great I think being it in that framework would be important. And I would just conclude by just thought I had when value chain Miranda were presenting is that if any of the reporting by the 24 25 company applies to India, I think that should be showcased. Because the application of AI development and AI deployment is cross-jurisdictional it's across borders. So that could be a good starting point. Of course. The companies would need to accept and consent to that being publicized in another jurisdiction. I don't know how to approach operationalize that, but that could be a good starting point. I'll stop there, but happy to take any more questions and looking forward to discussing more.

KERRY: Thank you. So, we are gonna take questions momentarily. I will distill one question for the panel coming from some of what we've received by way of questions online. But so, if I'll come to you in a moment. Shatakatru, I was encouraged to hear you talk about, the opportunity to use the hype reporting framework in India. And it, that I mean I think one of the great advantages of this and of this process is to fill in some of the blanks on principles of transparency, of accountability that are recurring principles in AI frameworks and part of the India guidelines as well. So, helps to that flesh out the meaning of those principles.

So let me put to all of our panelists how do you see the process of promoting further adoption of the AI principles we've heard in the presentation from Valerie and Miranda as well as in questions online about, the applications SHA grant you talked about, the contextual issues of ai. So how do we adapt this and diffuse it to across principles and ecosystems around the world? Keita, you want to start on that? I know. But Japan has worked with developing a Friends of Hiroshima network. And it was that a channel to spread the reporting framework.

AZUMA: Yes. Thank you very much. Regarding the significance and the current status of future process, though, I would like to highlight three points. The first, we need shared global references that policy makers and developers can rely on. The high reporting frameworks helps field this gap by offering a common point orientation. And second, the first three reporting cycle has revealed clear areas of improvement, including greater clarity and consistency in technology to enhance capability across airport.

Third, the hype health value well beyond countries with advanced AI development capabilities. So, for countries without extensive domestic AI ecosystems, the frameworks, hype frameworks can solve other principal guide for safe adoption, helping to prevent post regulatory fragmentation and the widening of the digital device. So, looking ahead, the similar priorities stand out that we should further clarify definitions to make report easier for companies and more meaningful for the general public so that market trust itself become incentive for transparency in global south countries as well. So, we should also continue to scale the hype beyond the G seven, not as a model of rural exports, but as a usable and adaptable references for the rest of the world, including the global cells.

The Japan established safety needs to the following, the United States and United Kingdom, and we firmly believe that air safety and security and governance are not a trade off with innovation or competitiveness. On contrary, interoperability among advanced AI countries coupled with cooperation with Global Source is essential to achieving shared goals, greater transparency, improve comparability and sufficient flexibility for diverse actors. So, when Japan, Japanese government has worked closely with a summit host by the UK, Korea, France, and we strongly support the upcoming India Impact Summit. So, achieving safe and trustworthy ai while accelerating innovation requires deeper international cooperation and reverse each country. So, we really expect the leadership of India. Thank you.

KERRY: Thanks. Audrey, I know you talked a little bit about wanting to get wider participation. Can you say a little bit more about that?

PLONK: Yeah, and I'm glad that you brought up the Hype Friends group because we'll have a meeting in Tokyo in mid-March, and I think it's one of the, it's one of the tracks that we're. We're participating in order to expand participation in in the reporting framework. But also I think, we've linked somewhat intentionally, and I hope strategically these things together of the OEC AI principles, the global partnership in ai, and when the bringing the HYPE framework into the OECD and the global partnership is, also there's a strategy to try to, as we expand the membership and the participation in the, what we call the G pay, that also gives countries access to the hype so that there's a symbiotic and hopefully mutually reinforcing relationship between the, the attractiveness of participating at the

OECD and the GPE, as well as participating in the hype and that becomes a group of countries that is.

Sophisticated, but also and we're emerging in the way that everybody is and willing to come to the table to work in this, in the same way with an industry that will that will participate. And so, I think the hype friends group, which you mentioned Cam and Kai too is providing access to a lot of countries that are interested in and they're, and are emerging and may not otherwise necessarily know how to engage with us over here in Paris to hopefully, not just bring them into the framework, but more largely bring them to the table at the OECD. In addition, of course we try to have as many events as we can. We're not an event management place, we do we do analytical work here mostly, but we are trying to piggyback on the back of, big events happening. And thanks to India for their, willingness to work with us and host us that, their hosting and council meeting of the GP at the summit in a few weeks.

But also, we hope to have a site event there on the with Infosys and others on the reporting framework itself. So, we're trying to write about it to partner with NGOs and others like yourselves, to get the word out and to continue to work with the private sector because their participation and their cooperation is critical. And as always, and I know I always say it, but we're always. Open to, to feedback and new ideas of ways that we can continue to push it to push it forward. We, the holy GRA for us would be also to get our countries to include it in their policy approaches. Japan has done a little bit of that, but if our countries can think about if they're having voluntary commitments or they're having voluntary actions to use this framework instead of creating new things, we'll continue to try to, push on that door with both our member countries, but also other countries so that we can try to get as much, coherence and reduce the burden on organizations that want to share their experiences as possible. And in our context, we'll con we continue to do that and we've seen some, positive signs of how governments might think of using it in the future.

KERRY: Thanks, Audrey. I think you've answered or at least partially the, a question I had for Shata, which is, can this play a role? Can diffusion of hype reporting play a role at the India Summit? Shata, anything you want to add on that and on ways that India can help advance this fall?

SAHU: Of course. Thanks, Cam. I have a slightly different take to it. To it. I think the lens has to be changed and just being bluntly honest, the lens has to be changed from getting high adoption increased in other countries and just find the low, lowest common denominator in the countries you're going to, which is transparency and accountability. But I think it has to be more than just a presence at the India AIPAC Summit. And there needs to be some homework and groundwork, which I think OECD bookings all of us need to do because it'll be difficult to convince counties to adopt a framework. And just picking up from Audrey's point on just pushing the doors on the voluntary frameworks already existing.

So, I think there should be good effort done to understand what does that mean in the Indian context. And that translates to, now I'll give an opinion, is that within the AI governance guidelines by India, it has designated the AI Safety Institute, which India announced about a year back that would be the nodal agency to, so to say. So, it can lean on the nodal agency, which India has appointed for transparency and accounting work accountability work. And it has also highlighted how sectoral regulators need to take a play at it. So, if you're talking about high deporting, then for India. Maybe you need to slice in. Start it in terms of sectors.

For example, in the financial sector, we are, central Bank is the lead. This is our Bank of India, and there are some couple of other financial institutions also. So, they would have their repository of knowledge in terms of what AI developers and deployers are doing with what kind of data, et cetera. Similarly, for health, that's what India's trying to do, not create any new law, stay away from new codes as much as possible. So, in health, you have the National Health Authority, you have the Ministry of Health, so on and so forth. I think visibility and marking the moment at the India AI Impact Summit, it would be important.

But in the longer term, I think some more work needs to be done with the Indian stakeholders, especially those in the government for more cohesion between the high framework and what India is doing. And I'm saying this bluntly because having observed the Indian ecosystem over the past two years in terms of governance, I have not come across hype being marked as the go to volunteer framework. I, I think there is some distance to be traveled for that. But having the idea report, it seems like it's a great one for India. So, I think more socialization, more integration with the Indian

stakeholders and understanding where India is in its governance journey and understanding India's sensitivities towards prioritizing innovation over anything else would be helpful. That's just my take in terms of how to approach it, but that's just my take. I'll just say that.

KERRY: We have about one or two minutes for questions. I know we have one over here. Any others? So let me hear, both of those questions and I'll ask and then I'll ask the panelists to, to respond to the two questions to together quickly and wrap up.

AUDIENCE MEMBER: Hello. So, I had a question about the India models. The ambassador presented four models and also said they were grounded in the billion-person ID of India, which basically connects all the data sets. The fourth model he mentioned, which had the five layers from action apps to the AI models, to the data center infrastructure. So, in ai to the chips, to the energy. I haven't seen that model published before in India. All the other models are actually only in the action side. So, I was wondering if Shata could give a reference to where the five-layer model is. And the other question was in presenting the height report, you said you're drawn on...

KERRY: You get one question…

AUDIENCE MEMBER: You're drawn…

KERRY: Another, a second question on the -- sorry to cut you off, but we are running up on time. Question over here.

AUDIENCE MEMBER: This is more of a reaction. It might be a little bit more of an outside of the scope of the group, but thinking about global, systemic exogenous risk. And also, in a quote from the ambassador when he mentions about removing the bias out of some of the models, specifically we're talking about internet size, large language models or auto type of models, they use it that much data.

The removal of bias also adds a little bit of the subject ability of what can that be given that the who and the what can be interpret differently by different actors. And while thinking that the decision or utilization is gonna be done really in the market by who and which models are used. And already seeing some actors, such as one of the war richest person trying to modify a large model to have a

particular way of thinking. And as we are moving as a Canadian prime minister, Karen mentioned, to a new multiport war, more fragmented and a new reality. How are we to approach those type of sensitivities when we can have a new sense of heuristics developing a globally?

KERRY: It looks like we have lost two of our panelists. Shatakratu, you're on your own. I know the first, oh, Audrey, you're back. Okay. The first question certainly is one for you Audrey, the second, risk is certainly something that, that the OECD has focused on a lot. So, if you wanna jump in on that you're welcome to do that.

PLONK: I guess I'll just say that we, the, that the framework itself tries to, to address risk issues. And we recognize that that's one. One path. I think there's a lot of work going into the risk space. We're also trying to balance that out with the, with the, also the opportunity space and where, how do we make this technology useful and whatnot. I guess on the other risk piece I would say we're trying to develop complimentary products. I talked about the incidents work a little bit. I won't get into it deeply. But I think, as this technology grows, we need. We need more understanding of how the risks are playing out in real time and in real situations.

And so, we've tried to get ahead of that by starting to build methodologies for measuring certain things. It's imperfect, it's early days. But I think you see that that, that is, is, we take that really seriously. I don't know that we're gonna get consensus globally on what like governance looks like. I don't mean to be negative or pessimistic, but I think that's why some of these, voluntary procedures and listening to what other countries are doing are so important because we want to try to coalesce as much as possible and understanding the risks. We have a taxonomy of risks. We have a classification of risks. We have a lot of work that we've put into trying to harmonize how we can categorize risks and think about them.

And in line with things like, the NIST work and other standardization work around this. So, I don't wanna go on at length, 'cause I know we're out of time, but just to say I think it's a big important space that we need to not lose sight of in light of the global dynamics, but that, I do think there's, it's a top, it's, it remains a top priority for us.

KERRY: So, Keita risk is certainly what the ACS are focused on. Any responses do you wanna give?

AZUMA: Let me add one thing. Thank you very much for that. Great question. And then the hyperscale view of the G7's power, and actually I'd like to highlight that. The one very important things like the iterability and arrangement is policy coordination. That's very important. That's essential among national frameworks in international initiatives and emerging technology standards for countries beyond the G7, including those in global cells. The policy offers a shared reference point versus a prospective regulatory model. So many countries are adopting AI technologies without having extensive domestic AI development capacity or comprehensive legal frameworks. In this context, Hiroshima process provides practical guidance on risk mitigation and transparency that can support safe and the responsible AI adaption. Thank you.

KERRY: Thank you very much. Shata, you get the last word, the fifth mark?

SAHU: Yeah, I think Ambassador Kwatra was definitely to the components of AI as such a, not a particular LLM model or an SLM model, India trying to build. He was just trying to break down that, the five components of power models, applications compute and maybe data centers. He was doing. He was, I think, explaining that what he was telling in essence was that India is trying to build its own sovereign large language models, SLMS, et cetera.

And I would just like to inform the room that the Indian government has subsidized 12 startups. Has accorded 12 startups with subsidized compute one of the cheapest rates per hour to access compute. So that's what the government is doing helping Indian select, Indian startups with affordable compute to train their models on. And not all models are trained on all of the data that's available. Of course that's subject to respecting privacy of the citizens. And there is a public data sets platform also. It's called AI Co and it's up on the it ministry's website as well. So that's what we're trying to do. And if you zoom out from that also outside of these 12 focused startup trying to build sovereign models, India is really just trying to increase impact by using all open-source models out there.

Also using models of the us mostly and focusing on the influencing layer on the application layer. So, the end goal is access to resources impact on ground. And somewhere in between is your objective

of building sovereign LLM models because you wanna have your sovereign capabilities in the longer run. So that is what Ambassador was referring to. In terms of what are the details of these startups and what will get announced at the summit. I think we'll just have to wait for the summit and see. Thank you.

KERRY: Thank you very much. Thank you to all of our panelists. We will now take a 10-minute break.

BOGEN: Gonna give them, we'll get started in just a minute so folks could start taking their seats. All right. We'll get started in just a moment. All right. Thank you everyone for sticking around. There's obviously so much to talk about in these topics. Thank you. So the last panel left off on, I think, a really interesting note, which was there are a lot of different choices developers are making around how they're building these models, how they're handling risk which is information that's useful for a lot of different folks to have, to be able to really compare, what are the choices that one, the one provider is making vis-a-vis vi bias, for instance, what are others how are others handling risks throughout the supply chain, throughout the stack?

And I think we're seeing a lot of examples around how providers are communicating those thought processes, including the hype process. One interesting new example was just I think yesterday or perhaps the before anthropic just released a new version of its constitution which in a previous area you might think of as platform policies, but really the guidance on how they expect their system to behave and some of the guardrails that that are trying to be built in. So, we're really seeing a lot of different efforts here, but I wanted to start with some, excellent panelists here specifically David from Google and Amanda from Microsoft, respondents to the hype reporting framework to really understand how do your companies think about participating in efforts like these.

Also how are, how is the hype effort situated amongst other efforts you might have model documentation, six system documentation, general comment, general communications about the work that you're doing what role does it play within the company? And I'll start with David since we're have a pleasure to have him here in person, and then we'll turn to remote participants.

WELLER: Great. Thanks. Thanks for having me. And thanks to Brookings and CDT for convening this conversation. I think we are all struggling a little bit in, in the dark or at least the fog over the past few years about how to bring alignment around a very complex and heterogeneous landscape. And maybe just to take a step back and to state the very obvious, we are living in a time of intense change across, I think, three vectors, which are relevant to this whole conversation. One, of course, the technology and the incredibly rapid advances in AI over the past several years.

Secondly with respect to international coordination rules of the road, to what extent there should be rules of the road internationally and what those should look like. And then thirdly. How this rapidly evolving technology should be regulated. And we see debates going on in across many different countries about more of a precautionary approach, more of an innovation-oriented approach. And how governments are thinking about these issues are all in flux. So, three areas where we have a tremendous amount of flux. And one could be tempted to just throw your hands up and I heard, AAU Audrey say before from the OECD, they're saying let's keep our expectations in check around governance and fair enough.

But I really give the OECD and the other governments who have participated in that process a lot of kudos for not simply throwing up their hands. And I think from an industry perspective, these efforts are super important because. In this time of flux and change, we do see many governments at the national level and in the US sometimes at the state level moving in to develop their own either soft or hard law regulation. Often getting at the same ultimate objectives but pursuing very different paths to that. And this is commonly called the problem of fragmentation. And of course, again, going back to the theme of the India ai impact summit, which is this is all about, this technology is all ultimately about deployment.

And that's going to be how we're going to harness the incredible benefits of it. If you have this very fragmented regulatory landscape that is, is a fundamental impediment to deployment. And then secondly, in terms of the goals of. Ensuring responsible development and deployment of this technology. When you have that kind of fragmented landscape, that's a real problem. So, I think the hype, and again, also commend the Japanese government for kind of leading that process and stepping into the breach and recognizing all of those uncertainties that I mentioned at the beginning

and trying to develop a code. So, I know that's a bit broad, but I think that's why this effort is worth the candle and why. Certainly, we at Google have invested in the process and supported it.

BOGEN: People talk a lot about transparency artifacts as boundary objects, which basically can facilitate the types of conversations that you were talking about to even say what are we collectively talking about and what are areas of divergence that maybe we want to deliberate about and guide progress forward. But I know the same can be the case inside of companies. And Amanda, we've talked about how participating in processes like these can really be a good forcing function for companies to think about their own values and processes and what matters. And so, can you talk about your experience, Microsoft's experience working through this process or, any others that involve collating, combining refining organizational risk management processes?

CRAIG: Yeah, thanks Miranda. And I, that's a really helpful tee up because I think the thing that I wanted to add in terms of. David's framing is absolutely, we have this heterogeneous and dynamic landscape and we need alignment. And the other thing that we need is emerging like consensus best practices on how to do this risk management work well. And so, your question I think is going very much in that direction. Of course, our organization many other organizations are doing a lot of work internally to evolve our, best practices for managing risk in this space. And that's important. And for the ecosystem, it's going to have limited value if it doesn't get out into the universe.

And if we don't work together across industry and across the multi-stakeholder ecosystem to really drive sort of common expectations for what those best practices are. And we really see this as something that is very connected to the, to where the AI governance conversation is today, because we certainly have the alignment challenges that have already been referenced, but we also do have this challenge where in a very helpful manner to some extent, the AI governance conversation has really been quite active and defining sort of high-level norms and expectations. And there's this real gap, right? And the kind of a consensus best practices about how to implement those. And so, this is really where we see the. Reporting framework the Hiroshima AI process reporting framework, playing a role and across industry, the work that, that we are doing that you mentioned to figure out what we think works at best in practice, what others are doing.

And. Surfacing that and really working on trying to figure out, where there is emerging consensus. So, if we think about, the reporting framework, addressing the gap of trying to get to the next level of detail beyond the kind of high-level norms and ambitions that have been defined. And really thinking about, okay, what are the operational practices to, to do that work? And then thinking about how also it can bridge between the other pieces of that puzzle that we have also worked through other sort of pathways. I, you talked earlier today Miranda, and the report acknowledges there's all this other kind of work that's happened around, product specific documentation, for example.

We have other pieces of the puzzle. And so, the other thing that I think is really important that we really valued about the Hiroshima AI process reporting framework is that it is very purposefully trying to bridge across those existing practices and provide connectivity across them. And I think this also relates to Cam's point about how they see the governance ecosystem being a bit of a network of networks and you need that kind of connectivity. And so, it, it was really important that from the get-go OECD facilitated a process that was about seeing this as a tool for interoperability and seeing the opportunity to linked to existing best practices for risk management, even structurally with the reporting framework.

But then also seeing it this reporting framework as something that does need to continue to be iterative because the ecosystem is going to evolve. As I said, we're in the early days of defining best practices, especially as an industry with that kind of like consensus best practices. And so, we're gonna need the reporting framework to evolve just for that purpose. But also, it's a very dynamic technology ecosystem, so you're gonna need the reporting framework to evolve, to continue to drive alignment.

BOGEN: In the research I think we saw that the disclosures from different companies, we, there was a real mix of practices that were clear, were adopted, whole SCA wholesale and were really embedded in an organization with research that was emerging and cutting edge and maybe pilots and things like that. And I just think that speaks to exactly this point. If this is so dynamic it is it is hard to parse sometimes in those documents, like what is an adopted best practice and what's an emerging one? But I think, hopefully the evolution over time through multiple iterations will help, demonstrate

which practices are maturing and which are getting adopted and which were, maybe helpful at one time but maybe no longer.

I'll turn to Emily. Who, has worked across industry in a number of roles, has, has been involved. You've been involved in some of the thinking around what good documentation of AI systems looks like. Can you give us your perspective just across the field? How are different institutions thinking about documentation, disclosure of information, how does this fit into the broader governance ecosystem? What in, what's the theory of change behind these efforts and do we see that bearing out?

MCREYNOLDS: Yeah, thanks Miranda. And thanks all for this great discussion. I think there's a lot of opportunity here. When we start getting into theory of change, that's really the big picture. Having worked on documentation for a long time now dating back to papers in 2018, figuring out what it looks like to provide the information that different audiences need to be able to understand what it is that is going on inside these systems, how they're being built. I think a really important perspective that I've been focusing on lately.

Having worked across multiple companies, multiple types of companies, and done a lot of research into how people understand the technology and the building of it. For me, one of the things that this report does really well is understand that there is a distinction between builders and deployers so that there's almost two big pictures we need to pay attention to here because those incentives are different. And I think a lot about the incentives when we're talking about this kind of governance and framework and what the right documentation is. I've been breaking it down into three buckets around these kinds of company incentives because an organization that is looking at selling to other businesses.

So, business-to-business needs the trust of that business and that business's customers and that's a very different perspective than a second type of company that is looked looking at a foundation of making their financial goals based on advertising. And then really, now being in a place of more business to consumer perspective, thinking about what those individual expectations are and how that plays into balancing out. The different aspects of the expectations for the types of use. And I think in,

in that deployment perspective, I think we've spent a lot of time thinking about what the builders need many years on how we're gonna document the different models involved, the data that goes into the models, what we're going to share out about that internal perspective.

And I think an Amanda really highlighted that how to gap the high-level principles we're getting pretty good at, in the building space. Even if there is. A proliferation of frameworks, and it's great to see the Hiroshima framework getting this kind of response in terms of the reporting. But that how to gap for deployers, I think is coming into a lot more focus. We had one announcement in the last six months that a particular technology has eight mil hundred million monthly active users. And so, you're talking about deployers who might not know these frameworks exist, or if they do, how to leverage those.

And so a for me, the big point around documentation and governance right now is that the deployment guidance really needs to be that practical how to thinking about, early days back when I was at Microsoft and building a practical standard for each audience, and then thinking about what gets shared out from not just the builders, but those deployers, it's often really hard to highlight deep specifics that might be helpful to that audience reviewing documentation because of how fast things are changing internally, I've seen, we've all seen certain coding companies go from, nobody's heard of them to hundreds of millions in three months. And so those specifics right now with this are changing so fast that big picture framework is so important to have. But we also need better insight and understanding of those specifics.

BOGEN: Yeah, the deployer side, definitely something we saw an area of growth, and I think some of the research my team's been doing recently is just on how much change an AI system can undergo when it's modified for implementation, whether that's fine tuning or connecting to tools or just simply being deployed in a certain context. So definitely as part of the risk management big picture, that's a part that I think we haven't seen as much focus on since, a few years ago when the motivation to create the hype process and other similar things were really around the foundation models and the capabilities that they might bring.

And now we're seeing how those diffuse and how we need a collection of information across that ecosystem. Emily, you talked about the developers and the deployers, but there are, other

stakeholders and so I'll turn to Phil. Phil, you're at Armilla AI, which is actually an insurance provider in the Usher ecosystem. Folks trying to incentivize practices that provide confidence that systems are working as they intend that they aren't introducing risks. And so I'm wondering if you can talk about from your perspective, what role do exercises like hype add to the overall risk management picture, but also is the type of information that assurance actors like you all need coming through in those, or does that high level information help guide you towards practices that you might wanna see adopted by folks who are looking for insurance coverage or, that instill confidence that we aren't gonna run into those risks? What's your perspective?

DAWSON: Yeah, thanks so much. I think Emily's comments were a really great segue as we as both provider of third-party evaluations for typically applications have been deployed, developed or deployed not as much on the frontier model side. And now as an insurer we have had a need and for data from our customers for several years. And the advent of frameworks like hype and other regulations and code of co codes of conduct or even certifications are a huge help to us as they help to do a few things. All that contribute to really the scalability of, a trusted information exchange about these systems.

So, one thing that is a set common requirements that the market as a whole can digest and adapt their internal practices to. But the reporting framework and the way in which the data is communicated as well is important. For scalability for transmission. And also, really for companies out there ultimately who can develop way technologies and other ways to automate the collection and transmission of these data's, whether through GRC platforms or ML ops tools, if we're talking about a system specific data. But yes, we have had as a provider of third-party evaluations, we sometimes had difficulty collecting this information. I would say it wasn't typically because of.

The sensitivity of the data, if we're talking about application developers or deployers beyond really standard questions about storage of data or any or servers or data protection sharing agreements which were surmountable. It, it was far more about the, sometimes with larger companies the difficulty of providing us some of the information we request as a matter of logistics. The person the persona who is interested in the third-party evaluation might not have immediate access or knowledge of some of the information that we would require. Whether it's, it was a policy level information or information

about the past even validations that have been done internally. If we're talking to product teams or legal teams and there was a new product, for instance, the, that type of centralized effort at collecting all the information that we might want to see hadn't been undertaken yet.

And I'm talking about two to three years ago. This has changed as the market has matured overall. And this is also the same in underwriting where we do ask about adherence or even certification with standards such as ISO 40 2001 or any efforts to implement and oversee the oversight of and other internal risk management frameworks, maybe based on the NIST framework. For companies that have not gone through that type of either certification or enterprise-wide governance effort they sometimes just don't have haven't developed either the capability or even muscle memory for any one individual stakeholder to be able to collect this information, provide it in, let's say a timely manner that is, makes it worthwhile for our customer and as well for us as a business.

So, frameworks like hype frameworks, certification frameworks as well that that where a third party has a mandate to essentially centralize the collection of this data is a huge help for third party evaluators as well as for the emerging cohort of underwriters of AI risk and liability. I think maybe a comment I'll make is that there is, I think there's not enough of focus in some of the frameworks or let's say granularity. If we're talking about what em, Emily said on the deployer side, to get some of the information that you might want to see, you might want to see that, that is unique given the architecture of a system, given the sector that is used in, and the specific regulations or the uses and the types of testing that would be different than in, in, in other industries.

So, the, the short answer is there's probably never gonna be enough. We're far, we're not really close, I think, to getting sector specific and use specific requirements for reporting on system evaluations. That is definitely an aspirational direction for the assurance ecosystem and ultimately for underwriters as well.

BOGEN: So, we've talked, you've talked a little bit about risk management about incentives for, and the benefit of companies going through this. I wanna just ask everyone to reflect on what impact you've seen organizations going through efforts like Phil was talking about, to do that collation to get the house in order. Does that tend to change practice? Does that tend to lead to more responsible

practices? Or is it more reflecting practices that that are being incentivized by other sources? Whether the need to build trust with customers or some of the other standards. How do these things fit together?

I know when I was in industry and when there was a vacuum of sort of structure and documentation, the simple act of needing to write it down highlighted gaps that then we could identify, needed to be filled and develop those processes. So, I just, I'd love to hear what folks have seen in, in, in your experience or, where you see incentives coming from to do more of this moving forward. I'll start with David.

WELLER: Or a couple things. I think one real, positive development over the past few years is the move from many companies working within their own spaces on these issues to, to more coordination both across companies. And then with governments as they've become much more active in this space, as well as of course civil society. And Microsoft, Google, certainly many of the companies that have been deeply engaged on AI for several years have their own principles approaches for, almost a decade we've been issuing annual report on how we implement our AI principles and have detailed reporting about that.

But I think over the past few years the move towards coordination among the Frontier Labs, for instance, in the formation of the Frontier Model Forum three years ago to ensure that we're comparing what are our own practices, for instance, to deal with frontier safety. Some of those issues, because we're talking about proprietary models and also risks related to, Seaburn, chemical, biological, radiologic, nuclear that you're not going to necessarily have full transparency on. You might talk about your overall approaches and processes, but there's importance for. Comparing your homework among companies in and other experts in the field in more protected venues. I think that's all more that's all important and something that processes like hype of encourage.

I think the second piece that I would say is and maybe picking up on Cam and Brooking's, broader point on kind of network of networks, I think a recognition that this is going to be a multi-speed bicycle, right? There's not gonna be one place, and we are putting way too much pressure on something like hype as the, as the place that we're going to all align on best practices on any of these issues. Or for

that matter, a particular subgroup at iso or at NIST or the Frontier model form. All of these things need to work together. And I think what one piece of homework for us in a place that I see good evolution for the hype process is showing what the connectivity is between those broad principles and best practices.

The 11 or 12 and that code of conduct, what are the specific technical implementations that are being developed that go to those. Broad best practices. So, for something like model cards, there there's a requirement in the code of conduct in terms of reporting on capabilities and evaluations and risks. And again, a number of us for many years have been working on how we report on that but aligning on what are the best practices for that. There's both work at NIST and ISO to develop actual technical standards on that hype. OECD is not the right place to do that. But to the extent that we can map these are technical standards that we are developing that show that we are progressing these best practices and hype can reference, I think it would be, a helpful next step on that.

BOGEN: I'll jump to Emily 'cause you nodded excitedly when I asked the question.

MCREYNOLDS: It there's so much going on and having been, across. Four companies now in this era of a large jump in the advancement of ai we would, the number of times we have wished for a standard like hype with a reporting like hype in order to be able to make an internal point about what a best practice is and having that support. So, while there's, a proliferation of frameworks and, people have heard me say for almost a decade that we need those really practical how to technical implementation specifics. The driving incentive the factor of having these out there and having to answer them is so impactful.

And when we talk about that technical implementation. The thing I always come back to is putting it where people are. So, if you're talking about, when I was working with teams that were building the technology, you don't want to have them go read a principles and framework document. You want it in the hub or the portal where they are launching that model or framework. And when you're talking about deployers, you want that reminder in however that assistant or language model or image model is there. When they're writing in a prompt, it can give them the reminders they need to use this in the best practices way. So, whenever I hear that, technical implementation perspective I go to these

practical points that I think are really helpful. As a starting point when we might get overwhelmed by taking x number of principles into x number of best practices into what is the individual person on the day they're at work doing.

BOGEN: Such an important perspective. I'm sure Amanda and Philly, I imagine you have amazing points. I'm just gonna maybe take two questions for starters, and then folks can respond to, any more thoughts on this question as well as start speaking to other folks' interests. Yes. Mike, we'll get over to you just morning.

AUDIENCE MEMBER: Thank you. Hi there. Colin my question is about what sort of risk management practices are ripe for maybe standardization or that how to point in their lifecycle? Where are we with some of these things that we're reporting on in hype that we can build consensus on to something really concrete?

BOGEN: Great. Maybe I'll actually stop there, and I'll turn back to Amanda, because I know you run the team at Microsoft who really thinks about encapsulating good practices into internal standards at the very least, and thinking about what is ready. Over to you.

CRAIG: Yeah. Thank you. And this is a I think a helpful question to take in parallel to the one that you were talking about before, because I do think when you think about a, what does the, something like providing a report on this framework achieve that's different from something, other processes as David mentioned. And all of our organizations, I think have a lot of robust internal process where we're working on as you mentioned there, Miranda, like how do we operationalize our principles and what are the internal standards, the internal sort of requirements and practices that we need to implement this technology, to develop this technology consistent with our principles.

We're always gonna have continuous iterative improvement of what those practices look like internally, and we have lots of drivers of that, but we also do need to be in conversation more broadly with others in industry and with stakeholders across civil society, academia, and with policy makers. And I do think that is like an area where there is such opportunity through the high reporting

framework. And it really gets to this question about like, where is there opportunity to actually drive progress right now in defining, in more detail, like those shared expectations about practices.

Even I think honestly today at the level of like terminology and what we are associating with different areas of practices is there's real value and working on standardizing that. And so, through the process of. Really working on the specifics of the hyper reporting framework and what is associated with, risk identification and evaluation and what practices align with that. What are we talking about in the term in the context of risk management practices, risk mitigation practices and this piece that Emily raised around who's doing what and how like all of those practices need to be implemented in a way that's coherent for ultimately achieving the risk management outcome that we have.

There's a real need to deepen understanding about shared responsibility across the AI value chain for those practices. I think that's the opportunity that something like the hyper reporting framework, it's ripe in this moment to really make progress in that way. And there's a real need and something like the hyper reporting framework, especially to the extent that it evolves toward providing more clarity about, the kind of pathways for reporting for different parts of the AI value chain.

And we can maintain kind of coherence across like how different parts of the value chain are providing context on their practices that accrue to the kind of risk management areas that, that the high reporting framework enables you to report on. That's a real opportunity for the ecosystem. And again, we're gonna have. Our own internal processes. It's always helpful to be able to also have this kind of internal external feedback loop on what expectations are, how we are aligning those and what those look like across the ecosystem.

BOGEN: Phil, on that same question your organization just adopted ISO 30 40 2001 certification as a concrete indicator of risk management that you actually take into account in your underwriting. So that seems like it's a signal of a sense of maturity of certain practices. But I wonder if you could speak a little bit to why did you make that move and how would you determine whether a practice is mature enough to actually consider from a monetary perspective as to addressing risk?

DAWSON: Yeah, that's a great question. And we do look at, we do ask about ISO four 2001 adherence certification as part of our application form for our insurance. As well as other frameworks too. For instance, NIST, AI, Risk Management Framework or whatever else they have been using as a reference point for us. And in part, as I said before, to hopefully nudge the customer in the direction of some of that readily available information. Because of that internal effort that has been under undertaken, ISO four, 2001 has over 200 companies that have now Certify have achieved the certification. It's not actually a huge number but we're also we're trying to incentivize the adoption of some of these leading frameworks. So that, that is one NIST as well.

And at least it's a starting point today for what we hope will be even greater adoption of that standard. Others as well, what we would love to see is at the end there is some work at ISO already 4, 2, 1, 1 9 others where there is more system level standards being developed. And David is quite right that, hype might not be the forum in the vehicle to answer all of these questions that we have, either as a third-party evaluator or underwriter. But there are some other efforts that we would like to encourage as well. And given the early stage of AI insurance and assurance, I think more broadly we also see ourselves as having a role of signaling what will be important. Even if it is not giving us the full picture or as much as information as we'd want or precise or as deep we will continue to signal and try and drive the market in directions to, to support that adoption.

BOGEN: Thank you so much. I think we're running up on time, so I'll wrap up here and just say thanks everyone for joining us today for this overview leading up to the AI Impact Summit in India. Coming up for the release of this research on the hype reporting framework and for these really robust discussions. Thank you to the panelists. And I'm sure there'll be many more such conversations to come both up to the summit and beyond. Have a great day.

TABASSI: We are getting close to the last session and the closing. My name is Elham Tabassi. I'm the Director of AI and Emerging Tech Initiative and a senior fellow at Global Economic Development at here at Brookings. And I joined Brookings in March, and it from, I used to work at NIST. It really warms my heart every time I hear NIST name in ending of the conversations. So, with that I am actually very pleased to have a conversation with my former colleague Jesse Dunietz at NIST. It was mentioned that this is also working on some sort of a pre standardization documents and effort.

We call them zero draft on transparency requirements. And given that this the focus of this effort and this conversations and why we start looking at the hype is this broader transparency documentation. And we heard it this morning from many of our speakers that a there are many different aspects to this from transparency of the organization's governance to things such as data sheets for data sets, to transparency at the data level, at the model, at the system risk management. And the need for actually also be clear on the audiences of we are doing the transparency reporting for what purpose and to whom. So, I thought of joining Jesse on hearing about what NIST is doing along this line. So, Jesse it is now about a year or so that NIST start this project of zero drafts, and one of them is on transparency requirements. How about be feel free to introduce yourself and tell us about, first let's start about motivation for zero draft and some sort of overview of the document.

DUNIETZ: Sure. Thanks. Thanks very much. And wonderful to be here with you at...

TABASSI: So, then we can't hear you well...

DUNIETZ: Lemme see if I can fix that. I can try speaking loudly, if that helps at all. Is that any better?

TABASSI: It's not it didn't, it's the same. Can you hear him? Can the, now we have to do something with the audio.

DUNIETZ: What about now?

TABASSI: It's good.

DUNIETZ: It's good. Okay, wonderful. Apologies. Very good to be here with you all, at least virtually. As Elham said, I'm Jesse Dunietz. I am a computer scientist with the AI team and the Information Technology Lab at NIST, overseeing a lot of our standards work, and that includes the Zero-draft project. Maybe I'll just give a moment of background. As Elham knows we had a lot of engagements with the community in 2024 on what the needs were for standardization in the community. And we kept hearing that everyone wanted standards quickly and also involving everybody they could think of in the process of development which are both very hard things to accomplish and they're extra hard together. And in addition to that, NIST is just one player in this broader ecosystem. We don't even run

the process of standards development. So, zero drafts as a concept was our answer. Really largely Elm's answer to that question of what we could do and blame on...

TABASSI: now you know where to put the blame on, but go on Jesse...

DUNIETZ: And no, no blame there. I'll credit and we decide, so the concept was that we take the pen early on to come to. Something that is rich and thorough and relatively mature with a lot of community input that is not a standard, but that we can then submit as a proposal to say, we've gotten you a long part of the way toward an actual consensus standard and now go through the actual consensus processes to get the rest of the way there. And the hope is that our process would be able to pull in more kinds of input and also hopefully speed the process up end to end. So, it's a pilot to see if we can accomplish that. We chose the documentation topic as one of our two initial pilot topics for a couple of reasons. One is that there is as we've heard a little bit about already today, a large body of work that's already been done on what makes for good documentation.

And so, there's, there is something to standardize that we can probably get to consensus on, at least to some extent. And that there was a sense of real significant need in the community as again, we've heard already to have something about this. And so, we really, we decided to maximize the extent to which those things were true of what we're focusing on. For the moment we're focusing on documenting AI related data sets and AI models leaving the system piece for future work.

TABASSI: The, so you said that you are working with the community. Can you give us an idea about the who are you engaging? What are those people or those people? And also, what's the stage of the document where it, where we are now?

DUNIETZ: Absolutely. We have been getting lots and lots of emails to an open email inbox, which you can find on our website with comments on what the materials we've put out at various stages so far. And so that's anybody and everybody. We've also gotten we've worked through the NIST AI Consortium, which has quite, quite a variety of large companies, small companies non-governmental organizations academic organizations. And then we've also been working with the us national Body to ISO IECs committee on ai to make sure we're reflecting all their views and aligning with them in

advance of ultimately turning document over to them. Where we are now is we have a pretty thorough extended outline that we've put out a few months back.

We've collected a large amount of input on that and are hopefully we've found resolutions to some of those key outstanding issues that now we're in a position to. We're drafting at this point, we're actually putting together draft text that is gonna come out. And the way that we are working through the document, the way that we're thinking about the document is it has a couple of key components. It has a piece about potential outcomes of documentation, meaning. Dimensions that documentation can affect positively or negatively, depending on whether and how documentation is done. Then some guidance on how to do it well so that we get toward the positive versions of those outcomes more than the negative versions.

And then keeping all that in mind, actual templates for both AI data sets and AI models that are first of all designed to reflect all of that guidance in the rest of the document, but then also to really provide some flexibility for people to then further apply the guidelines to tailor what they do and don't put in to, for example, really meet the needs of exactly who they're targeting with this documentation. And so, we try to specify in the guidelines piece how to do that and the templates provide the practical mechanism to do it in a way that is comparable across providers and across documenters.

TABASSI: Looks like to me that this is a piece in the broader transparency documentations. And going back to a lot of things that we heard, we need to connect all of these things together. We need a combinations of this type of, best practices, zero drafts towards standards to cover transparency of the different parts of the AI lifecycle. So, I wanna hear from you how you see all of this connects to hype. And then it also can give us a sense of the timeline on as hype is gonna be revise and discuss as we heard it from OTRE and how you, your timeline of sending it to ISO.

DUNIETZ: Sure. I think our focus is really the kinds of things that had typically been in, included in a model card or a dataset card or, documentation, artifacts of that nature. And so, we're looking at trying to provide that common structure very practically speaking for reporting the details that some audience needs to know for a purpose. Can I use this model for my purposes? Can I trust this model for, for, has it checked the boxes for whatever policy requirements I have in place? And so, trying to

provide the mechanism for that kind of communication at a fairly granular technical level. Actually, there's even a, I could imagine even the part, there's a line in the HYPE framework that has a question about what, whether you have such documentation, one could imagine simply pointing to one of an artifact of the type that we're talking about producing MA in the template that we're providing and saying there, that's it.

We provide that type of documentation. Go check that out. So, I think there is a connection there. It isn't quite the same form of transparency and documentation. In terms of timeline and next steps we are again, currently in the stage of drafting this. I would expect that we will have a draft out for comment within the next couple months. And then beyond that, we will hopefully not too long thereafter take all the comments that come in and turn that into something that we hand over to the US National body and the US national body who represents the country, not just the government, but the country in ISO IEC will then have to decide what to do with it and hopefully in, in relatively short order, propose the content or incorporate it somehow into the ISO IEC suite. I would expect that could happen by the end of this year. Although it depends what process is it might happen as, in the fall. It might take even shorter. It might take even longer than that, depending on precisely what bureaucratic process is followed to get it into the ISO suite.

TABASSI: Fall 2026 plus minus one quarter. I wanna open the floor for question for Jesse. We have one question here. Please Go ahead. I don't see a microphone so I can give you a microphone.

AUDIENCE MEMBER: Sorry. You said you were doing zero reports on documentation. Could I ask what? Documentation is it only paper based or could it eventually be a model itself, which might have a paper report at any particular time, but which you could update by putting in further papers. So, what documentation?

TABASSI: Jesse, did you get the question? So, the question is that, is it is it document for documentation or that document for documentation can be interactive and online? I think that was the question.

DUNIETZ: The, what we're producing is agnostic to that. So, what we're producing is what the information structure should be. We are not trying to codify what form it should take. What are the machine-readable formats, what are the, whether it's a PDF or a website or anything else. And so, you could take the same information and put it to any of those forms. That's the question on that. Yeah.

TABASSI: Any other questions for Jesse? All right. Thank you so very much and all the best with this process, and we follow this. Thank you.

DUNIETZ: Thank you.

TABASSI: I believe we are coming to the end of our program. I wanna thank everybody that join us, either in person or virtually. Special thanks to all of our speakers for sharing their time and insights with us. I think if I want to quickly summarize it is that yes, we do need transparency, and transparency can takes a lot of forms, and there are a lot of people that working on this, and we need to connect all of those things and try to do to be better and hopefully get all of them towards some sort of a standards or sets of standards that can be the backbone of better responsible ai, but also a lot of the policy and regulation conversations that's going on. So, thanks again and I'm sure that this conversation is gonna be continued. So, stay in touch and congratulations to the authors of the report and my special thanks to all of the Brookings and CDT team for making this event possible in both in terms of the content and topic for having conversations, but also all of the logistics behind the behind the scenes. Thank you all and we'll see you soon. Bye-bye.