Center for Strategic and International Studies

TRANSCRIPT

Event

AI for Food Security Forum

Plenary Panel IV: Implementing AI for Food Security – From Inspiration to Action + Closing Remarks

DATE

Thursday, April 30, 2026 at 2:00 p.m. ET

FEATURING

Judy Chambers
Director, Program for Biosafety Systems, International Food Policy Research Institute (IFPRI), CGIAR

Olivia Graham
Product Manager, Google

Crystal Haijing Huang
Senior Economist and Director, IDinsight

Amen Ra Mashariki
Director, AI and Data Strategies, Bezos Earth Fund

CSIS EXPERTS

Caitlin Welsh
Director, Global Food and Water Security Program, CSIS

Aalok Mehta
Director, Wadhwani AI Center, CSIS

Zane Swanson
Deputy Director, Global Food and Water Security Program, CSIS

Transcript By
Superior Transcriptions LLC
www.superiortranscriptions.com

CSIS | CENTER FOR STRATEGIC & INTERNATIONAL STUDIES

Zane Swanson: With that, it's now my pleasure to introduce the last panel of the day, From Inspiration to Action. We've heard all of these amazing stories about how AI is transforming data, and creating tools, and connecting tools, and connecting people for the purpose of creating a more food-secure world. Let's talk about now how we actually actualize that, and really drive food security. For that, I'm welcoming today's emcee and the director of the Global Food and Water Security Program here at CSIS, Caitlin Welsh, back to the stage, along with her panel. (Applause.)

Caitlin Welsh: Hi again, everybody. It's great to be back here to moderate this fourth and final – I can't believe we're at the final already – plenary panel discussion, "Implementing AI for Food Security – From Inspiration to Action," where we want to bring it all home, bring together all the things that we've talked about throughout the day. Throughout the forum, as Zane was mentioning, we've heard about many of the ways, the exciting ways, that AI is being applied across food systems today. Many of the technologies still on the horizon, and some risks alongside opportunities.

And in this panel I want to explore what's needed from critical sectors to develop and promote the use of AI-enabled technologies that minimize risks and maximize benefits for the greatest number of people. Because at the end of the day, we want to use AI to improve food security. So how do we do this?

I am very pleased to have gathered a star-studded panel with representatives from the sectors that I think are critical to this task, funding, technology development – funding, tech development, policy making, regulation. and M&E. We have with us Amen Mashariki, who is director for AI and data strategies with the Bezos Earth Fund; Olivia Graham, who's a product manager at Google – hi, Olivia, thanks for joining us online; Aalok Mehta, who is director of the CSIS Wadhwani AI Center; Judy Chambers, director – Judy – (laughs) – who is the director of the Program for Biosafety Systems with IFPRI, International Food Policy Research Institute; and Crystal Huang, who's senior economist and director with IDinsight. So I have a great panel. Really excited, as I said, to bring it all home in this final panel discussion of the day.

And I will start actually with the M&E sector. We've heard a lot from representatives of great research and tech development firms and everything. And so it might be a little bit counter intuitive to start with M&E, but Crystal has been doing really fantastic work in this regard. And so my first question to you, Crystal, is about a product that you've launched recently in collaboration with CGD and The Agency Fund. It's a playbook for the evaluation of AI models for development, including for

food security. So can you talk a little bit about this playbook and its framework?

Crystal Haijing Huang: Absolutely. Yeah, thank you for having me.

So this generative AI evaluation playbook was actually built off an earlier blog post by J-PAL, Agency, and CGD. And now IDinsight is involved. It was really made in response to the fact that AI interventions are now being deployed increasingly across sectors of development. But there's no unified kind of consensus on what evaluation should look like. And in fact, like, what a definition of success looks like. And so the playbook outlines four levels of evaluation.

You know, actually, I'll kind of backtrack, and actually say that the word "evaluation" itself can mean different things to different people that you talk to, right? If you're talking to somebody that's in tech or a product developer they mean evaluation as benchmarking, kind of performance metrics. If you talk to a development practitioner that's involved in the M&E side of things, that means impact evaluation and measuring kind of downstream outcomes. So how do you get these different camps to speak to each other in kind of a coherent framework?

And so this playbook lays out four levels of evaluation. And I can briefly walk through each. The first level is model performance evaluation. So that's level one. And this level really asks, is the AI system working as intended? Is it accurate? Is it producing kind of factually, contextually, culturally relevant responses? And is it, you know, consistent across, like, languages, for instance? So there's different kind of model performance metrics that you'll want to evaluate your tool against. Level two is this tool actually engaging users. Like something that's super accurate, like if a farmer had gone and gotten on chatbot, is not very useful if the farmer kind of abandons it after two sessions. And so to what extent is it engaging and retaining users? Can you maybe run some A/B tests through, like, rapid experiments to design aspects of the user interface that is actually helping to engage users.

Level three is product value – sorry, level two is user product evaluation. Level three is user evaluation, which is, I guess, more close to kind of social science, kind of intermediate outcomes. And so this level asks, is interacting with the tool actually changing a user's thoughts, feelings, and behavior? You know, are they – is a farmer that's interacting with an advisory tool more confident in their kind of input decisions? Are they actually changing their kind of behaviors around farming in certain ways? Is it increasing their knowledge of different kind of crop disease? And so this can use, you know, things from, like, sentiment analysis

based on in-app data. You can consider, like, embedding in short surveys within the flow of the conversation to kind of track these outcomes.

And then level four is impact evaluation. So is this tool actually changing development outcomes down the line, whether that's yields, income, or kind of food security indicators? And, you know, the crucially, like, once you get to this level it's really important that you're not just kind of investing in, like, a $1 million RCT, but, rather, you're moving the product through these different stages of maturity and attaching these evaluations along the way. And I think my last point here is that it's really important to think about continuous evaluation. It's no longer enough just to do this like big RCT down the line. AI is a tool that's different than many other interventions that are static. It's constantly changing. The model is changing. The foundational model is changing. And so evaluation needs to be kind of embedded as a continuous effort versus a one-time task.

Ms. Welsh: Thank you. Thanks so much. I assume we can find this online?

Dr. Huang: Yes. It was just released, like, last week. And so –

Ms. Welsh: Yeah. OK. Congrats. Look forward to looking at that a little bit more. We talked also about – in the application of AI-enabled technologies for development, including for food security, that there can be unintended consequences. So how do you propose that policymakers, implementing partners, tech developers, and others think about this idea of unintended consequences?

Dr. Huang: I love this topic because I feel like there's a lot of rightful concern in the public discourse around these, like, big apocalyptic, dystopic outcomes of AI. But I think what's equally important and maybe more directly preventable and tangible is to think about at the micro, programmatic level, what are the consequences when you're applying AI to a given problem? So I can name three.

The first is selection bias. And this has kind of come up in some threads earlier, where, you know, these tools, whether that's, like, an AI advisory tool for a farmer or something that helps them better make these decisions based on kind of a predictive algorithm, that requires a certain amount of motivation and literacy to be, like, engaging with the tool. When your information seeking now becomes very self-directed, compared to the status quo where, like, a field extension worker is running these meetings and giving you information when you're kind of receiving it, right? So who does that benefit? That benefits farmers who are more motivated, who are more technologically literate. And that can

eventually lead to these kind of disparate outcomes in who benefits from AI.

I think related to that also is, like, we need to think about what accountability relationships are getting lost. Like, it's very natural for humans to feel accountable to another human. If you're seeing this extension worker at the next collective meeting, you're more likely to adopt the practice that they told you during the last meeting, compared to feeling accountable to an app. And so if you're kind of moving – substituting in any way from human-directed receiving of information to an app, like, what elements of accountability, of trust, which is something that was mentioned earlier, gets lost.

The third one is what I call creative constraint. And so there's an example in a paper that came out recently in this prominent economics journal where they looked at – there was a study on call center workers in the Philippines, where they were giving these call center workers scripts to better field these questions. And we've all, like, been there, where we've spoken to someone who's really bad at their job and it's really frustrating, right, or someone who's really good. And they found that overall this AI, like, scripting improved outcomes, but for those who were really top performers and good at their jobs previously, it actually worsened their performance because they were relying on this rote AI-generated data versus their kind of human intuition and skill.

And so I do wonder – I mean, I know that's not an ag example – but there are crossovers to when you're providing information, maybe a field extension worker is using that versus their own kind of relational judgment. We know that information is just one piece of the puzzle. So, like, what are you disrupting that's not just the information barrier.

Ms. Welsh: Yeah, very, very interesting. Is there somewhere we can go for information on that too?

Dr. Huang: Yeah, there's a thought piece that I put out on it. Yeah.

Ms. Welsh: Great. Thank you. Well, I'll move now to speaking with Olivia Graham. Olivia, thank you so much for joining us online. We had a really great conversation in advance of this panel so I really look forward to talking to you about what you think needs to happen to take us from inspiration to action. Before then, though, can you talk to us – can you just explain the model – the program that you've been working on with Google?

Olivia Graham: Yeah, definitely. And I think this is actually a perfect transition, because I come at this problem space from the perspective of a researcher, of a software engineer, as a developer of technology. And so that's really

what we do at Google Research and Google DeepMind. I'm a product manager. And my focus for the past three years has been developing more accurate weather models. And if you spend enough time working on training accurate weather models, you begin to ask yourself, right, where can these models really drive outsized impact? And agriculture was the clear answer to that question.

And so I've spent a lot of time on the continent talking to met agencies, learning more about how we can use AI in helpful ways. You know, it's not a panacea. AI can't solve every problem. But these AI weather models do have unique advantages. They can make sense of vast amounts of data, non-traditional kinds of data, like you heard from Philipp with the World Bank's project on forecasting food insecurity. So we've been asking ourselves, how can we play to our strengths as an information company? How can we develop new technology? And how can that technology then serve the world and increase global food security?

And really what we've done that through is collaboration. So we are focused on the tech piece, but we then find wonderful partners out there in the ecosystem – folks like Brian Miranda at TomorrowNow; I can't see you, but I hope you're there in the audience – who are able to complement the work that we're doing and really drive it into fields, into farmer-level impact.

Ms. Welsh: Wonderful. Thank you, Olivia. And just one note. Olivia mentioned met agencies, which is meteorological agencies, because she works with weather data. So that's what – we'll be talking about met data, and that's what we – what we mean by that. Olivia, in your work, what's one of the most important factors you think in your ability to develop usable and impactful AI tools?

Ms. Graham: Yeah. And thank you for catching me with that. (Laughter.) I spend so much time in the meteorological world, and it's hard to say meteorological so I avoid it when I can. (Laughter.)

For me, it's data, data, data. And you heard that in the previous panel, but it really does hold true. With one of my earliest projects at Google Research – again, we were trying to build models that would be more performant in an African context. And the first challenge that you run up against is that the ground truth data that you need for validation is not openly accessible. And so my team spent probably close to a year sending out emails calling meteorological agencies, asking them if we could get access to their data, if we could pay them for access to this data. And we just weren't able to get much of it at all. There's some trust questions there, which I think are really interesting. But if you don't

have access to good data, you simply cannot build a good model. You cannot know if your model is performant for the use case that you're hoping that it serves. And so data really is the beginning and end of model development.

Ms. Welsh: Yeah. Thank you, Olivia. Do you have thoughts on how we – we're calling it reversing reluctance – how to reverse the reluctance of these agencies that are reluctant to release this data for public use?

Ms. Graham: Yeah. I think you have to articulate the value in a compelling way. And what we're seeing over time is that you can – hmm. As much as there is this desire to protect this data, to ensure that it doesn't get shared widely, there are risks that are associated with that. If the next generation of models – call them weather models, plant phenotyping models, whatever models – if they're developed without your data, they are not being developed with you in mind. And so finding ways to articulate that loss, that risk, that is inherent in protecting this data, and finding ways to motivate contribution of data to the open ecosystem. Whether those are calls to action, hackathons, Kaggle competitions, anything that can motivate folks to really understand and see what value this unlocks for them.

Ms. Welsh: I just want to ask one final follow-up question, because I think this is something that's really interesting that came out of the conversation we had previously. Can you give us an example of something you have in mind for, like, a benchmarking competition. Where, let's say, one country makes its met data publicly available. And then you were saying that perhaps then there can be a competition among model developers. Can you pick up – like, can you explain what you have in mind there?

Ms. Graham: Yeah. Definitely. So there's a great platform called Kaggle, where people will post datasets and then, essentially, issue a call to action, telling not just the meteorological community but the machine learning community at large, look, come help us solve this problem. I've seen that be a really powerful way to drive development. The other thing that I found pretty inspiring, motivating over the past couple months is actually the work that the European Center for Medium Rrange Weather Forecasting has done, to say, look, we know that sub-seasonal to seasonal forecasts are the most critical in agriculture. And therefore, we're going to host a public competition. Come get the glory. Come solve this problem. Come make an impact. I've seen that be really effective as we think about researchers and what motivates them.

Ms. Welsh: Yeah. Thanks for planting the seed for that idea. Perhaps we'll see more of that in our sector. I'll take it back into the room now. Thank you, Olivia.

Turning to my colleague Aalok Mehta, who is director of the CSIS Wadhwani AI Center, but I'd like to platform not only your role as my colleague here but actually I want to draw on your experience before you came to CSIS. So at OpenAI, SeedAI, and Google, as well as your time on the Hill. Really, there are a few people with such deep experience in the tech community and in the policy community. So in a conversation we had leading up to this forum you mentioned that you think it's important for companies to develop better and faster use cases on the ground. So why do you think that this is important for the work that we're all talking about today? And then how do we incentivize companies to do this?

Aalok Mehta: Yeah. So I think one of the things that's really important here is that AI tools are most effective when they're tailored to the specific situation that they're trying to address. And so we see, for example, China is offering – is engaging with governments and sort of really trying to dig down on specific problems that countries are facing, and then come back and develop turnkey solutions. So this is sort of a very produced sort of solution that a government or organizations within a country can adopt with minimal technical knowledge. And so what this means is that they have looked at the problem, they've come up with a bespoke solution, and they're offering an easy way to access that. And I think this is a real key thing that both the U.S. government and U.S. companies should think about doing more.

And I think that there are a few ways to incentivize this. The government, the U.S. government, for example, is working on a sort of export control approach to sort of exporting U.S. technology around the world. I think it's important when doing that to focus on building out sort of the equivalent of turnkey solutions. So minimizing the level of technical knowledge that is required to implement these solutions, especially when you're – when you're thinking about sort of agricultural communities, which may not have either high levels of technical knowledge or high levels of connectivity, or experience with using these kinds of tools at scale.

I think there are incentive programs that the government can use. So it's used sort of prize challenges and other sort of challenges as a way to engender sort of public or a private sector competition in development of solutions. Some of these are sort of solutions for space technology, but there's no reason they couldn't apply to agricultural technology as well.

Ms. Welsh: Thank you. Your note there about the space technology actually made me think of the Genesis Mission, which is something that the White

House launched in November. It unveiled the Genesis Mission to accelerate AI for scientific discovery. It focuses largely on AI infrastructure, but can aspects of this mission, you think, be leveraged to advance AI tools for food security?

Dr. Mehta: Yeah. I think there are two ways that are really important to think about when it comes to the Genesis Mission. So one of them is that generally what Genesis is trying to do is accelerate scientific discovery overall. So the use of AI tools so to sort of speed up, streamline the process of engaging in scientific discovery. This will have a lot of spillover effects to all kinds of agricultural and food-related research. But I think the second thing that a program like this can do is really focus on developing specialized models for food-specific applications. So we've seen a lot of success in the scientific field from specialized models.

The most notable of these is AlphaFold. And so the latest generation of AlphaFold not only predicts protein structures, but it also is able to predict what happens when you introduce new kinds of molecules into the system and how that protein might change. And it's being used by, I think, more than a million scientists worldwide on things like drug discovery. And so, again, this is an approach that can also be used, I think, with some great success in things like researching crop varieties and thinking about how to tailor new varieties of crops to deal with specific kinds of geographical challenges, including challenges related to pests, water, and other kinds of natural resource availability.

Ms. Welsh: Yeah. Thank you. Thanks. One final question before we move on to the next panelist. Again, you have really high-level experience in the White House and also on Capitol Hill on policy development, policy implementation. Are there certain policies you think that exist on the books that can be better implemented, or policies that are not on the books that should be thought of to advance the development of AI for food security and the use of AI for food security?

Dr. Mehta: Yeah. I want to double down on the issue of data. So I think data is really important here. We've had some discussion about weather data. There's all sorts of other data that the U.S. government has access to that can be really relevant to food security applications. So the Agricultural Research Service has, for example, significant amounts of data around things like crop yields and what farmers are planting. And what's really important here is trying to get that data made as widely accessible as possible, so making sure that that data is available via API's and other ways that allow people to get easy access to it.

I think the second thing the government can do is sort of survey the broad areas where it collects data, really identify gaps that exist in the

current data collection apparatus, and then figure out ways to collect that data. A lot of the data that we're talking about here is data that is only collectible by the government, or that the government has existing infrastructure to collect in a way that private sector companies can't. And so I think this is – this is also important. And I think the third thing that we can do is think about how to standardize the metadata and presentation of this type of data across the world so that you are able to collate and leverage datasets from around the world, not just from a handful of companies, without requiring significant retooling or technical expertise.

Ms. Welsh: Thank you so much. I do have follow-up questions for you, but I'll move next to Judy Chambers. We've now helped answer this question about how do we move from inspiration to action from an M&E perspective, policies perspective focusing on the United States. We'll talk about regulation, and I want to focus on the work that you've done in several African countries – actually, across many, many countries, but your work is focused on, I think, four in particular. And before we get to, you know, what needs to happen to improve regulation within and across these countries, can you explain what you see as the ideal state for regulation, and then we'll – then we'll work to how it is that we get there.

Judy Chambers: Thank you. That's an interesting and difficult question. So I come to this not from the perspective of the AI practitioner, but I have spent decades – and that should tell you all you need to know – working on a similar policy climate for ag biotech, where the regulatory environment, in particular, became the single most consequential and contentious issue that affected whether countries would actually end up accessing and adopting and benefiting from the technology. And I've seen it from multiple institutions as well, from a small biotech company, from a large biotech company, from the perspective of the U.S. government, specifically USAID, where I was working at a time where the biotech issue was very much in the national competitiveness focus. And now I'm at an international public goods institution and looking at it from that perspective as well, particularly around the food security issues.

And so I have a simple message, which I think most people in the room will appreciate. Which is that regulatory frameworks cannot be an afterthought, particularly for transformative, controversial technologies. They have to be a precondition for it. And they have to be proactively addressed with the right resources, expertise, and capacity-building for policy, not just the technology, and also while respecting the fact that not everybody is going to view the technology as you do. There's going to be differences of opinion.

So the ideal framework, hmm. Well, I've identified five attributes, first of which is balance. That's very important. You have to have a balance between evidence-based risk assessment and the benefits versus assessments that are done on the perception of risk and precaution. And our experience in the biotech industry has shown that when perception – risk perception drives the policy without the accompanying evidence, that the resulting regulatory regime is precautionary in theory but paralytic in practice. And so decades later we've seen issues from these kinds of policies resulting in the fact that many countries around the world have not had the access to the technology. They've not benefited from it. And they have had to revise their policies, which has been costly, time consuming, and laborious.

And then second is the issue of context sensitivity. So I'm skeptical, I would love not to be, that we can achieve a universal AI regulatory template that works identically everywhere. I think you have to take into account the local institutional capacity, the existing legal frameworks, the infrastructure realities and the local agricultural challenges. And however, like Ruben said before, I'm an optimist. And so at some point I think we are going to get to a point of a more harmonized policy than what we're currently seeing at the moment. But we have to work on that.

And then the other thing is some mutually enforcing aspirational policies, which, like, they go together, transparency, predictability, enforceability. So regulations that exist on paper but cannot be implemented or enforced absolutely create a worst-case scenario. You end up with compliance costs but not the accompanying safety that everyone is aspiring to. You end up with uncertainty that discourages investment, and a trust deficit between both innovators and the public. The fourth is an issue of coherence. You cannot develop regulatory domains in silos – and this is really a lesson for governments in particular – because you'll likely end up with frameworks that are internally contradictory and practically unworkable.

And then the fifth and the most critical thing, I think, and we have talked about this, Caitlin, is the early and sustained – early and often sustained stakeholder engagement. So when you have a policy that's advancing ahead of the practical experience on the ground, as is currently the case in many African countries that we're seeing, there's a risk of regulating based on assumptions and not on the evidence. And so engaging a broad group of stakeholders is critically important. It's how you build the knowledge base, the practical frameworks that are both legitimate, functional, and trusted.

And so for all this, the operative word is "enabling." It's a word that gets thrown around a lot in the policy community, but it is truly what's needed. From my experience with the ag biotech situation, the ideal state of a regulatory framework is one that's informed by an inclusive approach that facilitates innovation, engenders trust and competence, while managing the real risks across the whole of government, and it's both responsible and ethical.

Ms. Welsh: Thank you. We love to see that, that encapsulates so much right there. So thank you for that – for that succinct answer. We probably won't have time to get to the question about next steps, but can you just briefly explain where you found we are today with regard to regulation of AI technologies for food security? And in particular, can you tell us about how you find influence from the United States, the EU, and China informing regulatory states across these countries?

Dr. Chambers: So we had some support from the CGIAR Digital Transformation Accelerator. And we did a very small, limited desktop study looking at the AI regulatory environment in Africa. And it was focused on four countries. And what we found in general is that the policy development situation in Africa is very fragmented. It's moving quite quickly, but it's very uneven, the movement. And so there's 15 countries that have published national AI strategies, but only a handful that have actually turned those strategies into a policy. So that's an issue right there, because the technology is moving faster than the policy climate is.

And then the other issue is that, in the absence of that, the data regulations have become the de facto regulatory instrument by how countries are governing these technologies. And they all are relying on different structures and different geopolitical forces. So the four countries that we looked at were South Africa, Kenya, Nigeria, and Rwanda. And all of them have been influenced by the EU's GDPR policy. But they all then adapted it to meet their own internal conditions. And so we looked at also what the geopolitical influence is of the three primary actors, China, U.S., and EU. And the EU is – basically is the template for most African data laws. And so they are also supporting these strategies through their development activities in both Kenya and in Rwanda.

And then, in the case of the U.S., we are shaping the policy by our market environment, through our commercial companies, who are working broadly across Africa through the CLOUD Act's extraterritorial data access provisions, and explicitly, more recently, through a very trade-oriented AI strategy. And then China is working through their infrastructure investments, and increasingly through the applications that you were talking about. And so that creates some structural

arrangements and relationships that could actually supersede the data laws that are currently on the books. And so there's a lot more that needs to be looked at with respect to this. We've just scratched the surface. But we do need to do a deeper dive to figure out what's going on.

Ms. Welsh: Great. Thank you, Judy, for that very comprehensive overview.

We've now heard about what it takes to move from inspiration to action from the perspective of M&E, tech development, policy, regulation. And now let's talk about funding. So we have Amen Mashariki, who is director with for AI and data strategies with Bezos Earth Fund, as I had mentioned. Regarding the needs we've discussed so far, what do you see as the role of philanthropy? Generally speaking, but then how does Bezos Earth Fund approach this specifically?

Amen Ra Mashariki: Yeah, yeah, yeah. So I used to work in D.C., way back in my days. And I have a Ph.D. in computer science, so I'm primarily technical. I completely forgot how, you know, being around AI researchers and scientists, how complex policy and regulation is – (laughter) – compared to the work that's being done on the ground in the AI space. It is quite impressive, the complex things you guys have to think about.

I will say this, so being new to philanthropy I hesitate to sort of jump in there. I came to the Bezos Earth Fund from NVIDIA. And so, like I said, I'm primarily technical about this. Here's how we think about this at the Earth Fund, specifically around AI, which is a horizontal at the Earth Fund across our many verticals. We start with the horizontals first, right? And so when we executed the AI Grand Challenge for Climate and Nature, it was really unique in that it was two phases. The first phase was we just put out an open call globally and said, any organization that has a really, really powerful idea on how to transform their work in four focus area – biodiversity conservation, sustainable protein, which is relevant here, and then power grid optimization. And then our founder, he basically was, like, yeah, add a fourth one, wild card, which basically superseded all of the other three, right? (Laughter.)

And so we said, give us – I do not mention AI. We could care less about any interesting AI thing. What we want to understand is how are you thinking about the problems in your space and how to maximize impact. And tell us about that. And what we did was we got over 1,200 submissions over a short period of time. And then we dug through. And then we funded, to the tune of 50,000 (dollars) a piece, 30 of the more compelling ideas to now take that interesting idea that you offered, and the $50,000, and then now think about what role AI could play in executing that work, right?

So we started with the problem and the idea, and then we actually created what we refer to as the innovation sprint. So for three months we put them in touch with sort of my former colleagues at NVIDIA, Esri, there's a whole list of organizations – tech organizations. Because we were not funding tech organizations or AI organizations. We were funding on the ground climate and nature organizations. And so we really saw it – the way we wanted – what we wanted to fund was the AI support of solving a particularly powerful problem.

In my career, I started my career at a company some of you may remember. It was called Motorola. (Laughter.) It's almost nonexistent to some extent. And I've been in tech companies my whole career. And I can tell you, every time – every tech company I've been in they've always tried to solve the domain problems without the domain people at the table. And so we wanted to set the table with the domain experts and then fund the domain experts to bring in the AI resources. And then from that 30, we selected 15 to fund 2 million (dollars) a piece to actually get the work done.

Ms. Welsh: Wow. Very, very interesting. Can you give examples of funding decisions that you think were ultimately successful, whether in the food security space or in a different sector?

Dr. Mashariki: Yeah. No. So the way we thought about the funding decisions was that we needed to understand that they understood, to your point, what data they had, but also what data they needed, right? And so you weren't required to have your arms wrapped around all the datasets you needed in order to execute this, but we needed to understand that, you know, you knew what you didn't have, and what partners you could bring to the table.

We also needed to understand that they understood the commitment to AI. When I was at NVIDIA, we used to have a joke which was – and this is the honest truth – we used to have a joke, which was, you know, a lot of people back then were talking about AI, but really didn't understand the need for compute and infrastructure to support AI. So if you were actually doing something on your laptop, eh, that was more like data science, and classical machine learning, and so on and so forth. Really didn't move into the space of AI. And so what we needed to also understand was these organizations understood the infrastructure need to get there.

And then lastly, we needed – we needed – the question that we needed them to answer was, I cannot solve this problem if it were not for this AI capability, right? And so we call that AI solution fit. Sort of play off of

product market fit, right? You couldn't – you couldn't say, well, I could solve it. It would just take longer and be more laborious, but I could solve it without AI. No, it had to be, like, we could not. And so there's some examples with food systems innovation where they have billions upon billions of datasets that they're using to train and build an open source model on how to predict the types of food that people are actually looking to – that need – they need.

We worked with University of Leeds. We funded University of Leeds to take waste and transform that waste into microbial protein, and other organizations sort of like that. And it's been fairly successful at the early.

Ms. Welsh: Yeah. Thank you. I want to, in a few minutes, give the audience, in person and online, the chance to ask questions. And I also want to give panelists a chance to comment on what they heard from other panelists. But I have one final question for you first, Amen. When we were chatting you mentioned something about this concept of move 37. Can you just explain what that is and how it is that it informs and inspires your work?

Dr. Mashariki: Absolutely. So there's a really amazing documentary called AlphaGo. So many of you familiar with AlphaGo, where Google DeepMind developed this model called AlphaGo that competed against Lee Sedol, which is the world's greatest go player, right? And we can get into the complexities of go and so on and so forth. But this was really an interesting competition where they competed in five matches. The second match in the 37th move, AlphaGo, the model, did a move that made everyone who was watching either laugh, made them confused, or, in the case of Google DeepMind, concerned that the model wasn't functioning correctly. (Laughter.)

Because they ultimately did research and found out that move 37, there was a one in 10,000 chance that a human would have actually made that decision to make that move. And Lee Sedol himself was just so confused. He was just, like, what is going on? Why would you ever make this move? Later, come to find out, everyone saw that move as sort of genius and beautiful and elegant, right? And so this was the first instance in which we saw AI actually be creative. Now, Demis Hassabis sort of posits – some people disagree – but posits, that's the instance in which, you know, what we refer to as hallucination, where models are trying to be creative but just haven't really gotten into that space of being creative and so you see sort of hallucinations happen. But this was a real instance in which it was quite creative.

So just take that concept. And it's this concept of without AI we would not have made that decision. So let's take in restoration, right?

Biodiversity, let's take in restoration. Imagine you have the world's – a room full like this of the world's leading restoration Ph.Ds., and conservationists, and so on and so forth. And imagine an AI model making a suggestion to that group, and everyone in that group saying: That is the dumbest idea we've ever heard. We would never do it under any circumstance. And then three to six months later, that same group saying, huh, that's actually quite genius, right? And so at the Earth Fund we think about move 37 in the way we think about what are decisions we can make with AI that we would have never made, had we not made AI available to the grantees.

Ms. Welsh: Yeah. Well, thanks. Thank you. Thanks for sharing that. Panelists, is there anything that you'd like to comment on that you've heard so far? Including Olivia, of course. If not, I might have some follow-up questions. But I saw you nodding. I saw – OK. (Laughs.) I don't want to put anybody on the spot.

Ms. Graham: I can chime in.

Ms. Welsh: Oh, great. Thanks, Olivia, yeah.

Ms. Graham: Yeah, happy to. First of all, the AlphaGo documentary is really worth a watch if folks haven't watched it yet. It's very moving. (Laughter.) And it sort of distills, for me, this feeling of what is possible now that we're at the frontier, right, now that we're opening up new areas of discovery that wouldn't have been thought possible only a few years ago. So I just wanted to say that I love the way that that was teed up. I thought that was very beautiful. And I appreciate the comments.

Ms. Welsh: Thanks, Olivia. Thank you. This might be a good time then to move to questions from the audience, in person and online. I think I'll take a couple at a time, actually. I see one hand and then, yeah, if there are questions online. Great. I actually see two in the room. So we'll start here in the middle and then we'll take a question in the back.

Q: My name is Abhishek. And I'm a lawyer as well as an engineer.

I have a question for Olivia. That when you are developing these agentic tools, you know, like all the AI models are prone to hallucinations. And in the decision-making process, if you have hallucination at one prediction point it can cascade through the entire decision-making process. And the losses or errors can cascade. So, like, what are the – what are the mechanisms or the solutions that you look for when you are deploying or developing these products, the agentic tools specifically, so to avoid the indemnity issues, basically, for the losses that can cascade? Thank you.

Ms. Welsh: Thank you. And we'll take a question in the back.

Q: Thank you. It's kind of similar, but I'll take a different tangent. Is just this tension on kind of generative systems versus deep evaluation. If you have something that's generative, it's going to come with something new. And obviously if you have the data that backs it, you can go and do a prospective valuation, in the same kind of move 37. But if you're moving into fields that significantly are data scarce how do you go about this tension between both of them, especially as AI systems become bigger and bigger and kind of more ubiquitous?

Ms. Welsh: Great. Thanks to both of you. Olivia, you're welcome to take the first question, and then open to the panelists for the second question.

Ms. Graham: Sounds great. Yeah, thank you, Abhishek, for the really thoughtful question.

One thing that I will flag is, although I work on AI I don't work on large language models and I don't work on agents. And there's this whole field of AI that is asking, you know, how do we take what previously was only done with physics and train models to interpret large amounts of data to get strong predictions? And in those kinds of models, it's not like LLMs that spit out a bunch of text. There is a quantifiable answer, at the end of the day. So we can back test. We can say, look, we're only going to train our model with data up until 2023. And then we're going to see how well our model would have performed for 2023 through 2026. And that gives us the confidence to see how the models are performing.

I think this touches a little bit on Paul's question, too. This idea that we have to be careful as we deploy. We have to ask, is this actually meeting a need? Is it solving the problem in an effective way? Is it delivering value to its users? And as it becomes easier and easier to generate answers, the burden falls to us to ask, are these good answers? Are these answers that we're proud of and that we want people to be receiving?

Ms. Welsh: Thank you, Olivia. Any other thoughts on the first question, and if not happy to move to Paul's good question. It was in essence around AI being used, both for generative purposes and for evaluative purposes. Is that – can you – can you summarize your question again one more time?

Q: Sure. So the question is, can you – generative AI is going to be good for bringing out new things, but it's harder to evaluate. So how do we resolve that tension as we deploy it more and more. Yeah.

Ms. Welsh: Crystal, do you have thoughts there?

Dr. Huang: I didn't catch the last part. As you something more and more?

Ms. Welsh: As you use it more and more.

Dr. Huang: As you use it more. No, I think that's a great question. And that's something that we're definitely grappling with in creating this evaluation playbook. And it touches on the point that I made earlier, which is that Gen AI is changing, like, so fast, right? The underlying foundational models are changing. It's getting better with more user engagement. And so how do you think about evaluating a tool that changes? Like, a lot of the assumptions of RCTs, where you have, like, a static treatment unit, right? It doesn't hold anymore with AI.

And we touch on this in the playbook. There's definitely – so the other thing I forgot to mention is that this is a living playbook. This is not like a one time, this is it, these are the methods. Like, these are – the playbook is meant to be updated as we do more evaluations, as more novel tools, and techniques, and measurement innovations come down the line. So, yeah, I guess, like, some ways of thinking about how to adapt, maybe, like a conventional RCT tool to AI is there's been some exploration of using, like, surrogate indices. So using kind of intermediate proxy outcomes, maybe some user-generated data. Like, if a user is interacting in a way that is more engaged, like, does that predict, say, like, learning outcomes? Admin, backend data. Like, how much do those kind of surrogate indices predict outcomes? This is still a pretty open field of inquiry.

Other suggestions are basically just a timestamp. You know, at the minimum, all of your – all of the versions in the model that – the versions of the model that you're running an experiment with. Maybe creating a holdout group of users that's not getting – that's getting just the base version zero models that you can compare effect sizes thereafter. And then really using, like, rapid testing. So maybe you want to do an impact evaluation first to determine base impact, and then you do like A/B testing on these, like, user engagement outputs as you iterate on model improvements. And that gives you some directional evidence that you're going in the right direction and that your input – your impacts are compounding.

And I was just speaking with Mike from Dahlberg who was telling me about these, like, synthetic personas, which is something that I hadn't heard about before. And that might be another way to kind of reduce

data collection costs. It doesn't replace primary data collection, but you might be doing more rounds of data collection with, you know, the less cost. So those are just some thoughts.

Ms. Welsh: Thank you.

Dr. Mashariki: Caitlin, mind if I – just to give an additional, because I really appreciated that. To give an additional sort of perspective. At one point in my life, I was the chief analytics officer for New York City and ran the mayor's office there. And we used data analytics and data science to sort of solve operational challenges around the city. And over time, we got to a point where – and this is – we didn't come up with this term, but I used this term a lot when I was talking to agency heads, which is – because at the time, you know, machine learning and data science was really new. This was 2014. And so all models are wrong. Some models are useful, right? And so when you think about how we think about these things at the Bezos Earth Fund, at the core of the evaluation is if you were trying to mitigate against deforestation in a particular region, the evaluation is on were you able to be successful there?

I think there is conversations that need to be had around evaluating models, and AI, and so on and so forth, and really digging in. But at the core of the work is, what was your intended end goal? Because looking for perfection in those models, you're going to be there for a while. And that's going to be an iterative process. But if you're – like I said, all models are wrong. Some models can be useful. And if we can get to – if we understand the problem that we're looking to solve, and we can move the ball and ameliorate a lot of those challenges, and move the ball in the ways that we want to, I think those are things that – you have to understand those tradeoffs. Lean in to model evaluation and AI evaluation, but make sure you have your eyes on the prize in terms of why you're doing that work in the first place. And that is to mitigate against deforestation in the Amazon. And the question that we will always ask is, did you do that? We could care less about interesting numbers in a spreadsheet around the efficacy of the model.

Ms. Welsh: Thanks, Amen. Thank you.

Dr. Mehta: I'm happy to chime in.

Ms. Welsh: Please. Thanks, yeah.

Dr. Mehta: When it comes to evaluating large language models, I think – so there are three points I'd like to make. So the first is that we've seen rapid saturation of benchmarks and evaluations. And I think what we really need to figure out is a way to iterate those evaluations faster so that we

have things to measure and we don't have evaluations that are saturated and don't provide useful information about models.

The second, which is related, is we need to figure out how to keep models from cheating. So if you have a dataset that includes the evaluations then the models can optimize for the data that they have instead of trying to measure what we're actually trying to measure, which are the capabilities. And the third is that we do a lot of testing of models in sort of what is functionally a laboratory setting, and not the way that those models are used in real life. And so trying to figure out better ways to do testing that is closer to real world settings is also going to be important. And so moving away from abstract or sort of theoretical benchmarking will be helpful.

Dr. Huang: Can I just say one more thing?

Ms. Welsh: Please, yeah.

Dr. Huang: So I just find it really interesting that the way that we all answered that question really kind of belied our disciplinary backgrounds, right? Like I was talking about evaluations as, like, RCT, as in impact evaluations. And I think the question maybe was maybe more intended about benchmarking and performance evaluations, or whatever, maybe, like, spanning the spectrum. But just kind of goes to show that it's important to have dialogue across the spectrum of what an evaluation could look like.

Ms. Welsh: Yeah. To your very, very first point, actually. (Laughter.) I think we have a question from online.

Q: Thank you, Caitlin. So a question from our online audience is: What does meaningful local ownership of these AI tools look like in practice? And how could it be measured?

Ms. Welsh: What was the second part of the question, in practice?

Q: And how can it be measured?

Ms. Welsh: How can it be measured, hmm. Meaningful local ownership, what does it look like in practice? Actually, Olivia, can we start with you?

Ms. Graham: Yeah. I can share some thoughts on that. What I love to see, what really warms my heart, is when we do some kind of foundational scientific research, we publish a paper, and we open source a model, and we share a dataset, and people then make that their own. They fine-tune the model for the specific circumstances that they have in mind, for the use

case that matters most to them. They do an evaluation that we wouldn't have thought of because maybe it's a problem space that we are not familiar with. And so when people with the know-how – when people who understand the problem better than anyone else also have the know-how of how to improve these systems and benchmark these systems and validate that they're working well in situ, I think that's, again, where magic can happen.

And so we see things like Gemma, for example, our open source large language model, being used by people all around the world. Our weather models can be run on an individual laptop rather than on a $100 million supercomputer. And so a lot of this technology is now more accessible than ever. And it's incredibly exciting to see what others are able to build on top of that foundation.

Ms. Welsh: Thank you. Thanks, Olivia.

Dr. Mashariki: I would just say that –

Ms. Welsh: Amen and then Judy, yeah.

Dr. Mashariki: What we've seen in terms of meaningful local ownership – and I love that word, "meaningful" – we've seen NGOs really stepping in, in that space, and partnering with and engaging with indigenous communities locally. So an example would be the Nature Conservancy did a project that we funded where they were in the Pacific Islands. And they're tracking vessels – they're not tracking vessels. So on these vessels that are fishing, these fishing vessels, it is illegal to haul in certain type of fish and certain amounts. And so what would happen is the captain of those vessels is charged with, literally, pen and paper, kind of taking notes of what fish fishermen are bringing in. And if – you know, if it's on the list, hey, you got to compel them to put it back.

Well, when you're out on a boat in the middle of nowhere you can kind of slip the captain a couple of dollars and then that comes right off the ledger, right? And so what the Nature Conservancy did, was using edge AI technology from NVIDIA and small models, was put these cameras on the boats that are tracking 24/7 fish that are coming in. And then it juxtaposes what the AI saw and what's on the ledger, to make sure that we're keeping true with the laws in that area. And so this is the Nature Conservancy stepping up and playing a key role in supporting the AI ownership of these respective communities, which are this sort of conglomerate of communities in and around that area that agreed that we will do this to protect the seas, right, the fish. So that's where we've seen it meaningful in terms of NGOs really stepping up.

Ms. Welsh: Yeah, thank you. Thanks for that example. Judy.

Dr. Chambers: Yeah, I was going – I'm having a déjà vu moment, because this used to come up with the biotech space all of the time. And it's really resulting from the fact that, you know, we have concentrations of power with the technology, both, you know, within companies as well as at the government levels. And they're all – you know, there's tensions around whichever way you look at it. And so one of the ways that we try to create a more democrative space for the technology in the biotech scenarios is exactly what you were talking about. Which is to have use cases. In this case, the U.S. government funded many use cases for the technology. And that meant that, you know, some of the proprietary issues were somewhat off the table. People had a much better feel for the technology. They weren't so concerned about the risk because they could see the benefits firsthand on issues that were of important local context for them.

So I think doing something along those lines would be very effective. The difference in the risk scenarios is a little bit different between biotech and AI, in that there's more personal familiarity with the technology. I mean, people are walking around with it in their cellphone, you know? So we didn't have that situation with biotech. But we should take advantage of the fact that people do have a more intimate relationship with this technology right now, and capitalize on that to make sure that we don't go down this very risk averse, you know, precautionary attitude about the technology.

Ms. Welsh: Thank you. Thanks, Judy. I'm going to take the prerogative to ask one final question of all panelists. And I'll just go in reverse order, actually. So if we were to gather here in 2027 to discuss the same topic, applications of AI-enabled technologies for food security, what's one thing that you hope would be improved by, you know, a year from now, in each of your realms, to better enable the uptake – use of AI-enabled technologies to benefit food security and minimize risks? In each of your realms. Hope that's clear. And these can be really short answers. It can just be like one thing or, you know. So, you know, 30 seconds or so. We'll go Amen, Judy, Aalok, Olivia, and then Crystal. So I'll start with you.

Dr. Mashariki: Yeah, really quickly, I would say that one of the things we didn't account for was access to compute. Even though we funded each organization $2 million dollars, a large amount of that went towards AI expertise and compute. And then they still need more compute. And so I think a sustainable – and one of the things we don't want to do is contribute to a lot of the complexities around growing datacenters and its impact on the environment. And so if we can be more thoughtful and more – and

have a more – how we provide compute can be more sustainable. And I think that would be something that would be useful.

Ms. Welsh: Thank you. Thank you, Amen. Judy.

Dr. Chambers: Well, I hate to say this with a donor sitting right next to me – (laughter) – but more funding for the policy research component of this and capacity building at the policy level. Sometimes I feel that that is sort of, like, left on the table. And it's like, you know, the stepchild of the funding community. (Laughter.) And so I think it's so critically important. I mean, even now, 50 years after the introduction of biotech, look at the map of Africa. So few countries have actually adopted it. And what is the rate limiting step? It's their policy situation.

Ms. Welsh: Thank you. Thank you. Appreciate it. Aalok.

Dr. Mehta: So in two years, I would hope that we have better, more accessible datasets. I would hope that we have figured out a way to share lessons from use cases amongst each other. And I would hope that we have advanced in our ability to do scientific experimentation so that we're able to focus research dollars on things that are high impact and have a high likelihood of achieving the kind of result we're looking for.

Ms. Welsh: Awesome. Thank you. Thank you. Olivia and then Crystal.

Ms. Graham: Yeah, again, data. I hope that we live in a world where these foundational datasets that unlock the future of research are available to the ecosystem. I hope we have a shared understanding of what the gaps are, right? Because whether that's, look, large language models aren't performant enough in local languages, or actually we need better weather forecasts. Well, what is the need and how can machine learning models and AI models help to solve for some of those needs? Recognizing that we have to work hand in hand with the rest of the ecosystem to actually deliver impact.

Ms. Welsh: Thank you so much, Olivia. Thank you. Crystal.

Dr. Huang: Yeah. I hope that in two years we have a solid evidence base of what works and what doesn't, and that these, you know, pieces of evidence actually drive real funding and implementation decisions, so that we're ultimately scaling up sustainable and equitable AI solutions.

Ms. Welsh: Yeah. Great. Thank you. Thank you so much. Great answers from everybody. What an excellent and wide-ranging discussion, exactly what I was hoping for to cap off our day. This brings the final plenary panel to a close. So, again, thanks to all of my panelists for a fantastic discussion,

and for telling us, essentially, our marching orders, what it is that we need to do to move from inspiration to action. So Crystal, Olivia, Aalok, Judy, and Amen, thank you, again. And this brings our – (applause) – thank you.

This brings the AI for Food Security Forum to a close. Today's conversations, I think, have been informative, honest, and even sometimes controversial, which is exactly what we were hoping for. The stakes of food insecurity are high and the potential of AI-enabled technologies is vast. So we hope that today's discussions will leave all of you better informed and better prepared to use AI for the benefit of food security.

And before we conclude, I have just a short list of very deep thanks. To my own team, the CSIS Global Food and Water Security Program, today's conference was made possible through your constant dedication and hard work. So really, when you see my team members, please thank them at our reception. (Applause.) Also in the very background is CSIS streaming and broadcasting, who are making all of these remote conversations possible and also making this viewable to people all over the world, in the room, everything. So our streaming and broadcasting team, I think, is the best in the business. So thank you so much for your work.

To Google.org, thank you, again, for your partnership for today's forum, for supporting my team, and creating a community of experts who can do this important work together. To our audience, thank you for joining us in-person and online. And finally, to the CSIS catering team, who's also the best in the business, I think, who kept us well fed and well caffeinated, and who will host the reception that will begin now. So thank you, everyone. (Applause.)

(END.)