Raw file

Workshop on Embodied AI and Multimedia Technology Standards

October 10, 2025,

14:00 to 18:30 p.m. CET

Services provided by:

Caption First, Inc.

P.O. Box 3066

Monument, CO 80132

www.captionfirst.com

>> MODERATOR: Ladies and Gentlemen, we are going to start soon. I invite you to take your seat.

(Pause)

>> SEIZO ONOE: Okay. Ladies and Gentlemen, good afternoon. Very pleased to initiate this workshop with you. The topic is very, very exciting, embodied AI. We will have a demo just at the end of this first session.

But before we kick off any of the ITU events, we are always very pleased to have our director with us to provide an opening speech.

Thank you for removing the echo.

So from Kigali Rwanda where the director is enforcing cooperation with the General Assembly, I'm very pleased to invite the director from remote for his opening remarks for this workshop.

>> SEIZO ONOE: Thank you and hello, everyone. I wish I could be with you in Geneva.

>> MODERATOR: Just a moment, we cannot hear you.

>> SEIZO ONOE: Can you hear me?

>> MODERATOR: Yes, we can. Please go ahead.

>> SEIZO ONOE: This exciting field of innovation raises key questions for technology experts as well as experts in governance, economics, social well-being, and much more.

To achieve trusted AI and realize its true value, we will need a wide range of [?]. Embodied AI was in focus at our recent AI focus world group summit. We discussed technical challenges and standards priorities, but also the social and ethical [?] of humans and robotic -- robots working together.

We identified real need for differential and taxonomy, interoperability and [?] benchmarks. And here standards will be essential.

We also noted the importance of social, technical standards with standards developing on certain technical questions as we are a human rights, accessibility, and environmental sustainability.

Published [?] robotics in areas from [?] conversational robots to autonomous driving management, we are now working on new standards to provide framework for embodied AI and benchmarks for capabilities such as perception, decision-making, execution, and safety.

This workshop has been arranged by the [?] driving this work. Our Expert Group for multimedia content and cable TV, ITU Study Group 21 is responsible for [?] AI and apparently there are 17 more. These standards on various topics including [?] systems, assessment criteria for foundational models, AI-generated multimedia content, AI platforms, and agentic AI.

And at this meeting of SG 21, we aim to complete our new standards providing comments and framework for embodied AI.

I thank our SG Chair and everyone involved in making this workshop possible. All of our speakers, moderators, and everyone participating in Geneva and online.

I also thank the AI Industry Alliance for sponsoring our coffee break and for the robotics demo during the break.

As we gear up new standards work, open discussion is so important to help ITU's expert define the priorities. That's what this workshop is all about. And I wish you every success.

Thank you.

(Applause)

>> MODERATOR: Thank you very much for being with us from Kigali Rwanda today. We appreciated you opening this event. It's a very important event especially from the standards angle of the Study Group 21 on AI which the study group is taking. And it's taking that especially in Question 5 that as you mentioned developed standards related to AI and has about 70 open work items which are draft standards being discussed as we speak in this week and next week during this Study Group 21 meeting.

We are pleased to have with us the Chair Mr. Noah Luo that will provide closing remark this evening at the end of the event.

At this point I'd like to stop the opening ceremony, thanks again the director Seizo Onoe from Kigali to come to the podium and to start the first session. Please, I don't need to name all of you, you know who you are, please come on the podium, you'll have your name plate with a seat right now all together.

I will just introduce you the moderator, Mr. Yuntao Wang from China who, as I said, is a very pivotal person because he's the rapporteur of Question 5. So I can fairly say is a very thorough expert on the topic and very pleased to leave the floor to him. Just let me give some little guidelines.

We have many speakers all today for the three sessions. We don't have much time. So we have limited to 15 minutes time for each speaker to give their presentation. We have a timer at the end of the room that everyone can see.

I would strongly recommend speakers to think before they start what message they want to deliver and deliver it up front so that after the 15 minutes if they're still talking, the moderator might cut them off.

So please, think of the key message you want to deliver so that at the end of this session there might be some time for Q&A that will be moderated by Yuntao. Thank you very much.

The floor is yours.

>> YUNTAO WANG: Thank you for this opening and I'm humbled to be here. This is the biggest theme ever for me as a chair in ITU.

I think our session 1, as we can see, is the title enable AI, exploring the intersection with multimedia services and emerging use cases.

Actually at the end of this session there are two key questions we want to answer. The first one is what is embodied AI and why it is important for ITU.

And the second question is how those EAI could be used in vertical industries, especially in telecommunication and ICT industries.

So keep in mind we have those two key questions to be answered in session 1. We are very glad to have six speakers for session 1. So I think we can start with the first speaker. It's Mr. Wei Kai. And also the Secretary-General for the technical committee on artificial intelligence from industry and information technologies in China.

Let's welcome Mr. Wei Kai for the delivery of his presentation. What is embodied AI and why it matters to ITU.

>> WEI KAI: Distinguished Colleagues and others in Member States outside ITU, it's my honor to have first speech.

So let me start with a simple thought experiment. Imagine teaching a machine, not only to think about -- not only to think, but also walk into a room just like this. And scan this room, update and collaborate with all of us here.

This is more than artificial intelligence. This is more than robotics. So this is the beginning of an embodied AI, and it's aware of the future of the intelligence is truly been shaping.

So it's my great honor to speak with you about the what is embodied AI and why it matters and how ITU can lead its standardization.

So in coming 15 minutes, I will elaborate my presentation in three questions -- to answer the three questions as the moderator raised, what is embodied AI, and what the challenges ahead of us, and what the standard gap and what the ITU can play in the industry, the leading role of standardization work.

So when we look at the involvement from traditional robots to smart robots, and now to embodied AI, we see a fundamental shift in capability, in intention.

Early robots such as Shakey in 2016 could only follow strict instruction written by programmers. And they could not truly understand.

Later on, smarter robots like AIBO and ASIMO had the sensor capability allow them to react to their environment. But still within pre-defined rules.

So embodied AI goes far beyond this. It combines sensory inputs, exclusion ability, cognition collaboration, and continuous learn in single integrated system.

So it is not -- it's not just doing what we programmed, it is deciding what to do based on its own understanding of the situation in the real world.

In other words, we are moving from machines that we got control to machines that are autonomous.

Some people called embodied AI with humanoid robots. Many humanoid robots are example of embodied AI. But core idea is not about machine, it's about embodiment. It's about having a body that can act in physical or virtual world and a brain that can learn, adapt, and collaborate.

This is the integration of embodied intelligence and what makes the next generation of the AI system.

So to truly understand what is embodied AI, it is helpful to compare it with other -- other things. First, I want to compare the traditional non-embodied AI such as ChatGPT or some other image defines.

So these systems are described to brain in vat, they exist purely virtually as a software running on servers and waiting for passive -- for inputs from the task. So their intelligence operates entirely in a digital domain.

Embodied AI, on the other hand, is an embodied -- embodied brain. It is an integrated into a physical or virtual body that can sense the environment, make decision, and take action. Instead of passive processing data, it acts proactively in the real world, navigating from a room, [?] objects, addressing human driven car in traffic.

So this shift also change how much -- how statistician learn, traditional AI learns from static dataset.

But the embodied AI learns through interaction much like a child exploring the world. It learns by moving, touching, falling, trying. So in summary, embodied AI equals AI physical body with action and experience and online learning.

So that's the comparison with the AI and -- traditional AI and embodied AI.

And another thing I want to compare is the embodied AI and robotics. I think the three different aspects we can see from the two different -- two concepts.

First from a programming to learning. So traditional robots operate based on coding by engineering. They can acquire abilities through real-time interaction within the environment by learning.

Second, from perception for action to perception cognition, action integration. Classic robot use sensory data only to exclude pre-defined tasks.

But embodied AI go further. They contact internal model of the world, and understanding that this is my messy room, recognize objects and priorities. So this cognitive layer allows them to risen before action -- before acting leading to more intelligent [?]

Third, traditional roe bats have a fully [?] environment. Real-world application rarely offer such certainty, embodied AI on the contrast is designed to operate in dynamics and structure and certain environments.

It must handle and see objects adapt change and deal with ability and deal with ambiguity.

In this sense, embodied AI equals robots plus brain with cognition ability. So it enables robot to move beyond automation to our true autonomy. It will not only to exclude tasks, but understand, exclude, and improve over time.

This is the next generation of AI system. What is embodied AI? The transformative nature is recognized by leading institutions worldwide from attributable IEEE SMC and the society, they define it as system that integrate cognitive abilities with sensory and action capabilities integral machines to perceive reason and act autonomously in dynamic environments.

In Stanford University and others give their definitions to this concept.

So with a broad [?] of the existing definition proposed by different parties, we abstract the key capability of embodied AI. The first thing is cognition. The ability to reason and understand the context.

The second capability is learning. The capability of continuous adaption. Adaptation and improvements from experience.

The last one is collaboration. So the agents is cued to worked synergistically with other humans and environments. These are not isolated functions, they are interconnected, integrated, and enable systems in complex real-world settings.

So what the -- what can we define the term of the embodied AI. Study Group 21 gave their answer in recognition of this importance ITU-T Study Group 21 has been actively working for a common understanding of embodied AI.

Our draft of new recommendation F.RF-EAI requirements and framework for embodied AI provides a framework to understand this content.

So we define this embodied AI and we gave the conceptual backgrounds and key features in this draft of new recommendation. So that's our answer to have a common understanding of this -- this concept.

So here is a one-minute video to show some application of AI in the real world. So I don't know how to -- how to play the video. Yeah. Yeah, yeah, yeah.

So as we can see from the video, there are device application AI is not confined to a single shape or function. From humanoid robots performing household chores to [?] robots dogs carrying out inceptions. AI is truly diverse and variety forms.

This diversity is a strength but adds complexity for development and either for standardization landscape.

So there are many applications in real world the AI can feed. And the suitable shape is designed for those scenarios.

So the market potential is staggering. Some framers, organizations point out to the market size is very big. And the penetration is very fast in the near future. And the U.N. agencies also estimated we are facing a global societal needs as early -- elderly care increases and the labor shortage.

So the EAI is fit global societal needs. And also embodied AI is widely considered a promising AI development for AGI.

There's a Chinese proverb read ten thousand books, travel ten thousand miles.

Traditional AI has excelled at reading ten thousand books, having vast amount of training data, but to achieve the true intelligence in future, a system must also travel ten thousand miles to learn from embodied experience in the real world.

So this concept of embodied learning is about rethinking AI -- AGI road ahead.

This all sounds incredible, promising. However, the path to the future is not without its obstacles. This brings our second part of our challenge.

So Moravec's paradox is still valid for EAI. That means the AI system can win, can beat the human on very complex intelligent task like checkers. But it's difficult or impossible to give them the skill as a 1-year-old when it comes to perception and ability, such as moving -- navigating stairs or open the door, remarkably difficult for the robots.

And the challenges we are facing to advance the EAI technology and application, technically, theoretically, applicationally, and societally, governance, so many challenges are add if you look at -- if you're interested, there is some of the survey you can read.

So overall, EAI's challenges need the technical -- [?] innovation as well as standardization effort. ITU can play a critical role.

So to my team from CAICT identified seven key areas where it's needed to ensure a responsible development of EAI. I don't want to go in detail of that. It's now exclusive the standardization gap there.

I think maybe for example the data perception is needed to -- to standardize and networking and collaboration between different EAI agents need to be stand up and benchmark evaluation should need standards -- for example, at least but not exhaustive.

ITU plays a very important role in EAI development. I think already ITU play already form very good international open community to discuss this kind of very important thing. And collecting use cases and study the cases from all over the world.

And also play a very unique standardization work in U.N. agencies.

So ITU -- ITU-T starting contribution is there. We have already started a new work item and we are thinking about new things this time and for the coming -- for the coming meetings.

This time we are going to consent recommendation framework on embodied AI and propose the framework to define the capabilities and the building blocks of the system.

So last one slide that in summary, EAI is promising a pathway to AGI and we are facing significant challenges. Some of them can be effectively addressed through standardization.

ITU plays a key role in this effort with its Study Group 21 serving as an open, collaborative platform tool advancing in this field.

Looking forward, we recommend continuing workshop discussions to establish -- or to establish a focus group on this issue in future, to explore the opportunities for EAI for good, and to foster standardization studies inclusively, collaboratively globally.

Thank you very much.

(Applause)

>> YUNTAO WANG: Thank you, Mr. Wei Kai for this excellent presentation. I believe because it's a very important question

to be answering in this session, a little bit over time seems okay here.

So I think we go directly to the second speaker. The second speaker is Imad Elhajj, the professor from department of electrical and computer engineering AUB.

So his title would be embodied AI embedded with humans. Very interesting.

>> IMAD ELHAJJ: Thank you. Thank you for the invitation. I'm glad to be here. Faculty members, we talk in chunks of hours, but I'm only given 15 minutes so I'll try to make it.

We've seen this -- all of us in different forms expressed how AI is beating humans in different games as a benchmark. We've seen the gap close between the use and the complexity go up as far as AI beating humans.

We've seen this being a broad scope in many different fields. It's very difficult to think of a field that AI is not involved in.

And there's many indicators that show us that growth and obviously the industry's excited about this because of financial implications. We see here industrial robots installations, the curve going up, you know, very steeply.

And this is kind of what I want to focus on today. Is why do we care about this growth and this broad spectrum of applications that we're seeing?

And these stats from two different sources saying the same thing would highlight that kind of a concern that I'm pointing out, which is household robots market size.

So this is not anymore about industrial robots being caged or in zoned environments with trained individuals working with them. These are robots roaming our homes, kids, families are there, and you can see the numbers that the projections that this is going to be more and more.

And this is the ultimate thing that in the robotics world old enough to remember that we've overpromise and underdelivered long ago. Now we're hoping we don't do this again in robotics field. And the slide that my colleague previously showed that robots are still at some tasks failing miserably and AI is picking up.

And I tried sometimes disconnect these two by the way from each other, robotics and AI, they're not necessarily always the same thing.

And so why am I saying this?

Because now we have robots really alongside humans. We have the issue of trust, liability, how do we mitigate technical failures and bias that we keep hearing about. What about the unstructured environments that they have to work with.

And so there is a need for high-level intelligence to be included in this discussion. And the issue of supervised learning coming in.

And so if you don't remember anything from this presentation, if I don't get a chance to complete it, as the Chair of the session recommended, this is the slide to remember. This is the key message today for us.

What we're proposing is that humans should be in the loop. Humans should be the center of robots, networks, sensors, and AI is with an S. Agentic AI, embodied AI, whatever we want to call all of those, those are just a bunch of AIs on top of each other that we are also now interacting with just like we used to do with the networks.

So what loop we're talking about, those of you in control field know exactly what I'm talking about. You've seen this slide before. This is the feedback loop that we traditionally had with sensors.

Then back in the '80s and '90s, teleoperation came about and said let's operate robots over the Internet. Now the Internet allows us to do this remotely, very exciting applications. We overpromised, underdelivered there also.

And then AI came into the loop and said, let's replace the humans with AI agents that could do this better than humans do.

What we're advocating is actually to add these together. And what we end up with is something like this. So we want to look at systems that senses AI, humans, controllers, all of those are connected together through network or networks to enable us to do more interesting things and safer than other ones.

So some of the opportunities that I will point out and some of the projects that I would highlight, just to give you an idea of the things that we've been working on for the past few years. I think you've seen this video on YouTube, some of you. If you've bought a drone, you could relate to this because most of us break the drones in the first couple hours that we have it because it's very difficult to control.

It's several degrees of freedom, you need a bit of training and it's not intuitive.

So the question that came to our mind is, what if robots understand and guess and know what a human is trying to do. So understand your intention, and suggest to you that intention and complete it. Just like we do with auto complete on your keyboard. Now as you're typing, even when you're replying email now, it does it on the whole email. It used to do it word by word, phrase by phrase. Now sometimes when you're replying, AI will recommend and suggest the whole reply to you.

And what we're suggesting here and advocating for is that the robots could do this. As you're controlling robots, a robot could say, oh, you're trying to go in a straight line or a curve, do you want me to complete that for you.

Now the challenge is that we've demonstrated that could be done and we've demonstrated that performance could be improved. But the issue of generalized issue of AI. Another AI comes in, behavior is different and AI fails and that doesn't do well.

And we build systems where based on the user accepting or not accepting the suggestion from the AI, the AI actually learns from that. And says, okay, this is a different profile, I need to do online training. And by doing that, we've shown that after a relatively quick period of control, the AI will adapt to the new user.

I think this is very important to build in our AI systems is online training. They need to be able to adapt to different users so that we can at least mitigate this issue of lack of journalizability that we see in several applications.

So you can see what I'm talking about here is a system where we have a robot that was flying in this case a drone, we're driving it. And the operator is given by the AI, a visual indicator of what it thinks the operator is trying to do.

And if the operator agrees, accepts that, then it goes into autonomous flying. And when the robot realizes that the operator wants to change course, it will give control back to the human.

So this is kind of a hybrid control proposal here where we are having the robot and the human codrive the system, sometimes autonomously and sometimes interoperation. But we've shown that improves performance drastically and extend it to 3D complex navigations like a building or any structure you can see in this simulated real-time operation of this drone.

As the operator is flying it around, -- the structure, the AI recognizes this is a cylindrical motion. And if they say you guessed it, the AI is not only able to complete the path, but complete the geometric measurements of that math. So it does that geometry to complete it as accurately as possible.

We've done human subject testing with different -- with systems with and without auto complete and we've seen that over and over again that such systems favor well with human operators on all NASA TLX indicators. We've seen that quantitative results also favor such approaches.

Average time of completion, average distance traveled, average smoothness of the path. All of the above tend to improve when we're working alongside the AI.

I'll shift tracks from the auto complete concept to something else is where actually the human is teaching the robot in a seamless way.

This is a robot walking on different terrains and trying to detect the terrain it's walking on so it can change the controlled strategy that it's applying. And we're using AR for humans to just by looking at the robot can understand the internal state of the robot, including what the AI is guessing.

And if the AI is wrong, the human can correct it. And with time, that improves the AI itself. And as you can see again, the human here is playing the role of supervision and playing the role of giving feedback for the embodied AI in real time and not needing to have to retrain the system all over again.

This is a bit of a long video, if we can jump ahead. I'm not sure if the -- excuse me. The lady can help us just jump it ahead forward.

This is -- no, just in the video, please, if you can jump like a third of it forward. Yeah, thank you. This is a new funded project where we did collaborative mapping between robots and humans. Yeah. I just want to see the visualization in the middle please. There, perfect.

And we've seen that we can improve the ability to map an environment in 3D by humans working close by several robots and we're able to do 3D mapping of the environment, correct the map, and be able to improve the digital map eventually obtained.

We're showing here the visualization from the operator's side using octo maps. This is the three D map built by the human and the robot and the human is able to see it and correct it.

The alignment in real time is better it's just when you capture on the device it shows a misalignment. And then the human can go into the map in real time and do the correction. An example of humans and robots working side by side.

So this -- let me go. I think I skipped a couple slides. One more slide back, please. Okay. Here.

So the XR agents and the human in the loop, building these digital twins, the idea is for us to be able to apply several AI algorithms also on the systems both on the XR device and on the robot to help us in developing more accurate digital twins.

We could see here the real time digital twin update. The digital twin is on the top and the real world is on the bottom. So this is a 3D map of the environment being synchronized in real time both dynamic and static objects are being you dated. Robots report on their location in real time and then if the human is walking around and through the XR device, they notice something had changed, those also get updated. Things like the chairs that you see.

Here we see 6D pose estimation. AI comes in and says I could detect this object automatically for you and snap it into place immediately. Or you do that manually by doing an overlay and suddenly you can see that snapping occur. And then the map is completely corrected in real time.

So this is -- these are 3D CADs or scans of an object that the AI is trained on to detect them. Then when we go to onna environment with an XR device, we're able to detect them and update that 3D environment that was created in the digital twin.

And obviously the digital twins we don't have time for it, but I think that's an interesting field for us, for example, here we show you temporal play back of the twin itself. We can navigate in time back and forth and not only in space and therefore in an industrial setting, which is where our partner Ideal Works is involved in, this would be very valuable for them to be able to replay the environment with 3D localization of assets.

This experiment was done in Munich and the Ideal Works facility. And we've done, again, experiments with human subjects for developing the digital twins with the AI involvement and without the AI involvement and using different modalities, using their own systems, using XR devices, laptops, note pads, and so on.

And we've shown the superiority of having XR coupled with AI for this particular application. Both on quantitative, objective measures like accuracy of placement, and the subjective results.

Again, the NASA TLX indicators shown that there's improvement in had human acceptance of the technology and the comfort in using the technology for this application.

I will fail if I don't mention the issue of privacy. Clearly when we're discussing all of these topics an issue of privacy comes in. We're very interested in real time demonstrability on mobile devices. What you see here is the algorithms. If we can play the second video, if you don't mind, the one on the right.

You can see here different algorithms where we select an object in the environment and we want to hide it. So opposite to what Zoom does now, it blurs the background, but actually blurring objects of interest, objects of privacy. In this case, in real time with AI on the mobile device front end.

So very significant concerns when it comes to processing and memory requirement.

In this case what I'm showing is AI UPTA algorithm which is AI watching an AI. So we've trained an AI to learn when it fails, under what conditions based on motion, lighting, and so on that it detects a failure in object detection. And in that case, that second AI will tell us that your AI that you're following is actually going to fail.

And in that case we blurred the whole screen, we blur the whole image just as a mitigation. I think this is maybe an underresearched domain is where we have AI working with AI to actually improve or at least give us indicators of potential failure areas or conditions.

I don't have time to go into this, but we've also worked on environmental applications where we've used AI in sensors for measuring oil so that -- can we play also the other video please in real time -- so that we can mitigate difficult conditions for measurements and instrumentation.

You can see the waves on the right and the dragging on the left. These are very difficult challenging cases for sensors measuring oil thickness that's floating on the surface of oil and AI there did a great job to mitigate these challenges.

Can we move to the -- yeah. And this is us working on the same application, also applying AI but we're measuring oil under ice noninvasively in this case. This is an instrumentation application versions the robotics application.

I know I'm out of time. Humans have to be at the center of this discussion. Current questions we're looking at is what's the role of physics and knowledge. What are the human factors involved and can we have human-like robotic AI, VLM roles, and the context. But when we talk about human behaviors, how do we even test it. That's a challenge. What about cultural differences, personal preferences, and explainability for the operators.

Because as of now, explainable AI needs explanation. Explainable, I don't understand explainable AI as intuitively as we like to show what it is for the outside world. So we need to do a lot on the issue of explainability.

Communication challenges, we here at ITU-T, privacy, bandwidth, processing, memory, availability, and heterogeneity, those of you who are old as I am will know these are not new challenges. Every application that comes out and uses the network we complain of the same thing. So similar good old challenges remain there regardless of the technology we're referring to today.

We don't want to compete with AI, we're looking to hopefully bridge this need to actually work together so we can get more market acceptance of some of these applications.

Thank you very much.

(Applause)

>> YUNTAO WANG: Thank you very much, Professor, for this excellent presentation. I'm very glad that human has to be in the loop of this one, because actually with the help of embodied AI, I think human has to be embedded already.

Okay, thank you.

Let's go to the next speaker. The next speaker is Kashif Ikram, the President for Europe, MicroPort scientific. Welcome.

>> KASHIF IKRAM: Thank you very much for the invitation. I'm representing MicroPort MedBot and I'm here representing Dr. Alex He who was, unfortunately unable to make it.

So my topic today's going to be about surgical robots where humans are not only in the loop, humans are also on the table underneath these devices.

So where did the need for telesurgery arise?

Well, if we take China, for example, where our robot was designed and built, the further east you go the more rural it becomes and less surgeons are available. The further west you go -- sorry, the further east you go, the more the population is and the more pressure there is on the surgeons to perform and the number of procedures that needs to be done is quite overwhelming.

Therefore, it was a huge need to drive efficiency, automation, and to bring AI into help solve some of these issues.

Which is what MedBot said itself to do.

If we go back to the original history of telesurgery, the first telesurgery procedure was performed back in 2001, 24 years ago. When is when Professor Jacques Marescaux in France performed telesurgery from France telecom in New York and his patient was in Strasburg in France.

He was using a fairly advanced technology for its time. The Zeus robotic system that was bought by the company, intuitive and incorporated into the current da Vinci system. The dance was 7,000 kilometers and the latency was 310 milliseconds, which was very inhibitive.

Also in order to do this, they actually shut down all other traffic on the line between the U.S. and France so that they wouldn't have any interference in the system.

Apart from some work that was done then in Canada, we had a huge gap between then and today before telesurgery, again, was brought to the forefront.

Initially at the beginning with what we call second generation telesurgery, we had a dedicated line, optical line from point to point, done in China to test the system and to do the first telesurgery in a situation where we tried to remove all the risk.

And this proved to be so successful that today if we take China, for example, we have multiple robots in multiple cities, in multiple hospitals interconnected where the surgeon in A can do a procedure on a patient in multiple locations.

We can even have two surgeons from two different locations today operate on a single patient as well. So you can do a tag team type of approach as well. These technologies are coming in.

So the advantages of this are quite clear. So when it comes to if you live in a remote area, now you have access to the top doctors. You don't need to necessarily travel to where those doctors are. And those doctors no longer need to travel to you.

There's a huge humanitarian aspect to this, providing health care into war zones, into famine areas, into disease-struck areas. And then we have some other fantastic opportunities as well.

I think we're taking this from telemedicine where the doctor can talk to you, sort of examine you to a certain degree, but here you can actually have the surgeon operate on you. And I live in Sweden, which is a very long country. If you flip Sweden on the map, the top of Sweden will be below Italy. And four major cities where all the health care experts have concentrated.

So if you live in the very north of Sweden and you're a minor, you have the same right to that health care as someone in Stockholm and with this technology we can provide that without you having to travel.

How does the technology work? In the operating room what you need is obviously the surgeon console where the surgeon sits and he operates the robot. The robot is the patient surgery platform which has four arms and four laparoscopic instruments that pass through four holes into the patient's abdomen or chest.

And then you have a tower which has the imaging processor, it has the energy device for cauterization, cutting, and sealing. And connected to that you have the remote communication workstation that sends a signal out, 5G or LAN to dedicated lines or fiber. And that's picked up by another remote communications workstation in point B where the surgeon is sitting at another console.

We're able to have achieved now is ultra low latency. Anything under 150 milliseconds is not perceived by the human eye. So you want to keep the latency below 150 milliseconds. So we have ultra low latency, a lot of it done by software how we compress the signal before sending it.

We have multiple network fusions. So we're compatible with current infrastructure in the hospitals, we've done telesurgery with 4G in some third-world countries and using dedicated land lines, the Internet, and we collaborate obviously with the Chinese mobile communication providers.

We have a low latency requirement as well in order to send the signal that we've compressed and sent out. And remember, we're sending out the instructions that the robot has been given by the surgeon. We're sending out the image backwards and forwards in real time, and we're also sending sound communication between the doctor and the nursing staff, between the nursing staff and the doctor, between the various machines that are sending information to the doctor like heartbeat, blood pressure.

So there's a lot of information that's passing backwards and forwards all the time.

So system and of course safety is very important. So we have dual end hot stand by. So any point a surgeon in the same room as a patient can take over instantaneously with the other console. So we have a zero-second dynamic switching should it be required. Should you get a catastrophic bleed and the surgeon needs to jump in, that's in the room with the patient, they can do so.

Of course, data encryption for safety. We're sending very sensitive information across the Internet and we want to make sure that that's encrypted. And no one can hack into the system and control the robot.

And then real time monitoring as well. Monitoring and handling of the network status so we know where we are in terms of milliseconds. And we can switch to a faster network or a faster route should it be required.

So in our work that we did, the first real telesurgery was done across China for 500 kilometers of distance, from Shanghai to Xinjiang. The latency was high, 220 milliseconds and the procedure was decapitating renal cyst. Relatively normal procedure.

That worked well and we continue and did 100 telesurgery cases in this study which were numerous, complex procedures such as gastrectomies, liver resections, pros indicate cancer, hysterectomy, bear yacht tick procedures, pediatric procedures as well. These are the most complex procedures you can do after a broad range of indications and number of different hospitals.

This proved to be, again, very successful using a 5G network cable.

For time I'm going to skip over that.

So basically the experimental stage was done and we were ready to go big time across the world. Telesurgery basically reached a point where it could be utilized across our world. And what's happened since?

In the last couple of years, telesurgery has exploded. We're at over 400 cases.

The first column, that should be America and Africa at the bottom. That's a typo. We've done cases in Europe, Belgium where they've performed telesurgery where the surgeon and the patient were in two different locations. They've done hysterectomies as well as prostatectomies. We've done from China back to Kuwait, multiple number of surgeries showing this can be done over and over again without incident.

We've done from the United States from Orlando to Angola. We have three robots in Angola and you can see the humanitarian effort there of bringing the surgical expertise of one of the biggest prostatectomy centers in the world to the people of Angola so they don't need to travel.

And of course we've done the longest telesurgery in the world which was from Shanghai all the way to Casablanca and the surgeon was in Shanghai with his patient -- prostatectomy patient in Shanghai.

So that's -- most of that was I'm going to say done in the last year. The Angola case -- we did a case in Angola from point A to point B within Angola between different buildings and the latency was we low 7 milliseconds, so there's a typo there, I apologize.

And again, this showed Professor Patel who is one of the leading surgeons in prostatectomy in the world that they could bring this technology to areas where these patients would typically have had to travel or would not have had access to this type of surgery.

And what's important here is we can up-skill the local teams as well.

That's the live surgery from Orlando that I talked about. Here's the route that was taken for the longest telesurgery in the world, from Shanghai to Beijing and the signal sent to Frankfurt, Paris, Madrid, and eventually Africa. More than 30,000 kilometers, straight line distance 12,000 kilometers, and the latency here if I remember, because I was in the room, was 172 milliseconds. So it was perceptible but the surgeon was able to compensate with that so that went quite well.

We took a look at the robot is [?] for example, the telesurgery equipment, the telecommunications equipment is C marked. But telesurgery itself is not C marked. So that's one example.

How do we go to the authorities now and get this type of surgery approved. We started with guidelines making national standards, we have obtained regulatory approval in China at the NMPA. We have an IDE-approved study going on in the U.S. And also here in Belgium in Europe.

And we're hoping to get the CE mark for telesurgery at the beginning of next year. Meaning that hospitals can do this without ethics approval from their local management.

We've published numerous articles on this telesurgery and spread this to a number of places around the world. And what we're trying to do is establish three levels of telesurgery.

The first being local. Within city this can be linking up hospitals that are part of a network typically in London, a London University hospital will have three or four subhospitals. We can link those up together. Maybe only one of those hospitals needs to be on call on any particular night.

Then we can look at from city country level we can go cross-border.

So we can look at hospitals that have multiple hospitals in multiple countries and link them up together. And then of course we'd want to look across the globe.

And what about not only across the globe, but we can look to the space station, this is the Chinese space station. We have the European space station.

We are looking at building a lunar colony on Mars. We're looking at the moon as well.

Last year, two astronauts flew up into the international space station. They were supposed to go up for a week. There as an issue and they were there for eight months. If one of them had had appendicitis, it would have been a death sentence. Now with this technology, we can look at those issues and hopefully solve them.

And also in my final slides, we're looking at low satellite and high satellite and we've actually proud to say that we've done the first cases in those. Here we've done a high orbit satellite case from Lhasa to Beijing. We had two liver procedures and we did low orbit around the world as well from Beijing to Xiongan. Which is 105 kilometers. The 16 minutes and 72 minutes with a 100 milliseconds delay which is unacceptable.

These are challenges that we need overcome, but the thing is we can offer multiple different ways of doing telesurgery regardless of where you are in the world. Which is quite exciting.

So thank you very much.

(Applause)

>> YUNTAO WANG: Thank you, Professor, for the wonderful presentation. We see this use case is very convincing and we see that communication is a vital key enabler in embodied AI.

So let's go to the next speaker, actually there are two -- there will be two speakers joining from remote. The first speaker would be Julya Rebstock from information governance practice manager, the John Caras. And the topic they are delivering is very interesting as also and very important is security and privacy aspects of embodied AI and robotics.

Please welcome. .

>> JULYA REBSTOCK: Thank you. First I want to say thank you to this body for allowing us to join the group today.

There's a lot of potential collaboration between our goals. There is so much information that we could go over. We're going to try very hard to keep it to the 15 minutes.

Next, please.

So there -- John, sorry.

>> JOHN CARAS: What did I do? Oh, you want me to talk, okay, that's fine. So sorry --

One of the things we did for this meeting, I'm going to speak rather quickly, is that we looked at the scope for, you know, Study Group 21 in terms of embodied AI.

And also because I come from Study Group 17, there are some overlaps. So we just want to highlight things like digital inclusion, active assisted living, human interfaces, digital health and culture and mobility, metaverse-related issues, and audio visual content, immersive multimedia applications.

I'm going to skip to the next page. For Study Group scope, Study Group 17 scope is on cybersecurity, security manage management, and security of ICTs and the data they transmit.

Here's a list of those. I think that pretty much -- I think many people in this room kind of know what Study Group 17 does. And I'm going to keep more focused on the actual presentation.

Go ahead, Julya.

>> JULYA REBSTOCK: So I just wanted to call out, I've heard a couple of other definitions of AI here, and I wanted to call the body's attention to a document that we have in draft. And it is an artificial intelligence security standardization strategy.

So there's a number of definitions called out in there, and we have provided a link to the document that you'll be able to go to. We can keep going.

And the appendix has key areas for standardization. Y'all are building the communications and there's a lot of talk about the hardware layer, I feel. But security is something that Study Group 17 is kind of focusing on. As well as from a privacy perspective.

Next one, please.

So ours is -- we're kind of focusing on the progression. So obviously generative AI brought in a lot of security concerns, privacy concerns. And it has always been a little bit more of a one-to-one if you think about the user experience, what is each individual user doing. They're going to be opening a browser and they are initiating a session with ChatGPT, for example.

But it's very one-to-one. The input is typically the keyboard or maybe voice recognition. But when you add that autonomy and permissions, and that's what I want to be specific about, you are now giving it authority to do something on your behalf.

So we kind of have moved into agentic AI. With agentic AI, it now is acting on your behalf. It's being, for lack of a better explanation, it's being a secretary. But maybe you've given it -- you want it to book travel and you've given it permissions to your credit card or your login ID for the airlines. You've given it your passport information.

Well now if we take that same model and we embed it with mobility and sensors, we're letting it interact with the world, again, on your behalf. So these physical systems now, like the health -- many health care robots that we've seen. But with the interaction between humans and the robot, you start increasing the risk.

So the escalating risks that we've got, number one, it's collecting and using a ton of data. But also with that, think about the interaction if the robot is in the room and there's three humans, how is it going to distinguish which human it's interacting with?

It has to take a picture or it has to have some kind of scan of each person so it can distinguish who it's working with. That's where we start bringing biometrics into the story.

With biometrics we have privacy, safety, and we have trust.

Next.

>> JOHN CARAS: Okay. For embodied AI interaction, embodied AI is effectively downloaded. The AI agent -- the AI model is downloaded into the system. And from what we've seen, there's actually multiple AI, I would call, embodied AI models inside of these robots. Some for audio, some for video, some for memory management.

With this, I always like to refer to this three laws of robotics. This just comes with the 1940s, has to do with the -- there was a trend from the 1920s to the 1940s about robots and the fear of robots taking over and things like this.

So you know, as Isaac Asimov came up with three laws of robotics. It's interesting because you find it a good correlation between literature or Sci-Fi literature in standards for technology that's going to be developed in the next 50 -- you know, 20 to 50 years.

So I just wanted to mention that we're in an age where we have to have embodied AI robots or autonomous systems have to interact with humans. The problem that we've got in standards is we don't recognize life. We don't really have standards for recognizing life in species and -- we have biometrics, but we don't technically really recognize life.

So there's a big gap. At least AI -- some standards of biometrics would argue that. But with AI we don't have protections between robots and humans. And so it's not just robots, but it's also autonomous vehicles. We've got a vehicle, we got a picture down here the bottom left, and it shows actually just shows three people, but there might be a minimum of four people. We're not sure because one of those cars might be, you know, like a Waymo or an autonomous system without a driver or person inside.

The challenge that we're finding is that not only does it need to recognize life to protect it, but it also needs to understand the actions associated with it.

So kind of like a vert. We've got one person walking and another person walking or driving a motorcycle, and we have somebody riding on a bike. If this image was better, we'd show a dog walking in the -- in the pedestrian strip.

And so we need to be able to -- in order to protect life, we have to recognize life. And then we need to have rules for how we interact with it.

Study Group 17 has started working on that, and let me scroll down. We deal with telebiometrics. Actually deal with that, I've been doing this for a while.

If you look at in our different level, we just talked about this biological properties. It has to recognize if something is biological. Then it helps determine the species. Kind of like a car driving. It would be nice to know if it was a pig, a human, or a deer. And it's going to interact differently.

We're having situations where cars are getting -- autonomous vehicles are driving, and they have to choose between hitting a pedestrian or killing the driver. And then once that decision is made, it has to go to court, right.

So there's unique identifying action. And the prioritize secure interaction, how we protect those people in the street.

I'm going to keep going.

This is a consolidated diagram of what we call X.1081. It was developed back in 2007 when the standard first came out, but it's an interaction model where the human is at the center.

It basically gives the body borders and we're able to come up with something called bio signals. So in the physics domain, if you're measuring somebody's temperature, that's a heat signature coming out of the body. If you're getting somebody oxygen, that's a chemical interaction going into the body.

And we also have it mapped into different multimodal -- multimodal like senses. So for taste, touch. And this builds into a protocol called B2M proposal. And B2M protocol is like machine to machine, this is biology to machine because biology are computational actuator systems. And actually computers are modeled after the human brain.

The architecture is modeled after the human brain. Even drones are modeled after things like birds.

Most computational systems just copy biology. They don't do it very well, which is fine. But biology machine protocol was developed in Study Group 17 to tag -- basically there's metadata. It would identify -- let's say a biological entity, it would have certain properties about it and it would say what it was doing.

It would have certain measurements and become as a bio signal.

Now this is the language that you require to talk to and interact with autonomous systems. And that's what we're currently doing at our company.

Go ahead.

>> JULYA REBSTOCK: So if you think about each one of these, as John said, there could actually be multiple AIs embedded inside of these robots or in the embodied AI.

And each one of those can actually communicating between itself. So you have a lot of risks that Study Group 17 we've identified, and things like data leakage. You've got one model is passing raw input or context. It could be personal data, it could be credentials, in the case of the surgery we could have medical records, something like that, between two models. May or may not be filtered.

This of course opens the gates for prompt injection of -- or some kind of malicious input being taken over. So this could manipulate model A. and it has tainted output and cascades in to model B.

This brings in data integrity and authenticity.

If it perceives a fact from model A but doesn't have the Providence and trust score, then the downstream reliability erodes.

And we're talking a lot about transmission here this morning. So if models are exchanging data via APIs, if they have message cues or if communications are in any way intercepted or interrupted without proper encryption authentication, or the, you know, picking up the pauses like previous speaker was talking about latency in the model, what happens?

And of course, intercepted model outputs potentially interrupted or replayed, and that could manipulate downstream actions.

Context contamination is another one. These context, the cascade risk and the amplification of those risks. So a single corrupted message could ripple into multiple output actions. If orchestrators are not validating intermediate results.

So we want to make sure there is some kind of checks and balances in there.

Next, please.

This also brings us to, of course privacy risks. If you in my previous example, if there's three people in a room and the AI must somehow distinguish person A from person B from person C.

I found very interesting the previous example of trying to put privacy on the cup. And have it interactive. You've got one person in the room that is the subject of the interaction, how are we going to maintain the privacy of the people in the background?

So that's always an issue. And of course, biometric signature, the only way for the AI to distinguish is going to be somehow to scan me. It's going to scan my face. It's going to take some kind of a reading on me. That is a biometric signature which, of course, is covered under a lot of privacy regulations right now.

It's considered high-risk data. So there is a big risk of unauthorized collection, overcollection. We, of course, training datasets are always a big issue. What is actually in the training data? Do we have provenance for all of the training data?

If anything was unintentionally ingested into the training data, then it could harm or influence downstream information.

Data leakage is another one. For an AI that is interacting with the human, if you think about a conversational robot and it is talking to someone, it is, of course, recording the conversation. It's recording and trying to read possibly emotion sensors, things like that.

Well, we also have to now consider where is the data going? So is the data being maintained in the body of the robot? Likely not. It's probably interacting with a cloud API of some sort. So that data is moving somewhere, being stored somewhere else and who's going to have access to it, all these things.

We have to make sure that these models and their communication is secured.

And of course all of this brings us to the regulatory misalignment. Because as of today, we know that like the right to be forgotten, for example, if I'm walking down the street and I interact with just a marketing bot that wants to know my opinions about something, do I have the right to give my consent or to remove my consent?

If I say no, I do not consent, I don't want to talk to you and I keep walking, is my interaction and my consent recorded? Have I just -- by even looking at it, it has gathered my bio -- it has gathered my presence, my location, my face. There is no way for us to actually delete the information out of the models today without retraining the model.

So that's something that is in -- currently we do not adhere to our existing privacy regulations when models are put into and used with production data.

Let's go to the next one, please.

So this is just kind of a look at the existing privacy rights. They're all over the place. And the way models are being used now, we have to find a way to fix this.

Go ahead and go to the next one.

Safety and trust is the next one, of course. There is even embodied AI is more exposed because if you think about it, the AI is predominately automatically trusting the sensor output.

So the AI logic must trust the sensors that it is receiving information from. If you think about all of the components, the sensors, actuators, all of this is a risk spot because we have from a supply chain perspective we have a very global environment, of course, and one failure in a sensor from vendor Number 1 put into a robot. If they have to have a firmware update or something like that, what is going to be the compromise chain?

So you know, we've had a lot of press here lately about supplier failures and updates and things like that that have taken over entire systems or brought down entire systems.

So we -- we very much have that cascade risk. And like the is a big one. One malicious or faulty sensor, bad data put into a perception model, embodied AI makes a dangerous decision, an actuator executes a harmful action, anything like that, it definitely amplifies risk. Because of course, bad data means bad physical outcomes.

And then we have a trust escalation as well. Each stage with a GenAI it's just a one-on-one conversation. It's me and whatever I'm saying to the AI. Well agentic AI, I've entrusted it with credentials. Embodied AI has credentials, authority, and it can impact physical safety.

So of course --

>> YUNTAO WANG: Sorry for the interruption, but I'm not sure if there are timer remotely, but there's a timer in this room indicating 15 minutes.

>> JOHN CARAS: We can't see it.

>> JULYA REBSTOCK: We can't see it, I'm sorry.

>> YUNTAO WANG: Okay, thank you.

>> JULYA REBSTOCK: I'll skip through, I just talked about that.

>> JOHN CARAS: Okay.

>> JULYA REBSTOCK: Skip -- we have some mitigation strategies exist with models today, but not -- not as many. I'll leave them up there.

Let's go -- okay.

So as far as where ITU can fill gaps, intermodel communication is definitely a need. Machine unlearning is something that would be great to -- and supply chain and component trust. Go on ahead.

You want to cover biological.

>> JOHN CARAS: I'll cover this real quick. The gaps for biological is missing life recognition protocols. The second one, the absence of biometrics. We developed X.1094, you can find that. That would help. Bio signals are how we communicate between the two entities. The autonomous system and humans.

Security misalignment, data protection versus life protection. Standards are typically data centric. We are protecting data in standards.

We rarely protect life. And if this is supposed to be human centric or life centric, we have to amend or add to the standards to help protect life.

Of course that subject alone don't want to get into it, it's a very touchy subject. But we need to be able to recognize to protect life, we need to be able to recognize it.

Also ignoring interaction complexity. That's what X.1081 and

X.1080.2 deal with is interaction complexity. That's the language that that communicates between both of them.

Then we give here an example very quickly of AI and Google systems. AI is launch -- sorry, Google is launching an AI in their Google home and say did FedEx truck arrive today? So it will tell you which truck arrived. It will say did the dog dig in the backyard? It's associating verbs with nouns. This is the new -- great example of embodied AI.

Let me go down here, telehealth. You guys talked about health a lot. You cannot do telehealth without measuring biology correctly and protecting and transmitting that data over a network. That's all I'm going to say on this slide.

So being on the measure of biology, protect it and transmit it, that's critical. We're working on that in Study Group 17.

Also this is just a small example, heart rate monitor we're seeing used in the field. It gathers certain biosignatures. Same thing here for elderly care. We're able to develop devices like sensors on the wall to detect activity for people living by themselves.

Julya.

>> JULYA REBSTOCK: Just closing off, we had some other samples of medical robots but we've seen all of those. Most of this is we want to make sure that we don't forget the human signature and protecting the human signature with the technology.

So sorry for going over time. Thank you very much. We've got some additional things that will be available in the slide deck.

>> JOHN CARAS: That's it. Thank you, guys. Appreciate it.

>> JULYA REBSTOCK: Thank you.

(Applause)

>> YUNTAO WANG: Okay, thank you, Julya and John for this wonderful conversation. I'm sorry to be the bad guy to cut off this wonderful --

>> JULYA REBSTOCK: That's all right.

>> YUNTAO WANG: -- presentation. I think that security and privacy aspects of embodied AI and robotics is very important. So encourage everyone to take a look at this slide and maybe there will be something very interesting to be studied from there.

So I think the next speaker is from Beijing academy of artificial intelligence, Mr. Zhongxia Zhao, the research fellow.

Also the visiting scholar for Peking University from China. And his presentation will be observations on robotic field and the multimodel sensors. Let's welcome.

>> ZHONGXIA ZHAO: Thanks. Hi, everyone. It's such an honoring to here at this workshop.

I have spent around ten years in AI and the robotics, but now I and my staff observe embodiment AI in the Internet space.

Why?

Because there is something new and exciting popping up. Today I want to share some of my recent [?] and evaluations from China with you.

First off, let's start with some macro-level observations. Here are the question, what -- what's the biggest difference between 2025 and 2050? You might remember ten years back robots were not -- okay, sorry.

You might remember robots were not general purpose. AI could only do things like physical recognition and the world was filled of emerging trends. But now they have merged into one.

It feels like everyone is gearing up for AI and super globals. Once they become a reality, it will like the [?]

The circle where we start.

This is a robot for our store. I believe that my new robot stores will open in China soon. It will bring new devices.

Right now this data allows the shift from horse wagon to cars. And also similar to the Motorola era. But that's what makes it so exciting.

Next up, it's about robot body and supply chain.

There are two questions we often ask. Is design necessary for robots and the Dexterous hands. It's no. How can we upgrade solutions that are starting to emerge in China.

For example, Shanghais from youth design to make their robots work lighter and faster. Phybot has joint [?] there are tons more solutions out there. [?] and so on.

Human and robots are also redesigning new ones. There are robo sense has a [?] solutions and cameras from the nation and local code built embodiment AI robotics innovation center in China.

It's called a [?] model.

And English in accents has a high-resolution tactile sensor. You can see the result.

Okay. About robotic arms. Robotic arms are how you go from instructional to collaborative. And now to the academy data for embodiment hybrids.

We need them to be more Dexterous. And how human interactuate robots better.

It mean it's cheaper and faster to respond. It's more compact. It has greater flexibility. You can say it's four, five, six joint.

The robot from byte dance. It has ball wrist. And some use linkage wrist. It's offer greater flexibility.

Embodiment -- embodiment is also pushing teleoperation forward.

Teleoperation has always been used in professional fields. Like space exploration, the da Vinci robot, but as embodied intelligence need more data from human-controlled row bats. The software and hardware for teleoperation has also evolved.

Teleoperation is to robots what steering wheels and cockpit systems are to cars. Embodied intelligence is like autonomous driving.

Once teleoperation complete a [?] robotics product with [?]. This is some demo. Tactile intelligence from Beijing, our AgiBot from Shanghai.

All these use teleoperation.

I think that will make it much easier for us to set standards, once truly mass market production emerge in the industry. We can develop standards for human robots and embodiment. And just as we do for automobiles and autonomous driving.

Similar to autonomous driving, I think a embodiment intelligence will emerge from a stronger interaction shadow mode.

Okay. Similar to autonomous driving, a better embodied intelligence will emerge from a stronger human-robot interaction shadow mode.

Just like the professor said, human in loop.

But unlike cars, robot steering wheel and its body are separate. Robots get different designs based on their environment, while teleoperation is built for human use.

This is why teleoperation might evolve into traffic gateway for the Internet, letting people and organizations connect to different robots.

The future of teleoperation is a super platform like PC and mobile phones.

Now something about embodied intelligence.

It can be toolkits as backed. Hierarchical resolution and unified integration. Take the academy as an example. We adopt a brain-like structure, Large Language Model as a brain. It's [?] action model acts as a brain. In chart, our action generation we add model similar to the [?] and the refers learning to connect the brain and the [?]

And unifying integration as AI glasses and the robots become more popular, a huge amount of eco data will emerge and the ways of [?] teachers. Our unified model might just appear, maybe a [?] model or [?] model.

Let me give two more examples. Dyna-1 is the first foundation model featuring all-with weather had high-efficiency and Dexterous autonomy. This model has the first time proven that Dexterous operation is commercially viable. So in my wheel, the brain show the hand perception and planning. The brain show the take charge action generation.

Reinforcement to learning should be responsible for imploring the risk and the control to the whole body.

Then about commercialization.

The current robotic body is not sufficient to replace humans in completing assembly-line-level tasks. Most of the time it's a task that adapts to the body rather than the body adapting to the environment.

The body and algorithms that truly belong to the industry are still in the making, and even the industry itself needs to define a universal body design that can solve most of the needs.

It is recommended that the industry side collect more requirements, conduct research on robotic designs that adapt to their environment and build the infrastructure for data collection.

AGI is coming.

Had here are two examples. One is a human robot that makes full use of its front and back sides to assist the elderly. Its backside work as a wheelchair. Wheel the front side have to stand up. So the other is a 2.2 meet-tall robot designed especially for operation in the power sector.

And it truly needs a lot of stress to complete power-related tasks.

I think the opportunities in the wave of robotics innovation should first lie in safety applications in extreme or enclosed environments, as well as the use, management, operation of human-used equipment. Only after that will they address the dirtiest, most difficult, and most delicate tasks that automated equipment leave for humans.

Both robotics and [?] Ferrari of the future in the industry. The emerge of embodiment intelligence is more like [?] driving by multiple agents including the robots supply chain. Robot bodies model [?] and communications.

Super intelligence is bound to emerge -- super intelligence is bound to emerge, we just don't know when or where it will happen. Sorry for my poor English.

Okay, that's all.

(Applause)

>> YUNTAO WANG: Thank you. Thank you, Mr. Zhao, for the excellent presentation. I think some of the observations are very interesting and the use cases are also very interesting as well.

So the next presentation is from the director of HRI, Boston Dynamics, and he will be joining us remotely.

And the title is designing industry or applications of embodied AI.

Let's welcome Mr. David Robert. .

>> JOHN CARAS: David, you got take off mute.

>> YUNTAO WANG: Yes, we can hear you clearly.

>> DAVID ROBERT: Great, good afternoon, everyone. Can you hear me clearly?

>> YUNTAO WANG: Yes, very good.

>> DAVID ROBERT: Excellent. So first, I just want to thank you for including me in this important workshop. And I also wanted to send my greetings and respect to all the Distinguished Colleagues assembled here today. My name is David Robert, I'm the director of human robot interaction of Boston Dynamics United States-based robotics company.

And what I want to do today is just to provide, in addition to everything else we've heard today, just to provide a very direct example of how we use embodied AI today in our various products.

But I will be focusing on our main product that's been out there the longest so that everyone can get some context of how this gets applied on the commercial side.

So many of you may be familiar with some of these robots, but I just want to describe them so that everyone has a reference point of what I'm about to discuss.

Spot is the yellow quadripad on the left side. This is the basic model of Spot that does not have anything on its back, no payloads.

But normally our customers will purchase this robot and configure it for their particular application by adding different types of sensors on to its back.

That's important to talk about later on because of course, each sensor generates different types of data and these are things that we might want to consider when thinking about standards and moving forward.

In the middle here you see our warehouse robot called Stretch. And currently Stretch is used in logistic operations to assist in unloading large containers and trucks.

So you drive Stretch, it's on a mobile base, up to a warehouse door and then you back up a truck full of boxes. And when that door opens, the robot is able to detect all the boxes and move them.

And this is a great -- actually a great robot in terms of helping folks not have to do this task themselves. And as other people mentioned, we all believe that robots initially will be taking on dull, dirty, and dangerous jobs. This job is not only dull, it's dangerous.

And most human beings that do this job currently do not last that long in doing so because of the conditions. The heat inside of the container, as well as how much it hurts them physically to do this repetitive action.

So Stretch is a great warehouse robot that's been deployed all over the world of as well.

And on the right I think lately based on a lot of the current interest in humanoids, this is Atlas. Atlas used to be a hydraulic robot, and it's really been around for a long time. But recently it's become -- it's on the path to being productized and it's turned into an electric Atlas robot that's a humanoid that will likely be working in many industrial environments probably in a car assembly facility. Since Boston Dynamics, the U.S. company, is mostly owned by Hyundai, the Korean car manufacturer.

So just to talk a little bit more about these sensors and the things that go on top of Spot, if you look any top left, you'll see that there's a large black bulbous shaped object. Actually that one's pretty easy to explain. It's a camera that's more high-resolution than the cameras that you already get with a robot.

And it's able to turn and be able to zoom in to different objects. Because we realize that one of the really core competencies of this robot is that it can do things in a very repeated sort of dependable way.

And that's very useful for industry in order for the robot to be able to inspect warehouses, different production lines, and just make sure that things are running as they should.

But we also realize that even though Boston Dynamics have spent a lot of time making sure the robots can go anywhere, which is why we have legs so that we can walk over obstacles and uneven terrain, but we also realize that using a camera with higher zoom abilities would actually have to sometimes go all the way to the place, we can just position the robot and then look through the camera's eye.

So that's a very helpful camera on the robot that's an add-on. And that generates a ton of data for your consideration.

The middle sensor is actually a really -- a really neat object that is able to use ultrasound. So able to detect sounds above the human hearing range that through a variety of techniques we can analyze that data and pick out failures in things like ball bearings and also in motors.

So one of the things that's really core to our business is to do preventative maintenance. What that means is that you can imagine in all different types of facilities there's a lot of machines and gauges and pumps and ball bearings and motors. And you know, the way the world works today, these things mostly are on 24/7. And the robot, what it does by being able to detect a potential failure or anomaly from how things should be operating, is it enables the customers to find out, hey, there's an issue here that you might want to have checked out. And do that maintenance before you have to do the repair.

And in case it's not obvious, doing maintenance might be under $50, but doing repair might be in the tens of thousands of dollars. Providing that kind of service gives people peace and it's something that we've been pretty successful in.

In the back there's a classic Lidar which I think we've heard described a couple times too, and that allows the robot to have more awareness in 3D of what's going on and it creates a large point cloud. And that point cloud is actually pretty substantial too.

So I also wanted to point out that in terms of applications, yes, the main application is preventative maintenance and industrial inspections. But there's a variety of Spot robots configured for public safety.

The New York fire department has a Dalmatian-themed Spot robot that goes into buildings and does search and rescue. You can see here on the bottom left there's a suspicious package that, you know, may or may not contain a bomb. So we have this robot configured to do bomb squad work.

And overall, what I want to talk about a little bit is that, you know, AI is not new for Boston Dynamics in the sense that actually the productization of our efforts is only a recent development in the past five or six years. But that the Boston Dynamics effort has also been focused on advancing AI and embodying it.

And actually what you're seeing here on the top left is the difference between what happens when you try to walk this robot quadripad over a slippery surface without AI and then on the right you see the improvement of having the robot trained to walk on slippery surfaces like ice and also dirty warehouse environments sometimes have oil and things on the ground that the robot can slip on.

So this is something that we keep on doing. But on the bottom left is something fun. We are really proud of what we call athletic intelligence. And just to zoom out for one second and talk to some of the things I heard today that I thought were really interesting, athletic intelligence and being able to have a body and giving an AI a body, I do believe, is a pathway towards artificial general intelligence. And the athletic intelligence in our company has I think personally a beautiful origin in that the current CEO Robert Plater, while he was studying MIT, he was also a very competitive gymnastics expert who participated in gymnastic meets all over the world. When it came time for him to declare what he was going to do for his PhD, he decided to teach the robot how do backflips and front flips.

And I believe that's really beautiful in the sense that it's always nice to see a creator putting themselves into their work. And the benefit of that is that we built robots in a way that creates a lot of real value for our customers and hopefully for society.

But at the same time, we have fun. And that's a really important thing that we are enjoying designing these robots. And sometimes overengineering them, which gives them a competitive advantage so that they're not just done -- they're not just for one task, but rather can be applied to different tasks, even tasks that we can't predict.

On the right here you're seeing some images coming in of the old Atlas on the top. And on the bottom right here is a newer Atlas, we call it eAtlas with the head and this is the one that's currently under development as a product.

So I just want to focus a little bit on Spot, because it is the most mature product. It's been out there for a while. It's been deployed in several countries. And there's thousands of them out there. But like I said before, think about spot as sort of the horse for sensors, but it's moving so the difference is now instead of, perhaps, setting up your warehouse facility with hundreds or thousands of sensors, you can get a fleet of mobile sensors that are actually just Spot robots. And the way we're talking about it today is how much data is being generate and recorded and transmit and all that that you might want to consider when you're creating standards.

So what happens when you have these different sensors on the robot?

Well, for example, with the camera, the robot can walk up to a gauge to check the pressure. And with computer vision send an alert if something is not looking correct.

Also, of course, it can look beyond what a human can see with the human eyes. It looks through infrared and detects different kind of thermal anomalies. So if things are running hot.

And like I mentioned before, acoustic inspection is really interesting. You know, because these are things that are invisible to humans. But the robot is able to -- when you program it, to move around autonomously. It's able to do this very, very reliably, day after day as much as you want.

And I think the last point here is the data analysis. And that's something I'm going to get back to in a moment. But just to say that, you know, a robot with sensors generating all this data, you know, then where does it go? What do you do with it? How do you analyze it and generate insight is another question.

And again, to that point of, you know, repeatable and standardized collection, what I wanted to show you is just sort of a mosaic of how, you know, you put -- you put the robot to work and every day it wakes up and it goes and it takes the exact same photograph or sensor reading from the exact same location which is highly reliable and leads to really much easier analysis of extremely large datasets to create insight.

And this is happening across all industries, oil and gas, power, utility, food and beverage, pharmaceuticals, aerospace, automotive, of course. So just to think about the scale of this and how it keeps on growing I think will help give you food for thought in terms of what to think about when you're starting to create standards.

And you know, I would be remiss if I didn't mention that, A, as the human robot interaction director I do believe in the usefulness of the ITU and standards and I understand that international standards, global standards interact with local standards. I'm involved with setting standards at the national and local level here in the United States.

But I also feel like it's important that we really, as a species, as humanity, we really allow ourselves to take the benefits of all this new technology. Given all the potential risks of it, which we need to think about really and consider very deeply, I think it's very important for us to not stifle industry and not stifle innovation so that we can continue to provide value and build robots that help humans, not harm them.

So about that data, where does it all go? So, you know, in the past couple years we have launched a platform called orbit. And what it does is a lot of different things, but it's mostly a robot fleet manage we are a lot of extra features. Because the people that operate Orbit are able to see different alerts as the robots are moving around these different facilities. And if they see something that's wrong, the Orbit is going to alert a human user or send a message to another system.

But while it's monitoring these site conditions, and even connecting to different other systems that exist, it's also storing all this information in the cloud and it's also looking at time series data.

So we're not talking about just looking at one picture, we're talking about looking at a lot of pictures over a really long period of time in order to be able to extract and mine the information that we need to take an action.

So here's an example on the right. You might have the robot at night walking around your facility and you may ask it to just double-check that, you know, nobody accidentally left a door open, right.

Or you might -- it might also realize, hey, there's water on the floor, it could be slippery and dangerous and this is something I need to do something about.

So that kind of smart rounds is what we call it, enables the robot through AI, embodied AI, computer vision and different techniques we have to basically see entities and produce warnings for people to act on them.

So an example would be like I just mentioned in the middle here, the door was opened by accident on the left. We have these lockout stations and so this really provides a lot of safety.

There's also the idea that, you know, if, for example, you don't have a fire extinguisher where you really need to have one, the robot can help with compliance to make sure that you're complying with all the safety standards.

So this is a very useful thing. But with regards to what we're talking about today, this is generating a ton of data. You know, it's protected in the cloud, but at the same time just the transmission of it in order to make this a useful system, you know, that's -- a lot of -- a lot of information is passing around and I think that that's something for you to consider in terms of this sort of industrial application of it.

So I'm with the HRI group, like I mentioned. Human robot interaction is about user experience, about the design, and considering and making sure the human is in the middle of the loop. And so what -- and so I agree completely that the human needs to stay in the loop because I'm looking at a future where I'm trying to create a partnership that benefits both humans and hopefully larger groups of humans and society as a whole.

But just to show you behind the scenes a little bit, within the corporate structure like Boston Dynamics company, the HRI team sits in the middle and is constantly influencing both engineering, the product design, the strategy, and then safety and ethics.

But I wanted to discuss in case it's of interest, the fact that we have a design philosophy that we want the robots to be as transparent as possible. Of course, it's beneficial for the robots to on easy to understand and intuitive to use, and then totally integrated to different environments.

So when we think about that intuitive aspect is because we want robots to benefit the most number of people. I think gone are the days where you had to have a PhD from Stanford MIT to run a robot. Luckily these things are getting much easier to use and the more people that can use them, we hope that the more people can benefit from them.

So when we -- we carefully -- when we go to design new robots, we carefully study every single touch point between the human operator or the user, and the robot. And we build these things called user journeys which you may have seen before.

And they help give us an idea of what are the different opportunities for creating various types of processes or devices even to interact with a robot. Will I be controlling it with a joystick? Will I just let the robot go and it controls itself autonomously? Will there be a console?

And also what happens when something goes wrong? How do we fix that?

So this is a user centric design process in contrast to a more engineering approach which would be the traditional concept of operations. So we try to meet in the middle and collaborate closely with the engineers.

Regardless of what that process, what I think is important in terms of understanding the user is not just recognizing that they're there, but also starting to eventually predict their intentions and their intent.

And that's something that I'm hoping will lead to a better partnership. And in order to do that, there are a variety of methods from using a camera to detect a person, but then you can also look at nonverbal cues. So people's facial expressions, their gaze behaviors, their posture, their movement. How they touch.

And so these things are, you know, well researched in the field. And these are also generating a ton of data. So you have -- and of a highly personal nature. So how do we protect people's privacy?

You know, at the same time create the functionality we need. I'll give you an example. In a prior job I created an embodied AI for learning. And it was deployed in the Chinese market. And we basically had some emotion recognition on the robot, and we first started academically thinking, oh, this robot needs to recognize the whole range of human emotions, but then through testing and just, you know, the inevitable solution ended up being that we realized that all we -- the most -- one of the most important emotions to detect was frustration. Because if the person using the robot to learn, in this case young children in China, if they get frustrated, then they're not learning anymore and everything breaks down. And the whole purpose of the robot becomes null and void.

So just to keep in mind that, you know, there are reasons to protect, and it makes sense to me of course, especially with vulnerable populations, to really insist on privacy. But there are also really good things about understanding certain more personal and deeper aspects of the human experience.

And what I mean by that is that when we realized, for example, that the child was being -- was getting frustrated, which didn't happen often, but when it did, then the robot would adapt and change its pedagogical model in order to present the material to the child in a different way. Right.

So moving on, I also want to talk about something that I'm very passionate about. And that is that as we design these systems, we -- we should really make sure to uphold human agency rather than impinge on it.

And I'm really seeing a world where very soon, you know, right now we're in industry and it's a little bit in service areas. But I imagine robots really sharing social spaces with us pretty soon.

And I want to make sure that everyone designing these robots and folks that are creating standards really put the human in the center, up front, and that human agency is respected. Even something as simple as sharing a sidewalk with a delivery robot. You know, in my opinion, you know, robots should never impede a human's ability to move around freely. And that's just a starting point.

But there's a lot of other aspects to human agency that I think we need to put, you know, put forward.

So --

>> YUNTAO WANG: Sorry, David, for the interruption. But I think we've reached the maximum of 15 minutes already. So could you wrap up so there will be time for Q&A.

>> DAVID ROBERT: Okay. Everyone wants AI to be transparent. You can see that it's telling you it has a low battery projecting on to the floor. We believe that giving the human the ability, even innocent bystander to be able to see into the mind of the AI in an iconic way or sorry you, using icons what the robot is intending to do next is really good for safety.

So if anyone's interested, take a picture of this QR Code and you can learn a little bit more about how we do that.

But I just want to end on this thought, which is that we really have to act now because think about the scale of this. Right now I'm HRI practitioner and professional, but very soon it's not going to be human robot interaction, it's going to be community robot interaction.

Thank you very much.

(Applause)

>> YUNTAO WANG: Thank you, okay. Thank you so much, David, for bringing this wonderful presentation for expressing the

designing philosophies and considerations for industry or applications of embodied AI.

So I just want to thank all the insightful contributions and active participations for the speakers. And I think according to the agenda, we still have some very little time for Q&As.

So I think the areas already requested from the floor, but unfortunately we only have one question opportunity, I'm afraid.

So I open the floor for one question and I see this lady over there. Yeah.

>> And in mind of the one goal initiative for governance. Thank you.

So about cognition plus learning plus collaborative as being functions, I would say collaborative is another category of things than the first two. In fact, collaborative is a goal, whereas, the other ones are features that enable any other goal also.

So maybe some -- and about Asimov, please, everybody, I used to be a science fiction writer at some point in my life. And fiction writers need to create drama. And when we -- when we design our -- you know, the structure within our story, including rules like that, they're designed to create drama.

So let us not follow anything in the fiction area. Thank you.

So consideration of conflicts and other damages created by the resources that need to be extracted to build those things, et cetera, would need to be taken into consideration when we want to assess their risks.

And when I say risks, as I say, I put the opportunities in the risks. We take both things together and there's sort of a compound risk that comes out.

And then about surgeons, that was interesting. The 150 milliseconds versus the 72 milliseconds. I say thinking a surgery when they see a latency they can do something but there's a latency that they don't see. So I'm wondering what the impact of that is because they cannot do anything about that.

And that is all. Thank you very much for giving me the time.

>> YUNTAO WANG: Thank you. Just so one participation with so many questions, yeah. So experts, yeah.

>> KASHIF IKRAM: Can I address the surgeon issue?

>> YUNTAO WANG: Yeah.

>> KASHIF IKRAM: I think you bring up a good point in that something I talked about yesterday at the academy for surgical

training, in that particular case, the surgeon was able to slow down his own movements. He said it was like swimming through tree core, basically working -- or walking through honey. But by working slowly he was able to overcome that 172 milliseconds obstruction that he had.

But what we need lay down is, at what point do we say to the surgeon, stop. Right? Surgeons sometimes are not good at that and they will continue. With all the obstacles and try and push through.

So we need something like a patient advocate or some rules and regulations that say when you get beyond this point, then you automatically stop and you have to hand over to the surgeon that's in the room where the patient is.

I think you raise a very good point that needs clear rules and guidance around.

>> YUNTAO WANG: Any other comments?

>> I think you mentioned collaborative is the nature -- one of the nature of the EAI we're talking about in this context. I think for the previous robotics of the virtual agents, they are siloed. They're not collectively interacted autonomously, so in the context of when we're talking about the EAI for the future, it is -- it will -- collaborative will autonomously between agents.

And also between human and agents. So that's -- that's meaning of this feature, yeah.

Thank you.

>> YUNTAO WANG: Thank you. Are there any other comments? Okay.

>> On the issue of delay and intelligence itself, I think it's also an interesting approach to have the AI know that there is delay and act accordingly on-site. This is fully operated telesystems I'm assuming in this case, there's no local intelligence at all.

But I think the transition between this and autonomous surgery, you know, decades down the line, is something that is collaborative surgery where the robot is also taking local decisions to mitigate this delay so that the robot is doing a certain command and then the robot is completing it and then the delay's out of the equation now and then they just have to sync up at certain points.

I think that could be interesting where the AI is also mitigating the delay.

>> YUNTAO WANG: Thank you. Thank you for all the comments and thank you for for these wonderful questions. Very unfortunately, we have more questions than answers, that's why we're carrying out this workshop.

I think actually before the closing of session 1, we have a very special guest who will be in here. And actually that is a [?] for session 1. I think let's welcome that special guest and let's anticipate his performance. Okay.

>> Thank you, please everyone stay seated and there will be a demonstration now. And I'm inviting Yuki to take the mic and talk to us.

>> YUKI: Hello.

>> Yes, we can hear you. Just talk.

>> YUKI: We can work the parameters, our humanoid robots up to center stage, and in the meanwhile I can explain a little bit about our robot boy.

So while the robot is slowly walking to us, I can start the explanation into that a little bit hello, everyone, my name is Yuki and integration partner for Switzerland and for the European market. And here is the robot baby that's six months out. But still is like a 5 to 6-year-old size. Is like 1.3 meters in head and the 35 KG in the [?]

This is a humanoid robot built for general purpose. And today I'm traveling from Zurich to Geneva -- today I'm traveling from Zurich to Geneva and my two friends, they are actually first time new to operate the robot. Maybe I need to support them a little bit.

And I want to introduce a little bit more about Hadawa and also the system on our robot.

We can still walk it to the center a little bit, it's getting dangerous. Kristin is first time to control the robot boy. I think she's still doing a good job. All right. To the very center. Here we can stop.

So this is Prometheus from Union Tree. And actually this is the first CE certified humanoid in the European market. And some quick facts. Here is the depth camera for perception system. And leader as well. And we have the model 23 degrees of freedom make it move just natural like human things.

And we also have like three system to power up the human noise. The first one is perception system, like I mentioned about the cameras. And then is the motion control system. And the last one is the AI or they say the brain function.

And the Unitree of two versions of humanoid. One is default while we control it with a controller. And we can also do the second developing like autonomous walking, interactions, and this one is actually integrated with chat GPT so basically you can talk to 40 languages with him, but we are not using this function right now in the event room because too many voices goes on.

Because later you can interact with him during the coffee break outside. We have three robots today, just to have fun with you.

One is the small size humanoid and the other is 1.65 meter robots upstairs, big guys. And robot dog together as our robot family.

And as integrator of Unitree, we are more focused on several scenarios. First of all is the educational scenario, which robot goes to the laboratories, to the research centers. He hears me.

And the second scenario is to industry, for example, the largest warehouse support, sorting, loading, uploading. So basically replace some human being work to combine -- it's usually for example when beings move into between the [?] industry robots, work can be replaced by human beings as well. By humanoid robots as well.

And then the third scenario is more to the public sector, like inspections, security. Health care we're slowly working on it, but it's really difficult at the moment.

And I would say we are facing quite some challenges, but also opportunities in the European market.

And when the humanoid robots slowly walks out from a lab to factories, [?] are still missing but evolving. And we are very proud upon part of process. We have seven workshops across Europe. This one comes in Geneva with my partners. And we also have workshops in Zurich and together with ACP we have other workshops in Berlin and other places.

I think for us the humanoid robot market is really emerging fast and crazy in a way, because back to last year, 2024, when I was in China, when I first met Unitree robots, but at that time the robot can stand without falling was a big win. And there is only like three players in the market.

And this year September 1 I go back to China there were around 200 or even 300 manufacturers. They produce all size of humanoid robots. And all legs, two legs, four legs, six legs, you can imagine.

And also for intelligence, and we are I think the model we trained the humanoids is like 80% done by the AI training, maybe the experts on stage can explain better.

And still 20% by coding. But in the future, I would assume we can offer no coding solution on the humanoids to European client, maybe probably within half a year. So we're quite promising on that.

So that's all. Basically some flavor to the event, and later you can during the coffee break you can chill, stay around, or just have fun with our robots upstairs.

Thank you.

(Applause)

>> YUNTAO WANG: Thank you very much Unitree to be with us and showing us a bit of those robots. So you will have two, three robots around. We are incredibly exactly on time, 16:15. We have full half an hour for coffee break waiting for you outside. Kindly offered by AIIA, artificial intelligence industry alliances. And actually Unitree is one of the members of AIIA. So AIIA is open for everybody and so sponsor this coffee break.

So thank you very much to all the speakers and see you again at 16:45. Quarter to 5:00. Thank you.

(Break)

>> STEFANO POLIDORI: Okay, we will start soon. So if we will have the speakers sit, moderator. One speaker is online, right? Or not. All speaker are here? Okay.

Okay. So we'll start now session 1 -- sorry, session 2. And I'm very pleased to introduce you to the moderator of this session, that is Mr. Justin Ridge from Nokia Corporation in had USA. Justin is also vice-chair of the Study Group 21 and co-chair of the Working Party 3 which is the Working Party responsible for Question 5 that we spoke about earlier about artificial intelligence.

I'm very pleased to give the floor to Justin to lead us into this session 2.

Thank you.

>> JUSTIN RIDGE: Thank you, Stefano. So session 2 is a little shorter session, and it's more focused on multimedia standardization.

So you heard this morning very broad topics, but one of the things that came up was that there's lots of data.

So this raises questions about how we transmit that data, how we interface between systems, how things are synchronized, and that's sort of the topic for this next session.

Now, you don't need to hear from me, so the way we're going to do this is I will introduce each of the speakers in turn. I won't go through their bios in great detail because they are on the events page. And you can look up and read their bios there.

But I will introduce them briefly. And we will hold the Q&A until the end of the session. So with that said, let's make a start.

Our first speaker for this session is Touradj Ebrahimi. He is someone I've known for a very long time and I must say it's a great pleasure to have him here presenting. Touradj is from EPFL. And he also holds a number of other roles, including the Chair of JPEG. And so his topic is towards efficient vision representation and coding standards for superior embodied intelligence.

Touradj.

>> TOURADJ EBRAHIMI: Thank you, Justin. It's a pleasure to be here. I was given the difficult task of actually going faster than the allocated time, so I do my best.

Let me remind ourselves that in session 1, this slide having my definition of what is the meaning of embodied AI, we also saw a number of challenges. I'm not probably using exactly the same words, but here a list of some of the challenges that were mentioned, all right.

So Justin just said it. We need not only efficient ways to code, but also represent the data. We heard about images and video and even Lidar, point cloud, et cetera being used.

We need to manage, basically, a very difficult thing that is called latency. Our latency are increasing in the case of embodied AI, and there are a few other things that we mentioned, right?

Privacy, security, and of course because we have interoperability formats, we need performance metric. Evaluation was also mentioned as one of the challenges.

Now, the most obvious, of course, challenge that we are facing is making systems that become interoperable, because otherwise we are going to be prisoner of a modelistic business models. And so I guess everybody agrees that standardization, and that's why this workshop I think exists, is very, very important.

But also it might be interesting to remember and remind ourselves that sometimes some of the things that are desired in some applications and some technologies already have been standardized.

So it would be sometime good to look and see if there are existing international standards we can leverage on, rather than reinventing the wheel.

And if they do exist, it makes a lot of sense do it jointly. If there are more than one standard for the purpose of whatever may exist, they weaken themselves and they weaken basically the ecosystem.

Now, it was mentioned I have been involved in one of the many standardization bodies that have been taking these objectives as a guideline to develop standards. And that is JPEG. Just want to remind that JPEG actually is a group, a committee, standardization committee under auspices of ITU-T. And in fact, even SG 21.

And it has maybe a kind of nice number of successful standards to its 30-plus years of existence. And I would like to focus on three of them that I think that are particularly interesting to consider. One talks about -- when one talks about the embedded or embodied, rather, intelligence.

And you know, the first one is a standard that was developed not because of embodied intelligence, it was developed because the number of applications that wanted to have low latency but still low complexity. And they were basically relying mainly on compressed data were increasing.

And I think that one of the such applications is also what -- what we are witnessing in the embodied intelligence. Artificial intelligence.

But there are, of course, many other applications. So this is one I'm just putting here because there are many, many applications. And we should not try to always look at applications and try to come up with standards for specific applications.

It's a better way to look at what are the features or benefits that we are looking for. And this is usually called requirements. And that's exactly what JPEG is and many, many standards and factors are doing.

They look at specific applications and extract requirements from them and from those requirements they come one systems that have good features that are interesting for a larger set of applications.

And I think JPEG XS is one of those. And in fact, it's quite successful. It's not only has been used in directly in a number of applications nowadays, like some that I just mentioned, like cloud computing and automotive, et cetera, but also it is used in other standards.

And this number is growing. And I think it's -- it's something to keep in mind. So if people are looking for ultra low latency and low complexity, which are two of the challenges that we mentioned in session 1, JPEG XS maybe could be a candidate to look at.

I want to switch to -- because embody the artificial intelligence, of course, has artificial intelligence in it. And the second standard that JPEG has been working on for a few years is called JPEG AI simply. And this one is an interesting standardization effort, because it is, first of all, a joint standard between ISOIC and ITU.

And its goal is to leverage artificial intelligence to not only do better coding and compression, but also a number of other things.

This is done, of course, as I mentioned, through a number of requirements that were identified. One, of course, is the high coding efficiency. So you want to compress because there's a lot of data, but also many applications and particularly embodied artificial intelligence, content understanding, robots that have to see what is going on and understand what is going on around them and objects, et cetera, and when they do actions see what is the result of actions are becoming very, very important requirements.

And of course also the -- the content enhancements. So the world is not perfect and sometimes we have to process and enhance the content that we are -- we are working on. AI is being increasingly used for this kind of thing.

And last but not least, JPEG AI wants to come up with a standard that does not really develop a standard necessarily only for machines versus humans because that might duplicate.

We heard many times people talked about user centric artificial intelligence or embodied artificial intelligence. It's in fact the same way.

And this is done through a standard that now is published since January this year that is following something called the triple purpose framework. Meaning that not only you can compress efficiently using artificial intelligence technologies content, but you could also in a very efficient way use artificial intelligence to process them in the compressed domain. And also run all sorts of content understanding again in the compressed domain. That will make things lower latency because you don't have to wait to decode where doing.

And at the same time, also more efficiently.

So as I said, the Version 1 of the standard is out since January this year, and there is a second version we are working on that is focusing more on this kind of how you could leverage computer vision and processing and enhancement.

And I think the timing is good. So it's not necessarily a standard that is ready you have to -- those of you who have an interest in the embodied intelligence, this could be done jointly.

I want to finish by telling you about something that is evolving and to do that I want to [?] artificial intelligence at least the current paradigm is very inspired by the biological computation as people put it. Deep neural networks is a good example.

But actually the capture is not really following the same approach, right. Most capture, most cameras are actually following something that takes advantage of human visual system, but it is not actually following the same principles as human visual system.

So the question is, can you come up with a human visual system or biological sensing inspired approach? And the answer is yes.

So there is -- in fact, there are a number of sensors called event sensing that are really different from frame-based video, for example. Cameras for the lack of time, I'm not going to be able to spend too much time on the advantages of doing it through events versus frames as it is today.

But what I want to tell you is that the standard is now being developed and hopefully if all goes well, it will be released, published next year in 2026. And in fact, this has been done in junction with Study Group 21. And it's going to be -- if all goes well -- a joint standard.

I know, but I don't know the result of the discussion, I know that, in fact, Working Party 3 is looking into that and I think this week or next week ITU will decide whether this is the path to follow or not.

Again, there are lots of requirements regarding how to code, how to store, how to compress and how to transmit interoperability with existing standards and interfaces for camera systems and capture devices is very, very important.

So this is one of those things that -- that this standard has been trying to do. And it's actually quite -- it's quite a good stage of development. There is already a CD and soon next week a draft international standard that will be available for that.

And it is also a little bit like the agentic AI that I mentioned, it's done at stages. The most urgent one is to come up with a lossy coding that will be embedded in devices. So really in the sensors in the is Phase 1. We're already done with the Phase 1, so we're just waiting for the standard to go through the process that is required to become a standard.

Our expectation is that by sometime next year, maybe about a year from now, it will be published. There is a second and a third phase that are in tool making.

Just want to say a few last words. I still have two minutes. So about -- I gave you three examples. But by listening to what happened in session 1, but also looking around, I think that there are a number of other things that JPEG has -- has created that standards that could be also relevant.

We heard about privacy and security many times in session 1. Actually JPEG has standards to protect privacy and security of imaging systems.

I think if you want to look at next generation pictures, there was also a lot of mention of photos and pictures, et cetera. I really recommend that people look at JPEG XL, because this is something that's hopefully not fulfilling all the features of the good ole JPEG that is now 30 plus years old. But it does it better, including in terms of complexity.

So it has complexity that is comparable to JPEG 1, and it's really worth looking.

I also heard a lot mentions of Lidar and 3D. JPEG has actually already standardized, so it's published, a standard on a learning base point cloud coding. The story's exactly the same as JPEG AI. It not only does good compression, but also latent space compress domain, processing, and analytics.

And of course performance also, there is at least as far as images are concerned, a number of standards, especially JPEG AIC that offers ways to do that.

I think I'm done and I think I was three seconds late, so thank you very much for your attention.

>> JUSTIN RIDGE: Thank you, Touradj:

(Applause)

>> JUSTIN RIDGE: Thank you, Touradj. As I said, we'll hold questions until we finish this session.

But I'd like to, in the interest of time, move straight on to our next presenter. Guoping Pan is from Zerith Robotics and his topic is titled how to transform data into knowledge.

Sounds very -- very interesting. So please go ahead.

>> GUOPING PAN: Thank you, moderator Justin. Maybe I need the controller.

Okay. Okay, I am Guoping Pan from Zerith Robotics. It's a pleasure to be here to share our insights.

I'd like to express my sign veer gratitude to the ITU organizers for hosting this wonderful event.

And the topic to focus today is how to transform data into knowledge. How what I share can be helpful to all of you here today.

And first, let's take a look at some of the popular open source datasets in the fields of embodied AI.

Among them the Open-X Embodiment were not the most [?] datasets, containing over 1 million trajectories and distribution is uneven, with a majority of simple skills.

Since it aggregate data from multiple robotic systems, they offer great diversity, but also suffer from new data collective and quietly unbalanced skill distribution.

Researchers of [?] such as [?] from the datasets for training purpose.

The has the most systematic clarifications of multiple occasion tasks. [?] object into action and shows understanding and so on.

The dataset coverage on for robotics environment, the size remain limited to 55,000 in total. The AGI datasets have over 1 million trajectory. Similar to OEX dataset.

However, they have better consistency in embodiment and data quality. Despite this, most of the existing datasets only contain to raw [?] data such as [?] discretion and robo actions.

They lack knowledge level annotation which are crucial for truly awakened intelligence.

And the importance of extracting knowledge from data can be clearly observed in the evolution of reality model. A typical example is the transitions from pi zero to pi zero .5. And pi relies purely on the robotics dataset and pi .5 is data commonly used for the foundational model.

This comes with data with understanding, grounding, and reasoning which helps mitigate the loss of general culpability and the [?] backbone.

In addition, number of models such as instructionary has also helped in this. Since it's inherently tied to specific tasks, enriching it with the most comprehensive knowledge level annotation can greatly enhance the model understanding of action, somatic and text context.

During the task definition phases, we clearly [?] the task description and defined subtask label. At the beginning of each task, experts have a description and assign corresponding annotation to enhance the association between tasks and object.

All objects immersed in each subtask are annotated with grounding label. The annotation aims to strengthening binding between language and model.

However, more compressed [?] to first high-level cognitive ability such as temporal [?]. Each aspect based on the same discretion, encouraging the models to [?] contact.

When it comes to task level planning or what we may call the model's imagination ability, it can guide the current behaviors to some extent. However, the assist [?] actions using frame by frame imagination is often redundant. Since one [?] has little inference on behavior.

We adopt the [?] to change the model predictionability. We labeled [?] from the [?] accordingly.

The design choice is closely related to our model architecture which I will talk about later.

To enable efficient data processings and iterative optimization, we have developed a comprehensive framework called the Zerith chain.

This system supports task and data collections involving hundreds of participation. And able to [?] human in the room and automation for generating high-quality data.

The platform is also int -- integrated for benchmarks [?]

Based on where we start, the checkpoint which can be deployed to robust for further revocation.

The end to end greatly accelerate the model development and iterations enabling further progress from data collections to development.

Based on our knowledge enriched datasets, we changed the -- which a system architecture combines a cognitive system and behavior system. The design enable real time as 30 Hertz and control of 600 Hertz with an action size of 20.

Within this framework, the latent, planner has Horizon manipulation tasks through [?] and prediction.

Next I'd like to show you the performance of [?] long horizon cleaning task.

Maybe I can play the video. Okay. That is the workflow of our robotics. The first one is pick up the dirty laundry basket and move to this destination location. And place the dirty towels into the basket.

Okay. Considering the time constraint maybe I will speed up the video. So that is cleaning of the bathroom.

And the last one is cleaning the bathroom wash bin. Cleaning the floor.

Okay. To surely validate the models [?] we focus on examining the four or five dimensions of performance. The first one is instruction following. The agent accurately executes tasks based on the natural language instructions. The capability thoroughly evaluates semantic disambiguation, interactive communication and reasoning skills.

The second is spatial understanding. The agent perceives and comprehends spatial relationships between objects, such as relative position and orientation.

And third long-horizon planning. The agent plans and makes decisions over extended sequences of action in complex environments.

This is essential for tasks like cleaning and cooking, where long-term coordination is needed.

And four is the force control. The agent precisely controls the forces applied during object manipulation to prevent damage to itself and the object or the environment. This is vital for tasks involving dense contact, such as wiping or door opening.

Five is generalization. The agent transfers knowledge to new, unseen environments or tasks, demonstrating adaptability and robustness.

To efficiently stimulate this, we have new dataset which is planned to be released in Q4 this year. The dataset contained a total duration around 2,000 hours and carry 100 tasks and 50 scene variations.

Total duration is more than those of some existing datasets. Each trajectory in Zerith home is annotated with label and datas maintain high degree of diversity.

The scene diversity focus on various household environments with cleaning and organizing tasks as the core categories.

We aim to further enhance the capability of embodied intelligence in home environment promoting its practical application in daily life.

Finally, I'd like to briefly introduce our robot, the real deal robot H 1 and humanoid row bat. Both are equipped with the suite and high degrees of freedom making them [?] for embody AI research and development.

However, embody intelligence it has seen some early development, but there is still a need for innovation and data utilization and model architecture and training.

This concludes my presentation. Thank you all for your attention.

Thank you.

(Applause)

>> JUSTIN RIDGE: Excellent. Thank you very much for your presentation.

It's -- we've gone from -- we're covering quite a broad range in this session as well with multimedia standards and robotics in action.

But again, we'll hold the questions till the end of the session.

So let's keep -- let's keep moving along and I would like to introduce to you our third speaker for this session. Andrea Cavallaro, who is also from EPFL, and his topic is multimedia content for embodied AI.

And I will pass the controller down. Thank you very much. And please, go ahead. Thank you very much.

>> ANDREA CAVALLARO: Thank you for your kind introduction and thank you for the opportunity to share our experience with multimedia data, AI, and robotics within this audience.

We talk about what we learned through an international collaborative project called CORSMAL where we looked at the problem of human robotic interaction through objects.

So it's physical, human-to-robot interaction through objects that the robot didn't necessarily seen before.

So the idea here is to enable what is called in the AI community open vocabulary, understanding of the scene, and enabling robot to interact effectively with humans.

I will cover the experience we have gained in generating multimedia data for robotics, sharing those data with the community, and helping the data being used for training in such a way that the models themselves that are built in different laboratories are then shared and the community can build on those.

And then I will talk about the power of AI when embedded in robotics. And in particular, the ability to give natural language commands to robots and how this information can be seamlessly merged with the multimedia senses that the robots like the one we saw just before the coffee break are embedding today in their body.

And finally, we'll see how today's AI enable us to condition the vision, the audio of robots through natural language in a way that allows us to is really in-home and assistive applications.

But let's frame the comments I will be making during this presentation. The task we developed some six years ago when we started the project CORSMAL, which was a collaboration between three European universities, one in Paris, London, and Switzerland was that of enabling robot to understand the behaviors and intentions of humans that were manipulating containers. And without having seen previously those containers, so without having 3D models, without having videos of those objects prior to the interaction with the humans, enabling the robot to understand what were the physical properties of the object, the weight, the stiffness, for example, and also the intention of the human.

Here we're talking about dynamic humans to robot handovers, in such a way that we could accomplish a safe handover between the human and the robot.

This project was defined in open access and open science story, the funding bodies that supported this. And since the project we are contributing our benchmark for human to robot handover to an international competition that takes place every year at the major conference of robotics that is called ICRA.

And at the other large conference in robotics called IROS where we share the experiences of the competition.

The importance of multimedia data for embodied AI might be summarized with this diagram that shows that the sensing abilities of a robot must be able to tell apart in this particular task the human hand from the object to be captured. Here we have a problem of safety in addition to the problem of accuracy.

And also in many cases we cannot use the traditional liters, because we have to resort to the stereo cameras to give the shape of the object to grasp.

What I'd like to focus on is also the sheer amount of data that this type of competitions are necessitating, and that was mentioned also by the previous speaker.

So through the project we generated a benchmark. In this case we're not only dealing with data, we're dealing with humans. So the benchmark has got to define how the human should behave in respect to the robot in order to have produceable results.

Here the idea is to reproduce the same experiments across different spatial locations at different times and with different robot embodiments, and therefore we define a benchmark that's now used for the specific competition that takes place every year at ICRA.

The benchmark and the tooling for the benchmarks are available at the links you see in the slide.

In this case, usually when we are talking about multimedia data, we think about only the digital elements of the data. But when we are thinking about embodied AI, when we're thinking about robotics, we need to have physical objects. So in addition to selecting the way to compare the behaviors of the different robots, we had also to select a range of objects that were available worldwide for the different teams to be able to acquire. And that was part of one of the protocols of the generation of this -- of this project.

And as I mentioned earlier, the protocol, the data, the objects were then used for different teams across the world to generate their own AI models based on these specific tasks. And these models are available open source for you to use if you're interested in benchmarking your embodied AI application.

And if you want to compare your results with respect to the results that are available worldwide, we have available a leaderboard that is updated with the results of the yearly competition.

And on the bottom right you see the link to the web page where all the resources about the project are available.

Now, I would like to frame embodied AI in terms that might be more familiar for any of you who are not necessarily some robot sifts -- roboticists. Here we are emerging the data from your ChatGPT input to whatever the sensing information from the robots are, not to produce texts, but actions.

Here we're going from a human language instruction to an action in the physical world through the sensing information that's coming from the robots.

Here we are merging natural language, multimedia data, and the embodiment with a particular robot.

And the idea is to condition what the robot is seeing through natural language instruction so that you can produce ultra aggressively like ChatGPT actions for the robot to have an influence on the world it operates in.

And this is what we've been doing in a couple of works that we'll briefly present. Here is about using a language model to seamlessly merge the semantic information that you can extract from the language content, and what you can see from the multimedia sensory information.

Using models like for example, CLIP that stands for contrastive language image pretraining, we are able to explore the Internet level training by using current views of the robot without having the robot seeing previously objects we are mentioning with the language.

And that's the power of tools like CLIP that has been trained on 400 million pairs of images and text.

And that is used over 500,000 words of the English vocabulary.

And all this allows us to really have proper multimedia not only in terms of sensory information, but also in terms of processing within the embodied AI.

And so we use the richness and the semantics of the human language to select what we are interested in in the sensoring information coming from the robots.

One example here is a work we'll be presenting in ten days in China at IROS 2025. We trained a robot purely in simulation to pick some particular objects and place them in a point of interest.

So there are different action there. The selection of the object, the grasping of the object. So a synthesis of grasps to be undertaken. And then the proper positioning of the object perhaps over another object in the scene.

And we were able, with zero shot, to simply pass what we have trained synthetically into a real robot.

So in this case, the typical synthetic to real gap is no longer an impediment, at least for this type of application where we're using natural language to condition the sensing information coming from the multimedia sensors available in the robot and there is also sound in a number of applications to be able to achieve rather complex tasks where the robot gets instructed by humans through natural language.

If you're familiar with AI pipelines, here's just an example of how this specific work has been implemented. So on the bottom left you're seeing the human command. And the view from the robots.

And the rest of the pipeline is taking care of merging the information from the human and the sensoring information so the robot is pick and place a certain object.

A second work in here could be related to the previous presentation where we had the robot entering a certain room and putting in order different objects.

Unlike the state of the art where other robots would need a 3D model of the objects or a video of the objects they are having to interact with, in here just a picture of an object is sufficient in the robot to have the ability to locate the 6D possess.

So the position and the orientation of the object in order to be able to understand how and where to pick it and place it correctly.

So in this case, we are using the open vocabulary ability of visual language models in order to enable our robot to operate on an object that the robot didn't see in training.

And this is what is made possible by multimodal models.

I would just close with a couple of examples. In here you have on the left the picture of an object of interest. Actually of a scene of interest. We are giving as an instruction a specific object in a new scene. And because of the ability of the system to associate a natural language word to a particular object in the scene, we are capable not only of locating that object in the scene, but also of the [?] 6D pose, meaning position and orientation so that the robot can go and interact with that particular object.

And I will close with a set of examples there where you see on the very right the results from the model we have developed. And in other columns you see other solutions and the coding you see there is the coding in the RGB color space that is mapped on to the specific estimation of the 6D pose of the scene.

The novelty of the work in its ability to enable the robot to see in a 6D possess of an object that it didn't see in training.

Based on the experience we gained with this project, what's next? Well, as we show in standardization, the importance that was mentioned by previous speakers is interoperability. But I would like to stress three points that I believe are fairly important for AI robotics and the humans as we mentioned today.

So transparency. We saw in our presentation in session 1 that robot was informing users about their battery status. I think we could now go several steps farther compared to that.

Robotics embedded in a normal environment like we saw just before the coffee break, means that the robots are looking at us most of the time and in order to interact effectively with us, they will need to understand our emotional state.

How do we manipulate that information and how do we ensure that information is communicated to the users, that's extremely important.

How choices are made and how these choices are communicated to the user is very important. So that's transparency, the explainability of the decisions made by the robot.

Robot will fall. Robot will create some safety issues. So it's also important that there is some forms of traceability of how decisions have been made, and perhaps on collection of multimedia data that are corresponding to those decisions. At least for a certain period of time so that we can recover back the reasons why a certain mistake had been made.

And finally, the last point I want to make, robots are with us and we saw it today during the coffee breaks and there will be more and more embedded in society. So we have to find ways to standardize how we expect those robots to uphold human rights and also to adapt to the local norms of the communities that they will be embedded in.

And this is a very interesting interdisciplinary problem. We have started recently a larger appeal project with the 18 PhD students working on aligning AI to human values. This is a very, very exciting area where we are benefitting from the experience of philosophers, social scientists, cognitive scientists, technologists like me to try and understand how we can align the ability of AI that we saw in this presentation, to human values and principles which is becoming one of the most pressing opportunity for research and important risks for society today.

Thank you very much for your attention and these are my contact details.

(Applause)

>> JUSTIN RIDGE: Thank you very much. And there was some links in those slides, and I think we will have the presentations available later on.

So if you -- if you were interested in those links, you can get them online later.

Now Stefano, we've passed the time for this session. Do you want to move straight on and hold questions -- people can ask later on?

>> STEFANO POLIDORI: If you want to have the one question, you can do it because we started ten minutes late after the coffee break. Thank you.

>> JUSTIN RIDGE: Would anyone like to ask any of the panel a question? If so, please indicate. But state it very briefly.

Anyone with a question? I've scared you off enough. Okay. If you have questions, I'm sure the speakers will be available after the session. Please state it briefly.

>> Briefly. Which conditions are liter not useable? Liter.

>> JUSTIN RIDGE: I think this was Andrea. Under which conditions is liter not useable?

>> ANDREA CAVALLARO: If you have some transparent objects, then the ability of some of the depth sensors to understand the distance at which objects are is somehow limited. And stereo cameras will work best, that's in indoor settings.

>> JUSTIN RIDGE: So it's due to the transparency of the October.

>> ANDREA CAVALLARO: Yep.

>> JUSTIN RIDGE: Okay. All right. As I say, I'm sure the speakers will be available at the close of the session. I'm sure you can approach them with any further questions.

So let's thank them once again, all three speakers for session 2. And very much appreciate it.

(Applause)

>> STEFANO POLIDORI: Thank you also to you, Justin, for running us through session 2. And we continue directly to the session 3. So I'm inviting the session 3 moderator and speakers to come up to the podium.

While your name plate is being set. Very good. Okay.

Okay. I'm very pleased to invite Lukasz Litwic to moderate session 3 to lead us through. He's from Ericsson, Sweden, and he's also vice-chair of Study Group 21.

By the way, I think I never mentioned who I am. I am Stefano Polidori, I'm the counselor of Study Group 21, in case you were wondering. Anyway, Lukasz, thank you.

>> LUKASZ LITWIC: Thank you, Stefano, and welcome to this final session. This session title is entitled embody AI, the future of collaborative multimedia connectivity.

In this session we turn our focus to connectivity and in particular to some of the questions. What are the requirements for connectivity, which are foreseen which developing embodied AI towards collaborative communication.

And how provision of such connectivity can enhance the performance and, in turn, and quality of services of embodied AI.

I'm very pleased that we have three distinguished speakers with us today who will guide us through this topic. Our first speaker is Mr. Hang Liu. He's from SparkLink alliance in China. The topic is integrating embody the AI with short range communications. Challenges and standardization opportunities.

Mr. Liu, please.

>> HANG LIU: Thanks to Lukasz for the introductions.

So today we will talk about the integrating embodied AI with short range communications, challenges and standardization opportunities.

Since I'm presenting iSLA, international SparkLink Alliance, some of you may not be familiar with that, please allow me to give brief introduction.

The SparkLink Alliance was officially established on September 2020 and further as international SparkLink wireless short range communication alliance, iSLA, 2023 in China.

So the goal of this is to promote innovation in 'next generation wireless short range Communication Technologies and prosperities of the industrial ecosystems.

From the left, we can see now many of the indications placing high diamonds on low latencies, [?]. Based on these requirements that the SparkLink technologies was designed to -- and considered at least the following features.

Low latency, high reliability, precise synchronizations, and high concurrence.

And with a target to meet the ultimate performance requirement.

This picture displays the organizational structures of iSLA. The General Assemblies serve as the highest decision-making bodies, in the operations council which means the Board of Directors represents general [?] in decision-making.

We have the expert committees, academic committees. The expert committees provide valuable suggestions and guidance of this. The academic committees is responsible for the research -- the technology research and providing proof for standardizations.

There are four working groups and four industrial promotion groups in this cadence. Among these working groups, the requirements and extended working groups is responsible for the formulating of standards.

There are four industrial promotion working groups which focus on the industry promotion of smart automotive, smart terminal, smart home, and smart manufacturing.

Okay. Let's go to -- dive into the integration of embodied AI with short range communications.

Now the research of AI in communication domains is mainly divided into two parts. One is AI for network and the other is network for AI.

For AI for network is to enhance the communication systems and network intelligence or performance via AI engineering issues.

Which is little bit emphasis on engineering issues or engineering implementations. However, it should be noted that the success of engineering implementations is reliance -- besides the success, contributions of the AI algorithms, it's standardization in the face between the AI applications and the communication stacks.

Because the communication -- this interfaces enables the communication stacks to provide, reach, and unify informations for AI applications.

Regarding network for AI, the key is to improve Communication Technologies and network architectures to meet the diverse technical demands of AI services, which is the focus of today's topic.

Of course, that we know that AI for network and network for AI, promotes each other and jointly driven better networkers and better services.

There are three communication scenarios that are considered for network for AI. Human AI large model interactions, internal communication of AI agents, and collaborative communication among AI agents.

It can be expected that with the rapid development of AI technologies, short range communications will play important roles, especially in collaborative communication among AI agents.

This page goal to we'll talk about the main challenges. So first keywords I would like to mention is the multimedia attribute. From the left side we can see there are already lots of products, AI products in the market in smart terminals, smartglasses.

That all these products processed heavy attributes, multimedia attributes with considering embedded AIs that these multimedia attributes is even more pronounced due to the diverse variety of sensors of embedded AI.

One of the main challenges is how can we achieve a unified, flexible, encapsulations of the multimedia informations that traditional mainstream media encapsulations may not fully meet. Especially due to the variety of the type of data.

Also that for the characteristics of business may be also different because that now the mixed transmissions of burst -- with post burst and periodic data.

The second keywords I've yet to mention is task-oriented perceptions that we know embedded AI is the perceptions, it's not merely simple information transmission. But task-oriented perceptions, why we have the traditional streaming medias may not fully meet the requirements.

Take voice communication as examples. In traditional voice communications, we know that the sampling reads as well as reliabilities are not quite high. Hour, if you embedded AI, the voice is used to trigger the command. Then the reliability should be guaranteed.

The key, the importance of the perceptions of -- the perceptions of embedded AI is to understand the environment, including identifying objects, the events, and states reasoned the key is to understanding the user, fundamental intentions and instructions.

And we only need to provide the necessary data for decision-making.

Additionally considering the different modalities exhibit complex relationships such as dependency, substitution, and complementaries to meet the task requirements, how to define the quality of each modalities considering the task requirements. And to meet these challenges.

The third keyword I want to mention is compatibilities and regulatory compliance.

For the compatibilities, we know that there are already traditional comprehension or representation [?] as well as AI-based -- AI-based methods in order to define unifying -- unifying encapsulate format, we need to consider both traditional methods. It will be compatible with ecosystems, but also need to accomplish the need of AI-based methods.

For the regulatory compliance, it should be noted that the rapid development of the AI technologies that AI risk also need to be taken into consideration. Nowadays, many countries introduce regulatory requirements. Especially in labeling requirements of AI-generated content.

Which means there are strictly need -- strictly need to distinguish from AI-generated content and anomaly generated content. So they need to notify the users that this content is AI-generated. Also maybe that the producers and also disseminators of the AI content need to be considered.

Therefore, for embedded AI, if some of the content are generated by AI, we need to also meet the regulatory requirements.

About the main challenges of the embedded AI, now in response now the SparkLink is working on one standards with a titled AI multimodal information encapsulation and transmission optimization started with this August.

So with a target to define AI multimodal information in encapsulation and transmissions of multimodal informations, these standards will cover flexible and unifying standards. Also we're considering to compatible with traditional [?] format as well as AI-based metas.

Also need to consider regulatory requirements of -- especially labels of AI-generated content.

Of course we're also considering to optimize the transmissions of multimodal information to reduce the transmission of [?] and improve the efficiencies while ensuring the qualities.

That's all for my report. Thanks so much for your attention.

(Applause)

>> LUKASZ LITWIC: Thank you, Mr. Liu, for excellent timing of the presentation. And as in previous sessions, we'd like to propose that we open the floor for questions at the end of the session.

And with that, I'd like to move to our second speaker, Mr. Jorge Pena Queralta. Dr. Pena Queralta is works at center for artificial intelligence in applied sciences Switzerland. His talk is embodied agentic AI, distributed applications from edge to cloud.

>> JORGE PENA QUERALTA: Thank you very much for the introduction. There's been a lot of talk today already about embodied AI, especially on the talk related to privacy and security. There was a lot of mention about agentic AI.

I would like to talk a bit about what's at the intersection of those and where I think we might be leading, what are some key challenges that are still unaddressed from the perspective of connectivity, but also mainly deployment, orchestration, and management of potential future agentic embodied AI systems.

I'll go very quickly through this slide because I think there's been a lot of context of what embodied AI is and where we are today. But I'd like to start with the relatively wide claim that we might be close to robotics ChatGPT moment today.

So most of these demos are already slightly old from earlier this year. Figure for example as new robot from yesterday and the videos aren't playing fully well here. Nvidia is working on VLAs as well. And we have a lot of [?] we have seen Unitree's robot a moment ago during the coffee break. Many Chinese companies working and driving at the forefront of this technology. Humanoid robotics, deep robotics, and one thing to highlight, because the focus is different, not just scale for the sake of scale, but showing emotion. This is an example of reinforcement learning, in my opinion.

The basics I want to take away interest here is today the landscape is fragmented. There's very impressive skill sets. They're getting better and better. We are progressing towards longer horizon tasks, things that like like more complete systems, but we actually have a very fragmented landscape of what I would call skill sets.

And this one's come from mainly two technologies either reinforcement learning or vision-language learning.

What I would also like to emphasize here is that my belief is that even if we might be close to the ChatGPT moment in robotics, we are still relatively far from general robotic intelligence. What I do believe, however, is that we might not be so far from generally useful robotic intelligence.

And I will tell you in a moment what I think that is and how we might reach there and also the challenges that we still have to reach to deploy this at scale.

To put it with other examples, physical intelligence, theres a mention of pi 0.5 a moment ago and the skilled AI are two very prominent start-ups in the U.S. in this area. They have various in funding, billions in evaluation. They can do impressive things. You can ask the robots from physical intelligence to clean the room, any room, and in an apartment it does a great job.

But when it stands, did you say where is the plant in the room, did you see my keys, it can't answer because it was a specific skill that was deployed on the robot. This is still very far from general intelligence.

A skilled robot you can ask give this balloon to this person. It can interact with people potentially safely. You cannot really say wait here and greet my friends.

It could probably do it, it will do it in a very inefficient way with a large foundation model running while the robot is just there waiting doing nothing. This is not an efficient use of computation, potentially.

I want to also highlight this actually an agentic word missing here in the title. But there are at the forefront prominent open source projects on agentic embodied AI, specifically tied to ROS.

The first work that came out was Rosa from the ROS agent. It's not active project at the moment. But it's a European project from Polish start up AI called RAI framework that enables agentic AI.

What I mean by that these systems are frameworks are a combination of multiple agents working together and orchestrated. And I think this is where we are actually largely leading.

Also I'd like to borrow this famous now image from Nvidia on how AI or where the AI frontier is going to. It's been already over a year that physical AI or embodied AI, they're not the same but we could use both interchangeably for a moment. The next big step, next trillion dollar industry or much more, and I'd like to put the focus at the intersection of agentic AI and physical AI.

Now, agentic AI, this slide, it's my understanding, refers mostly to what I'd call digital AI agents. So agents that control your browser, things that ChatGPT can do, create documents, search the web.

We can also deploy agents in a robot. You've seen the demos outside during the coffee break. You can talk to our robot. But this is not physical AI. It's still interaction. It's not very,

very different from talking to a computer.

Now, to deploy agentic AI on a physical system, I would claim that this is several degrees of magnitude harder than on a digital system. So there's multiple challenges from safety and privacy, but that has been discussed already today.

But controlling a browser by an agent is very, very different from taking control over a robot that might operate in very different environments with a variety of people more or less crowded and do very, very wide number of tasks.

I want to give quickly a couple of examples of things that we are working on in applied sciences in agentic AI. These are projects under heavy development. But just to give an idea what I mean.

This is on assistive robotics or assistive mobile manipulation. We're working with a start-up that develops this [?] and adding intelligent robot arms. Unfortunately robot arms for People with Disabilities and wheelchairs are really, really limited today despite the advances in AI and robotics.

And we are trying to bridge that gap and build interfaces to make this more intelligent. And easier to use. It can be through voice or through the joystick by making the AI system better understand user intentions.

Now, without going into details of the diagram on the left, what we are building here is not one AI agent, this is not just adding ChatGPT to the robot, it's a collection of mobile agents.

We have a specialized model. It could be a large model, it could be a specific AI, it could be traditional robotics.

A system that takes care of for example the navigation safely. We have another system that takes care of the manipulation. Another system that takes care of the interaction. And potentially many more things.

And all of these this is orchestrated. The amount of communication that goes in here, I'm not touching the topic of privacy and safety on purpose today, but basically you could theoretically have all these agents deployed anywhere in between the robot at the edge, a computer sitting in the same room nearby, or all the way to the cloud.

So the larger that these systems or multiagent systems become, there's also the challenge of where are they deployed. If we want something really advanced today, even with the best hardware, not all robots will be equipped with such hardware, but it's also not really available today.

The second quick example, we're working on deporting this to humanoid. I have here a small demo with a robot dog. What we're working on here is chat and voice interfaces. Still a multiagent system that this orchestrated.

Specialized agents from manipulation, navigation, interaction. And you can do things like are the prompt here is an educational example. So wait here until you see a person sit down. Then stand on your legs followed by laying down. It's just a silly example.

My point here is this is not, let's say, in the training data. This is not predefined. It's something that comes up. Our agent can adapt.

You cannot really do this with many monolithic AI approaches today. In theory you could, but you would have this Large Language Model in a loop which is inefficient. My claim here is that we can solve this task easily. And now we are building monolithic frameworks to take care of the robots that cannot take care of these tasks so easily.

I believe the solution is at the intersection of both. And what we are building here is an AI system that can reconfigure a traditional robotic system. So can reconnect different perception, control, interaction, actuation models in different ways depending on the task.

So it's a combination of foundation models with traditional robotics.

You could tell our AI system something like hey, from now on start counting any bike that you see. And after half an hour go and ask how many bikes did you see in the last half an hour.

This is not something that, again, monolithic AI systems or these foundation models can solve, because it's not the way they work. So we need something else. And I think that this something else is distributed.

And this is basically one of my final slides. The main message I want to bring here is that agentic embodied AI is distributed today and it's going to become increasingly distributed as the complexity and size of these systems increases.

So to start with, generative AI in general, there's already some key challenges in deploying generative AI at scale in robotics and IoT. One of them I mentioned was where you deploy this and is this deployed distributed across edge and cloud. Do you have smaller models in one place, larger models in another.

But also talk about multimodal data and specific data to robotics. This is a big difference between digital agents and physical AI agents. When we have an agent that controls your browser or does anything on your computer, I claim that you almost have enough with images and JSON data or text data. With that you can cover almost any use case.

In robotics we have a lot of different type of data. Multimodality, how to transmit this data, how to take care of this and standardize this is a big challenge.

I'd like to mention five key challenges that we are hoping to start early next year. On deploying embodied agentic AI at scale. This is much more challenging.

Challenge Number 1, agentic AI characteristics, this is where I would like to focus these five challenges.

We have today edge to cloud systems that are very efficient. The technology's mature. But usually when we deploy an application, we know in advance how much computing we need. We know in advance where the communication will be. We can get an idea of latency requirements, bandwidth requirements. This is not true anymore for agentic AI.

And the core reason is that context plays a very significant role here, and is not predictable. This is something that actually we have to start working on in learning for example to predict context. What I do mean by context?

Context means the query from the user. So we can ask simple or more complex things. Context means the embodiment, if we build an AI agent that can take over any robot, the actual amount of computation or computational resources, memory, or connectivity that the application might needle depend on the embodiment. So on the robot body. Context means the environment where the robot is deployed. Even the same application, the same robot at two different times in two different scenarios might require different amount of computation and different amount -- or different connectivity requirements and also the task complexity.

We might have a robot that can clean a room or the entire apartment. These don't have the same instructions.

These are not the same in form and in components, so because we are giving as was mentioned earlier, talks this agency to call tools to act on the environment, the amount of things that will be running in computers in cloud servers, the amount of things that will be talking to each other are not predefined in advance. They may say now I need to run this other conceptual model or control system and we cannot necessarily know or predict in advance how many of these and where they will run.

So I see this as some key challenges that are completely unaddressed today.

I'll speed up a little bit now. But without going into detail, some of the key innovations or processes, I will focus on cognitive orchestration.

We believe that we need to build an orchestration system in terms of connectivity and deployment and managing these applications and where they run. That is also an agentic system itself.

So one way to approach this is to build an agentic AI system to manage, deploy, and control other agentic AI applications.

This is -- this will need to support dynamic infrastructure interaction and we will need to define the new learning paradigms. So federated learning has been an EU priority for a while now.

In addition to learning, let's say, fine tuning foundation models for the specific applications, we might need to start thinking about learning usage patterns in terms of computational resources and communication resources of such agentic AI applications.

So we might have more let's say metadata or at the level of meta operating system, more things to learn.

And our ambition is that this will lead towards an autonomous AI infrastructure for AI applications that we will hopefully democratize such distributed AI systems and that will bring to reality the robotic things in a sustainable and efficient way.

I skip over this, but this embodied AI, physical AI is a growth area that we've identified that we are pushing forward significant with significant efforts at the Zurich university of applied answers and with the new lab I'm leading with infrastructure.

This is just a small list of the robots that we have. And I take this opportunity to mention that we are co-organizing or co-initiating the DAVOS tech summit in summer next year. If you're interested, reach out, look at the link, we have a promo video here. So I'm not fully involved myself, but I can put you in contact with key persons there. And we're hoping that the suggest of the WUF, of course, but on the tech area and especially little on physical and embodied AI also takes place and builds up to this.

So thank you very much.

(Applause)

>> LUKASZ LITWIC: Thank you very much for the presentation. Moving on tour final speaker, Mr. Abhishek Gupta. Mr. Gupta is founder and CEO of Open Droids Robotics based in United States of America and joining us remotely for this session.

The topic of his talk is building an open source embodied intelligence future.

Mr. Gupta, please go ahead.

>> ABHISHEK GUPTA: Thank you, Lukasz, for giving me this opportunity and, of course, it has been an amazing session and it was like so much learning. I really feel blessed to be part of this whole ecosystem now.

Let's start off real quick here. And real quick, Lukasz, are you running the slides or I can actually share my screen and then run them over? .

>> LUKASZ LITWIC: Both ways work for us.

>> ABHISHEK GUPTA: I can actually share my screen and we'll take it from there. Just give me one quick moment. Technology.

All right. Well, Ladies and Gentlemen, we are Open Droids. Let me take a quick moment to talk about ourselves, just to give context and like a set context for the conversation today.

We are developing a single dual arm as well as special purpose robots. Of course, embodied AI is a core to our presence and our being and it's very true to our mission and vision of creating a world where robots can potentially unlock opportunities for individuals for businesses and literally transform how we will interact with these amazing machines in the future.

So with that being said, let me actually talk very quickly about, like, how I got associated in this space and why this is very core to our existence and being as well.

I'm hardware design engineer by background and have been developing these kind of robotic systems, sensor technology, work considerably on IoT, used to work for the U.S. Department of Energy as well.

What we realized when we started the company is that robots typically have been looked out as, you know, unintelligent beings where, you know, they embodied intelligence aspect was kind of a foreign concept or an alien concept in the past.

It was typically predetermined in terms of sequencing that it's like the controlled systems are feeding into the sensor fusion of the robots and then that's how they're interacting with their environment.

But I think the story flipped on its head with this ChatGPT moment and now suddenly everybody with AI and AGI coming in play, everybody's talking about it and it's actually the true unlock of how robots will be acting as companions, as assistants, servers, as well as potential friends and who knows even lovers in the future.

And how this is happening with respect to embodied intelligence is that this interaction is now, as the story flipped, now it is about the interaction with the environment and not just about preprogrammed sequencing.

So how does this typically like look like and why this is so relevant even in our case.

With the aspect of physical interaction, it really accelerated the process of learning. We talk about reinforcement learning or even the learning loops is actually feeding into the behavior loops which are going to be kind of pretraining the robots to be getting more and more towards autonomy.

Where we are today, and I speak in full disclosure is that, you know, with most of the companies out there and I mean I have to give credit to Hang and Jorge as well, they talked a lot about some of the things that I'm going to be talking in our presentation as well.

The unlock moment, we all saw what happened when they released their latest robot, they're also working extensively towards the aspect of autonomy. Everybody, in fact, Jorge covered considerably about that aspect as well. And this is where we actually feel that the -- currently where we are at is with respect to collecting and aggregating considerable amount of data.

Now this comes through doing a lot of datasets that we are going to be potentially collecting over this period of time. Our core to existence is open source. That's what we truly believe in and we believe in the power of community towards this unlock.

And that's why we are definitely going to leverage heavily on some of the open source capabilities and tools available. Which is pretty much happening at this point and at an exponential pace.

And with the community behind it, this is our thesis, our methodology as well as our ideology and vision for the company is that the community and the society as well as hobbyists, researchers, just as they did in the early stages of computing, is exactly what's going to happen for this unlock moment when we talk about robotics.

So with that, we talk about initially collecting those large volumes of data through teleoperation-based control. In fact, our team has developed a teleoperation capability, the robots can be controlled remotely across the globe.

And with the consideration amount of learning that we get and the collection of the data through those datasets, leveraging some of the models that are at our disposal, we want to get towards that level of autonomy.

In fact, this is where I actually talk about what it will mean for our business as well as moving forward for considerable amount of, like, companies in the robotic space.

Service robots is turning to reality. This is where we focus our attention towards starting off with light industrial, because our core heavy industry is not our focus right now. And that's where we actually draw the line and that's where we see that the embodied aspect of intelligence in robotics will lead us towards these light industrial applications, assembly and the initial part, all the way towards leveraging this for businesses.

I talk about hospitality, road service robots coming as a robotic server in restaurants within hotels, and then potentially into health care facilities as well.

And how this differentiates is because with the aspect of embodied intelligence, these robots will have the capability to learn from that environment to actually be accelerated learning loops so that they can be used effectively and functionally within a very short, condensed cycle towards deployment in such facilities.

And where the true value lies is because we believe that, you know, one of our thesis was that we want to be building robots with arms.

So when we talk about robotic arms, and collection and leveraging of that massive amount of data, we can actually be close to using these robots as potential replacements for human labor.

This is what truly exciting and what we've started to see as well. Assistive technologies is another huge area.

One of the deployments that we're doing is we got an initial LOI for this, deploying 300 of our robots across multiple health care, particularly the addiction as well as rehabilitation facilities all over Florida U.S. And what these potential robots are going to be doing is through the aspect of embodied intelligence, they're going to be collecting the data within their environment, getting framed in real time, and then being used in these facilities for two tasks.

Collecting blood pressure and blood oxygen. And of course with having the arm, they are able to navigate into patient's rooms, open doors, and get in front of those in-house capabilities as well.

How we leverage the aspect of community is we want to leverage -- of course ROS is now becoming the global standard. And of course -- that is what I truly believe is where unlocks happen. Is something that gets into the standardization. And this is what I'll be talking about as well in a little bit, is we need to get towards more and more standardization so that rapid infrastructure development happens. And that adds to better hardware, better infrastructure, at lower costs.

Which is absolutely the ChatGPT moment for embodied AI.

And of course, leveraging the power of the community. Just as we did for the early days of computing, this is what we potentially are going to see in the future for robotics as well.

Now, I see a lot of current challenges, and I'm not saying that I'm being pessimistic about the fact, but of course these are the things we currently solving for and where we actually can leverage a lot from forums such as these. I love the initiative today. We talk about connectivity.

I think one of the biggest challenges of our times is proper connectivity. It leads to latency issues, because we have actually typically seen that we can experience latency of one second when the bandwidth is not appropriate.

And we've also seen on the flip side that if it's like 100 MPS speed, we are able to see almost real time navigation, perception, and manipulation capabilities within robotics. Even if it is done cross-country or cross-continents.

And then there's no unified -- current no unified global standards that exist. And that is also potentially creating kind of a bottleneck effect which we can unlock potentially once we start creating more standardization.

And I also believe that, you know, ITU can be instrumental and quite exponentially helpful in that aspect as well.

So what it leads towards is those enablements that we talked about. We need reliable low latency networks right now. Need better capabilities toward edge computing, and more and more collaborative projects. And that's like I said, it's the core thesis of how we formulate and created this company is we want to get towards that aspect.

I truly believe with the capabilities, with the community behind us and of course with the help of ITU, that is absolutely possible in the future if we have a strategic vision and we're working towards that goal.

And we talk about two aspects here. We talk about the integration of embodied AI into the blueprint of how telecommunication networks will be formulated and created in terms of blueprints for the future.

And create a trust framework for embodied AI adoption. Because I think that is essential, because at this point, it is a huge aspect if there's leakages and security situations where we can actually be much more prudent if we had a trusted framework.

So thank you, everyone. And I'll -- anyone who's interested to talk about robots, who wants to know about some of the cool things that we are doing, as well as how we're working on -- in this direction, continuing conversations so that we can build better embodied AI capabilities or anything in general, you know, if you want to have chat I, reach out to me. I'm available at the coordinates that you see here on the screen.

We are available on our website and you see our social media handles as well.

So thank you, everyone.

(Applause)

>> LUKASZ LITWIC: Thank you, Mr. Gupta. I'd like to thank again all our speakers. So please join me in thanking people for excellent presentations.

(Applause)

>> LUKASZ LITWIC: And I'd like to open the floor up maybe for one question, if we have time. Please.

>> This is a question for everybody who talked about unlocking moments or this technology. Why don't we meet the robots halfway? Why don't we use BCI for the people who the robots are supposed to help?

Why don't you just put a graphene ribbon on my brain, hook me up so I can transmit the information I need to so that the robots learns is so much quicker.

>> LUKASZ LITWIC: Thanks for the question. Would anyone like to take it?

>> I think this is outside of my field of expertise, but we're still far from that kind of expertise. Not my field of

expertise. But I would say we're quite far from that and learning in that way is not something that we can do today.

>> LUKASZ LITWIC: Thank you. Thank you very much again.

And I now leave the microphone to you, Stefano. Thank you.

(Applause)

>> STEFANO POLIDORI: Thank you very much. At this point, I believe session 3 is closed and we are going to go towards the end of this event. And I'm going to invite Noah Luo to come on the stage to give us his closing remarks.

So Noah Luo is from Huawei and the chair who is running this week and next week. We will conclude next Friday the long standardization work of Study Group 21 of these two weeks.

Mr. Luo, the floor is yours for your closing remarks.

>> NOAH LUO: Thank you very much. You hear me? This is Noah Luo. Okay. Good late afternoon, everyone.

Distinguished speakers, delegates, colleagues, and friends, so it's my great honor to deliver the closing remarks of this workshop on embodied AI and the multimedia technology standards.

I would like to remind you that this event has been organized by the ITU Study Group 21 in collaboration with AI for Good and having heard all the excellent presentations, I can safely say this has been a very successful workshop.

Three excellent sessions, excellent speakers, we have witnessed truly inspiring exchanges from visionary use cases and services to innovative contents and application frameworks.

And the forward-looking perspective on the future of collaborating embodied intelligence.

Yes, those distinguished speakers, industry and across the world have an understanding of how embodied AI is reshaping our communication, creativity, and human machine collaboration. And many other things, of course.

Your contributions have demonstrated that embodied intelligence is not merely about technological advancement, it's also about enabling machines that can sense, interact, and learn in harmony with humans.

It's about building systems that extend human potentials, while [?]

I really like one sentence from a talk. So keep humans in the loop. I think this sentence is also a centerpiece in many other presentations as well.

Today's dialogue reflects the spirit of global collaboration and defines ITU and its Study Group 21 as a home of multimedia related to standardization.

We'll continue to serve an open collaborative and inclusive platform for experts to develop standards that are interoperable, responsible, and of course respective of human needs and benefits.

Together we will ensure that embodied AI technology involved in ways that are trustworthy, sustainable, and beneficial to all. As we look ahead, September 1 is committee to expanding the frontiers in organization, exploring new interface between AI, multimedia, robotics, and the communication networks.

We will deepen collaboration across study groups, strengthen ties with partners and industry and academia, and welcome Johannesburg to join in the journey.

Let's work hand in hand to ensure that AI [?] mandate, not the other way around. [?] Sustainable Development Goals or SDG, our collective efforts embodied AI will have built a future where innovation I believe society connects community and empowers every individual to master technology for good.

So in closing, I would like to express my sincere appreciation to all speakers, moderators, organizers, and AI Industry Alliance, organizations from China who supported our coffee break. And all participants. Your patience, your dedication, are all that make this workshop a greater success.

Finally, I think my thanks also go to last but far from being the least important, to Excellency Kim, the composer for Stefano, Hiba, and probably many others. I think they're effective work is a key element in making today a success.

So thank you all for your valuable contribution. I look forward to seeing you all participation in the framework of Study Group 21, standards and activity. Do not hesitate to come to me if you wish to join and come to me embodied AI standards in Study Group 21.

I also on a present note favor to today's event. You may curious to know why. Let me tell you.

My son, 15 years old, a high school student loves robotics very much and he is active member of his school colab and captain of his small team. They are working very basically and in preparation for some international competition.

Okay, he really loved to join today's event online, but due to the conflict with his school activity, he couldn't. So I will bring back the good presentation for him.

Okay. This is my personal reason to are thankful.

Finally, I think I would like wish you very great weekend. And I know some of you are, you know, trying to leave and catch a flight this evening. Then I wish you very safe journey and bon voyage.

Thank you very much.

>> STEFANO POLIDORI: Thank you very much, Chair. Yeah, perfect in time. I think everyone had enough of our event, but if you didn't, including your son, you'll be able to watch it all over because we are going to put the recording on the web page of the event.

So this will be accessible in addition to having the PowerPoint presentation, maybe posted in PDF and so you will be able to come back perhaps next week on some of them and if you wish the recording of the event.

Thank you, everyone. Have a great weekend. It has been really an enriching event, I learned a lot and I think the same is with you. Thank you.

>> NOAH LUO: Let's give them round of applause.

(Applause)

>> NOAH LUO: Thank you.

(Session ended at 18:30 p.m. CET)