

# Zero-Shot Learning - Episode 10: Inhi Cho Suh and Eugene Chong

*AI enters the world*

**Hosts:** Nancy Wang, Chief Technology Officer, 1Password

Dev Tagare, Senior Director and Head of Engineering, Gemini Enterprise and Business, Google (personal capacity)

**Guests:** Inhi Cho Suh, CEO, Niantic Spatial; Eugene Chong, Director of Product Management, Niantic Spatial

**Nancy Wang:** Hi everyone. Welcome to Zero-Shot Learning, a podcast about the reality of developing with AI from the people who are actually doing the work. I'm Nancy Wang and I'm the chief technology officer at 1Password. I co-host this show with Dev Tagare, who is a senior director and head of engineering for Gemini Enterprise and Business at Google. Dev joins us in a personal capacity and his views are his own. This is our season finale. We thank you for joining us for all of these conversations. We had a really fun time recording season one, and we're excited to bring you even more conversations with AI leaders in season two. We're going to be taking a few weeks off, but we'll be back in your feed soon.

**Nancy Wang:** For today's episode, we're going to be talking about world models, spatial intelligence, and physical AI with two guests from Niantic Spatial: CEO Inhi Cho Suh and Director of Product Management Eugene Chong. Before working at Niantic Spatial, Eugene was the president of DocuSign. We also explore physical AI through Inhi's earlier work in medical imaging at IBM, where AI helped radiologists find X-rays that needed a second look instead of replacing their judgment. We start with the assertion that language models only know what a person thought was worth writing down, but nobody has ever published a sentence about where the chair in your office is located, or how hard you have to push it just to move it an inch.

**Nancy Wang:** Yet that data is crucial for AI to understand our physical world. And as Inhi points out in the episode, 80% of the global economy happens in physical space, not on a screen. Eugene makes the case that Niantic Spatial isn't building a world model. They're actually building the layer underneath one. They start from ground truth, the real world as it actually is, meaning there's nothing for the model to hallucinate. And that matters for safety, not just accuracy, because a language model that's wrong gives you a bad answer, but a spatial model that's wrong sends a robot crashing into a wall. Let's lock in. This is Zero-Shot Learning.

**Nancy Wang:** Thank you so much, Inhi and Eugene for coming to our studio today.

**Inhi Cho Suh:** We're excited to be here.

**Eugene Chong:** Thank you very much for having us.

## 1. Physical AI beyond robotics

**Nancy Wang:** Today's topic is going to be, I would say something that I'm super excited to dig in with Dev, my co-host here, which is around world model spatial intelligence and physical AI. So maybe let's start with, you know, for our audience today, they might have maybe a surface level understanding of what is physical intelligence, right? Or physical AI. Maybe give our listeners like a primer of why are we seeing this category take off? Like how is it different from maybe the 3D modeling that we've heard of before?

**Inhi Cho Suh:** When I think about physical AI, most people first think of robotics. I think that's too narrow. We don't make the robots or the robot brains, but we really think about all things on this planet that are physical. And the most important physical assets are often the places where we live and work. And that's actually where 80 plus percent of the economy is and the GDP globally. And we're so focused on the digital screen, on around language and texting and typing.

**Inhi Cho Suh:** And we kind of forget, you know, the myriad of work, whether it's engineering, navigation, logistics, airports, refineries, think about physical places to enhance for safety, reasons for collaboration, reasons for efficiency reasons, human plus machine collaboration in a way that's aligned and it's machine readable. So physical AI, it's making the world a lot more, um, collaborative and machine readable, essentially for AI, autonomous robots and better embodied AI as well, devices and so forth.

## 2. What language models leave out

**Eugene Chong:** Yeah, I think that's exactly right. I mean, language models have been amazing as it happens. Taking a corpus, ingesting a corpus of everything humans believe is worth saying to each other can actually provide you a pretty good understanding of the world. I think the point we're at now is that we're realizing that there are some things those language models leave out that they've been unable to capture. Um, their first thing is someone has to believe something is worth saying in order for it to make it into the model. Um, I can imagine that perhaps this is a good assumption. There's no writing on the internet that says where this chair is located in this room, and what direction it's facing, and that means it's absent from the model.

**Eugene Chong:** I think the other thing, though, is more fundamentally, can you actually capture the experience of interacting with this chair in language alone? Do you know how the cushions feel when you push back against it? How hard you have to push the chair in order to move it an inch that way or this way? I think that's the moment that we're encountering right now, is how do you bring that experiential data back into a machine readable form, such that the 80 to 90% of the world's work that happens in the real world can actually be enabled?

**Nancy Wang:** And maybe like that actually brings up a good point, which is maybe we have been like overestimating the importance of language right into models. I mean, LLMs language is in the name of the model, but maybe like, I feel like what you're proposing is a little bit different, right? Which is how do you capture the understanding of the physical world to your point, where you don't really have captions, right, or annotations to describe placement or movement. So like, how does that work?

**Eugene Chong:** The reason we're focused on enterprise consumers and enterprise customers is because we want there to be a feedback loop between what agents. And I mean that broadly, people, robots, machines navigating space and embodied AI forms. We want them to be able to tie their experiences in the world with outcomes that they're trying to drive. So in this particular case, um, the outcome could be my ability to move this piece of furniture to where I need it to go.

**Eugene Chong:** And we want that to be a core part of the loop such that you're not just ingesting, um, I guess, um, ambiguous or un anchored data on how hard it is to move a chair, but you want to know that doing so in a certain manner drives a certain outcome for your business, and then that gives you the ability to build upon that, learn more and more, such that these agents get better and better at performing the jobs that they're assigned to do.

### 3. Ground truth in the physical world

**Dev Tagare:** Taking this example a little bit further, right. Let's say there is a there's dense fog in the mornings. I'm trying to cross the street. Uh, I don't really see the lights on or off. I don't see the little man blink. Um, so help me understand a little bit to the next level of detail. What does a spatial model need to represent that a language model is deficit in in, like a real world setting where there could be a wind element, there could be, you know, temperature, that could be fog, there could be so many things.

**Inhi Cho Suh:** Maybe I'll start with a really good example. One of our first partnerships was with a company called echelon, and it was really to support the Coast Guard for search and rescue. So you can't Google where should I land or dock a boat, or land a helicopter, or pick up an emergency situation? For a number of reasons? A it doesn't understand the physics of the equipment that's moving around. It doesn't understand the geographic terrain that might be uneven on a coastline. It doesn't understand the time dimensions. It doesn't understand the weather forecast. So you can't not only not Google it, you can't put it into an LLM. And often it's based on historical human judgment of like super experienced people. That's one example.

**Inhi Cho Suh:** Another really good example is we've been working with in um, in the industrial category for oil and gas, for looking at data that's actually subsea and recreating,

**Inhi Cho Suh:** um, machines that may be subsea. And, and when you think about how that is historically captured, which would be autonomous or unmanned vehicles underwater, but it's at areas that are murky. Dark sand is shifting. But what you want to do is maybe maintenance around a manifold or pipe or a valve, or just look at the last, you know, compliance check. It's not a place that a it's easy to navigate to for a human be the data sets are going to be complicated. Um, obviously there's a lot of different types of sensor data. We happen to have a very unique computer vision stack. Um, but in addition to computer vision stack, you can have obviously sensors to calculate frequency, you can have thermal, you could have, um, Sonar. There's a number of ways.

**Inhi Cho Suh:** And so when you think about the physics of things. The, the vast property of sensors are so much bigger so than just language alone. And hence why even amongst the emerging physical AI companies in the world model companies, as you see, different companies sort of specialize in one entry point, because even trying to make sense of this at scale is a huge challenge.

### 4. Persistent and real-time data

**Nancy Wang:** And speaking of scale, right. So even with, for example, you know, what does that work, which is maybe doing post training, right? There's some data that's always going to be persistent and there's some real time data, like how does that work in world models. Right. What's persistent and what's real time?

**Inhi Cho Suh:** That's a good question. I don't know about the general world models categories, but for us we have a unique starting point where we have what we call we start with a ground truth and ground truth, meaning the planet. We're living on different points of interest that we've captured over history. Seasons, right? Times of day. Um, as structures

change and there because we understand that, we also understand how you might react or the flow of movement around those objects in certain times. And that becomes really important to things such as the way we develop our depth algorithm, which is super unique.

**Inhi Cho Suh:** Therefore, you don't have to have as expensive of a leader camera in order for us to process our models, because we've been able to handle what we call really poor quality, messy data. So we handle messy data extremely well. That might be in 1D, 2D and convert it to 3D. That is highly unusual versus your persistent question. Most people are asked, well, why don't you go and capture it new? And we have the ability to both capture new using standard phone off the shelf, you know, iPhone, Android phone or 360 camera as well as most enterprises have enormous corpus corpus of data. Right? So we're just kind of working with them to figure out, hey, can we make it more useful given their history?

**Inhi Cho Suh:** And the last piece is real time is something that hasn't really surfaced yet because it becomes much more complex. I don't think the architectures exist quite for it, for the database. I mean, if you think about real time, even for like traditional US SQL and no SQL queries, I mean, this is a whole order of like, uh, magnitude challenge for multimodal sensor data.

**Dev Tagare:** I've heard this term like mean time to reality being floated around quite a bit. Uh, kind of like the AGI verbatim in the language model world. And you raised a fascinating point, which was like, you know, you begin with the ground truth, which also has some equivalence in like the language model world. to my mind, you know? Baseline question is like at what confidence levels are we in the journey of physical models that we can trust them? First, some baseline of use cases.

**Inhi Cho Suh:** The great news for us is because we actually start with the real world. We don't actually have hallucination in our model, so we tell customers they can start with us today. We're not projecting what the next best frame should be. We're not. We're actually starting with like a very highly, um, precise geo referenced position model. So that's 0.1. The second piece that just becomes a little bit more challenging is, you know, your phrase like mean time to reality. That came up in a set of innovation discussions we had with ExxonMobil. So typically people talk about mean time to resolution for maintenance fix of different things.

**Inhi Cho Suh:** I mean, you use it in tech, use it in a PDR, you use it in a lot of places. they talked about as meantime, the reality was about how can they go from something that in the way they're operating in a very human and digital way to like making this thing working and functioning, and the thing that they're talking about is an entire refinery that might be the, you know, three times the size of downtown San Francisco.

**Dev Tagare:** It's not like there's one.

**Inhi Cho Suh:** Little asset on a table. And historically, they'd be using like an AutoCAD design. They would put potent. I kid you not. You know, certain places might be like post-its. You've got digital devices. But how do you allow everyday workers that are more blue collar workers navigate spaces and kind of achieve the mission to be done, which is they want to go and inspect something?

## 5. The spatial intelligence layer

**Eugene Chong:** I'm glad you brought up the comparison with language models, and the way we view ourselves is we're not necessarily building a world model, but we are building the grounding layer for a world model. So similarly to language models, we provide the essentially the retrieval layer, the retrieval function. So our view of how that manifests is that world models are not in competition with us. We believe that we are deeply complementary to each other. World models answer the question of what happens when I do this thing. They answer the question of cause and effect, but you still need to know your starting point. You still need an accurate understanding of exactly what the world looks like and needs to be to metric scale.

**Eugene Chong:** It needs to be placed on the globe as we know it. And so that's what we view to be our role in this. And we are focusing on enabling people to bring in these insights with the cheapest and most available sensors and pieces of data that they have, so that we can bring as much of this intelligence into as representations as possible.

**Dev Tagare:** Is it fair to say that you're building the grounding layer, you're beginning with a solid source of truth, and then you're giving the ability to end users to basically bring in their it's called their own on Golden's via either data streams or signals or impulses, what have you and types of collecting telemetry. And then you could adjust the ground truth that you have for the use case that they're trying to solve.

**Eugene Chong:** Yeah, that's 100% right. So recently we there a project with a company called flexion which creates, um, a robot brains essentially for cross embodiment, um, or across embodiments. Um, and one of the things we provided with them was a reconstruction of an office that they were trying to deploy one of their models in using just a commodity capture, like ten minutes of 360 camera footage. What they did from there was they fine tuned their policy within that environment. And to in these point earlier, the key features of that were that it was to metric scale.

**Eugene Chong:** Obviously obviously it looked photorealistic, but the textures and surfaces on the ground, this might sound basic, but flat surfaces were in fact flat, such that the signals that they derived from navigating these spaces were accurate to what it would be like in real life. So they found that when they deployed the robot in the real space, even when they modified it, it was still robust to the environment itself. And then from there, they can now start incorporating certain generative aspects to it. What happens if it's dark in here now? What happens if it looks different? But we're providing that grounding layer, such as the strong foundation for them to experiment upon.

**Dev Tagare:** So my original question was going to go in the direction of are you building a world model? But they sort of, you know, address it heads on. Um, so if I interpret this right, like, are y'all moving in a direction where it's going to be the first RL platform for world models to be trained on, and you're already giving the grounding layer. So is that roughly where we will see the proliferation of world models on top of your platform?

**Inhi Cho Suh:** Yeah. We want to build the spatial intelligence layer like earlier conversation about, you know, the leapfrogging in AI are accelerating faster and faster and people talk about AGI. However, you can't ever get to AGI unless you have spacial understanding. And that is our core. And it's a fundamental infrastructure core. It's also a reasoning core. And there's an intelligence element because you're querying it all the time. And the intelligence pieces, you know, there may be a tripod here or there's and you're trying to guide a machine to it. One of the aspects is it might only see it as one dimension, when in fact there's a lot of edges and you want the edges of the mesh aligned to what it's seeing via the camera lens.

**Inhi Cho Suh:** Another good example on the flexion partnership was most robots struggle with glass. We saw the same thing even with Cocoa Robotics in our last mile delivery for downtown LA streets. When you know those plexiglass benches right on sidewalks, they're all glass. Or there's billboard, you can see it. So it's a chronic problem. You can't always be precise in it, but you have to be able to measure when, okay, where is an edge? How far does it extend? You know, poured concrete on the ground. Is it always even. So you're going to have uneven slopes, especially in construction sites and places where there might be a slight change. And that's a balance piece that a humanoid has to like account for in its hip and rotators and sensors.

**Inhi Cho Suh:** And that actually plates places the most weight issues. So that can't be just a plausible scenario. It has to be much more precise. And those are the aspects that we're building in.

## 6. Enterprise data and vertical applications

**Nancy Wang:** The kind of reminds me like the first time that similar to inflection, you know, saw the demo from the skilled AI team, right? Where I saw robot actually climbing the stairs, being able to take in different points around it's, you know, through a 360 environment. So just like, for example, maybe my next question is folks were able to experience Lms through a chat interface, right? And of course, now we've progressed so far into autonomous agent. Right. What's going to be that first unlock for physical intelligence in your opinion?

**Inhi Cho Suh:** Oh, this is a good question. I've been thinking about this a lot because in the physical AI space in the world model space, it's not a, um, these aren't generalized services, although there is a bit of generalized training with some degree of video. And therefore you could in the language space, you can have a ChatGPT moment. But I would say that we've actually had many versions of it and not the whole planet. The whole planet has yet to experience it, but anyone in San Francisco that has ridden a Waymo has experienced it. Probably anyone that's ever called their very first Uber ride. You're using, like your phone as a remote control.

**Inhi Cho Suh:** How did like this car, even with a driver or without a driver, find me in this moment that you're kind of redirecting the physical world through, Um, a sensor, a mobile set of devices. That's how we're rethinking the world very differently. And I think it's going to be an unlock for every unique use case. That's 0.1 versus like a single generalized service. The second piece is a majority of the data set is not on the internet. So unlike oh.

**Dev Tagare:** You can't just scrape and just keep training on the internet. Um, and.

**Inhi Cho Suh:** Most of the most meaningful high quality data is actually locked in enterprises in the most important places. And they're discreet and they're messy and they're 20 years old.

**Dev Tagare:** And.

**Inhi Cho Suh:** Someone has, you.

**Dev Tagare:** Know, become an expert in.

**Inhi Cho Suh:** Where to find things and.

**Dev Tagare:** What they are and how.

**Inhi Cho Suh:** They look and how to dissect it.

**Dev Tagare:** And.

**Inhi Cho Suh:** You know, detect alerts and triggers. And so trying to compute all of that as sort of a new challenge. And we're just at the beginnings of it.

**Dev Tagare:** So with that in perspective, what sort of field or customer area is like the strongest pool today and you think would be like actual production use cases deployed en masse.

**Inhi Cho Suh:** We've seen it in three areas. It actually started in the industrial space, specifically oil and gas energy customers, because they have massive assets, physical assets, they have a massive workforce, and they have spent billions of dollars on digital twins. And they're thinking to themselves, is there much more AI for native way of operating that also protects the risk, security, collaboration, efficiency aspects that they've like highly tuned for their environment? So that's one. Um.

**Dev Tagare:** Uh.

**Inhi Cho Suh:** The area of, of governments is an evolving space for different reasons. Everything from city planning to much more, um, focus on different types of defense given geopolitical dynamics. But I would say in the city government planning area. We have a partnership with the city of Rancho Cordova, just about 30 minutes outside of Sacramento, and they are creating the very first 3D reconstruction of the whole city. Simulation. Yeah, for simulation, for.

**Nancy Wang:** Traffic.

**Inhi Cho Suh:** Training, for robots, for other types of activities.

**Nancy Wang:** On Google. Try to do that with Sidewalk Labs.

**Inhi Cho Suh:** Exactly, exactly. It's finally it's still taking a while. I don't know that we're actually done with it, but we have started with just under about under five square miles. The whole town is 35mi<sup>2</sup>. And so that is a different order of scale of reconstruction versus a super small reconstruction that's been that's highly tuned. And then the last category where we actually see the biggest future growth is in robotics and embedded embodied AI. So the partnership with Flexion Coco are two of the great examples. We've got many more.

**Nancy Wang:** Yeah. And that actually, you know, draws, like I would say, even more parallels to what we're seeing now with foundation models, which is going after vertical strategies. So you mentioned robotics maybe kind of segmenting that a little bit more. Right now we're talking about sort of B2B side robotics like warehousing right. Versus now home robotics. Like in your opinion what's going to be the first market to really explode is on the consumer side or the B2B side.

**Inhi Cho Suh:** What do you think he's been spending the most time on robotics on the team?

**Eugene Chong:** I think it's going to be on the B2B side, and I think it goes back to Dev's earlier point about sort of like producing almost like per enterprise RL environments. So if you take the example of cocoa, for instance, our initial deployment with them is to use our localization service to help them keep better track of their robots when GPS is bad. It's it's actually a very straightforward use case in a downtown area. GPS is very noisy. Sometimes the robot forgets what's block it's on. Very hard for it to make a delivery, let alone even, like find home after that.

**Eugene Chong:** But the question I think the big insight we've had there is we realized that the greater precision and confidence that you can provide them, the more use cases that were previously not contemplated can be unlocked. So the initial use case right now is stop my robot from getting lost, tell them what side of the street they're on. But what if we can provide a level of precision that's much greater than GPS, even in good conditions, but rather it's down to the centimeter? We're able to tell you your six degree of freedom pose precisely within a city. What does that unlock? What that unlocks is you're able to drive your robots directly to the narrow pickup area. So that's not blocking the wheelchair ramp. So that's not getting in the way.

**Eugene Chong:** Pedestrians on sidewalks is not getting hit by cars. It's able to intuit what an appropriate drop off location would be at an apartment building that it's never been to before. So these are sort of the use cases that we're trying to build up, and they lend themselves really well to this vertical strategy because those are very specific kind of problems and pain points that if we tried to be horizontal and serve everyone on ones, we would never get up to that level of detail. So that's why I think it's most likely to happen in the enterprise 100%.

## 7. Capturing environments for AI

**Dev Tagare:** I completely agree. Um, just so switching gears a little bit from like, we sort of covered the, the market perspective, the human interaction perspectives. I'm very much intrigued by like, what is the developer experience in working with with your APIs, with your tooling? Uh, can you like, describe life in the day of a person who is, you know, up taking your products and like trying to build something out of it? How would that look? Um, let's say I'm an Exxon or pick pick your favorite, uh, oil and gas company.

**Eugene Chong:** Yeah. Right now we're starting with capture as the entry point. Environment capture. And that takes two forms. You can either use commodity devices, your cell phone to scan your location. Bring it in. We'll map it for you. We'll get a sense of where it is on Earth and what it's composed of. Or you can provide the data that, as you mentioned, you've laboriously collected over the years for various services. So the first step for us is bring in all your data and align it and make sense of it and understand how it works together. From there, the question expands and it becomes, what are you going to do with it? Today, what we're focusing on is deploying our visual positioning system so that agents acting within that space know precisely where they are.

**Eugene Chong:** You can also work with that representation to run simulations in a platform like an Isaac SIM, or something along those lines to predict how your workflows would go. But what we're really trying to get to the platform we're trying to build towards is a loop based on that, such that the usage of the thing becomes the data capture in itself. A query for where I am updates the map. The fact that your wheels slipped in this corner and you fail the work order that gets captured within the latent understanding of the space as well. So that's that's where we're trying to move from with the, I guess, the digital twin paradigm, the perfect, hopefully up to date version of your world.

**Eugene Chong:** And we're trying to bring in the experiential and operational data in the long run.

**Nancy Wang:** So it's almost like you're building a comprehensive set of evals, right? Again, using your analogy of what to capture, what's the end result?

**Inhi Cho Suh:** Yeah. And we're at the beginning stages. We haven't I would say, you know, this category for a long time. You've had investments by many companies and in VR, AR as sort of the frontier and quite frankly, consumer has not materialized. Right. For a number of reasons. Hardware is hard, form factor is uncomfortable. The cost is just super high. Also, for the enterprises you have different categories and classes, enterprise developers, some used to and some not used to developing in this kind of way with 3D versus a traditional data workflow in a business process and enterprise APIs that they're used to.

**Inhi Cho Suh:** And so what we want to do is maybe express our SDKs as well as APIs in a way that makes sense, that are going to be a little bit more tailored toward the enterprises and their use cases. The second piece, which also makes us a little bit more unique as we realize that their services where the customer will want to operate. Leveraging our cloud structure architecture. And they'll be services where the customer says, you know what? This is highly sensitive. I actually want to run it on prem, and we're able to test and potentially containerize our models, which is almost the opposite end of where you would think for world models and real world physical models. But that's actually where we are.

**Inhi Cho Suh:** So we have, you know, a set of extremely large models that you cannot put in any kind of container too. We have a set of models that we're actually working on, containerized to be on prem for highly sensitive data and customer sets. Yeah. So the deployment, I always think about it as unique deployment patterns are dependent on the use case that we want to serve, and also the cost and efficiency that requirements that they have. Because for most customers too, They're not going to have like the luxury of what we have in Silicon Valley and the access to the highest end GPUs. We've actually been doing this. We've been doing everything almost counter. We have like crappiest data, you know, poorest GPUs. And it does it still work. And the answer is yes.

**Inhi Cho Suh:** Like can you run it on device. Yeah. Like yes. So we're I mean have we fully productized these things yet. No. But we're like working on that. But I'm confident that we're going to get there. And that's what's so exciting because that means we're actually going to be able to unlock it for more people.

## 8. Building companies around hard problems

**Dev Tagare:** When you solve for diversity, you solve for prosperity.

**Inhi Cho Suh:** Yes, exactly, exactly.

**Dev Tagare:** Solving for adversity. So, you know, you've been like like some amazing companies IBM, DocuSign, Niantic. Special. So in your experience, what's the difference between a really impressive far out there demo that's like, you know, everyone's standing and clapping and a solid product that just works.

**Inhi Cho Suh:** What's the difference? A lot. Yeah. Have you tried to scale a platform with a billion users on it, who, you know, are trying to sign a DocuSign from Antarctica? Uh, the answer is yes. No. Prototypes and demos are exciting to trigger the imagination of what's possible, but what inspires me most is actually working backwards from a fundamentally different point, which is, is this problem that we're working on a generational problem. So I, you know, DocuSign and that simplicity of the magic of knowing that you're about to complete a transaction of some sort, it's meaningful. I'm, you know, buying my first car, my first house. I'm about to adopt a baby. I'm doing.

**Inhi Cho Suh:** You know, there's, like, these meaningful life moments, and it's going to happen from your birth to your death. And it's going to happen every generation. And so when I look for like magic, I look for a unique aspect of the service that is going to stand the test of time in different ways, and the tech stack can evolve. So I don't get married to this tech stack. I don't get married to one architecture piece because DocuSign, much like many other services I love, often have been architected many, many times, and they're able to support a very interesting use case. And they they hide the hard stuff. And that's kind of what I'm interested in here with Niantic Spatial is we have a lot of unique people that are

**Inhi Cho Suh:** super deep in a very specific category, and there are these moments where we are able to, for example, rethink mapping and rethink physical geospatial AI through a transformer and embedded encoding lens that is very different than prior, and the priors were either mathematic algorithms or machine learning techniques. And we are inventing this new path. Mapping, whether I'm a human or robot, is going to be needed for every generation to come. My kids are going to need it. My grandkids are going to need it. And that's why that's like it's a problem, that's generational. And then I think about like, how do you express that in such a unique, magical way? That's simple, but also rethink the architecture in a way that's current for the times.

**Inhi Cho Suh:** And so that's why I said yes.

**Dev Tagare:** I think of a generational or like a multi generation problem in this case. Yeah. And then break it down into a set of implementations that are relevant today that keep sort of moving the needle forward but constantly. Exactly. A lot of sense. Like I remember my first DocuSign, I was I was a teenager of.

**Eugene Chong:** Sorts.

**Nancy Wang:** And DocuSign really changed the industry, right? For signing major contracts, you know.

**Inhi Cho Suh:** It's like a good example. Like, I'm in an elevator and someone says, oh, where do you work? DocuSign I love DocuSign. All of a sudden the first person shares like the story, right? Of like the experience that they were going through. And what was that experience? And it was memorable. And I hope that every customer we hit with nine Tech Special will be we'll be seeing the same in every worker that gets to touch it in a different way. We had this one, um, uh, technician that was actually in, uh, ExxonMobil that was looking at a specific asset to go fix. And he, you know, the best quote was, oh, if I had to find a needle and a stack of needles, I would find that needle with you guys. This is like a refinery is a stack of, like, all metal pipes.

**Dev Tagare:** All right. Yeah.

**Inhi Cho Suh:** And to get this person to that one blind for a turnover is like trying to find a needle in a haystack of needles. And so when he said that, I thought, okay, this is amazing quote. Yeah.

## 9. Trust, safety, and human judgment

**Nancy Wang:** Well, and I would be remiss if as a security person here, maybe to also kind of get your thoughts on trust and safety, right in that sense, because, you know, for a lot of things of what we do, right, it's really confined in the digital world. But let's say, you know, when you're now adjudicating actions for physical systems, right? Or directing robots. Well, there's also physical ramifications. So how do you think about, like for example, actions that are governed by the model versus, you know, actions that have to have a human in the loop?

**Inhi Cho Suh:** You know, first of all, you can't, um, you can't delegate all of that responsibility of what's been developed through decades of risk compliance measurement in a single model. Um, and many business processes are codification of like protecting the human risk. And you have really advanced technologies like time series. Database has been around for a very long time. And energy because of the sensitivity, the the precision. If you think about everything from simple frequency to seismic activity, it's in the industrial category has been pretty advanced. And I share that as an example of we're just at the beginning stages, much like the very first set of medical C work that IBM even worked on.

**Inhi Cho Suh:** I think about it in the same way for the work we have to do at nine tick spatial. So it allowed, for example, radiologists rather than to look at like 10,000 X-rays. Maybe they only need to focus on here's the 100 that really needs a second or third look. It's about concentration of the human capacity to think. And we had a similar situation actually in for nine take Spatial in an oil gas scenario where you've got lots of footage and video, and typically it would take, you know, a

20 year veteran having looked at that to diagnose a situation. And is there a way for us to look at that same, you know, pool of footage and video and compress the 13 hours down to two hours and say, here's where your two hours should be spent. Not all 13.

**Inhi Cho Suh:** And unlike machines, we all get tired. We're human. You know, it would be better to suit, um, that kind of thoughtfulness in the work. The security pieces are hard, I think, in this world. We're also thinking about it even in the build of our code base and access. And so it's not something that is a one time solved and done kind of thing. It's a continuous act. We have to be thoughtful about it. I think we have to also be upfront about our position of, of different aspects and, and the guiding principles that we're using for different types of decisions because we don't know what we're going to actually encounter. So I think the guiding principles of our values, culture, leadership decisions are going to be important.

**Nancy Wang:** Why don't you take us.

## 10. What changes next

**Dev Tagare:** On a couple of, like, fun questions? Yeah. Real quick. So the next two years, what do you predict will be the biggest change to your stack? Well, what do you think will disappear? What do you think? Evolve.

**Inhi Cho Suh:** Well, we're still building our stock. Okay. So we're super excited about what's next. So we've been we've been pioneering 3D reconstruction. Our team has actually created an open source, the Gaussian SPC format. And think about it, because the scale of like the data so massive, we needed to compress it and also make it highly performant and high fidelity. I think the work we're doing right now on the alignment with the polygon mesh, uh, mesh and polygons is, you know, in the context of the robots and the training policies to work in those spaces. The next piece, which is untapped for everybody, is solving for

**Inhi Cho Suh:** much more semantic understanding and not object detection segmentation based on traditional forms of labeling, but truly embedded in new ways. And so that is kind of where we're headed next.

**Dev Tagare:** Um, so it's like interpretation of something in the physical space, but I could see.

**Inhi Cho Suh:** It through a visual lens. Yeah. Non non non text language based. Although you can then apply language text to a concept like um and understand intent. Ultimately you know the ability to solve that spatial reasoning piece of what I would consider subjective relevance is probably the the ultimate goal, subjective meaning. You know, often in robotics, people talk about egocentric meaning the the self being the the robot itself. But in a subjective relevance standpoint, there could be multiple parties.

**Dev Tagare:** And that's true. And the perspective is different. It's very different.

**Inhi Cho Suh:** And the question could be is the robot safe from the human? Is the humans safe from the robot? Is the humans safe from other humans?

**Dev Tagare:** You know, like.

**Inhi Cho Suh:** This is a subjective.

**Dev Tagare:** Realm. My dog versus someone else does.

**Inhi Cho Suh:** I mean, I'm just using safety as a lens on that, but it could also be intent, right? Subjective relevance is all about what is what is the objective you want to accomplish in this space and in this moment? Is it a highly efficient space to do it in, or is in this moment a non efficient space to do that in? And that can change because of time of day conditions, messiness. And those are once again subjective relevance questions depending on the various subjects in the scene. And regardless of who you are, we want to be able to work with the model builders, the robot builders, the next generation kind of inventors, as well as the enterprises that know their domains really well.

## 11. The future of physical AI

**Dev Tagare:** Last question. So normally we ask like if you had some free time, what would you build? But clearly your hands are full for the foreseeable future. This is a very complicated space or a complex space. Um, so I'm going to go instead with, uh, let's do a contrarian thesis for the next couple of years across the panel.

**Nancy Wang:** Oh, that's a great question. Let's see. So I actually think that consumer robotics is going to, I think, experience like major tailwinds in the next maybe 6 to 12 months. I agree with you. I would say maybe a year or two ago because there's just cleaner guardrails in, especially in warehouses. And, um, you know, also used to work with the co-founder of cobalt. Right. So thinking about robots and warehouses being able to follow patterns. But I just see this like groundswell of consumer demand, especially as you can see with like personal agents. Right. Personal chief of staffs really wanting to make AI relevant, right, productive in their personal lives. So I think the really push for home robotics is going to

be next.

**Inhi Cho Suh:** Well, we have one that we've been working at work for, which I genuinely believe is we've been helping on the SIM to real. But we actually think you should start with real and then go to SIM to real to go faster and have a lot more precision in the deployment for scale. I mean, that's probably the biggest piece. As everyone says, let's solve the sim. I think the sim direction is probably going to drive much like the LM direction, which is, um, open source. And there's going to be continuous, like innovation. And that gap on real world deployment becomes harder and harder.

**Eugene Chong:** I think, you know, I think there's been a certain degree of debate over whether prior mapping having priors is the approach one wants to be taking. You see that manifest in the whole Tesla Waymo debate. But you also see that in the robotics world like Velas and local Slam, only all the way. I don't necessarily think that's an incorrect approach, but I do think that there is a hidden value to mapping that will be unlocked as we continue to, I guess, move down the intelligence curve. And what I mean to say is that the prosaic understanding of a map is it's where things are located. It's where things look like, you look at it on a 2D piece of paper.

**Eugene Chong:** And I think the direction that maps are going are to go back to sort of a theme I've been touching on earlier, that they will encode outcomes, feelings, the zeitgeist. If you want to be pretentious of an area such that you are able to provide your operators in an area with an understanding that's efficiently delivered of what they should expect to happen there. And I think in this current world, like if you think about maps as 2D or 3D spaces with visuals and such and labels, you're very limited. But if we think about a world where we've extracted and distilled spatial intelligence and make that relevant available to you in the most relevant way as possible, I think we would be silly not to rely on prior mapping.

**Eugene Chong:** Why would you throw away the knowledge that you have from before?

**Nancy Wang:** Yeah. As you were kind of describing, I'm almost envisioning this world. I mean, we do it today as humans of, hey, I'm going to be in XYZ city, right? Not only do we look it up on the map, but we're also looking at things like temperature, other attributes of local attractions. What's going to be like? You know, the vibes Z, the zeitgeist. So if you could actually just bring that all in different layers, almost right in a spatial understanding of what it means to be in that location at X time, like that would be huge. I just don't know what that would look like. Right. And how will you deliver that information?

## 12. Mapping outcomes and robotics

**Dev Tagare:** My thesis was going to be somewhere along the mapping direction where I was like, we're going to start mapping the outcomes and actually like sort of time scaling them over whatever trajectory you wouldn't like. Instead, I'm going to do a developer focused one, which is I think robotics in general is going to have a machine learning developer style moment where everyone just starts using it because the barrier to entry is becoming so low with, you know, apps that you all gave SDK that you all gave and others to. So yeah, my my thesis.

**Dev Tagare:** Was going to be a robot developer.

**Dev Tagare:** Is everyone's going to be just like everyone suddenly became an ML engineer. Everyone's going to be a robotics engineer.

**Nancy Wang:** So instead of a, you know, software agents going to be physical agents.

**Dev Tagare:** I.

**Dev Tagare:** Would hope so soon enough.

**Inhi Cho Suh:** I think it's a whole new category.

## 13. Looking ahead

**Nancy Wang:** Well, let's look back on this moment in the next 12 months and see how many of our predictions came true so far.

**Dev Tagare:** Yeah. If you have like.

**Dev Tagare:** Exactly.

**Dev Tagare:** For the past year or so, I.

**Dev Tagare:** Would say.

**Nancy Wang:** Pretty significant.

**Eugene Chong:** It's a pretty good hit rate for, for just the past year. Yeah.

**Inhi Cho Suh:** Yeah. Given all the volatility in the market that's pretty good. You guys are ahead of the curve.

**Nancy Wang:** Well thank you both so much for coming in the studio today. And I'm excited to see all that you will build.

**Inhi Cho Suh:** Of course. Thank you for having us.