Transcript of NXP CEO Rafael Sotomayor's COMPUTEX 2026 Keynote

Introduction: What defines elite performance

Eight days from today, the best footballers on the planet are going to get together, and they're going to play in this tournament that mesmerizes nations. You know when their team plays, countries stop. Billions of people are watching years of preparation. The players on the field are elite.
So, what makes them elite compared to the rest of the players?
It's not fitness, because they're all fit.
And it's not knowledge of the game, because all of them know the game.
It's the mastering of the task.

It is executing at the highest level with the highest level of pressure. That is what elite looks like.

What does “elite” mean for intelligent machines?

So, here in Computex, as I walk the halls, there are plenty of examples of physical AI—they're everywhere. So, I pondered the following question: if we're going to bring all these intelligent systems and devices into our lives, what does elite look like for them? What is the equivalent of a world-class athlete for a machine?
I'm going to need you to follow me here, because that's a question we're going to answer today.

Where excellence becomes instinct

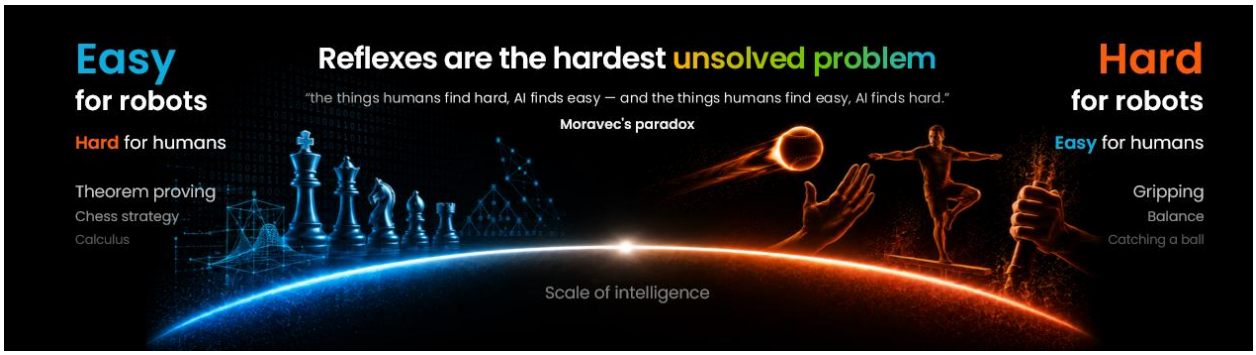
If you are a fan of football, you have probably seen Messi play. Now, if you're not a fan, watch Messi play and you will become one. But he's very smooth, the way he handles the ball so effortlessly.
When he is about to kick the ball, it seems that he waits until the last minute to decide whether he's going to go left or right, depending on what the opposing goalie is doing. And that gap—that tiny, infinitesimal gap between trigger and response—is probably one of the most sophisticated feats of intelligence on the planet. And yet he doesn't think about it. His reflexes take over.
We all know excellence is not about thinking harder. It's about mastering the task so deeply that thinking becomes unnecessary.

Reflexes dominate our daily life

Let me give you an example here with you. What's happening with you? I see a lot of you here. What's happening with your body? Some of you are already adjusting your posture. You're fidgeting with your hands, your feet, and you're doing this without thinking.
It turns out that most of our daily tasks—over 95% of the things that we do on a daily basis—we do without any conscious thought, without any effort, with very little energy expenditure. Think about that. That is brilliant. What we spend most of the day doing costs us very little.
So, if reflexes dominate our daily life, then I would contend that reflexes are foundational for robotics, are foundational for Physical AI, because that's the key. Reflexes are hard, and that brings us to Moravec's paradox.

Moravec’s Paradox

Moravec was the first one to highlight something that seemed very counterintuitive. What is really, really hard for human beings—reasoning, solving complex problems, playing chess—those are things that are really easy for machines.
But what is easy for humans, like walking, like folding clothes, that's not so easy for machines.
So, reflexes—not language, not reasoning—but reflexes are the hardest thing in robotics. I want you to hold this notion of reflexes being difficult for a moment, because we're going to get back to it, as that is key to unlocking the full potential of physical AI.



Innovating at the edge

Now, right now is a fascinating time. You can feel it here in Computex. All of you are here. We're living through the fastest technological evolution ever. I know you feel it.
Daily—almost hourly—something amazing gets disclosed. Something amazing happens and gets released. Whether it's new stuff in agentic AI, or new developments in software-defined vehicles, or machines that speak any language and solve complex problems—it’s amazing.
A lot of that is coming from innovation happening in AI in the cloud. But now AI is available and moving to the real world, and it's touching and transforming everything.

And the real world is what we call at NXP, “the edge”. The edge—that’s where we shine. Edge products are what we do. These are fit-for-purpose products. They have real-time performance needs, low latency, low power, high levels of security, high levels of safety.
Products that go into software-defined vehicles, smart factories, smart infrastructure—those are our products.
As we move intelligence from the cloud to the edge, what is so exciting is that now all these devices get intelligence.
You get to create new use cases. You now have the ability to create devices that make decisions, they're independent, maybe even autonomous.
But, in order to actually reach the full potential, we still need to solve the riddle: the Moravec's paradox.

Where Does Intelligence Reside?

So, now I'm going to talk to you about how NXP is well positioned to provide a solution to it.
To get clues on how to solve Moravec’s paradox, we need to start in a place that has already solved the problem of reflexes, which is us.
If I were to ask you: where does intelligence reside in the human body? I would guess over 95% of you would say the brain. You would be partially correct, which also means, you would be partially incorrect.

The Cerebrum

Let’s start with the greatest physical AI ever built: the human brain. The cerebrum is the largest part, representing 80% of the mass of the brain.
This is where reasoning, conscious thinking, learning, and decision-making happen.
But this is not where the fastest responses occur.

Let me tell you what happened to me this morning, and this is actually real. The NXP office is really a few blocks away from here. And listen, I'm a little jet-lagged and I was a little distracted. And I crossed the street and I was very close to getting run over by a scooter. And it woke me up.

And I'm looking at these 300 milliseconds of response time by the cerebrum, that's an eternity. What happened to me this morning? I jumped like a cat. The cerebrum didn't help me at all this morning.

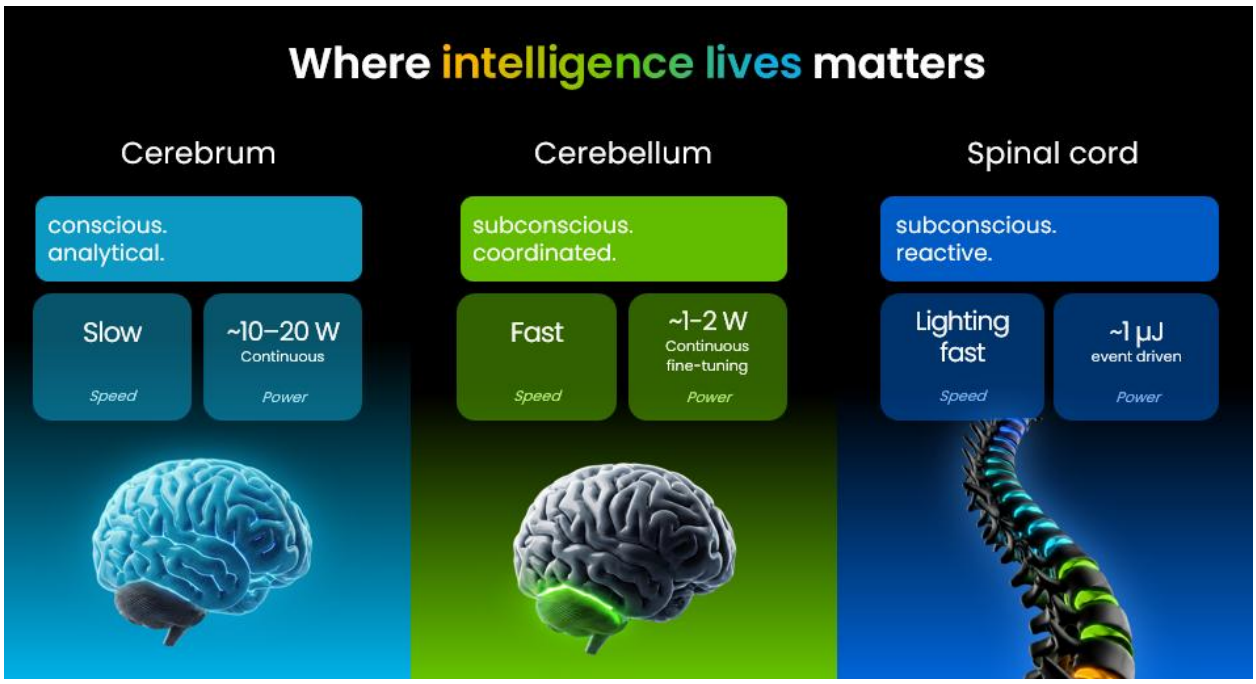
The Cerebellum

Now, the good thing is that the cerebrum doesn't work alone. It has a companion. And so underneath the cerebrum sits the cerebellum. And think about the cerebellum as the co-processor for motion, for motor control, for position and balance. It's still faster than the cerebrum, but not fast enough. Not for me to avoid the scooter this morning, I tell you that.

The Spinal Cord

So, we have to go a level deeper. So, we get to the spinal cord. And life-saving reflexes don't originate in conscious thought. They don't even originate in the skull. They live in the spinal cord.

I mean, you can see the stats here, right? 40 milliseconds. It's lightning fast. And this is nature's decision in design by choice, because in this case, latency is more important than intelligence. This is what saved me this morning. So, the spinal cord processes, decides, and acts independently.



Hot Surface Reflex

Let me give you an example to make it more concrete other than me trying to avoid the scooter this morning. And I know you will relate to this. You touch something hot. The stimulus reaches your hand and sends a signal to the spinal cord. The spinal cord gets the signal and sends a motor command that says remove the hand. And now the hand is away and safe.

Meanwhile, the brain is still processing and then eventually catches up, and it realizes, oh that was hot. Ouch, that hurts. And now you can see the spinal cord allows us to act before we think. The spinal cord keeps us safe.

The spinal cord is a stack processing center

Now before I move on, I want to highlight another design principle. I want to highlight that nature, in its infinite wisdom, decided to put reflexes where the action is. So essentially the spinal cord is a stack processing center that runs along your back, placing sensing and processing at the nearest possible point.

Again, brilliant. Closer means faster, faster means safer and also means that it is the lowest possible energy use. I mean what evolution is showing us here is billions of years of optimization.

Key Takeaway: You don't scale intelligence by making the brain bigger

So, where are we? We started with this conversation of biology looking for clues on how to solve Moravec's riddle. So, what are the takeaways? The takeaway is that you don't scale intelligence by making the brain bigger and bigger. You scale intelligence by putting intelligence at the right place.

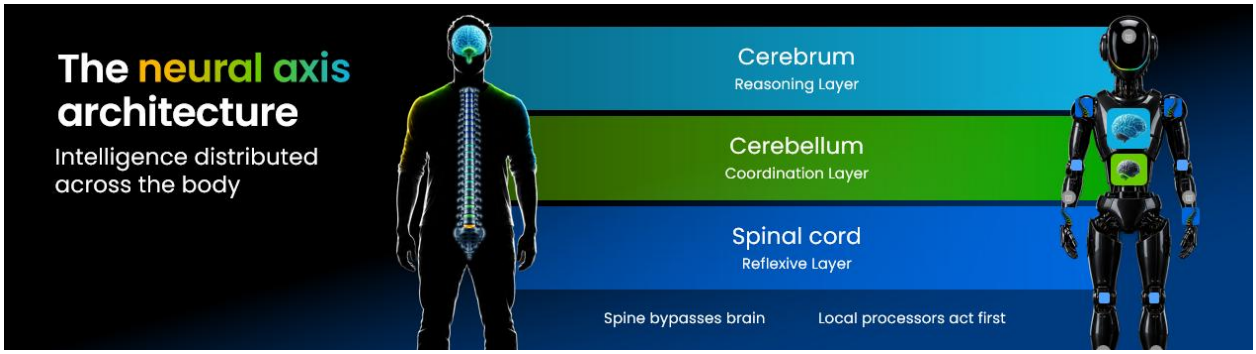
And there are three fundamental requirements of intelligence in real life:

- You need to have ultra-low latency responses to actually be able to live in the real world.
- You have distributed control so there's no single point of failure.
- And you have energy efficiency, extreme energy efficiency. So, you actually manage a strict budget to be able to extend life.

This is the playbook. This is the blueprint.

The solution: The Neural Axis Architecture

So, we contend that this is what is needed in physical AI. So, we now take this principle and apply it to robotics. And this is what we call at NXP the neural axis architecture. Intelligence has three layers: the reasoning layer, the coordination layer, and the reflexive layer. Three layers, independent but highly coordinated.



The Neural Axis Architecture in Drones

So, now let's see it in practice. We have three different form factors here, all of them in which NXP participates, three different implementations. So, let's look at the first one, drones.

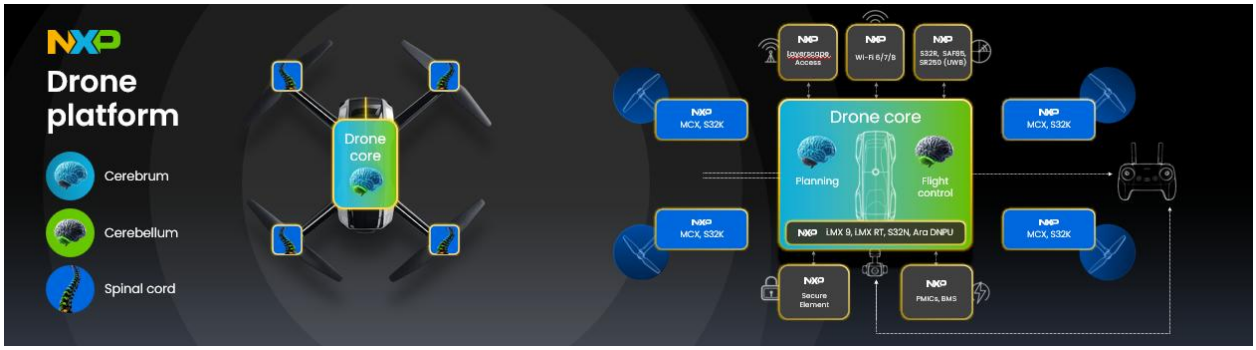
You can see also the neural axis maps very nicely. Here, you have the reasoning layer doing all the flight planning and path optimization, and then you have the coordination layer managing flight balance and performance. Both of them need to be operating independently, concurrently, and this is not a design choice, this is not a nice-to-have. This decision of having this independence is what makes this device trustworthy and safe.

Now the reflexive layer, which is now these nodes at the outer edge, they're connected to motor control and actuation, and these are fast, reflexive. They don't wait microseconds.

So, at NXP we deliver this full system, all the KPIs, all the things that are needed to build a world-class drone. And I give you an example. We track a metric that we call glass-to-glass latency, which think about it:

- camera captures, processes, transmits,
- the controller responds,
- then the drone reacts.

All that loop, 20 milliseconds. Miss it, drift, instability, potentially even a crash. Hit it, you're in control, the drone feels alive, safe. This is NXP's implementation of the neural axis in a drone.



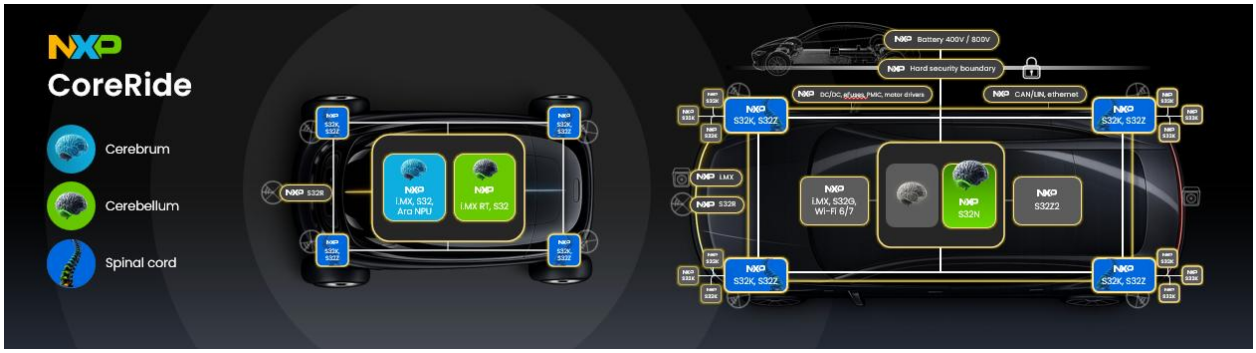
The Neural Axis Architecture in SDV's

Now, maybe this is one we're a little bit more familiar with, since we have a lot of customers in automotive, but the software-defined vehicle (SDV) is also a great example of the neural axis.

You have the reasoning, obviously ADAS and navigation—what is the vehicle going to do next. Then, you have coordination, which is a separate compute that manages vehicle dynamics, ensuring the car always has stable motion. This is an area where the central compute, our S32Nfive-nanometer processor—one that is unlike any other on the market—sits.

Now, on the reflexive layer, we have sensors managing mission-critical functions of the vehicle like braking and suspension. This is another area reflected in zonal architectures with our S32K family of products.

Now, this separation, both logically and physically, of the three layers is what makes this car reliable. Because in a vehicle, with lives at stake, you don't have margin for error.



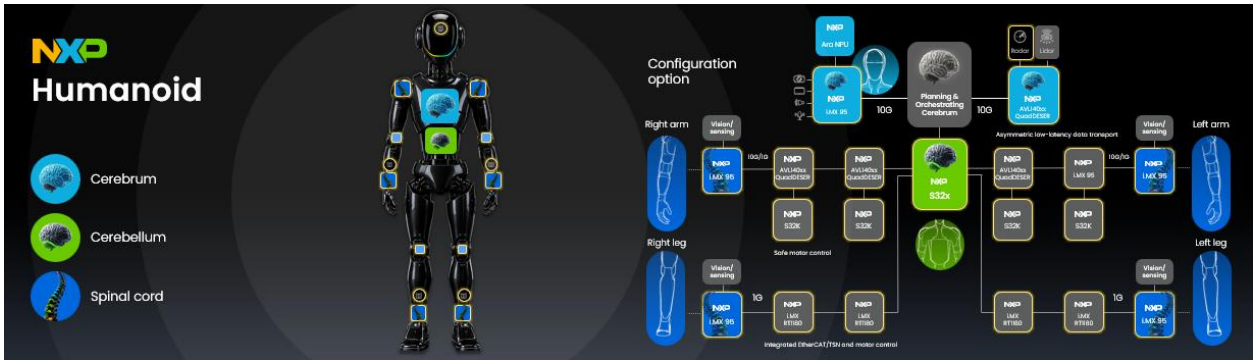
The Neural Axis Architecture in Humanoids

Now, humanoids, which is the most complex implementation of all. And I contend, because this is an emergent market and you hear a lot of different conversations about the architecture of humanoids, but our belief is still very firm. If this neural axis architecture is good enough for life, it's good enough for humanoids.

Now, let me maybe bring this alive with an example. Imagine a robot walking in a warehouse and carrying a very valuable and very fragile package. Halfway down the hall, somebody drops a pallet and it kind of brushes up the robot, and it kind of loses a bit of control.

Now think about what happens. The robot needs to recover balance, make sure that it still has the package, adjust the grip, understand where it is in location, recover, and continue walking. And all that needs to happen within 40 milliseconds.

No cloud call. No waiting for the model to respond. Okay. Intelligence at the body, at the joints, at the hands, at the feet. They must own the moment.



The Neural Axis Architecture is existential

You see, the neural axis architecture is not a nice-to-have, it is existential. And so, today, you usually see humanoids with a separation of the reasoning part and the coordination part, because robots continue to learn and continue to learn new use cases, but you don't want to disturb the motion of the robot. Motion needs to be stable.

And then, instead of a spine, you have all this distributed control at the ends, at the nodes. And these processors own their function locally. They own it. Hands know how to grab. Ankles know how to balance. Not waiting. Not asking for permission. Acting. Making sure the humanoid survives.

Stop thinking “brain”. Start thinking “neural axis”.

So, we just walked through three examples, three different form factors, three different levels of complexity, one system, one architecture, one blueprint: three layers of intelligence.

So, what’s the takeaway? Intelligence is not about a bigger brain. Stop thinking about a bigger brain. Think about a neural axis. Three layers of intelligence.

Beyond Reflexes: Toward Understanding

Now, we talked a lot about motion, but now I’m going to introduce kind of a new wrinkle. Motion and movement is not the same as understanding. The fact that a robot is able to execute a perfect reflex doesn’t mean it understands why.

And I would say that not understanding why, in a world as complex as the one we live in, is not okay. So, the question is: how do we teach a robot not only how to move, but how to understand?

How humans learn

We’re not born, even though some of us may believe so, we’re not born walking. We were not born knowing how to kick a ball or grab a glass of milk without spilling it. We learn slowly through mistakes.

But this process of trial and error didn’t only build a skill for us. It created a model of the world in our head. We learn the concept of gravity because we fall and it hurts, so we don’t want to do that again. We learn the concept of heat because we touch something hot and we get burned.

But obviously, the way we learn doesn't work for robots. You can't just let them go in the wild and say, "Okay, go ahead and learn and come back." That would not be safe, probably reckless. So, we have to take a different approach.

Perception vs Understanding

Now, one thing robots do very well is that they see remarkably well. Let me use this example:. The robot looks at the cup, probably identifies the bottle, understands that it has liquid inside, probably reads the label.

But does it understand what it's holding? Does it understand that the bottle is full? Does it understand that if you tilt it too much, you are going to spill the liquid?

If you don't understand inertia, it's really hard to estimate force.

If you don't understand friction, you don't know how that syrup is going to pour.

And without that understanding, the robot won't act in a safe manner.

So, you need a bridge between perception and understanding. Perception tells you what's in front of you. The understanding of the world tells you what is going to happen when you interact with that object in front of you. So, we need a bridge.

So, the question here is: **how do we teach robots physics?**

Teaching robots physics

And so today, we do it ourselves. And I really mean it, individually, we do it ourselves. People teach robots how to pick objects, how to move. They mimic. The robot adopts, learns. It works, but it has limitations. It's slow. It's very expensive.

World Models provide a short cut

So, this is where world models come in. The ability now to teach a robot to really estimate outcomes without going through the process of experiencing it. It's like you literally inject knowledge into the robot. Think about the matrix, right? Inject knowledge into the robot.

So, you inject knowledge, you inject experience without having to go through the physical act of acquiring that experience. And this is a fascinating field. World models are feeding VLA models.

VLA Models

And VLAs, stands for vision-language-action models. It is the bridge between what the robot sees, what it is told, and how it moves. This is the bridge between these three different layers. It creates the bridge that it needs, to see and perceive, and to understand.

And this is one of the most fascinating research areas right now in both academia and industry: VLAs. VLA models are evolving, they are getting better and better and better. And so, the question right now is: how do we take these VLA models that were created in the cloud, with essentially unlimited resources, and move them into the edge?

The eIQ ® Toolkit for model deployment

Because the edge is a constrained environment. It is a constrained environment in terms of power, latency, memory, and processing. A VLA model in a research notebook is not very helpful for a robot in a warehouse. So, you need to bring those models into real processors at the edge.

And so, our answer to that challenge, that problem, is right here. Our eIQ® toolkit. This is a software that allows you to grab those models, import them, quantize them, prune them, compile them. Essentially you take the model and you tune it for the target hardware and for the target use case.

Our goal is to remove objections from the customer. All these things are complex. So, as we look at the challenges, we want to make sure that we remove objections, we remove friction for anything that prevents physical AI from being deployed.

Usefulness vs Trust

Now, where are we? We talked about the neural axis as the architecture, and we talked about VLAs as the understanding, and all these two things come together and make robots useful. But is that enough? Is useful enough?

Because useful and trustworthy are not the same, and for physical AI to be widely adopted in the market, trust needs to be solved.

For Physical AI trust is not earned

Now, trust, for a human being, you earn it with time. You invest in your reputation and your rapport and your track record over months and years. You cannot have a machine trying to develop rapport with an operator for a year. We don't have the time.

So, you cannot design trust. You cannot earn trust. The machine cannot earn trust. It has to be designed in from the beginning. The machine needs to be trustworthy from moment zero.

Reality hits hard at the edge

So, trust is important. Nobody likes to think that things are going to go wrong, but they do go wrong. I mean, when I look at these pictures, and you probably cringe when you look at them, I can't help myself from thinking about two laws: Murphy's law and Finagle's law.

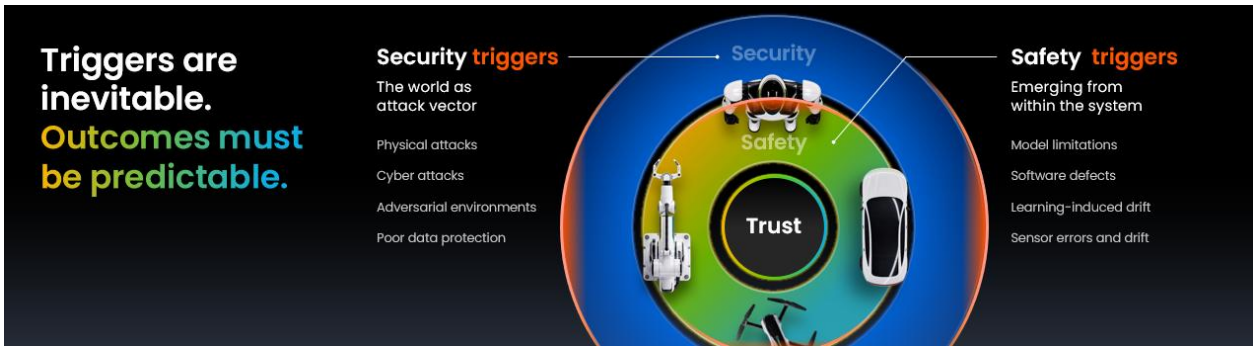
Murphy's law: anything that can go wrong will go wrong. And Finagle's law: anything that can go wrong will go wrong at the worst possible time. And that is the moment where trust gets defined. Trust doesn't get defined when everything is going great. Trust gets defined at the worst possible moment.

Security and safety triggers

So, how do we look at trust? Trust can actually be compromised through external issues, external triggers or through internal triggers.

External triggers could be changes in the environment. It could be malicious people trying to attack the system. Internally, you have design flaws, model limitations, sometimes wear and tear creates different behavior in electronics.

These are important issues, because as physical AI gets deployed, more and more mission-critical functions are going to rely on them. So, dealing with this is massively important.



NXP Trust Primitives

Now, we, at NXP don't come to physical AI from the cloud. We come from the edge. These are the things that we are already working on. This is part of what we do today, and now it is even more valuable for us. Some of the assets that we have at NXP to take care of trust are even more important with physical AI.

So, let me tell you how we, at a high level, think about trust. We have a framework today, that right now is getting enhanced, and it is a very simple one at a high level.

- **Step 1: Contain**

We contain. This means that we isolate the problem, we create redundancy, and we don't allow a single point of failure.

- **Step 2: Protect**

We protect. We have security injected directly into the hardware. We protect execution of the code, we protect against tampering, we protect credentials. And we are also post-quantum crypto ready, because the threats today are not the same as the threats five years from now. So, this future-proofs our devices.

- **Step 3: Verify**

The next primitive is to verify. Some of the certifications that we have, on SIL and ASIL, are not just paperwork. This is how trust gets measured.

Our SafeAssure program allows a hospital or a car OEM to get the one thing that is really important for them: assurances of how the device will behave when things don't go right.

- **Step 4: Adapt**

And obviously, adapt. Nothing stays still in this market. Threats continue to evolve, environments continue to evolve. So, we need to evolve. We need to have the mechanisms to constantly update these devices. These devices at the edge stay there for a decade, so you need to make sure that these devices are constantly inaugural, secure and reliable for years to come.

The real world has no undo button

We take trust very, very seriously at NXP.

You know, mistakes and errors in the physical world at the edge, they're not digital. You cannot do a software patch for a broken bone. You cannot do a system update to deal with a collision.

And I know we cannot design systems where nothing ever goes wrong. I know that. But we can design systems where if something goes wrong, it still goes right.

Trust. This is the core of our DNA. And the reason why is because the real world has no undo button.

Where Physical AI Meets Real-World Adoption

Now, we get to a part where I start landing on what's happening with physical AI today. Because our job right now is not really to push physical AI before our customers are ready for it. Our job is to make sure that physical AI has a place to be.

There's a reason why physical AI gets adopted. **It's about earning the right to be useful.**

I get a lot of questions right now from customers who are trying to adopt AI. And one of the questions I get often is: “Hey Rafael, how smart can physical AI become?”

And I ponder on that question, because I don’t think it’s the right question. The right question is: where can physical AI be deployed today so that it can be done in a safe and secure manner, where it creates value?

And I think the answer to that question is unfolding right now.

Physical AI in Factory Automation

Let me give you some examples. In factory automation, even today, in factories where physical AI-enabled robots are being deployed, they are already delivering around 40% productivity enhancements over regular automation. And regular automation is already very efficient, so this is on top of that 40%. Source: AI in Robotics Statistics 2026: Adoption, Efficiency & Future Outlook

And obviously we don’t work alone. We work with our customers. We are very proud of working with partners like Boston Dynamics, where we combine our hardware and software with their great robotics platforms to make sure that robots and humans can cooperate in a productive manner on the factory floor.

Physical AI in Healthcare

Now, healthcare. This is where safety and precision are extremely important. They define outcomes. Already in 2025. Diagnostics and lab robots are already increasing sales by 610%.

Source: World Robotics 2025 report – SERVICE ROBOTS – released by IFR - International Federation of Robotics

This is not a trend. This is a signal that healthcare is saying: physical AI, if deployed responsibly, will save lives.

We are very proud of our collaboration with GE HealthCare, where intelligent systems are being used in anesthesia. In that context, precision and safety really make a difference, and NXP is at the core of it.

Powered by Partnership

Now, this is a moment that I always take when I come to Taiwan, because it is truly amazing, the ecosystem in Taiwan. Because we know we cannot do this alone. NXP, without our customers and partners, we are too small.

The reason NXP punches above its weight is because of you. We cannot do this innovation alone, but through collaboration, co-creation, and delivering together to the market.

The people on this list are not just customers and partners. They are the lifeblood that makes physical AI deployment a reality. So, for all the companies here, thank you, because we do this together.

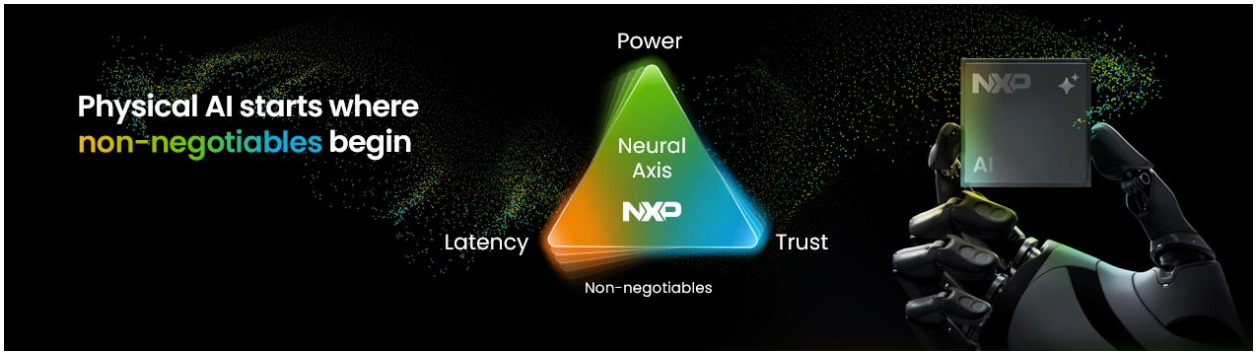
Conclusion: The hidden architecture of elite performance

So, I started with Messi, and I'm going to end there too. What makes this player so amazing and so elite, is not just the things that you see. It's all the things that you don't see.

Beyond the incredible commitment to his craft, it is the invisible architecture in his body that we call the neural axis. And that same concept applies to physical AI and robotics, engineered by humans.

We started with a question: what is elite for a machine? I think we answered that question. An elite machine is a machine that executes reliably, at high performance, in the real world, under real conditions, and when things go wrong, it keeps people and assets safe.

Machines and systems have three non-negotiables: **low latency, low power, and high levels of trust.**



Remember, you don't scale physical AI with intelligence alone. You scale it by **placing intelligence in the right place, using the neural axis as the backbone.**

So, for the push of physical AI into the market, we at NXP, we're ready.

Thank you.