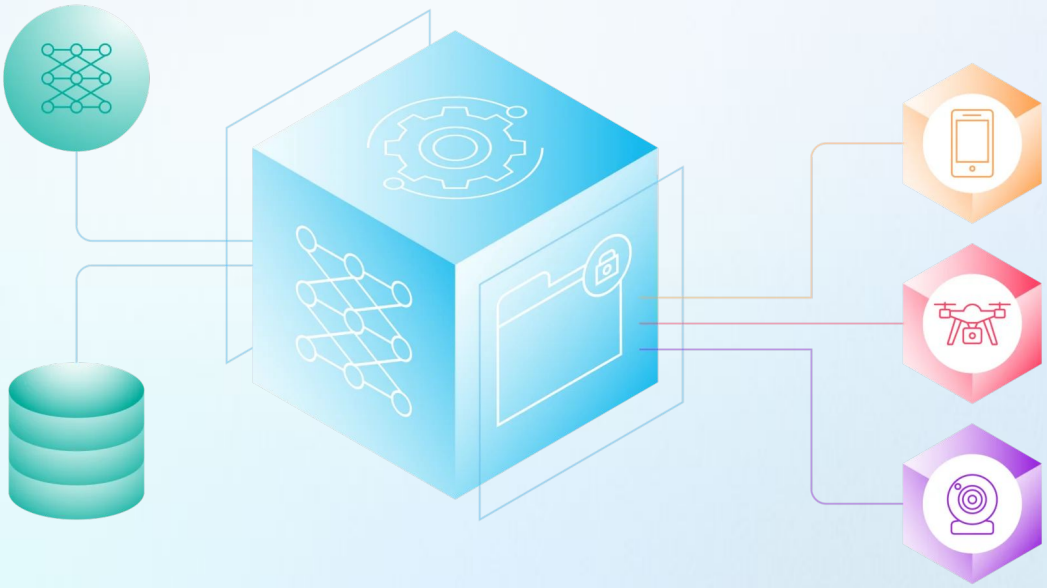
LatentAI

The Edge AI Factory

Build it. Run it. Or let us do it for you.



The edge is unforgiving

No cloud connection. No IT support. No second chance.

The edge is a factory floor at 3 am, a retail location in a strip mall, an unmanned vehicle operating in contested waters. It's every place where a decision has to happen right now, on hardware that was never designed with AI in mind, in conditions nobody fully anticipated.

Organizations have spent years getting AI to work in controlled environments. Getting it to work everywhere else is the problem that actually matters.

That's the problem we set out to solve.




NO CLOUD

CONSTRAINED COMPUTE

LOW LATENCY

POWER LIMITATIONS

Practical AI starts with the software layer

More compute won't solve it. Better quality data helps, but it's often not enough. The real bottleneck is the gap between a model that works in the lab and one that works in the world.

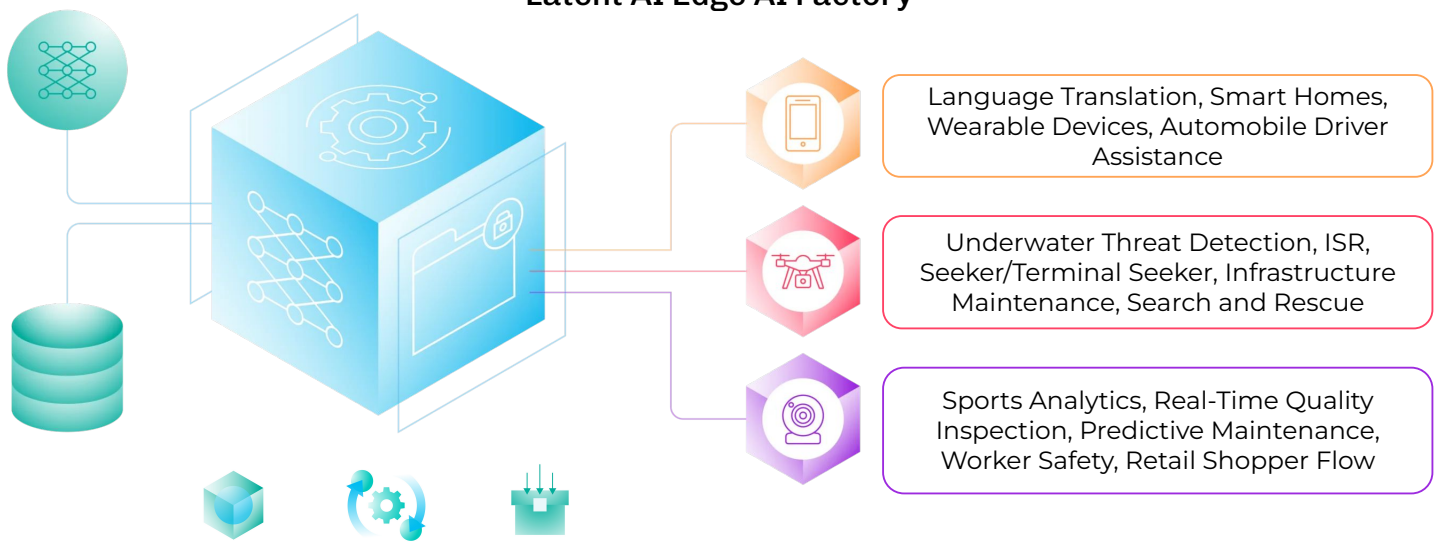
Think about what that means in practice. A threat detection model on an unmanned underwater vehicle that can't update fast enough to track new adversary tactics. A quality inspection system on a factory floor that loses accuracy when lighting conditions change. A retail shopper flow application that has to work across hundreds of stores running different hardware. In every case, the problem isn't the idea. It's the gap between the idea and the field.

Closing that gap takes software that can take any model, compress and harden it, optimize it for the specific hardware it will run on, and secure it at the model layer, not as an afterthought, but as part of how the model is built. That's what our edge AI platform does.

The result is AI that is smaller, faster, and more secure than what you started with and designed to run on the hardware you already have, not the hardware you wish you had.

> *"AI that's practical isn't just optimized. It's operable. By real people. In real conditions."*

Latent AI Edge AI Factory



But optimization alone isn't enough. We learned early that field teams can't call the lab every time conditions change or a model needs refreshing. So we built The Field Tactical Suite. A self-contained, deployable application that put AI capability directly in the hands of the people who need to use it, update it, and trust it.

The Edge AI Factory

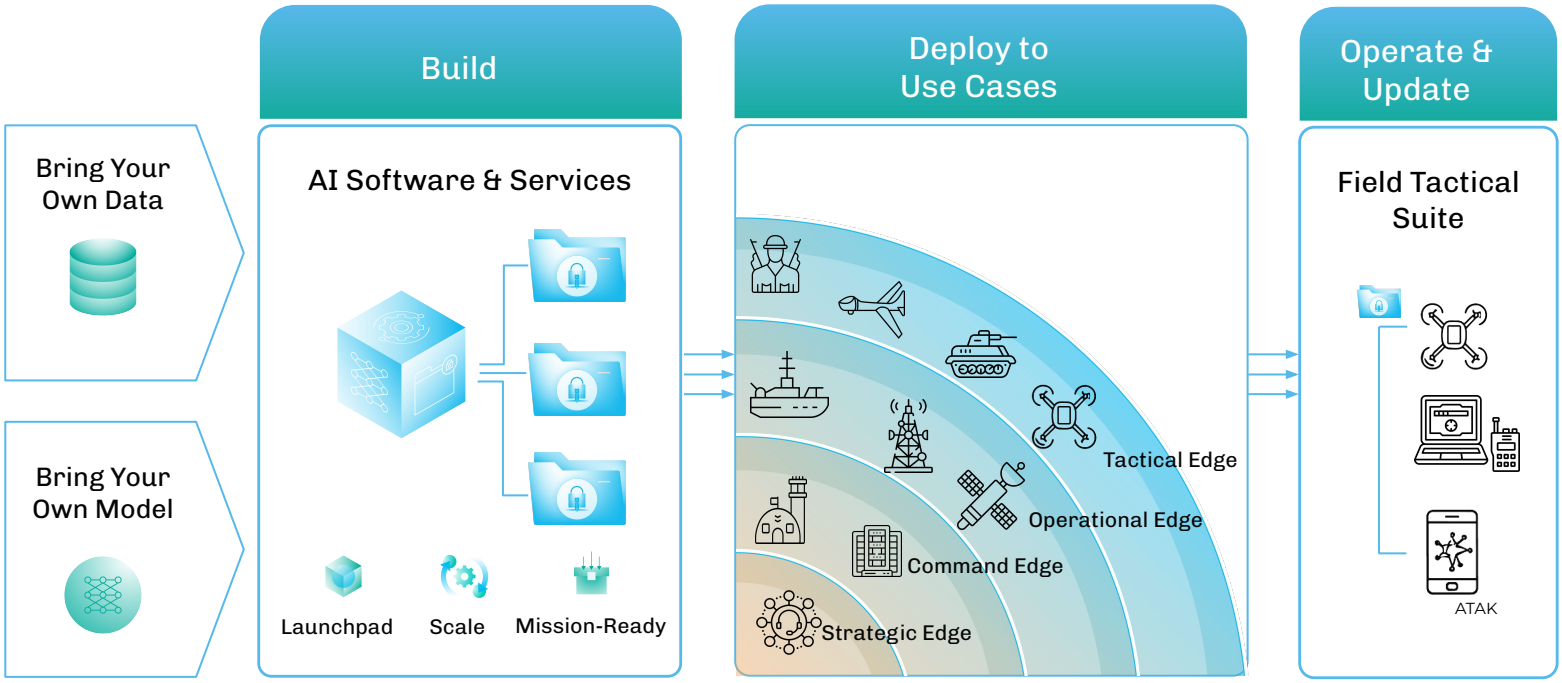
Our Edge AI Factory is a complete system. Here's how it works.

Your Data. Your Model. Start with what you already have. We meet you there.

- **The Platform** Latent AI Efficient Inference Platform (LEIP) optimizes, secures, and compiles your models for the hardware they'll actually run on. Any hardware. Compile once, deploy anywhere.
- **Services:** We help get your AI built, deployed, and into the hands of the people who need it. *Simple Development Launchpad* to get started, *Continuous Scaling Engine* to scale solutions already in flight, and *Mission-Ready Deployment* to refine your AI into lightweight, mission-specific containers.
- **Field Tactical Suite:** A Self-contained, deployable application that let operators run, tune, and update AI in the field, without a data scientist on call.

From your first model to full-field operation, the factory handles the entire journey.

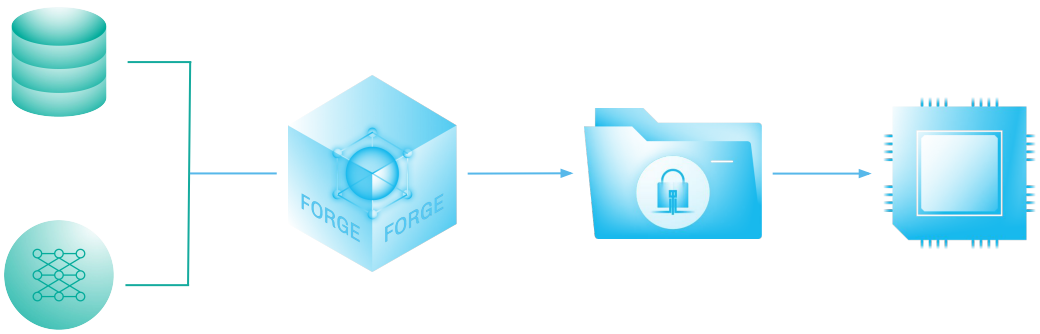
Latent AI Edge AI Factory ecosystem



Software that makes any hardware smarter

Our optimization engine compresses, hardens, and secures AI models to run on the device in front of us, not the device we wish you had. NVIDIA, Intel, Arm, Qualcomm — it doesn't matter. Because we compile once and deploy anywhere, you're not managing a different AI pipeline for every hardware platform in your fleet. You're running one factory that produces the capability for all of them.

How Latent Optimize works



That optimization isn't just about performance. It's about cost.

Latent AI Edge Optimization

- Up to 5x smaller model size on disk
- Up to 73% less RAM usage
- Up to 73% faster inference speed

Real cost cutting results

- 92% reduction in hardware costs on edge server deployments
- 75% reduction in device costs for edge surveillance workloads

The hardware you already own goes further. The hardware you need to buy costs a lot less.

Proof points

For Manufacturers

Running real-time quality control across the plant floor, that's the difference between a viable deployment and one that never leaves the pilot phase.

For Retailers

With hundreds of locations, each running different hardware, this means a single, consistent capability deployed everywhere without a rip-and-replace.

For Defense Programs

Running AI on unmanned systems in contested environments, it means models that adapt to new threats without pulling the system back to the lab.

Working with the U.S. Navy on Project AMMO, Latent AI helped reduce the time to update automatic target recognition models on unmanned underwater vehicles from six months to a few days.

18x
Faster model updates

4x
Faster inference speed

20%
Reduction in power consumption

97%
Reduction in model update time

"We demonstrated the ability to deploy and update our automatic target recognition models at the speed of operational relevance."

Commodore Shaun Lieb, Commander, Task Force 75

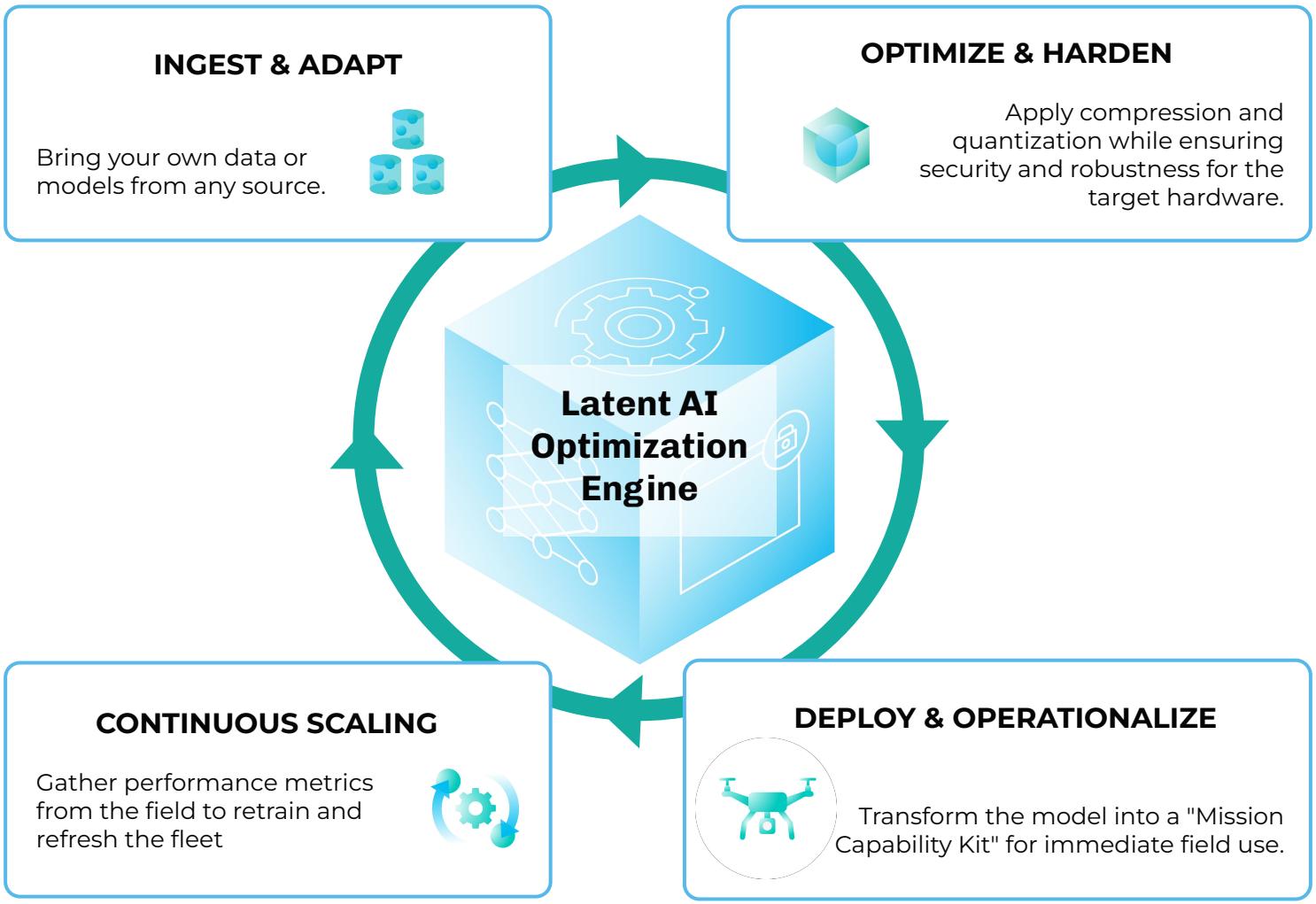
Speed and scale

"What used to take months happens in days."

The factory is designed as a continuous loop. You bring your data and your models. We ingest, optimize, harden, and package them into something field-ready. You deploy. Performance data flows back. The models get smarter. The fleet gets updated. The loop keeps turning.

What used to take months happens in days. And because we compile once and deploy anywhere, scaling from ten devices to ten thousand isn't a reinvention. It's a configuration.

The software layer



AI that works for the people who use it

Deploying AI is only the beginning. Someone still has to run it.

Field teams don't have data scientists. They can't wait weeks for a model update. They need AI that works in their hands, in their environment, on their schedule — and that stays current as conditions change.

That's what the Field Tactical Suite is built for. It bundles the optimized AI, the hardware integration, and the operator interface into a single deployable application. Teams can run it, tune it, and update it in the field. No lab. No specialists. No waiting.

What the Field Tactical Suite is running today:

Underwater Threat Detection
— U.S. Navy

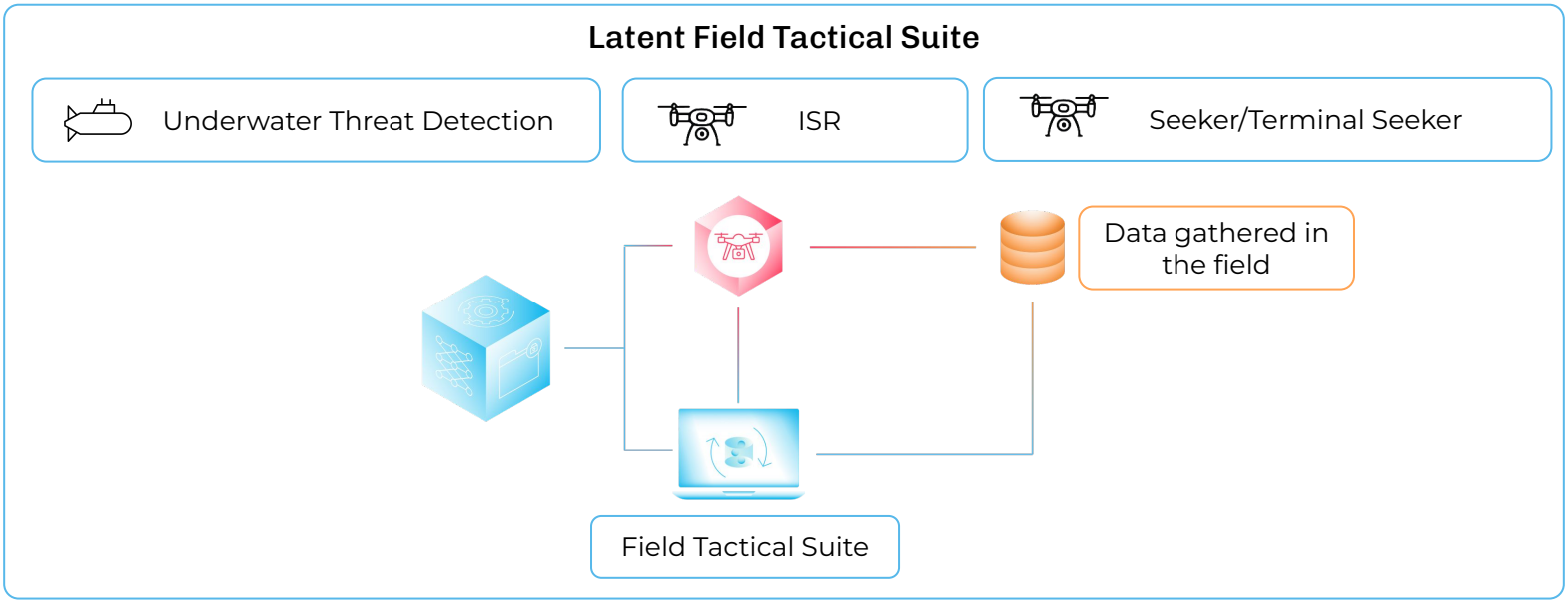
Deployed for Project AMMO, the Field Tactical Suite reduced model update time from months to hours, an 18x improvement, giving sailors the ability to tune and redeploy threat detection models at the speed of the mission.

Intelligence, Surveillance and Reconnaissance
— U.S. Army

In development for U.S. Army programs, the Field Tactical Suite will enable warfighters to gather new data in the field, tune their models, and redeploy, keeping ISR capability current without leaving the theater.

Precision Guidance for Autonomous Systems
— Defense

In development. The Field Tactical Suite is being applied to precision targeting for defense autonomous systems.



Multiple ways to work with us

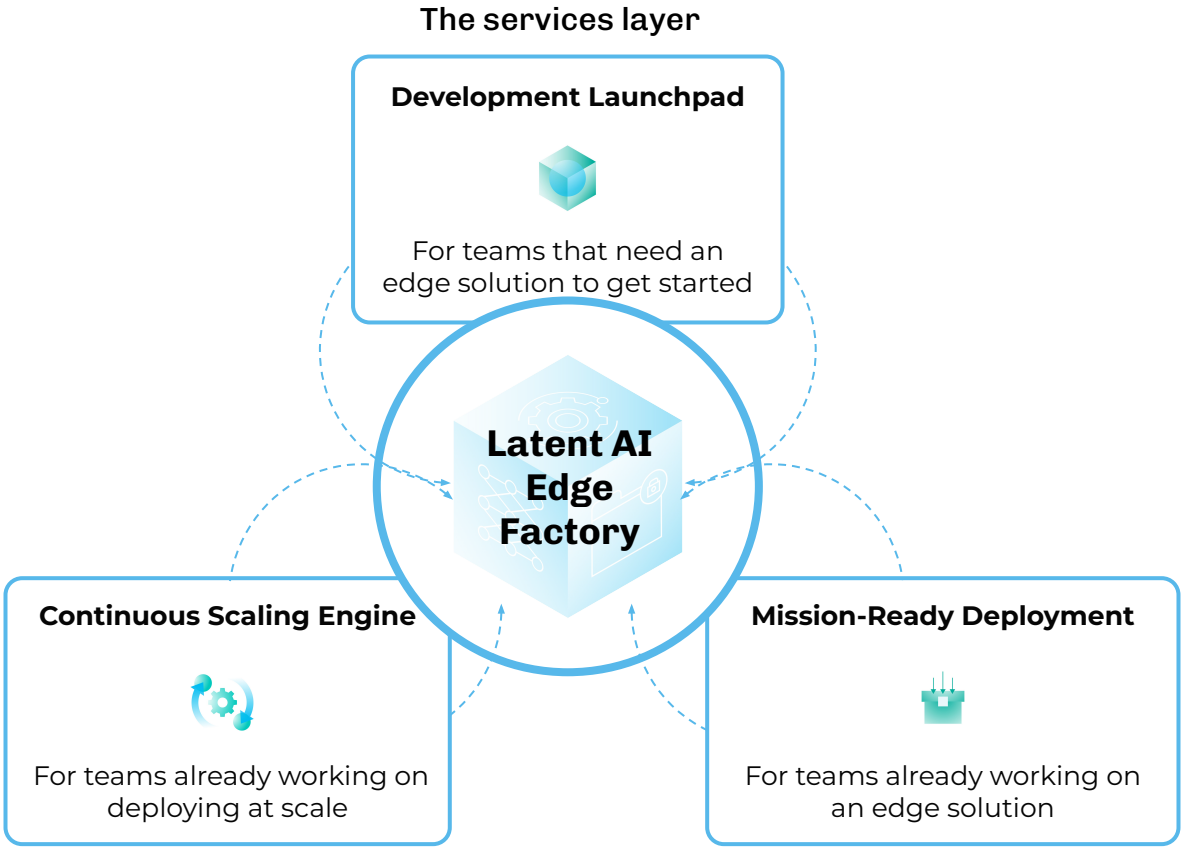
We know different organizations are at different points in their journey. So we built multiple ways in.

Simple Development Launchpad If you're just getting started and need to move quickly, Launchpad gets you up and running with a working edge AI solution without having to build the foundation yourself. You tell us what you need to accomplish; we deliver field-ready AI on your hardware, at your scale.

Continuous Scaling Engine If you have a solution in progress and need to grow it, we take what's working and deploys it further, faster, across more devices, more locations, and even more use cases.

Mission-Ready Deployment If you're ready for full production deployment, Mission-Ready delivers a complete, hardened, field-operational capability optimized for your hardware, secured at the model layer, and built to be maintained and updated by the people who actually use it.

License For platform builders, systems integrators, and hardware vendors who want to bake world-class edge AI optimization into their own stack our software is available to license.



Closing

The edge is everywhere now. The robotic arm on the assembly line. The camera above the checkout. The unmanned system operates where humans can't. Intelligence is moving out of the data center and into the world, and the organizations that figure out how to deploy, update, and scale that intelligence continuously are the ones that will define their industries.

The Edge AI Factory is how you get there.

LatentAI

Ready to get started? Whether you want us to build it, run it, or hand you the keys — let's talk.

latentai.com/get-started