



ModelCat: Accelerate On-Device AI Development

ModelCat is fundamentally changing how organizations create and deploy production-ready AI models for edge and embedded devices. Our patented AI-in-the-Loop fully automates ML model development, from data to deployment, industrializing this historically artisanal process. By replacing slow, costly, risky manual steps with AI, ModelCat delivers unprecedented speed and simplification.

The Bottleneck in On-Device AI

Recent advancements in silicon design have given us powerful machine learning capabilities for edge devices, paving the way for a new generation of intelligent high value offerings.

Unfortunately, tools and practices for developing AI models haven't kept pace.

The process of building and deploying efficient, high-performing ML models on resource-constrained hardware remains a significant bottleneck. Organizations currently deploying ML grapple with a number of challenges:

- **Talent Scarcity & Wasted Spend.** The soaring demand for specialized AI expertise far outstrips supply. This results in millions of dollars in additional spending due to, upfront investment in team building, ongoing costs due to extended project cycles and then at the end poor performance and high failure rates.
- **Protracted Development & Cost Overruns.** Developing and deploying a single, optimized device ML model is an expensive undertaking spanning 12 to 24 months. This lengthy process is built on top of disconnected toolchains and skill sets, limited knowledge, and excessive iterations.
- **Unmanageable Complexity & High Failure Rates.** Navigating rapidly evolving data science and hardware limitations to get to your end goals is inherently complex and error-prone. A high percentage of edge ML projects fail to meet their objectives.

ModelCat offers a fundamentally new and better way to overcome these challenges!



ModelCat, Your AI Model Accelerator

ModelCat fundamentally changes the model development game by transforming the entire ML model-building process into a **single powerful step: Data In, Model Out**. AI abstracts away all of the skills, tooling, and targeting complexities, allowing your teams — from developers and data scientists to product managers — to simply use their expertise to create highly optimized, ready-to-run AI models on demand.

Here's how ModelCat empowers you:

- **A Single Step to a Ready-to-Run Model.** ModelCat autonomously orchestrates the entire model building, evaluation, and deployment process (what we call *model production*). The thousands of steps, decisions, and iterations that define model production today become one single seamless step. How is this possible? We use AI- (as opposed to Human-) in-the-Loop.
- **Data in, Model Out.** Our AI has been trained on tens of thousands of model architectures, training methods, and individual model builds. Hand ModelCat a prepared data set and it will apply all of this knowledge efficiently, running thousands of precision experiments to arrive at a set of best-fit models to choose from — models optimized for your chip and specified speed, accuracy, memory, and power constraints.
- **Hardware Aware.** ModelCat deeply understands the capabilities and performance across an expanding range of chip vendors and chips (e.g., Qualcomm AI Stack, ST Micro N6, NXP MCX-N, Silicon Labs EFR32xG, and many more). Our AI uses this knowledge to hyper-optimize and test models for speed, accuracy, memory and power on your specified target hardware.
- **Increasing Model Performance.** In addition to the tens of thousands of model types already understood by our AI, we continue to train it on new architectures, methods, and it learns from its own operations as well. Every model we build is based on the best of the latest advancements. Our typical 0-2 day model build time means any user can regularly update their models to take advantage of these best-performing innovations.
- **Enabling Cross-Platform Migration.** ModelCat gives you the power to re-platform your model and deploy to any device/chip we currently support. Businesses can now



move to the latest chip innovations with minimal friction, giving teams unprecedented control over device cost and performance.

- **Empowering Your Team.** By embedding model development expertise directly into our AI, ModelCat democratizes model development. Now software developers, data scientists, and product owners can create models based on their specific problem using their own domain expertise. This unlocks opportunities previously inaccessible due to cost and talent gaps, and empowers your internal teams to take control of model development.

How ModelCat Works: Models Built to Order

ModelCat simplifies the journey from problem to model into just a few intuitive steps:

1. **Define Your Problem.** Simply provide your datasets, set your objectives and constraints, and select your target chip. You define speed, accuracy, memory, and power constraints for your model.
2. **AI Builds your Model.** ModelCat's AI then autonomously builds and optimizes a set of models that best match your criteria based on its deep understanding of model building and target hardware.
3. **Hardware-Validated Performance.** Crucially, we eliminate the risk that your model will not perform as expected. ModelCat validates performance of every model by executing them against actual physical chips.

To achieve this we have incorporated an innovative "hardware farm" into our AI. This eliminates the risk from AI hallucinations and confirms your models will perform exactly as expected for speed, accuracy, memory, and power constraints..

4. **Explore & Deploy.** Through our graphical interface, you can explore the generated models, review specific characteristics (like RAM/ROM usage or layer timings). Then download models for immediate deployment onto your own hardware.

The final result? You receive the best possible, custom-optimized model for your specific chip and performance criteria.

The entire process with ModelCat typically takes **0-2 days** depending on the model — a grand reduction from the 12-24 months using current methods — all without compromising model quality.



Built for the Future of AI

ModelCat is using AI to reimagine the future of on-device AI. Our vision is clear: to become the indispensable "model-build layer" for all AI development, so that any business across the \$118+ billion Device/Edge AI market and beyond can add AI models to their products.

Experience ModelCat in Action

To learn more, please reach out to us at sales@modelcat.ai or contact us through our website <https://modelcat.ai>.