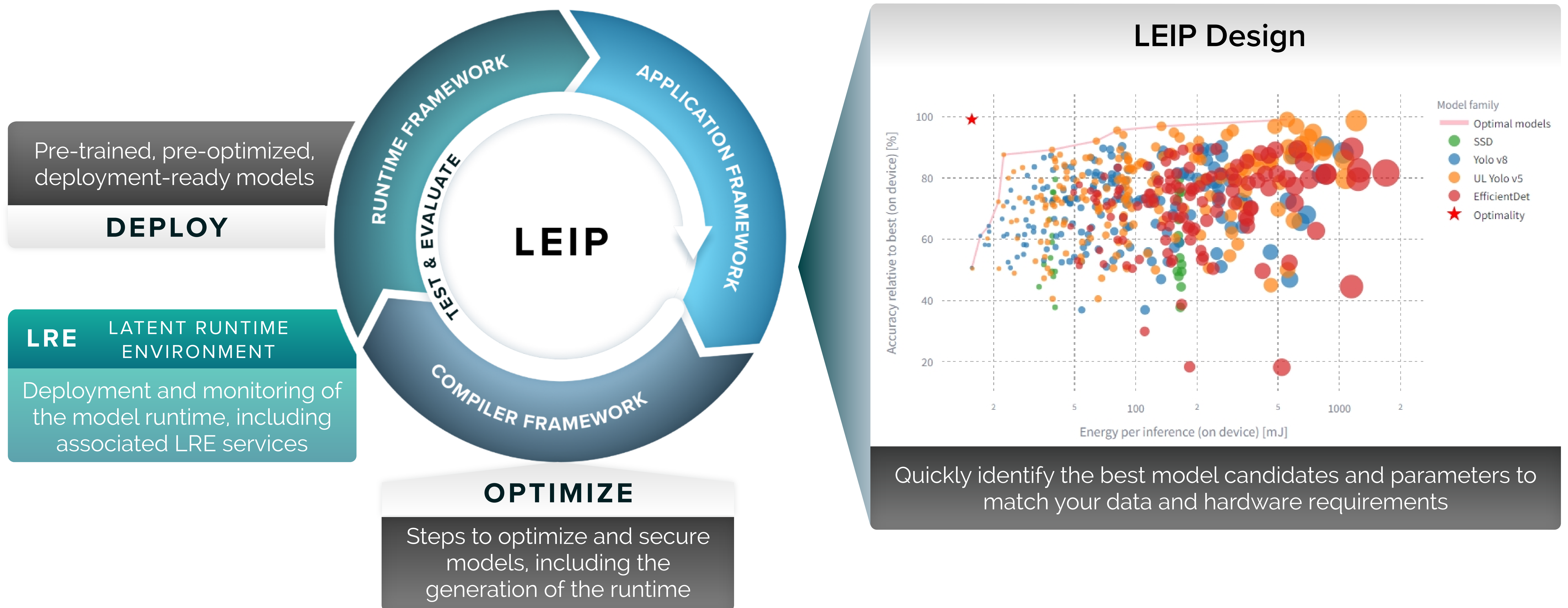


LEIP - Latent AI Efficient Inference Platform

LEIP is a developer-friendly toolkit that provides streamlined MLOps workflows for repeatable, reliable creation of ultra-efficient and adaptable edge AI. It empowers everyone from ML novices to AI experts to create and scale edge AI far faster than they can on their own.



LEIP delivers everything you need for faster edge inference while reducing your costs, complexity, and risk.

LEIP:

A complete MLOps platform that lets you enter at any point of the process

An accelerated path to begin training edge AI with your data

A toolkit that bridges the gap between developers and ML expertise

A single silo-busting DevOps for ML production pipeline all teams can use

A flexible MLOps workflow that easily integrates with your existing tooling

Less Stress. Faster to Market. Adaptable.

Faster Model Evaluation: Streamline evaluation across various model families with consistent Python APIs.

Lower costs: Reduce cloud processing and storage expenses while lowering power consumption with edge computing

Scale: Deliver reproducible, lightweight, efficient AI in a containerized executable via repeatable, reliable, dedicated MLOps processes.

Protect Your IP, Data & Brand: Bake security in with built-in model encryption and watermarking delivered via trusted and reliable processes.

Private Cloud and On-Premise Support: Deploy LEIP in your preferred environment, whether on-premise or in a private cloud.

Retrain and Redeploy: Leverage existing models and adapt to changing hardware or data requirements without starting over.

LEIP lets you build lightweight and powerful AI that moves decision making to where it's needed most - at the data source.



Faster to design. Quicker to ML. Guaranteed Performance.

LEIP Design enables you to develop efficient AI models faster. Its rapidly **growing library of 50k+ pre-built Recipes** provide pre-configured options for building and deploying high-performing models.

Recipes simplify development by allowing you to compare model characteristics across various hardware targets (CPUs, GPUs) and **ensure the perfect match for your data and processing needs.**

Lower your latency while lowering your costs
Close the gap between code and edge devices
Hardware research and ML expertise baked in
Create **scalable, repeatable, shareable** Edge AI

Make Informed Decisions:

Evaluate models based on performance and resource requirements before training.

Find the Right Fit: Ensure you are using the best combination of model and hardware for your data type and complexity.

Evaluate Existing Models:

Assess if existing models can perform on new hardware without starting from scratch.

Balance Performance vs.

Resources: Explore tradeoffs between hardware limitations and project needs (accuracy, speed, size, energy).

Path to GRDB
/opt/latentai/recipes/3.0.3

Golden Recipe Volume
ALL

Target architecture
cuda:AGX-Orin

x-axis
Inference speed (on device) [fps]

y-axis
Accuracy relative to best (on de...)

Color
Model family

Marker size

Marker size
Inference size (on device) [MB]

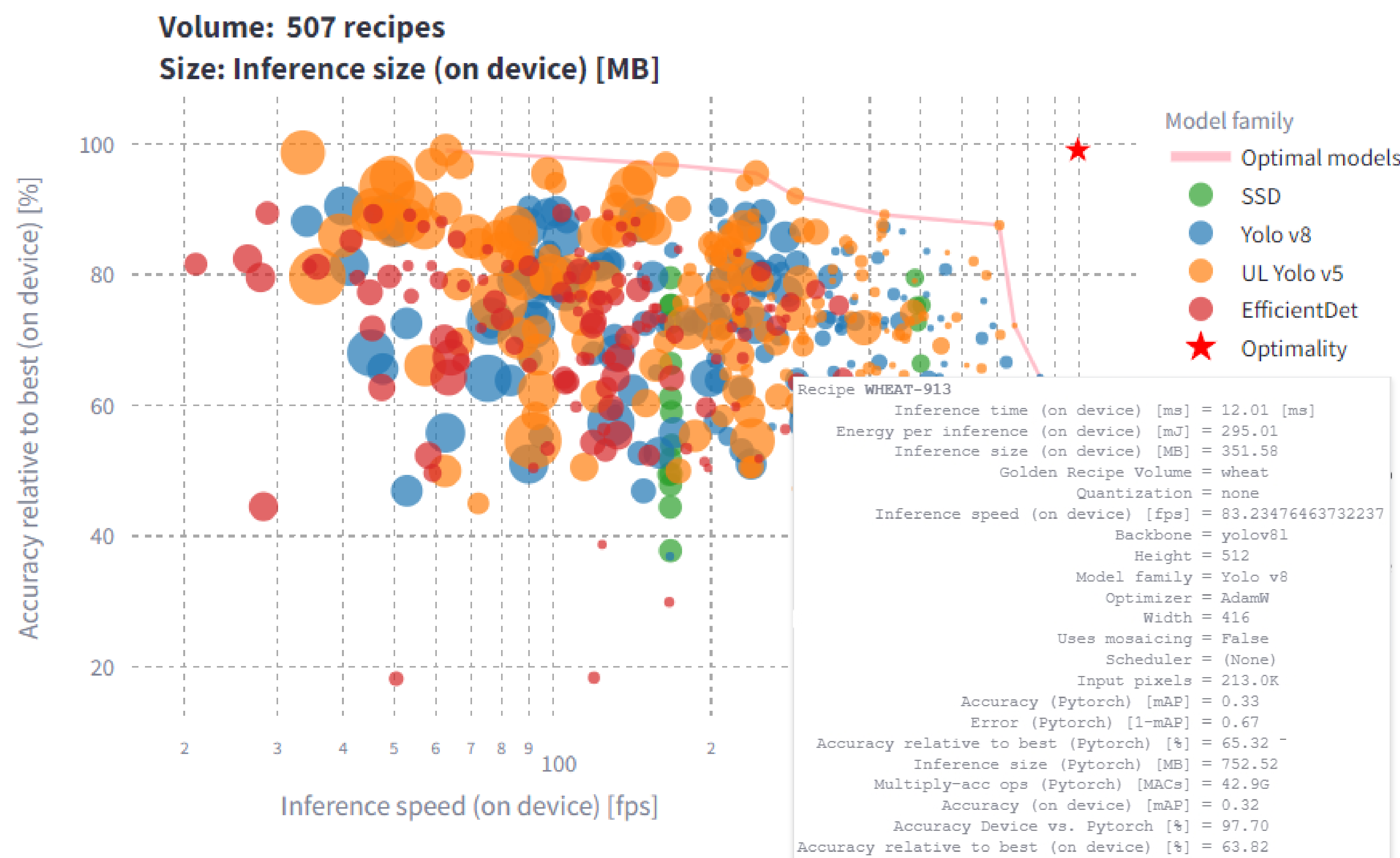
Plot Pareto-optimal models only

Log x-axis

Log y-axis

Comprehensive tooltips

LEIP Design



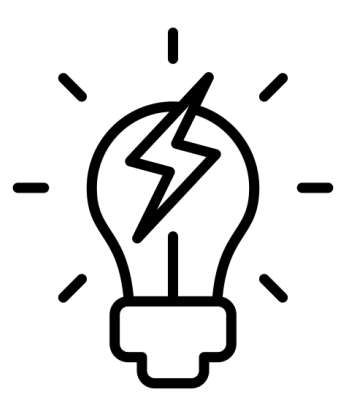
With LEIP Design you can:

Design and build shareable, repeatable ML model templates

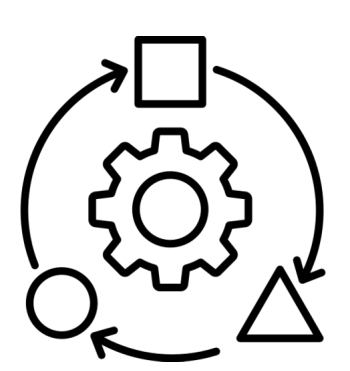
Quickly find your best model and hardware combination

Design a model and start training it within minutes

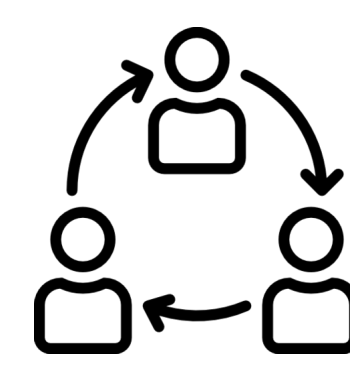
Understand how your model will perform before you train it



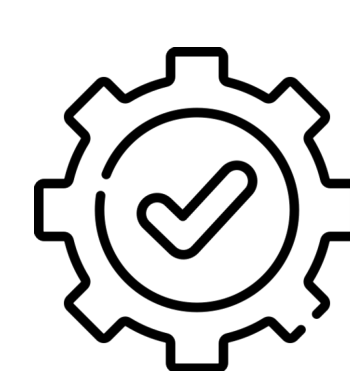
Predictive Deployment Performance: Recipes verified to work on many common hardware architectures with expected performance known before you start training



Hardware Flexibility and Optimization: Tools to quickly identify the best model candidates and parameters to match your data and hardware requirements



Collaborative Recipe Management: Easily share, re-use, and manage models across different users, groups, and teams.



Reduced Development Risk: De-risk your development process by guaranteeing hardware compatibility before investing time and resources into training.

Easily compare model performance across different hardware families

Quickly identify your best model candidate for training

Understand model performance before you commit

Start machine learning in minutes

Contact mlops@latentai.com to schedule an evaluation today! Scan QR code for PDF

