




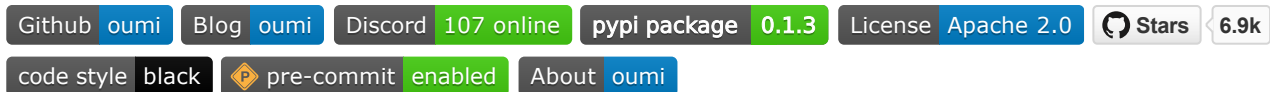
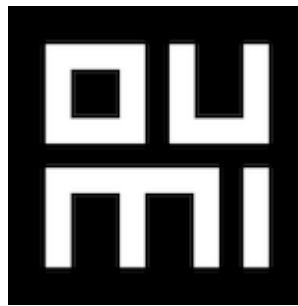




# Getting Started

## Contents

-  Getting Started
-  Why use Oumi?
-  Where to go next?
-  Join the Community!
-  Need Help?










Everything you need to build state-of-the-art foundation models, end-to-end.

















Oumi is a fully open-source platform that streamlines the entire lifecycle of foundation models - from data preparation and training to evaluation and deployment. Whether you're developing on a laptop, launching large scale experiments on a cluster, or deploying models in production, Oumi provides the tools and workflows you need.

With Oumi, you can:

-  Train and fine-tune models from 10M to 405B parameters using state-of-the-art techniques (SFT, LoRA, QLoRA, DPO, and more)
-  Work with both text and multimodal models (Llama, DeepSeek, Qwen, Phi, and others)
-  Synthesize and curate training data with LLM judges
-  Deploy models efficiently with popular inference engines (vLLM, SGLang)
-  Evaluate models comprehensively across standard benchmarks
-  Run anywhere - from laptops to clusters to clouds (AWS, Azure, GCP, Lambda, and more)
-  Integrate with both open models and commercial APIs (OpenAI, Anthropic, Vertex AI, Parasail, ...)

All with one consistent API, production-grade reliability, and all the flexibility you need for research. Oumi is currently in [beta](#) and under active development.

Notebook	Try in Colab	Goal
 <b>Getting Started: A Tour</b>	 <a href="#">Open in Colab</a>	Quick tour of core features: training, evaluation, inference, and job management
 <b>Model Finetuning Guide</b>	 <a href="#">Open in Colab</a>	End-to-end guide to LoRA tuning with data prep, training, and evaluation
 <b>Model Distillation</b>	 <a href="#">Open in Colab</a>	Guide to distilling large models into smaller, efficient ones
 <b>Model Evaluation</b>	 <a href="#">Open in Colab</a>	Comprehensive model evaluation using Oumi's evaluation framework
 <b>Remote Training</b>	 <a href="#">Open in Colab</a>	Launch and monitor training jobs on cloud (AWS, Azure, GCP, Lambda, etc.) platforms
 <b>LLM-as-a-Judge</b>	 <a href="#">Open in Colab</a>	Filter and curate training data with built-in judges
 <b>vLLM Inference Engine</b>	 <a href="#">Open in Colab</a>	Fast inference at scale with the vLLM engine

If you need a comprehensive platform for training, evaluating, or deploying models, Oumi is a great choice.