



# Power tools for your model development lifecycle

## Scale and operationalize your deep learning experiments

Determined AI is an **automated deep learning (DL) platform** that improves the productivity of DL engineers exponentially. With Determined AI, DL teams can rapidly build production-ready models and readily deploy them on the designated hardware, even on a device limited in resources such as an edge device.

Experiment 5  
mnist\_tf\_adaptive

Add label

Archive Fork

Configuration View

Model Definition Download

State: COMPLETED

Progress

Max Slots: Unlimited

Start Time: 2019/01/17, 4:57:17 PM

End Time: 2019/01/17, 4:59:42 PM

Metrics

Best Validation Metric vs Elapsed (seconds) graph

### Active Trials

ID ↓	H-Params	Start Time ↓	Steps ↓	Latest Validation Metric ↓	Best Validation Metric ↓
------	----------	--------------	---------	----------------------------	--------------------------

### Completed Trials

ID ↓	H-Params	Start Time ↓	End Time ↓	Steps ↓	Latest Validation Metric ↓	Best Validation Metric ↓
287	<a href="#">View</a>	2019/01/17, 4:57:17 PM	2019/01/17, 4:59:40 PM	4	0.02578	<a href="#">Log</a> <a href="#">Details</a> <a href="#">Continue Training</a>
288	<a href="#">View</a>	2019/01/17, 4:57:17 PM	2019/01/17, 4:59:40 PM	4	0.02816	<a href="#">Log</a> <a href="#">Details</a> <a href="#">Continue Training</a>
289	<a href="#">View</a>	2019/01/17, 4:57:17 PM	2019/01/17, 4:59:40 PM	4	0.02647	<a href="#">Log</a> <a href="#">Details</a> <a href="#">Continue Training</a>
290	<a href="#">View</a>	2019/01/17, 4:57:17 PM	2019/01/17, 4:59:27 PM	1	0.18690	<a href="#">Log</a> <a href="#">Details</a> <a href="#">Continue Training</a>
291	<a href="#">View</a>	2019/01/17, 4:57:21 PM	2019/01/17, 4:59:37 PM	1	0.12679	<a href="#">Log</a> <a href="#">Details</a> <a href="#">Continue Training</a>

Organizations using Determined AI have reported significant increases in both team and individual productivity, with some customers stating they save a minimum of one day per engineer per week. Additional benefits include:

### Stop waiting weeks for your models to converge

With Determined AI, DL engineers can readily take advantage of *distributed and/or parallel training* on their existing hardware infrastructure (e.g., a GPU cluster) with a few lines of code. Our specialized *hyperparameter search* optimizes the search space to get to quality models faster. Experiments are executed in containerized environments and all experimental meta-data is tracked. Collectively, these features enable seamless *reproducibility* and *warm starting* of future experiments to expedite subsequent results.

### Collaboration and sharing of GPU resources

Determined AI's industry-leading job scheduling algorithm enables DL engineers to maximize the number of experiments that can be run simultaneously. Our technology has the ability to resize jobs dynamically as new jobs come in or hardware goes on/offline. DL engineers can also share their compute resources anywhere according to their jobs' priorities. Your GPU resources will be utilized in an optimal way to accommodate all of your DL experiments.

### Automated hyperparameter optimization and neural architecture search

The Determined AI team has intimate knowledge about the deep learning problem space gained through research at **UC Berkeley** and **Carnegie Mellon University**. Our DL optimization technologies, including *hyperparameter search*, *DL job mapping and scheduling*, and *automatic architecture design* that *satisfies deployment SLAs*, are much more effective than generic solutions.

### Bring your own software and hardware ecosystems

Determined AI can be run on premises and/or in the cloud. It is compatible with:

#### Deep learning frameworks:

TensorFlow, Keras, PyTorch

#### Data storage and management systems:

NFS, Amazon S3, HDFS, GlusterFS, and more

#### Container Orchestration Frameworks:

Kubernetes, DC/OS, bare metal

# Key features of Determined AI for deep learning

## AutoML at scale

Deliver quality models in less time

- ✓ State-of-the-art hyperparameter and neural architecture search
- ✓ Push-button transfer learning and retraining
- ✓ Automatic data-parallel distributed training
- ✓ Data access accelerator
- ✓ TensorFlow, Keras, or PyTorch support

## Experiment tracking

Enable reproducibility and collaboration

- ✓ Integrated metrics capture and model version management
- ✓ Reproducible, containerized training
- ✓ Label and share experiments
- ✓ Multi-tenancy in a shared infrastructure
- ✓ Real-time experiment metrics visualization

## Seamless infrastructure

Optimize resource utilization and ensure experiment reliability

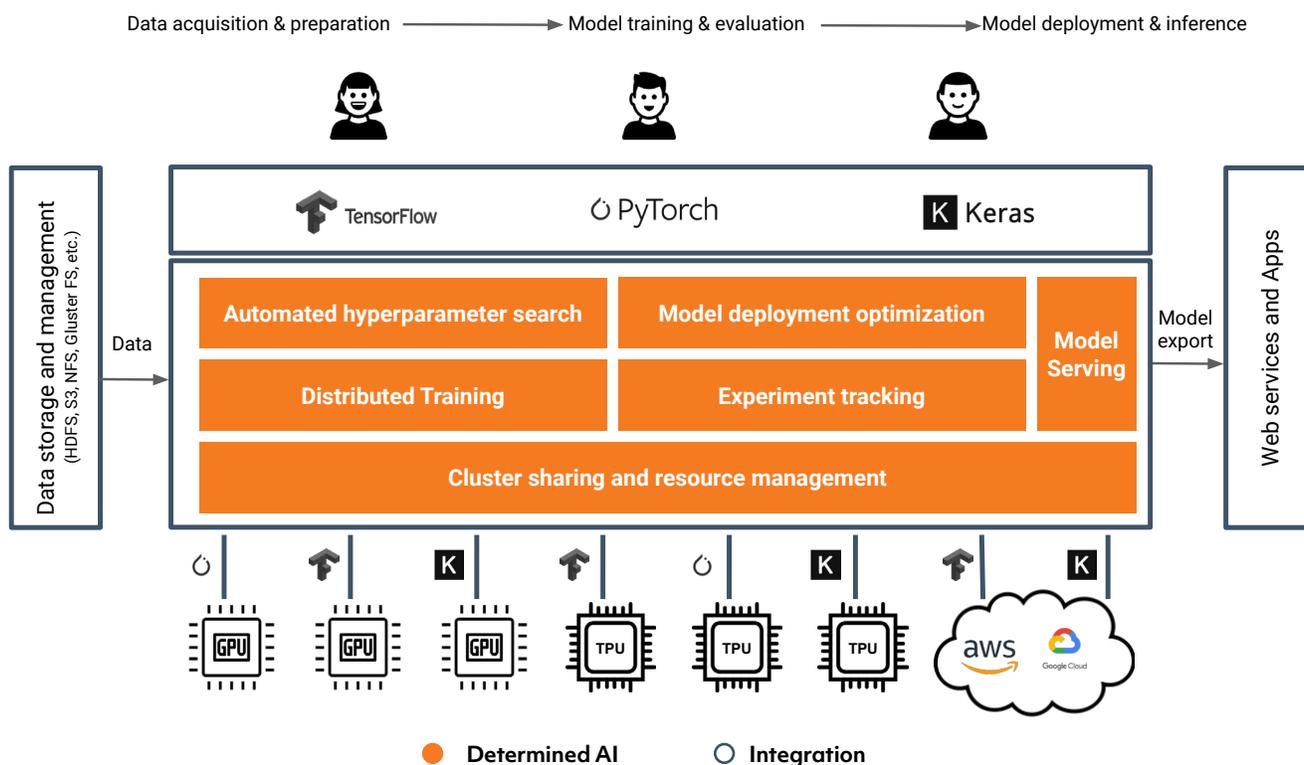
- ✓ Fault-tolerant training
- ✓ Efficient scheduling and execution of experiments
- ✓ GPU-attached Jupyter Notebook and TensorBoard integration
- ✓ Share GPU resources on premises, in the cloud, or both.
- ✓ Run on bare metal, and in Kubernetes

## Model deployment optimization

Ensure model deployability and reduce time-to-market

- ✓ Explore models with deployment constraints
- ✓ Simulate inference speed on deployment hardware
- ✓ Automated architecture search for the edge, cloud, and mobile
- ✓ Containerized deployment in a single click

# Determined AI offers holistic and specialized AI-native infrastructure



See Determined AI documentation (<https://docs.determined.ai>) for installation details and system requirements.