

Introduction to Ray AI Libraries

Emmy Li, Balaji Veeramani, Yunxuan Xiao





Welcome!

We're happy to have you here.





Meet the team!



Emmy



Balaji



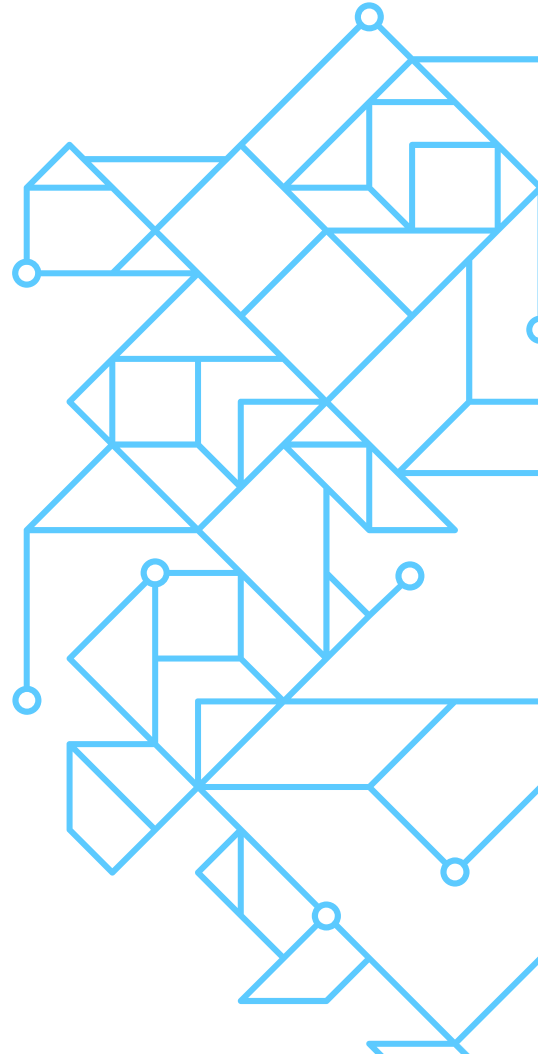
Yunxuan





The Plan

Here's what to expect today.





Today's agenda.

9:00am (15 min)	Talk: Introduction to Ray AI Libraries
9:15am (15 min)	Demo: End-to-end mini example
9:30am (60 min)	Coding Lab: HF Vision Transformer + Ray Train for image classification
10:30am (15 min)	Coffee Break
10:45am (60 min)	Coding Lab: Scaling out with Ray Data and Ray Serve
11:45am (15 min)	Talk: Resources for Further Exploration

Course Set-Up

Tech checks all around.





Tech check.



Participating via app.sli.do

- Join with code **#ray-ailib**
- Ask questions.
 - Pose your own and upvote others.
 - TAs will be answering questions on a rolling basis.



Tech check.

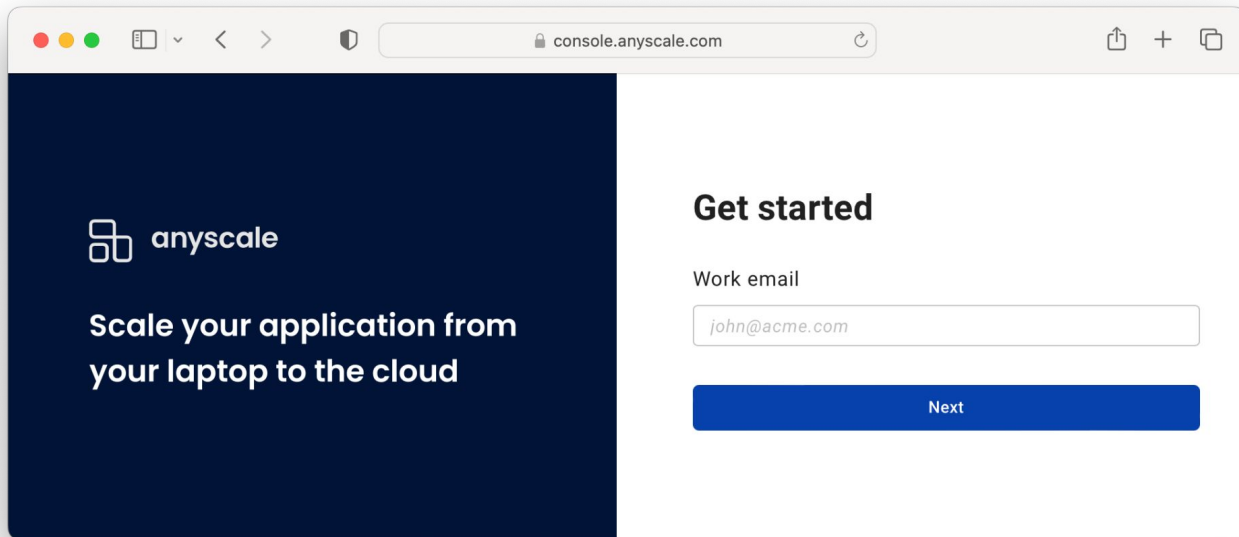


Accessing Anyscale clusters.

- All work will be in Anyscale provisioned clusters.
- Our GitHub repo will be mounted automatically.
- Access begins now.
 - Check your email for login information.
 - Step-by-step instructions to follow.

Anyscale login

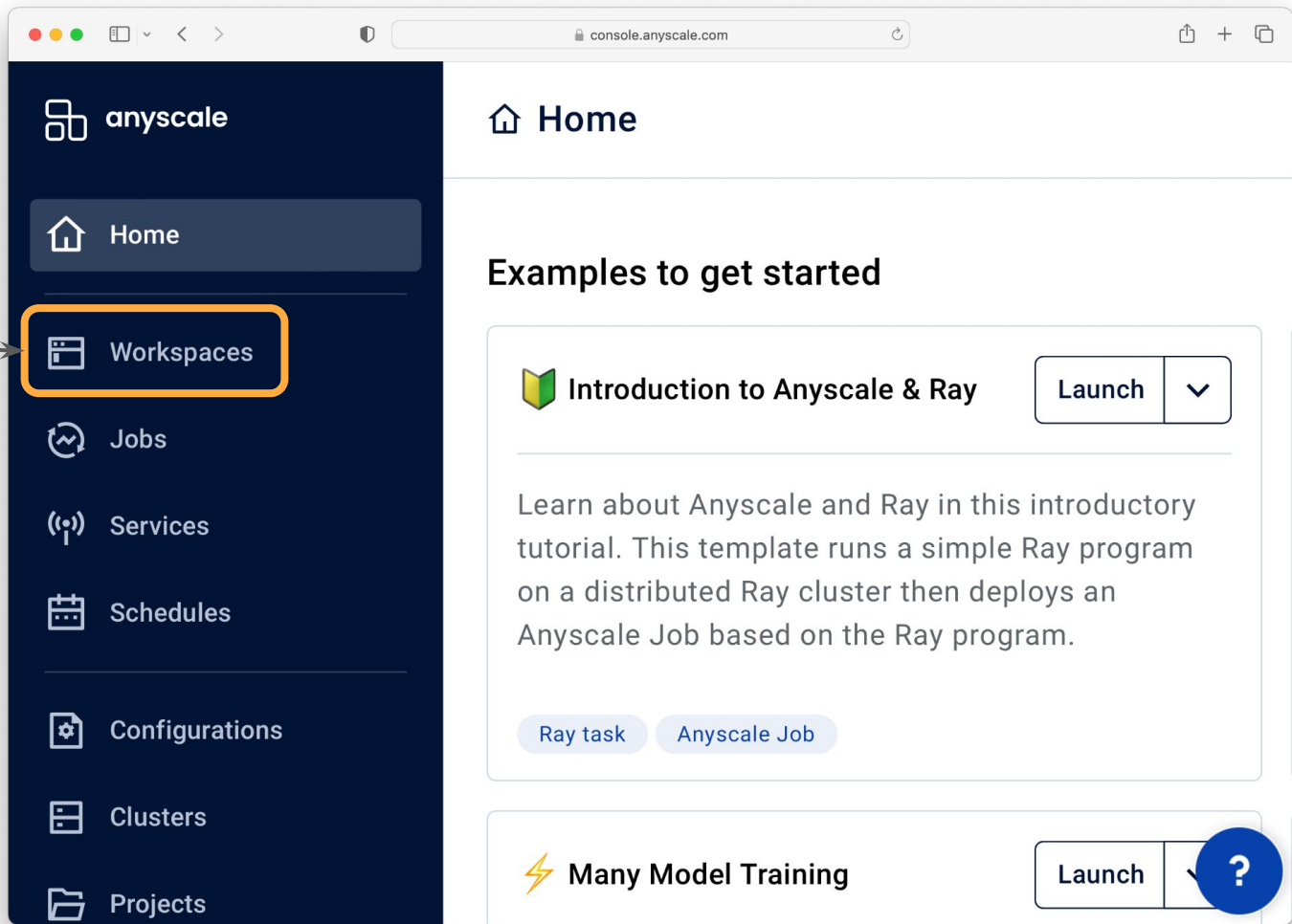
Link to Anyscale cluster: console.anyscale.com



A screenshot of a web browser window showing the Anyscale console login page. The browser's address bar displays 'console.anyscale.com'. The page is split into two main sections. The left section has a dark blue background and features the Anyscale logo, the text 'anyscale', and the slogan 'Scale your application from your laptop to the cloud'. The right section has a white background and is titled 'Get started'. It contains a 'Work email' label, a text input field with the placeholder 'john@acme.com', and a blue 'Next' button.

Enter the
**unique
credentials**
sent to your
email!

1. Select Workspaces



2. Select Your Workspace

The screenshot displays the Anyscale console interface. On the left, a dark blue sidebar contains the 'anyscale' logo and a list of navigation items: Home, Workspaces (highlighted), Jobs, Services, Schedules, Configurations, and Clusters. The main content area is titled 'Workspaces' and features a header with buttons for '+ Create', 'Start', 'Terminate', and 'Delete'. Below this is a search bar labeled 'Search names' and a filter button labeled 'Created by is me'. A table lists the workspaces with columns for 'Name' and 'Status'. The table contains one entry: 'my-workspace' with a status of 'Active'. The 'my-workspace' name is highlighted with an orange box. Below the table, the text '1 - 1 of 1' is displayed. A blue circular help button with a question mark is located in the bottom right corner.

Name	Status
my-workspace	Active

3. Click on
Jupyter
icon

The screenshot shows the Anyscale console interface. On the left is a dark blue sidebar with navigation links: Home, Workspaces (selected), Jobs, Services, Schedules, Configurations, Clusters, Projects, Emmy, Help, and Feedback. The main content area is titled 'm... workspace' with a status 'Active (Ray)'. Below the title are tabs for About, Files, Terminal, Logs, and Serve deployments. A table displays workspace details:

Status	Resources	Access ②
Active (Ray)		Everyone in your org
Created Sep 7, 2023 at 4:26:50 PM, by emmy+education@anyscale.com	Cluster environment summit:9 🔗	Compute config ray-summit-2023-gc
Network access Public with auth token 🔗	Job submissions None	Ports ② 36075

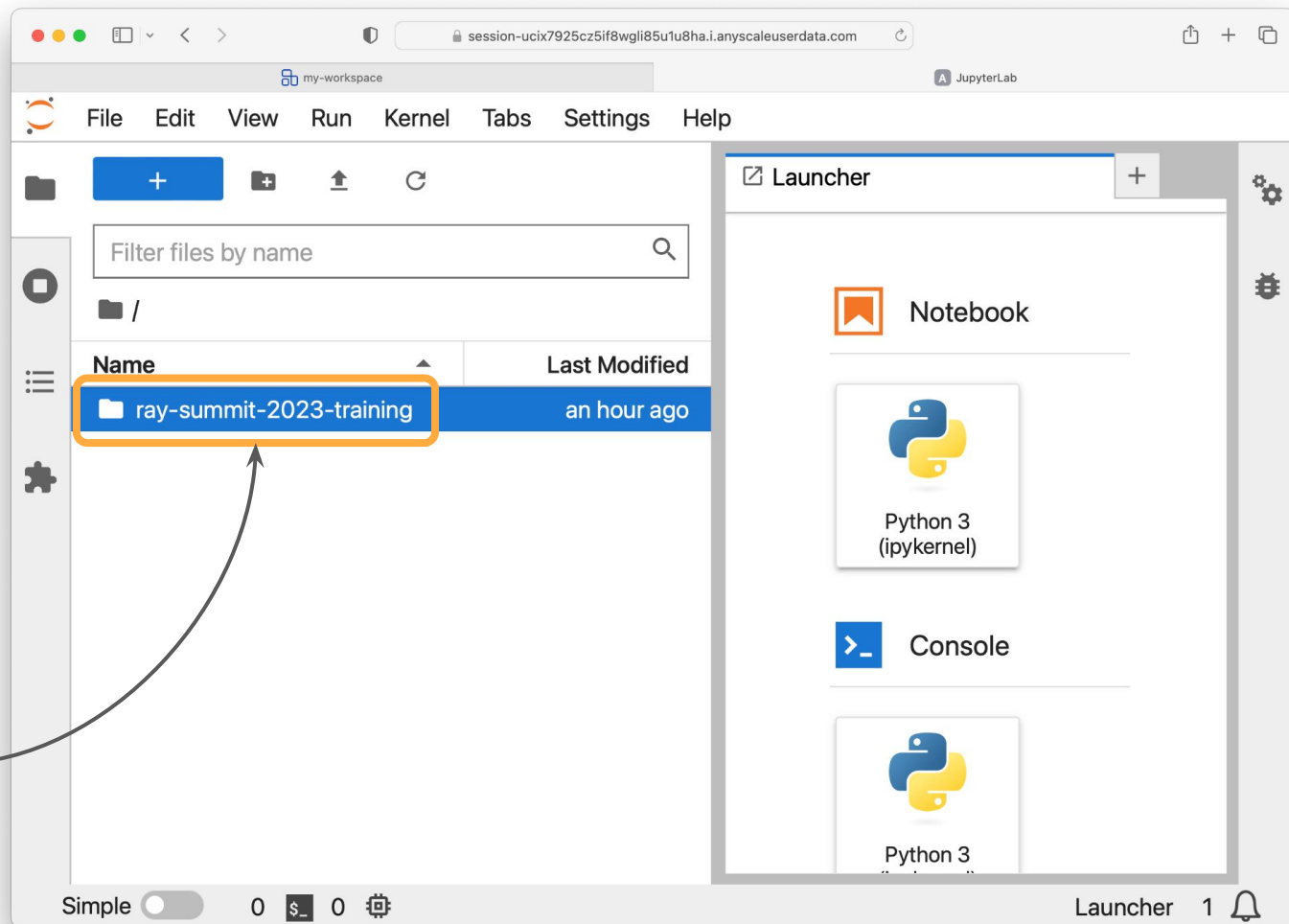
Below the table is a 'README' section titled 'Workspaces' with the following text:

A Workspace is a fully managed development environment focused on developer productivity. We enable ML practitioners and ML platform developers to quickly build distributed Ray applications from research to development to production easily, all within a single environment.

Workspaces provide a remote experience for programming your cluster while working with JupyterLab notebooks or Visual Studio Code.

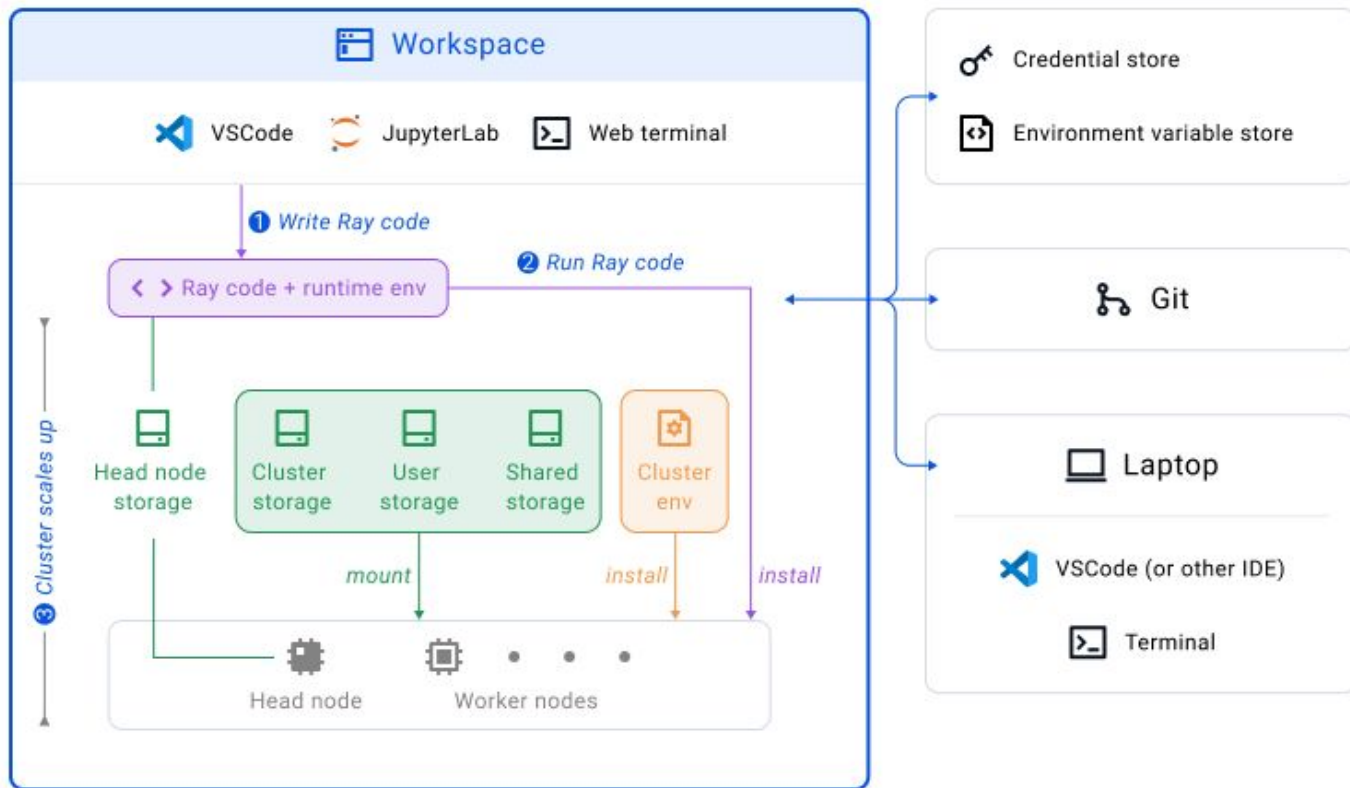
In the top right of the workspace header, there are icons for JupyterLab (highlighted with an orange box and an arrow from the text '3. Click on Jupyter icon'), Visual Studio Code, and a network icon. To the right of these icons are buttons for 'Terminate' and 'Tools'.

4. Find the
content for
your class
here.





What are Anyscale Workspaces?



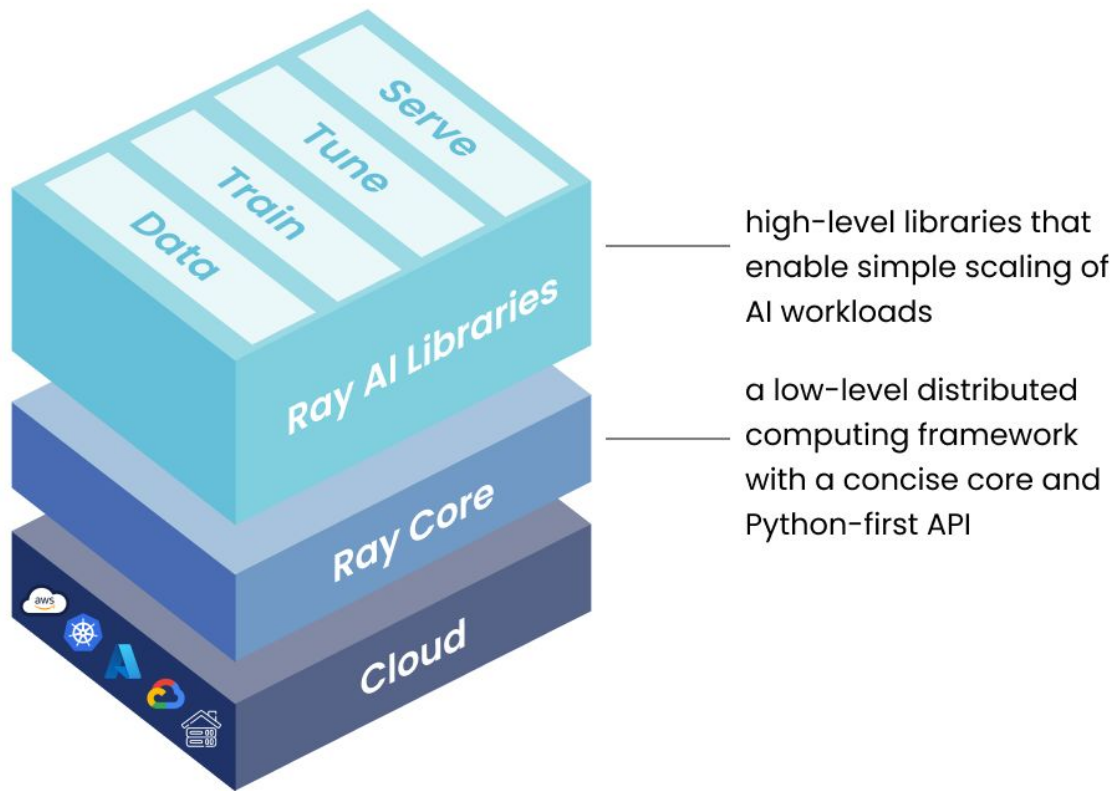
Ray AI Libraries

Your morning briefing.





Overview of Ray AI Libraries





Ray Data

A scalable data processing library well-suited for distributed data ingest, preprocessing, and batch inference.

Highlights

- + Streaming execution across CPUs and GPUs
- + World-record breaking scalability.
- + Peak performance for offline batch inference



Ray Train

A library for **distributed training and fine-tuning** with integrations with PyTorch, Tensorflow, XGBoost, and much more.

Highlights

- + Training for LLMs across 1000s of GPUs.
- + Fine-tuning LLMs with DeepSpeed.
- + Cost-savings on heterogeneous hardware.



Ray Tune

Easily scalable **hyperparameter tuning.**

Highlights

- + Integrations with Hyperopt, Optuna, Nevergrad and more with access to algorithms like PBT and HyperBand/ASHA.
- + Many model training for millions of workloads.



Ray Serve

Scalable **model serving** library for building online inference APIs.

Highlights

- + Optimizations for LLMs like response streaming, dynamic request batching, multi-node serving, etc.
- + Framework-agnostic serving from huge models to Python business logic.



The essentials.



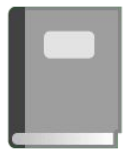
Scalable ML – Targeted libraries for large machine learning workloads.



Ease of use – Pythonic distributed computing primitives designed for smoother dev-to-prod.



Flexibility – Integrations with popular machine learning frameworks (e.g. PyTorch, Tensorflow, LightGBM etc.) and infrastructure (e.g. Kubernetes, AWS, GCP, Azure, etc.)



Let's make our way over to the notebooks!



**Time for a
Break!**

15 minutes.

More Resources

For further exploration with
Ray, Anyscale, and GenAI.





Today we learned...



Overview of Ray AI Libraries

Getting acquainted with each library for distributed ML.



Hugging Face ➡ Ray

Converting a vision transformer to run distributed.



Exploring Ray Train, Data, Serve

Constructing an end-to-end ML pipeline with Ray.



Sneak Peek: Self-Paced Ray & Anyscale Education



Access to sharable course materials will be emailed to you after Ray Summit.



Preview special technical content releases from the whole team!



Fill out the survey.



Go to bit.ly/ray-summit-feedback





Reading list.



Self-Paced Ray & Anyscale Education

Access bonus notebooks and scripts about Ray.



[Ray documentation](#)

API references and user guides.



[Anyscale Blogs](#)

Real world use cases and announcements.



[YouTube Tutorials](#)

Video walkthroughs about learning LLMs with Ray.



Upcoming events



Bay Area AI + Ray Summit Happy Hour

Today at 5:00p.m.

Cap off an exciting conference with lightning talks, new friends, and good times!

bit.ly/bayai_ray_meetup





Connect with the community.



Join the community

[Attend events](#), [subscribe to newsletter](#), [follow on Twitter](#).



Get support

[Join Ray Slack](#), [ask questions on forum](#), [open an issue](#).

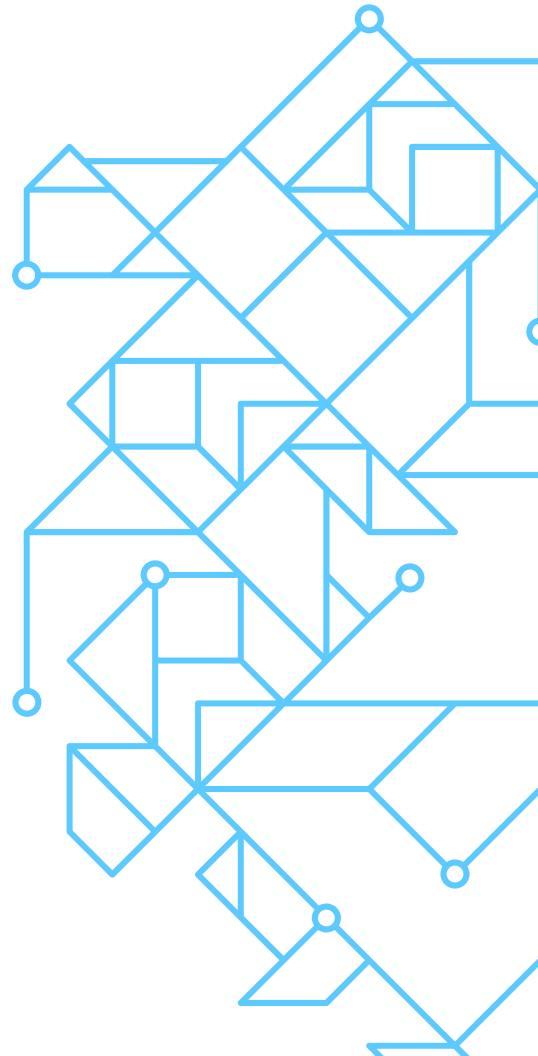


Contribute to Ray

[Read contributor guide](#), [create a pull request](#).

Thank you!

We hope to meet again.



Ray Summit 2023 Color Palette



7A8EA3



95EEF5



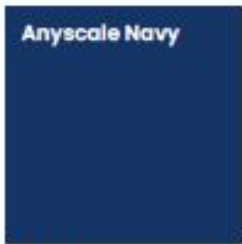
5CBCE



028CF0



234999



143566



4834B3

Here is a basic Dark Slide

Slide Template

Keynotes

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation

Here is an info card

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation

Slide Template

Keynotes

- Start to storyboard the keynote presentations
- Build out the stage design and presentation requirements
- Connect with external speakers on themes/topics

Production Costs

- Original estimates are lower than the quotes now coming in.
- Upgraded production quality results

Registration

- Registration will continue to be a main area of focus especially as we approach