



Building Real-Time Inference Pipelines with Ray Serve



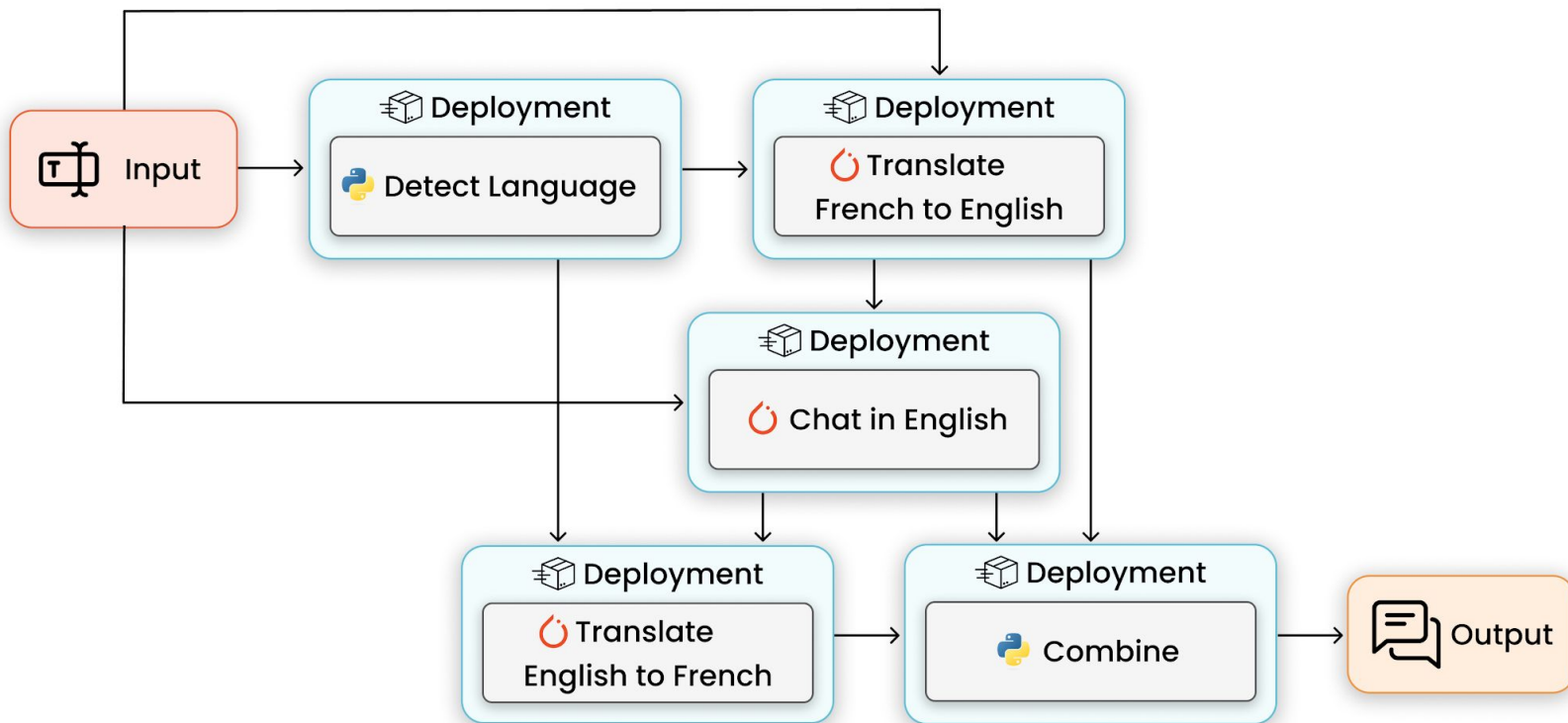
Who am I?



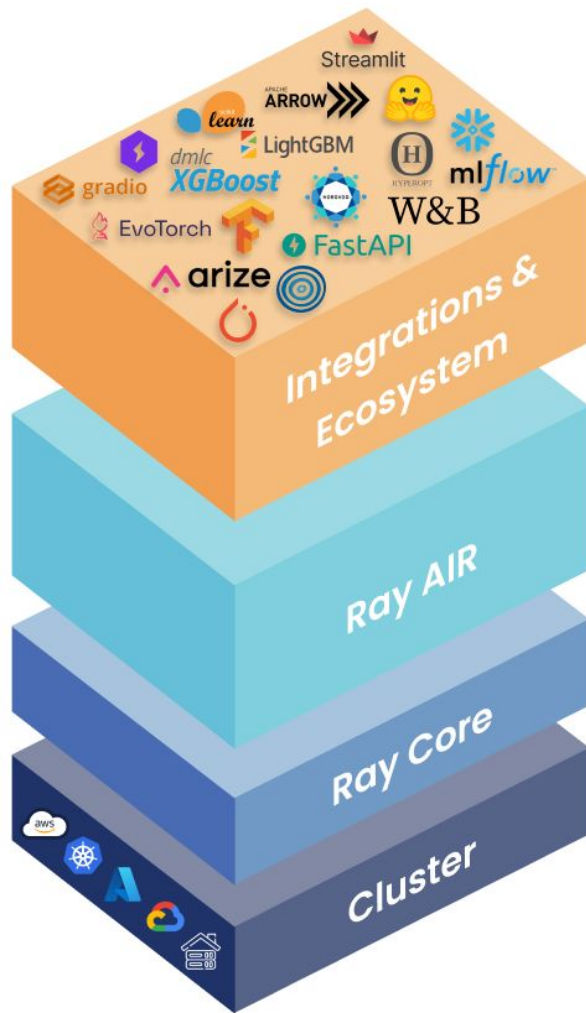
Adam
technical trainer @ 



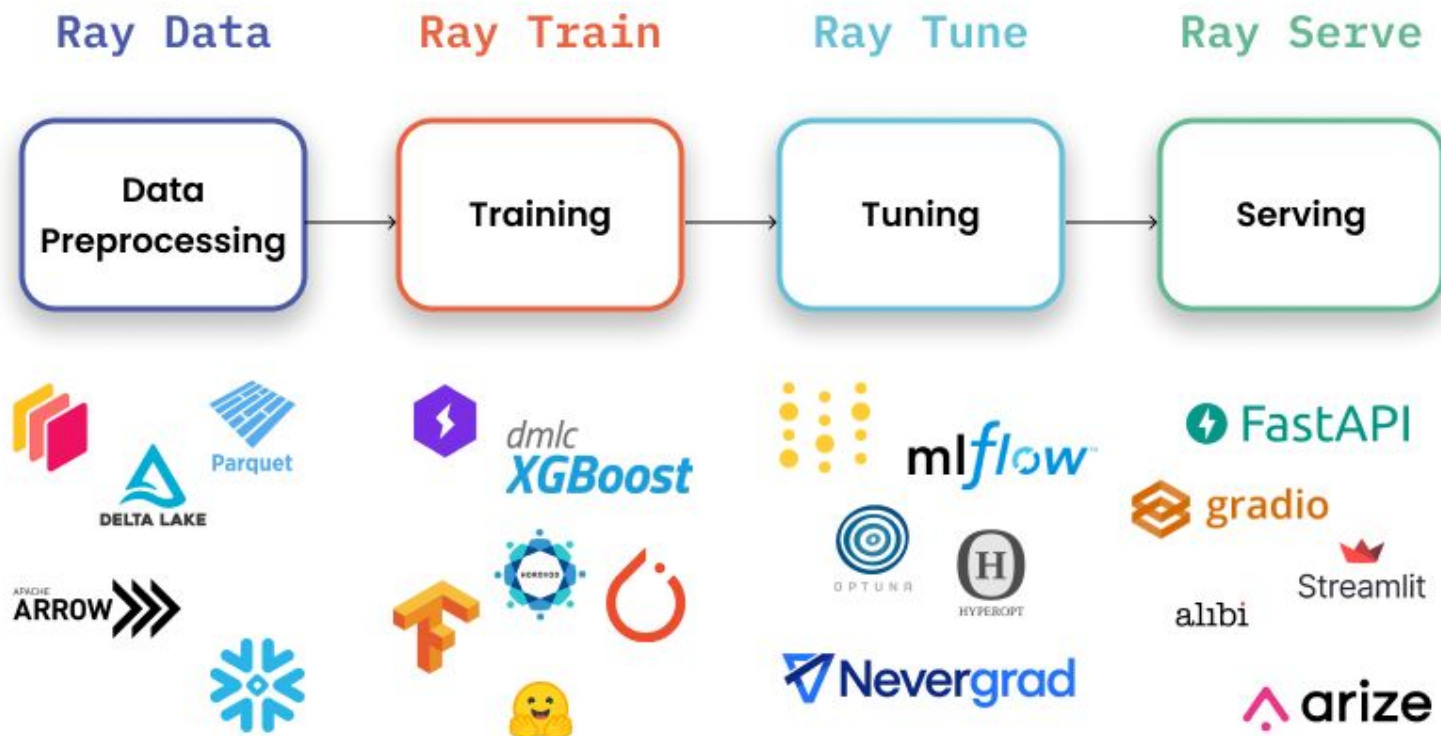
What will we build today?



What is Ray?



↔ End-2-end ML scaling





Ray serve for online inference

Ray Cluster

Node 1

 Deployment



Replica 1



Replica n

Node n

 Deployment



Replica 1



Replica n



Resources



bit.ly/arize-ray-serve

Access all the notebooks, slides, and scripts.



[Ray documentation](https://ray-project.github.io)

API references and user guides.



bit.ly/raysummit2023

The place for everything Ray.

Thank you!

We hope to meet again.

