

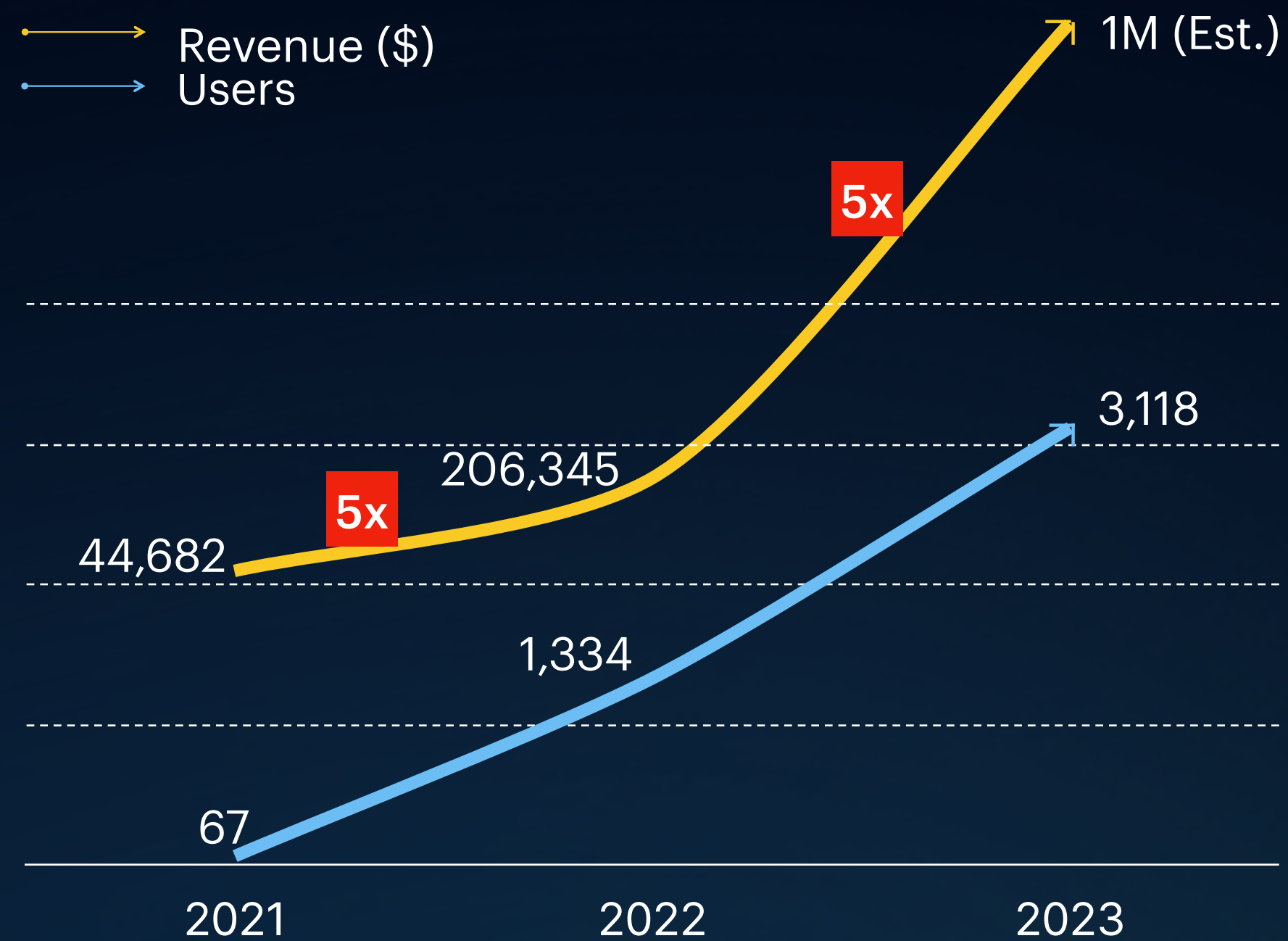


**We build the best MLOps infrastructure
for developing AI faster at scale.**



The most ambitious ML teams use VESSL

Numbers



VESSL is used for

- Fine-tuning LLMs
- Deploying enterprise-scale AI models
- Building ML Pipeline
- Research Platform for top universities

Key Customers



Hyundai Motors' Autonomous Driving Team uses VESSL as a collaboration platform for 100+ ML engineers.



ScatterLab trains their LLMs using 20+ A100 GPUs from multiple clouds services with VESSL.



The Graduate School of KAIST AI provisions over 1000 GPUs to 200+ ML researchers.



TMAP Mobility deploys AI models via VESSL and saves 60% of computing costs.

Partners

Accelerated by **SAMSUNG**

Google Cloud **Partner**
INCEPTION PROGRAM



Investors

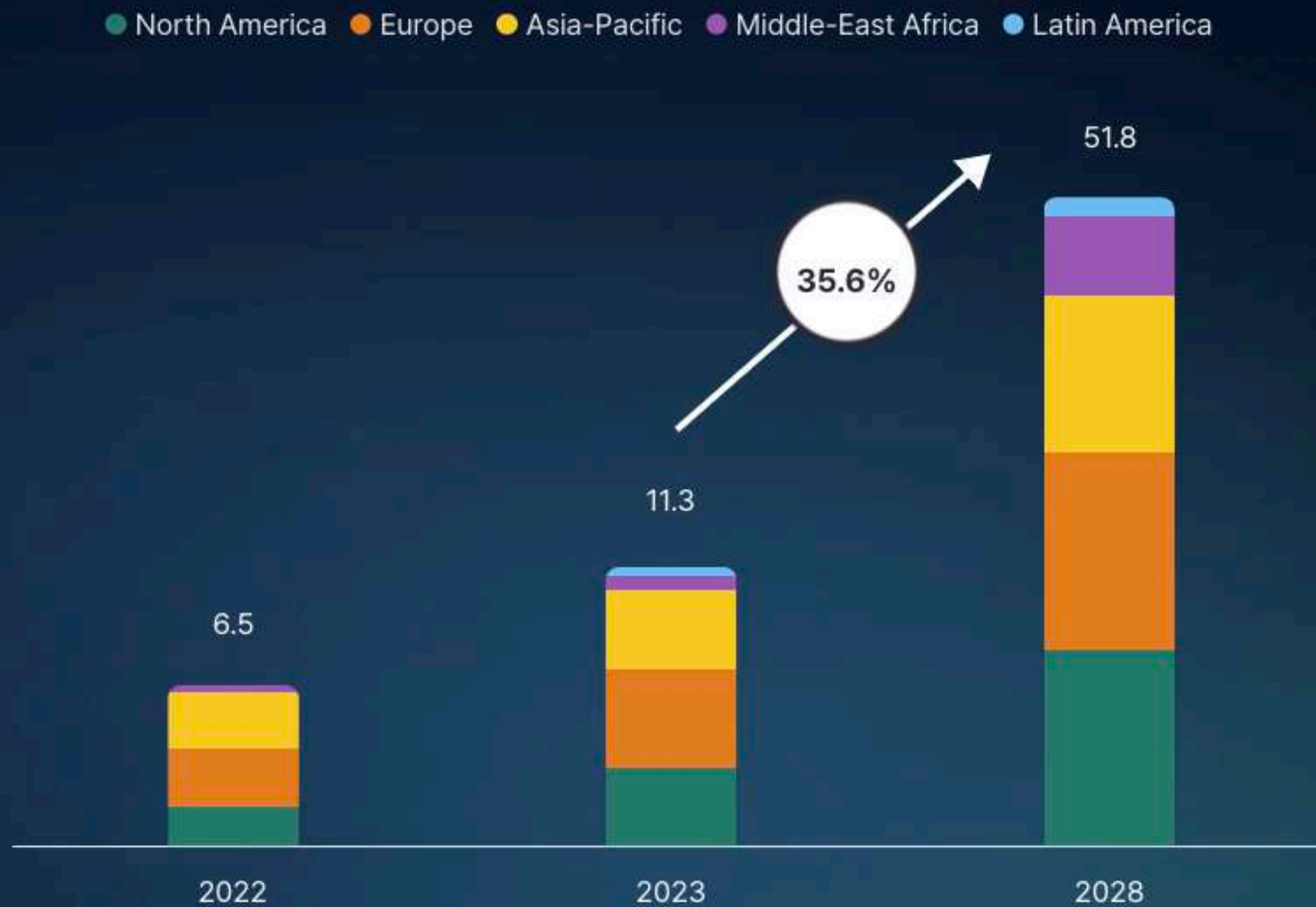
KB Investment No. 1 Commercial Bank

MIRAE ASSET Invested \$200M in SpaceX

springcamp Subsidiary of Leading Search Engine

The Era of AI: Transforming the World

Generative AI Market Global Forecast to 2028 (USD BN)

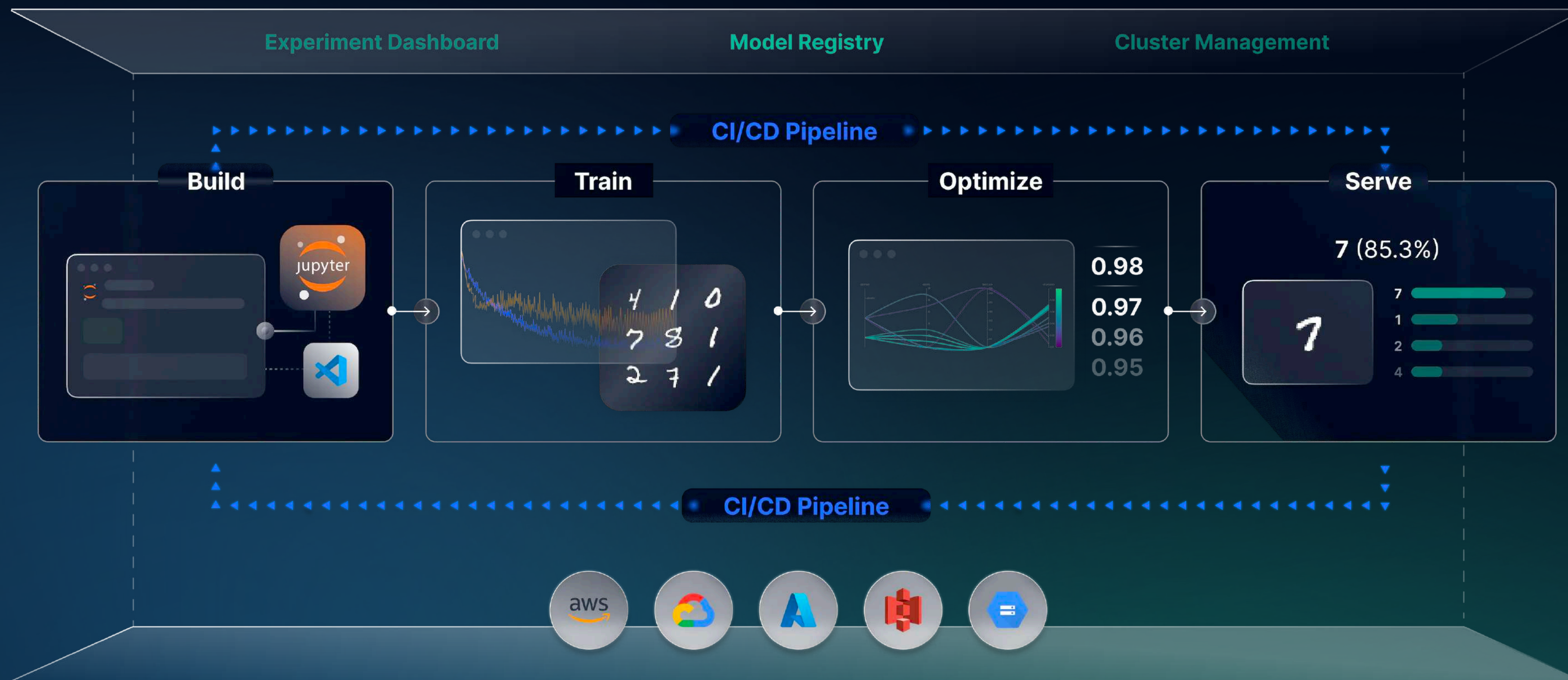


Percentage of respondents currently adopting or evaluating as least one use case



Problem

Running AI models in Production Needs Reliable, Scalable MLOps Infrastructure



Solution

VESSEL: Easiest Way to Run AI

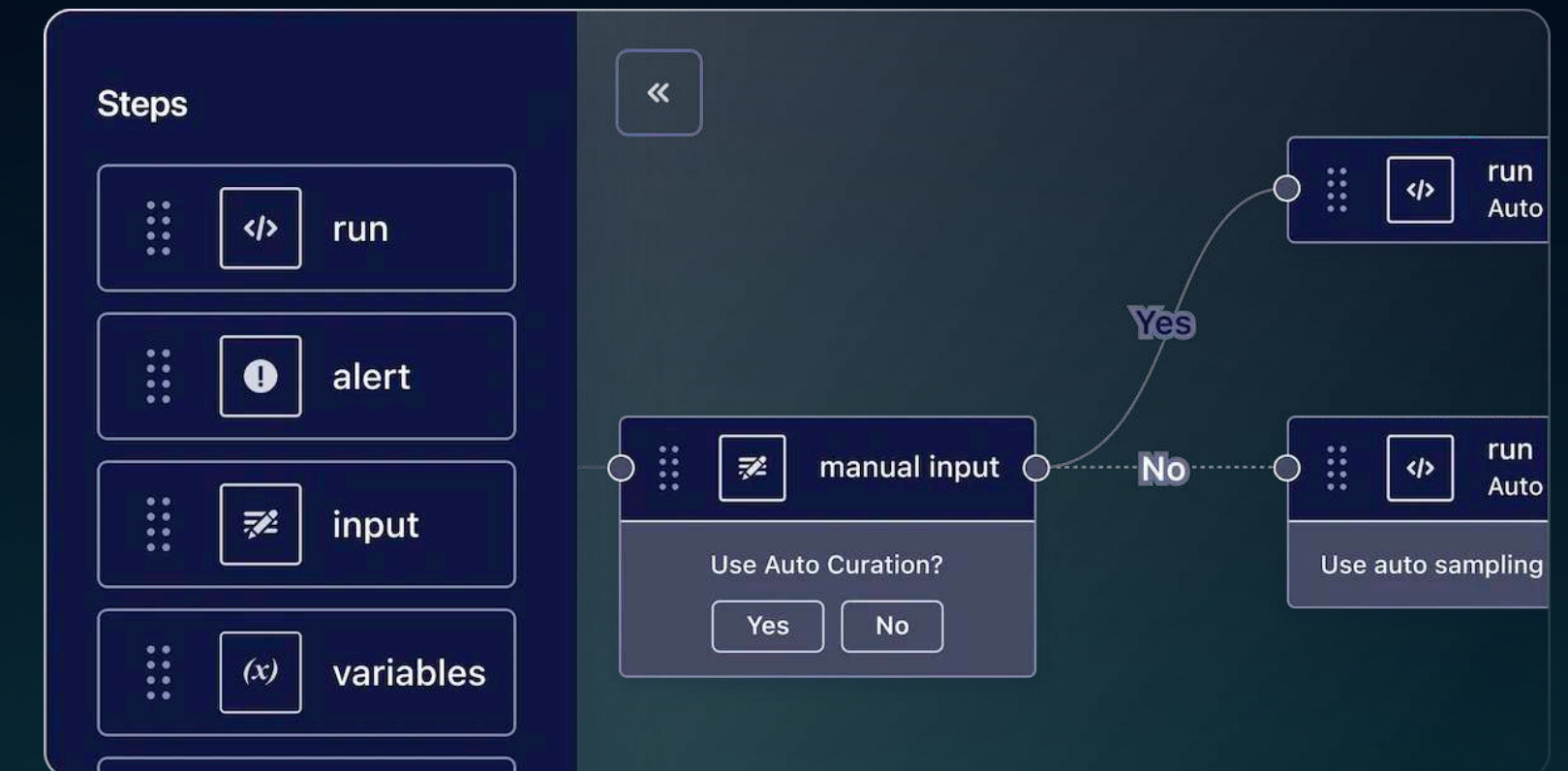
VESSEL is a unified interface for

- GPU computing
- Training AI models (including LLMs)
- Deploying AI models
- Simplifying enterprise ML development

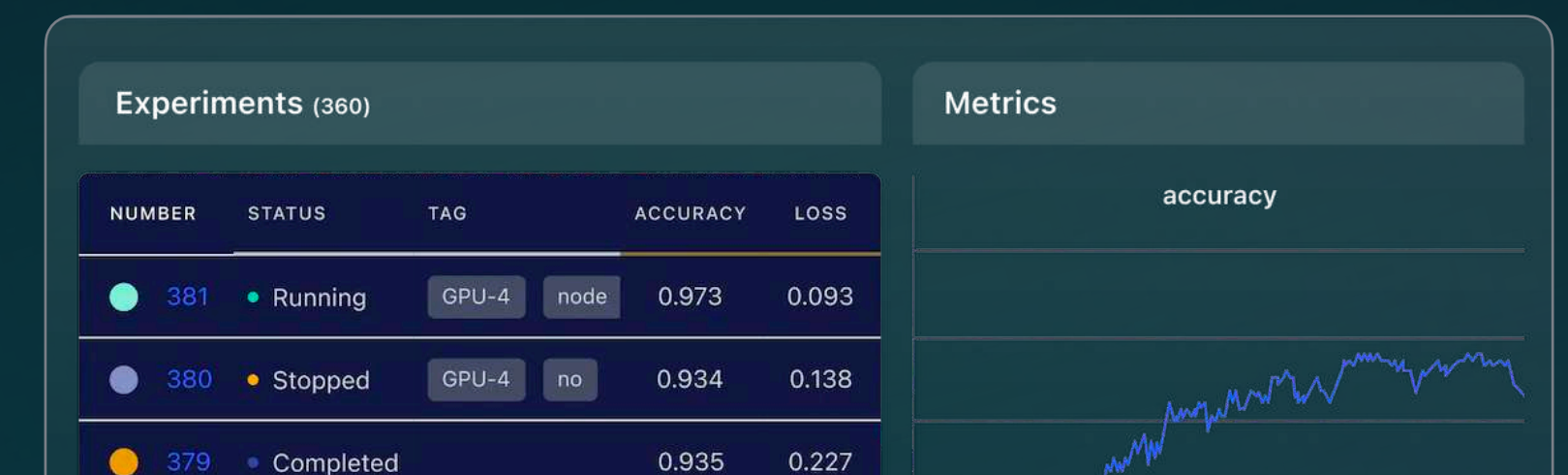
▶▶ VESSEL Run — Train and deploy in seconds

```
$ vssl run
  --cluster "aws" \
  --resource "t4.mem-163.spot" \
  --command "python main.py" \
  --dataset "s3://input:coco" \
  --learning_rate 0.001 --distributed
```

↻ VESSEL Pipelines — Automate workflows



📁 VESSEL Artifacts — Manage ML artifacts



Solution

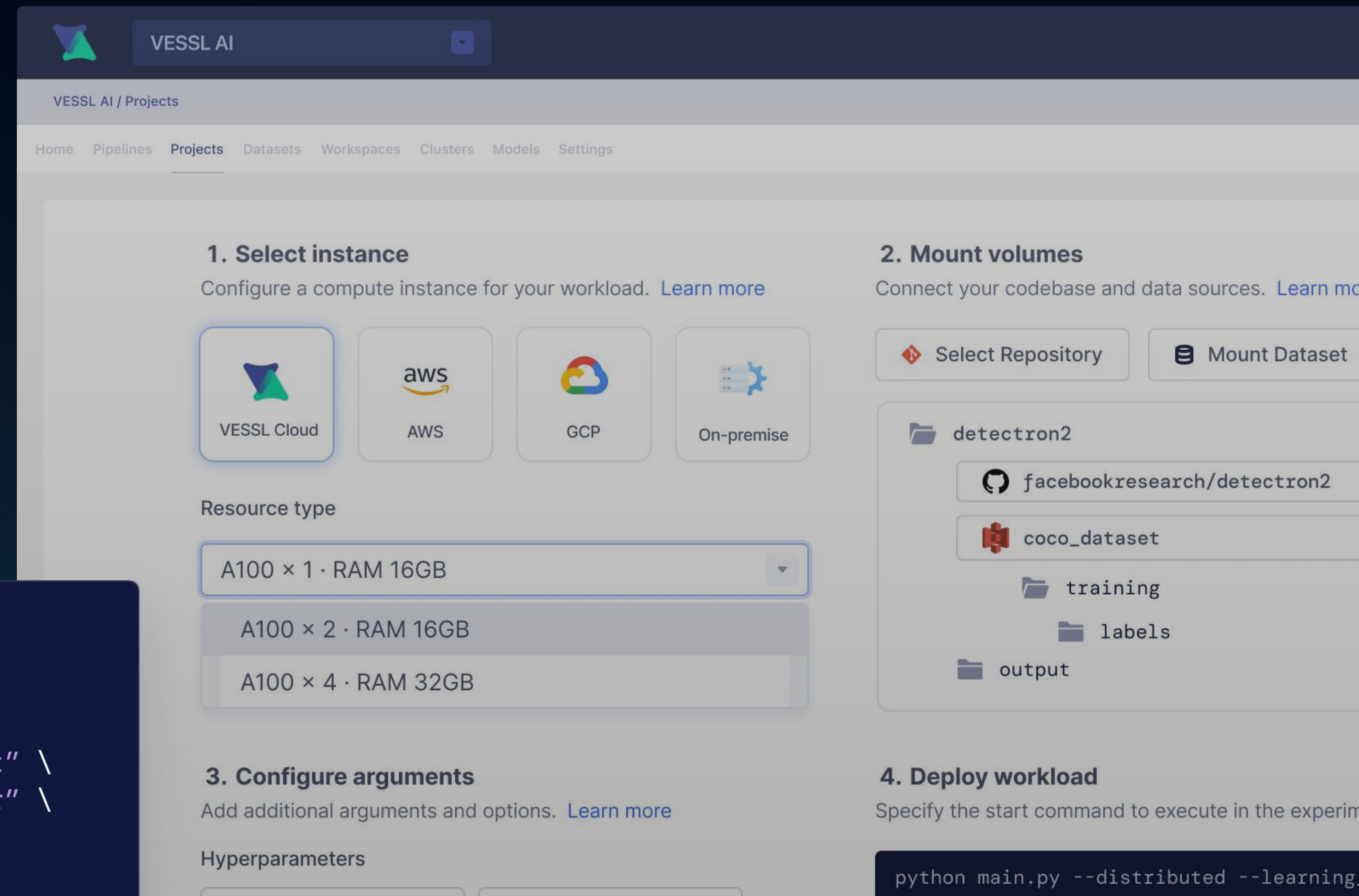
VESSEL Run

Run ML tasks in seconds at any scale, on any cloud

Use cases

- Fine-tune LLMs using 20+ GPUs
- Find cheap GPUs across multiple clouds
- Deploy AI models with Auto Scaling
- Coordinate 1000 GPUs for researchers

```
$ vssl run
  --cluster "aws-apne2"
  --resource "t4.mem-163.spot" \
  --dataset "/input:mnist-ocr" \
  --command "python main.py"
```



VESSL Run

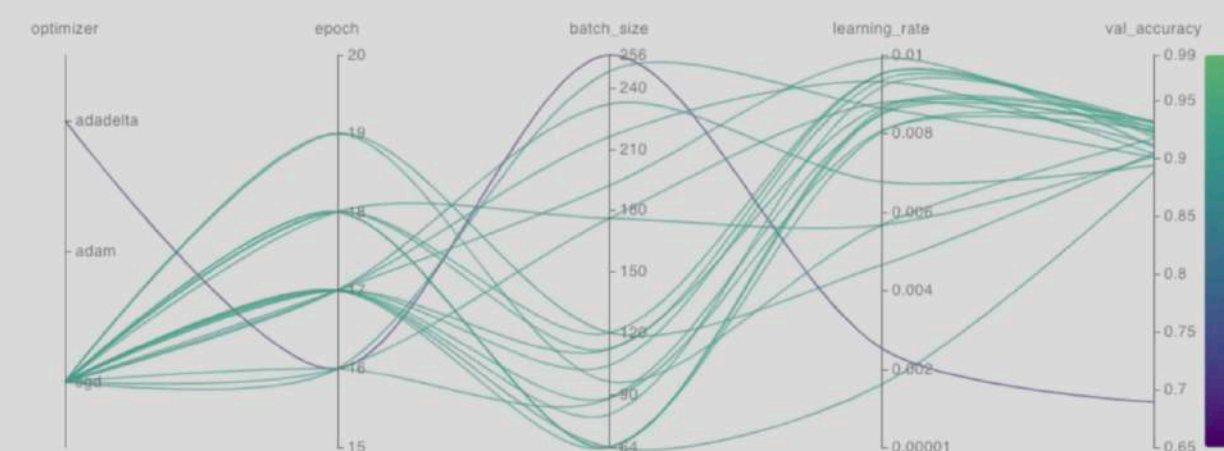
From LLM training to enterprise-grade model deployment

Notebook Servers

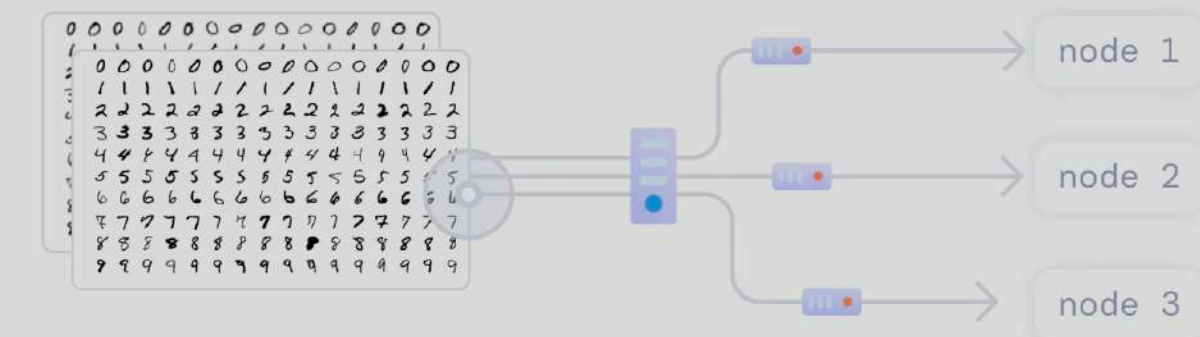


```
In [ ]: This is detectron2-coco  
  
In [ 6 ]: import numpy as np  
import vessl  
from detectron2 import
```

Hyperparameter Optimization



Distributed Training



Model Reproduction

v0.4	Completed
epoch	30
lr	0.044422
optimizer	adam
accuracy	0.969444
loss	0.1015583

Reproduce on the web

Reproduce experiment

Reproduce via CLI

```
vessl experiment create --
```

Local to Cloud

```
~ vessl run "python mnist/keras/main.p  
Organization: vessl-ai  
Project: ocr-mnist  
[?] Cluster: AWS-apnew  
> AWS-apne  
> On-prem-3090  
> VESSL-managed
```

T4 x 1 · 16GB RAM

K80 x 8 · 480GB RAM

V100 x 8 · 480GB RAM

Model Serving

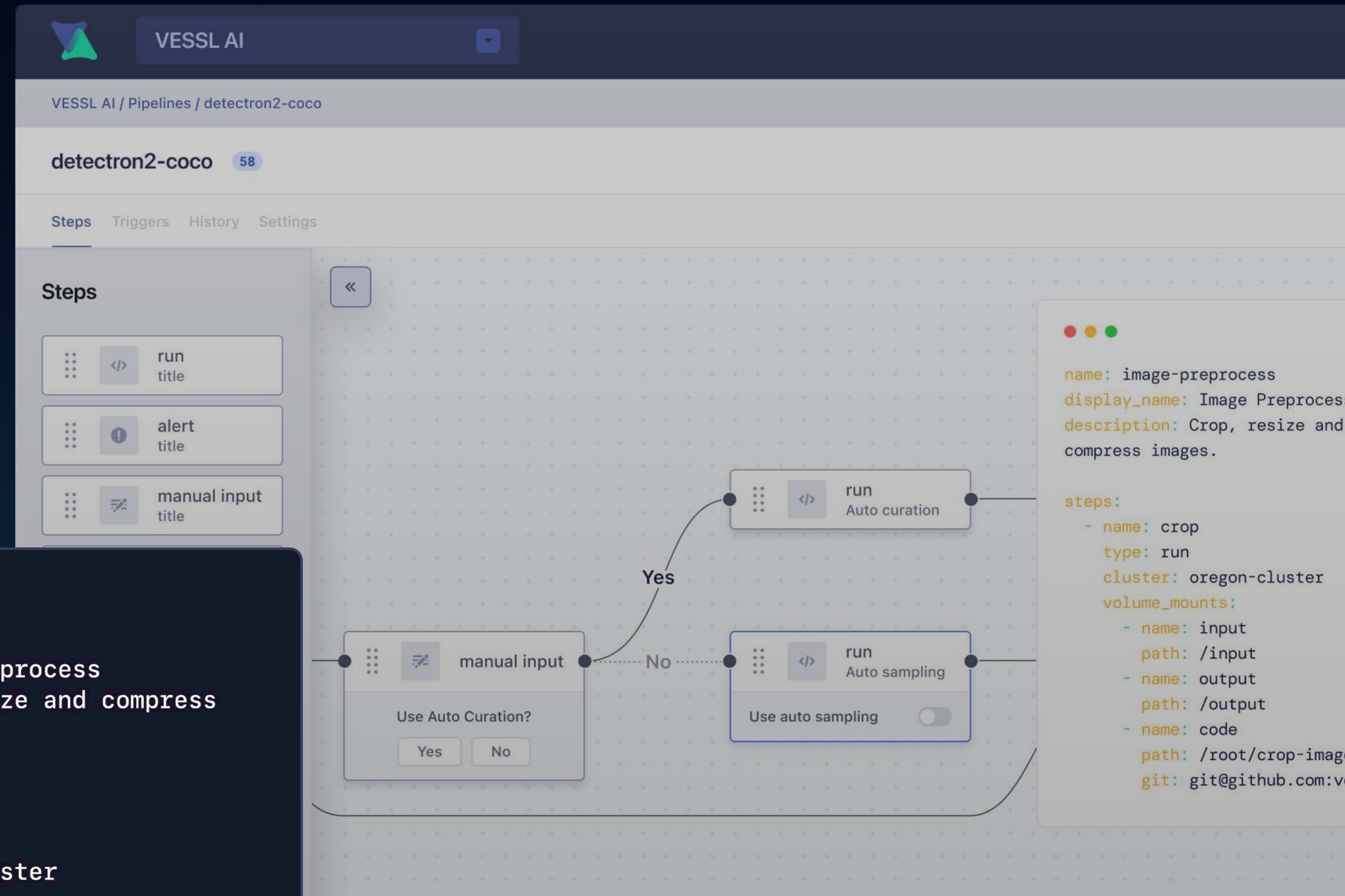
```
aws-detectron2-coco-base  
SERVICE STATUS Running  
AUTHENTICATION TOKEN *****  
SERVICE ENDPOINT https://service-f0r78b1f17ha-service-8000.  
CLUSTER aws-apne2-prod1
```

VESSEL Pipeline

Enterprise-level ML pipelines for data process, training and deployment.

Use cases

- Automated Data Pipeline
- Batch Inference for LLMs
- CI/CD Pipeline
- Schedule / Event Triggers



```
name: image-preprocess
display_name: Image Preprocess
description: Crop, resize and compress images.

steps:
- name: crop
  type: run
  cluster: oregon-cluster
```


Solution

VESSEL Artifact

Dashboard for datasets, models, and ML projects.

Central model registry

detectron2-coco-prod

Model	Status	Tags
v0.0.3	● Staging	epochs15 prod
v0.0.2	● Ready	avg_best epochs15
v0.0.1	● Ready	epochs10

Unified dataset repository

coco-test-s3

Source	Create
s3	1 month
s3://vessel-public-	

coco-test-s3

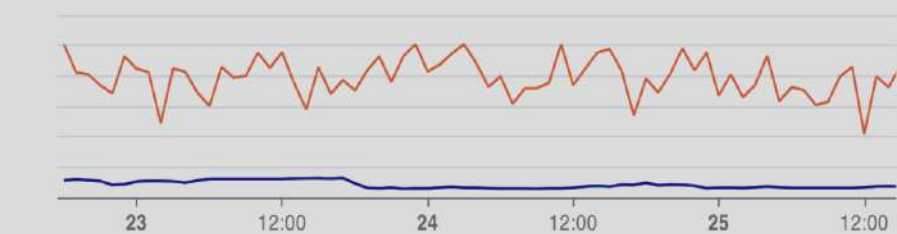
Branch	Run	Updated
○ main	1	3 hrs ago
○ af345ef	2	2 hrs ago
○ 06hsyh1	3	1 min ago

Comprehensive GPU monitoring

aws-apne

● Connected ● No issues

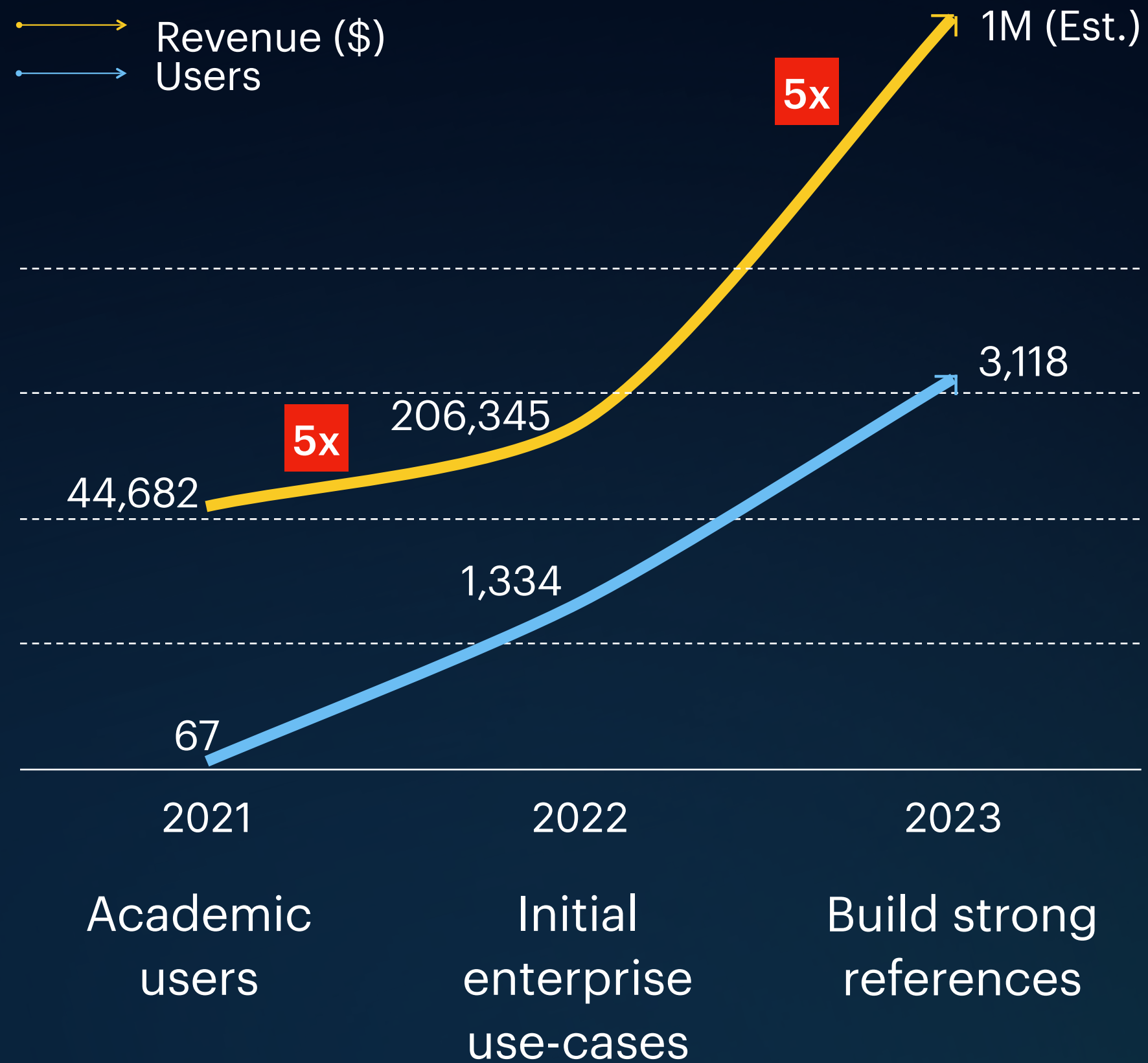
GPU



Traction

2023: Build strong references in top AI teams

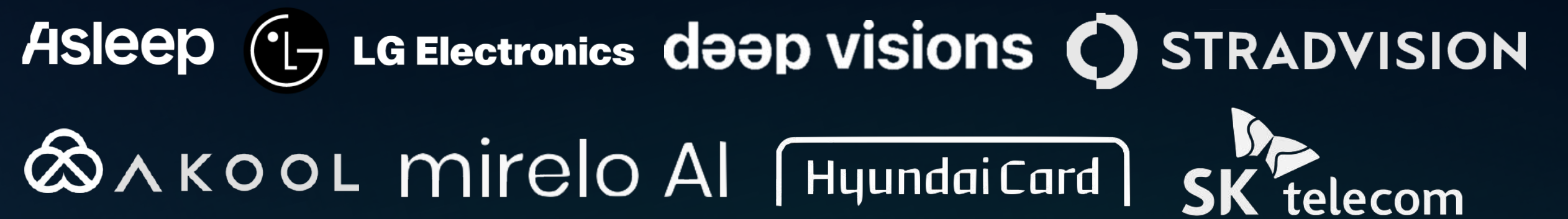
Numbers



Key Customers



Waitlist & PoC (200+ Companies)



Key Indices

Total churn-rate since first client **0**

From signup to onboard **84.67%**

2023: Get Excellent Customer Satisfaction

UPSELLING



Jungwoo Heo
Autonomous SW Development Team

“Integrates 100TB+ data from multiple regions, streamline data pipelines, and reduced model deployment time from 5 months to a week”



Wangjin Oh
Data Platform Group Lead

“We opted for VESSL over AWS Sagemaker”



Junseong Kim
ML Engineer

“One of the best tools for fine-tuning LLMs.”



Jiwoong Hyun
ML Engineer Lead

“First and the only mlops tools for testing and deploying our generative AI models.”



2024: Enabling Scalable Growth

	Qualified Leads (BY)		Deals (EY)	Business Activities	Revenue
2022	40		3	KR (Academy)	\$200K
2023	150		12	+ KR Direct Sales + US (Academy)	\$1M
2024	650	→ Series A	53 (Forecast)	+ US Direct Sales + KR Partner Sales	\$4M (Forecast)
2025	2500 (Forecast)		200 (Forecast)	+ APAC Direct Sales + US Partner Sales	\$15M (Forecast)
2026	6250 (Forecast)		500 (Forecast)		\$38M (Forecast)

2024: Building Unique GTM Engine in the US



AI Networking Event by VESSL AI (Quarterly)

Participants: 1500+ students & 500+ faculties from Stanford, Berkeley, Deepmind, Apple, ...



Partnerships w/ CSPs & Universities

Partnerships: Google, Oracle, NYU EFL, NYU Tandon, UT Austin (in-progress), ...



Engage w/ enterprises



Talks & sessions

Market

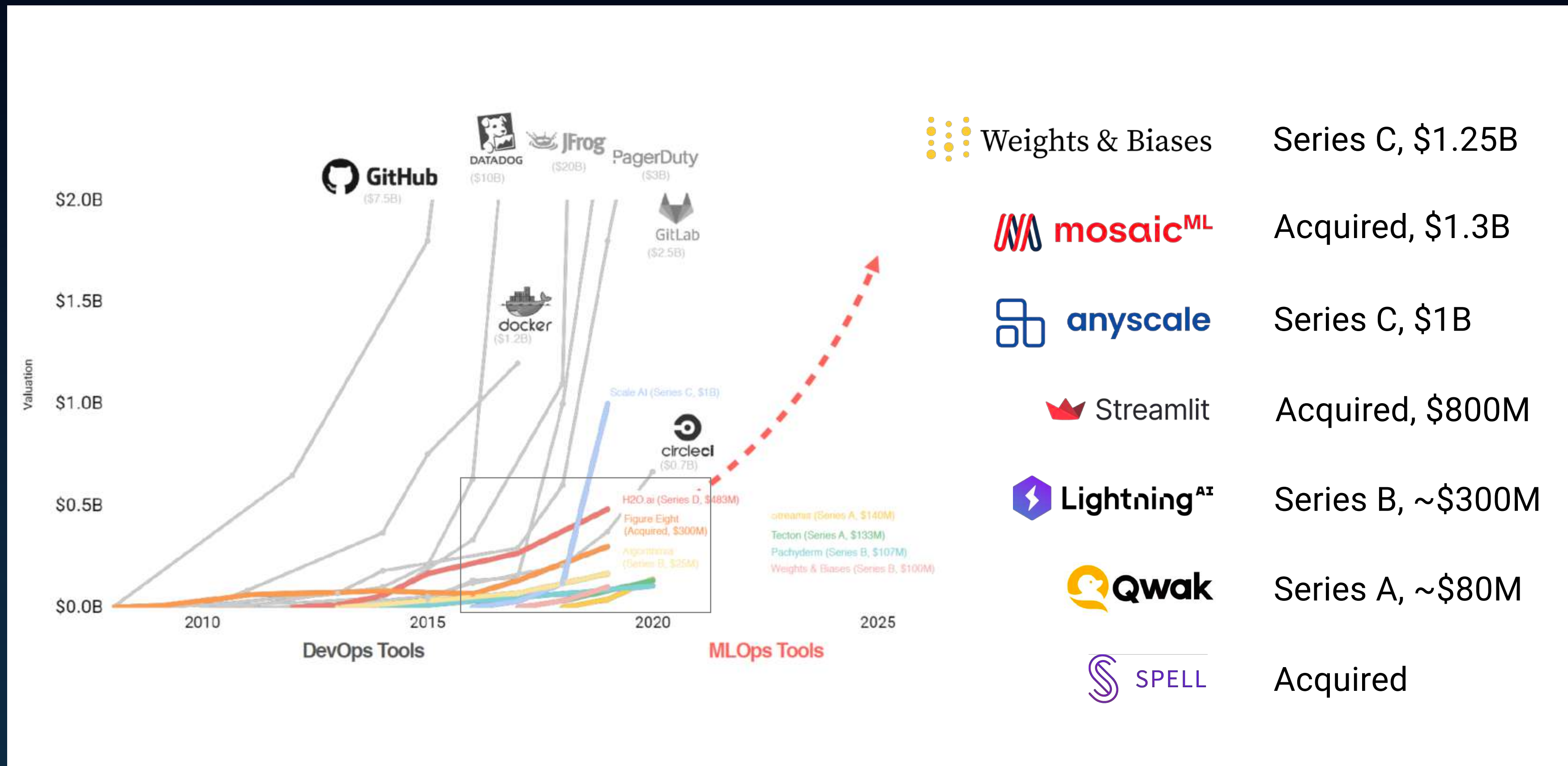
The next generation of DevOps

\$16.6bil

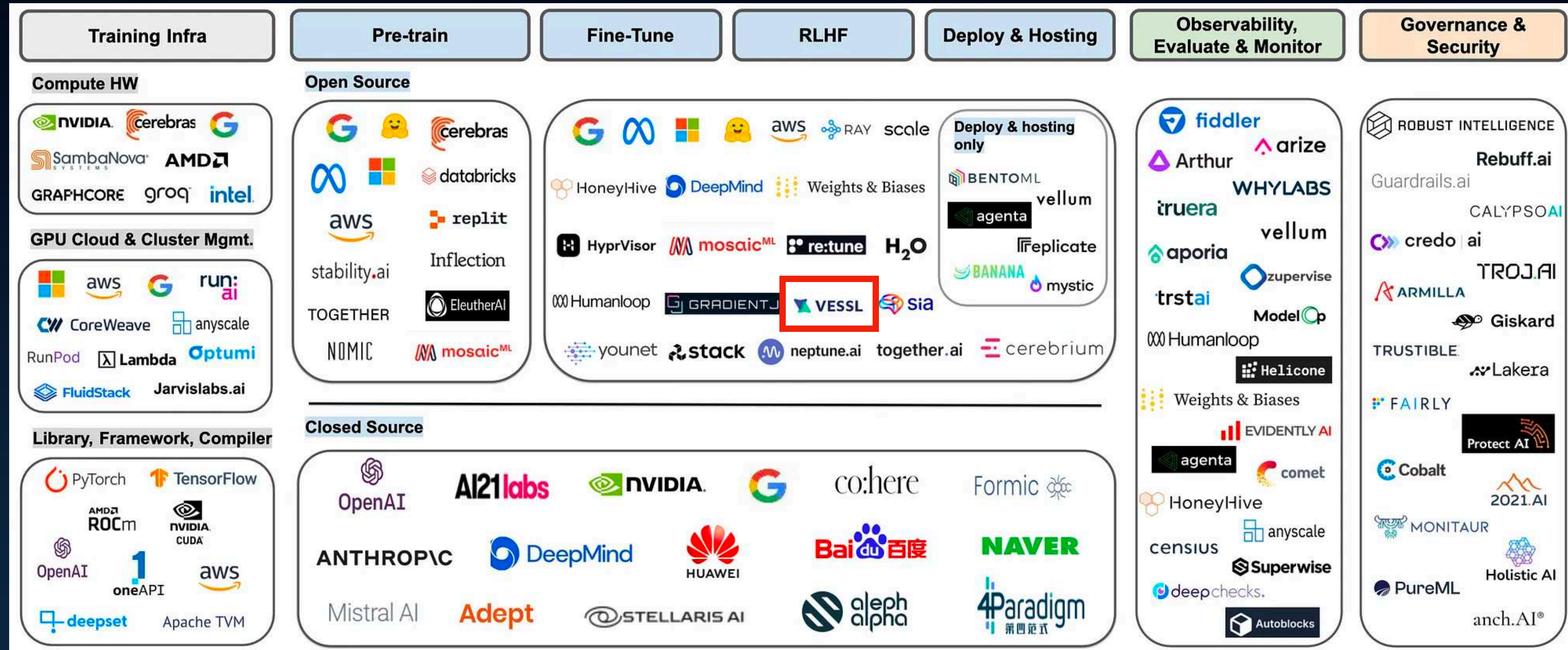
by 2030

39.7%

CAGR



Beyond MLOps: Pioneering the LLM & Generative AI Era



By Crystal Liu @ Bessemer Venture Partners

AI's \$200B Question

	Q2 2022 ACTUAL	Q2 2023 ACTUAL	Q4 2023 ESTIMATE
NVDA Data Center Run-Rate Revenue	\$15	\$41	\$50
AI Data Center Energy Cost	50%	50%	50%
Implied Data Center AI Spend	\$30	\$82	\$100
Software Margin	50%	50%	50%
AI Revenue Required for Payback	\$61	\$165	\$200

By Sequoia Capital



Unbundling the Public Clouds, Providing what AI companies need most

The next generation of DevOps

MLOps

Code / model warehouse

	DevOps
Name	GitHub
Business	Code warehouse
Founded in	2008
Growth	\$300M revenue, 56M+ users
Status	Acquired by Microsoft (\$7.5B, 2018)

	MLOps
Name	Huggingface
Business	Model warehouse
Founded in	2016
Growth	50M+ NLP model download
Status	Series D (\$4.6B, 2023)

Performance monitoring

	DevOps
Name	Datadog
Business	Monitoring & analytics for workflows
Founded in	2010
Growth	\$603M revenue
Status	NASDAQ : DDOG (\$28B, IPO in 2019)

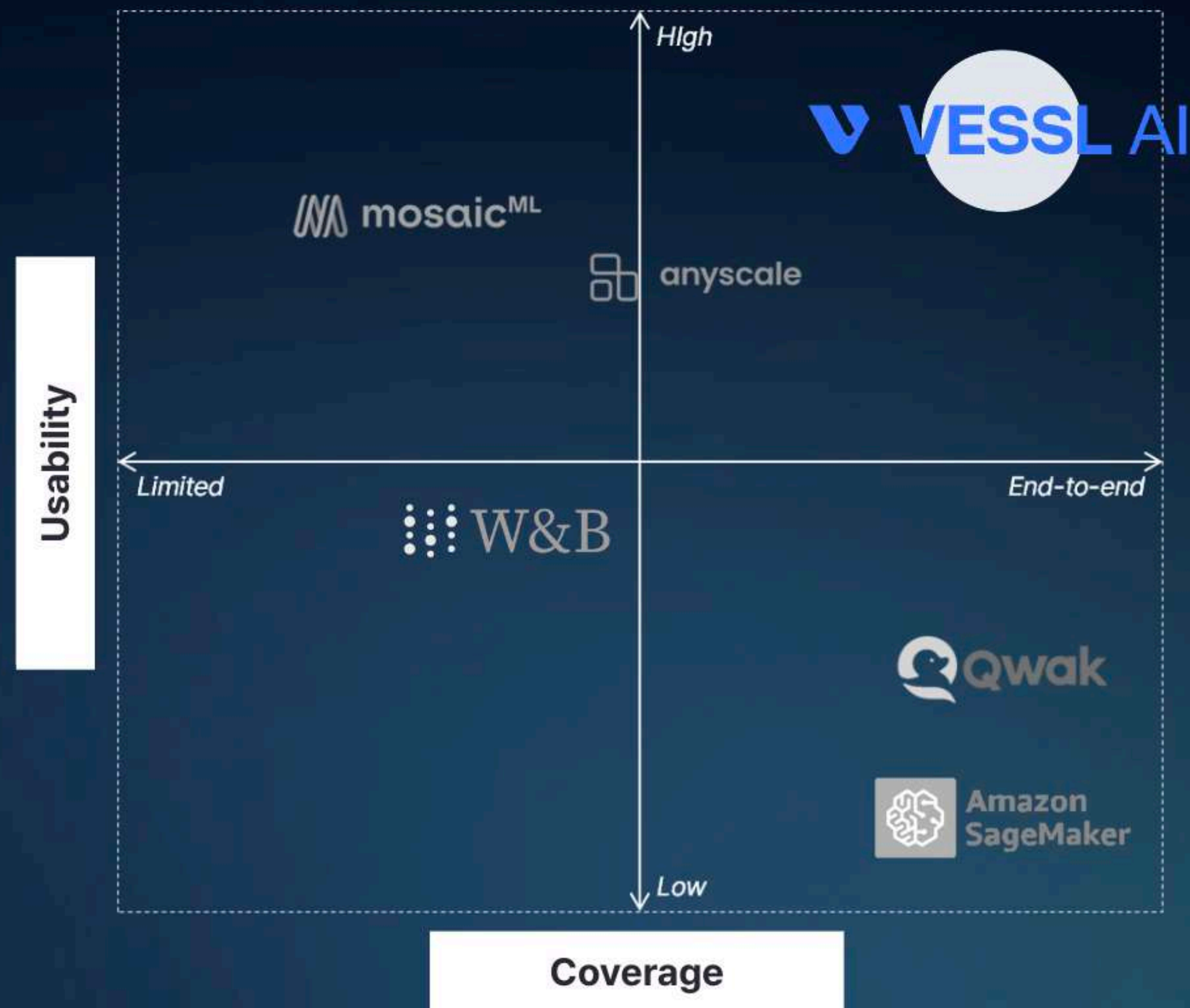
	MLOps
Name	Wandb
Business	Performance monitoring for ML works
Founded in	2017
Growth	\$10M+ revenue
Status	Series C (\$1.25B, 2021)

Engineering Platform

	DevOps
Name	Snowflake
Business	Data engineering platform
Founded in	2012
Growth	\$592M revenue
Status	NYSE : SNOW (\$73B, IPO in 2020)

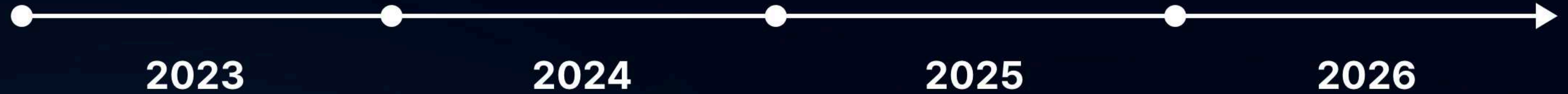
	MLOps
Name	VESSL
Business	ML engineering platform
Founded in	2020
Growth	-
Status	Pre A ~ Series A

VESSL: The only multi-cloud MLOps platform



	Managed computational resource	Multi cloud	No code-change	End-to-end	Automate Pipeline
VESSL AI	✓	✓	✓	✓	✓
Amazon SageMaker	✓			✓	✓
mosaic ^{ML}	✓	✓		✓	
W&B					
Qwak				✓	✓
anyscale	✓			✓	

No. 1 Interface for AI Computing within 7 years



Product

Run / Pipeline / Artifact
Multi-cloud

LLM Ops
PaaS (GPU as a Service)
Model Hub

NPU Cloud
No-code Gen AI
Enterprise Packaging

VESSEL Managed Cloud

Sales

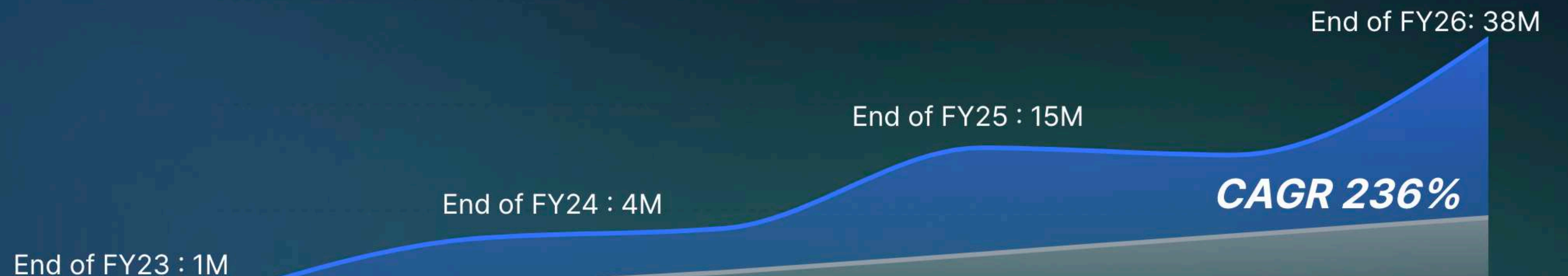
\$1M (Direct) @ KR
Make Key Customer
Enable Scalable Sales

US Sales Ops (Head of Sales)
\$1.5M (Direct) @ US (20 x \$75K)
\$1.5M (Direct) / \$1M (Partner) @ KR

\$6.5M (D) / \$2.5M (P) @ US
\$3.5M (D) / \$2.5M (P) @ KRAPAC
APAC(Japan) Sales Ops

\$16M (D) / \$6M (P) @ US
\$4M (D) / \$4M (P) @ KR
\$4M (D) / \$4M (P) @ APAC

Revenue(\$)



Best team for AI computing and SaaS sales



Jaeman An
CEO, Co-founder

KAIST, EE, CS, Math B.S.

VP of Engineering, Engineering Team Lead @ AITRICS, DEVSISTERS
"Developed 100M DAU mobile game, Recognized by AWS & NVIDIA"



Jay Chun
CTO, Co-founder

KAIST, CS B.S.

Engineering Lead, SRE @ Google, PUBG, Devsisters
"A pioneer of MLOps who built big data platform in PUBG and SRE in Google"



Sun Jeong
CISO

Chief Information Security Officer (CISO) @ Coinone, Uprise, KISA
"Cryptocurrency to legacy finance: A compliance expert adept in security."



Dami Yi
COO, Co-founder

Yonsei Univ, Business B.S.

BI, AE, CEO Staff, CFO Staff @ AITRICS, Sendbird, MeshKorea, and CVA
"Expert in scaling unicorn B2B SaaS startups with global expertise."

Becoming AI first — Productionize AI services in just weeks

<https://vessl.ai>

contact@vessl.ai

[YouTube](#), [LinkedIn](#), [Twitter](#)

[Docs](#), [Release Notes](#), [Blog](#)