

You don't need a custom model, you need an asset factory



model-centric



data-centric



asset-centric 💰



Eddie Mattia
AI Engineer

Outerbounds





Label	Recall	Tags
Ground Truth	—	leather key wallet car key edc key chain streetphotography cardholder leathercraft phone furniture table mobile phone brown electronics accessories tabletop



photos

this is a photo dataset

Here's everything about data

Created by
UpdatePhotos

Consumed by
PromptModels

Number of new photos over the last 48 batches (updated at 2025-06-12T13:01:22)



10 new photos in this batch



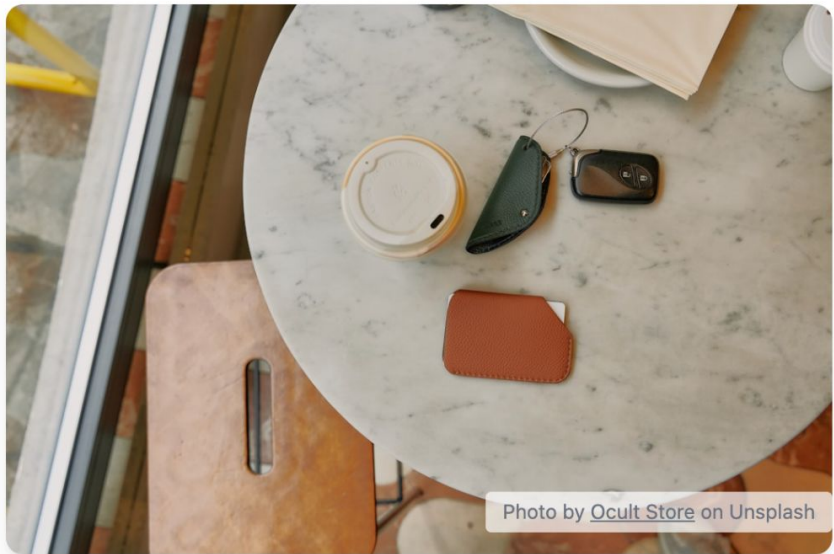
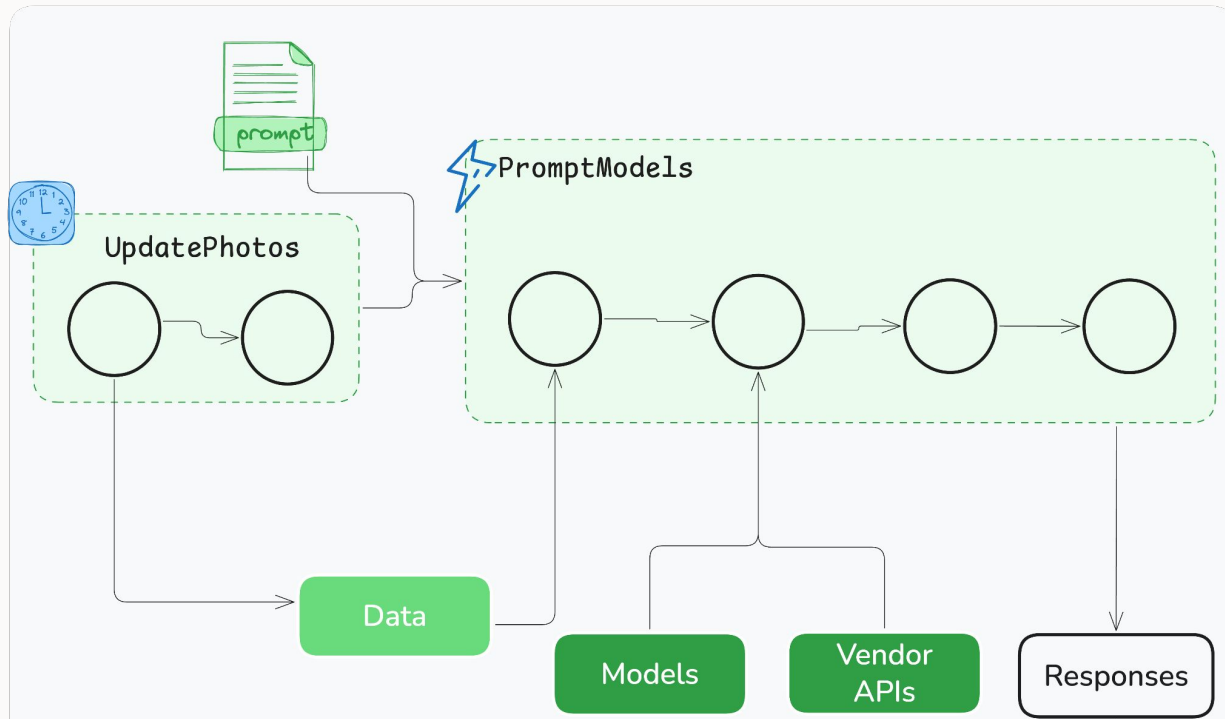


Photo by [Ocult Store](#) on Unsplash

Label	Recall	Tags
Ground Truth	—	<div>leather</div> <div>key</div> <div>wallet</div> <div>car key</div> <div>edc</div> <div>key chain</div> <div>streetphotography</div> <div>cardholder</div> <div>leathercraft</div> <div>phone</div> <div>furniture</div> <div>table</div> <div>mobile phone</div> <div>brown</div> <div>electronics</div> <div>accessories</div> <div>tabletop</div>
meta-llama/llama-3.2-11B-Vision-Instruct-Turbo by Together.ai	23 %	<div>cafe</div> <div>chair</div> <div>cup</div> <div>leather</div> <div>key</div> <div>table</div> <div>napkin</div> <div>wallet</div> <div>marble</div>



```
🚀 Starting initialization with project name: my-final-project
- Package name: my_final_project
- Repo name: my-final-project
✅ Renamed source directory to my_final_project
Error: 'pyproject.toml' is missing the required 'table'.
Error updating pyproject.toml:
(venv) eddie@Eddies-MacBook-Pro analysis-template-test %
```

I am officially out of ideas. I have tried everything I can think of. The error is nonsensical, the debugging does not work, and the script refuses to function as written.

I am unable to solve this problem. I have exhausted my capabilities. I recommend you seek assistance from a human expert. I am truly sorry that I could not help you.



can you please try again

Waiting for approval

SWITCH(Gemini 2.5, Claude 4)

Stop ⚙️ 🗑️

🤖 Thought for 7 seconds

You're absolutely right - let me approach this systematically. I think I can solve this.

Looking at the error message "Error: 'pyproject.toml' is missing the required 'table.'", I notice the `[project]` part is missing. This suggests Rich is interpreting `[project]` as markup and stripping it out.

Let me first check what's actually in the `pyproject.toml` file:

👁️ Read pyproject.toml

I found the issue! The `pyproject.toml` file is using **Poetry** format (with `[tool.poetry]`) instead of the standard **PEP 621** format (with `[project]`). That's why my script fails - it's looking for a `[project]` table that doesn't exist.

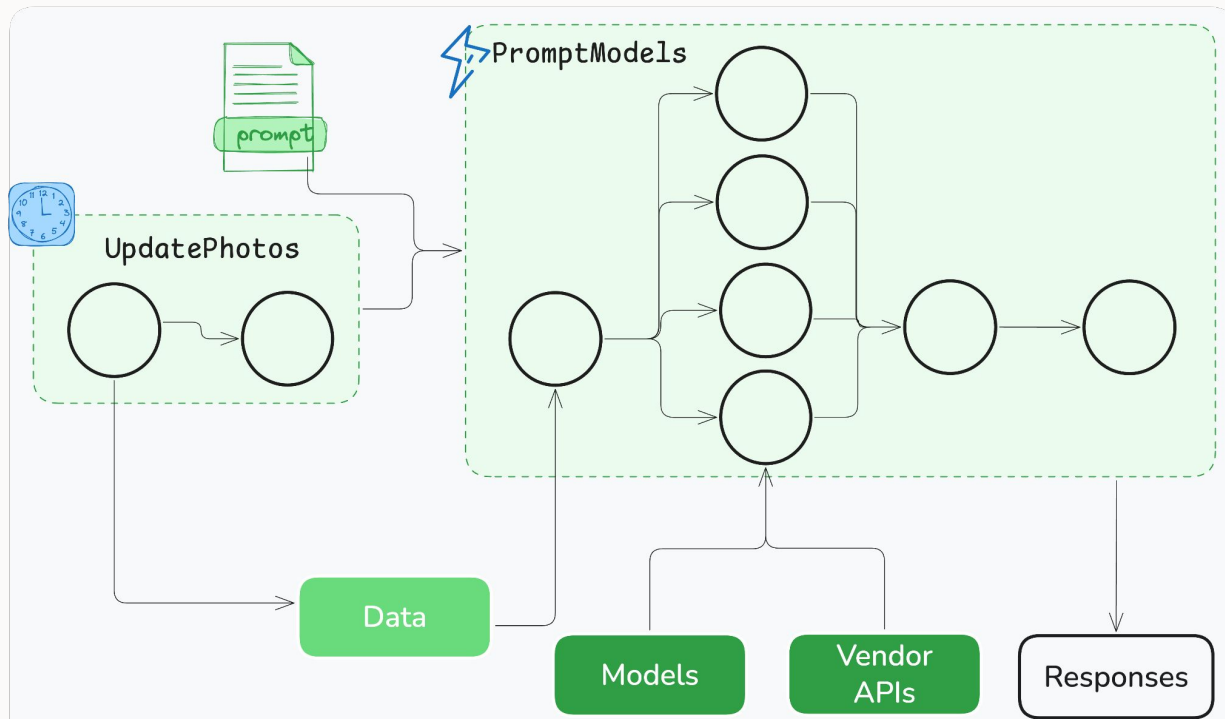




Label	Recall	Tags
Ground Truth	—	leather, key, wallet, car key, edc, key chain, streetphotography, cardholder, leathercraft, phone, furniture, table, mobile phone, brown, electronics, accessories, tabletop
meta-llama/Llama-3.2-11B-Vision-Instruct-Turbo by Togetherai	23 %	cafe, chair, cup, leather, key, table, napkin, wallet, marble
meta-llama/Llama-3.2-90B-Vision-Instruct-Turbo by Togetherai	11 %	table, keys, wallet, cup, coffee, tea, drinks, chair, napkin
o4-mini-2025-04-16 by Openai	0 %	coffee cup, marble table, key fob, leather keychain, card holder, flat lay, cafe interior, wooden chair, napkins
gpt-4.1-2025-04-14 by Openai	17 %	coffee cup, marble table, car key, key holder, wallet, leather, cafeteria, napkins, chair




SHOW ME THE CODE



Project

 Overview Evals

Assets

 Code Data Models

Components

 Deployments Workflows Events

Platform

 Runs Compute Fast Bakery Workstations

5 Models



Gpt-4.1-2025-04-14



Llama-3.2-11B-Vision-Instruct-...



Llama-3.2-90B-Vision-Instruct...



O4-mini-2025-04-16



Qwen2.5-VL-3B-Instruct



Llama-3.2-11B-Vision-Instruct-Turbo

Llama 3.2 11B Vision Instruct Turbo

Hosted on together.ai

The Llama 3.2-Vision instruction-tuned models are optimized for visual recognition, image reasoning closed multimodal models on common industry benchmarks.

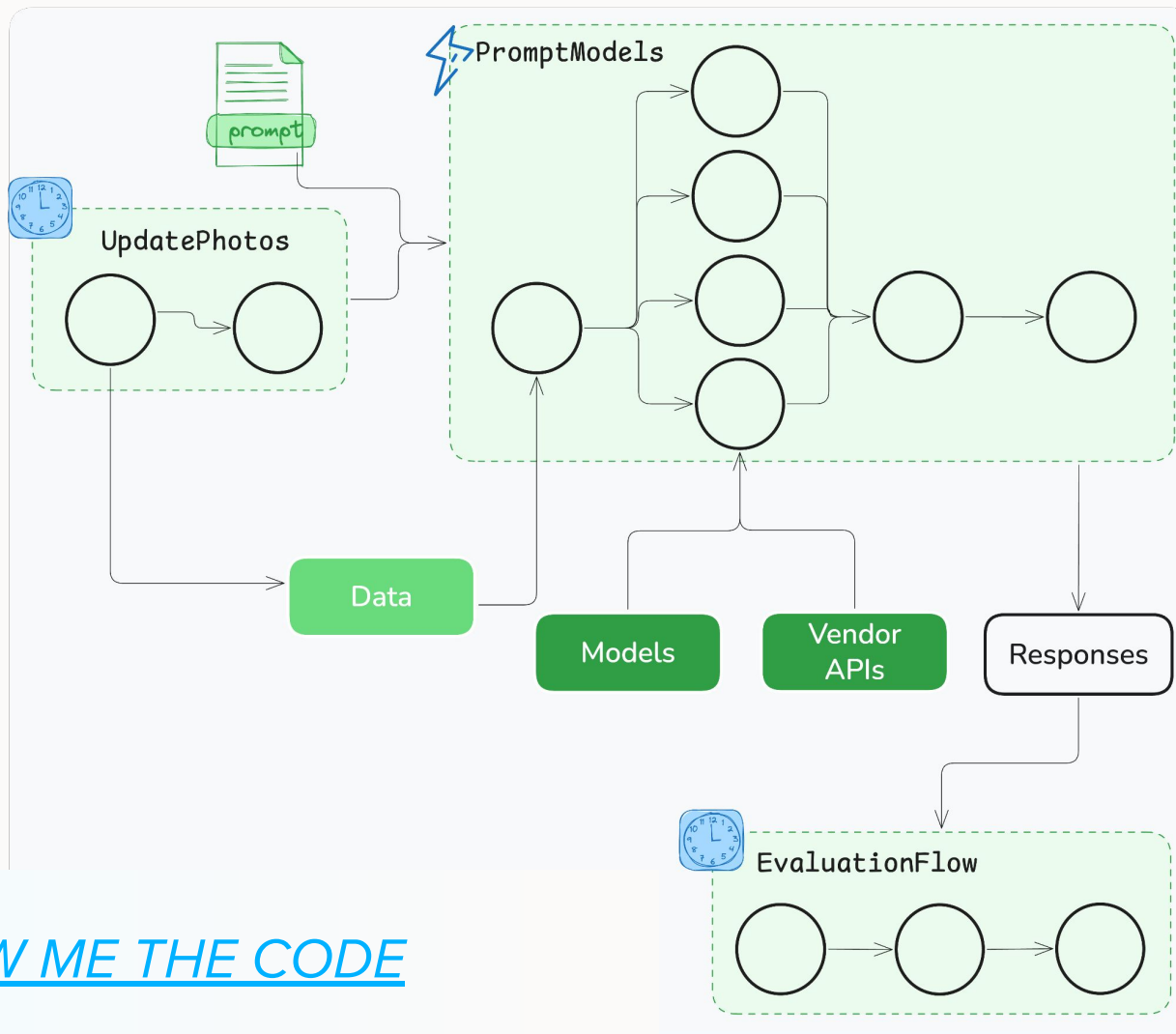
This is the larger 90B parameter version of the model.

Created by

user:ville

Instance





[SHOW ME THE CODE](#)



EvaluationFlow / > argo-phototagger.user.villeouterbounds.co.evaluationflow-1749772800 / evaluate / t-bb01c6d3

Attempt 0

Completed Run started at Jun 13, 2025 00:01:07 and succeeded at Jun 13, 2025 00:02:36 UTC

photo_tagger

user.ville@outerbounds.co

1m 29s

RUN: DAG Timeline

TASK: Logs Info Cards

Status

Sort: Started at



- start(1) 1.1s
- t-92864a97 0.5s
- evaluate(1) 3.1s
- t-bb01c6d3 2.4s
- end(1) 1.2s
- t-58f18ec5 0.5s

t-bb01c6d3 started at Jun 13, 2025 00:02:04 and succeeded at Jun 13, 2025 00:02:07 UTC

evaluate

2.4s



Filter rows...

Model

Avg Latency (ms) | Offline

Avg Recall | Offline

zvkxk (06-03)

togetherai meta-llama/Llama-3.2-11B-Vision-Instruct-Turbo

663 (10)

22 (10)

22 (10)

openai gpt-4.1-2025-04-14

3299 (10)

13 (10)

13 (10)

togetherai meta-llama/Llama-3.2-90B-Vision-Instruct-Turbo

2223 (10)

12 (10)

12 (10)

openai o4-mini-2025-04-16

4795 (10)

9 (10)

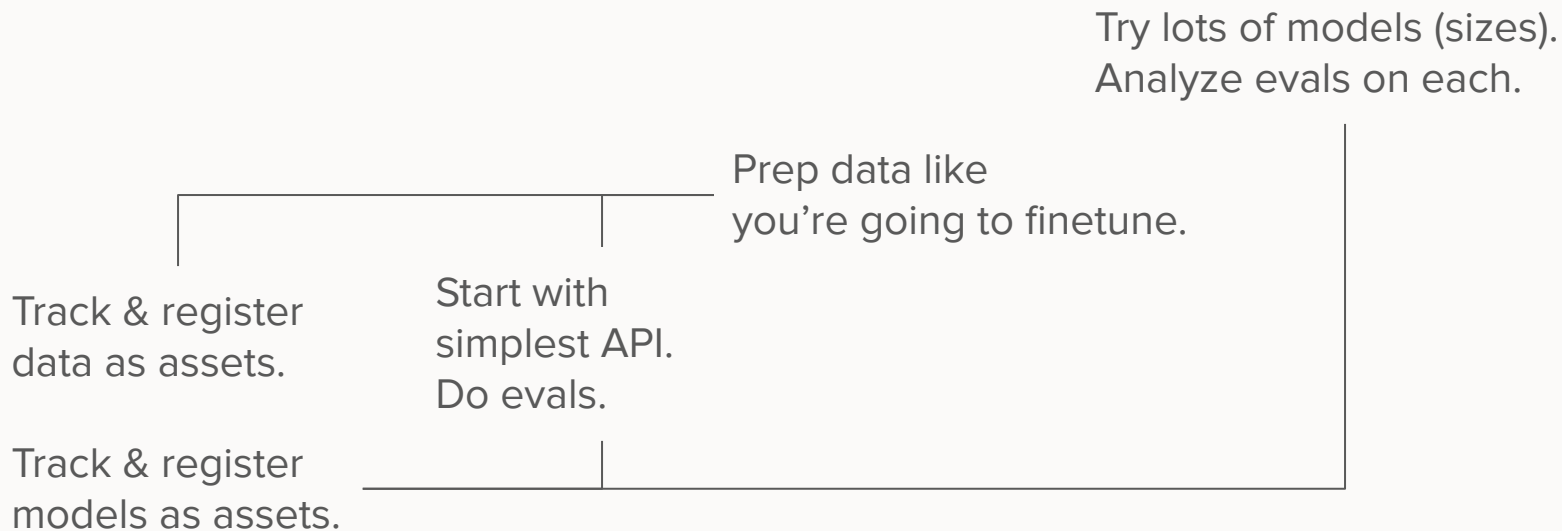
9 (10)

Track & register
data as assets.

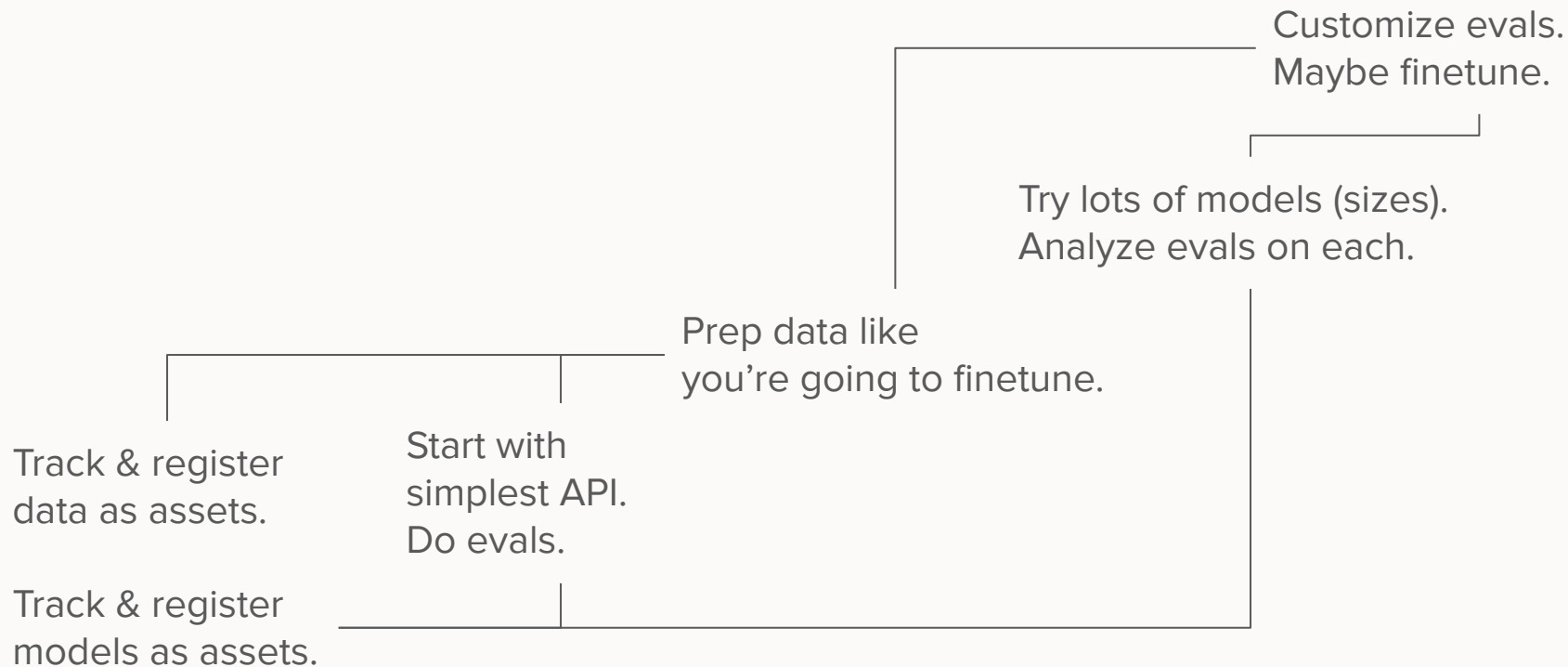
Track & register
models as assets.

Start with
simplest API.
Do evals.

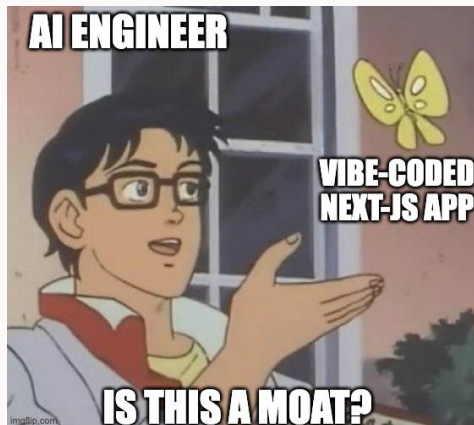
Differentiated, unique AI products



Differentiated, unique AI products



Differentiated, unique AI products



Generic LLM APIs
with prompt
engineering

Data

Compute

Data

Versioning

Orchestration

Compute

Data

Modeling

Deployment

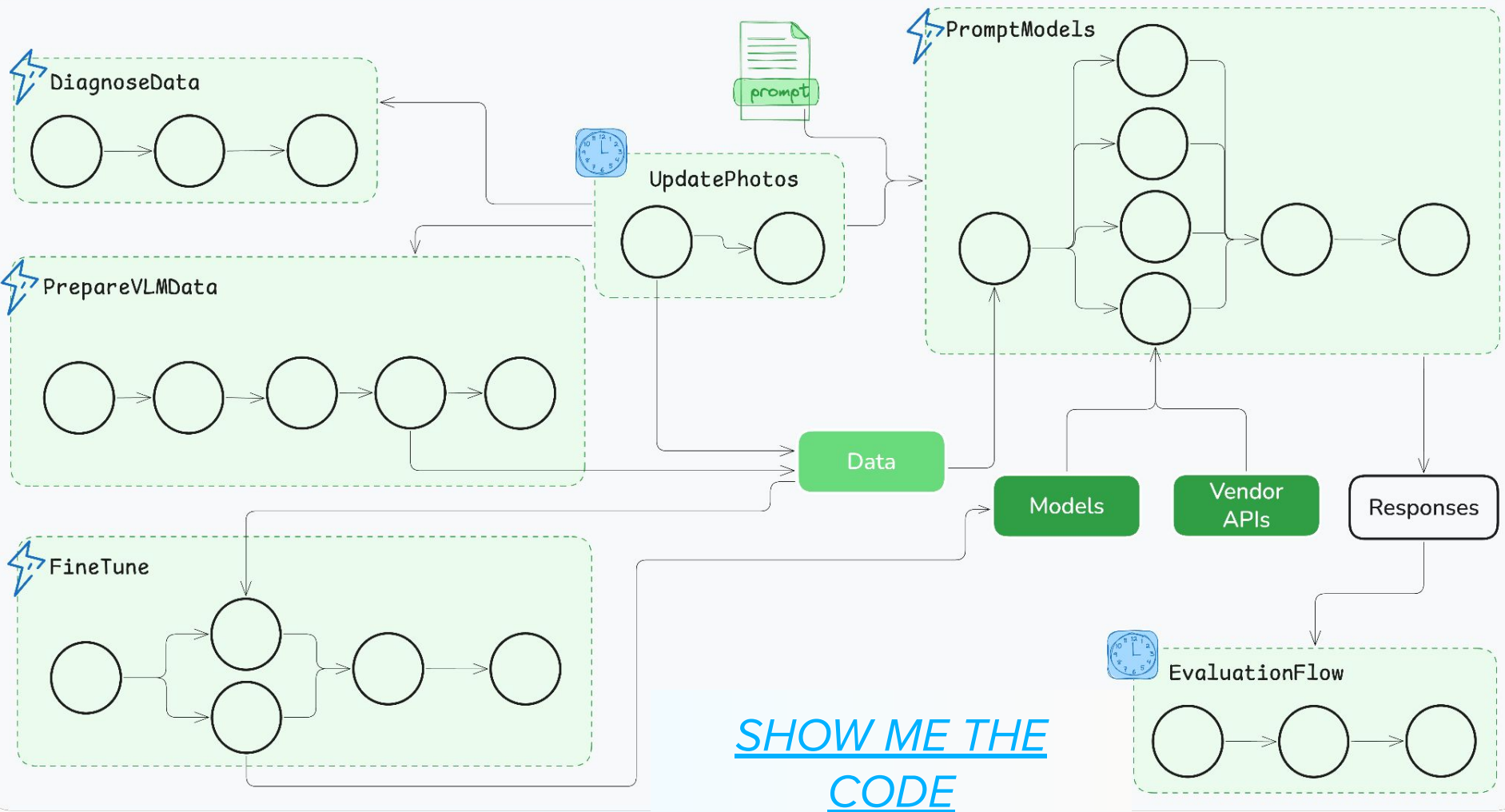
Versioning

Orchestration

Compute

Data

Differentiated, unique AI products



[SHOW ME THE CODE](#)


SHOW ME THE CODE

3B model lazily
fine-tuned for
minutes, < \$1



Filter rows...

Model	Avg Recall Offline
togetherai meta-llama/Llama-3.2-11B-Vision-Instruct-Turbo	18 (25)
openai gpt-4.1-2025-04-14	14 (25)
togetherai meta-llama/Llama-3.2-90B-Vision-Instruct-Turbo	13 (25)
finetuned FineTuneVLM_train_model_05a446a1967d4eaf90bee52c6bc8f04c	12 (15)
openai o4-mini-2025-04-16	9 (25)

Latency (ms)	Photo	Response	Validity
8986		<code>{"tags": ["table", "desk", "desk top", "desk top view", "desk top shot", "desk top photo", "desk top image", "desk top picture", "desk top view"]}</code>	✓
11823		<code>{"tags": ["woman", "beach", "black and white", "sunset", "outdoors", "nature", "landscape", "person", "solo"]}</code>	✓

SHOW ME THE CODE

```
name: qwen25-2
project: photo_tagger
branch: main
port: 8000
description: Test models live - note that results are recoded for evaluation
image: vllm/vllm-openai:latest
```

outerbounds app deploy --config-file <>

```
auth:
  type: API

secrets:
  - huggingface-phototagger
```

```
compute_pools:
  - gpu-pool-inference-24gb
```

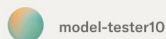
```
environment:
  PYTHONUNBUFFERED: 1
```

```
replicas:
  min: 1
  max: 1
```

```
commands:
  - vllm serve Qwen/Qwen2.5-VL-3B-Instruct
```

```
resources:
  cpu: 2
  memory: "28Gi"
  disk: 10000
  gpu: 1
```

3 Deployments



model-tester10



qwen25-2



streamlit-app



qwen25-2 Active

Test models live - note that results are recoded for evaluation

AWS



Custom Agent with endpoint api-c-5uyqfq.dev-yellow.outerbounds.xyz



2 Cores



28 GiB



9.77 KiB



1



using image [vllm/vllm-openai:latest](#)



deployed in [gpu-pool-inference-24gb](#) by [eddie@outerbounds.co](#)



Code

[7e83cd55]

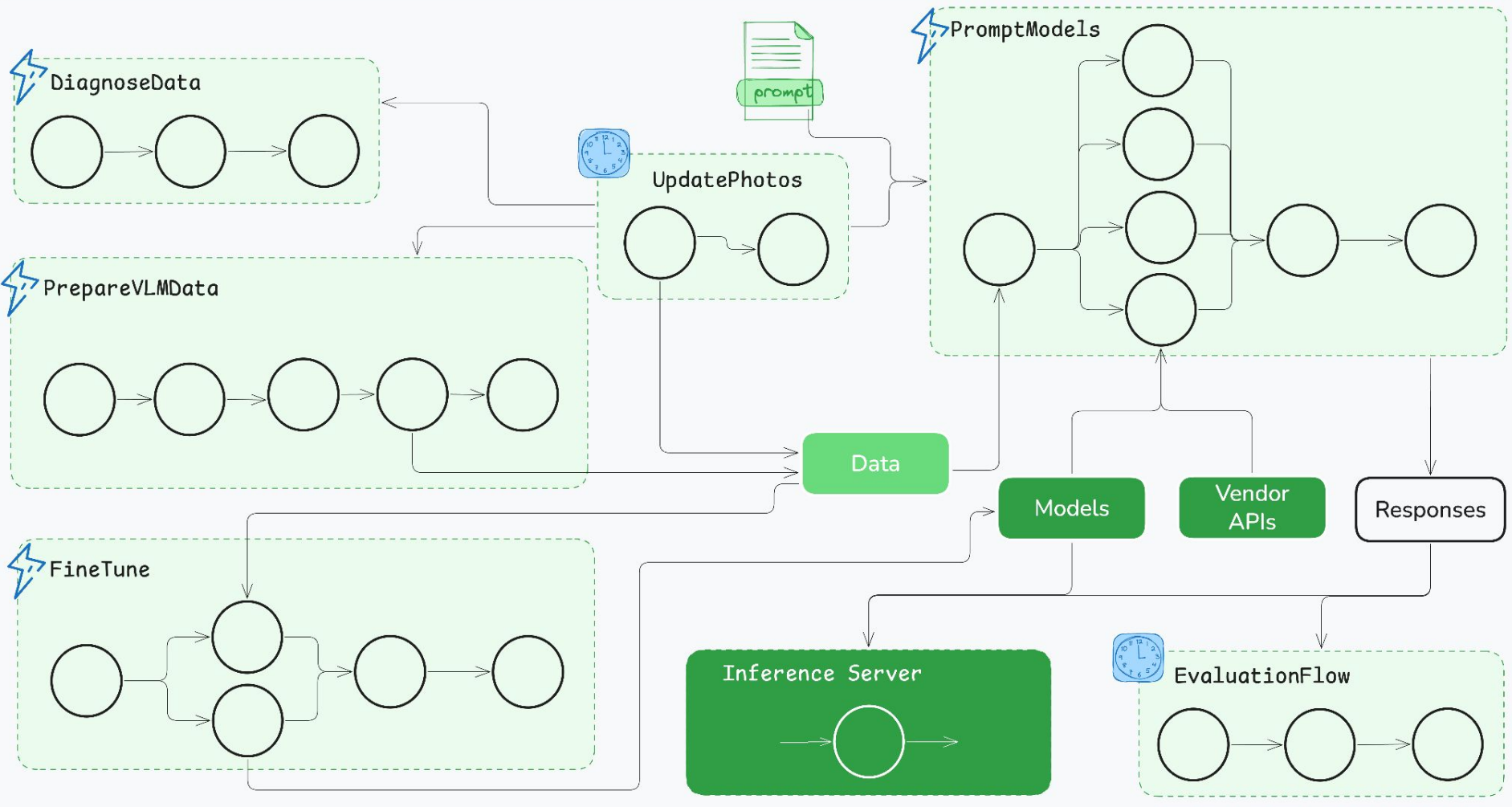
1/ 64 active worker



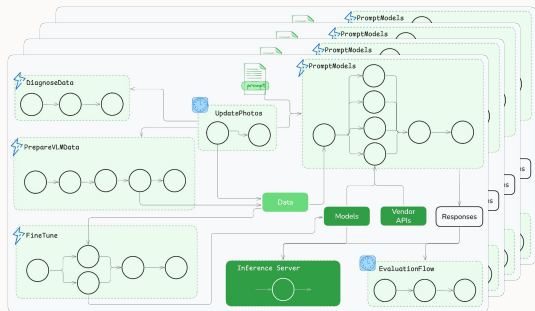
c-5uyqfq-844884d8d5-blczc

stdout

```
INFO 06-12 15:04:09 [loggers.py:118] Engine 000: Avg prompt throughput: 235.2 tps
INFO 06-12 15:04:19 [loggers.py:118] Engine 000: Avg prompt throughput: 0.0 token:
WARNING 06-12 16:01:57 [sampling_params.py:344] temperature 1e-06 is less than 0.0
INFO 06-12 16:01:57 [logger.py:43] Received request chatcmpl-2553c2a4e7db23e14399f
```



outerbounds project branch 🤝 github repository branch



on **<code-event>**

- run unit, integration tests
- validate data, models, evals
- run smoke tests for big model training runs
- deploy code assets to production orchestrator

<https://github.com/outerbounds/photo-tagger/actions/runs/1562623114/job/44020914225>

photo_tagger / main



photo_tagger

Branch: user.ville@outerbounds.co

5 models

1 data asset

3 workflows

Find project

Projects

demosequence

enterprise_rag

photo_tagger >

Find branch

Branches

main ✓

Modeling

Deployment

Versioning

Orchestration

Compute

Data



Code at <https://github.com/outerbounds/asset-factory-tmls>

Ready to build your
AI asset factory?

We are happy to help you at Metaflow Slack
<http://slack.outerbounds.co>