FROM RESEARCH TO PRODUCTION WITH PYTORCH

JEFF SMITH
ENGINEERING MANAGER
FACEBOOK AI

AGENDA

01
WHAT IS PYTORCH

02
RESEARCH TO PRODUCTION

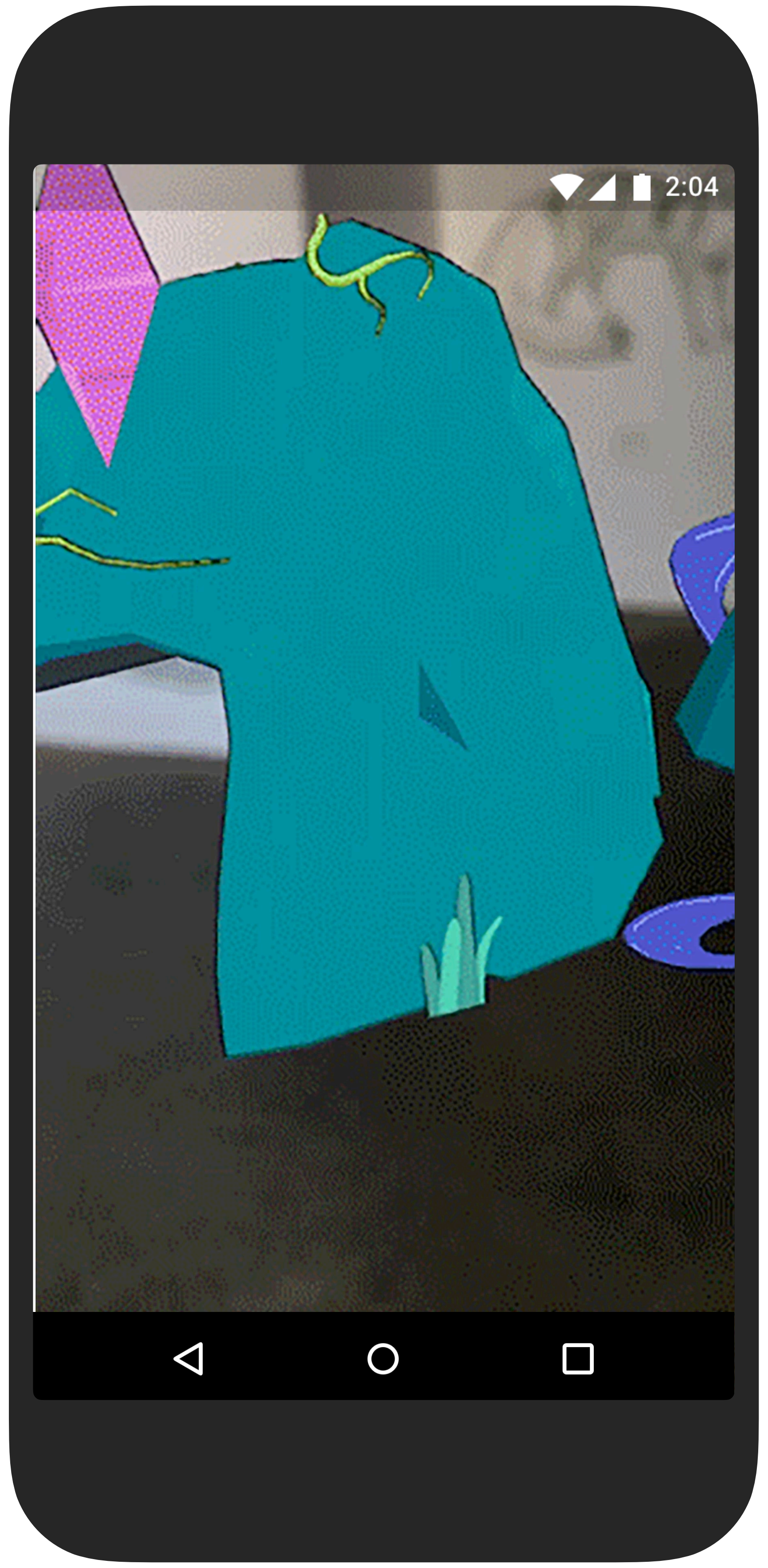
03
PRODUCTION PYTORCH

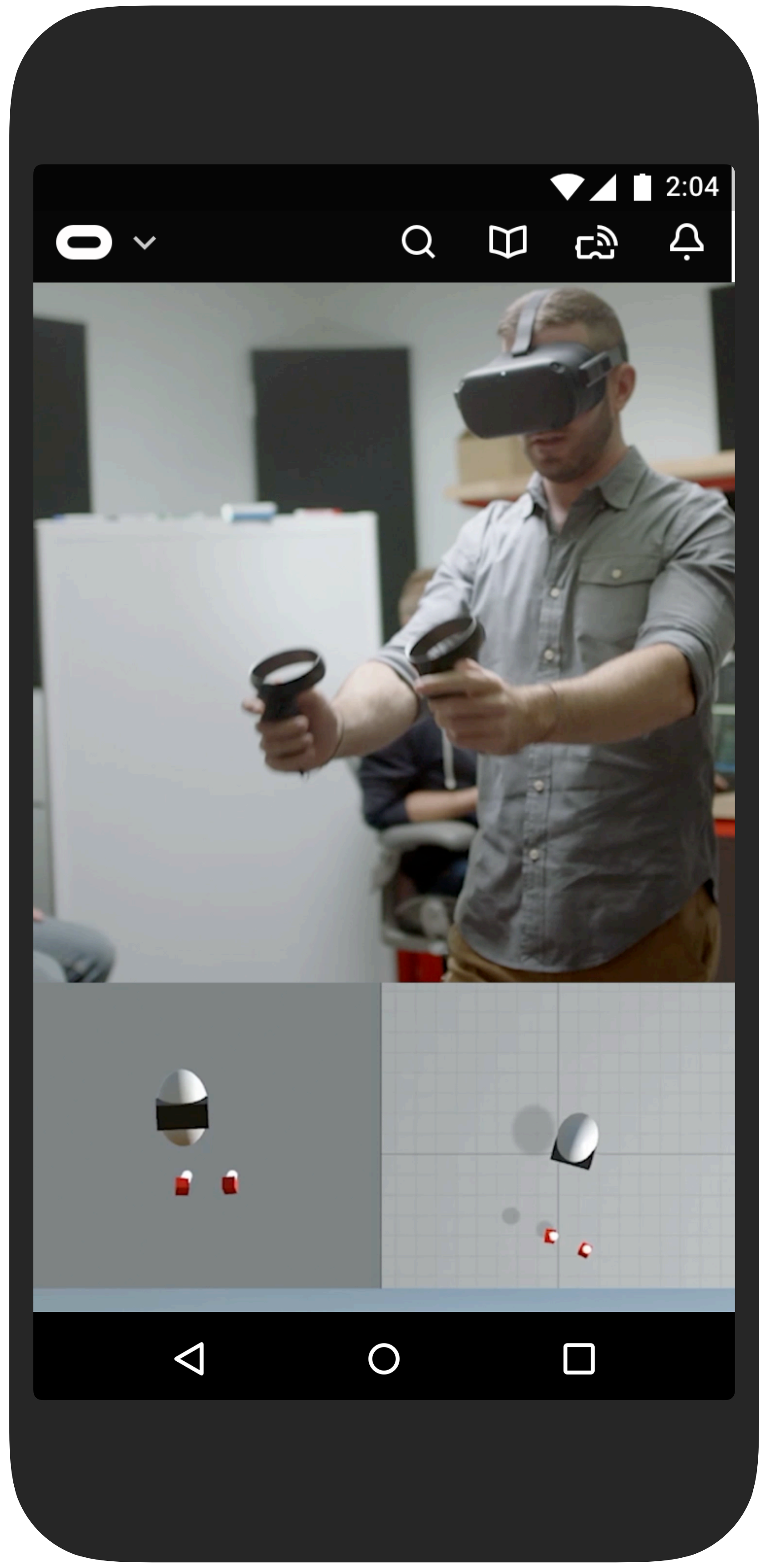
04
ECOSYSTEM

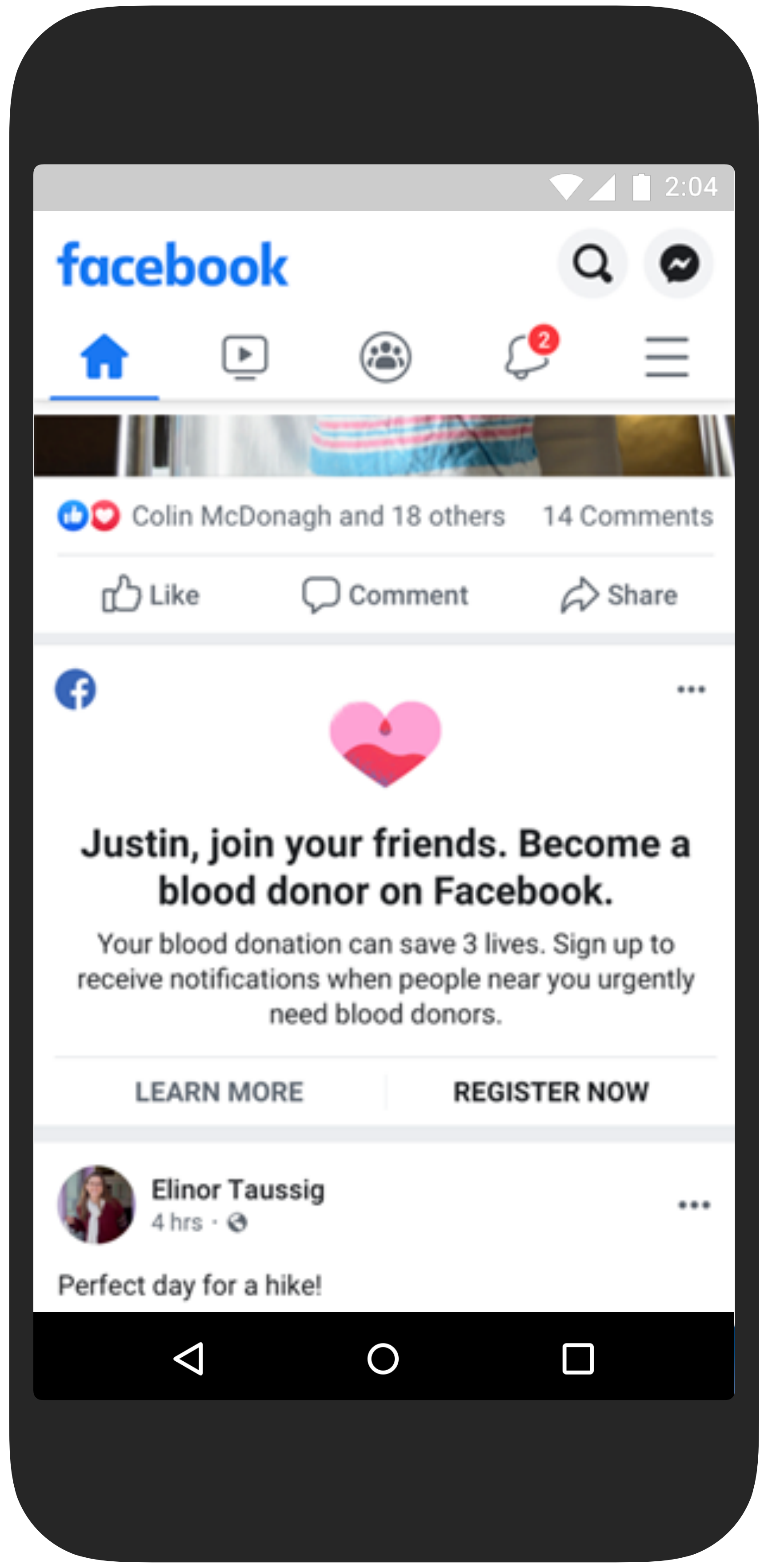
05
RESOURCES FOR DEVELOPERS

Sneakers (10.5) · $200
Marketplace

¡Hola! ¿Están esas zapatillas todavía en venta?





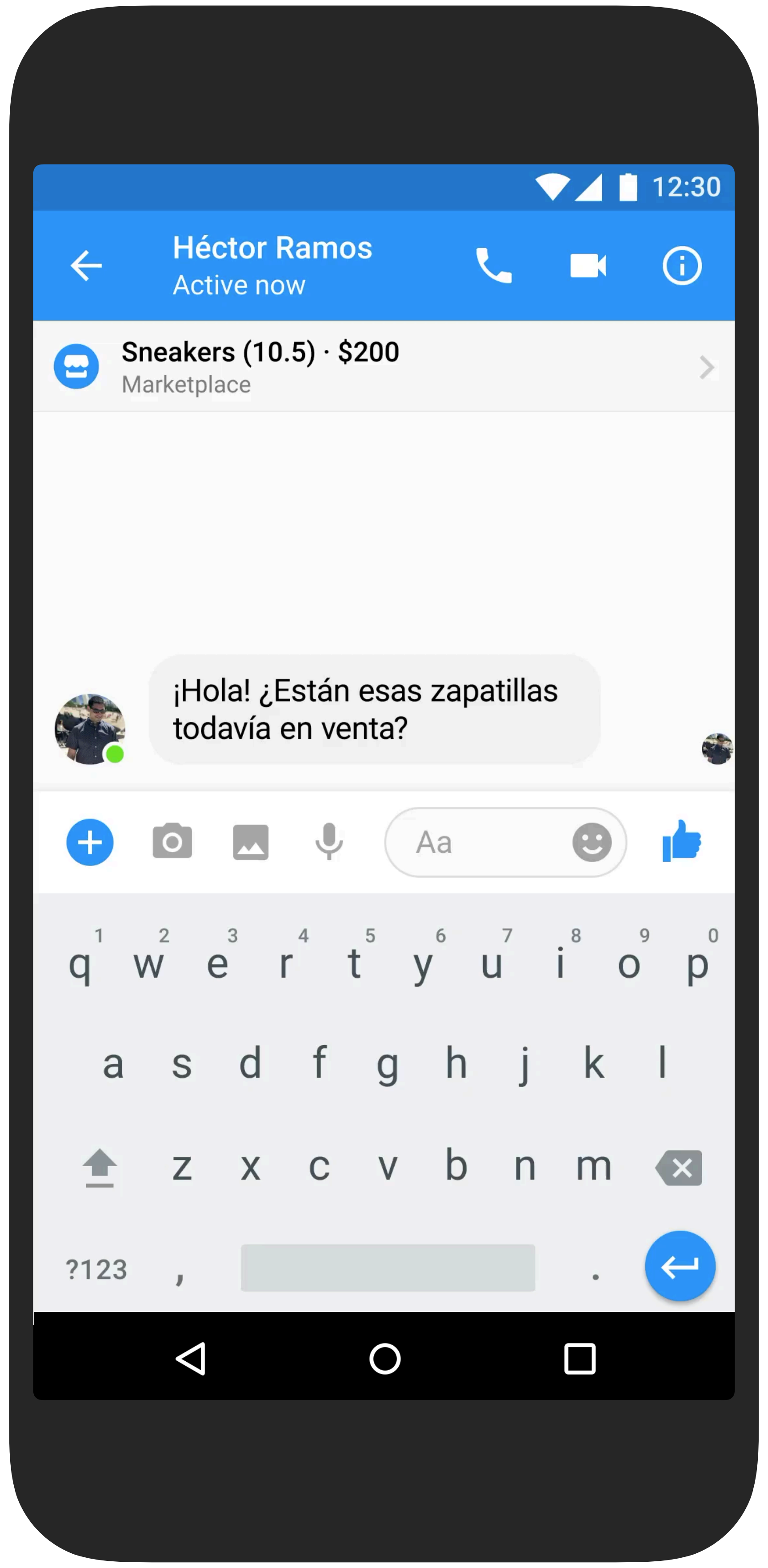


Justin, join your friends. Become a blood donor on Facebook.

Your blood donation can save 3 lives. Sign up to receive notifications when people near you urgently need blood donors.

Elinor Taussig
4 hrs ·

Perfect day for a hike!

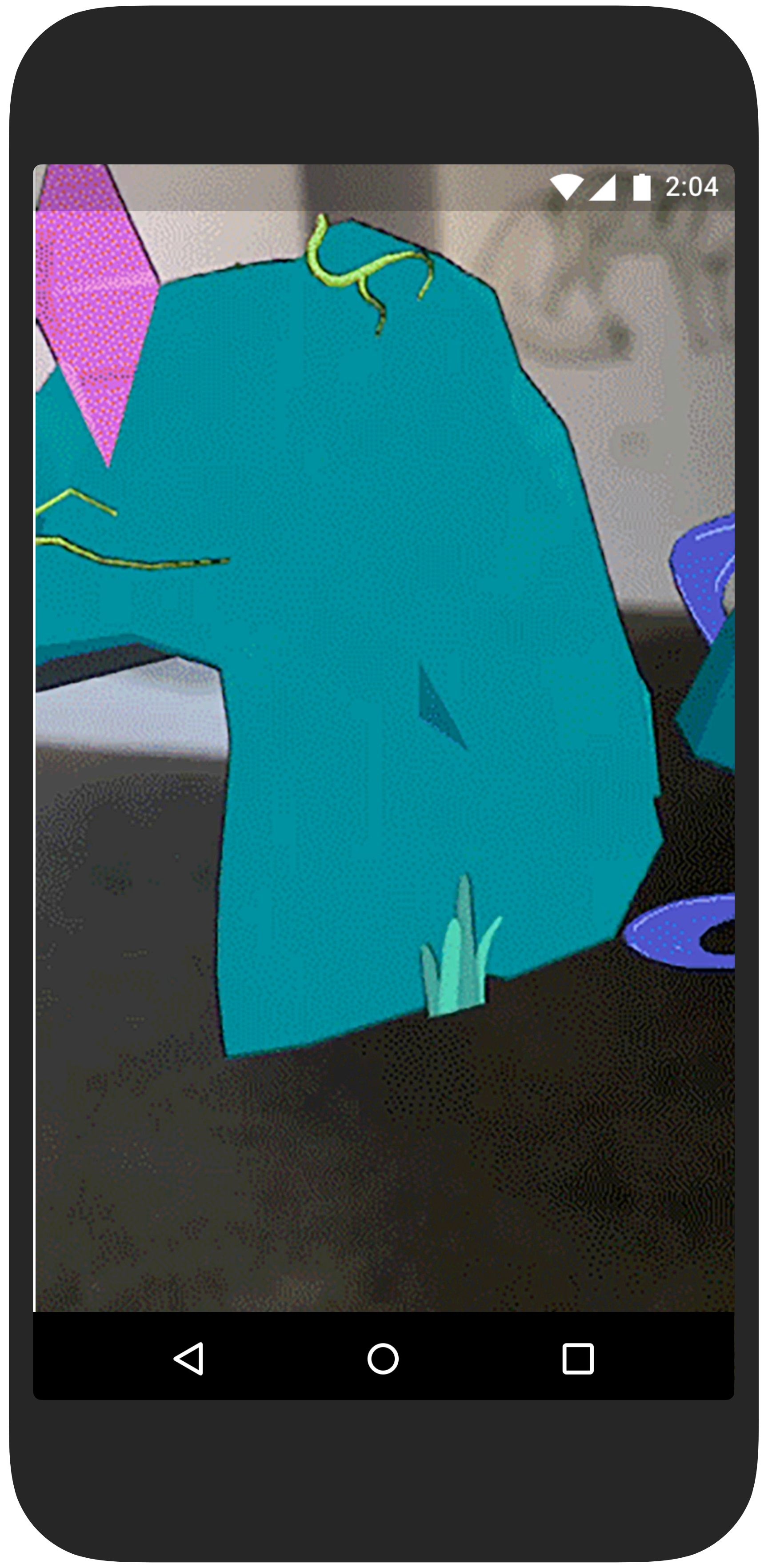


Héctor Ramos
Active now

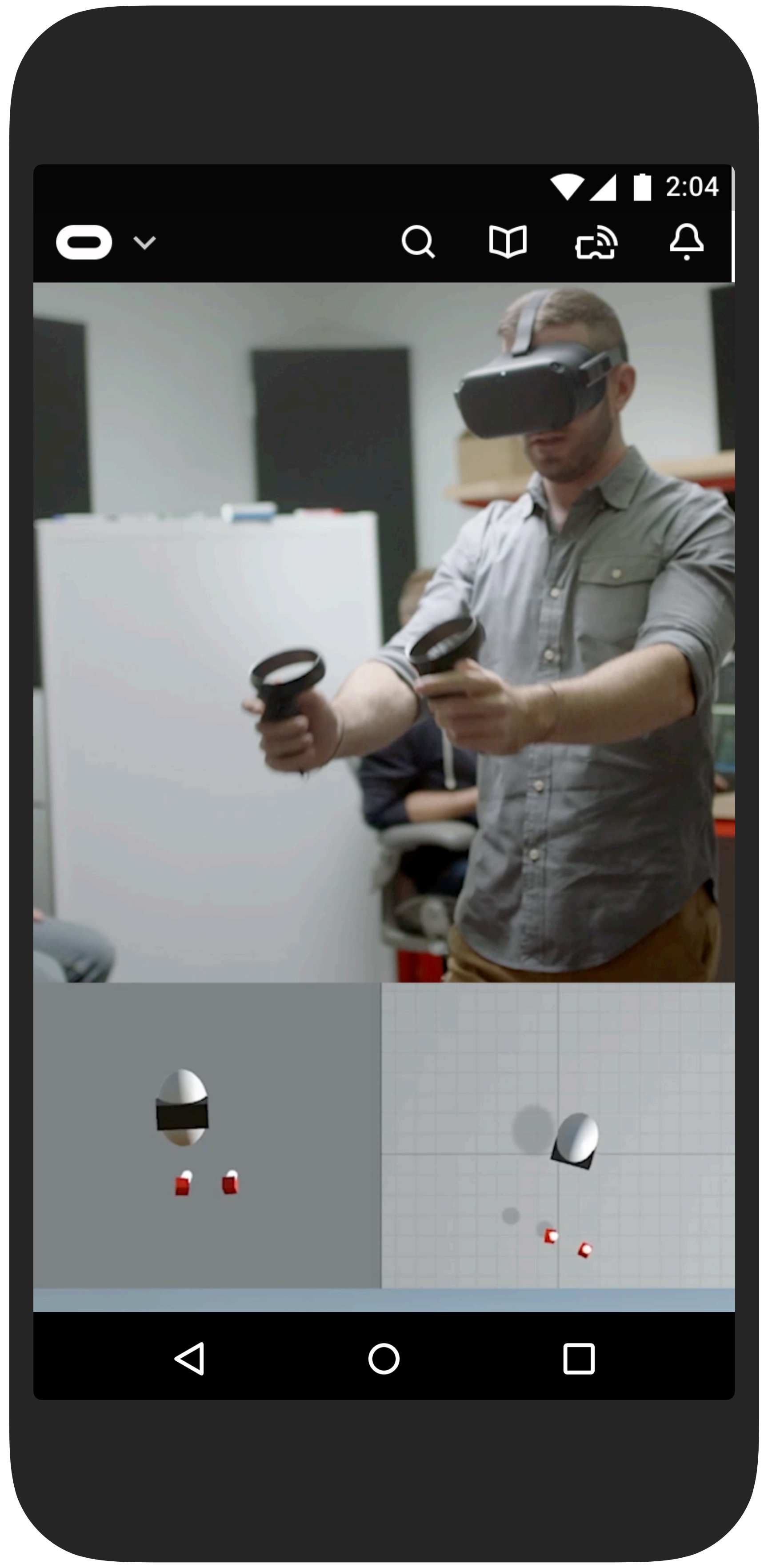
Sneakers (10.5) · $200
Marketplace

¡Hola! ¿Están esas zapatillas todavía en venta?

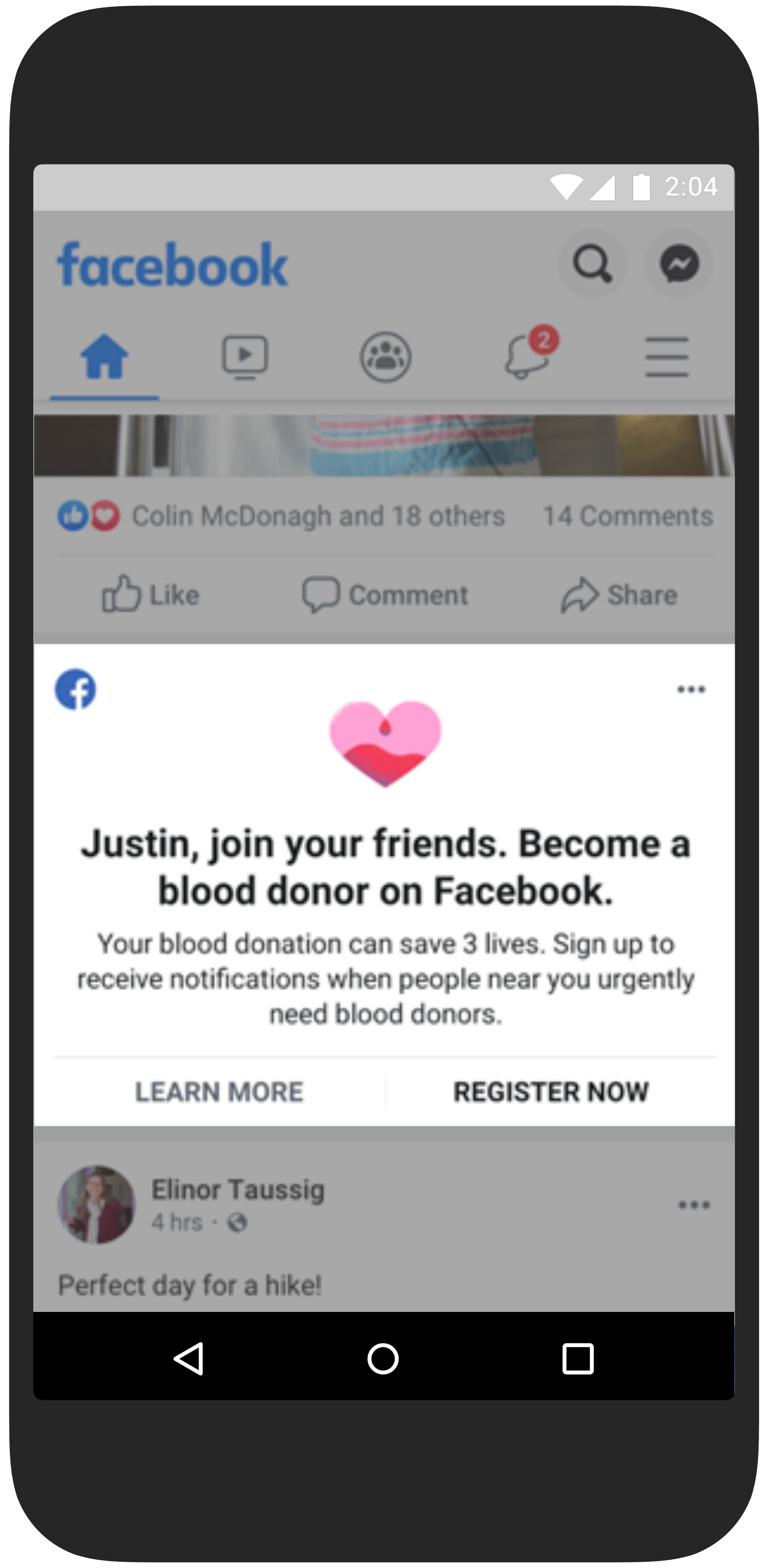
TRANSLATION



SPARK AR



OCULUS VR



Colin McDonagh and 18 others 14 Comments

Justin, join your friends. Become a blood donor on Facebook.

Your blood donation can save 3 lives. Sign up to receive notifications when people near you urgently need blood donors.

Elinor Taussig
4 hrs ·

Perfect day for a hike!

BLOOD DONATIONS

400T+

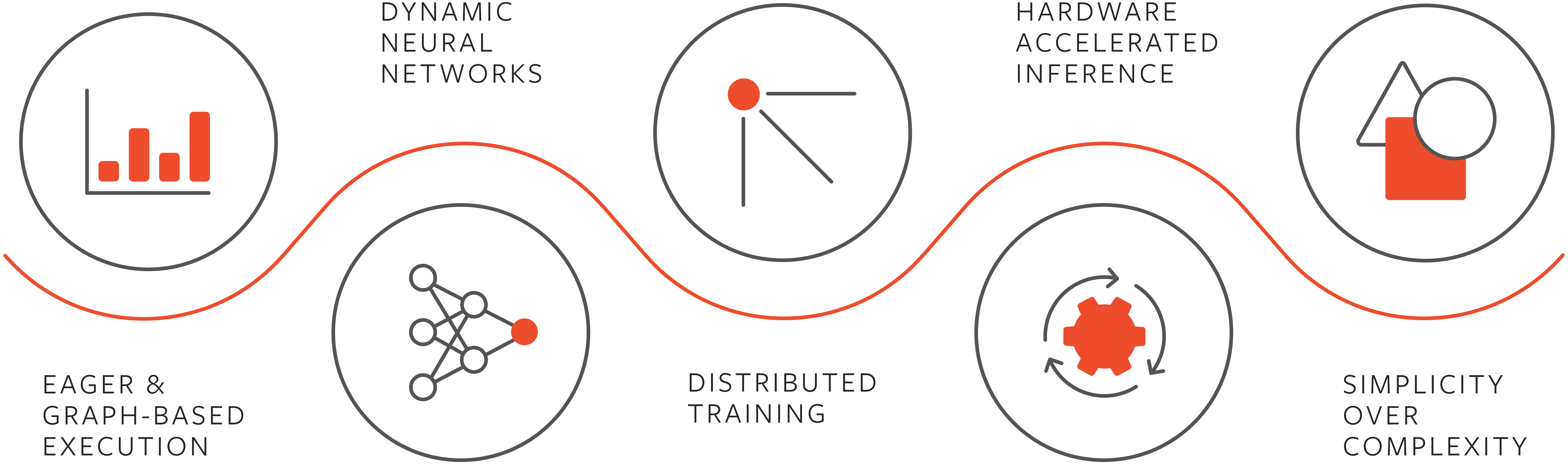
PREDICTIONS PER DAY

1B+

PHONES RUNNING NEURAL NETS GLOBALLY

WHAT IS PYTORCH?

WHAT IS PYTORCH?



EAGER & GRAPH-BASED EXECUTION

DYNAMIC NEURAL NETWORKS

DISTRIBUTED TRAINING

HARDWARE ACCELERATED INFERENCE

SIMPLICITY OVER COMPLEXITY

TORCH.NN

NEURAL NETWORKS

TORCH.OPTIM

OPTIMIZERS

TORCH.VISION

MODELS, DATA

TORCH.JIT

TORCH SCRIPT DEPLOYMENT

TORCH.AUTOGRAD

AUTO DIFFERENTIATION

TORCH.DATA

DATASETS & DATA LOADERS

DEFINING OUR NET

THE FORWARD PASS

LOADING DATA, MISC

TRAINING

```python
import torch

class Net(torch.nn.Module):
    def __init__(self):
        self.fc1 = torch.nn.Linear(8, 64)
        self.fc2 = torch.nn.Linear(64, 1)
```

DEFINING OUR NET

THE FORWARD PASS

LOADING DATA, MISC

TRAINING

```
import torch

class Net(torch.nn.Module):
    def __init__(self):
        self.fc1 = torch.nn.Linear(8, 64)
        self.fc2 = torch.nn.Linear(64, 1)
```

DEFINING OUR NET

THE FORWARD PASS

LOADING DATA, MISC

TRAINING

```python
def forward(self, x):
    x = torch.relu(self.fc1.forward(x))
    x = torch.dropout(x, p=0.5)
    x = torch.sigmoid(self.fc2.forward(x))
    return x
```

DEFINING OUR NET

THE FORWARD PASS

LOADING DATA, MISC

TRAINING

```
def forward(self, x):
    x = torch.relu(self.fc1.forward(x))
    x = torch.dropout(x, p=0.5)
    x = torch.sigmoid(self.fc2.forward(x))
    return x
```

DEFINING OUR NET

THE FORWARD PASS

LOADING DATA, MISC

TRAINING

```python
net = Net()

data_loader = torch.utils.data.DataLoader(
    torchvision.datasets.MNIST('./data'))

optimizer = torch.optim.SGD(net.parameters())
```

DEFINING OUR NET

THE FORWARD PASS

LOADING DATA, MISC

TRAINING

```
net = Net()

data_loader = torch.utils.data.DataLoader(
    torchvision.datasets.MNIST('./data'))

optimizer = torch.optim.SGD(net.parameters())
```

DEFINING OUR NET

THE FORWARD PASS

LOADING DATA, MISC

TRAINING

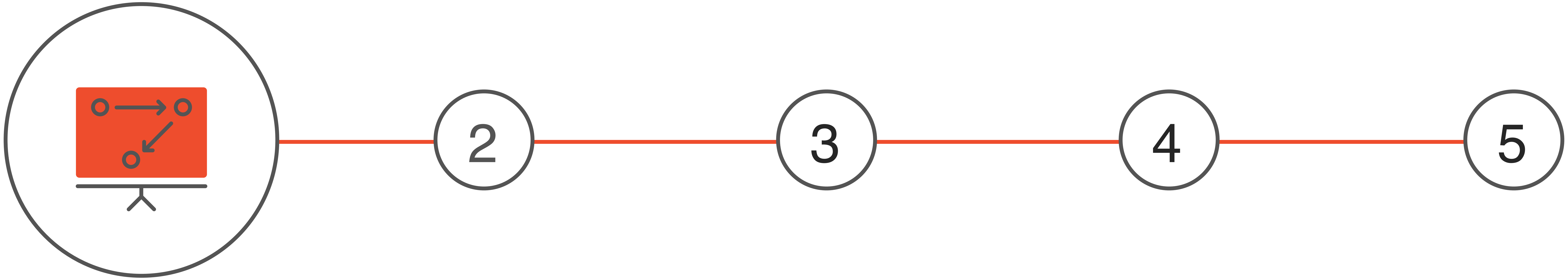
```python
for epoch in range(1, 11):
    for data, target in data_loader:
        optimizer.zero_grad()
        prediction = net.forward(data)
        loss = F.nll_loss(prediction, target)
        loss.backward()
        optimizer.step()
    if epoch % 2 == 0:
        torch.save(net, "net.pt")
```

RESEARCH TO PRODUCTION

RESEARCH TO PRODUCTION

RESEARCH TO PRODUCTION

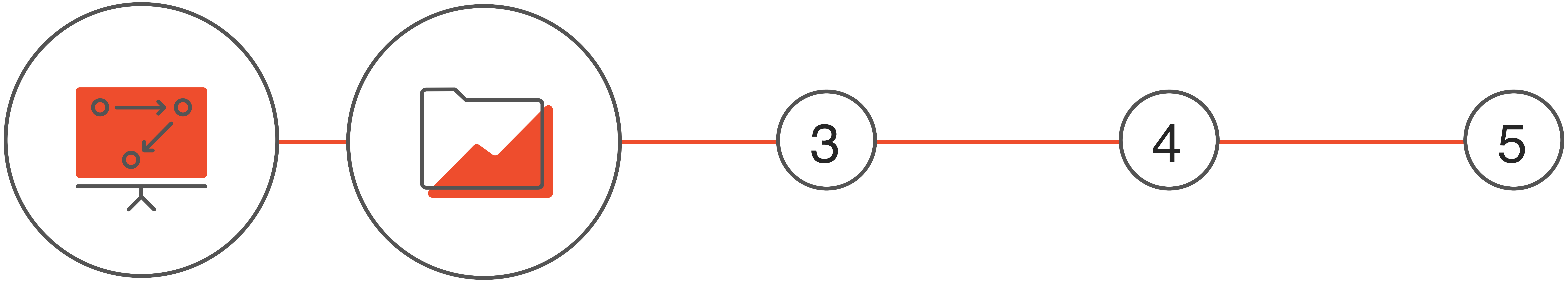
RESEARCH TO PRODUCTION



1

DETERMINE
APPROACH

RESEARCH TO PRODUCTION



3

4

5

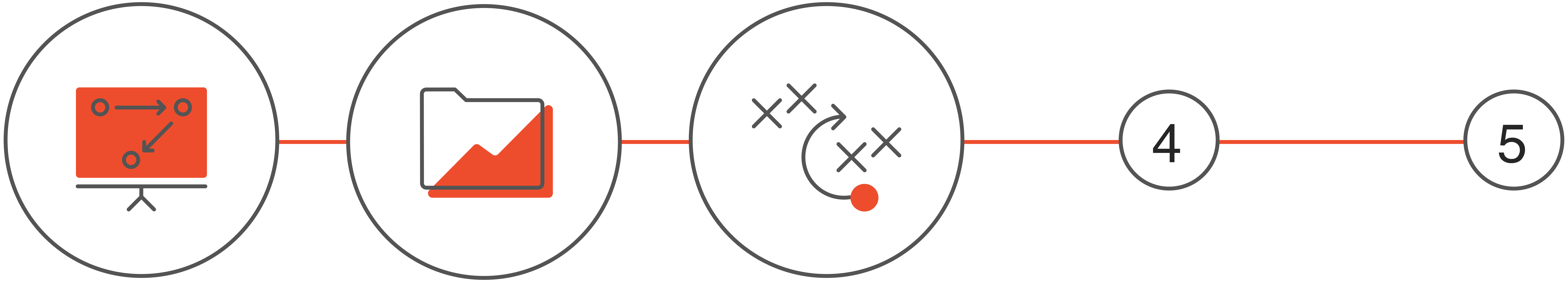
1

DETERMINE
APPROACH

2

PREPARE
DATA

RESEARCH TO PRODUCTION



1

DETERMINE APPROACH

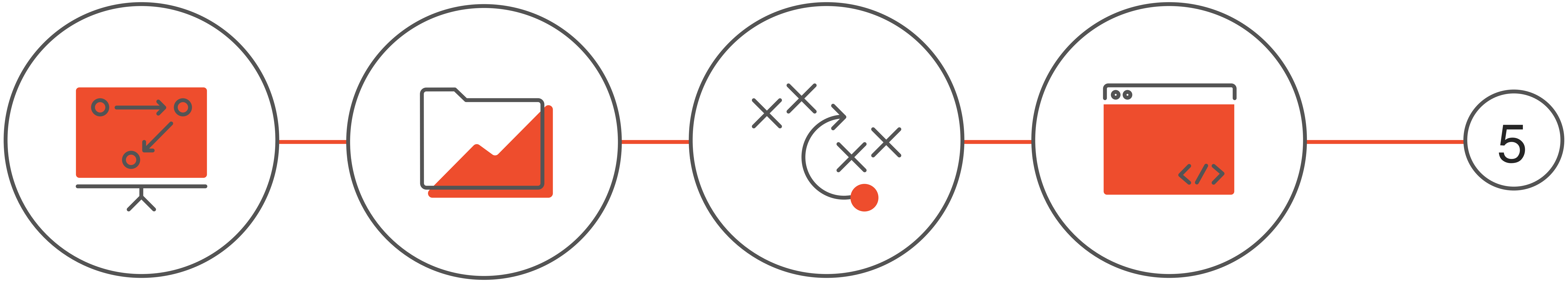
2

PREPARE DATA

3

BUILD AND TRAIN MODEL

RESEARCH TO PRODUCTION



1

DETERMINE APPROACH

2

PREPARE DATA

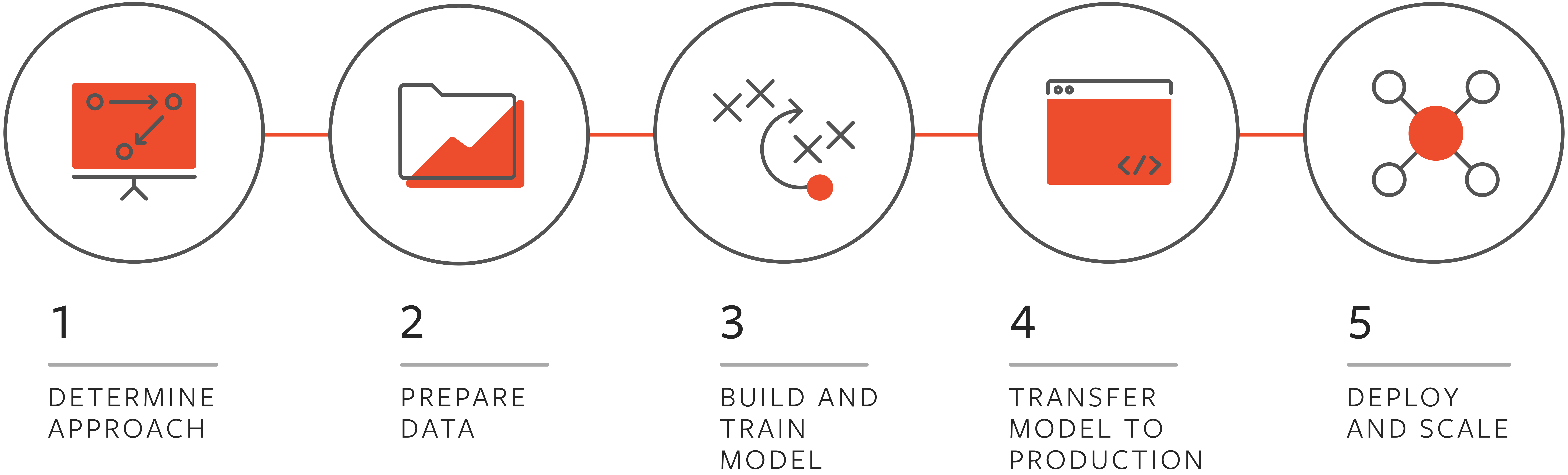
3

BUILD AND TRAIN MODEL

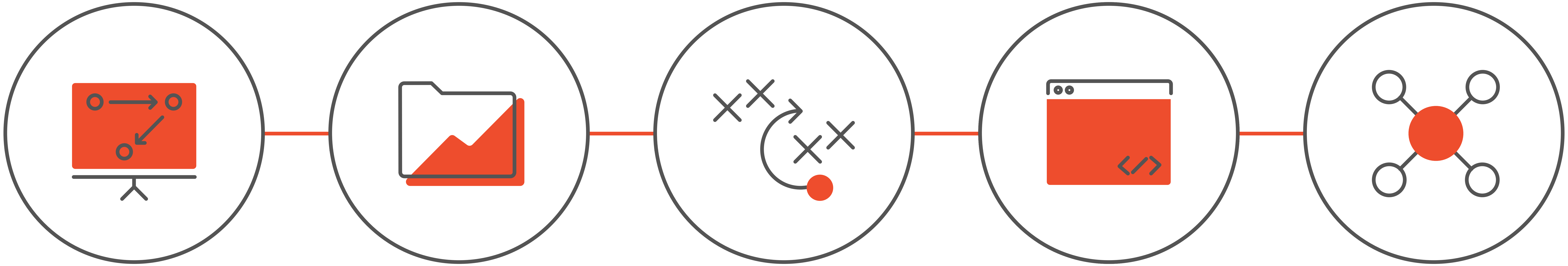
4

TRANSFER MODEL TO PRODUCTION

RESEARCH TO PRODUCTION



RESEARCH TO PRODUCTION



1

DETERMINE APPROACH

2

PREPARE DATA

3

BUILD AND TRAIN MODEL

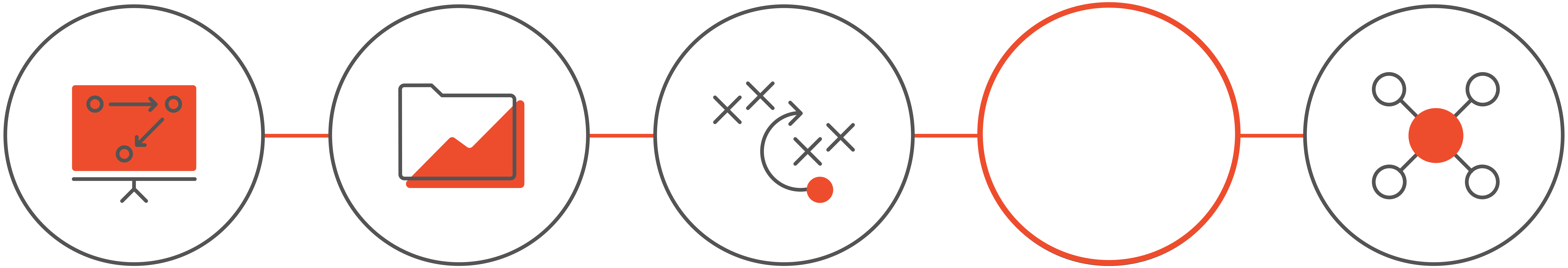
4

TRANSFER MODEL TO PRODUCTION

5

DEPLOY AND SCALE

RESEARCH TO PRODUCTION



1

DETERMINE APPROACH

2

PREPARE DATA

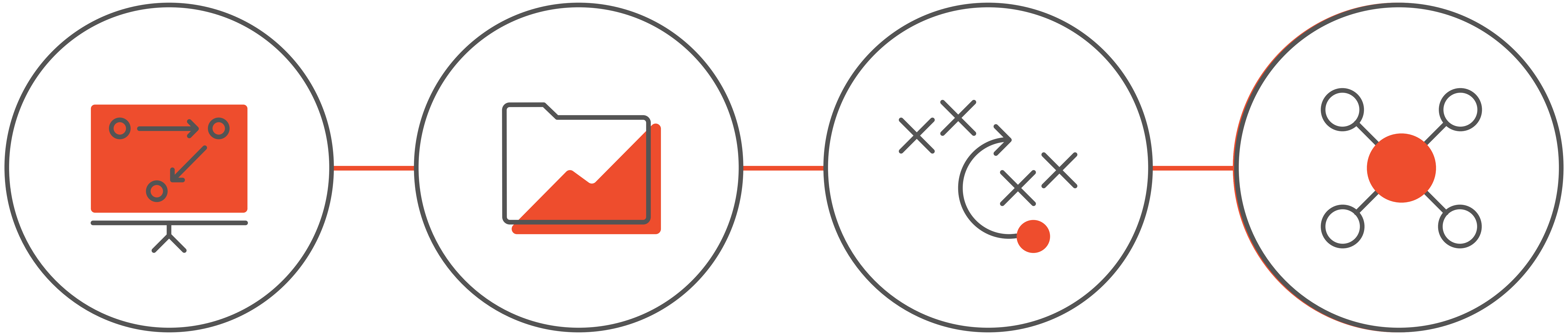
3

BUILD AND TRAIN MODEL

5

DEPLOY AND SCALE

RESEARCH TO PRODUCTION



1

DETERMINE APPROACH

2

PREPARE DATA

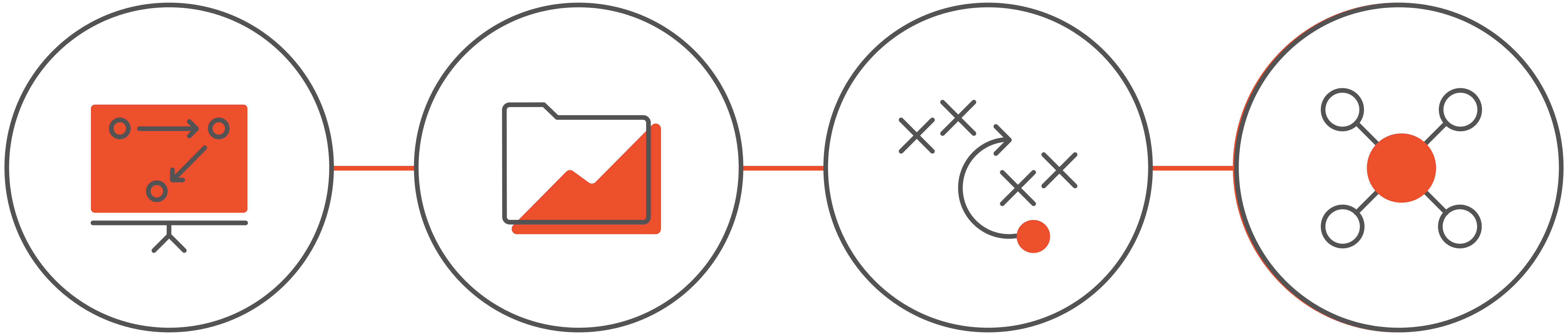
3

BUILD AND TRAIN MODEL

5

DEPLOY AND SCALE

RESEARCH TO PRODUCTION



1

DETERMINE APPROACH

2

PREPARE DATA

3

BUILD AND TRAIN MODEL

4

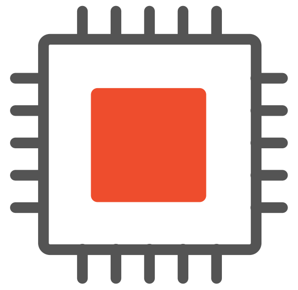
DEPLOY AND SCALE

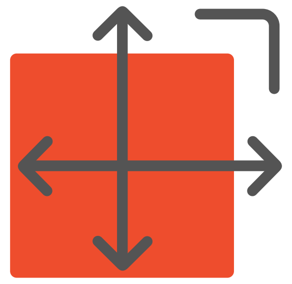
HARDWARE EFFICIENCY

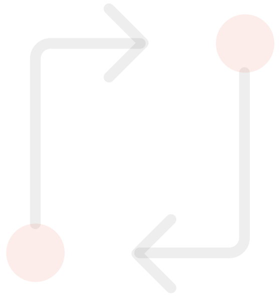
OPTIMAL PERFORMANCE ON CPU/GPU/ASIC

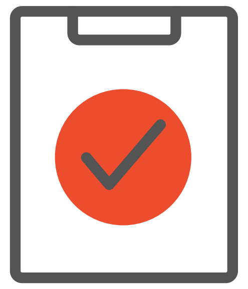
SCALABILITY

TRAINING ON HUGE DATASETS
BILLIONS OF INFERENCES PER SECOND









CROSS-PLATFORM

EMBEDDED OR MOBILE DEVICES

CONSTRAINED SYSTEM ENVIRONMENTS

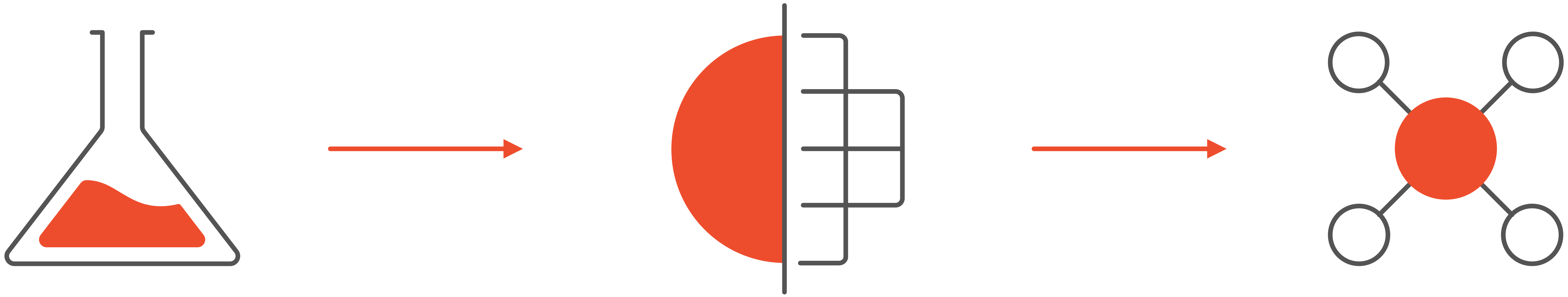
RELIABILITY

TRAINING JOBS RUNNING FOR WEEKS

100S OF GPUS

TORCH.JIT TO CAPTURE AND OPTIMIZE PYTORCH CODE

TORCH.JIT TO CAPTURE AND OPTIMIZE PYTORCH CODE



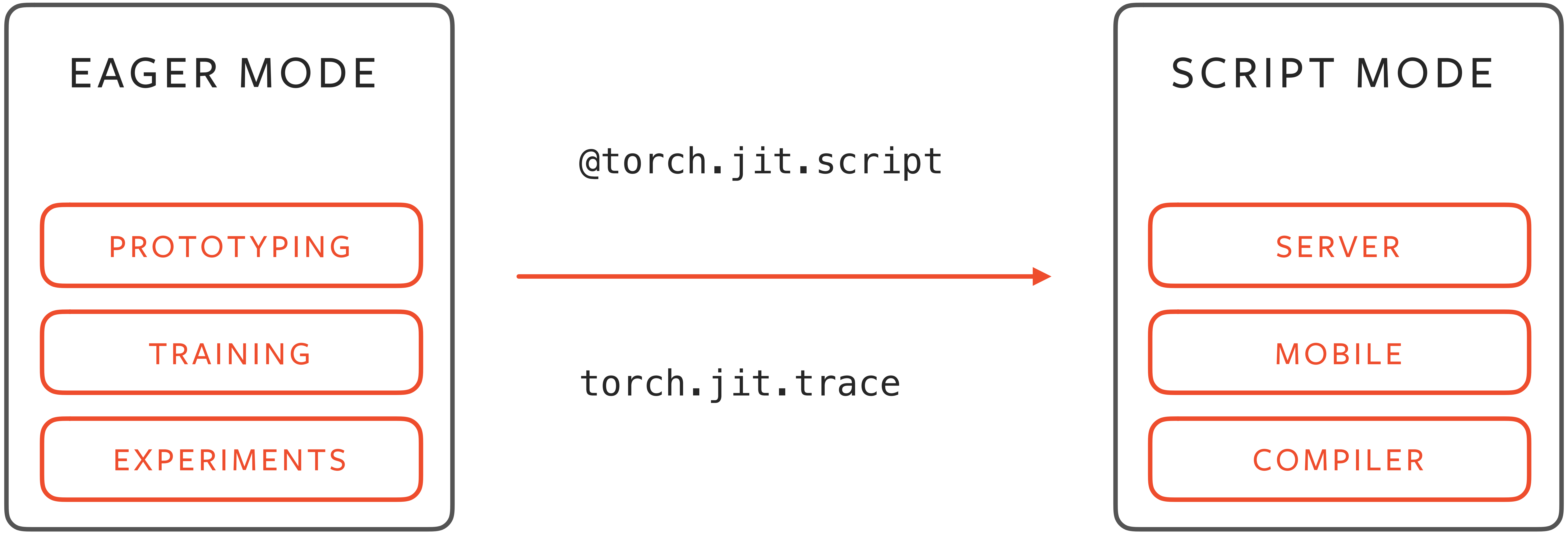
EXPERIMENT

EXTRACT TORCHSCRIPT

OPTIMIZE AND DEPLOY

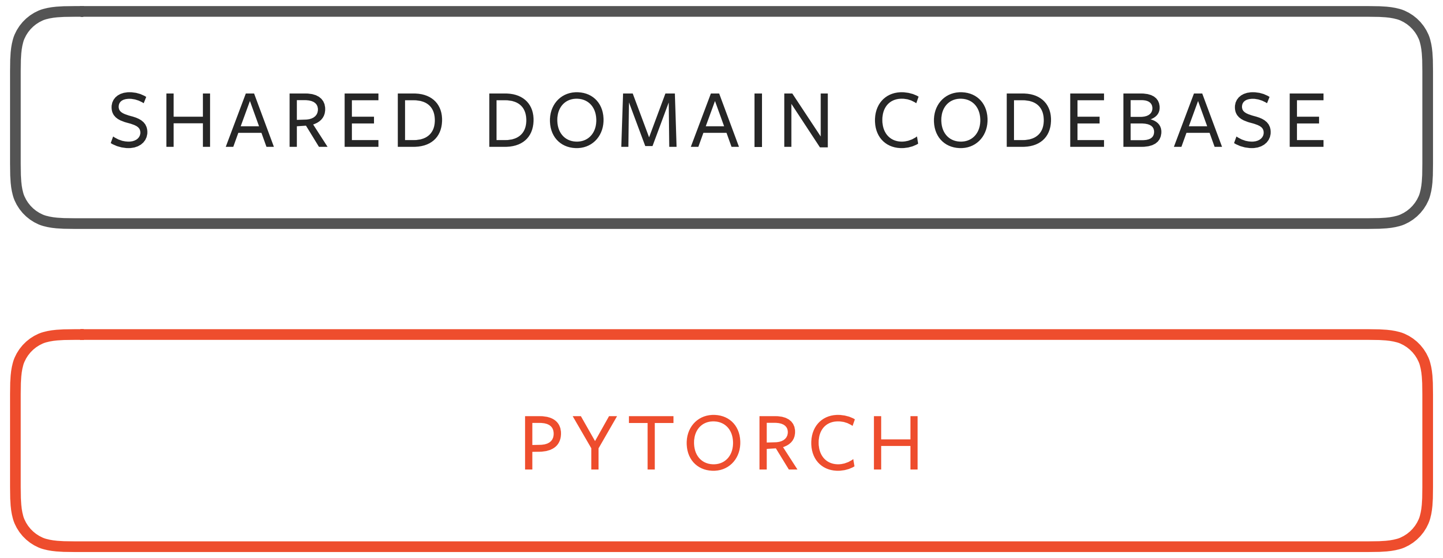
LEVERAGING TORCH.JIT

LEVERAGING TORCH.JIT



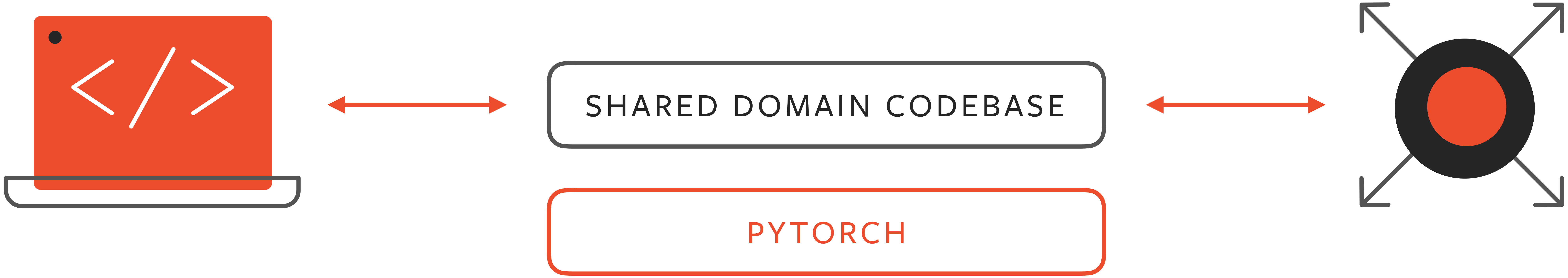
colab

DEMO



SHARED DOMAIN CODEBASE

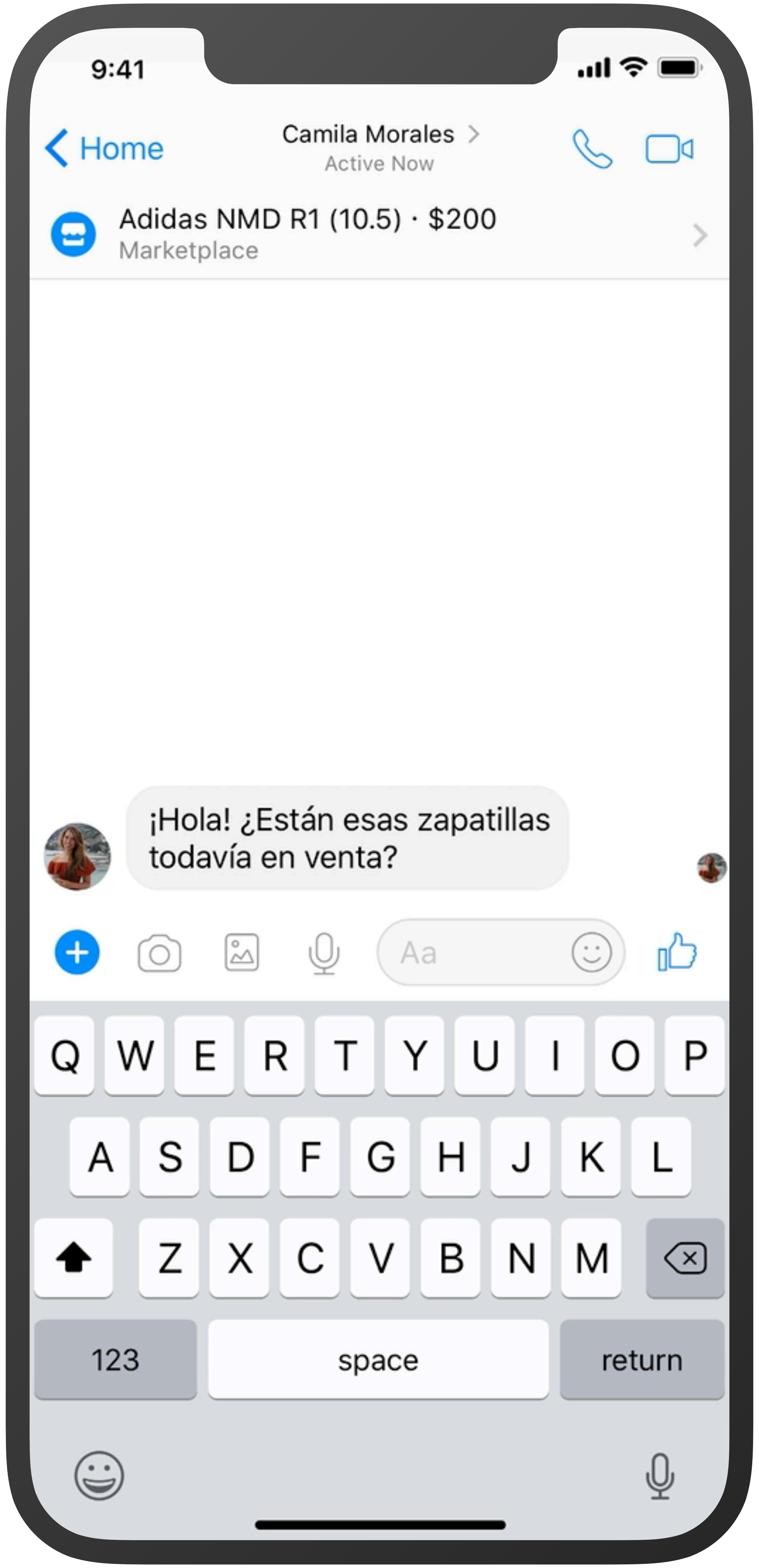
PYTORCH



SHARED DOMAIN CODEBASE

PYTORCH

PYTEXT IN MESSENGER



9:41

Home

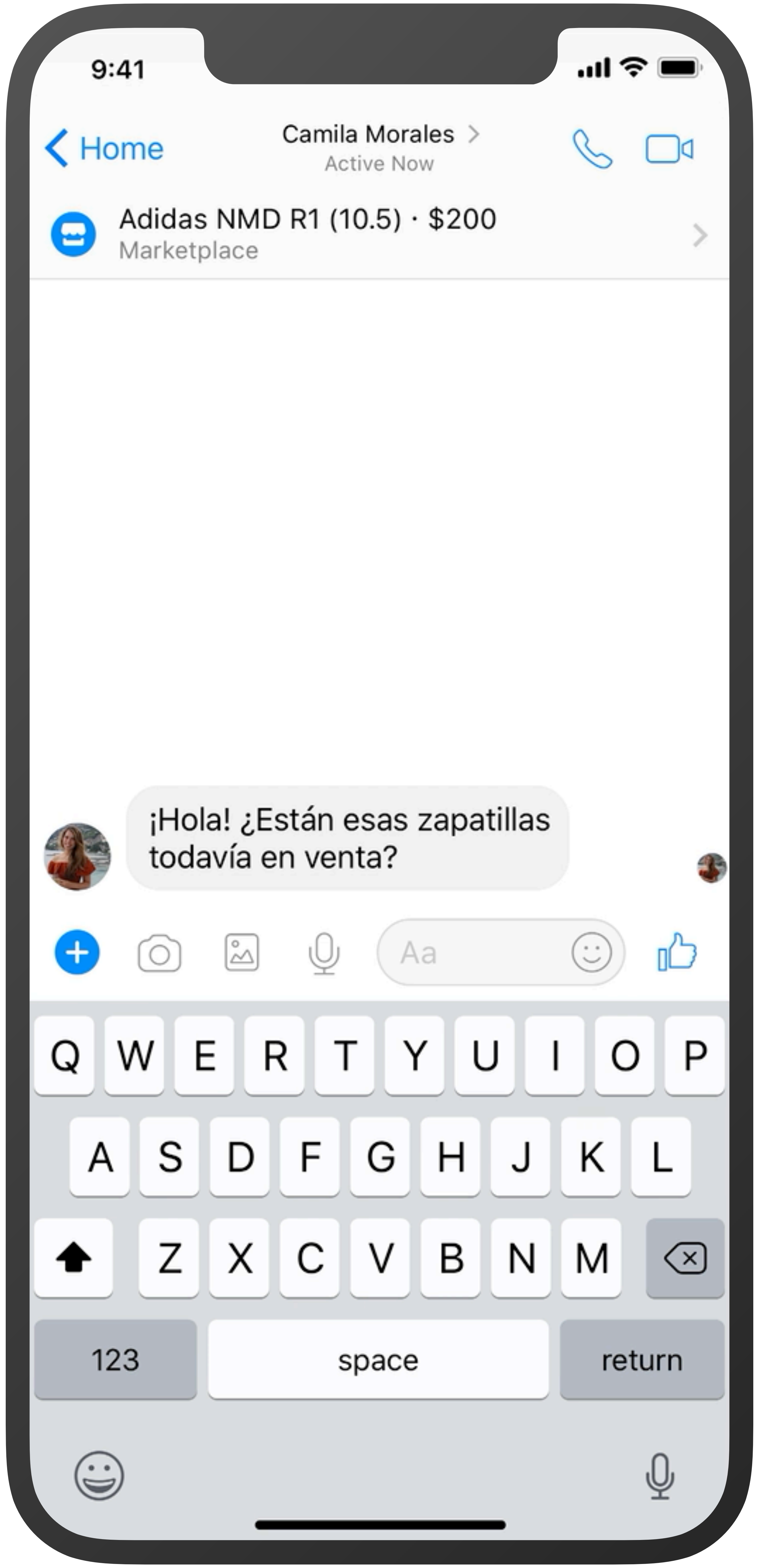
Camila Morales
Active Now

Adidas NMD R1 (10.5) · $200
Marketplace

¡Hola! ¿Están esas zapatillas todavía en venta?

PYTEXT IN MESSENGER

- ✓ Real Time
- ✓ Designed to Scale
- ✓ Different Languages



9:41

Home

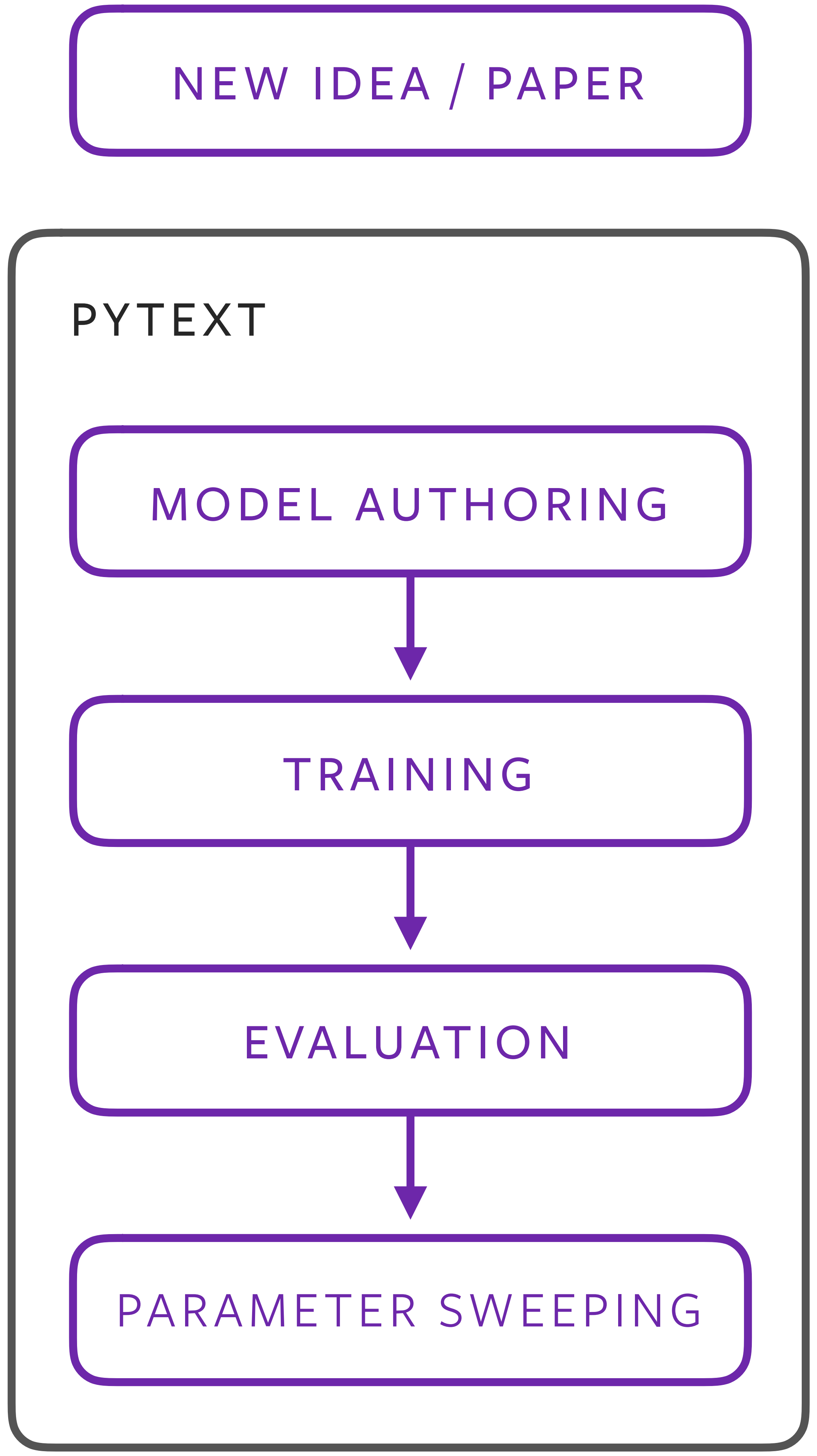
Camila Morales
Active Now

Adidas NMD R1 (10.5) · $200
Marketplace

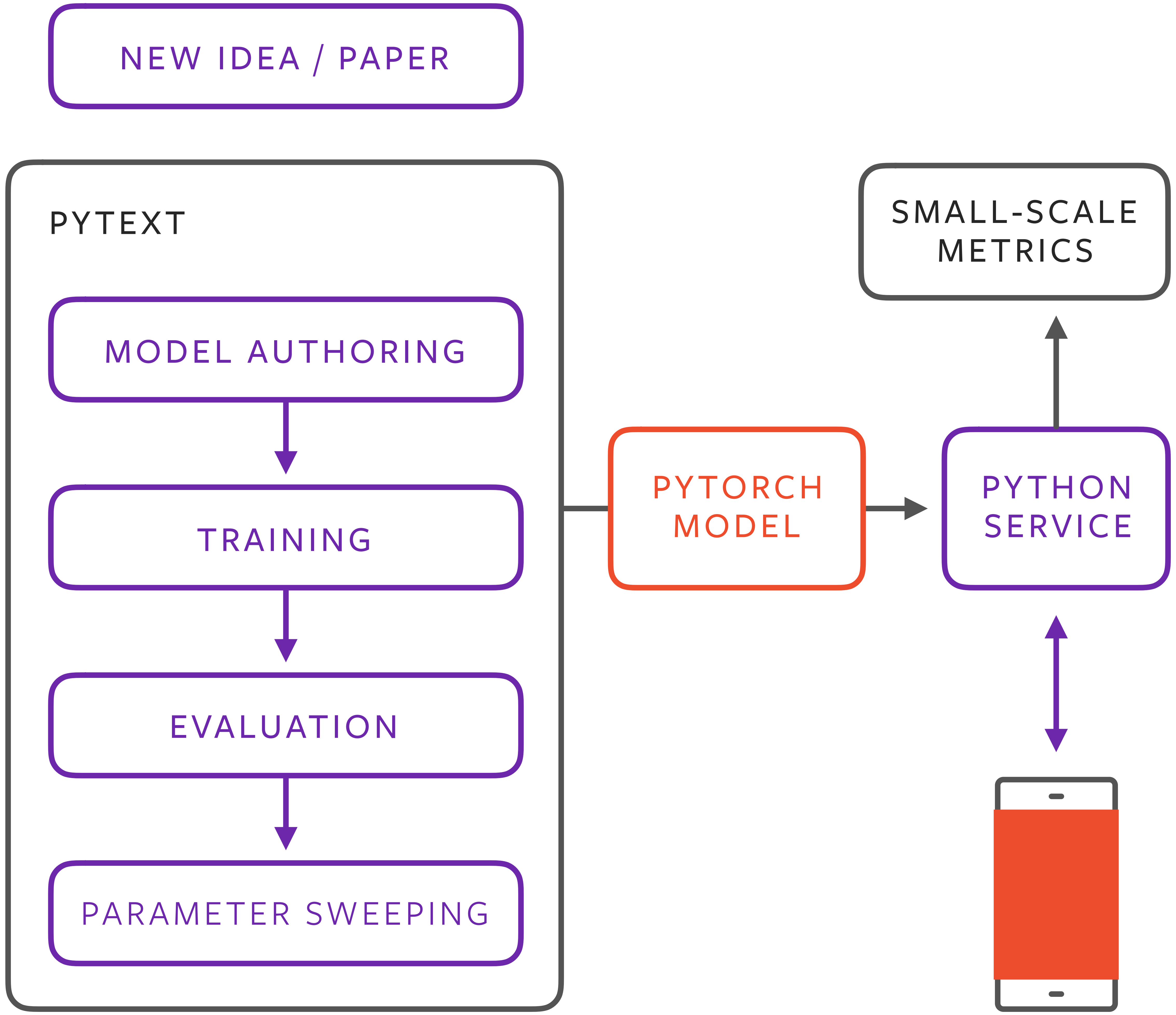
¡Hola! ¿Están esas zapatillas todavía en venta?

PYTEXT RESEARCH TO PRODUCTION CYCLE

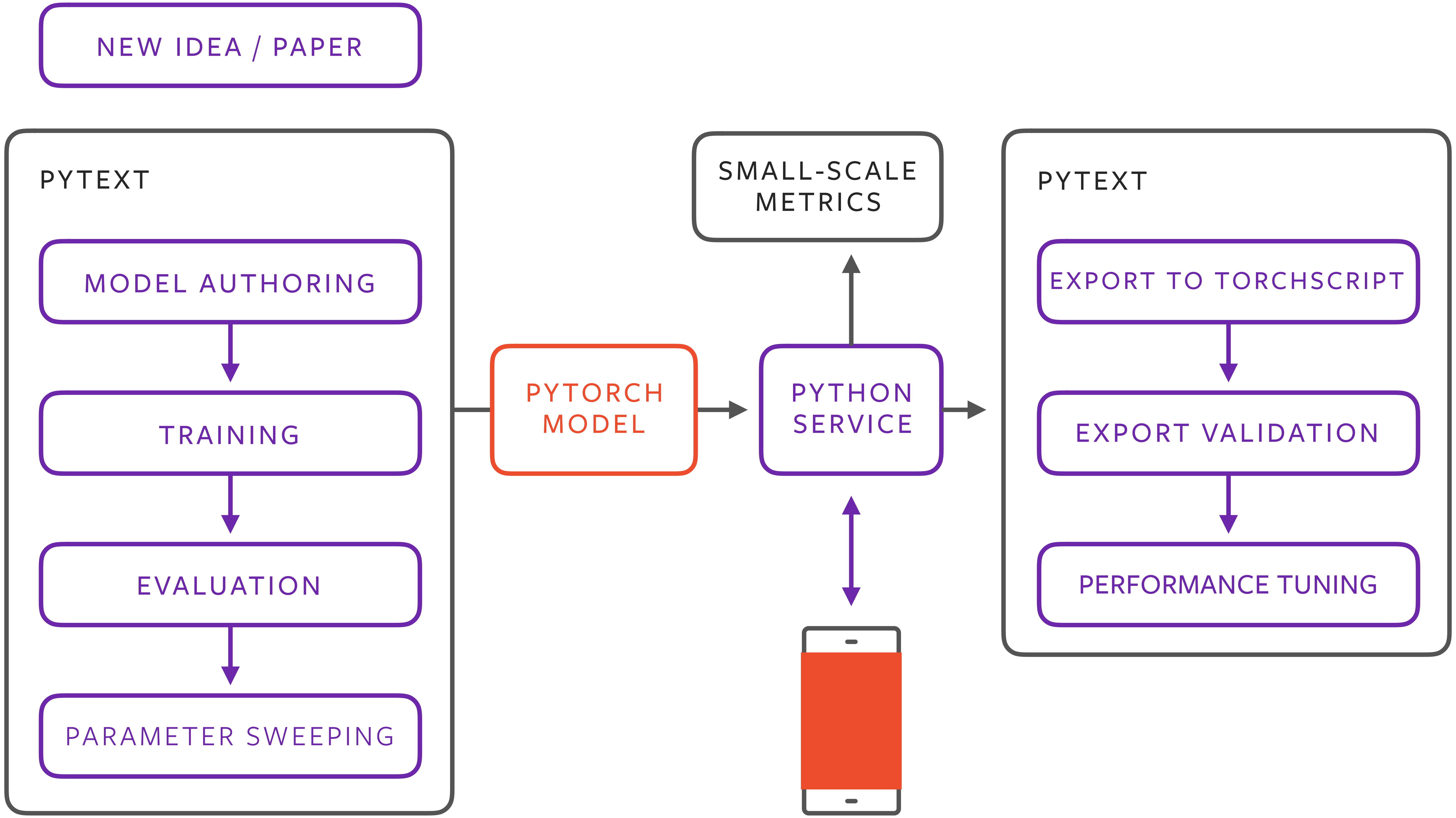
PYTEXT RESEARCH TO PRODUCTION CYCLE



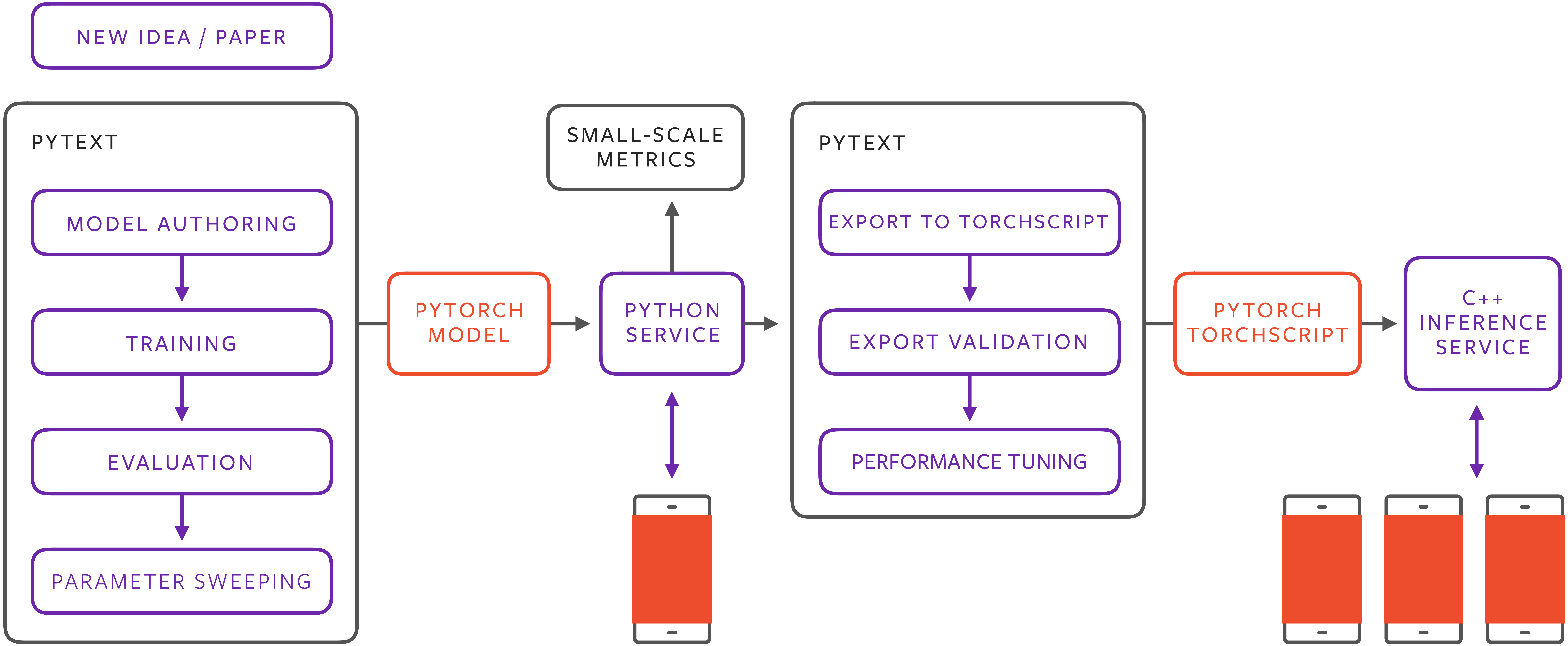
PYTEXT RESEARCH TO PRODUCTION CYCLE



PYTEXT RESEARCH TO PRODUCTION CYCLE

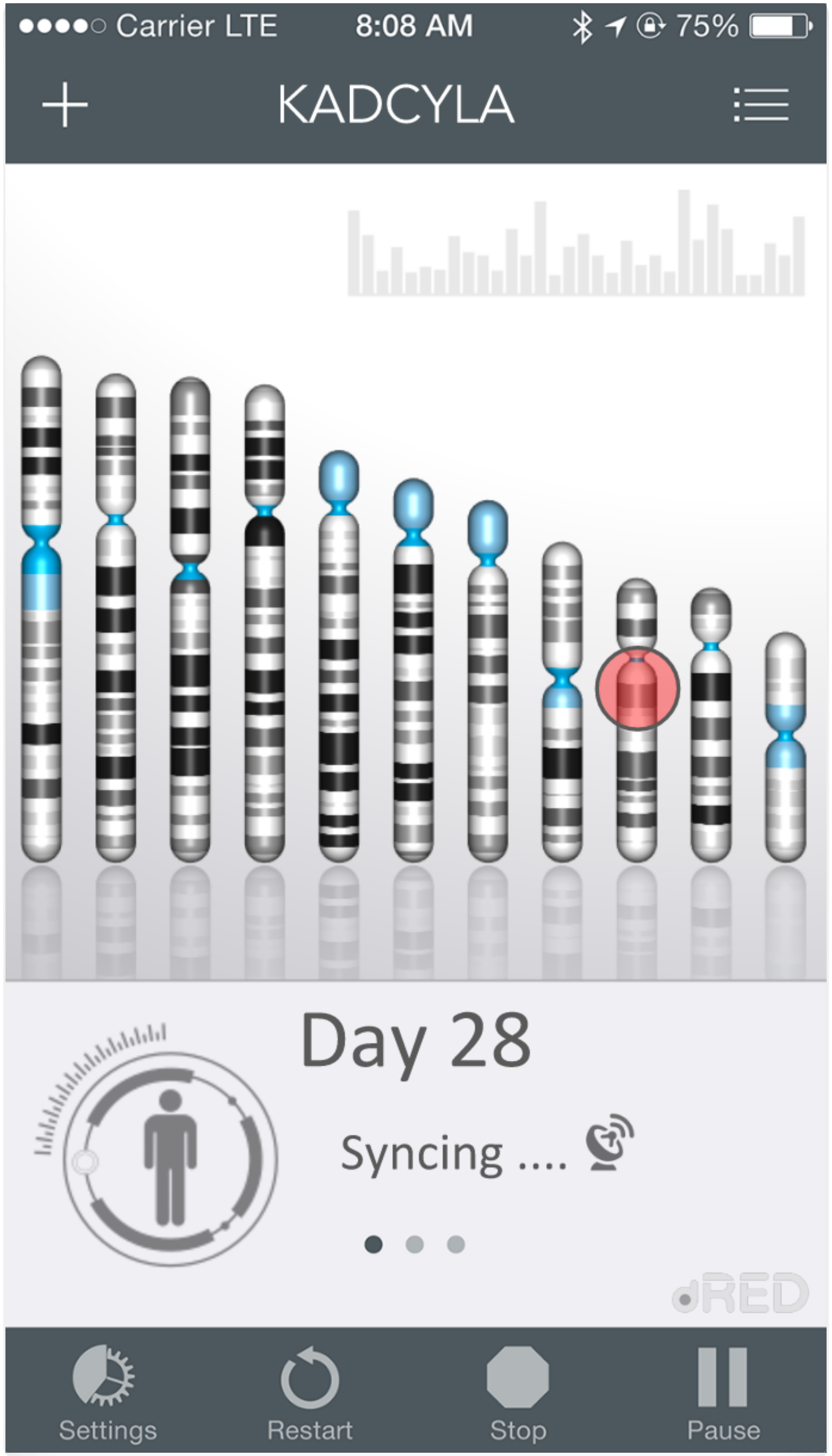


PYTEXT RESEARCH TO PRODUCTION CYCLE

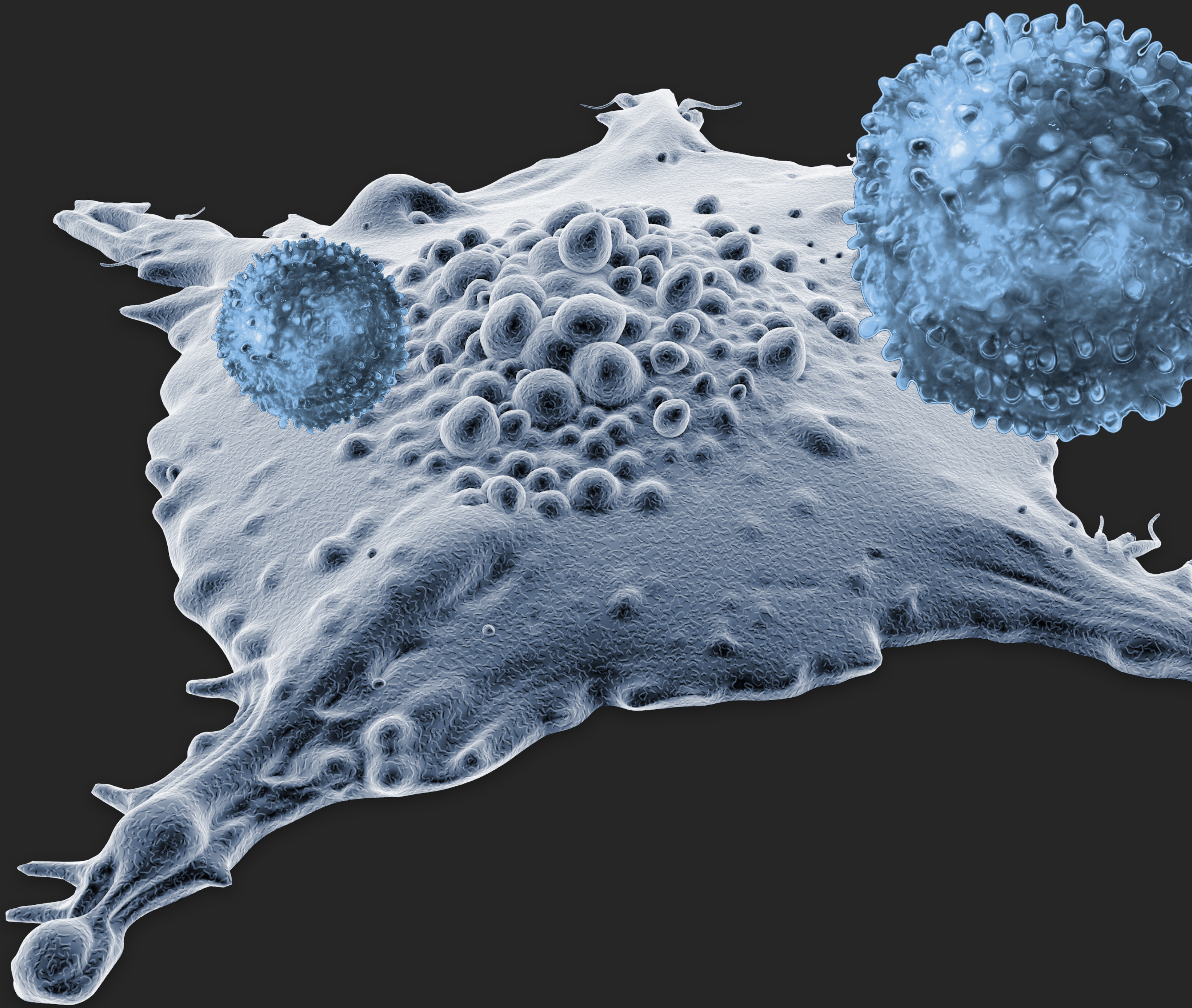


PERSONALIZED CANCER VACCINES

PERSONALIZED CANCER THERAPY



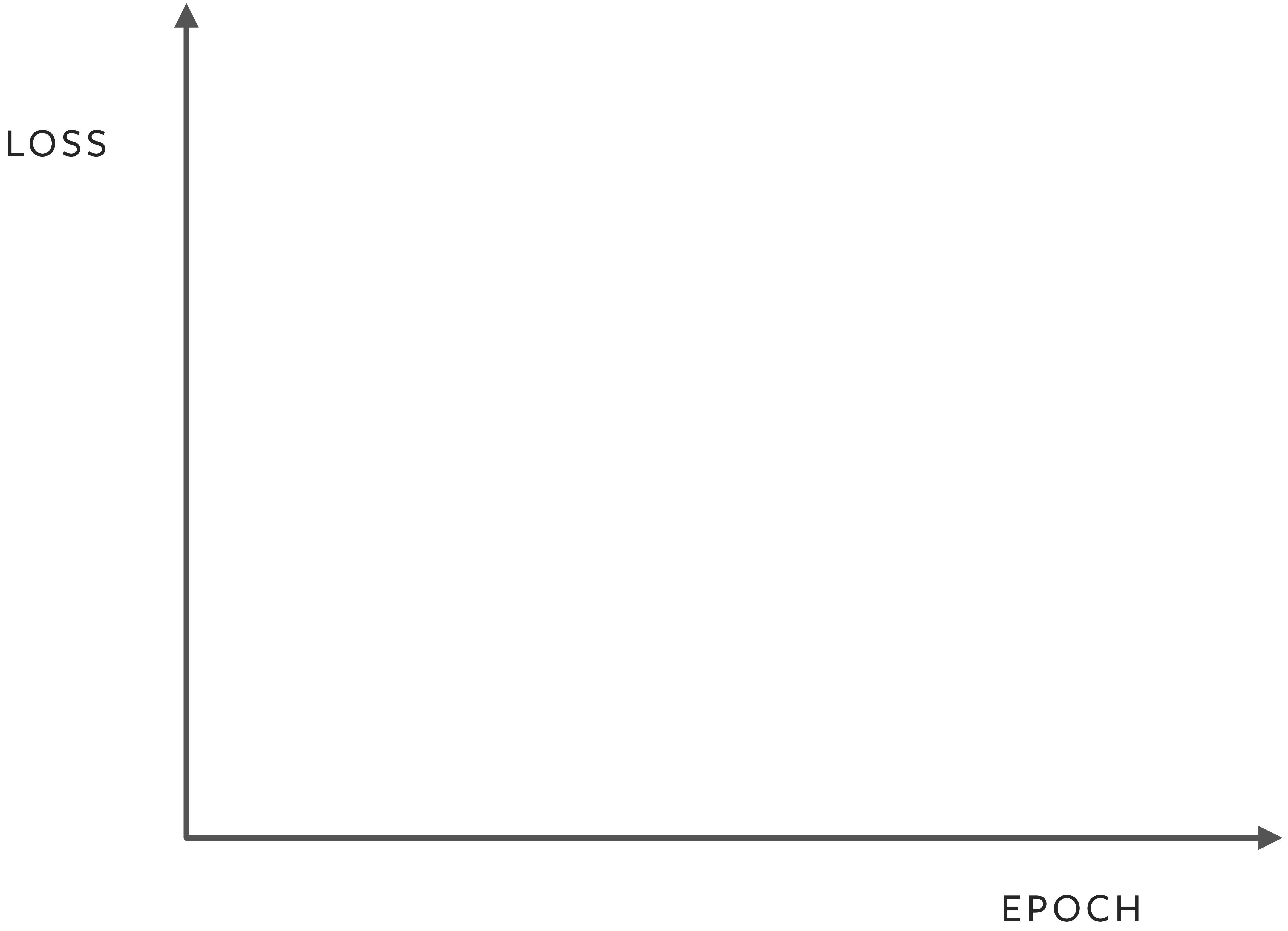
LEVERAGING THE IMMUNE SYSTEM AND AI TO FIGHT CANCER



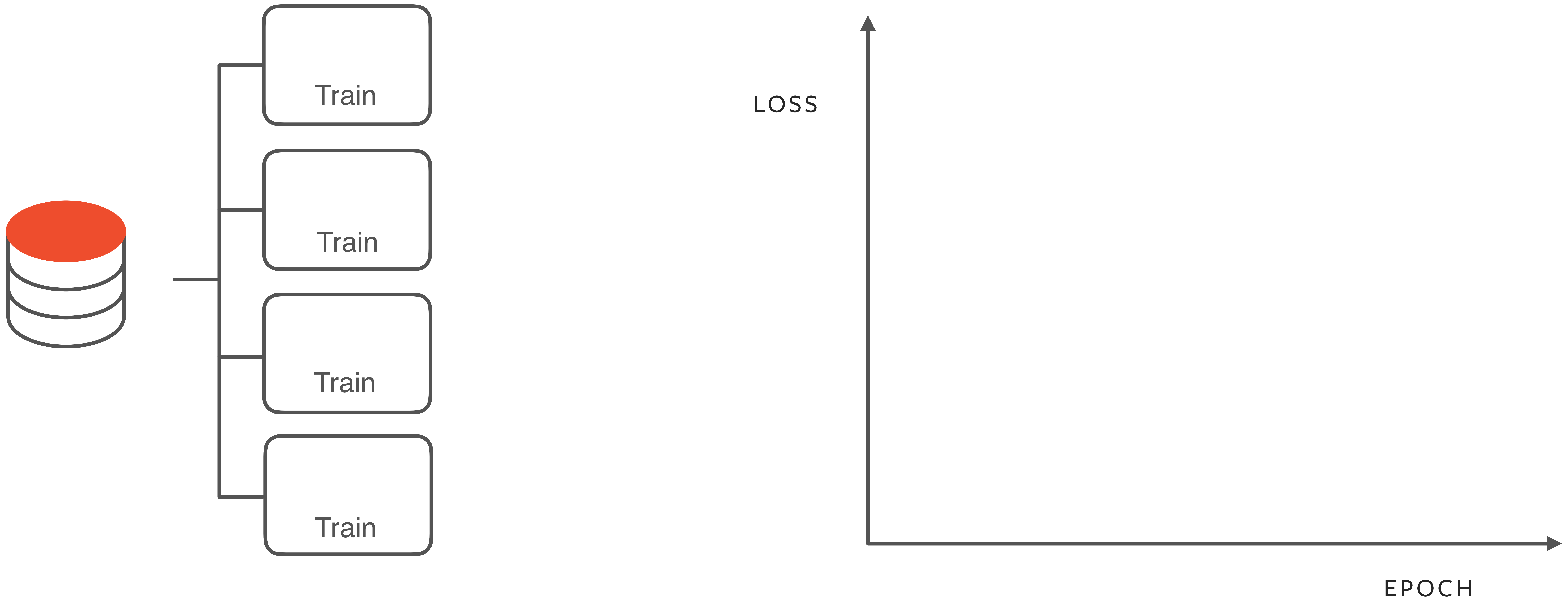
PREDICTIVE DRIVER ASSISTANCE

LIBRARIES: OPTIMIZATION

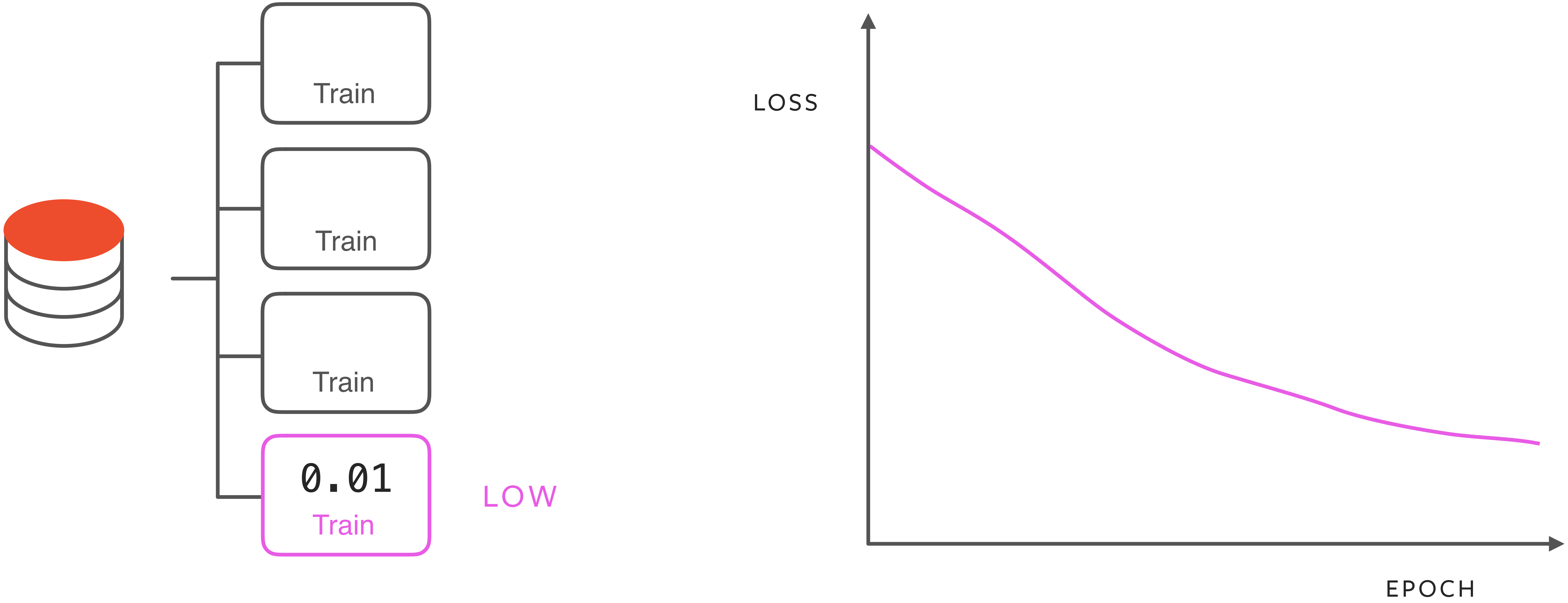
BAYESOPT AT FACEBOOK | HYPERPARAMETER OPTIMIZATION



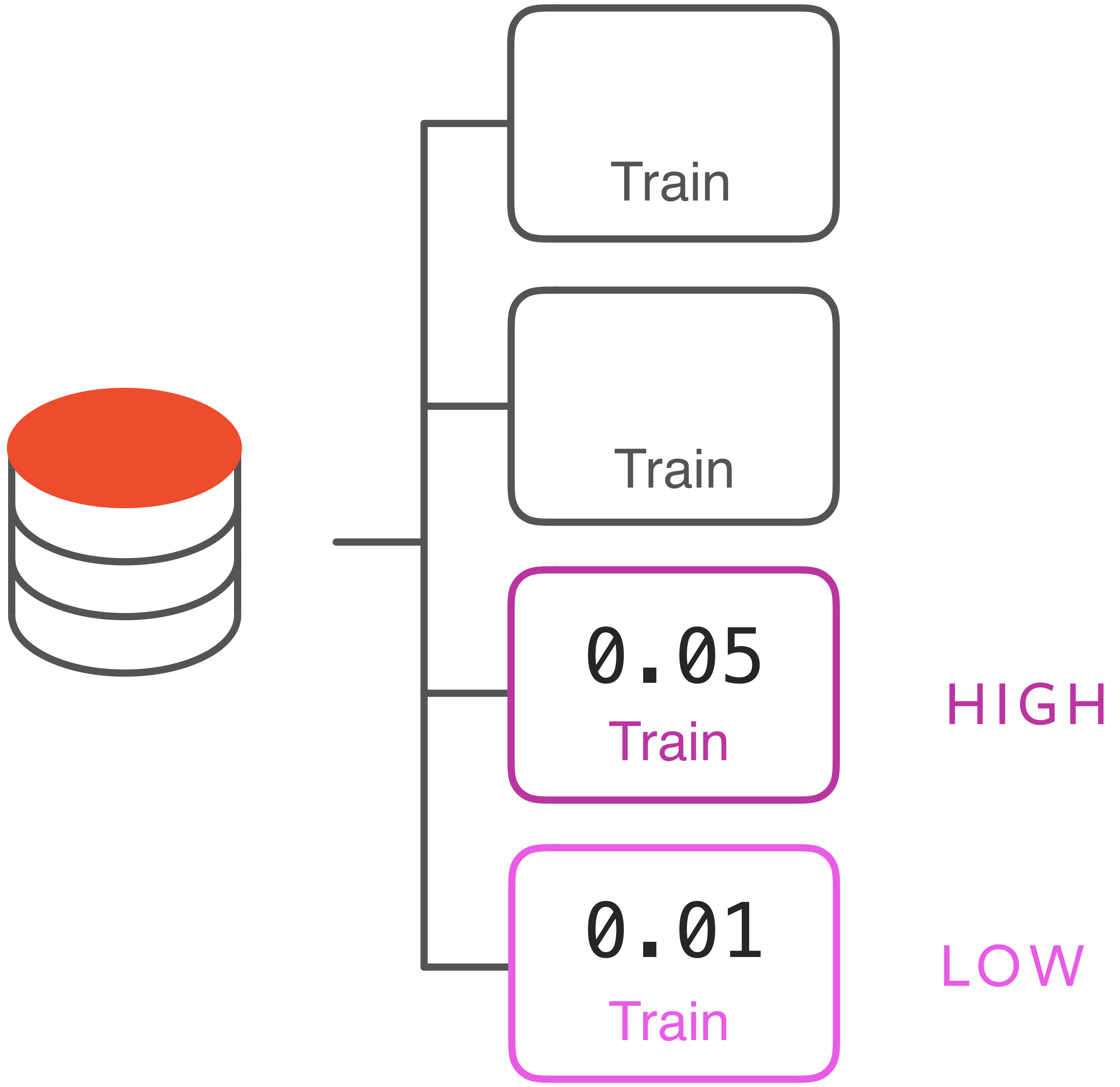
BAYESOPT AT FACEBOOK | HYPERPARAMETER OPTIMIZATION



BAYESOPT AT FACEBOOK | HYPERPARAMETER OPTIMIZATION



BAYESOPT AT FACEBOOK | HYPERPARAMETER OPTIMIZATION



Train

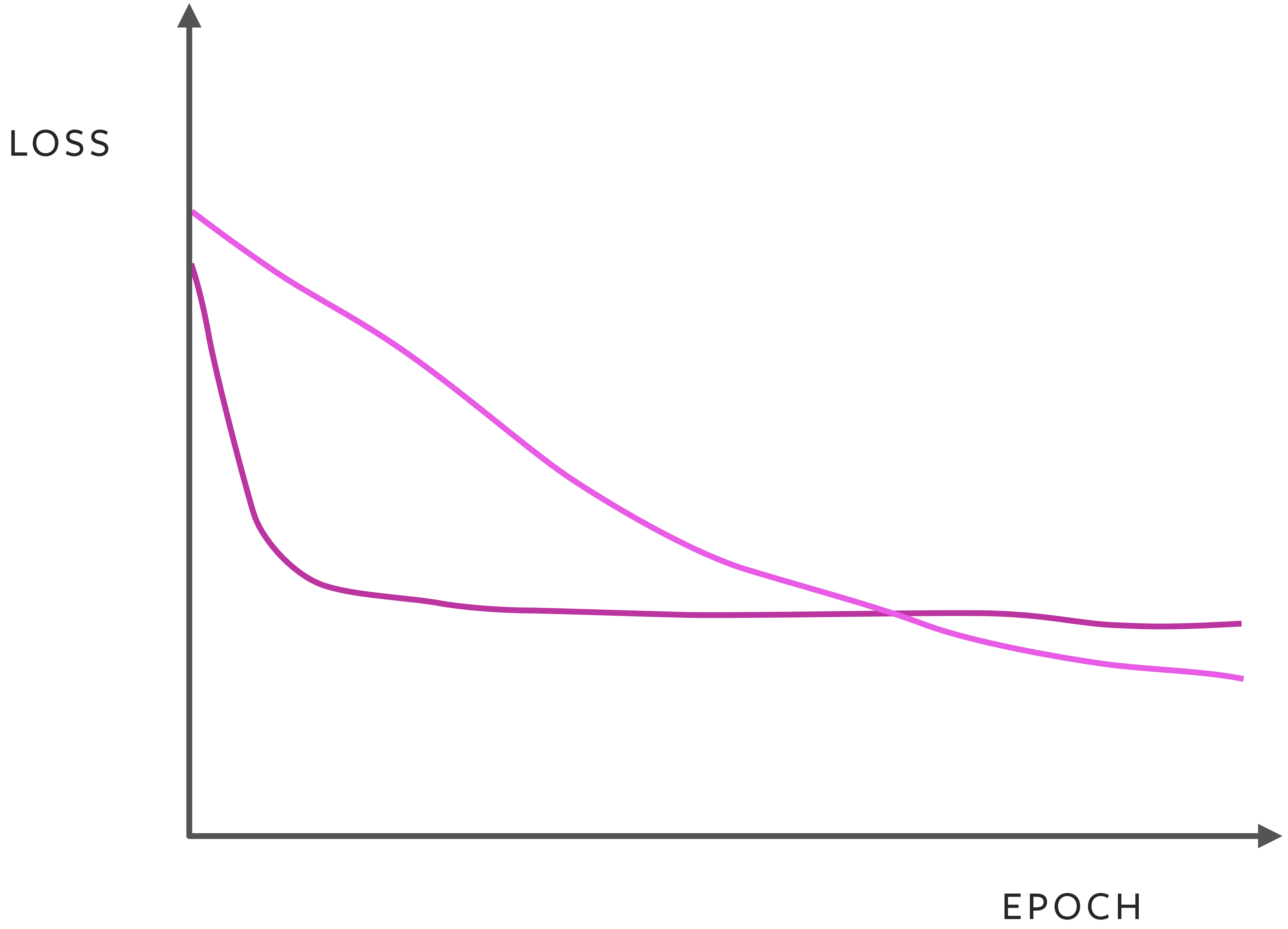
Train

0.05
Train

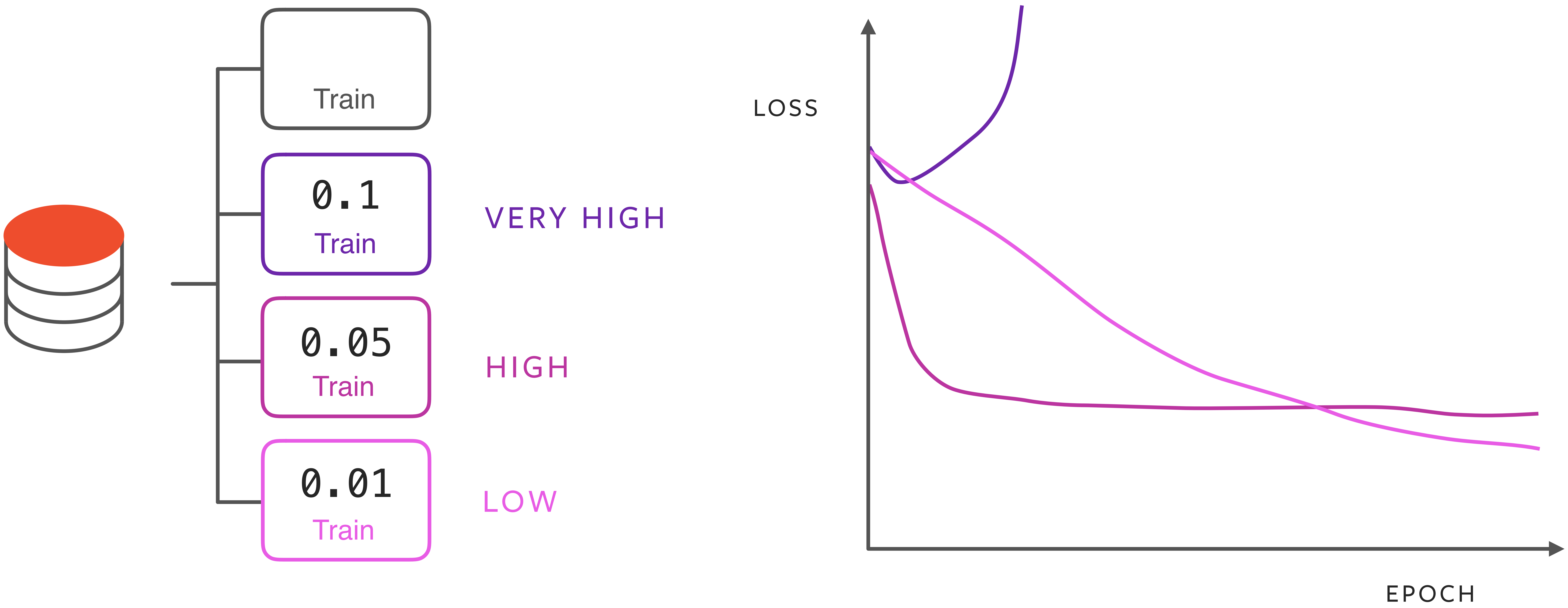
HIGH

0.01
Train

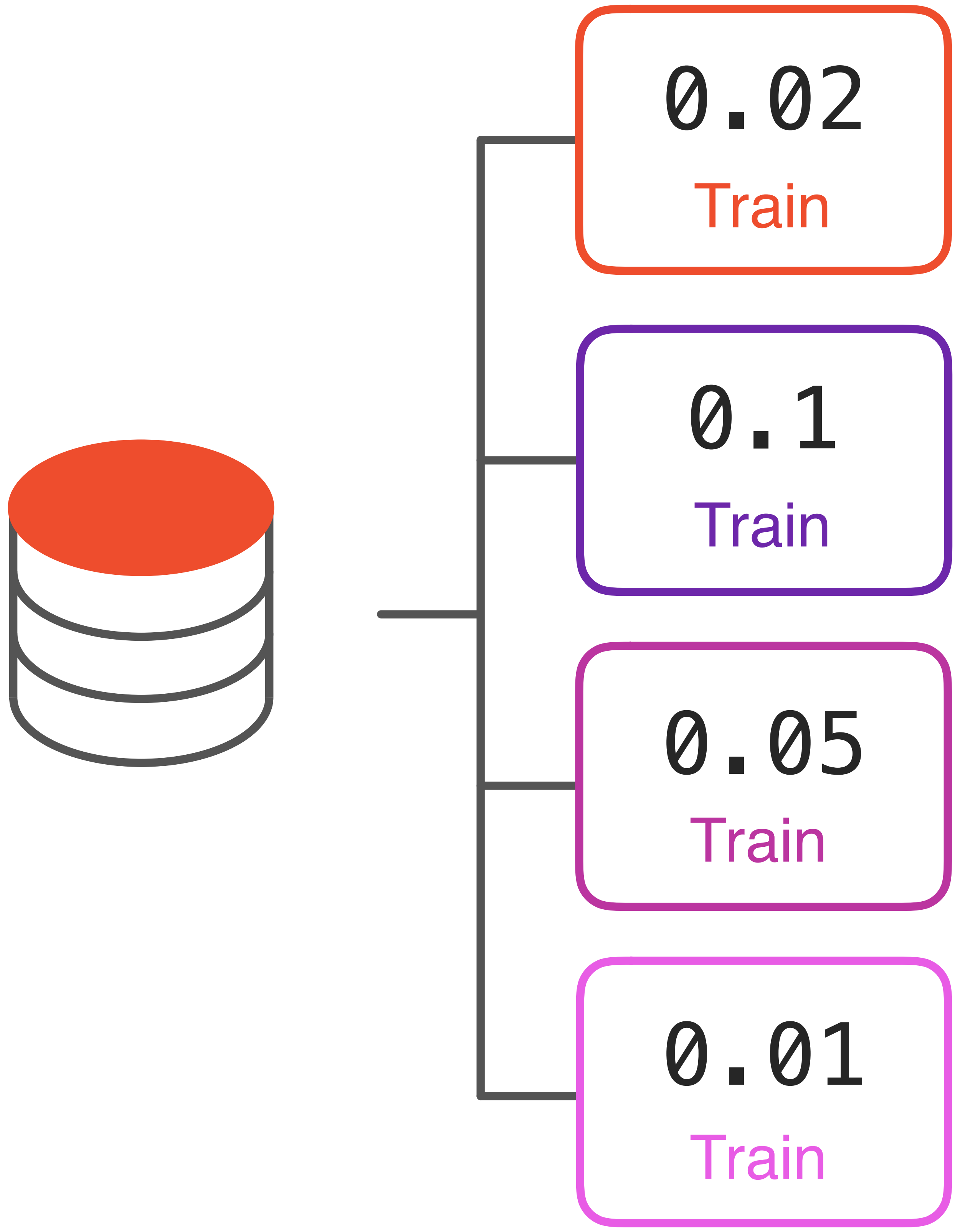
LOW



BAYESOPT AT FACEBOOK | HYPERPARAMETER OPTIMIZATION



BAYESOPT AT FACEBOOK | HYPERPARAMETER OPTIMIZATION



0.02 Train

0.1 Train

0.05 Train

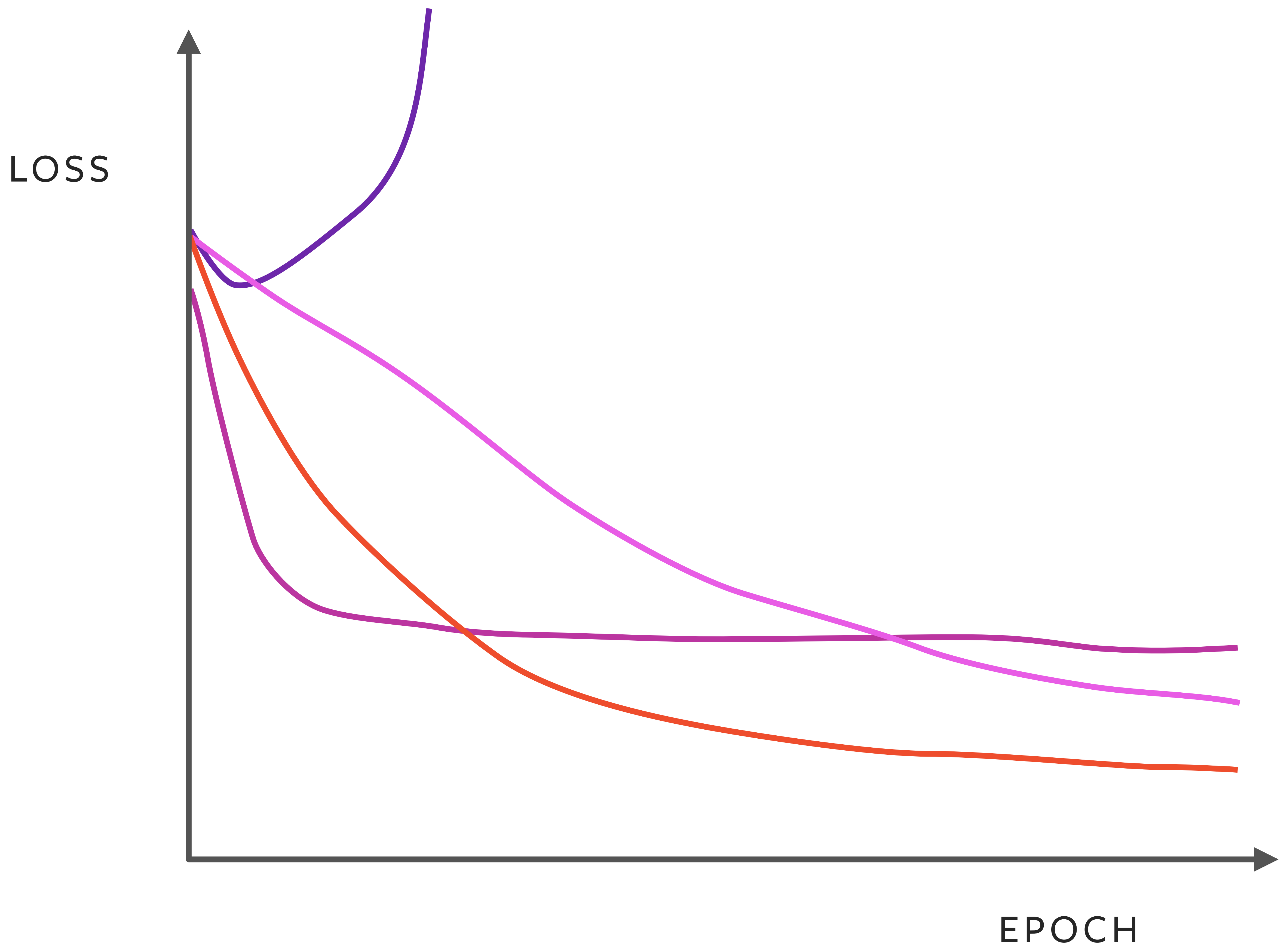
0.01 Train

GOOD

VERY HIGH

HIGH

LOW



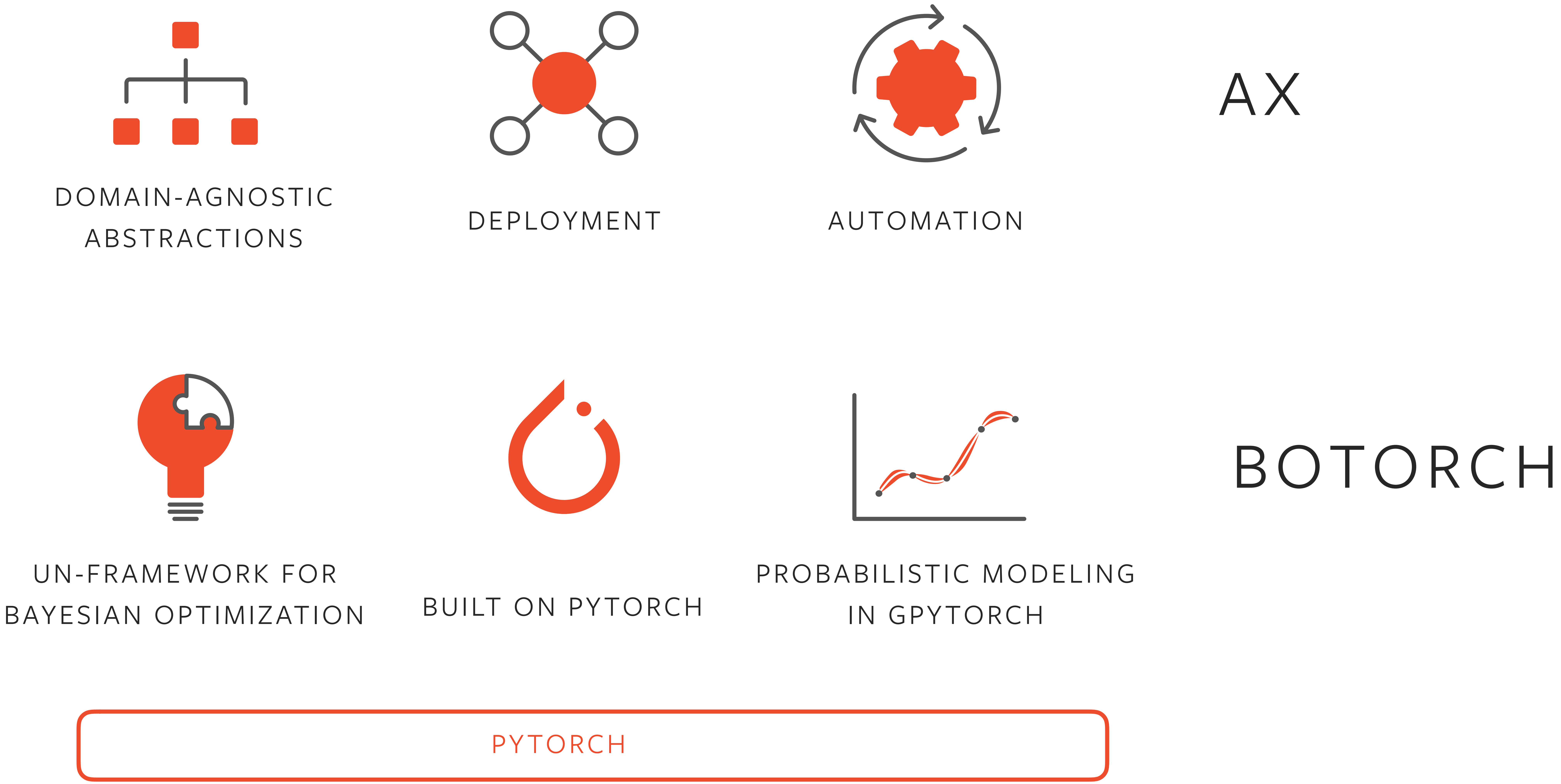
UN-FRAMEWORK FOR
BAYESIAN OPTIMIZATION

BUILT ON PYTORCH

PROBABILISTIC MODELING
IN GPYTORCH

BOTORCH

PYTORCH



AX

DOMAIN-AGNOSTIC ABSTRACTIONS

DEPLOYMENT

AUTOMATION

BOTORCH

UN-FRAMEWORK FOR BAYESIAN OPTIMIZATION

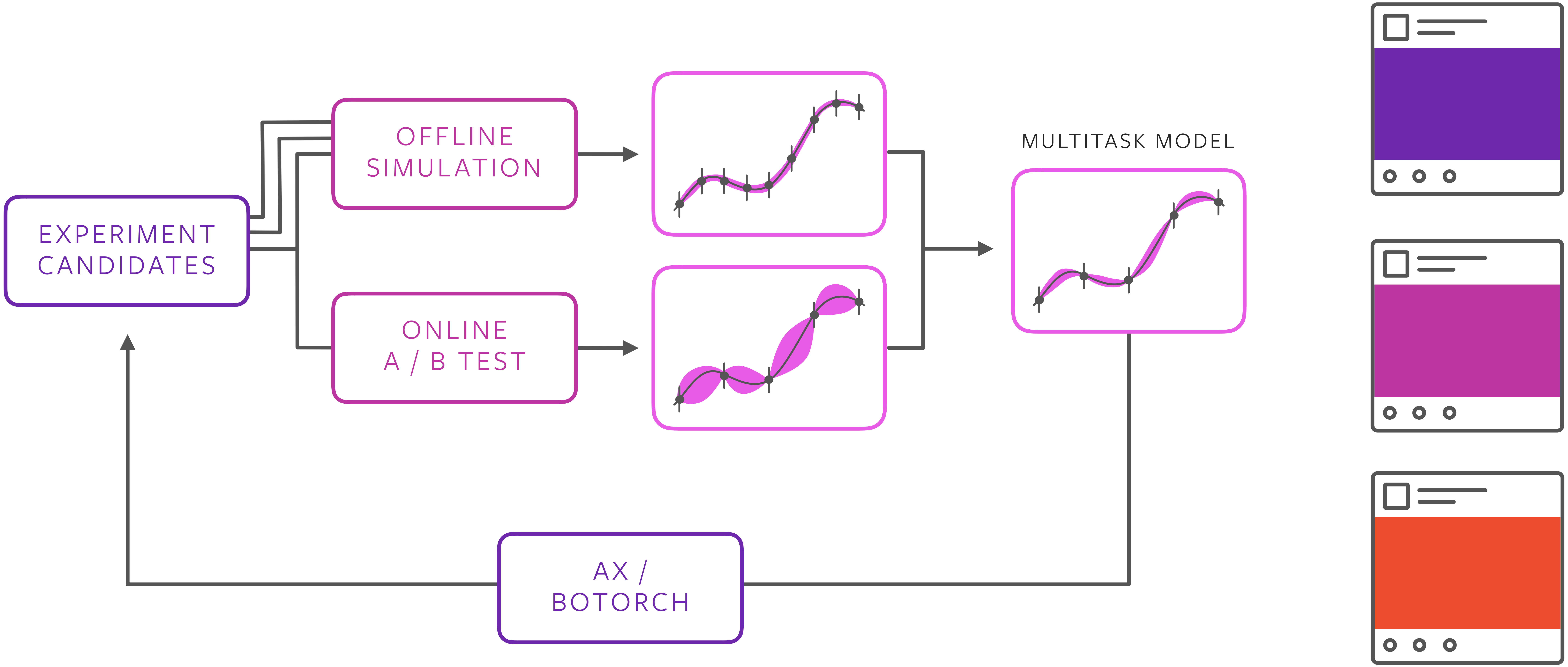
BUILT ON PYTORCH

PROBABILISTIC MODELING IN GPYTORCH

PYTORCH

ADAPTIVE EXPERIMENTATION FOR NEWS FEED RANKING

ADAPTIVE EXPERIMENTATION FOR NEWS FEED RANKING



EXPERIMENT CANDIDATES

OFFLINE SIMULATION

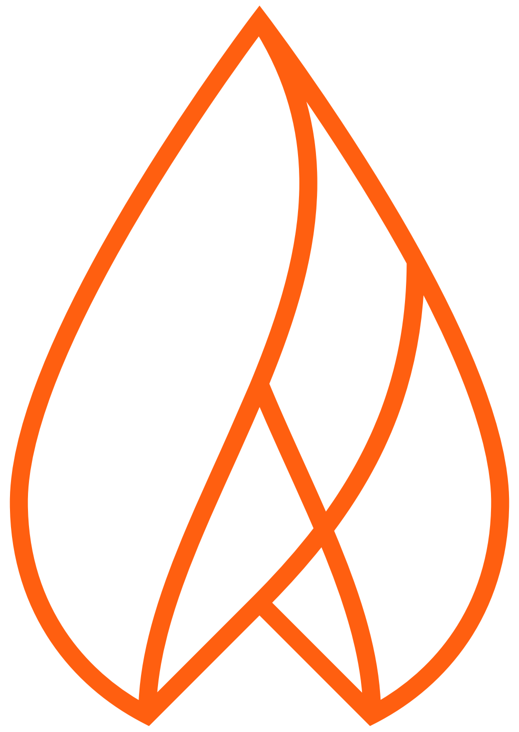
ONLINE A / B TEST

MULTITASK MODEL

AX / BOTORCH

AX

AX.DEV



BOTORCH

BOTORCH.ORG

GIVING BACK TO THE COMMUNITY

GIVING BACK TO THE COMMUNITY

PYTEXT

BOTORCH

PYTEXT

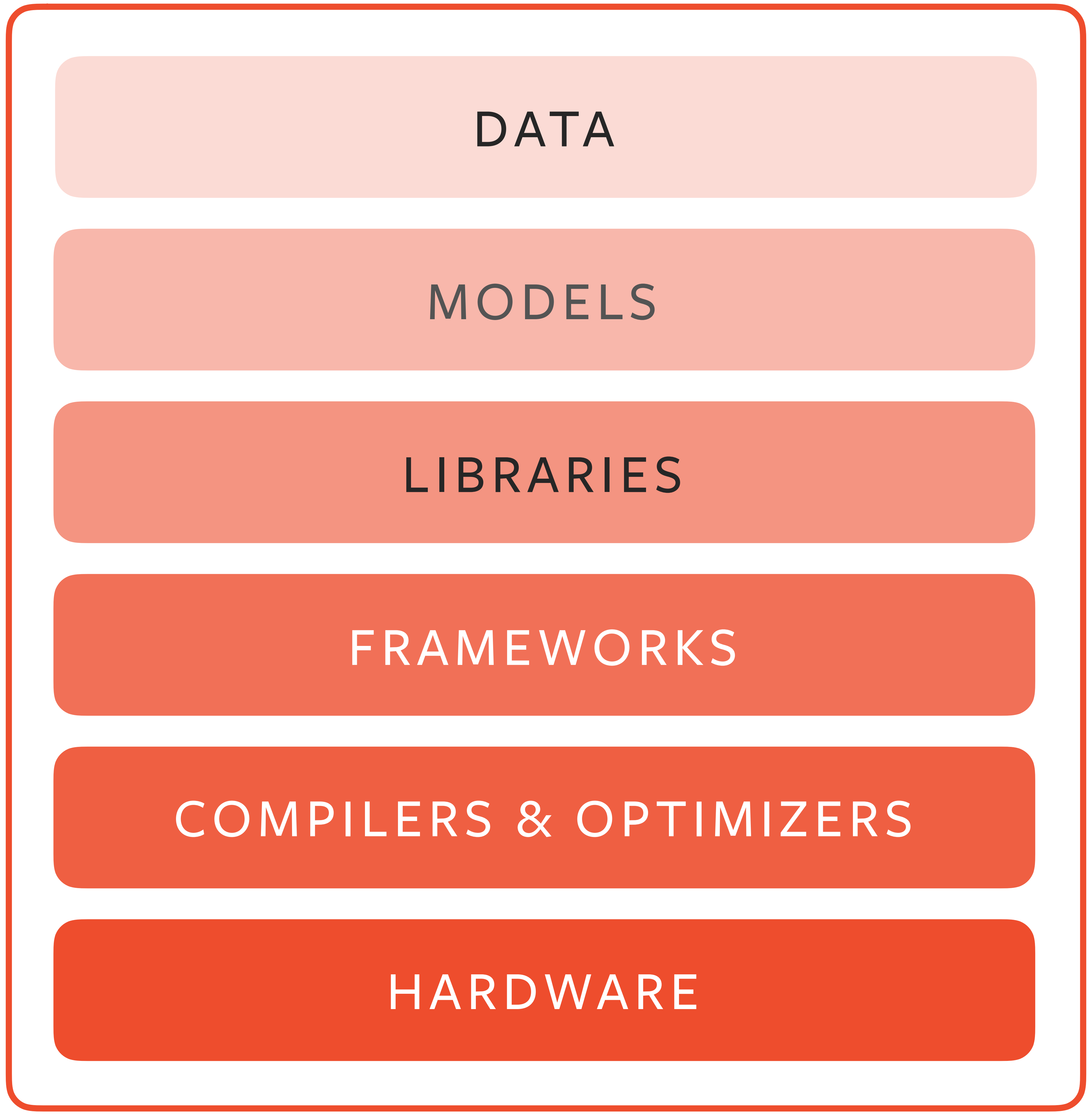
TRANSLATE

HORIZON

DEVELOPER RESOURCES

FULL STACK APPROACH

FULL STACK APPROACH



DATA

MODELS

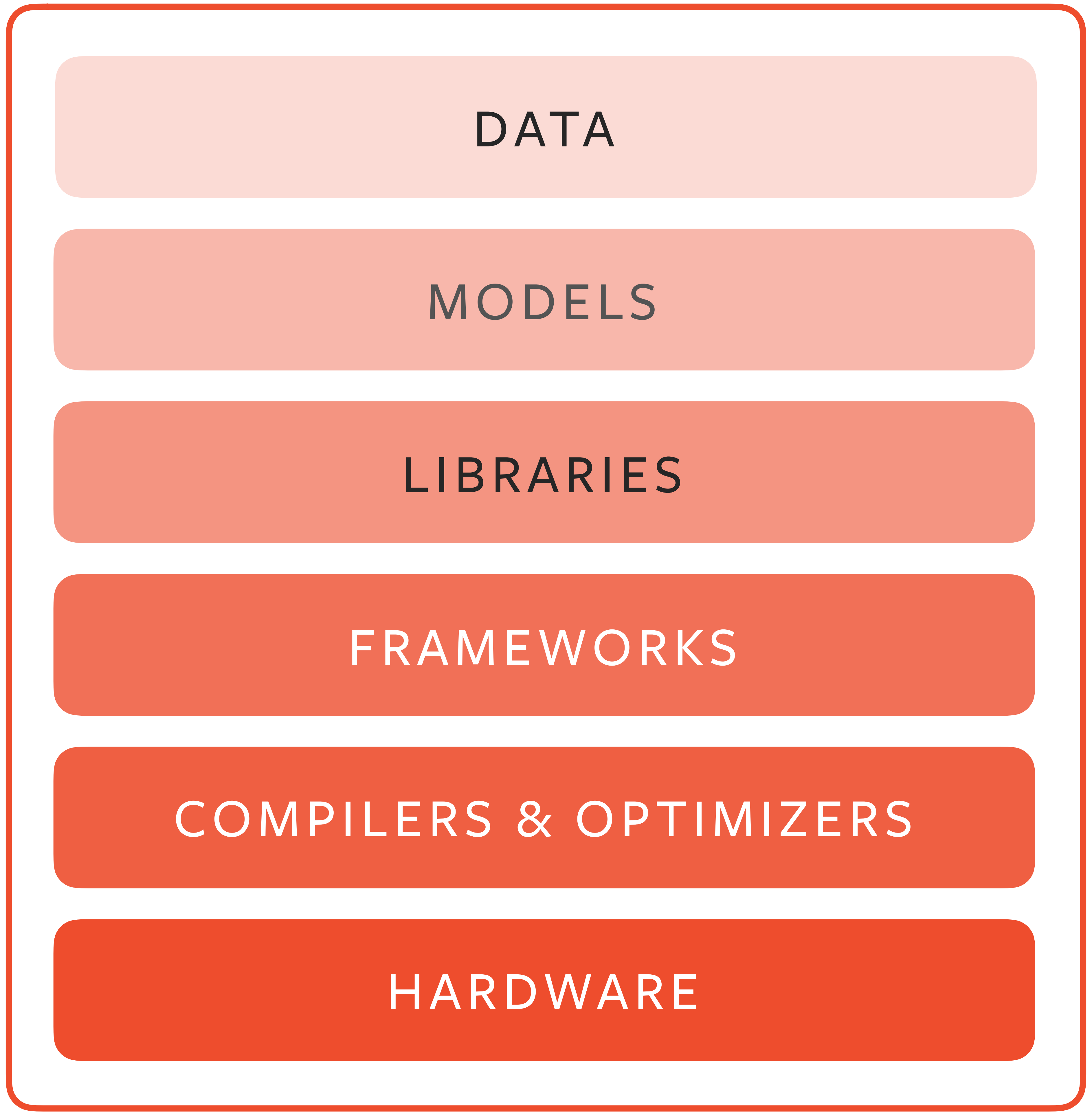
LIBRARIES

FRAMEWORKS

COMPILERS & OPTIMIZERS

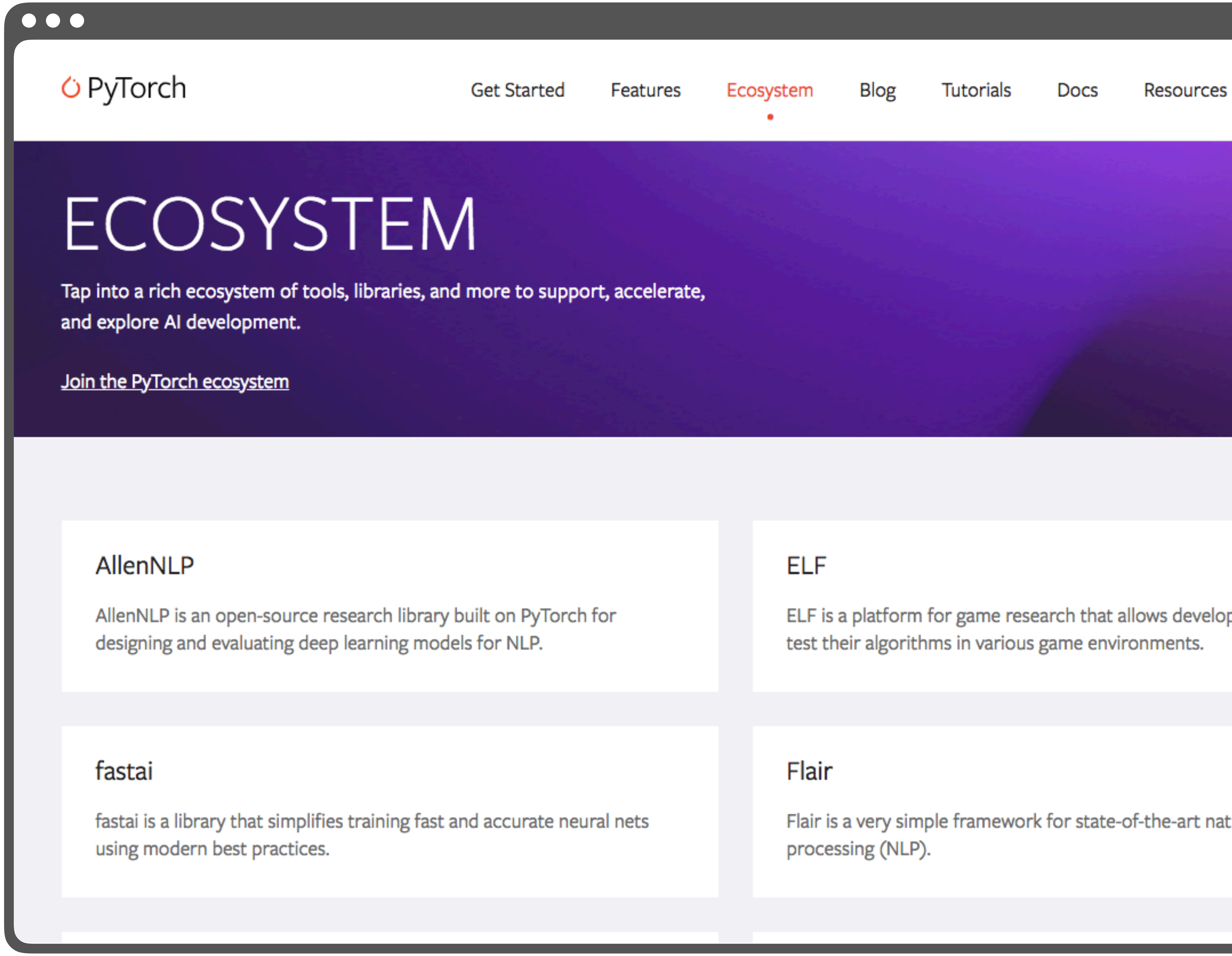
HARDWARE

FULL STACK APPROACH



DATA	HOUSE3D, CLEVER
MODELS	FAI3.5B, RESNEXT3D
LIBRARIES	PYTEXT, HORIZON, DETECTRON, TRANSLATE
FRAMEWORKS	PYTORCH 1.0
COMPILERS & OPTIMIZERS	GLOW, TVM
HARDWARE	ZION, BIG BASIN, TIOGA PASS

PYTORCH ECOSYSTEM



ECOSYSTEM

Tap into a rich ecosystem of tools, libraries, and more to support, accelerate, and explore AI development.

Join the PyTorch ecosystem

AllenNLP

AllenNLP is an open-source research library built on PyTorch for designing and evaluating deep learning models for NLP.

ELF

ELF is a platform for game research that allows develop
test their algorithms in various game environments.

fastai

fastai is a library that simplifies training fast and accurate neural nets using modern best practices.

Flair

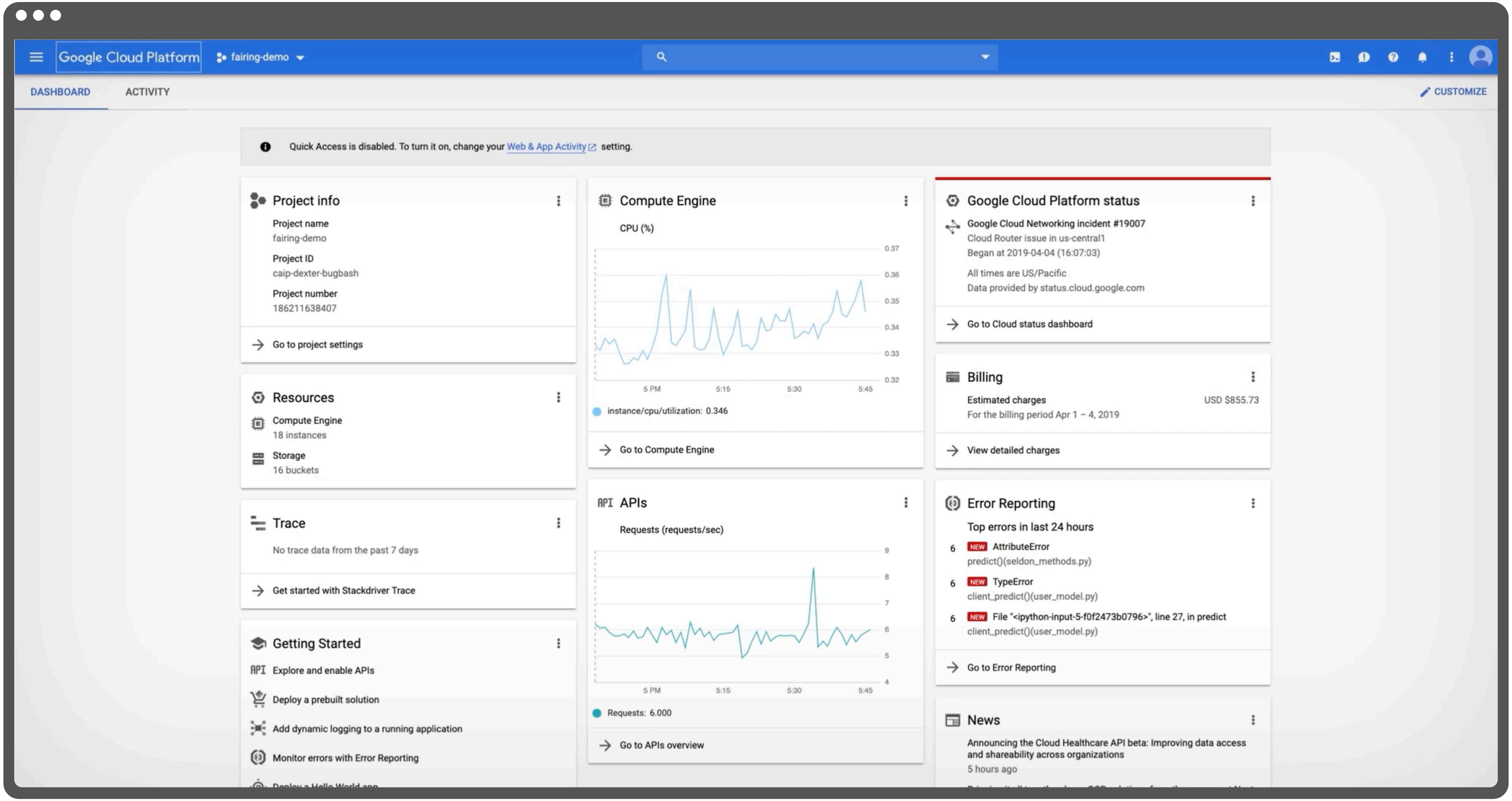
Flair is a very simple framework for state-of-the-art nat
processing (NLP).

EASY ON-BOARDING IN THE CLOUD

AMAZON WEB SERVICES

MICROSOFT AZURE

GOOGLE CLOUD PLATFORM



Quick Access is disabled. To turn it on, change your Web & App Activity setting.

Project info

Project name
fairing-demo

Project ID
caip-dexter-bugbash

Project number
186211638407

→ Go to project settings

Resources

Compute Engine
18 instances

Storage
16 buckets

Trace

No trace data from the past 7 days

→ Get started with Stackdriver Trace

Getting Started

- Explore and enable APIs
- Deploy a prebuilt solution
- Add dynamic logging to a running application
- Monitor errors with Error Reporting
- Deploy a Hello World app

Compute Engine

CPU (%)

● instance/cpu/utilization: 0.346

→ Go to Compute Engine

APIs

Requests (requests/sec)

● Requests: 6.000

→ Go to APIs overview

Google Cloud Platform status

Google Cloud Networking incident #19007
Cloud Router issue in us-central1
Began at 2019-04-04 (16:07:03)

All times are US/Pacific
Data provided by status.cloud.google.com

→ Go to Cloud status dashboard

Billing

Estimated charges USD $855.73
For the billing period Apr 1 – 4, 2019

→ View detailed charges

Error Reporting

Top errors in last 24 hours

6 NEW AttributeError
predict()(seldon_methods.py)

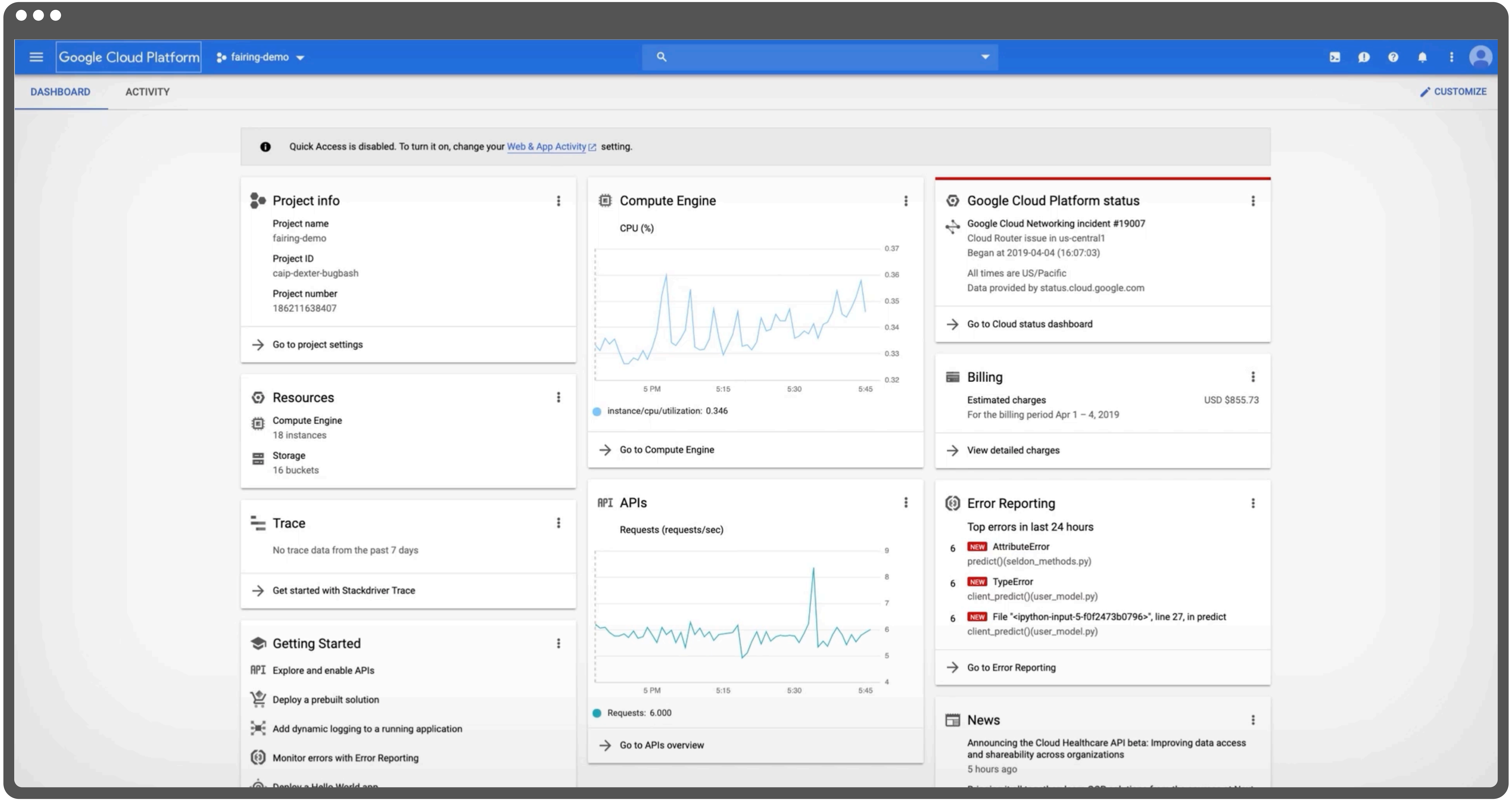
6 NEW TypeError
client_predict()(user_model.py)

6 NEW File "<ipython-input-5-f0f2473b0796>", line 27, in predict
client_predict()(user_model.py)

→ Go to Error Reporting

News

Announcing the Cloud Healthcare API beta: Improving data access and shareability across organizations
5 hours ago



Quick Access is disabled. To turn it on, change your Web & App Activity setting.

Project info

Project name
fairing-demo

Project ID
caip-dexter-bugbash

Project number
186211638407

Go to project settings

Resources

Compute Engine
18 instances

Storage
16 buckets

Trace

No trace data from the past 7 days

Get started with Stackdriver Trace

Getting Started

- Explore and enable APIs
- Deploy a prebuilt solution
- Add dynamic logging to a running application
- Monitor errors with Error Reporting
- Deploy a Hello World app

Compute Engine

CPU (%)

instance/cpu/utilization: 0.346

Go to Compute Engine

APIs

Requests (requests/sec)

Requests: 6.000

Go to APIs overview

Google Cloud Platform status

Google Cloud Networking incident #19007
Cloud Router issue in us-central1
Began at 2019-04-04 (16:07:03)

All times are US/Pacific
Data provided by status.cloud.google.com

Go to Cloud status dashboard

Billing

Estimated charges	USD $855.73
For the billing period Apr 1 – 4, 2019	

View detailed charges

Error Reporting

Top errors in last 24 hours

6 NEW AttributeError
predict()(seldon_methods.py)

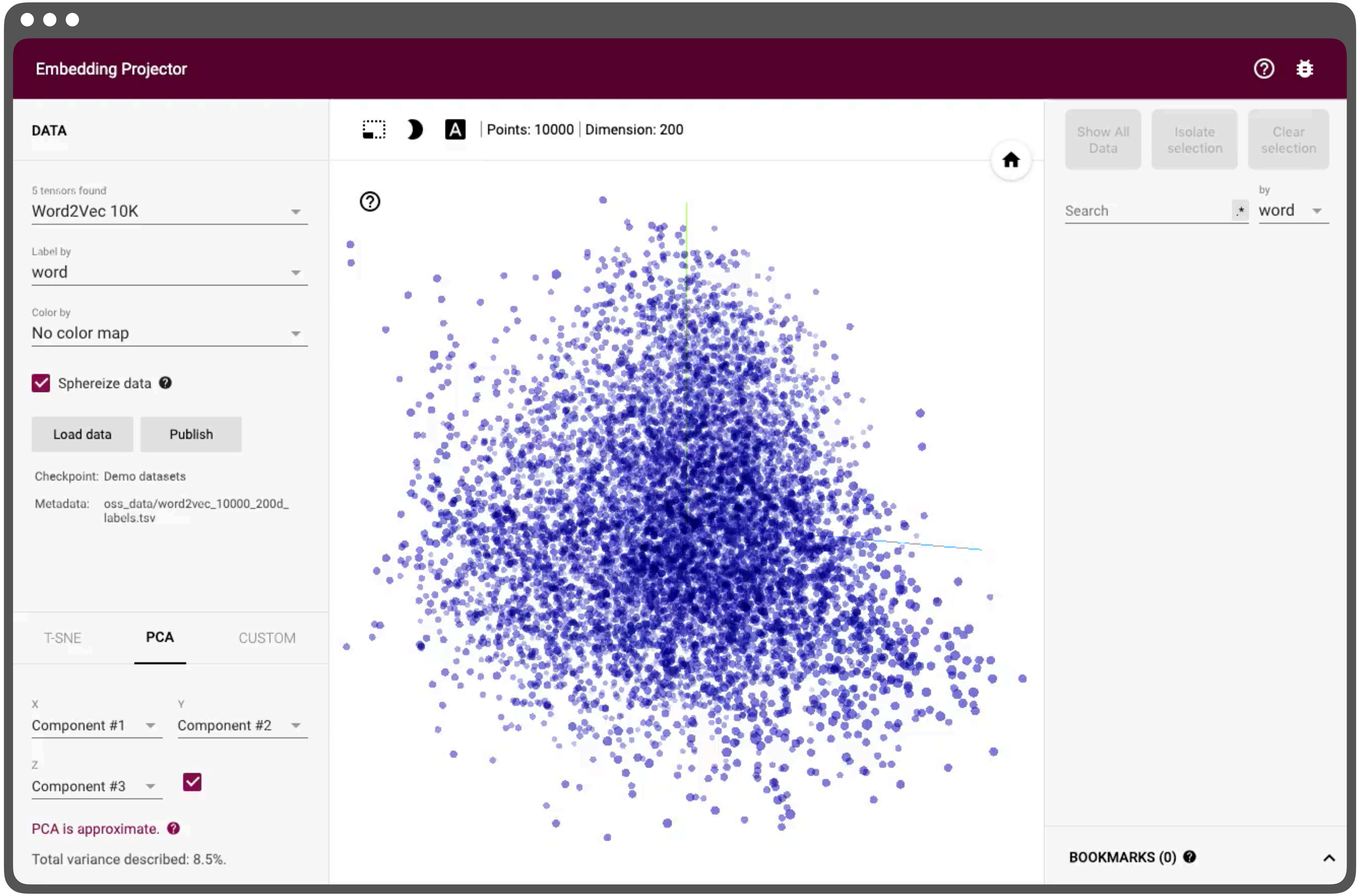
6 NEW TypeError
client_predict()(user_model.py)

6 NEW File "<ipython-input-5-f0f2473b0796>", line 27, in predict
client_predict()(user_model.py)

Go to Error Reporting

News

Announcing the Cloud Healthcare API beta: Improving data access and shareability across organizations
5 hours ago



Points: 10000 | Dimension: 200

Checkpoint: Demo datasets

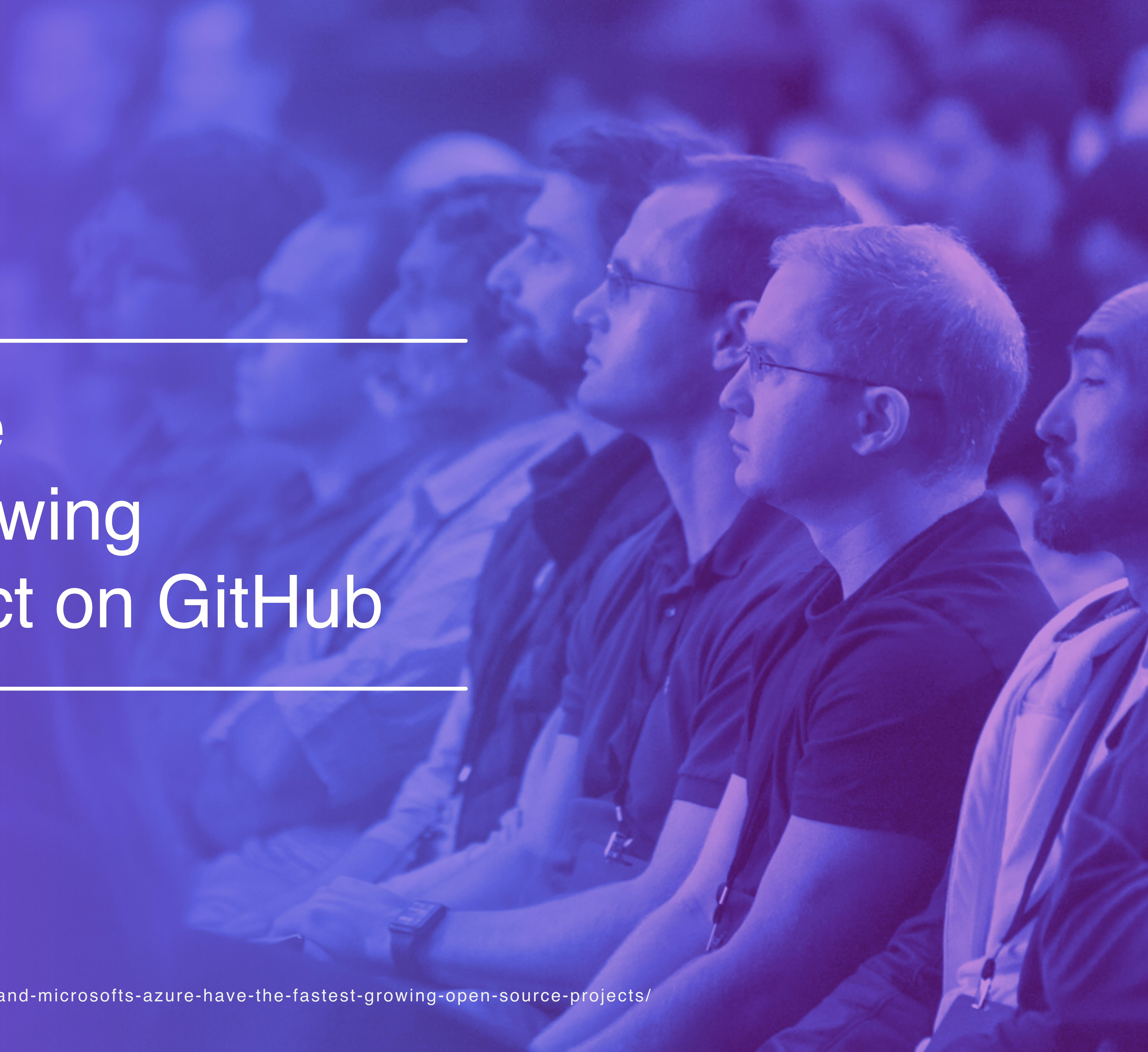
Metadata: oss_data/word2vec_10000_200d_labels.tsv

PCA is approximate.

Total variance described: 8.5%.

Thanks to Nick Felt (Google), William Chargin (Google), Mani Varadarajan (Google), Orion Reblitz-Richardson (FB), Tzu-Wei Huang

EDUCATION



PyTorch is now the second-fastest growing open source project on GitHub

Source: venturebeat.com/2018/10/16/github-facebooks-pytorch-and-microsofts-azure-have-the-fastest-growing-open-source-projects/

O'REILLY
Natural Language Processing with PyTorch
Build Intelligent Language A
Using Deep Learning
Deep Learning with PyTorch
Eli Stevens
Luca Antiga
MEAP
MANNING

UDACITY

UDACITY

✓ PRIVACY IN AI - FREE COURSE

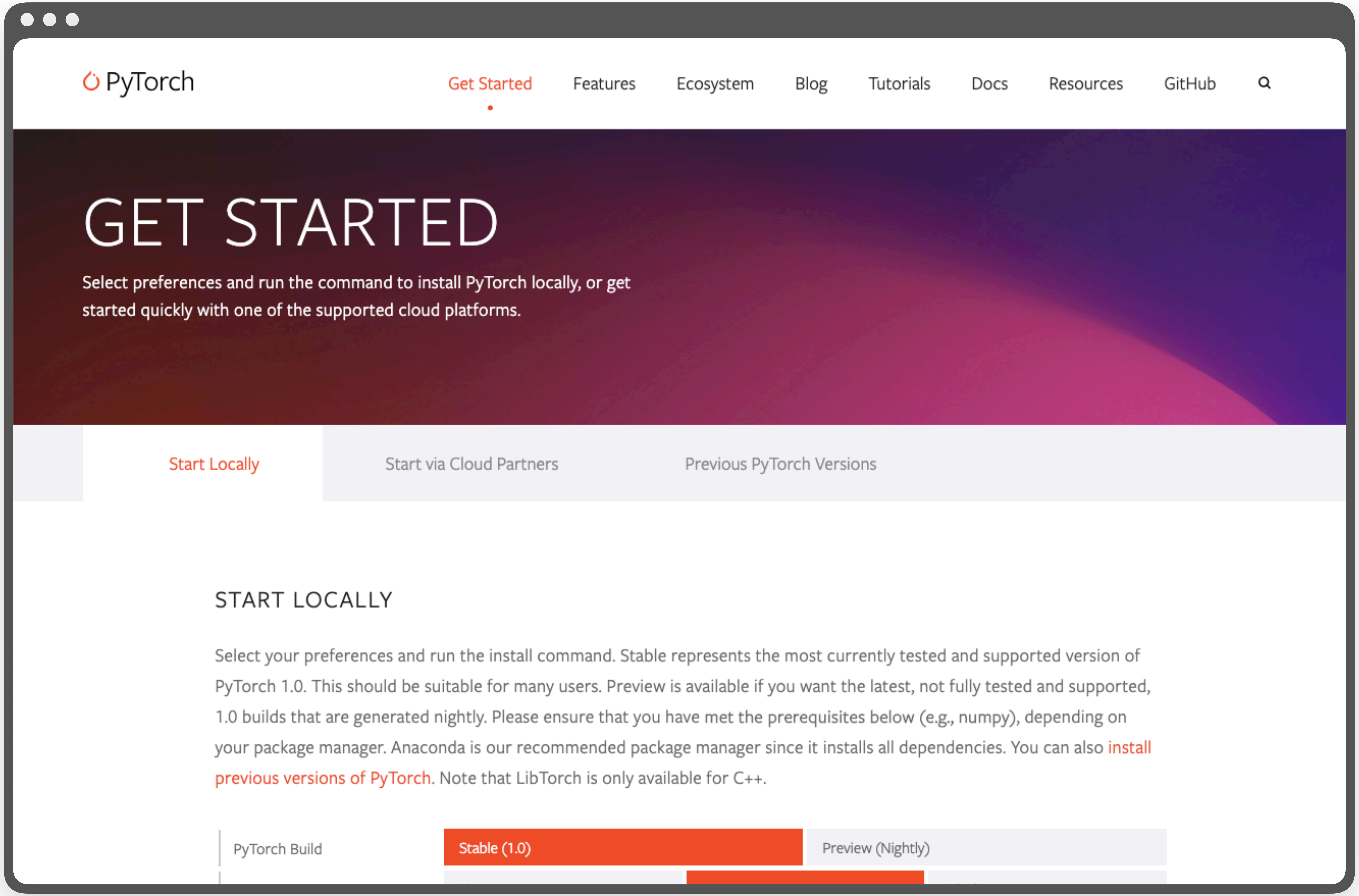
✓ FACEBOOK FUNDING 300 SCHOLARSHIPS TO CONTINUE EDUCATION

✓ FULL CURRICULUM OF DL COURSES TAUGHT WITH PYTORCH

SOME OF THE CLASSES BASED ON PYTORCH:

fast.ai

- ✓ OVER 1.5 MILLION MINUTES VIEWING TIME PER MONTH
- ✓ 100K MONTHLY VIDEO VIEWS
- ✓ GLOBAL COMMUNITY OF STUDENTS FROM THE US TO BENGALURU TO LAGOS



GET STARTED

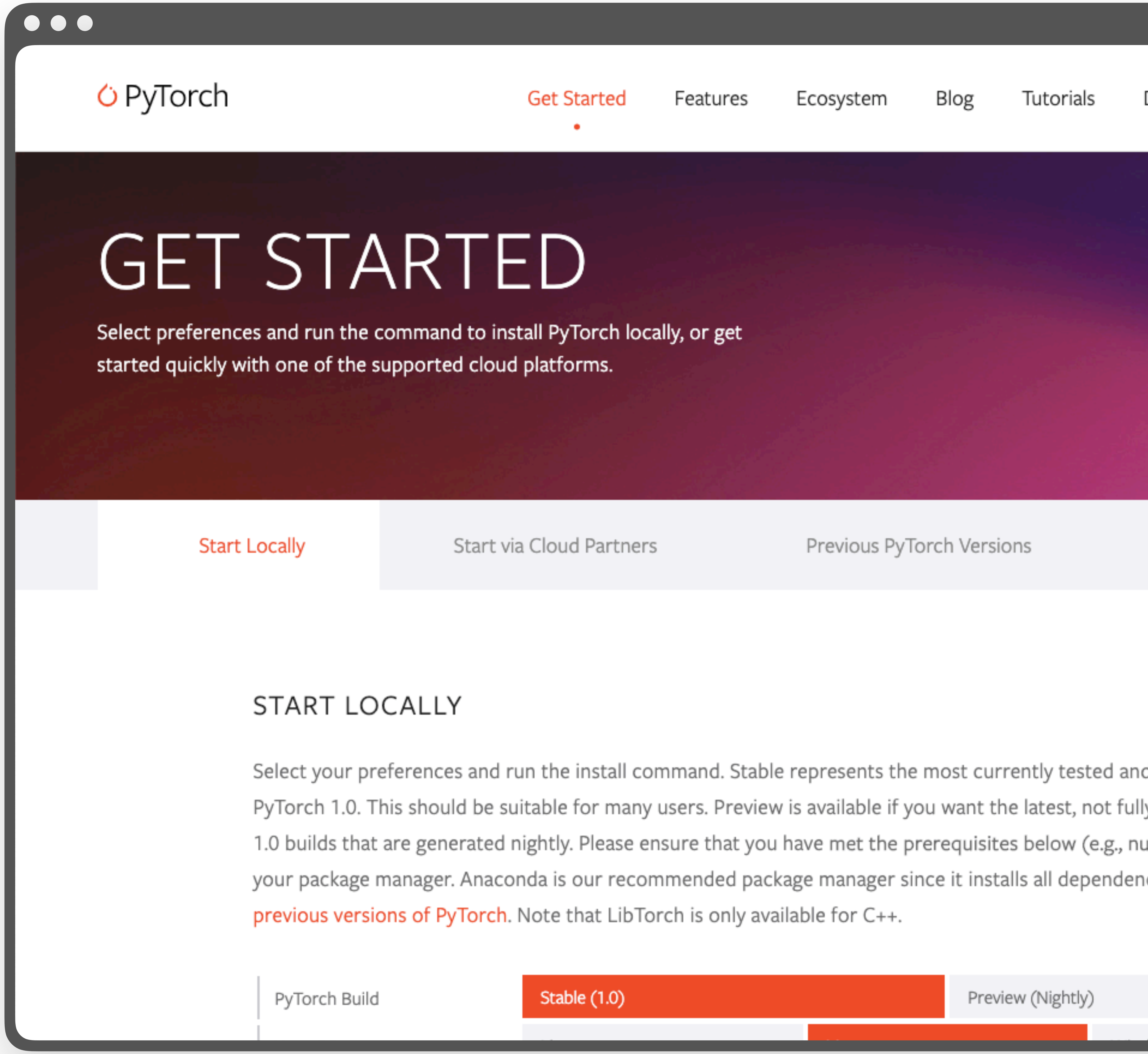
Select preferences and run the command to install PyTorch locally, or get started quickly with one of the supported cloud platforms.

START LOCALLY

Select your preferences and run the install command. Stable represents the most currently tested and supported version of PyTorch 1.0. This should be suitable for many users. Preview is available if you want the latest, not fully tested and supported, 1.0 builds that are generated nightly. Please ensure that you have met the prerequisites below (e.g., numpy), depending on your package manager. Anaconda is our recommended package manager since it installs all dependencies. You can also install previous versions of PyTorch. Note that LibTorch is only available for C++.

PyTorch Build: Stable (1.0)

PYTORCH



GET STARTED

Select preferences and run the command to install PyTorch locally, or get started quickly with one of the supported cloud platforms.

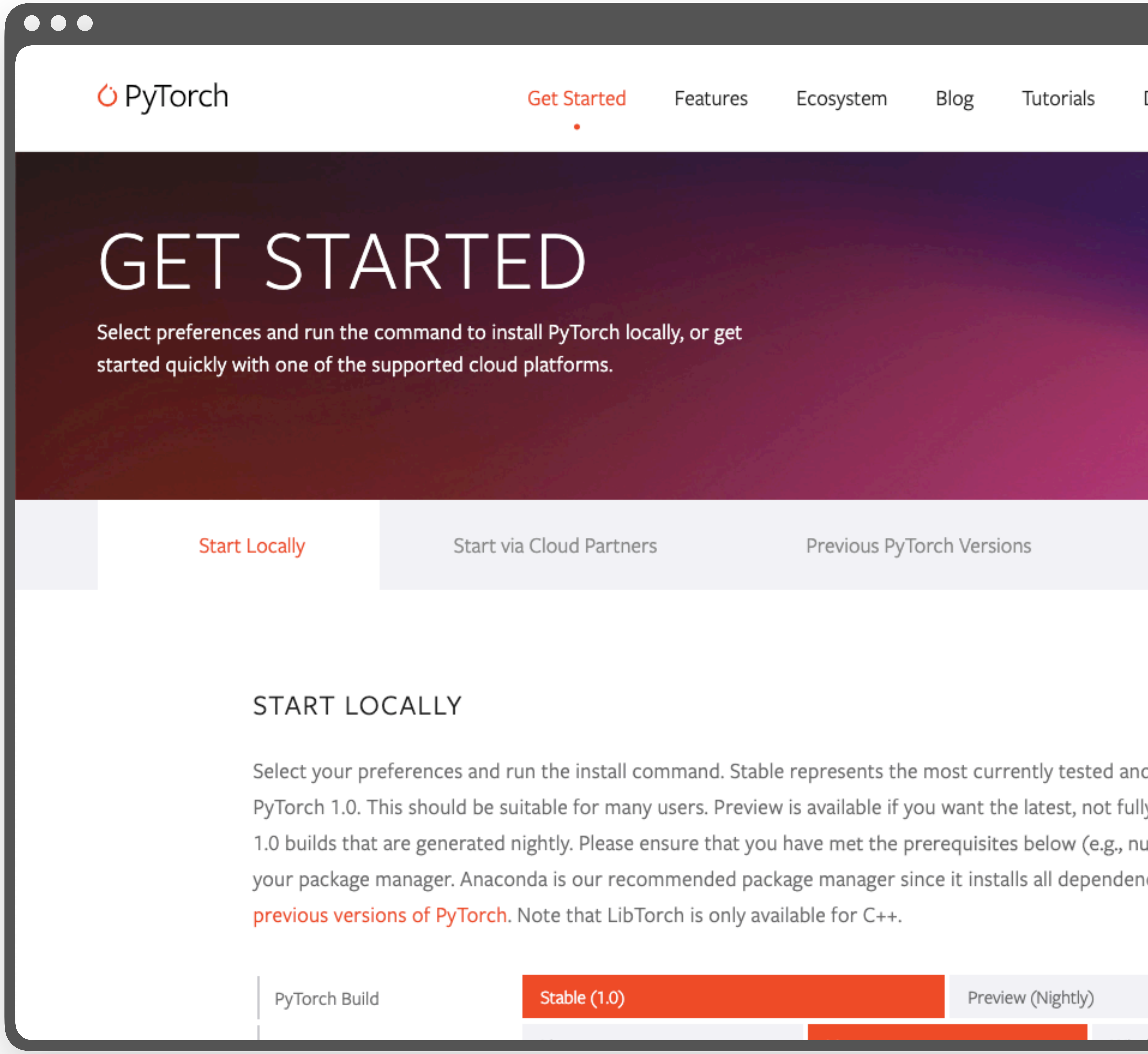
START LOCALLY

Select your preferences and run the install command. Stable represents the most currently tested and PyTorch 1.0. This should be suitable for many users. Preview is available if you want the latest, not fully 1.0 builds that are generated nightly. Please ensure that you have met the prerequisites below (e.g., nu your package manager. Anaconda is our recommended package manager since it installs all dependen previous versions of PyTorch. Note that LibTorch is only available for C++.

PyTorch Build	Stable (1.0)	Preview (Nightly)

PYTORCH

- ✓ EXPERIENCE INSTANTLY ON COLAB
- ✓ GET STARTED ON THE CLOUD
- ✓ INSTALL LOCALLY



GET STARTED

Select preferences and run the command to install PyTorch locally, or get started quickly with one of the supported cloud platforms.

START LOCALLY

Select your preferences and run the install command. Stable represents the most currently tested anc PyTorch 1.0. This should be suitable for many users. Preview is available if you want the latest, not fully 1.0 builds that are generated nightly. Please ensure that you have met the prerequisites below (e.g., nu your package manager. Anaconda is our recommended package manager since it installs all dependen previous versions of PyTorch. Note that LibTorch is only available for C++.

PyTorch Build	Stable (1.0)	Preview (Nightly)

JOIN THE PYTORCH DEVELOPER COMMUNITY

FACEBOOK.COM/PYTORCH

TWITTER.COM/PYTORCH

PYTORCH.ORG