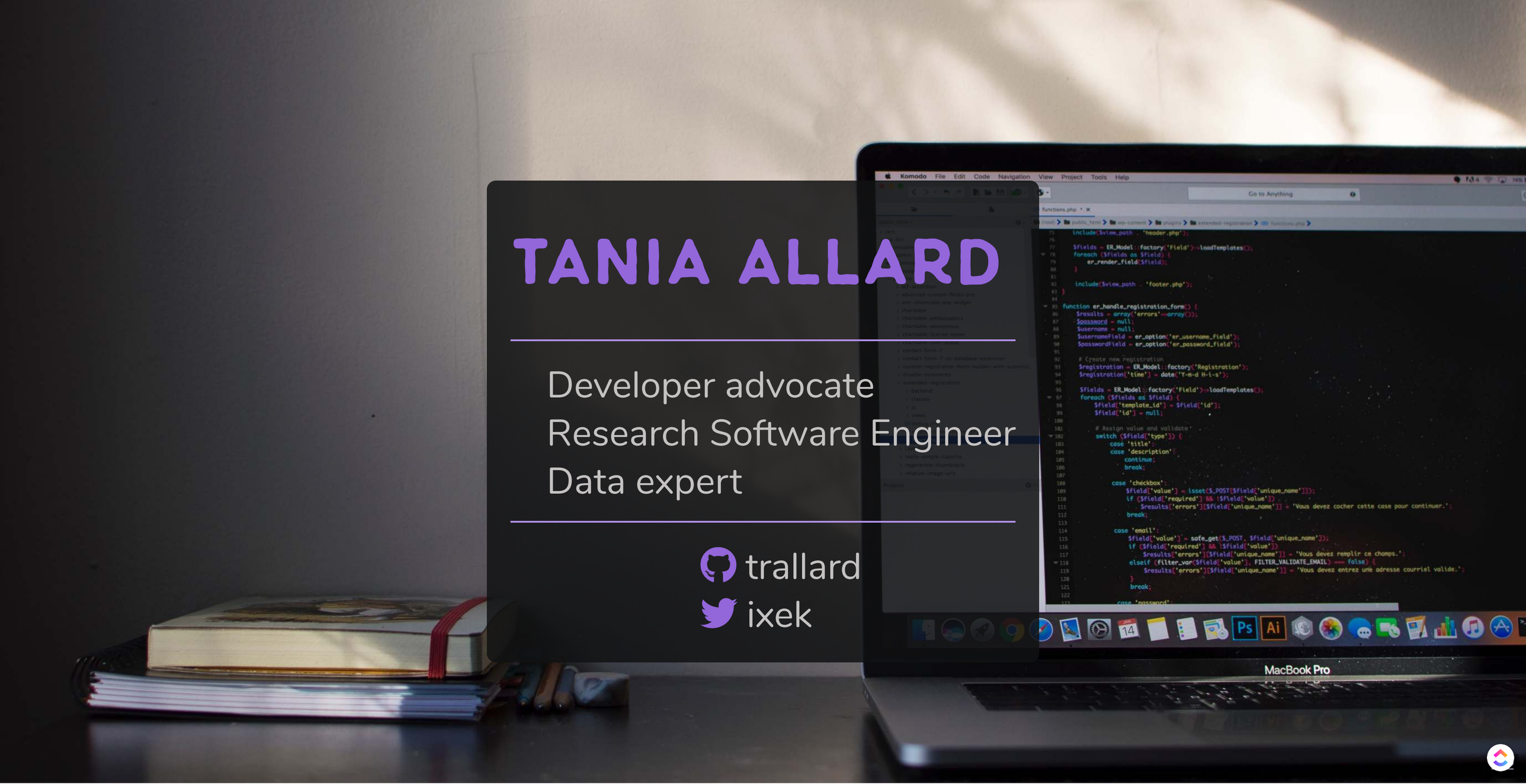

TUX

https://trallard.github.io/Talks/RSE-sheffield

THE STATE OF MACHINE LEARNING

RSE seminar, University of Sheffield

Tania Allard, PhD



TANIA ALLARD

Developer advocate
Research Software Engineer
Data expert

trallard
ixek

ixek

MACHINE LEARNING EVERYWHERE

ixek

MACHINE LEARNING EVERYWHERE

So much that it is starting to not make sense anymore... like when you say a word 50 times in a row

ixek

For good or for bad it is everywhere:

ixek

For good or for bad it is everywhere:

- Deployed in healthcare and warfare

ixek

For good or for bad it is everywhere:

- Deployed in healthcare and warfare
- In the creative industry (from music to books)

ixek

For good or for bad it is everywhere:

- Deployed in healthcare and warfare
- In the creative industry (from music to books)
- Reading CVs and judging your creditworthiness

ixek

For good or for bad it is everywhere:

- Deployed in healthcare and warfare
- In the creative industry (from music to books)
- Reading CVs and judging your creditworthiness
- Making us more Instagram worthy

 ixek

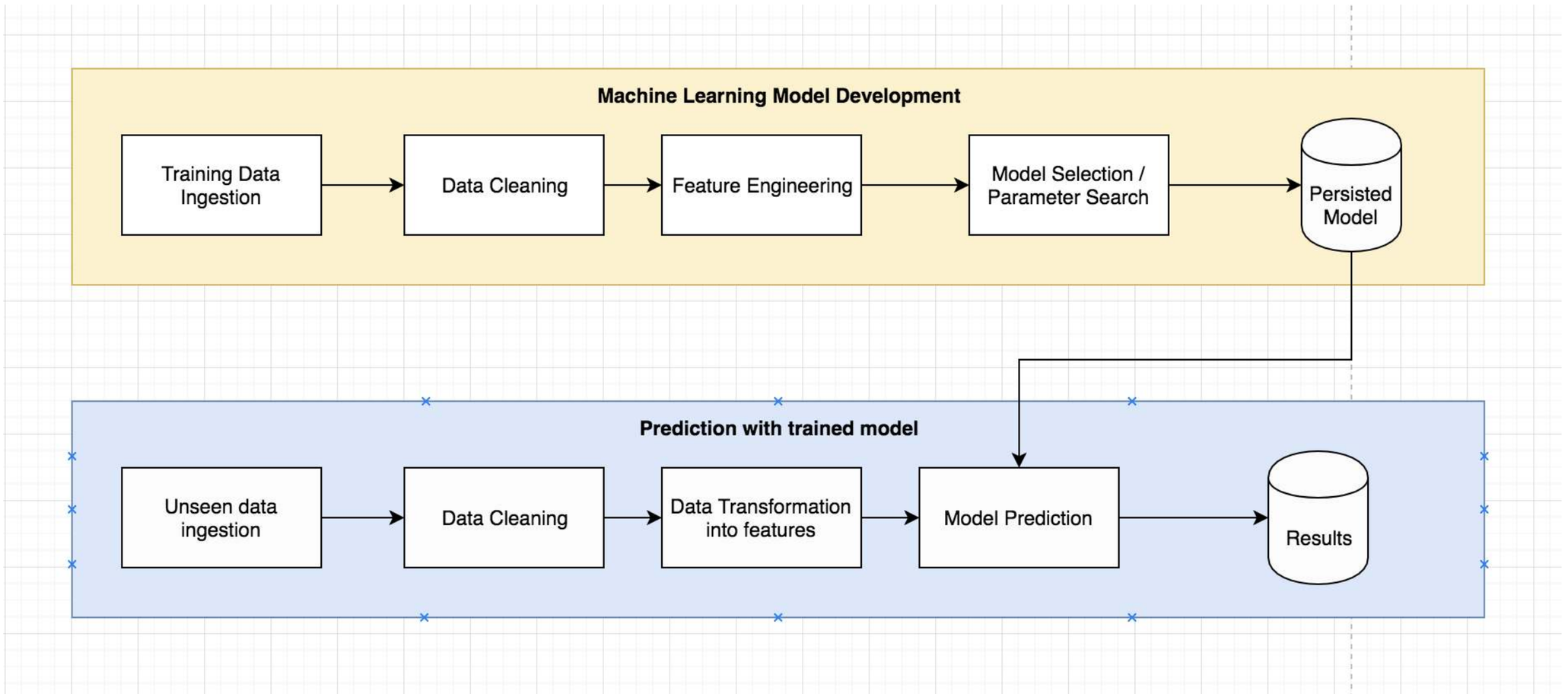
The big players:

- Apple
- Facebook
- Google
- IBM
- Intel
- Microsoft
- Nvidia
- Open AI
- Twitter

 ixek

MACHINE LEARNING GENERALISED IN TWO WORKFLOWS

- Model development (R&D)
- Model serving (production for customers consumption)


Machine Learning Model Development
Training Data Ingestion
Data Cleaning
Feature Engineering
Model Selection / Parameter Search
Persisted Model
Prediction with trained model
Unseen data ingestion
Data Cleaning
Data Transformation into features
Model Prediction
Results

ixek

WHAT ARE THESE GIANTS' ISSUES?

ixek

WHAT ARE THESE GIANTS' ISSUES?

Mainly scale...in multiple areas

ixek

If we have a small team we have a smaller number of issues... right?

ixek

If we have a small team we have a smaller number of issues... right?

- Small number of models to maintain

ixek

If we have a small team we have a smaller number of issues... right?

- Small number of models to maintain
- People have the knowledge in their heads

ixek

If we have a small team we have a smaller number of issues... right?

- Small number of models to maintain
- People have the knowledge in their heads
- They have their own methods to track progress

ixek

THAT IS THE SMALL TEAM PERFORMANCE FALLACY

We still need processes and best practices in place... so let me get back at this later

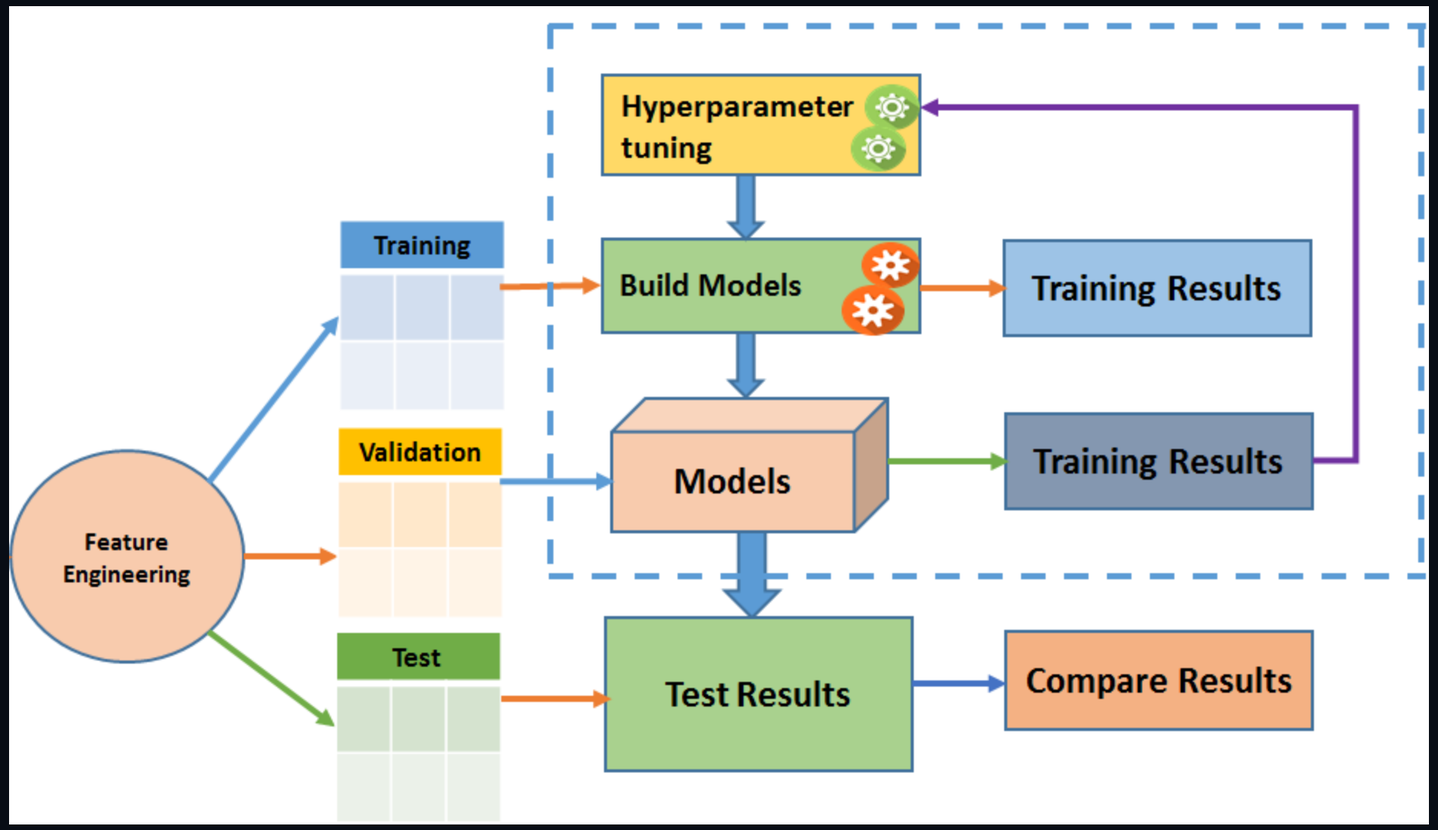
ixek

AS THE TEAM DEMAND GROWS THE PROBLEMS GROW

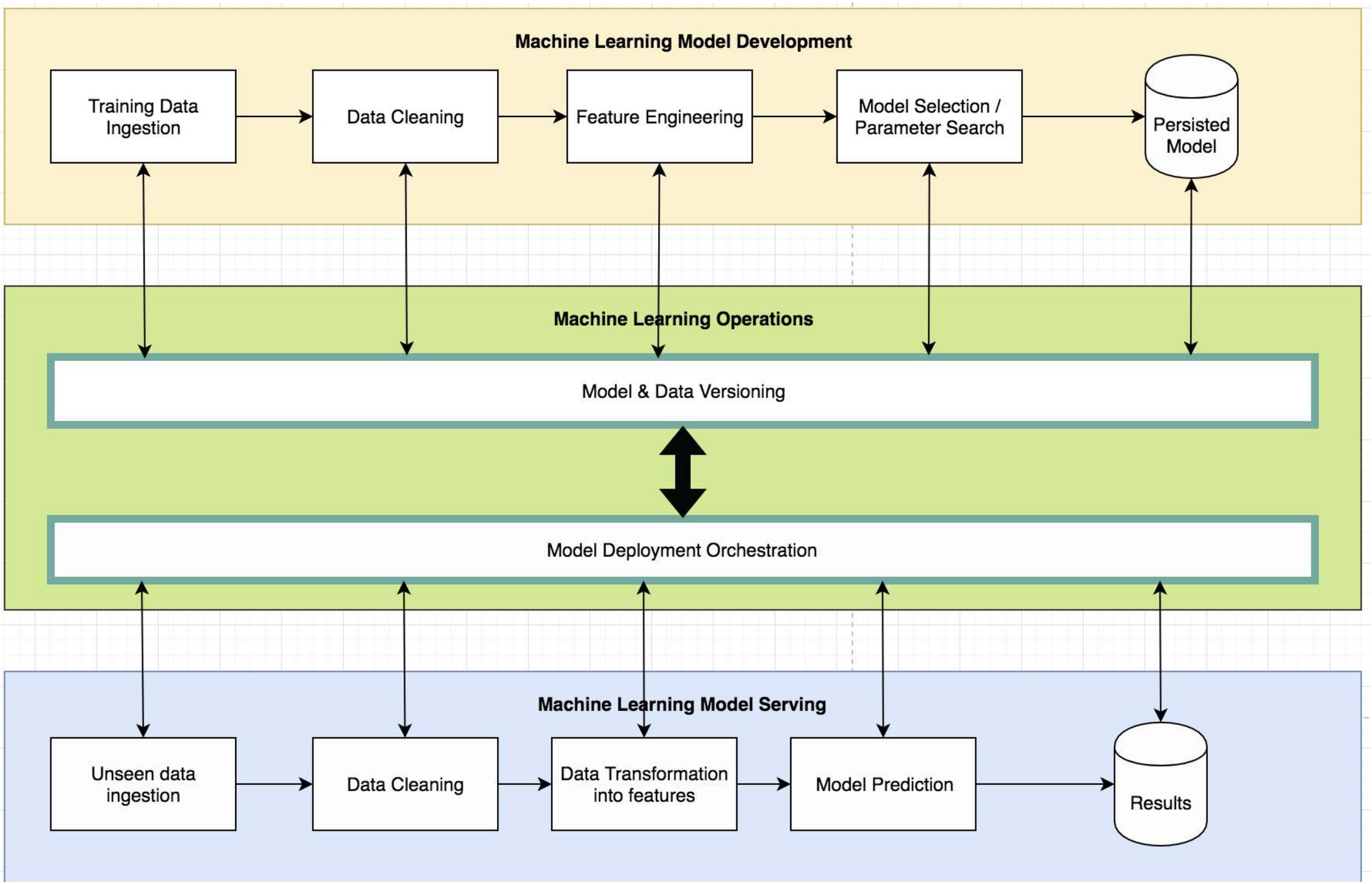
- Increased complexity of data flow
- Larger number of workflows
- Managing complexity of flows and scheduling becomes a nightmare
- Resource allocation has to be on point

ixek

SERVING MODELS BECOMES HARDER



ixek


Machine Learning Model Development
Training Data Ingestion
Data Cleaning
Feature Engineering
Model Selection / Parameter Search
Persisted Model
Machine Learning Operations
Model & Data Versioning
Model Deployment Orchestration
Machine Learning Model Serving
Unseen data ingestion
Data Cleaning
Data Transformation into features
Model Prediction
Results

ixek

HOW DO THEY SERVE MILLIONS OF

CUSTOMERS ACROSS THE GLOBE?

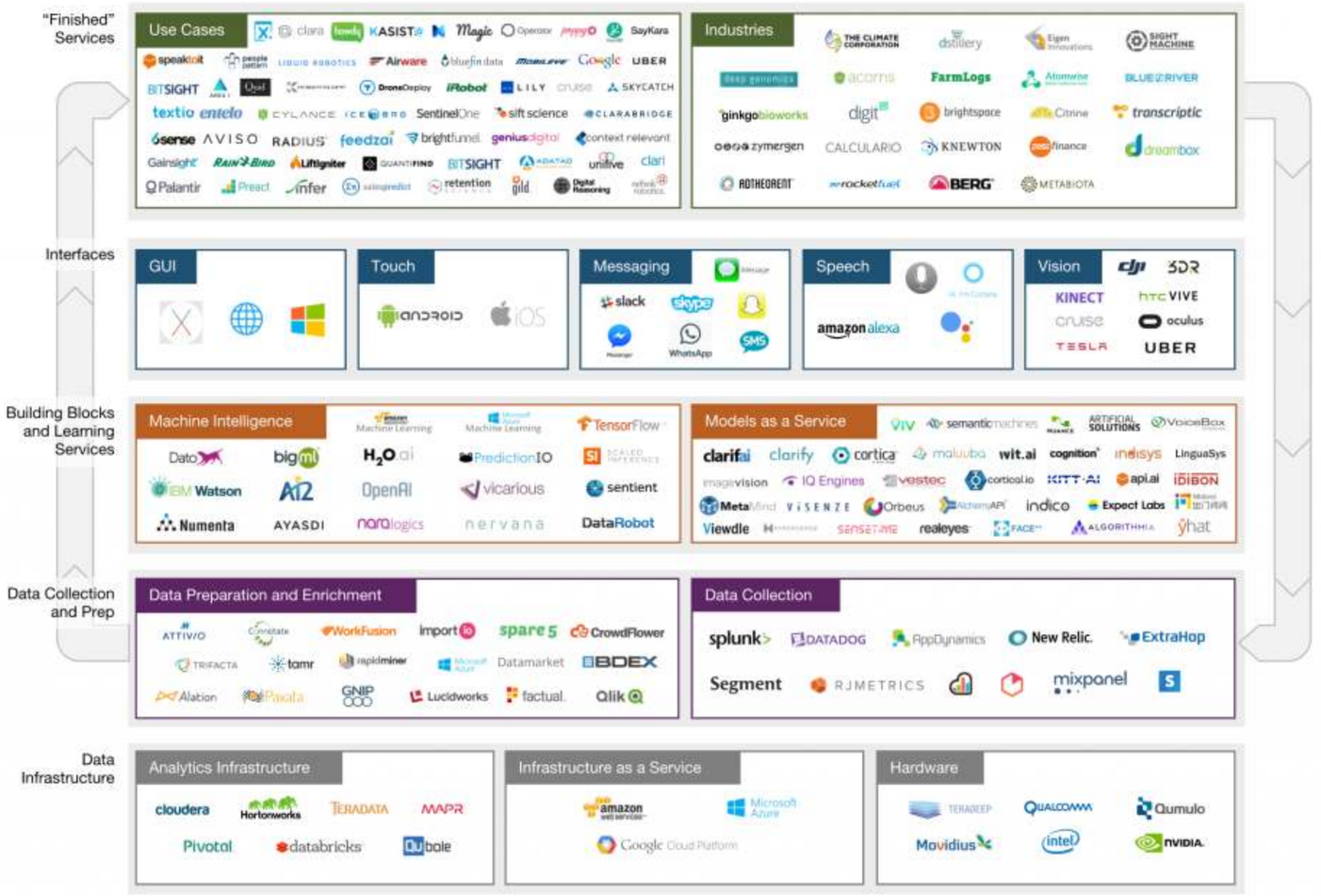
ixek

Three main players:

- Infrastructure / resources
- Processes
- People

ixek

Intelligent App Stack



Created by @MattMcIlwain, @TmPorter, @SSomasegar, @DanielxLi

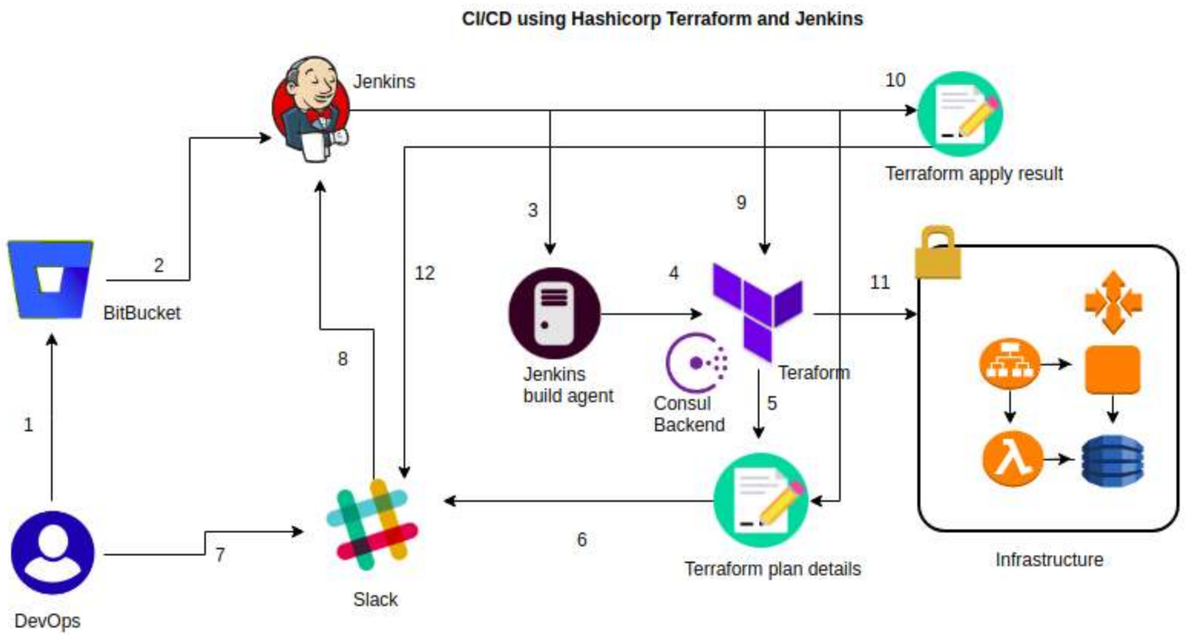
MADRONA VENTURE GROUP

ixek

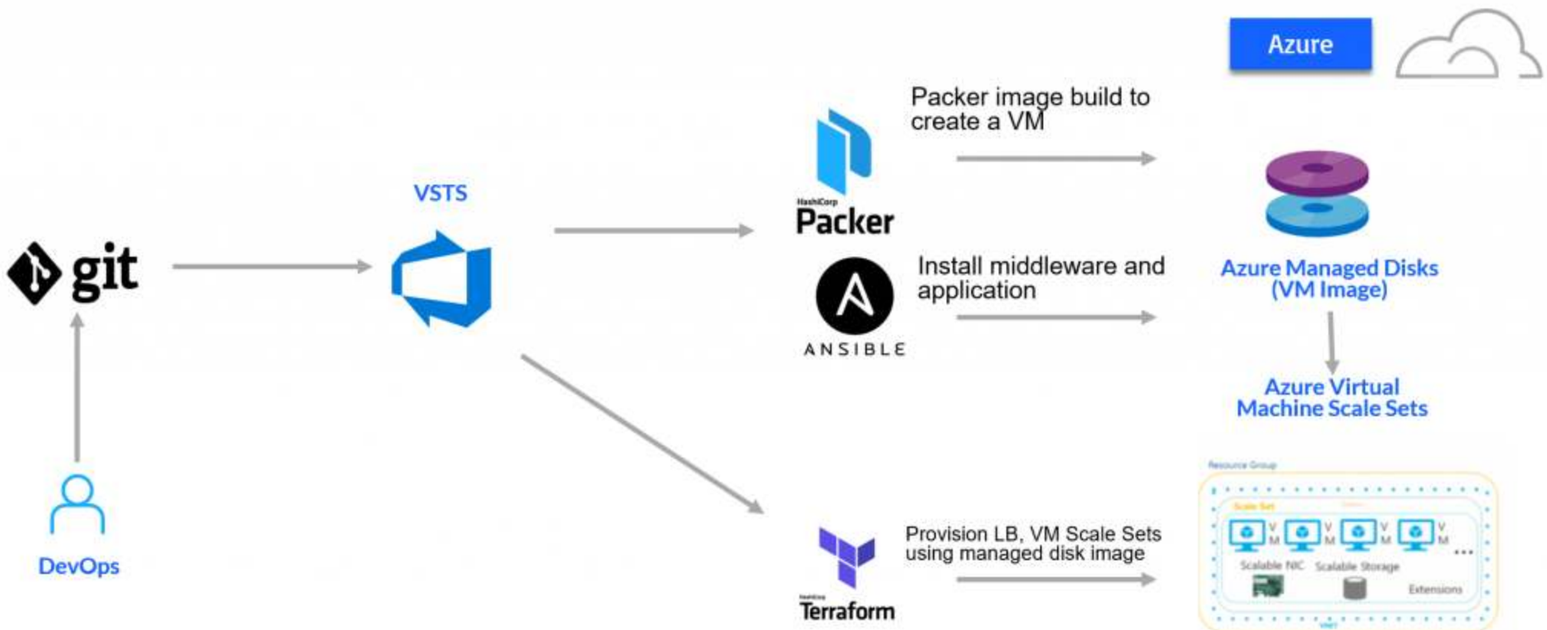
Analytics & Machine Learning

ixek

INFRASTRUCTURE AS A CODE



ixek


Azure
Packer image build to create a VM
VSTS
Packer
Install middleware and application
ANSIBLE
Azure Managed Disks (VM Image)
Azure Virtual Machine Scale Sets
git
DevOps
Provision LB, VM Scale Sets using managed disk image
Terraform
Scalable NIC
Scalable Storage
Extensions

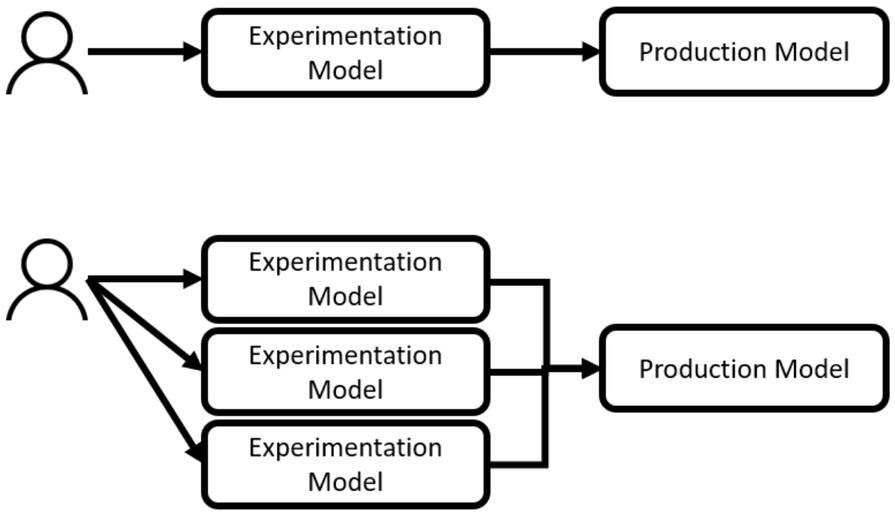
ixek

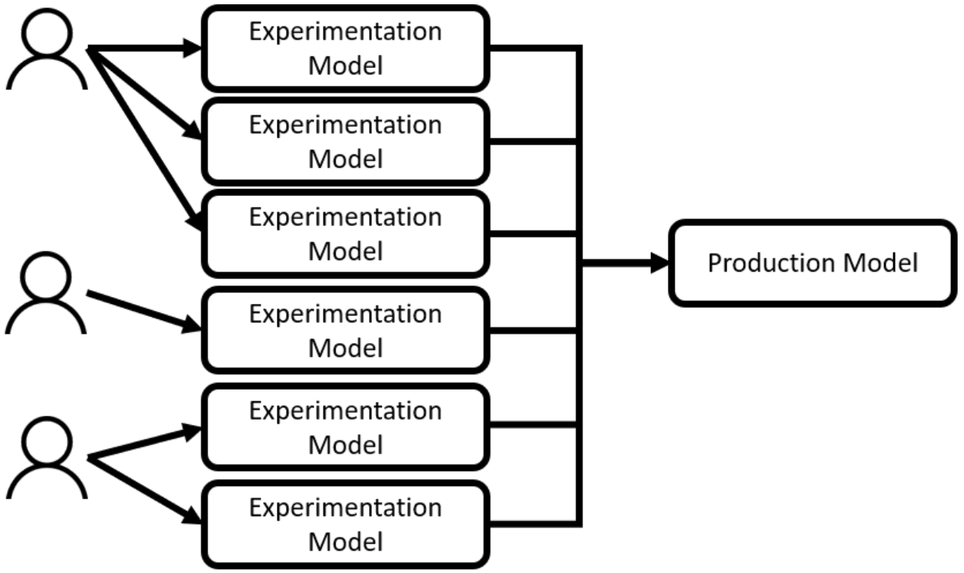
EVERYTHING AS A CODE

- Version control
- Less ambiguity on the configurations
- Shorter turnarounds
- Deterministic environments

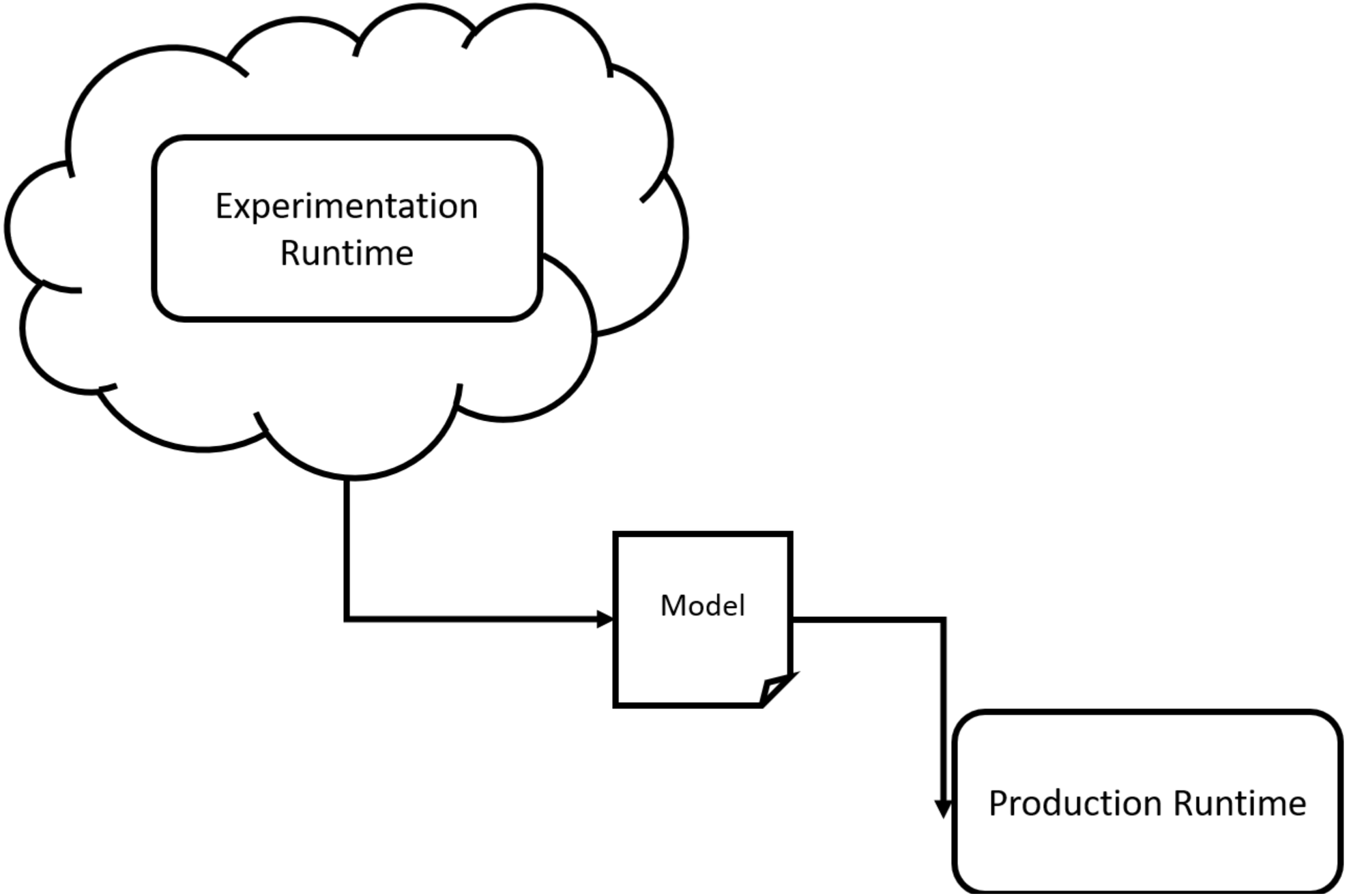
ixek

PROCESSES



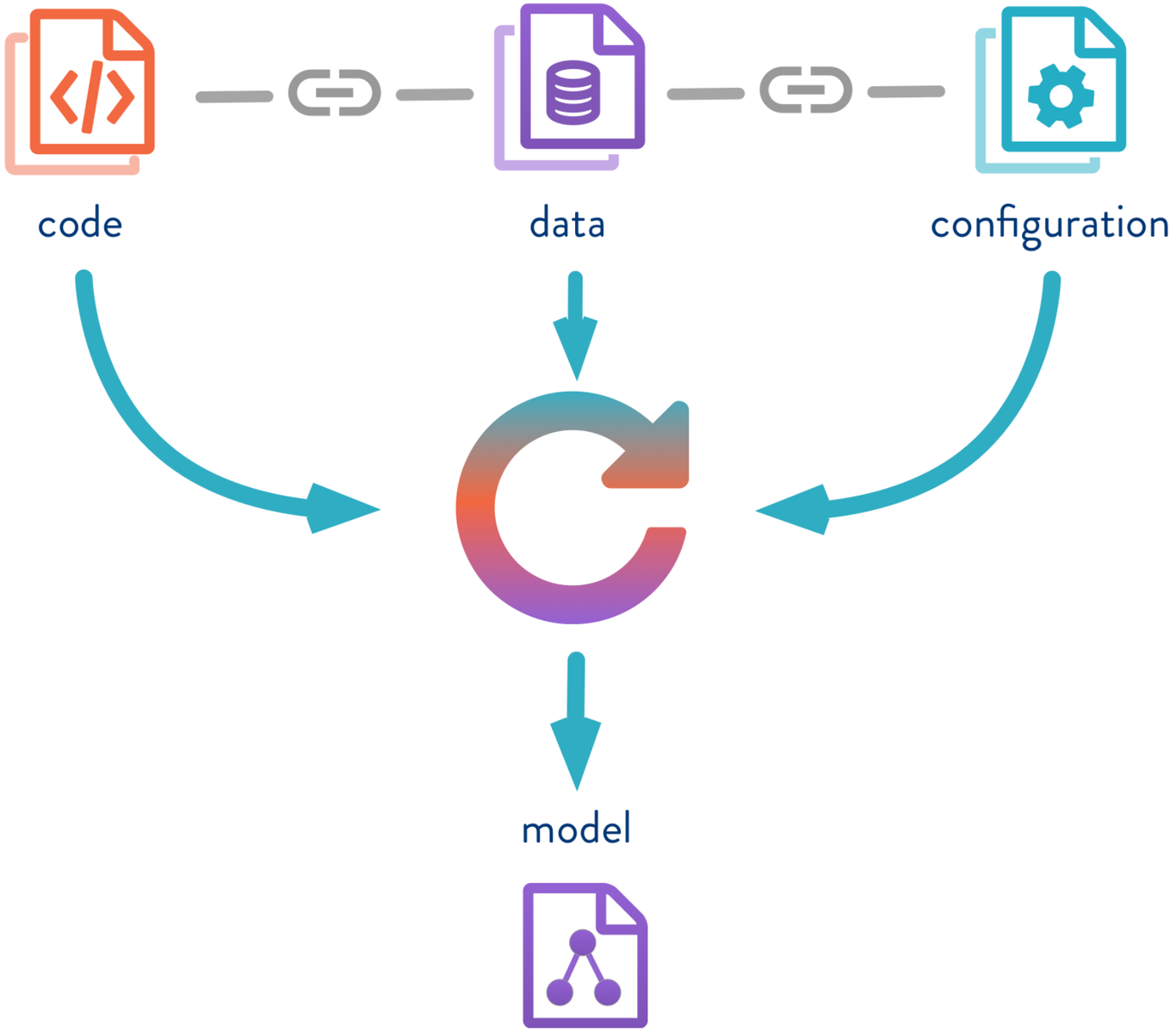


ixek


Experimentation Runtime
Model
Production Runtime

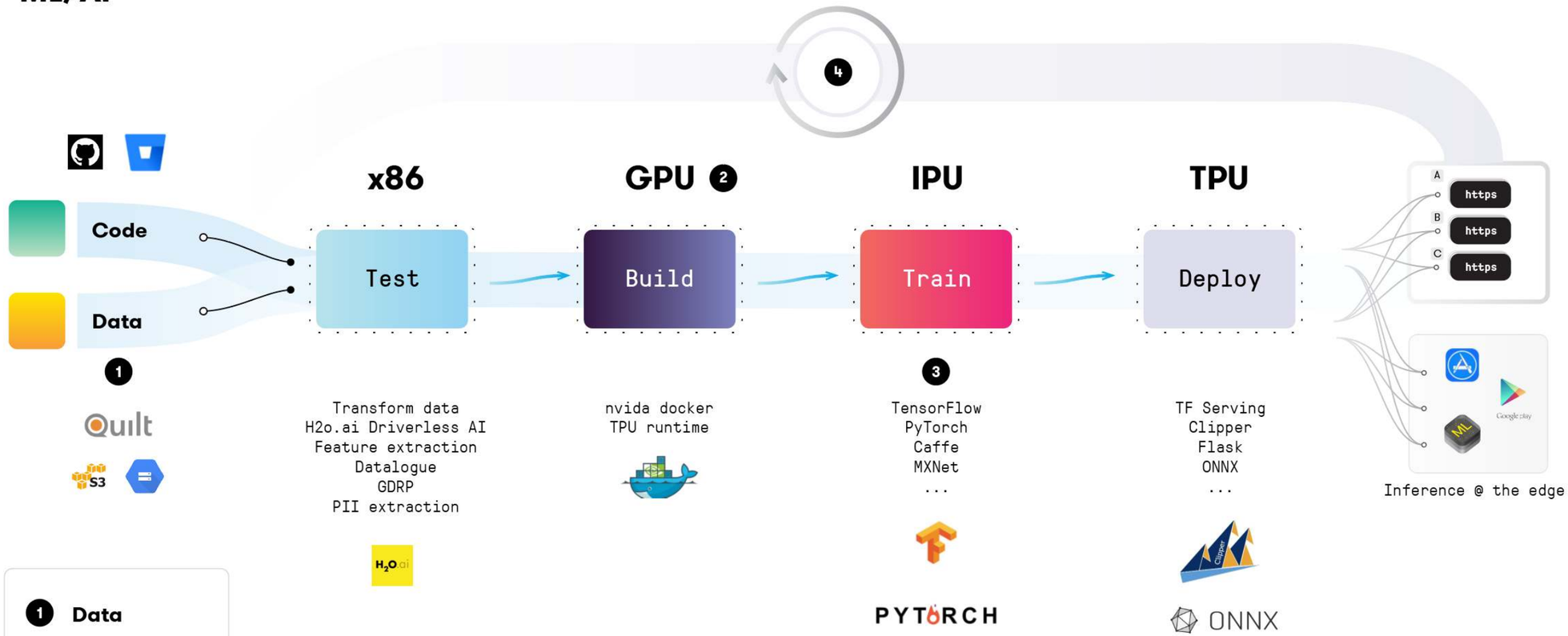
ixek

DATA AND CODE AS FIRST CLASS CITIZENS


code
data
configuration
model

ixek

ML/AI



1. Data
2. Accelerators
3. Train step
4. Refitting

ixek

PEOPLE

- Data scientist
- Data engineer
- ML Engineer

ixek

WHAT DOES ACADEMIA HAVE TO OFFER?

Much more than you think

ixek

PEOPLE

- Researchers
- Research software engineers
- Librarians

ixek

RESOURCES AND INFRASTRUCTURE

We still need to figure this out... it is pretty much an ad-hoc case

ixek

PROCESSES

- Scientific rigour
- Peer review
- Data management

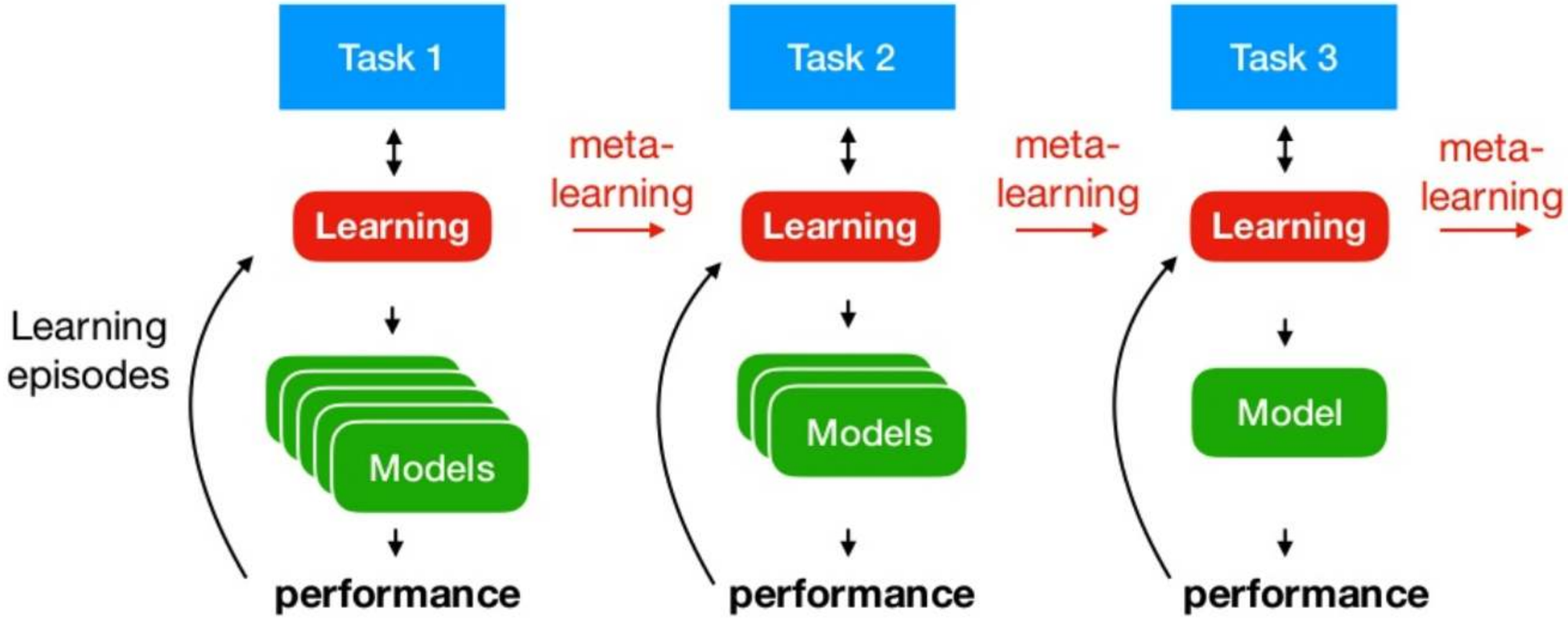
ixek

WHICH AREAS COULD BENEFIT FROM ACADEMIC COLLABORATIONS?

 ixek

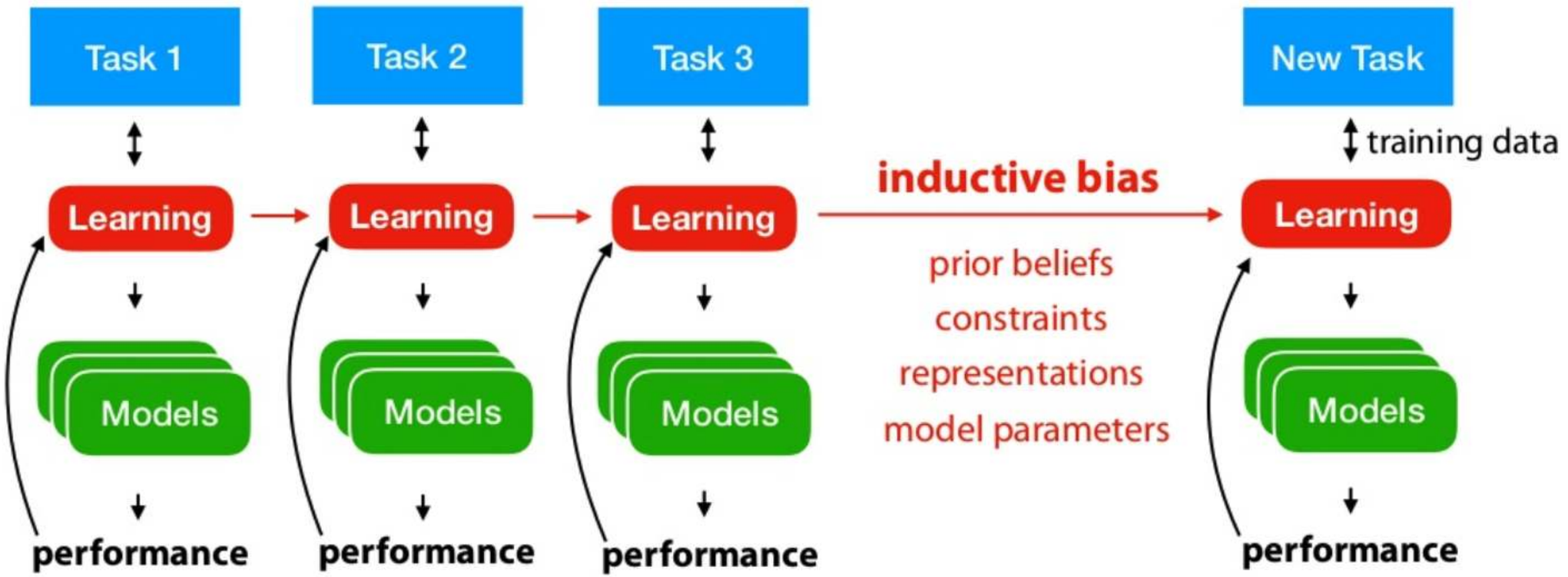
META-LEARNING

Humans learn across tasks (learn from experience)


Task 1
Learning
Models
performance
Learning episodes
meta-learning
Task 2
Learning
Models
performance
meta-learning
Task 3
Learning
Model
performance
meta-learning

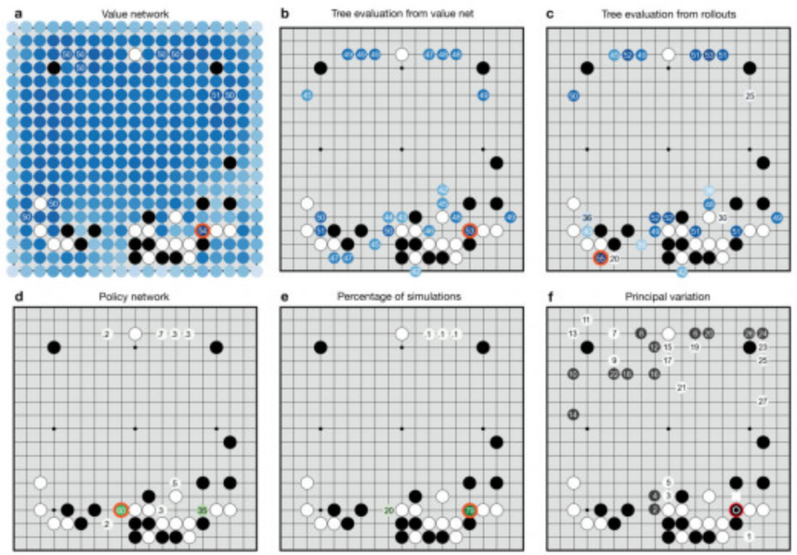
ixek

If prior tasks are similar then we can carry prior knowledge



ixek

AlphaGo uses some sort of meta-learning



AlphaGo

AI RESEARCH .COM

ixek

ALGORITHMIC FAIRNESS

It has become increasingly important to ensure that models are making justified calls that are free from unintended bias.

 ixek

ALGORITHMIC FAIRNESS

It has become increasingly important to ensure that models are making justified calls that are free from unintended bias.

The one way to make progress is through interdisciplinary collaboration

ixek

TOWARDS MODEL EXPLAINABILITY

Address the trade-off between performance and interpretability

ixek

REINFORCEMENT LEARNING DEADLY TRIAD

Following nature's paradigms RL agents receive awards and then learn to maximise success by performing optimal actions.

ixek

How to keep an algorithm learning if there are far too many potential variables or outcomes to be evaluated without being fed ridiculous amounts of data.

ixek

IN BRIEF

Focus on the 3 pillars:

- People
- Infrastructure
- Processes

THANK YOU

 ixek

 tania.allard@microsoft.com