

Software 2.0 | Fueling the AI Revolution



www.understand.ai

A young fast-growing start-up

Understand.ai



Founded: March 10th, 2017

Research dates back to 2014

HQ: Karlsruhe (Germany)

Heart of a key European Automotive & AI cluster

Investment: \$3.5M seed

From Valley, UK & German entrepreneurial VCs

Team:

53 associates, combination of AI & Automotive talent

Customers:

Automotive OEMs and Tier 1/2s

e.g. VDA V&V-Methoden (Pegasus successor)





Fueling the AI Revolution

Agenda

1. Where is AI used?
2. What is required to build an AI system?
3. Characteristics of Data
4. Training & Validation Data
5. Data Engine
6. Processes and People
7. Conclusion

Where is AI used?



AI as...

Personal Assistance

- speech to text
- night vision
- translation
- search & find
- ...



AI in...

Automotive

- Perception
- Planning
- Routing
- Behaviour Models
- ...



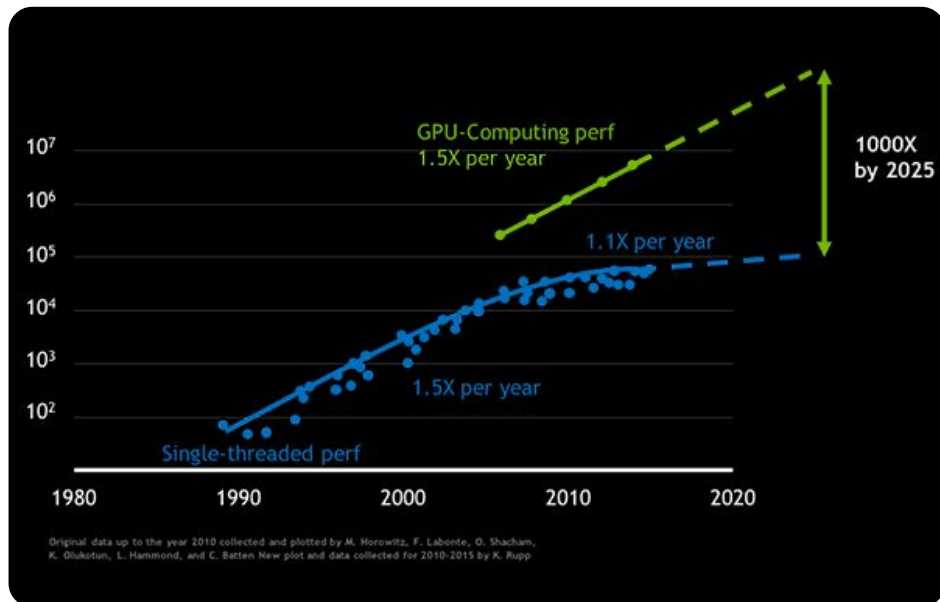
AI in...

Medicine

- Diagnose Diseases
- Develop drugs faster
- Personalize treatment
- Enable gene editing
- Chatbot as a personal doctor (X2.ai)
- ...

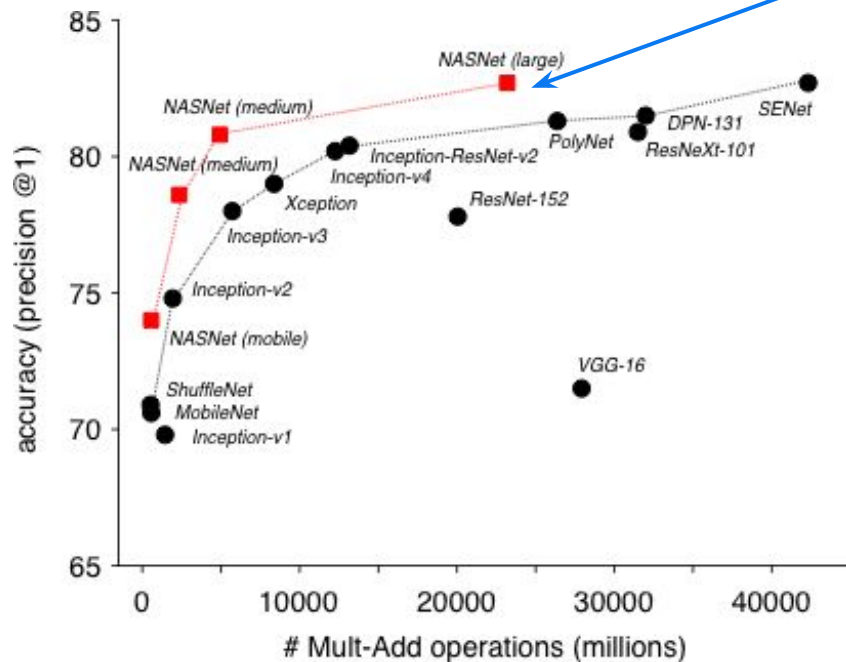
What is required to build an AI system?

Infrastructure



Algorithms

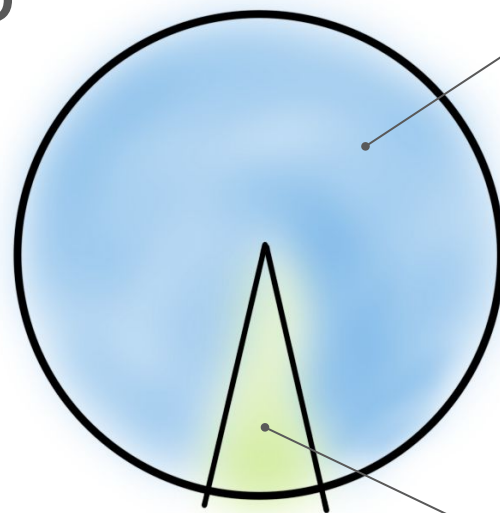
generated model performs significantly better than manually designed model



Amount of sleep lost over...

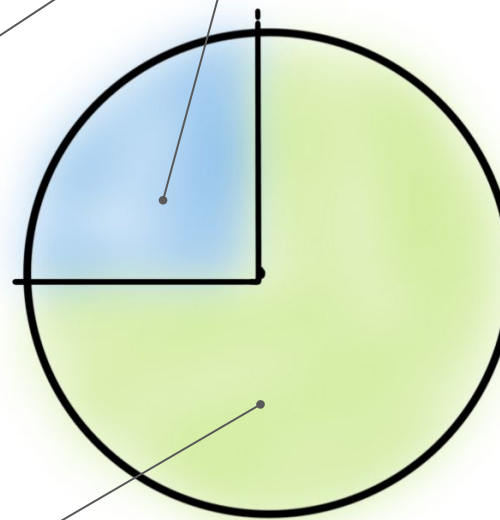
Data

PhD



Algorithms

Tesla

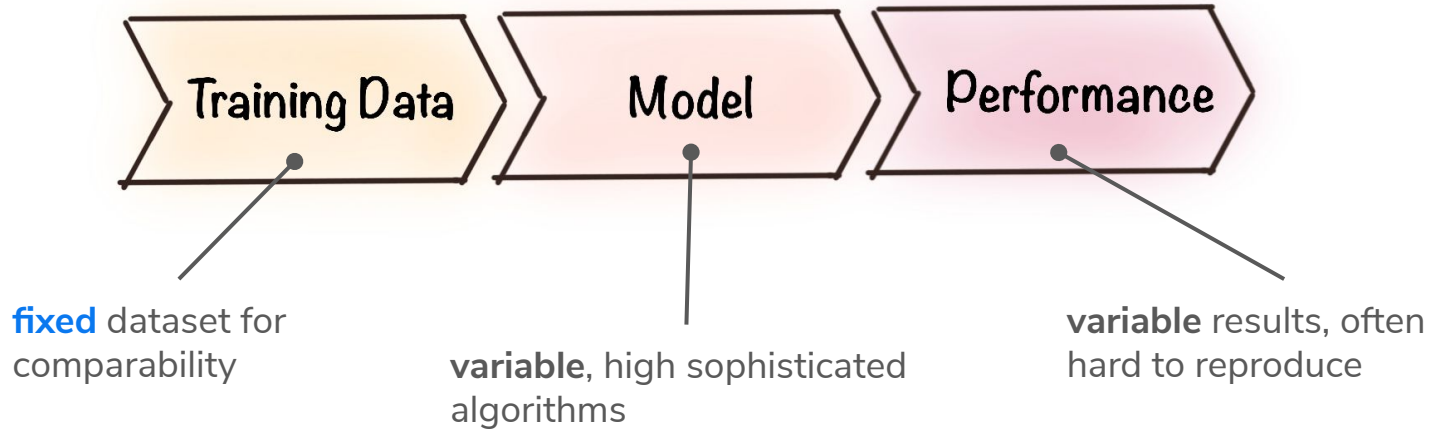


Data Sets

Source: Andrej Karpathy

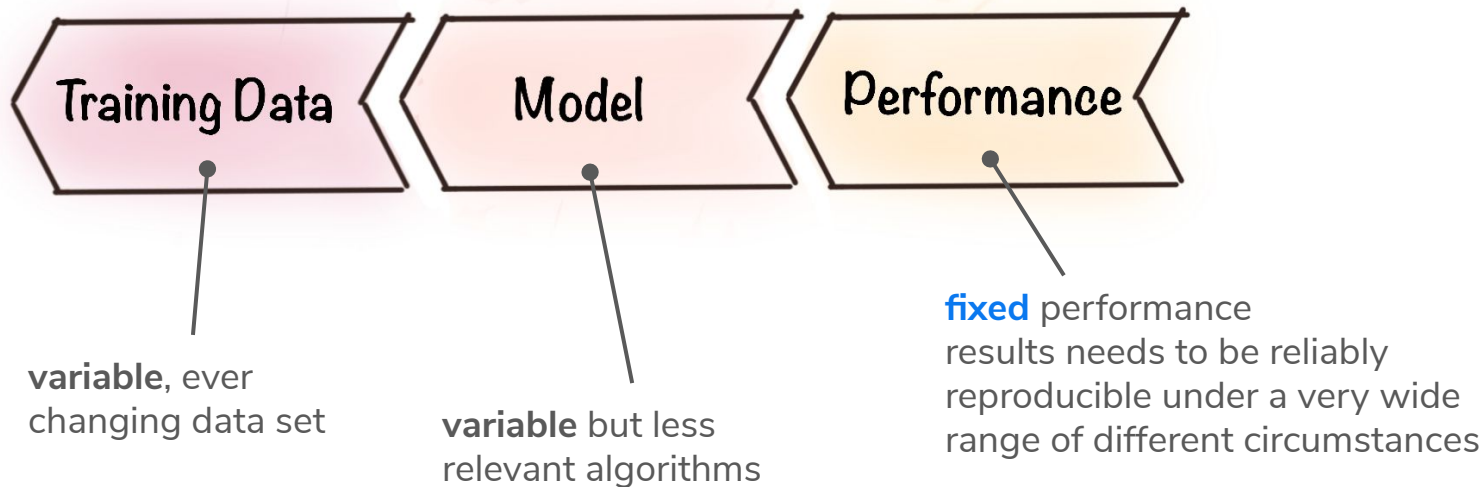
Academia versus Industry

Academia



Academia versus Industry

Industry



Industry

Fixed Model Performance



Characteristics of Data

Distorting VOC Bounding Boxes

Data Quality



ground-truth

small noise (0.08)

large noise (0.13)

Distorting VOC Bounding Boxes

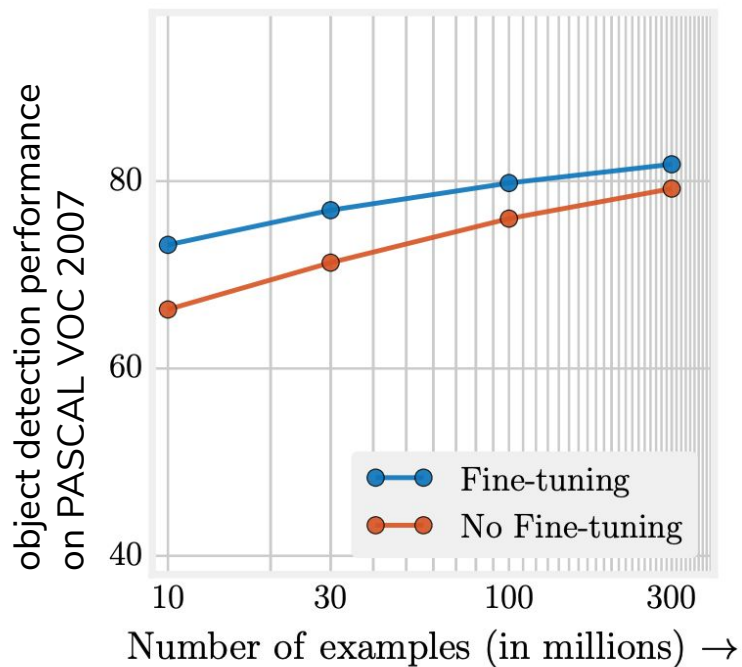
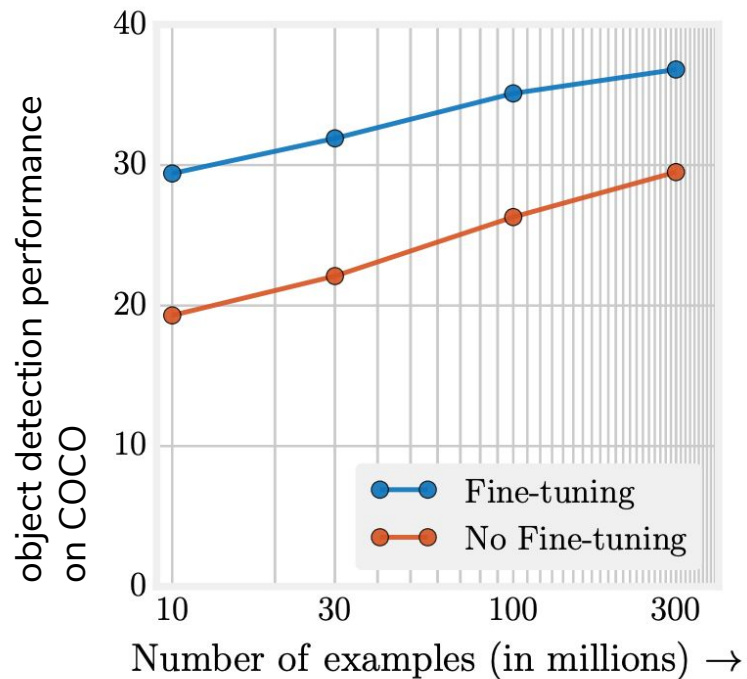
Data Quality

	at IoU 0.5	at IoU 0.8 industry standard
mAP VOC	0.6996	0.4626
mAP VOC 0.08	0.6794	0.2262
mAP VOC 0.13	0.6400	0.0562

-40,64 pts

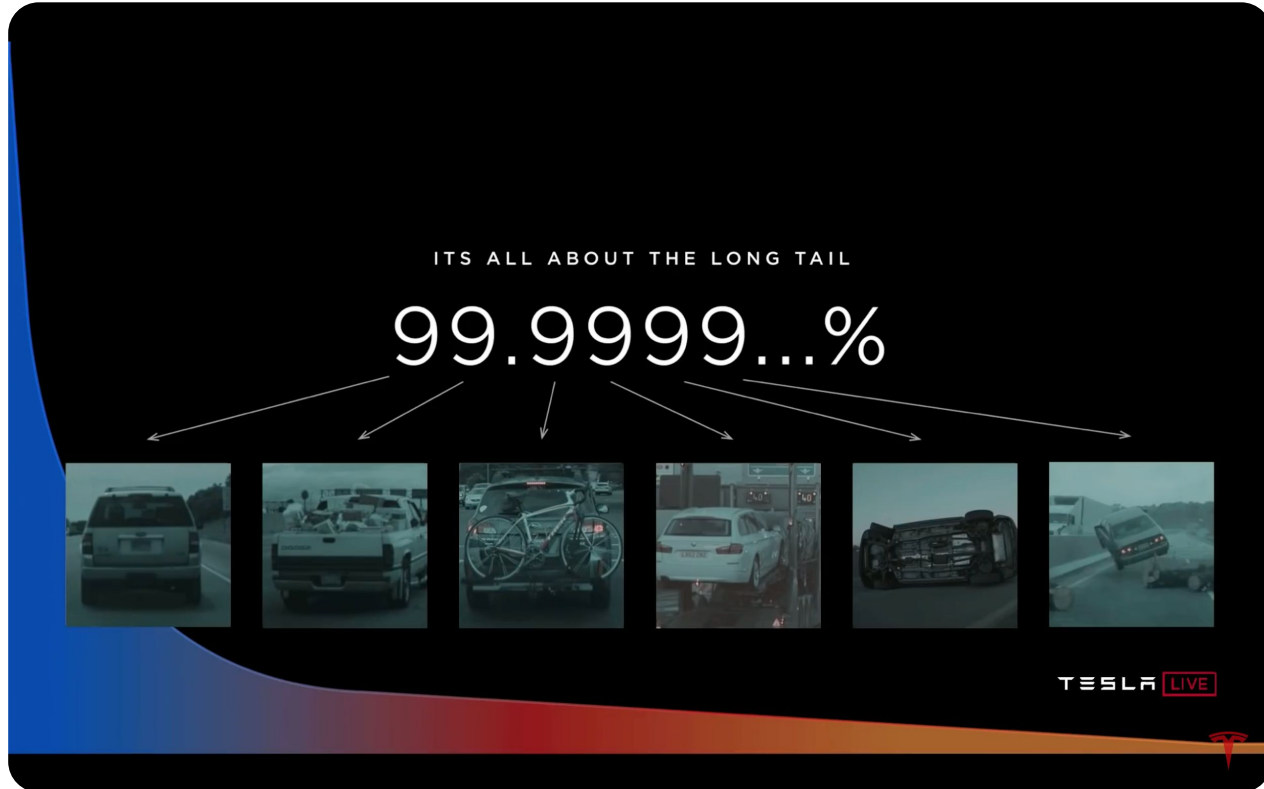
The more the better!

Quantity of Data



When is performance acceptable?

Diversity of Data



Source: Tesla Autonomy Day

Weird Edge Cases :)

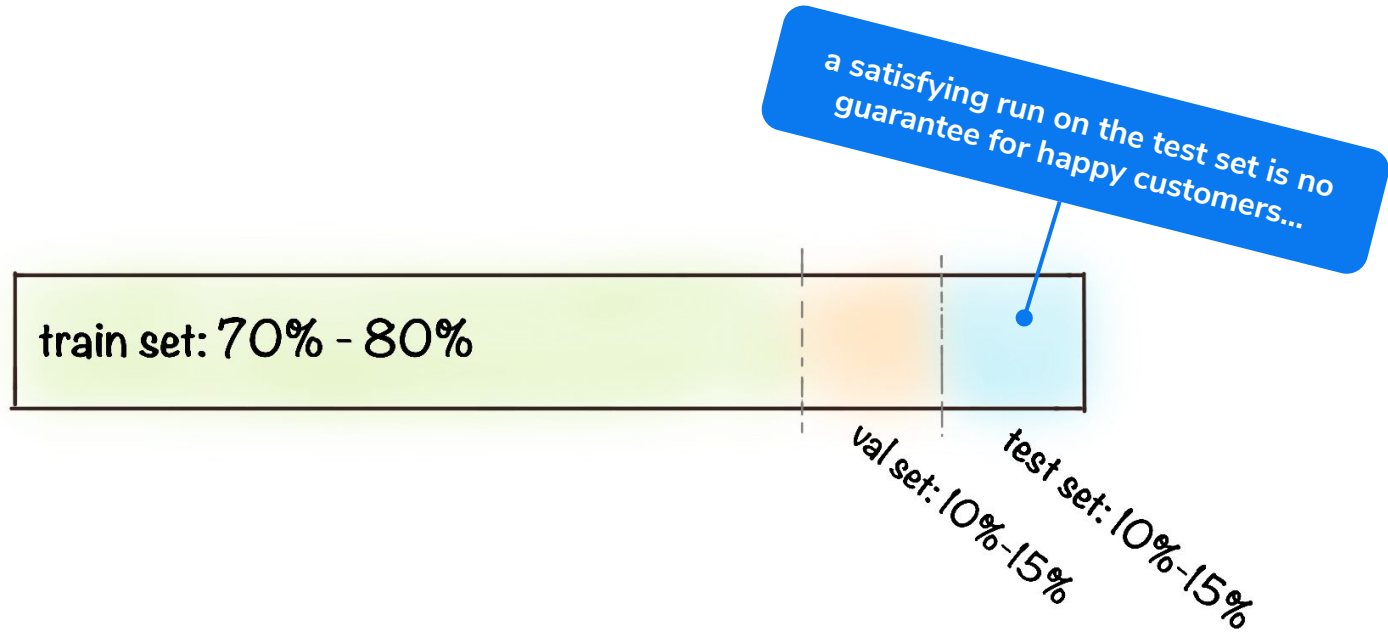
Diversity of Data



Training & Validation Data

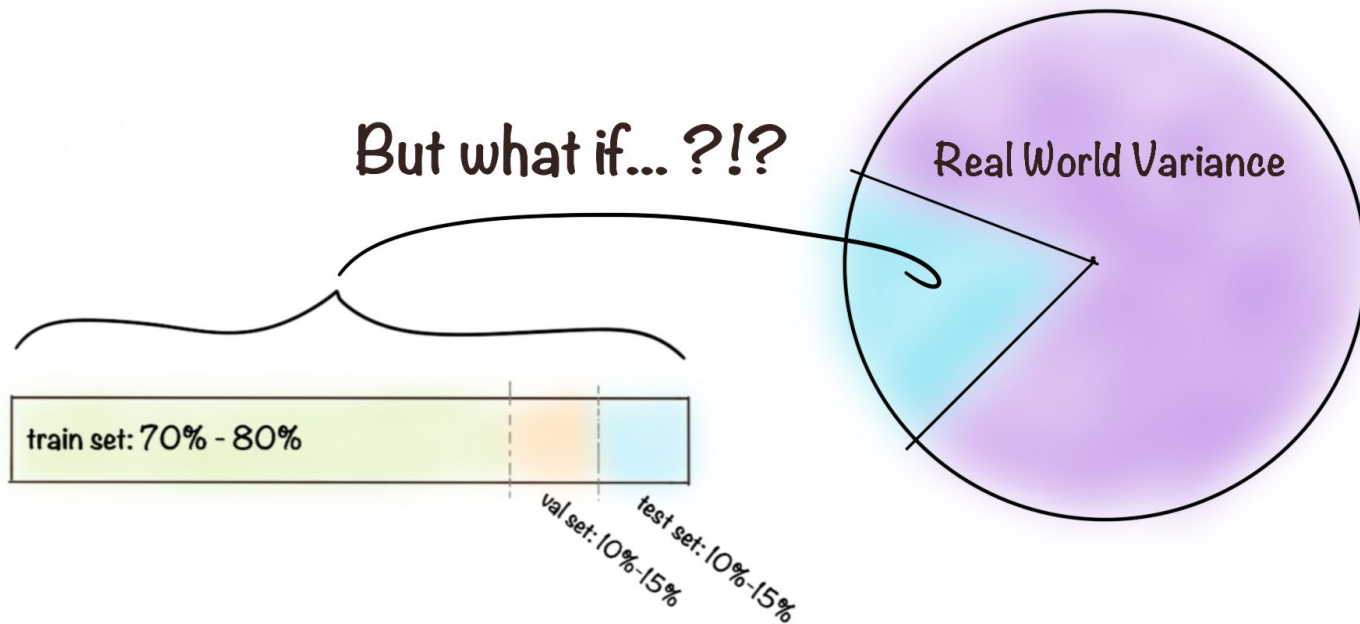
Best Practice

Validation Data



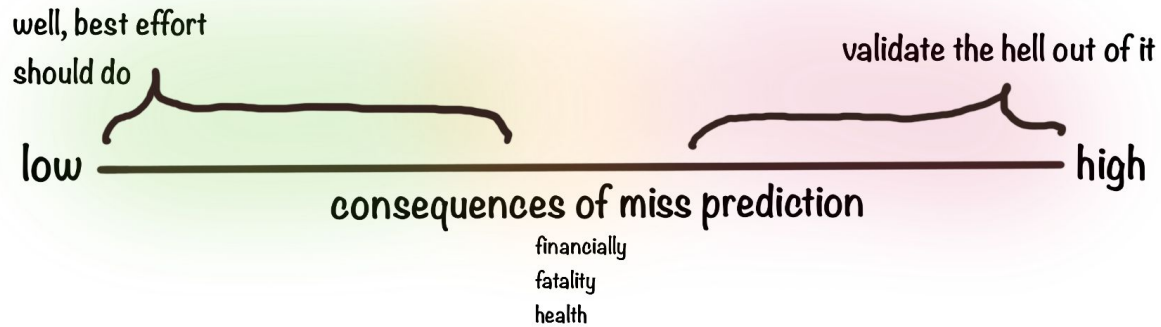
Best Practice

Validation Data



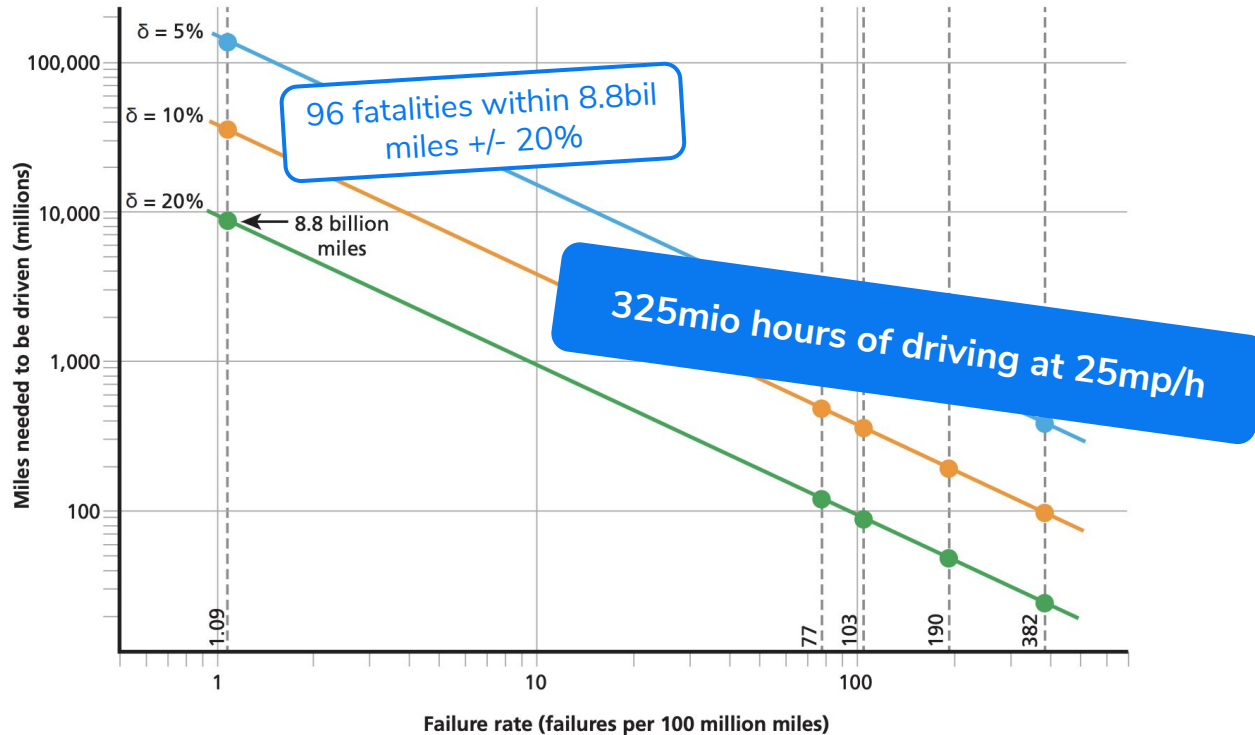
Validation effort might be insane...

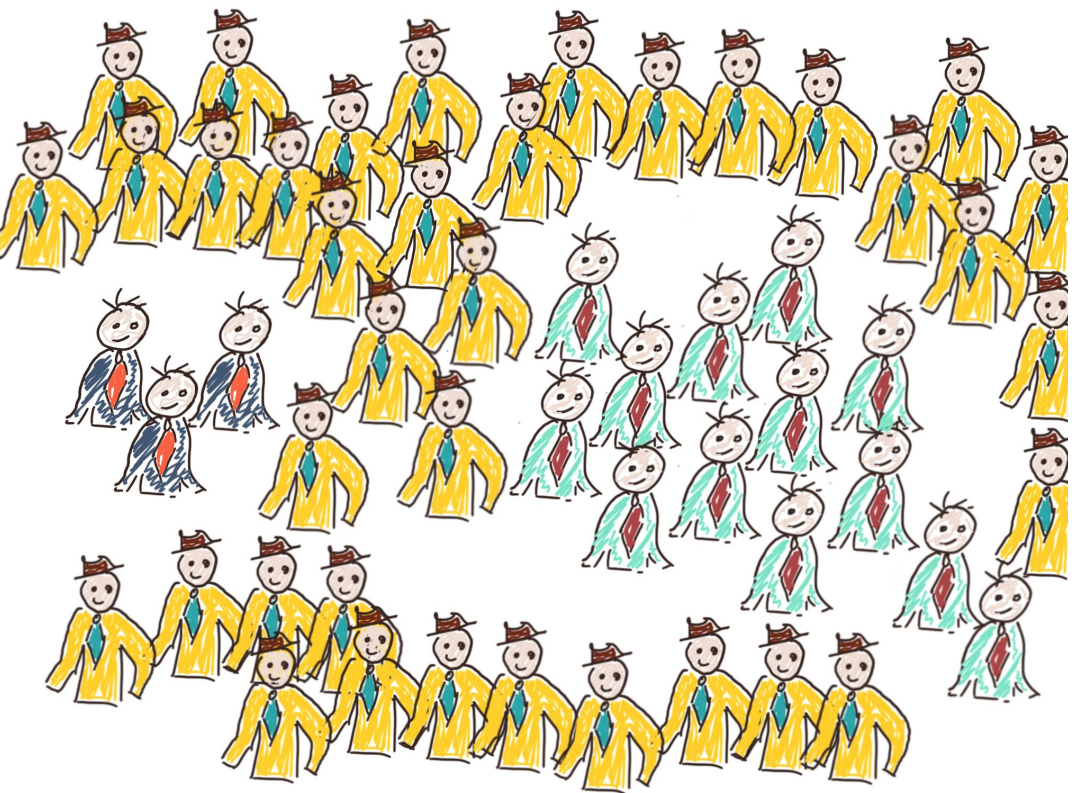
Required Validation Effort



Be as good as a human driver

Validate Autonomous Driving





Will we all be labelers?

Tesla: 1.000 labelers

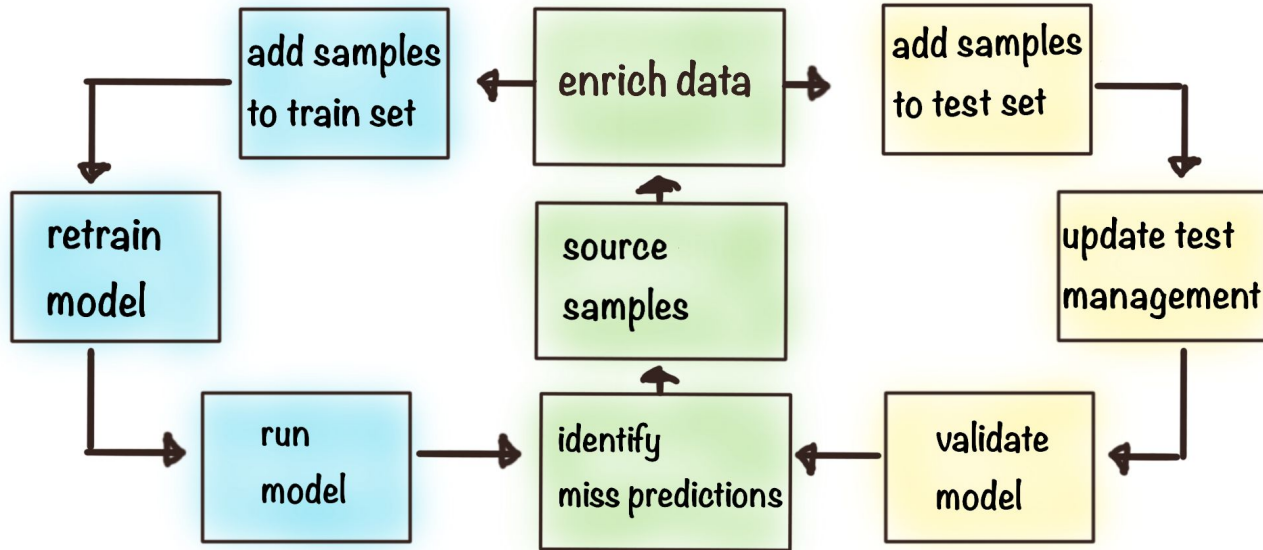
Google: 6.000 labelers

Baidu: 20.000 labelers

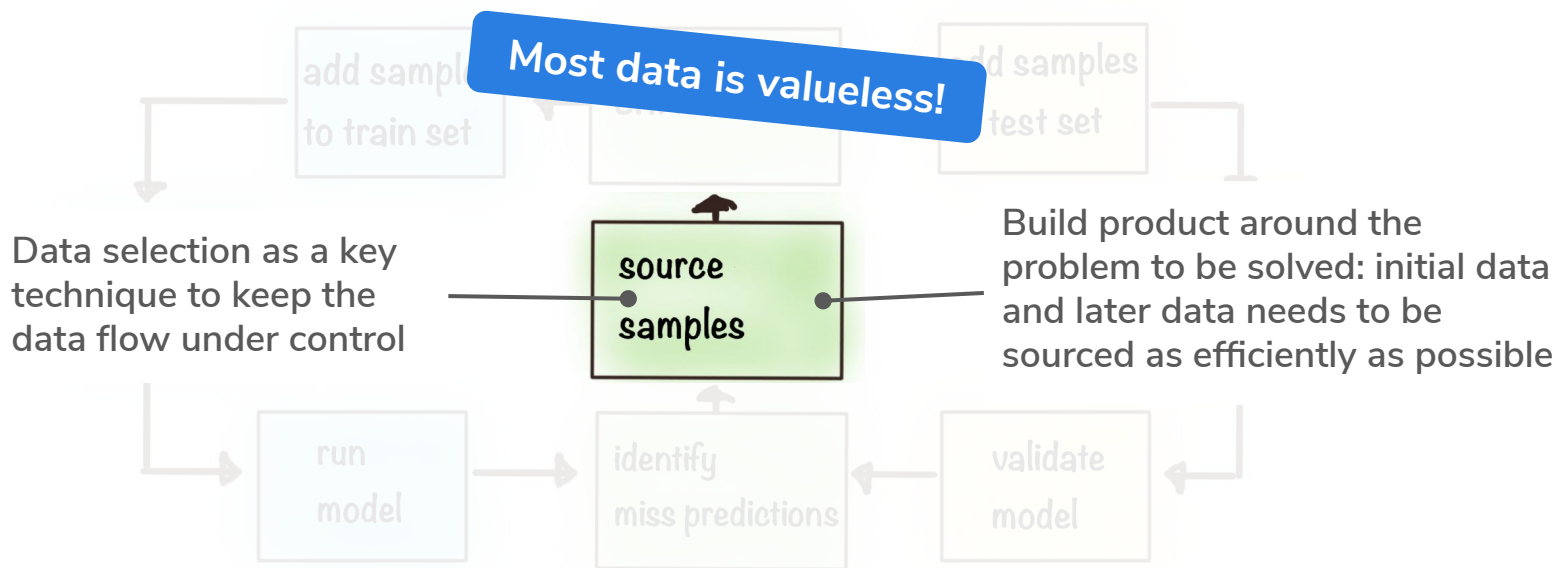
Tooling to Create Data Sets

Data Engine

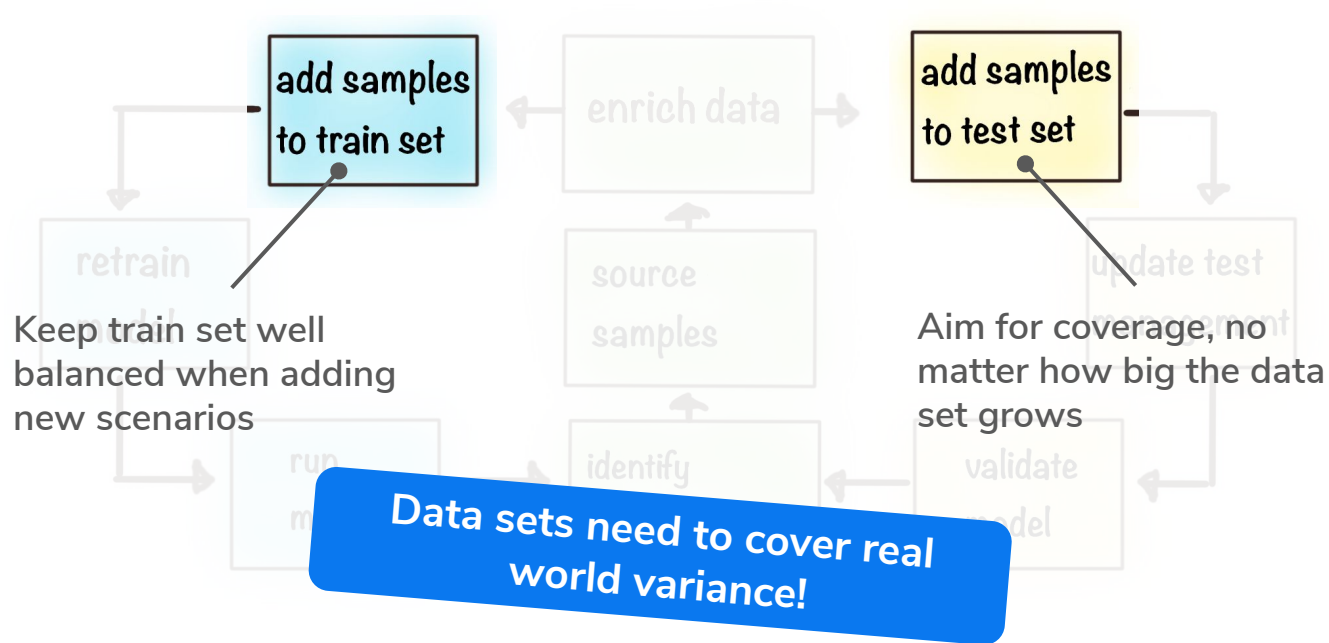
Data Engine



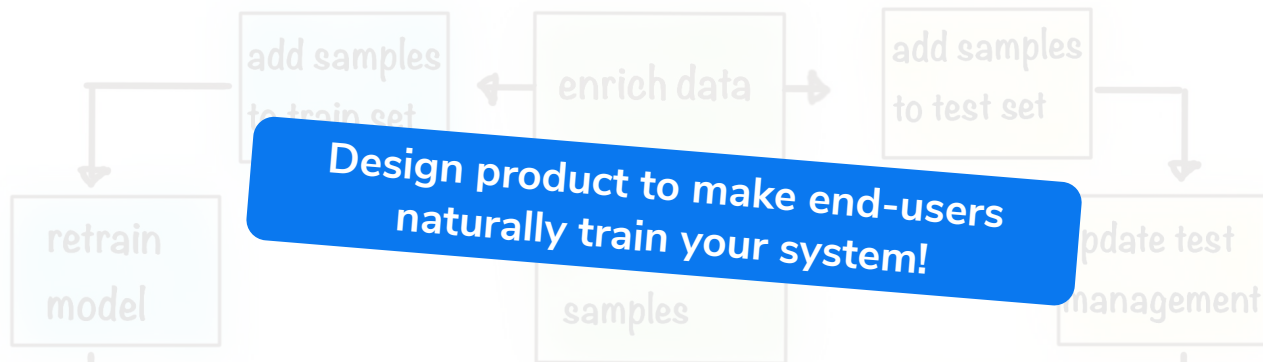
Data Selection



Training & Validation Data



Uncovered Areas



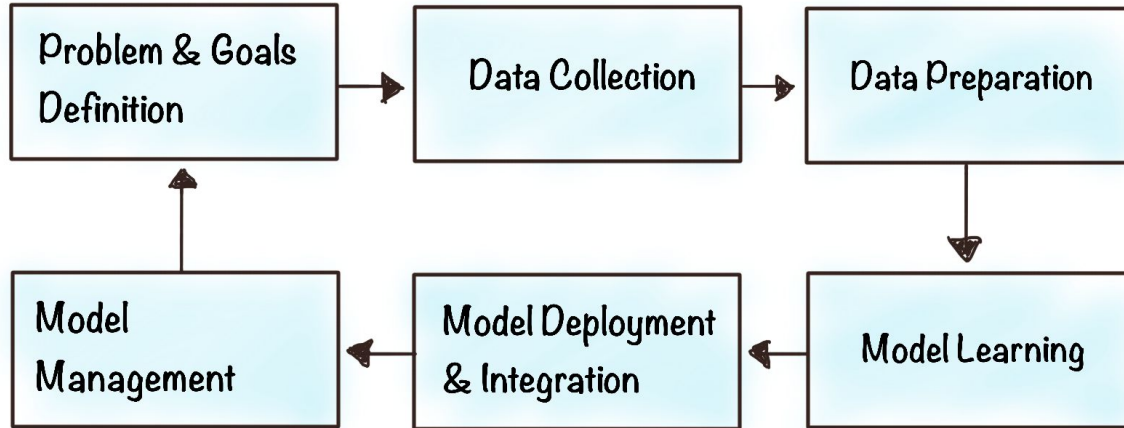
Design product to make end-users naturally train your system!

Visualize output, implement statistics, analyze outliers, run new models in shadow mode

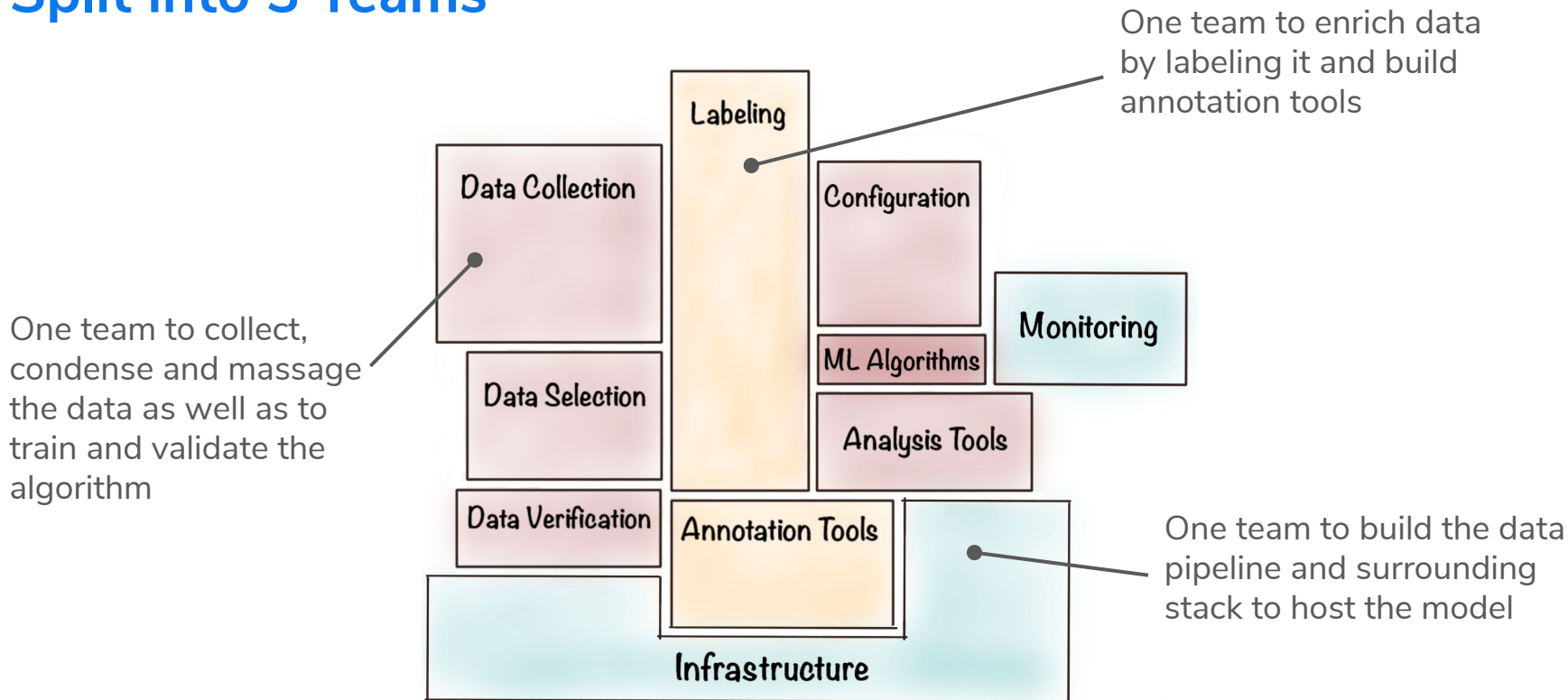
Initial data will not cover real world variance + over time the world is changing (model staleness)

Processes and People

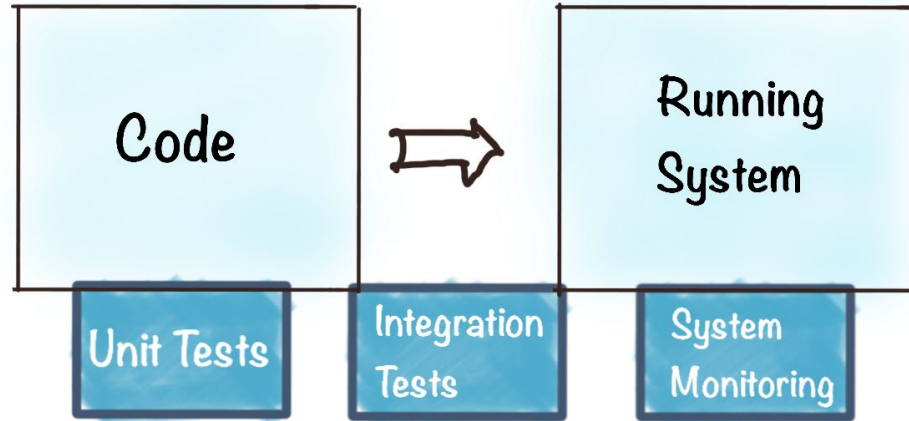
Machine Learning Model Development Process



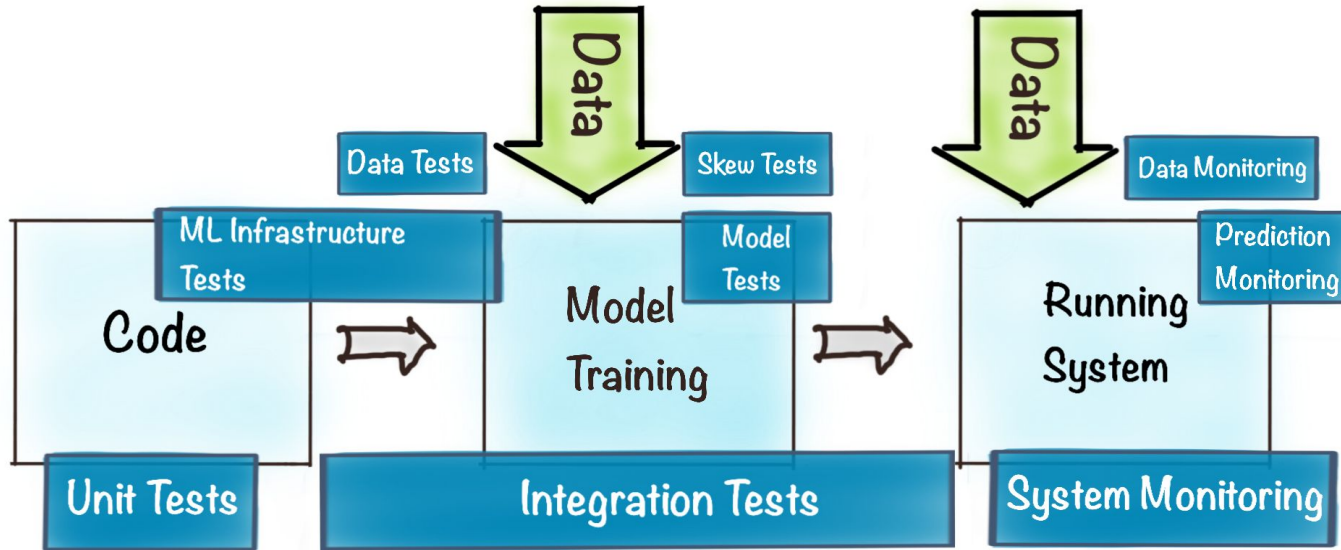
Split into 3 Teams



Traditional System Testing and Monitoring



ML-Based System Testing and Monitoring

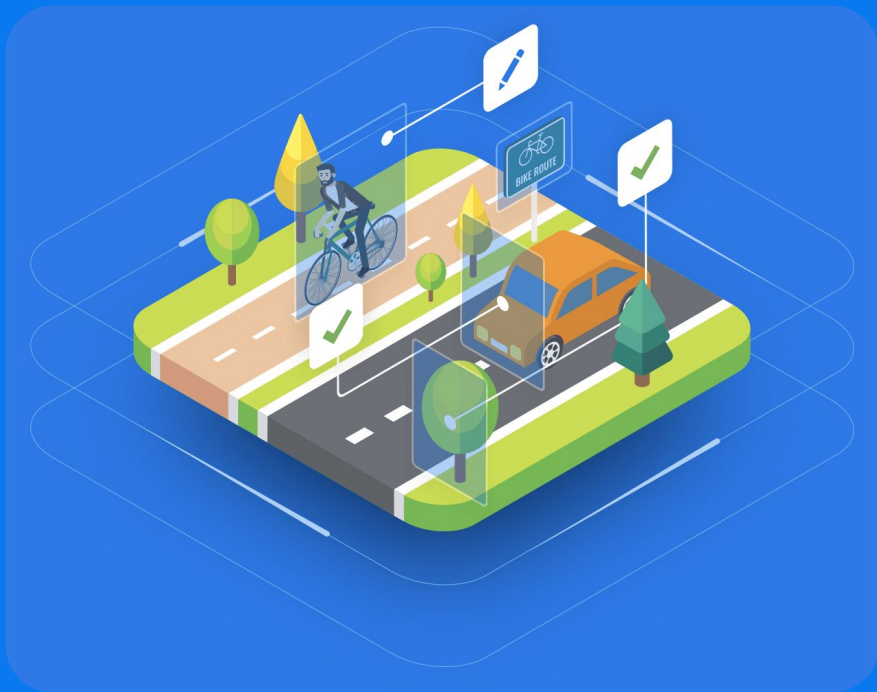


Behavior depends on dynamic qualities of data and model configuration choices

Conclusion

It's all in the data!

Conclusion



UAI is an awesome company

Quality, Quantity and Diversity of training and validation data enable high performance AI Systems

Validation requirements might have an insane impact on effort

Build a data engine!

QA for production grade ML Systems is highly complex

Feedback, please. Thanks!



Software 2.0
Fueling the AI Revolution!

Q&A



Philip Kessler
philip@understand.ai



Daniel Rödler
daniel@understand.ai