Microsoft

31/10/2018

Microsoft Future Decoded

Productionising deep learning:

The confused.com journey

Confused.com
On a mission to save you money

Microsoft

Who are we?

Dr. Tim Scarfe

Machine learning engineer, data scientist and entrepreneur.

Principal Software Engineer

@ecsquendor

Dr. Tempest van Schaik

Biomedical engineer, inventor, designer, researcher. Health + AI.

Software Engineer in Data & AI

@Dr_Tempest

Dr Tim + Dr Tem aka:

Tweat out to #TimTem !

Microsoft

What you'll hear from us today:

- What we do at Microsoft
- The story of deep learning
- Machine learning is becoming ubiquitous in software
- Many companies are becoming data companies
- Our collaboration with confused.com
- Our deep learning model for car insurance
- The emerging field of DevOps in data science
- Our deep learning DevOps pipeline

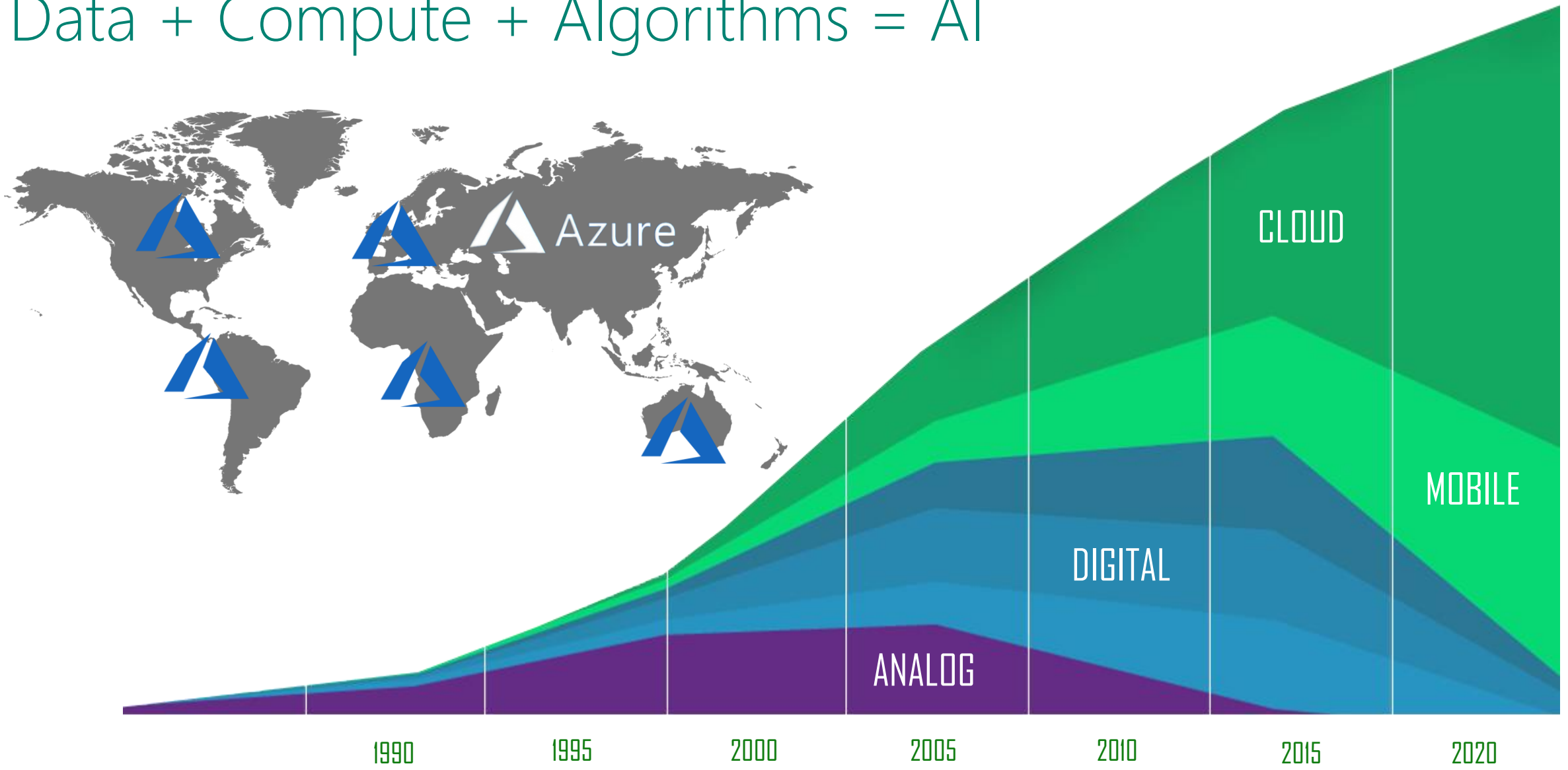
Commercial Software Engineering

Microsoft

Distilled Concepts of Deep Learning

Microsoft

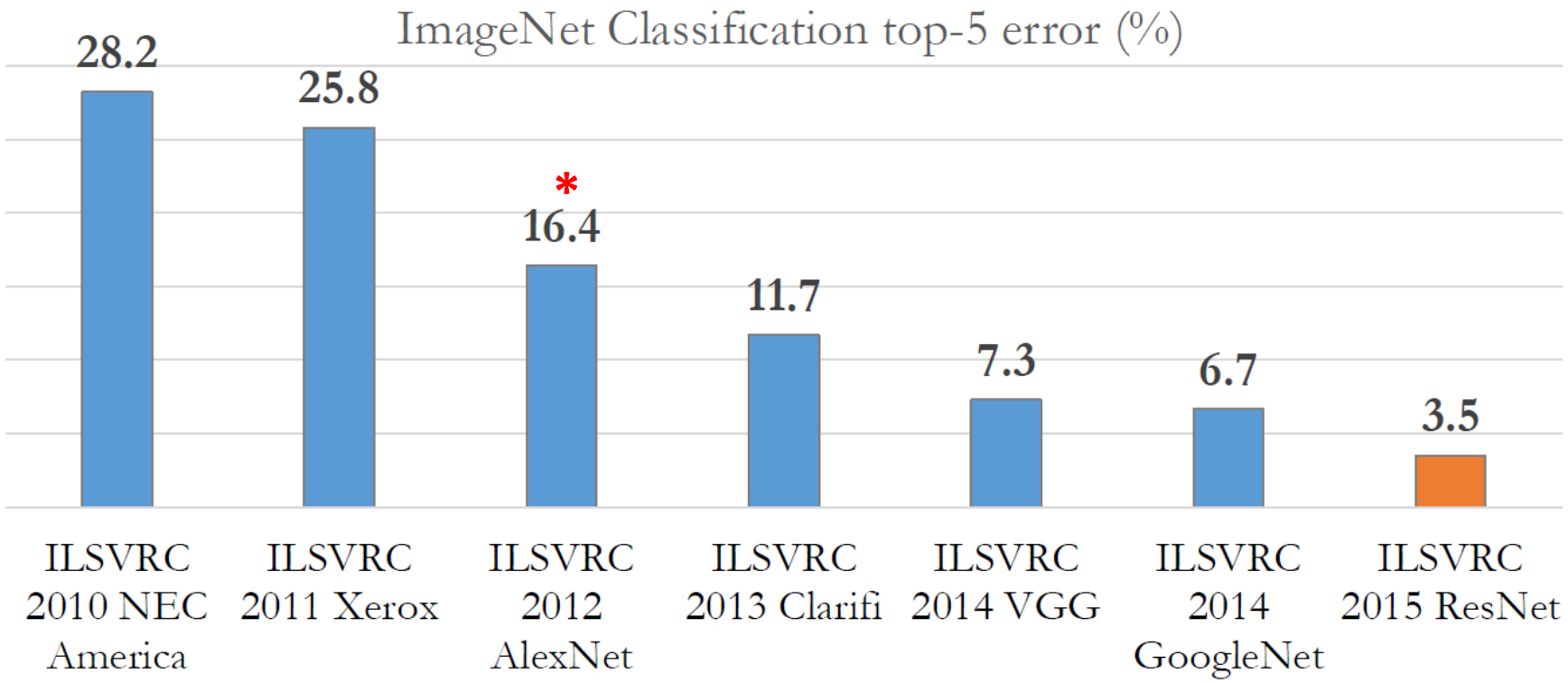
Data + Compute + Algorithms = AI



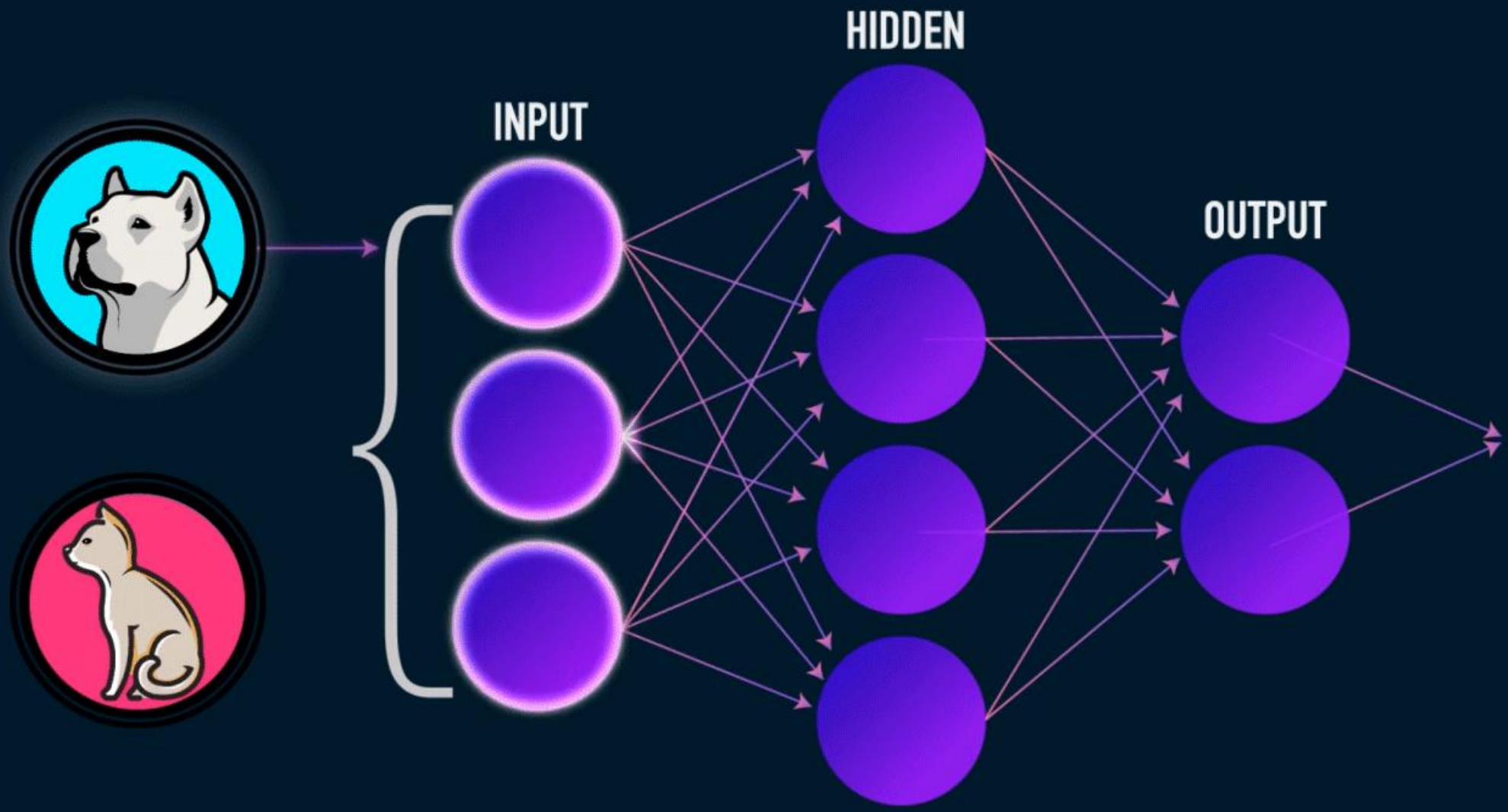
Microsoft

ImageNet

Alex Krizhevsky



- Alex Krizhevsky wins the ImageNet competition in 2012
- AlexNet
- Heralds a new era of AI
- Same story in language processing, speech recognition etc


HIDDEN
INPUT
OUTPUT

Microsoft

Traditional ML

Manual Features → Trainable Classifier

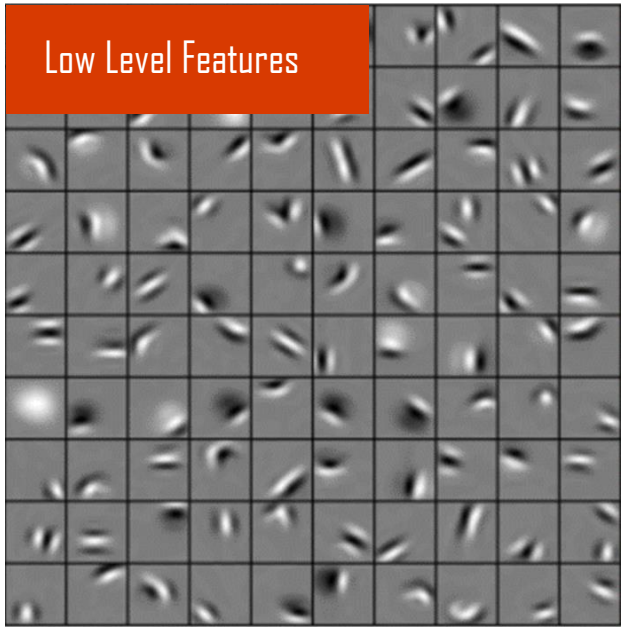
Deep Learning

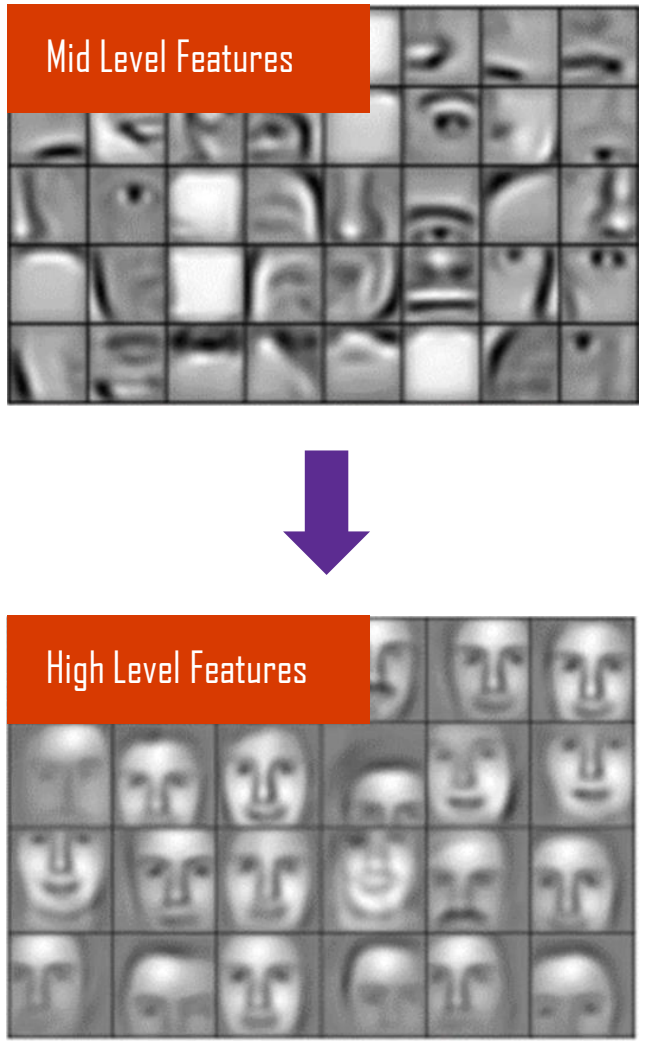
Representations are hierarchical and trained automatically

Low Level Features → Mid Level Features → High Level Features → Trainable Classifier

Deep Learning = learning a representation

Microsoft

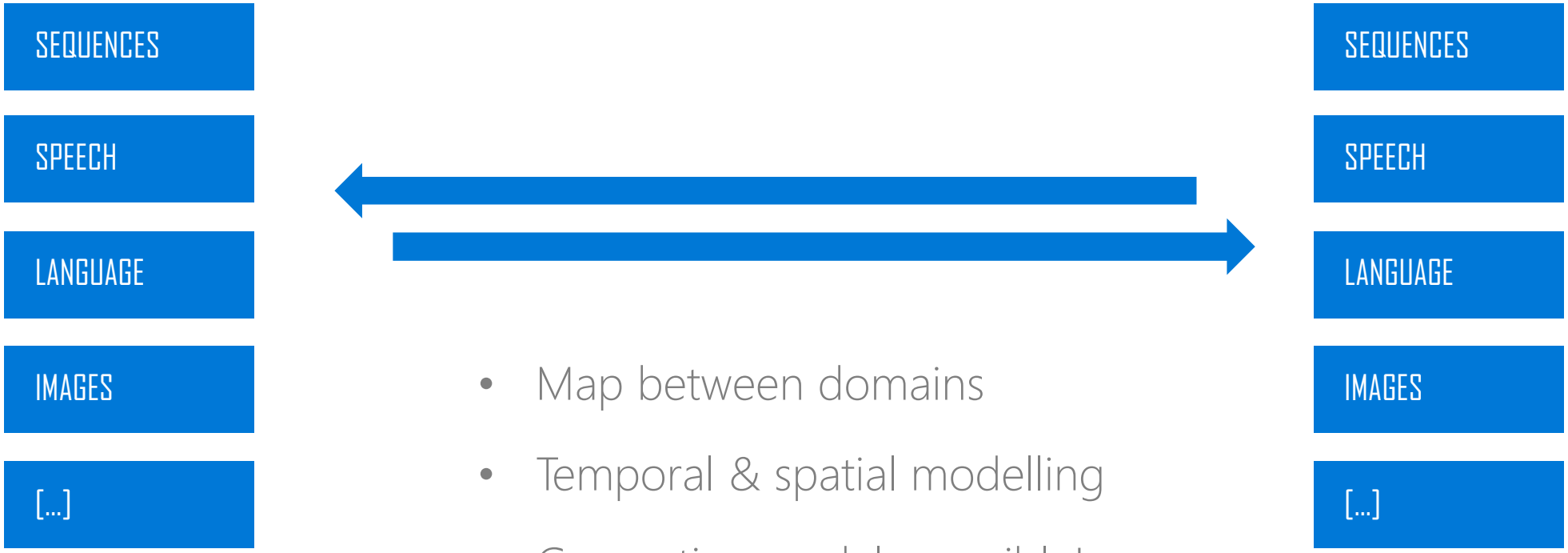




Representations are learned automatically

Microsoft

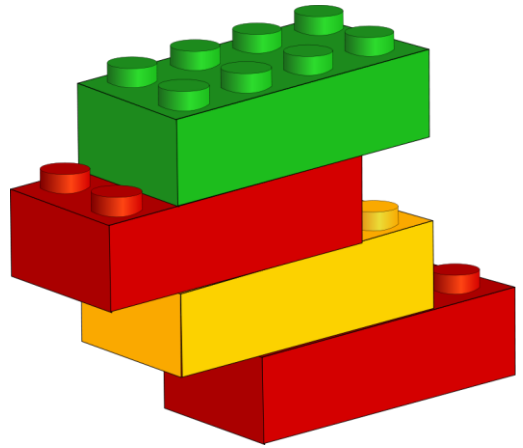
Universal functional approximators



- Map between domains
- Temporal & spatial modelling
- Generative models possible!
- Multiple inputs and outputs!

Microsoft

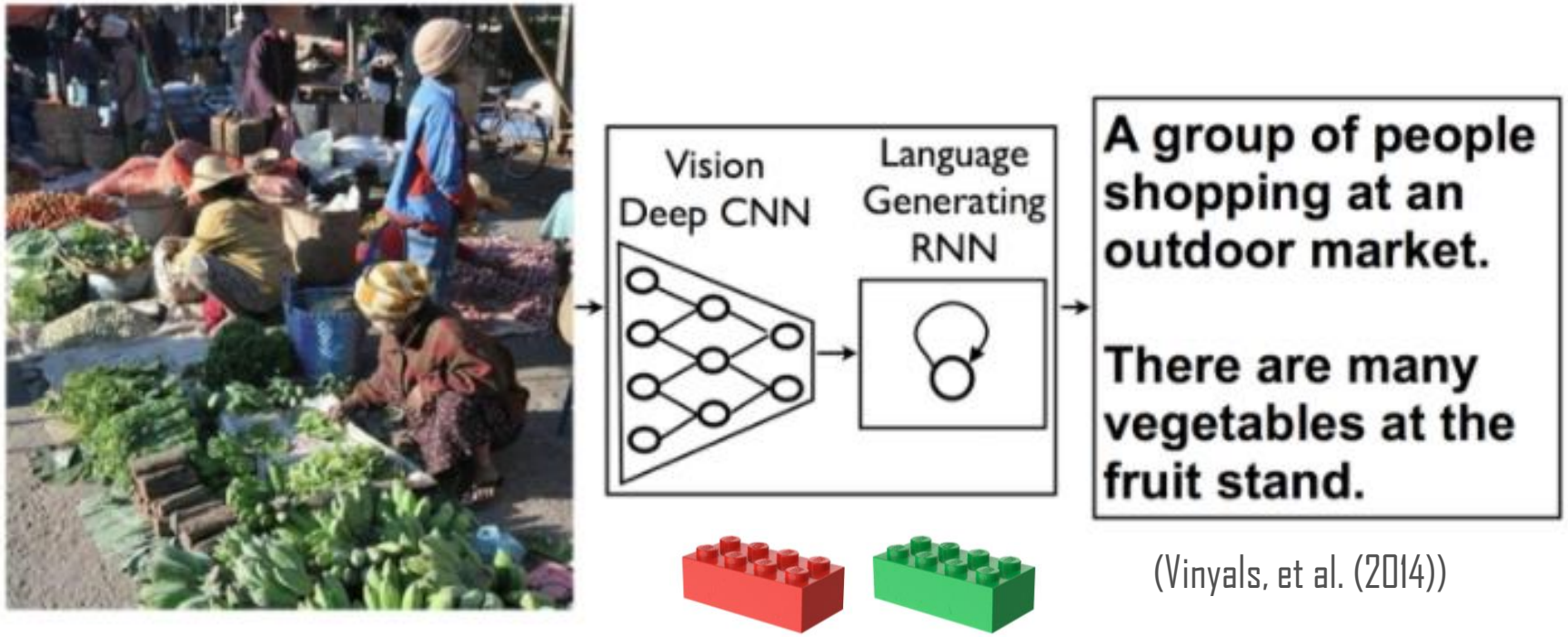
Composability



- Machine learning is becoming a form of software development
- Machine learning models are like software
- There is a dichotomy between the software engineering process and the data science process

Microsoft

Building predictive architectures like lego blocks

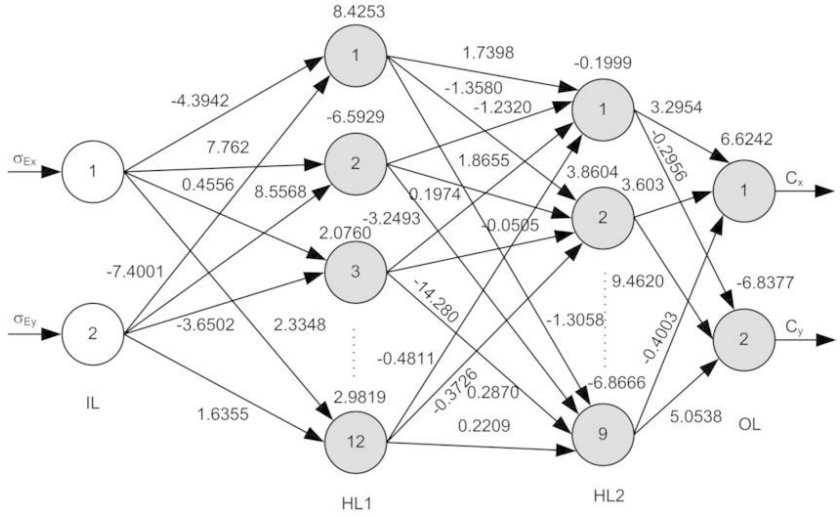


(Vinyals, et al. (2014))

Microsoft

We call this “Software 2.0”

- Software 2.0 is a confection of data generated code and handwritten code
- Some features are **impossible** to write with traditional code i.e. speech synthesis, machine translation, computer vision use cases, game AI
- Computationally homogeneous
- Simple to bake into silicon.
- It is highly portable.
- Composable
- Better than you



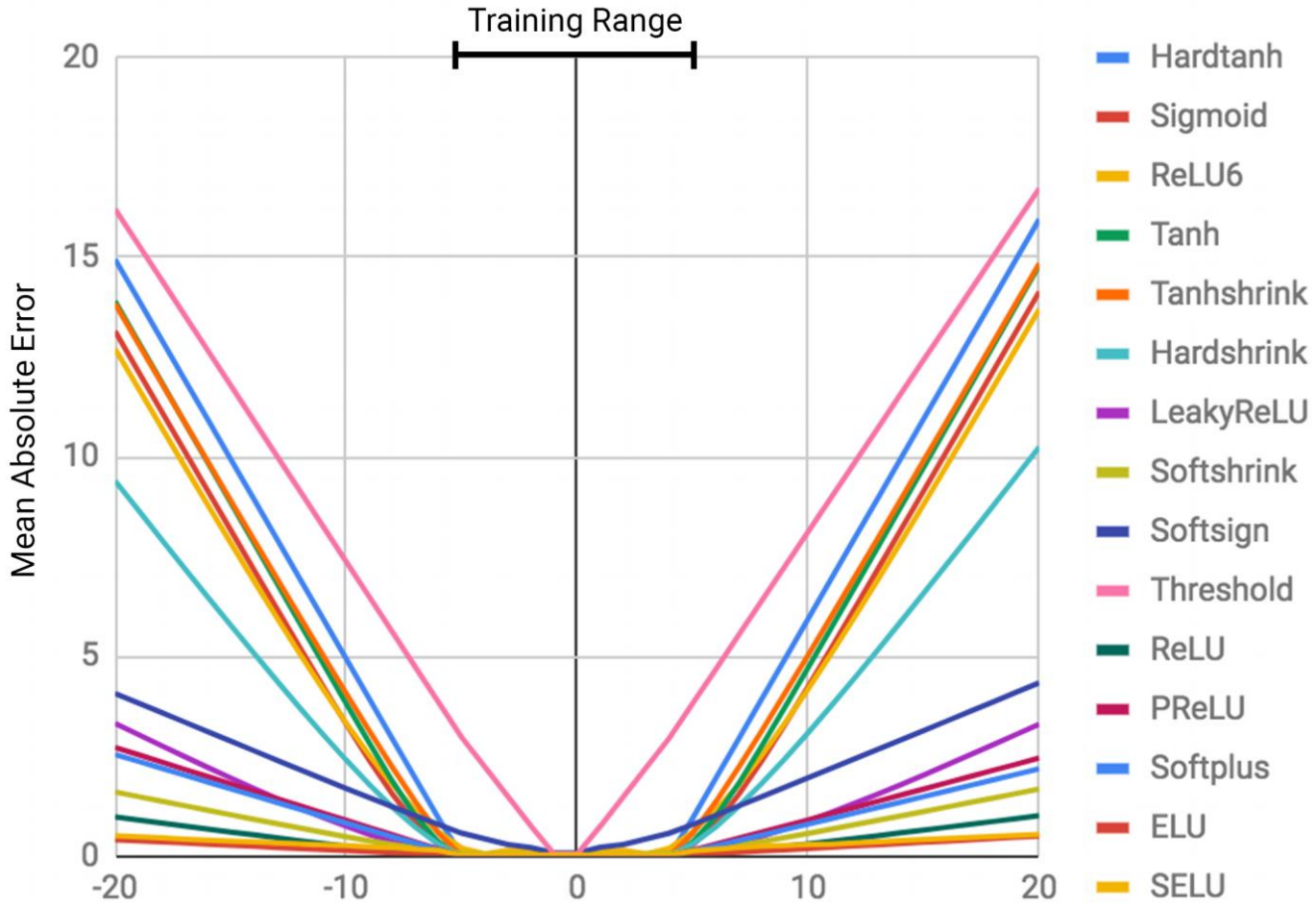
https://medium.com/@karpathy/software-2-0-a64152b37c35

Microsoft

Deep learning is not magic; interpolation vs extrapolation

ML algorithms perform very, very badly on data out of range. They don't generalise as well as you think. Trask et al 2018 (DeepMind)

f(x)=x



Microsoft

When to use deep learning vs classical machine learning

- Large volumes of data
- Unstructured data i.e. images, video, sound
- Modelling temporal or spatial dependencies
- Novel prediction architectures
- Composability, reusability, transferability of models
- Consistency of approach is more conducive to software engineering

Microsoft

When not to use deep learning

- When interpretability is important
- When the data is very small deep learning behaves like classical machine learning i.e. no representation learning
- When you need statistical guarantees

Microsoft

Machine learning models are becoming ubiquitous in software

Software is comprised of some things you can write with code, and some parts which must be generated from data i.e. ML

Microsoft

More companies are becoming data companies

- Companies are becoming data driven
- Diversifying, creating data products & services

Microsoft

Deep learning at confused.com

Microsoft

Confused.com
On a mission to save you money

Our collaboration with confused.com

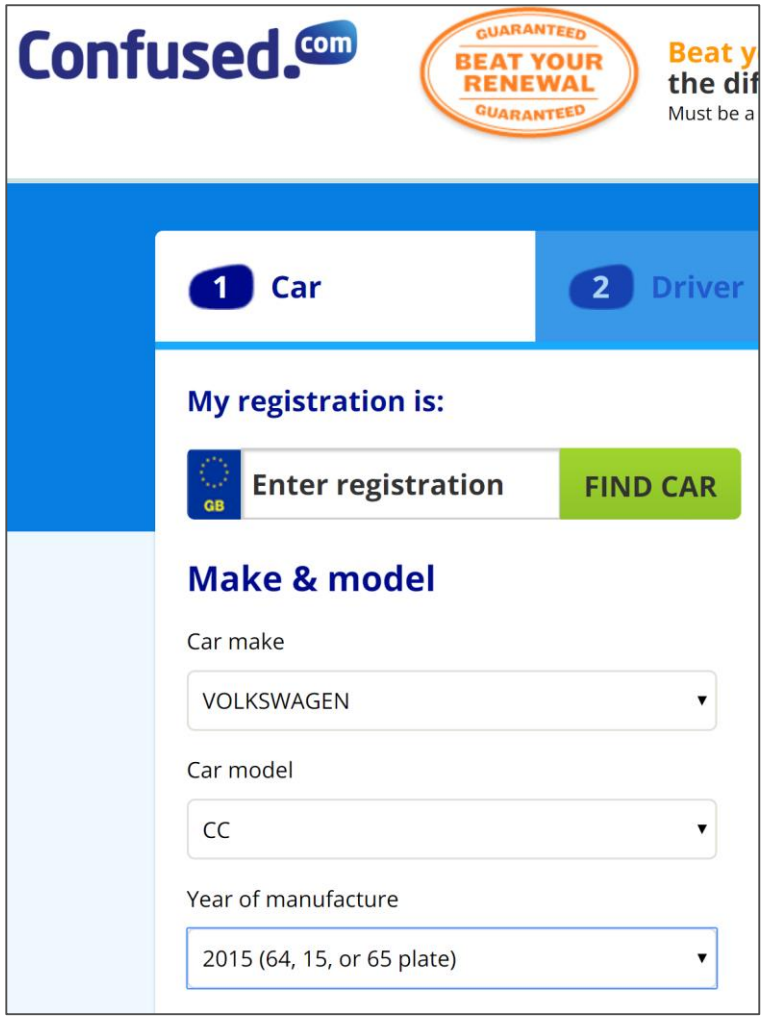
Microsoft

Confused.com
On a mission to save you money

Confused.com

- Who are confused.com?
- Evolution from BI experts to data science experts
- Old and new skills coming together beautifully
- Millions of records of structured data
- Building new products and services from this data

Microsoft



Car & driver details

100+ car insurers

Car insurance quotes

Microsoft

Potential services from data

- Only sharing customer data when necessary
- Getting a best-price ballpark in the car showroom with less data
- Annual musical chairs

Microsoft

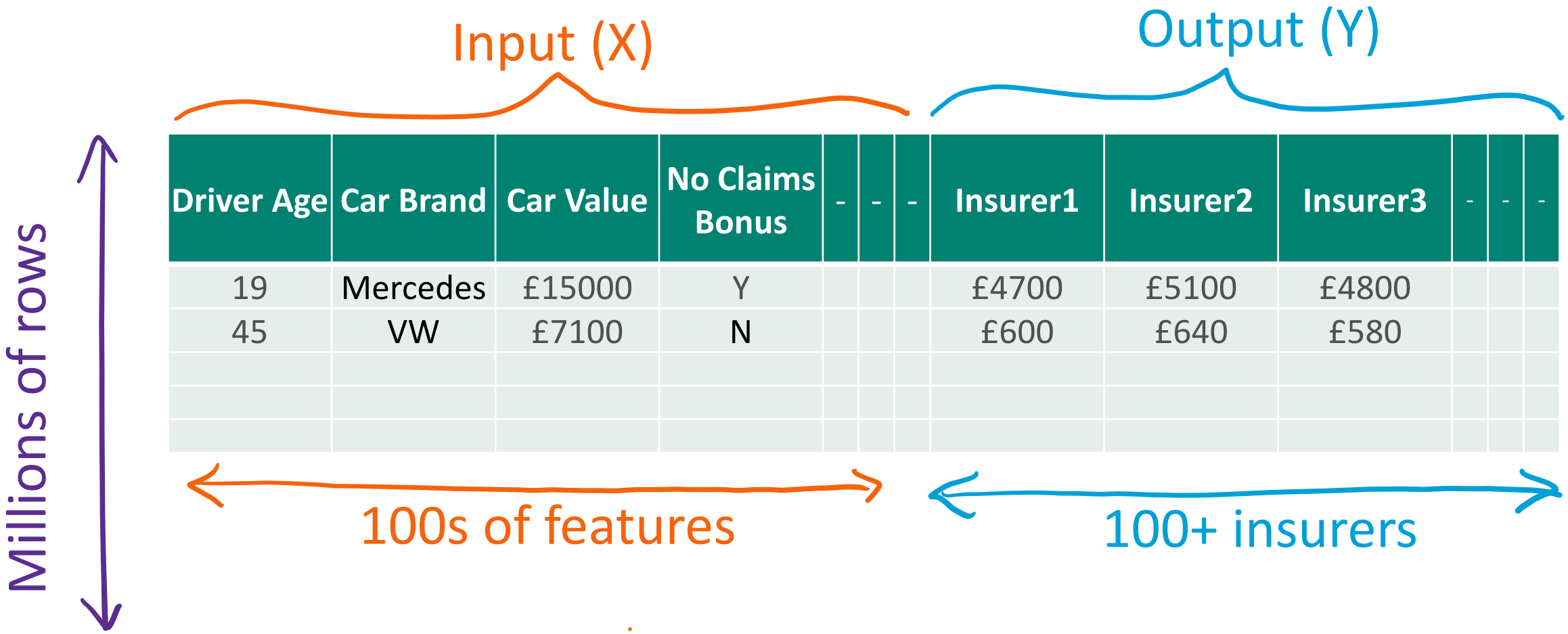
Helping customers get the best insurance premiums with a deep learning model

Our 2-part model:

- Whether or not a customer will be insured (Quote or not?).
- The best insurance quotes that a customer will get (Quote values in £).

Microsoft

Data structure



Driver Age	Car Brand	Car Value	No Claims Bonus	-	-	-	Insurer1	Insurer2	Insurer3	-	-	-
19	Mercedes	£15000	Y				£4700	£5100	£4800			
45	VW	£7100	N				£600	£640	£580			

Microsoft

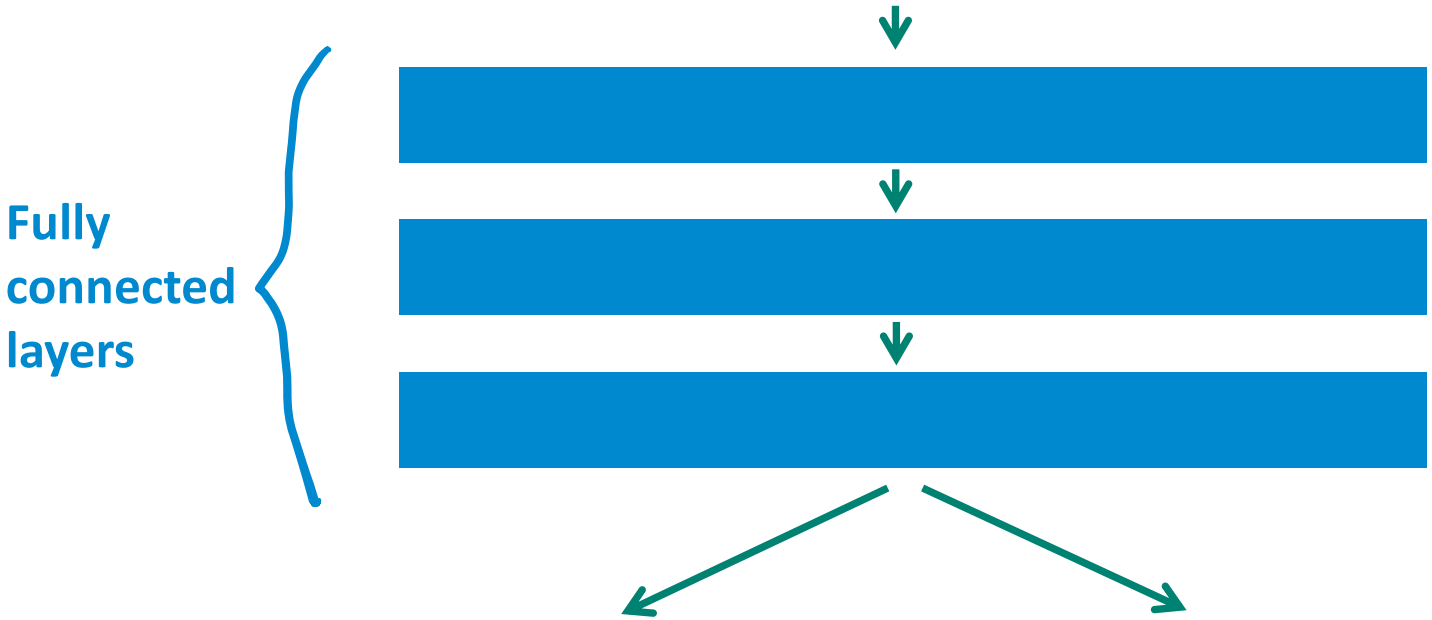
Prediction model

Model built in:

- Keras, Python, Databricks notebooks
- Keras Functional API to build novel prediction architecture: DNN with two branches

Microsoft

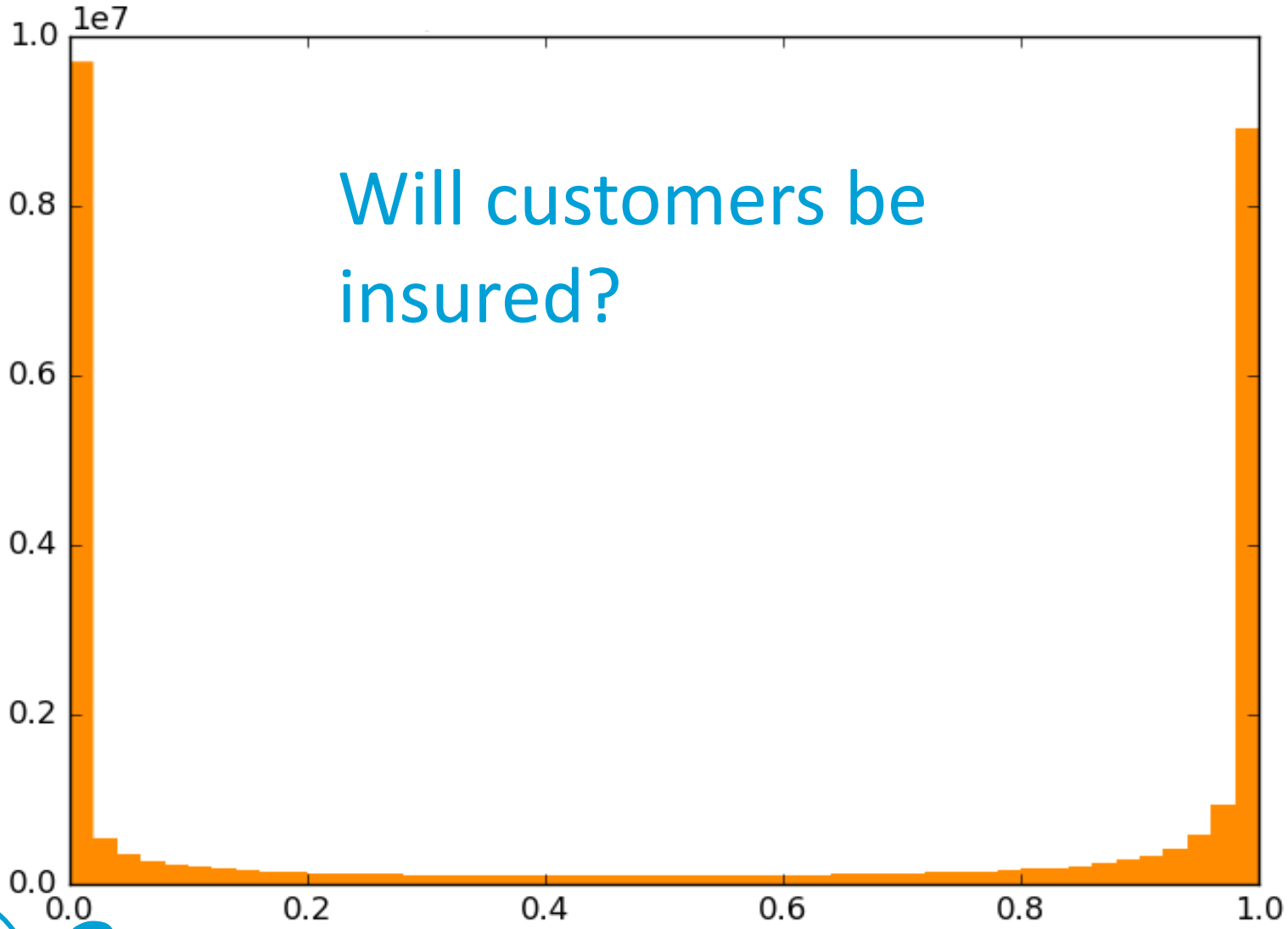
Driver age	Car brand	Car value	-	-	-	-
23	Audi	£12,000				



Insurer 1	Insurer 2	Insurer 3	-	-	-
Quote	No Quote	Quote			

Insurer 1	Insurer 2	Insurer 3	-	-	-
£600	-	£1200			

Microsoft



Will customers be insured?

0: No quote

1: Quote

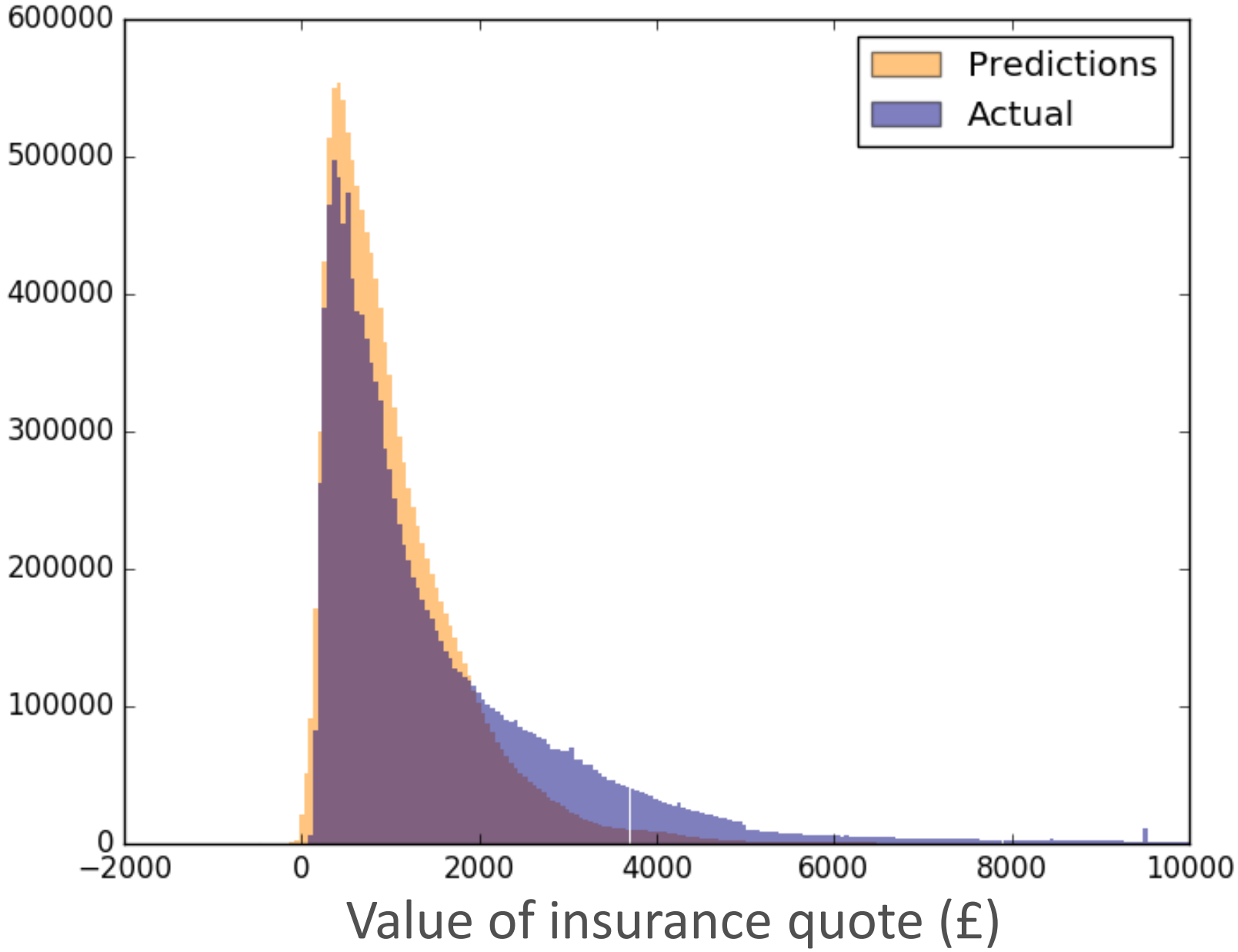
0

1

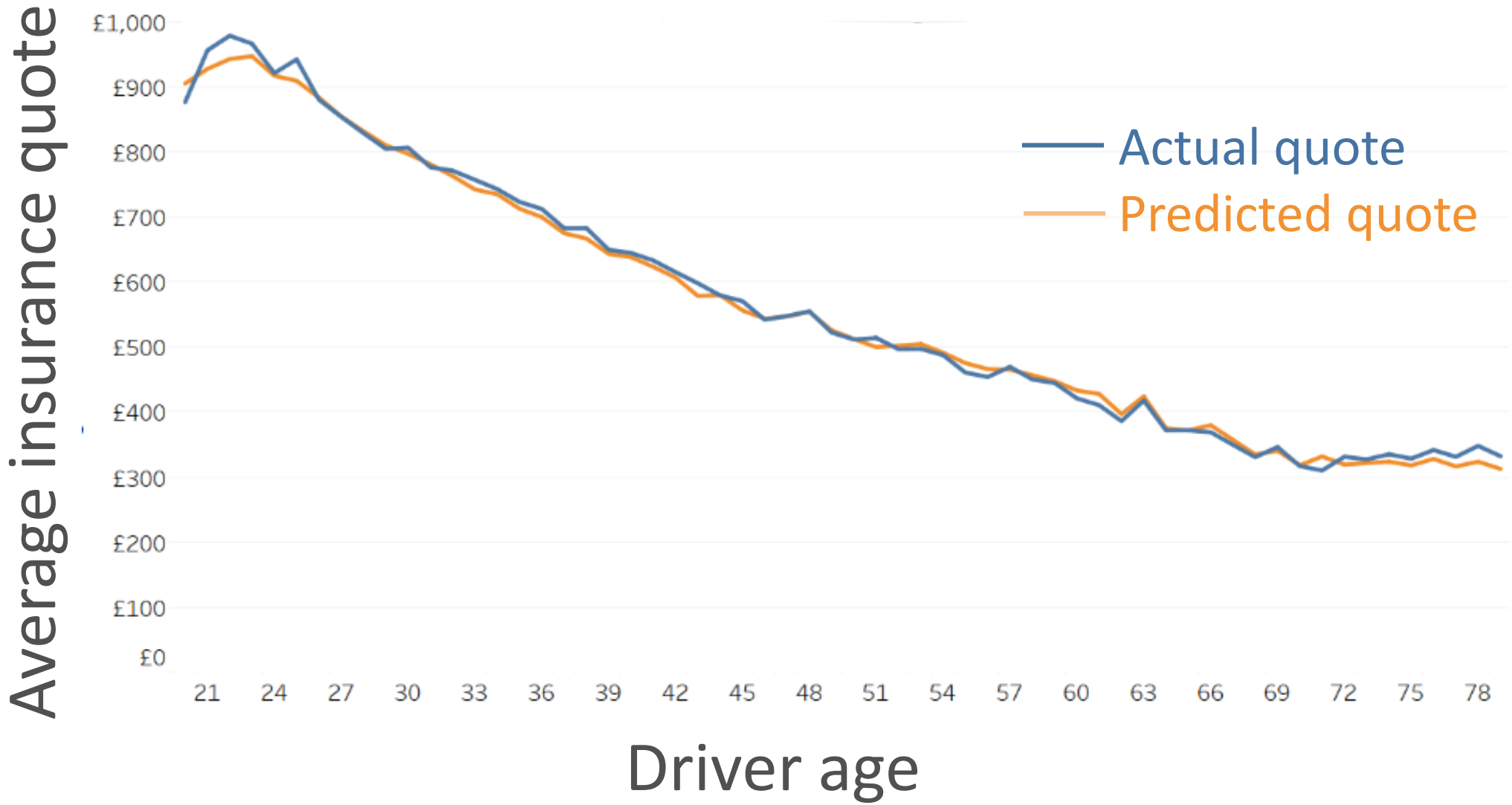
Probability of getting a quote

Microsoft

How much will customers be insured for?



Microsoft


Average insurance quote
£1,000
£900
£800
£700
£600
£500
£400
£300
£200
£100
£0
21
24
27
30
33
36
39
42
45
48
51
54
57
60
63
66
69
72
75
78
Actual quote
Predicted quote
Driver age

Microsoft

Nice model, now what?

- Now we leave the world of collaborative notebooks and data science, to the world of deployment.
- How do we get this model into a product?
- How do data scientists tweak the model when it is in production, without breaking everything?

Microsoft

Emergence of DevOps in Data Science

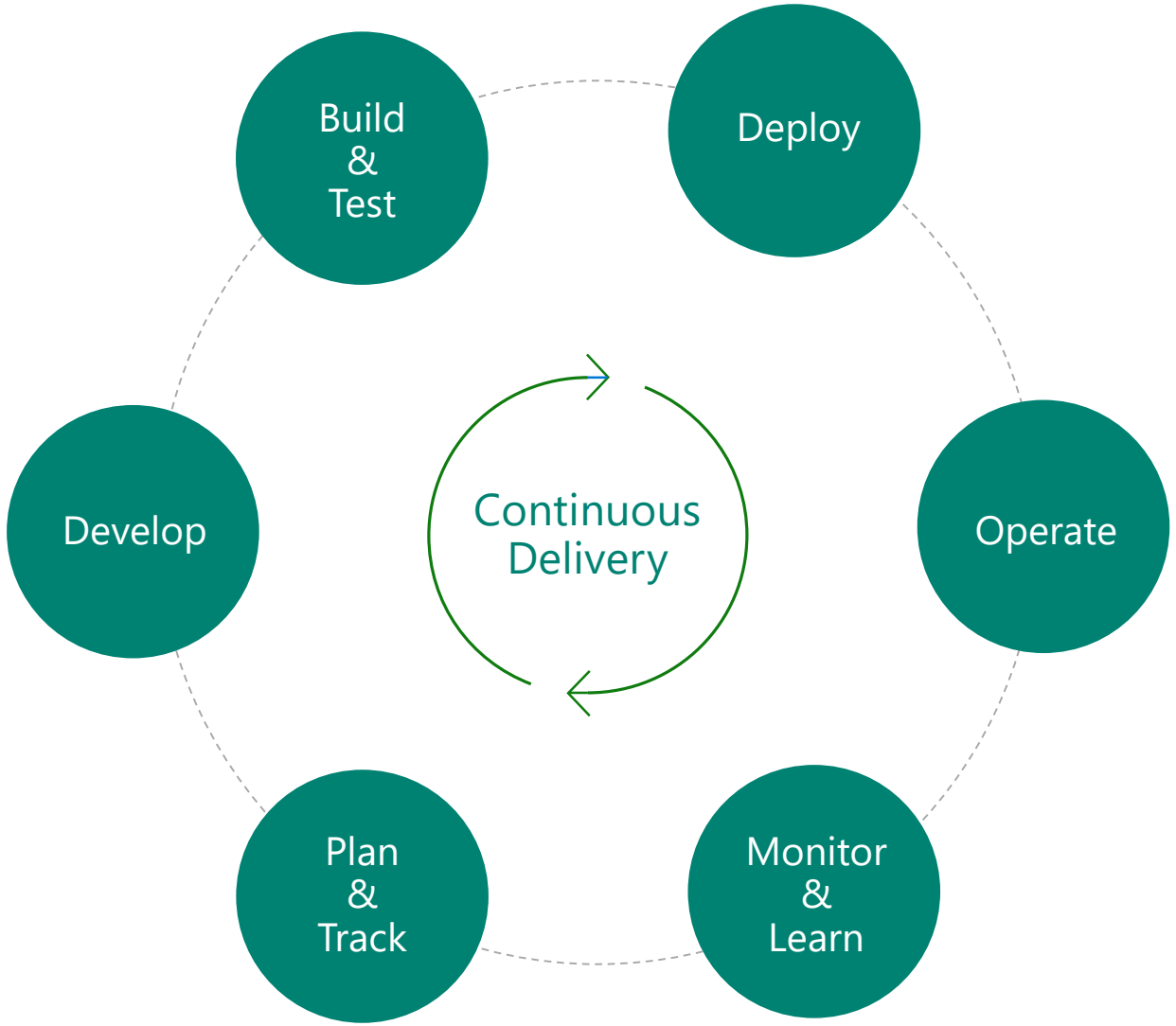
Microsoft

What is DevOps?

People. Process. Products.

DevOps is the union of **people**, **process**, and **products** to enable continuous delivery of value to your end users.

- Donovan Brown



Allianz
Allianz

Microsoft

Why DevOps is important

- Your competition is doing it already
- Increase velocity
- Reduce downtime
- Reduce human error
- Any language, any platform

Azure DevOps

Azure Boards

Deliver value to your users faster using proven agile tools to plan, track, and discuss work across your teams.

Azure Pipelines

Build, test, and deploy with CI/CD that works with any language, platform, and cloud. Connect to GitHub or any other Git provider and deploy continuously.

Azure Repos

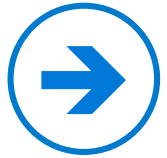
Get unlimited, cloud-hosted private Git repos and collaborate to build better code with pull requests and advanced file management.

Azure Test Plans

Test and ship with confidence using manual and exploratory testing tools.

Azure Artifacts

Create, host, and share packages with your team, and add artifacts to your CI/CD pipelines with a single click.



https://azure.com/devops

@DonovanBrown

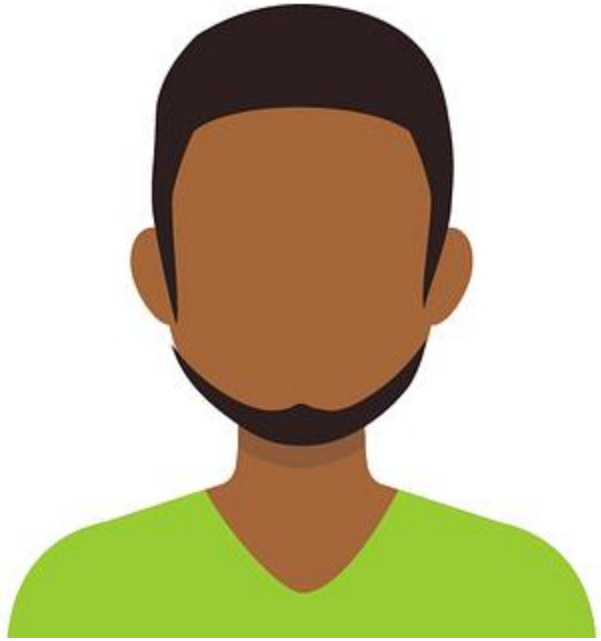
Microsoft

Adding data science to DevOps

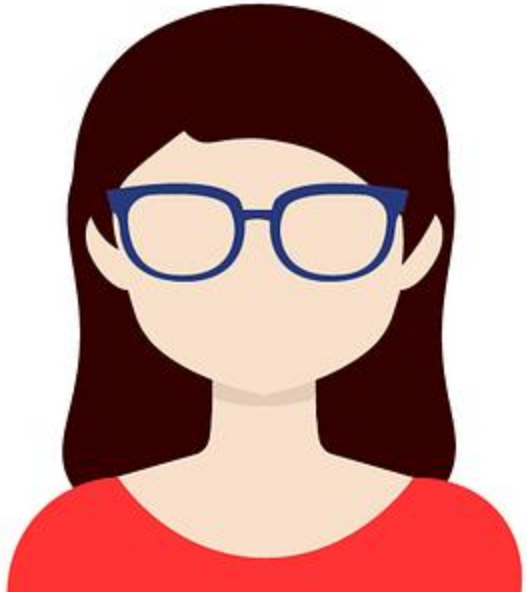
Culture clash

- Data scientists now ship directly
- DevOps in Data Science is *much* harder

Data Scientist



Software Engineer



Operations

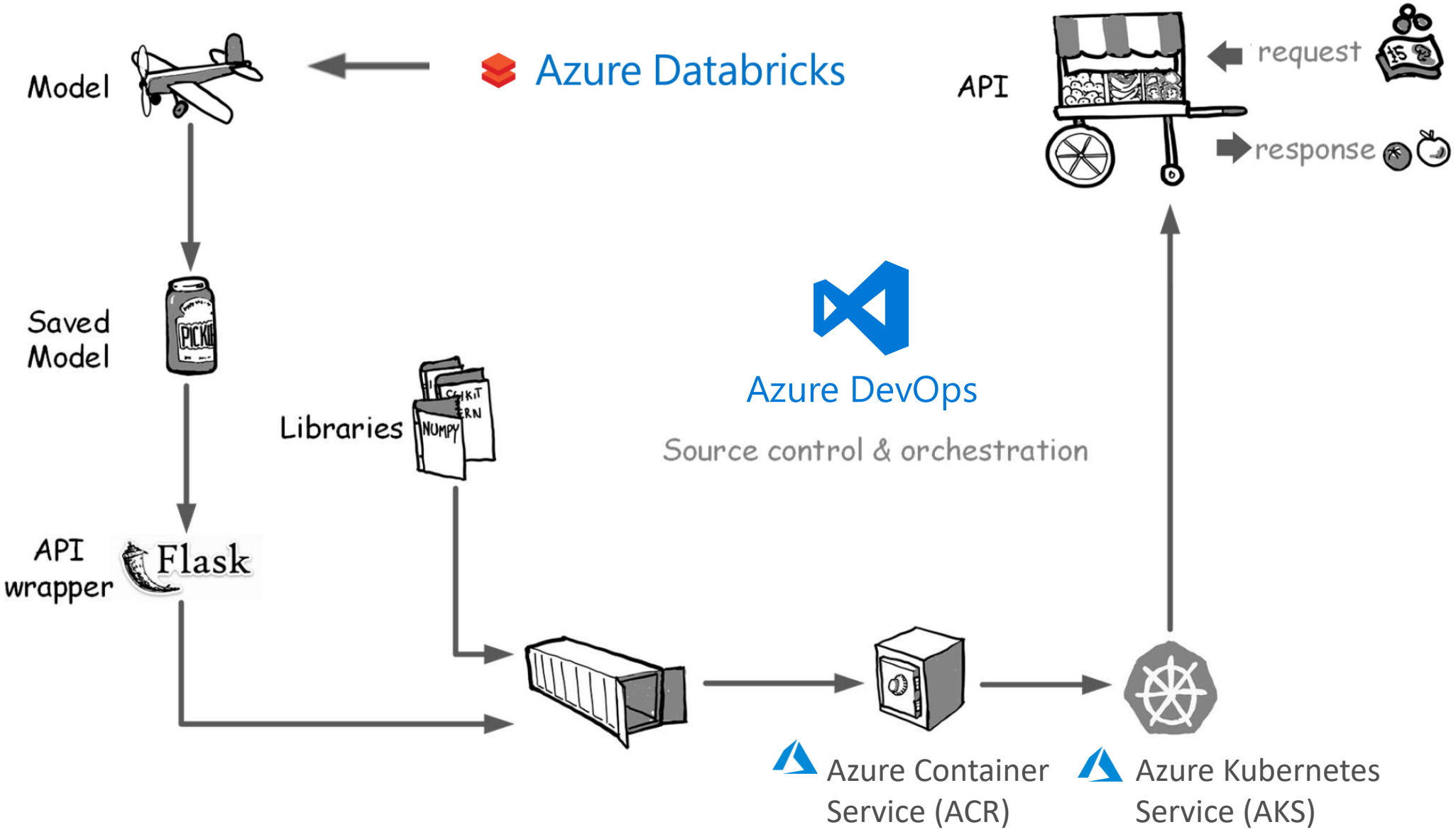
Microsoft

How did confused.com build a pipeline that enabled:

- Software quality/engineering
- Locked down dependencies
- Scalability
- GDPR awareness
- Monitoring
- CICD for models

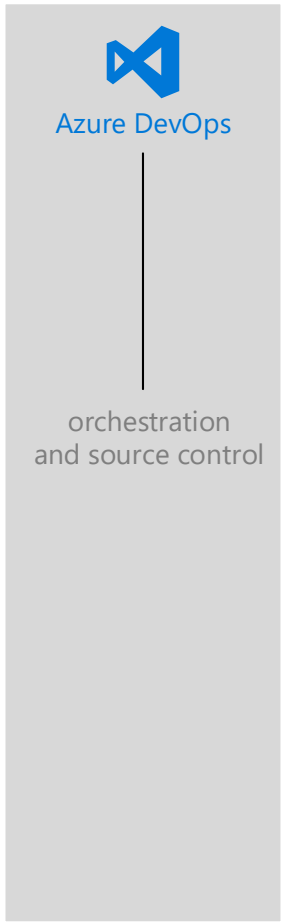
Confused.com
On a mission to save you money

Microsoft


Model
Azure Databricks
API
request
response
Saved Model
Libraries
Azure DevOps
Source control & orchestration
API wrapper
Flask
Azure Container Service (ACR)
Azure Kubernetes Service (AKS)

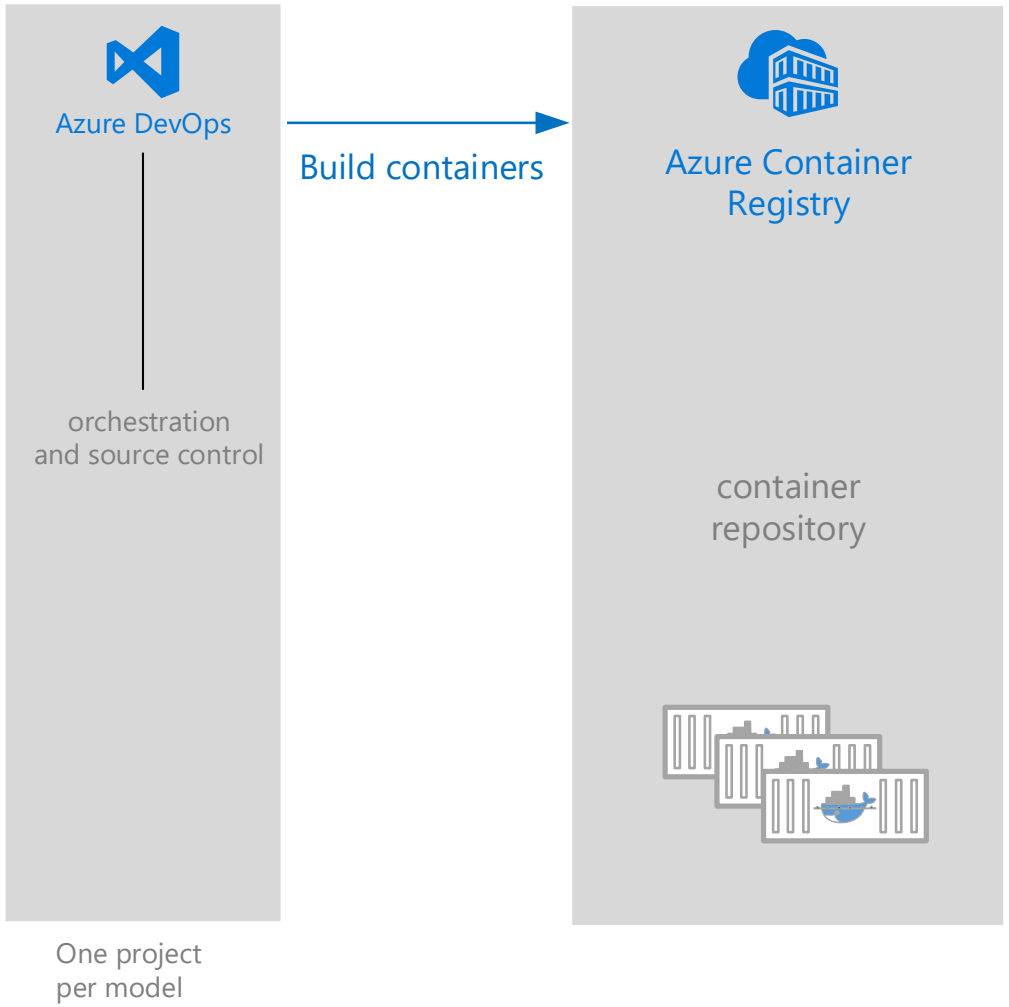
Microsoft

DevOps Architecture



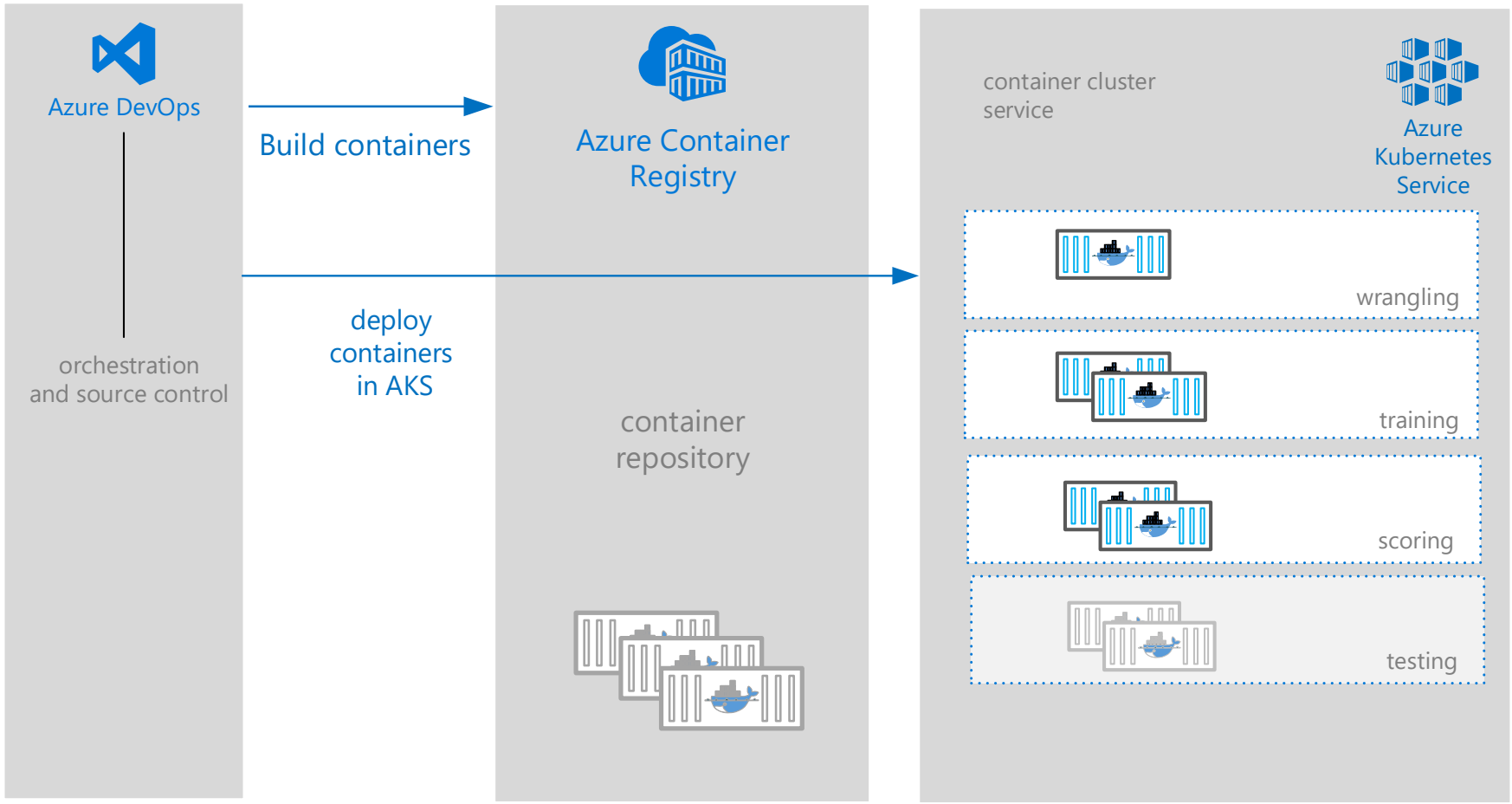
Microsoft

DevOps Architecture



Microsoft

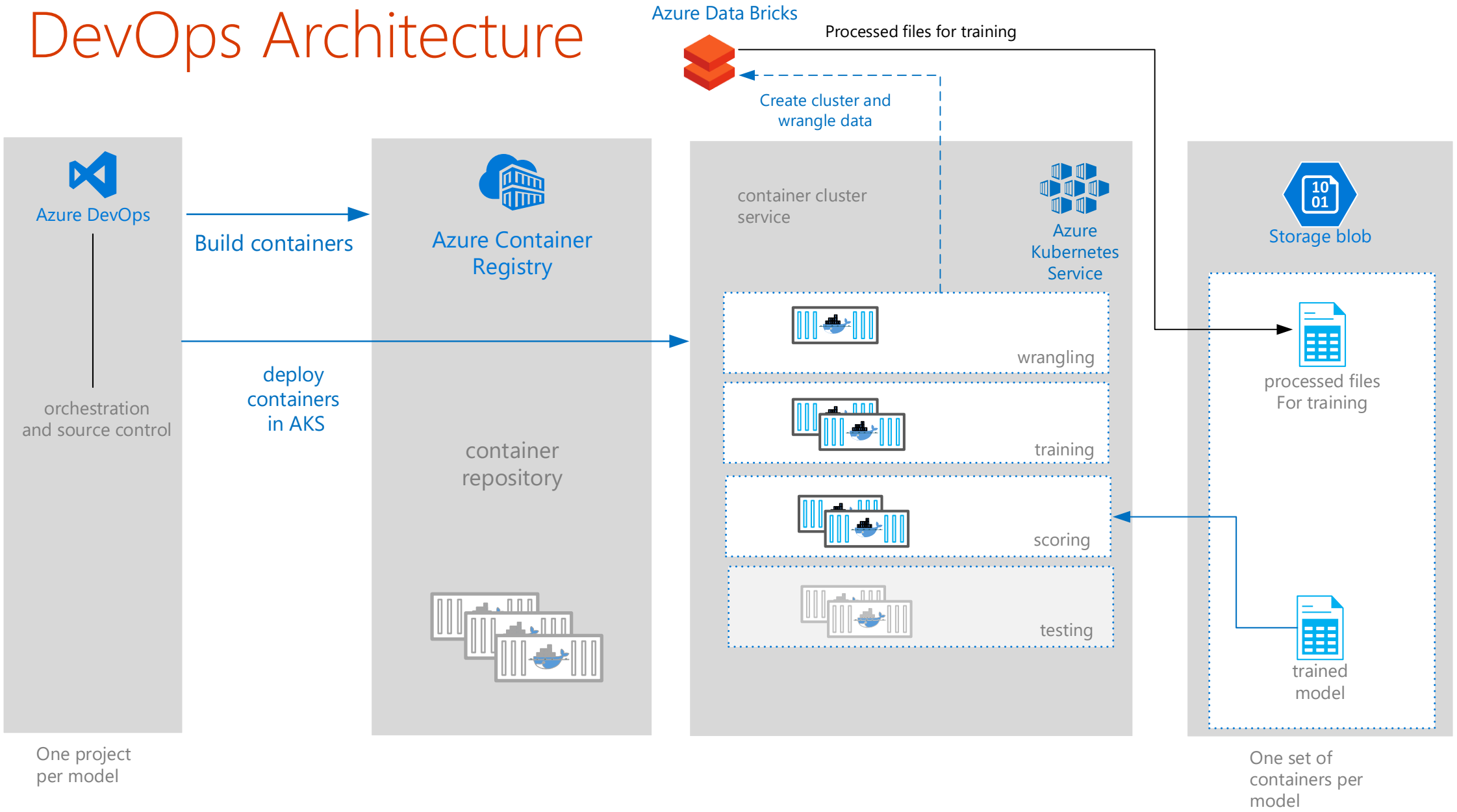
DevOps Architecture



One project
per model

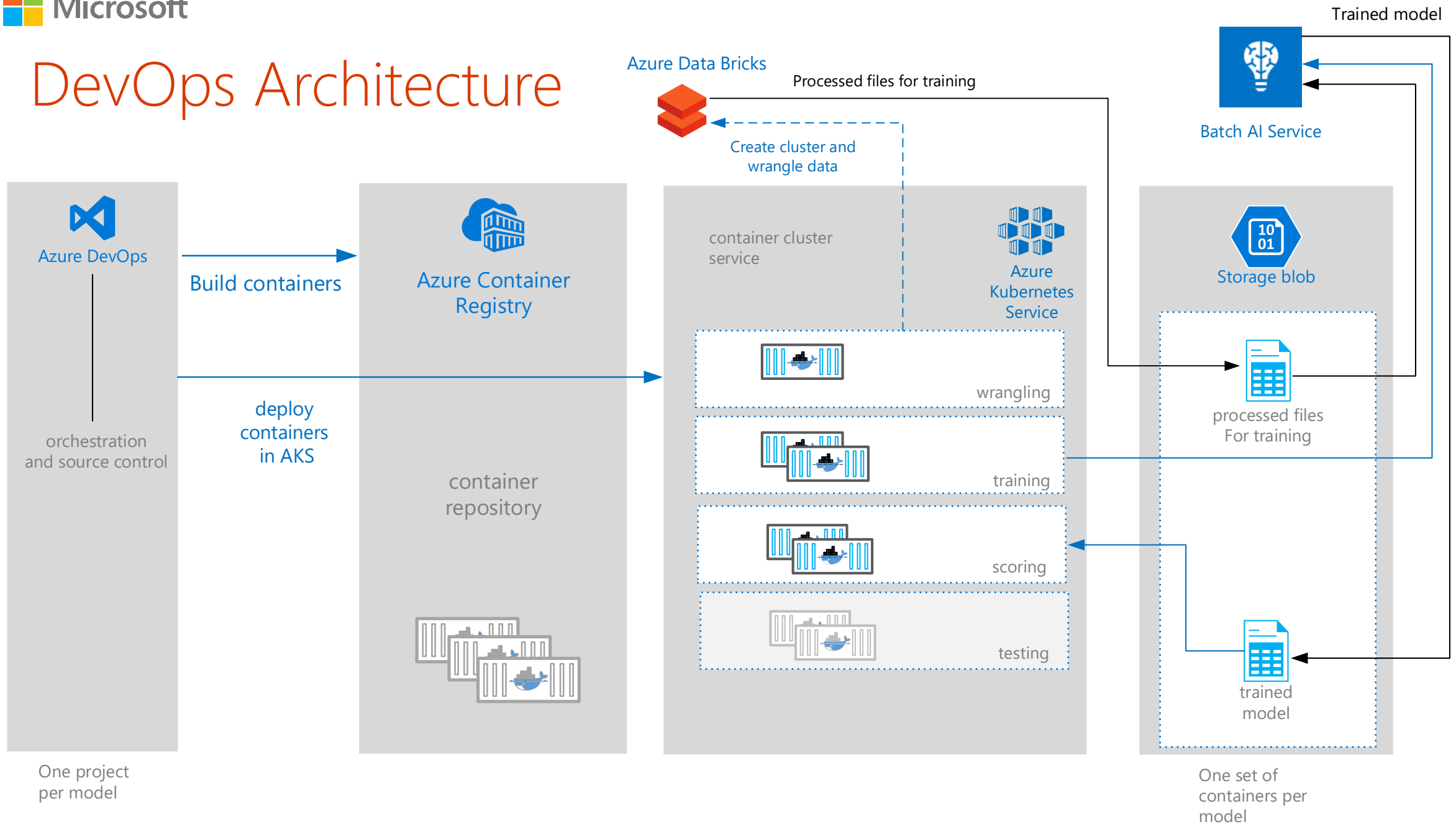
Microsoft

DevOps Architecture



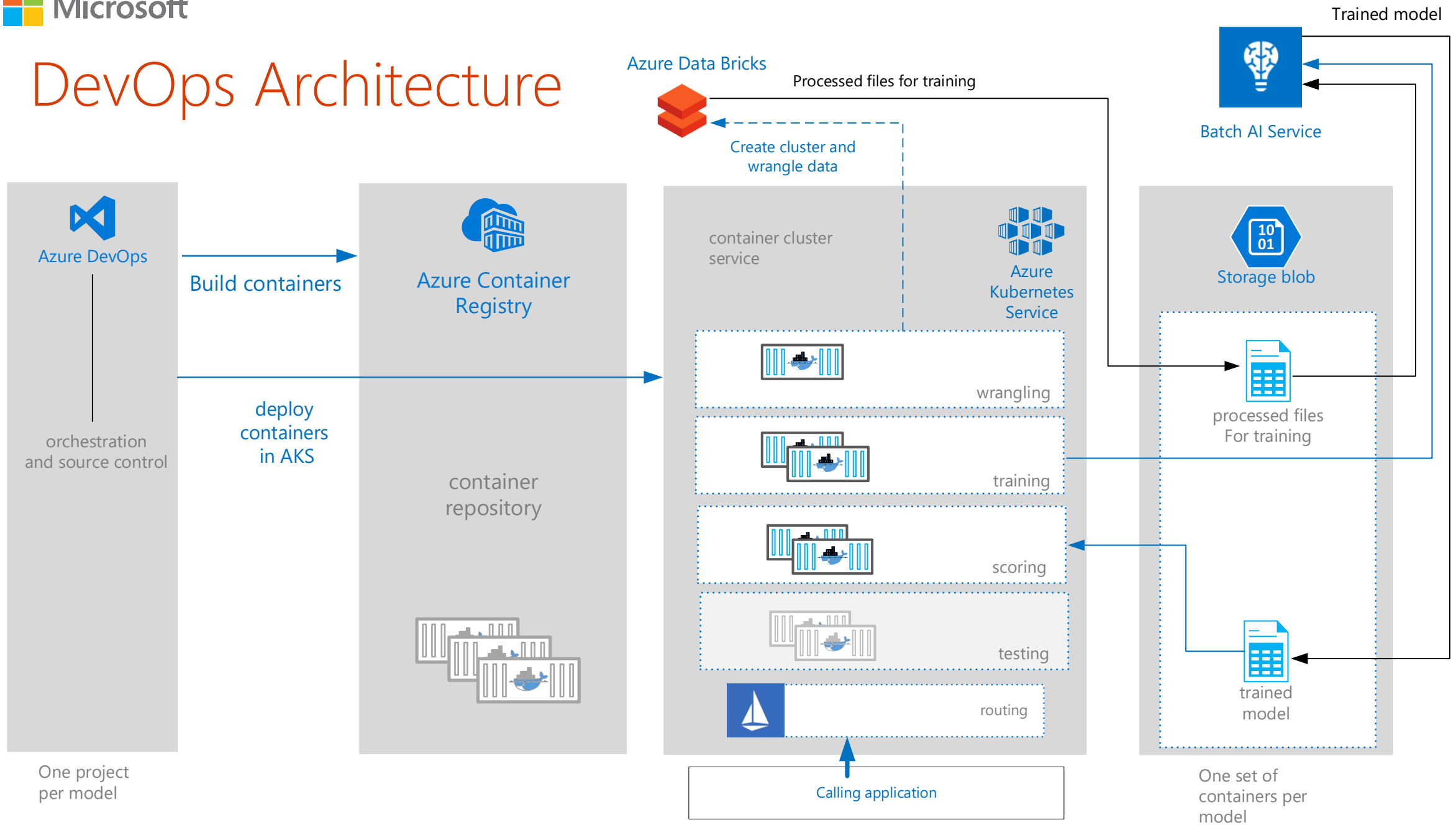
Microsoft

DevOps Architecture



Microsoft

DevOps Architecture



Microsoft

Confused.com's journey

- Data
- DevOps
- Team

Microsoft

Wrapping up

Microsoft

Conclusion & recap

- What we do at Microsoft
- The story of deep learning
- Machine learning is becoming ubiquitous in software
- Many companies are becoming data companies
- The confused.com journey
- Our deep learning model for car insurance
- The emerging field of DevOps in data science
- Our deep learning DevOps pipeline

Microsoft

What's your data story?

- Data?
- DevOps?
- Team?
- Ready!

Microsoft

Things to do next

Check out the Azure docs

DevOps: docs.microsoft.com/en-us/azure/devops

Databricks: docs.azuredatabricks.net/

Containers: azure.microsoft.com/en-us/services/container-registry

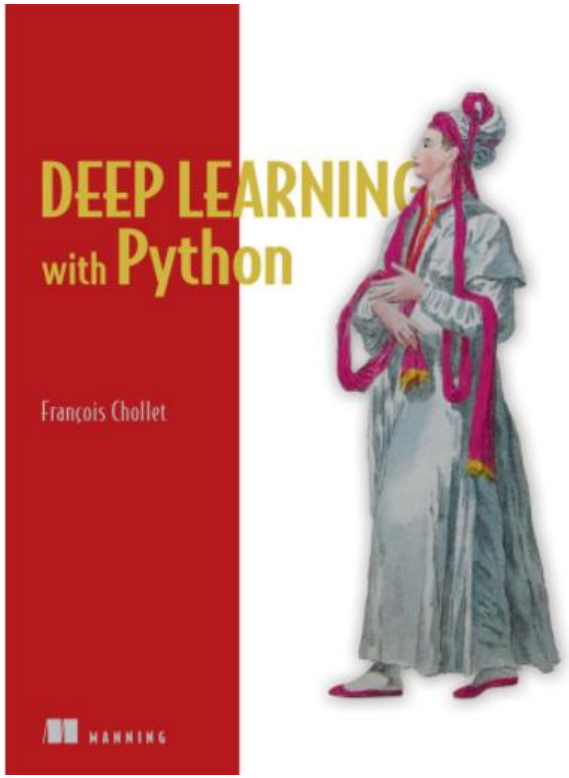
Kubernetes: azure.microsoft.com/en-us/services/kubernetes-service

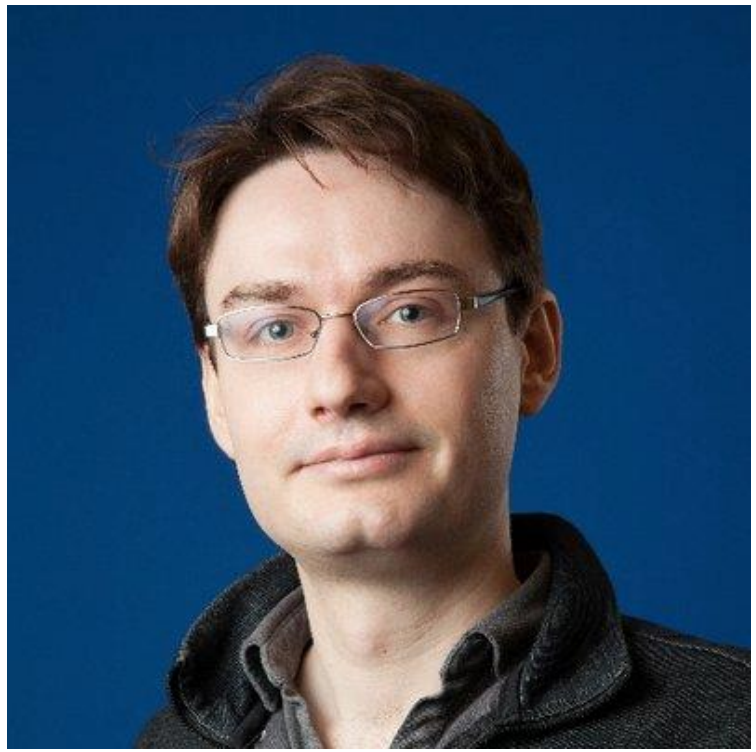
Session Feedback

Please rate this session in the Future Decoded app!

Microsoft

Best intro book on deep learning





Francois Chollet @fchollet

https://www.manning.com/books/deep-learning-with-python
https://github.com/fchollet/deep-learning-with-python-notebooks/

Microsoft

ank You!

@Ecsquendor @Dr_Tempest #timtem

ft CSE team:

, Tempest van Schaik, Andrew Fryer, Luke Devonshire, Anita Ramanan, Alysha

d.com team:

ding, Tim Lang, Ben Weatherburn, Vaughan Pugsley, Rex Johnson, Simon
h, Danielle Mansfield, Alice Burnett, David Trickey

Microsoft

Microsoft Future Decoded

Confused.com
On a mission to save you money

Productionising deep learning: The confused.com journey

Dr. Tim Scarfe
@ecsquendor

Dr. Tempest van Schaik
@Dr_Tempest

Dr Tim + Dr Tem aka #TimTem