ONE DOES NOT SIMPLY

DEPLOY ML INTO PRODUCTION

Henrik Brink

Machine Learning Engineering @ Wise.io / GE Digital

Wise.io

brinkar

Agenda

- From space to industrial machine learning
- Challenges
  - Optimization dimensions
  - Infrastructure
  - Development and deployment
- Solutions
  - The ML meta-algorithm
  - Containerization
  - Best engineering practices
- Wrap-up and questions

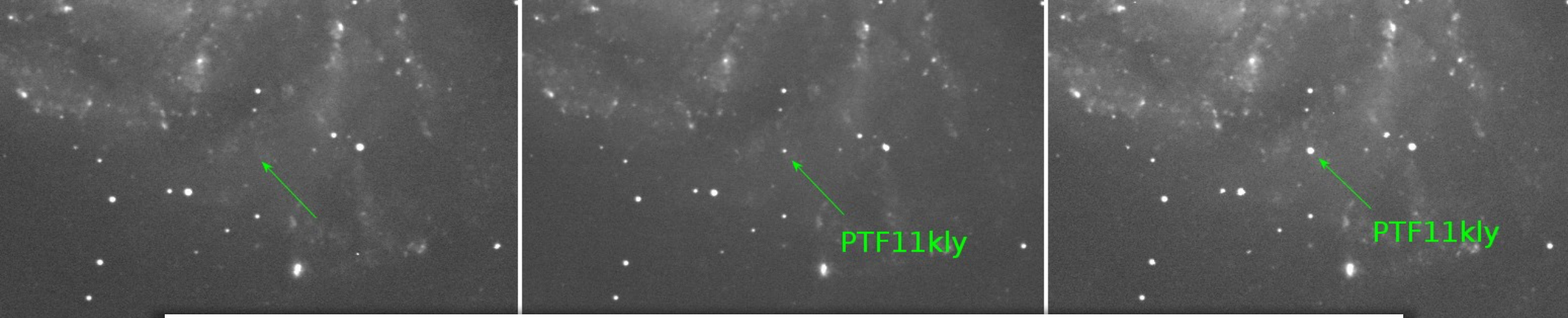
Wise.io

brinkar

From astronomy to industrial machine learning...

Wise.io

brinkar



MNRAS **435,** 1047–1060 (2013)
Advance Access publication 2013 August 29

doi:10.1093/mnras/stt1306

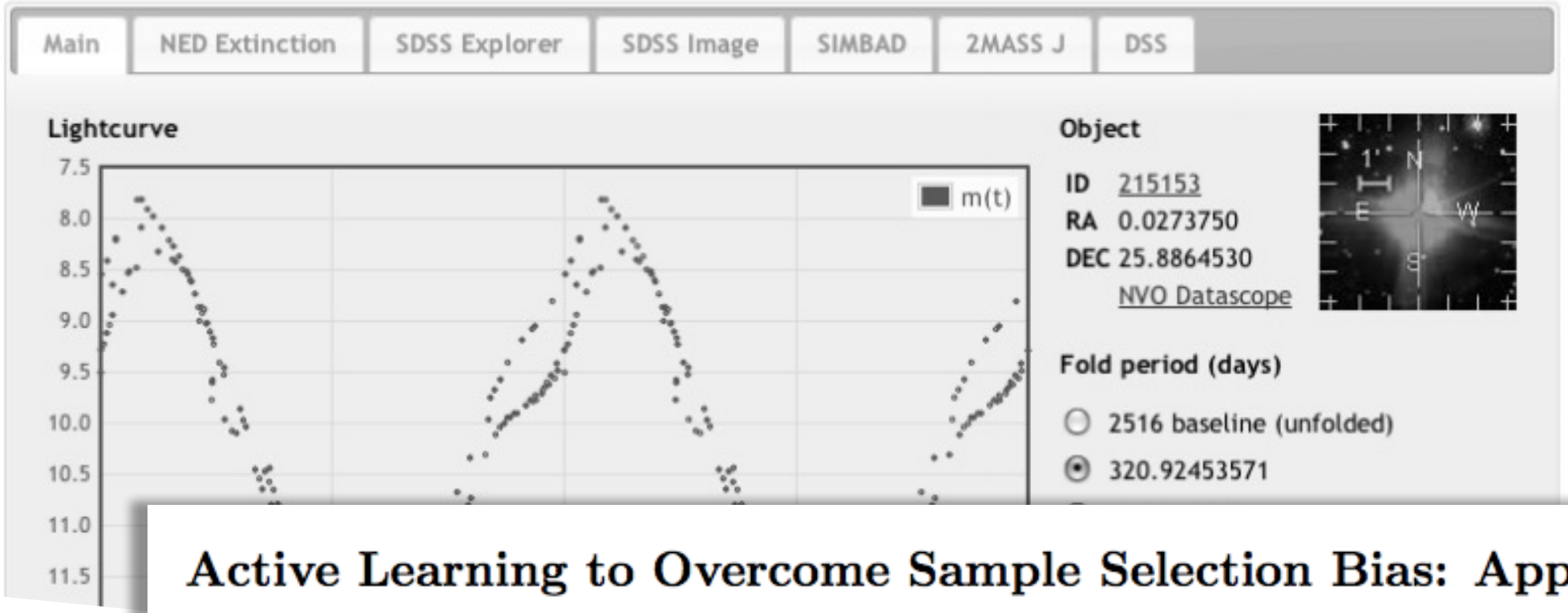
Using machine learning for discovery in synoptic survey imaging data

Henrik Brink,[1]★ Joseph W. Richards,[1,2] Dovi Poznanski,[3] Joshua S. Bloom,[1]
John Rice,[2] Sahand Negahban[4] and Martin Wainwright[2,4]

[1]*Department of Astronomy, University of California Berkeley, Berkeley, CA 94720, USA*
[2]*Department of Statistics, University of California Berkeley, Berkeley, CA 94720, USA*
[3]*School of Physics and Astronomy, Tel Aviv University, Tel Aviv 6997801, Israel*
[4]*Department of Electrical Engineering and Computer Sciences, University of California Berkeley, Berkeley, CA 94720, USA*

Wise.io

brinkar



Active Learning to Overcome Sample Selection Bias: Application to Photometric Variable Star Classification

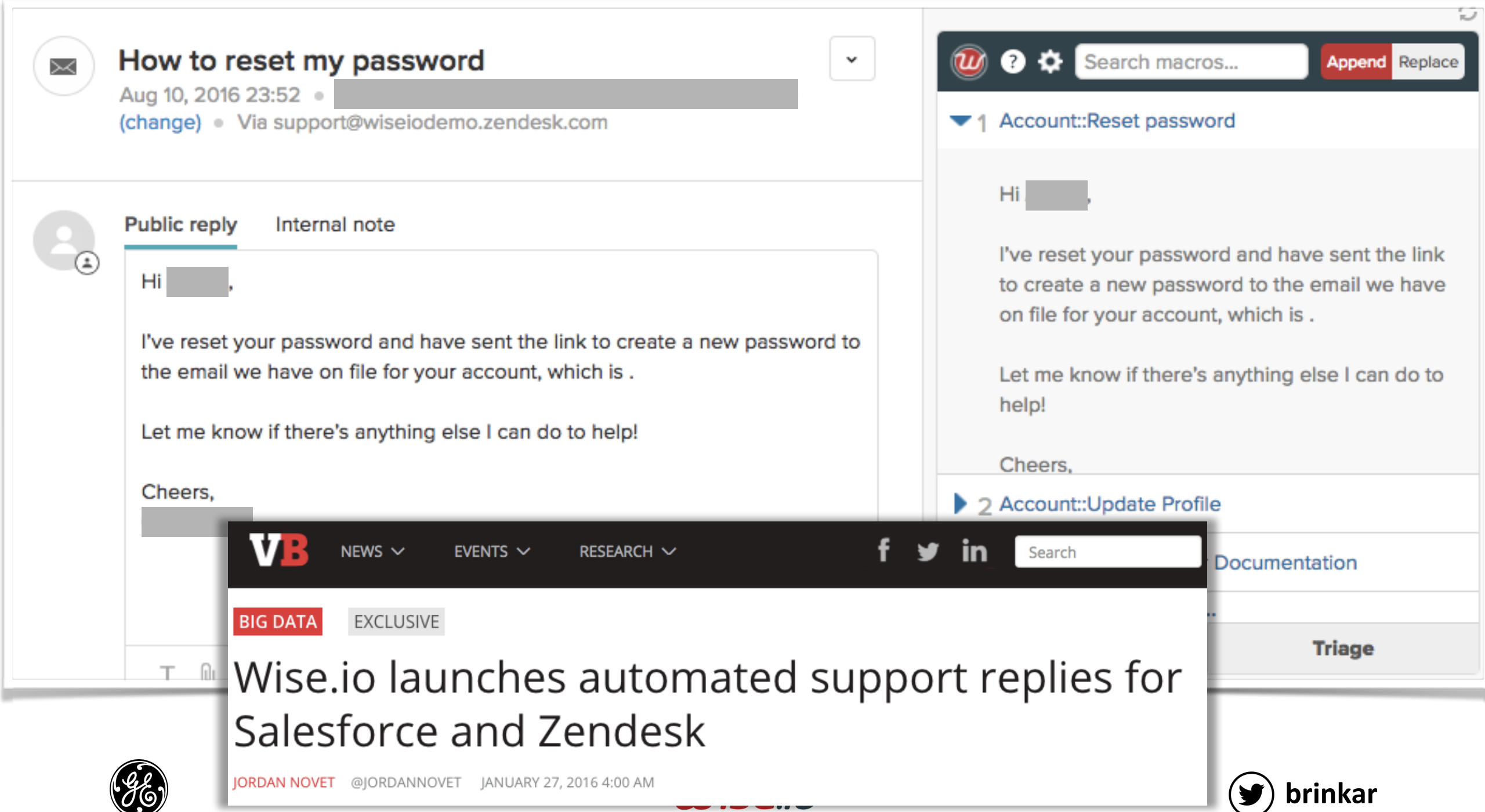
Joseph W. Richards[1,2,*], Dan L. Starr[1], Henrik Brink[3], Adam A. Miller[1], Joshua S. Bloom[1], Nathaniel R. Butler[1], J. Berian James[1,3], James P. Long[2] and John Rice[2]

Wise.io

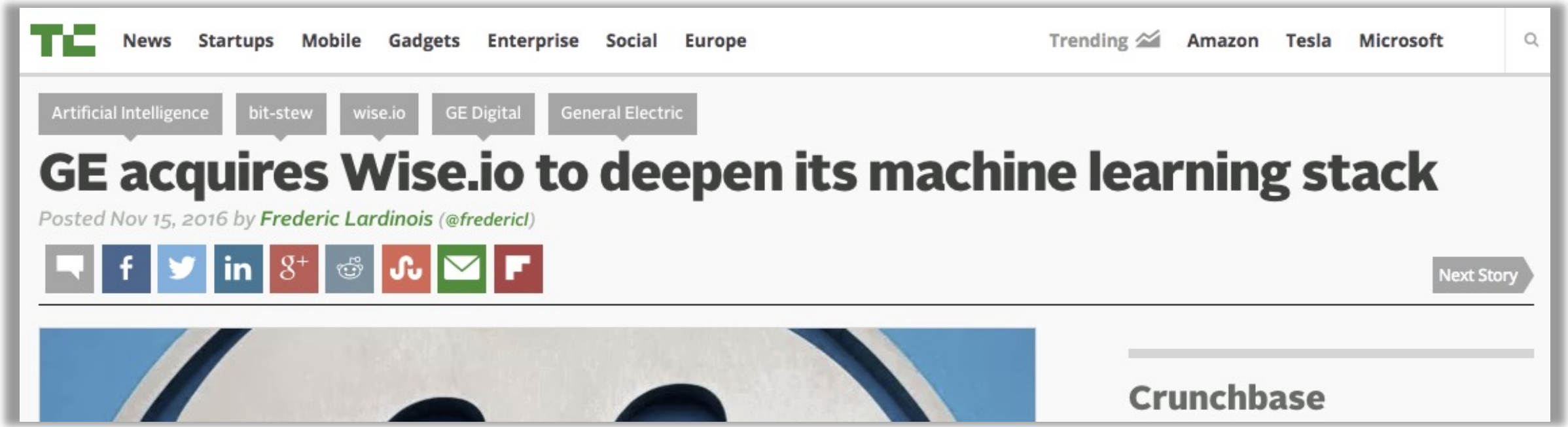
brinkar

Wise.io

brinkar

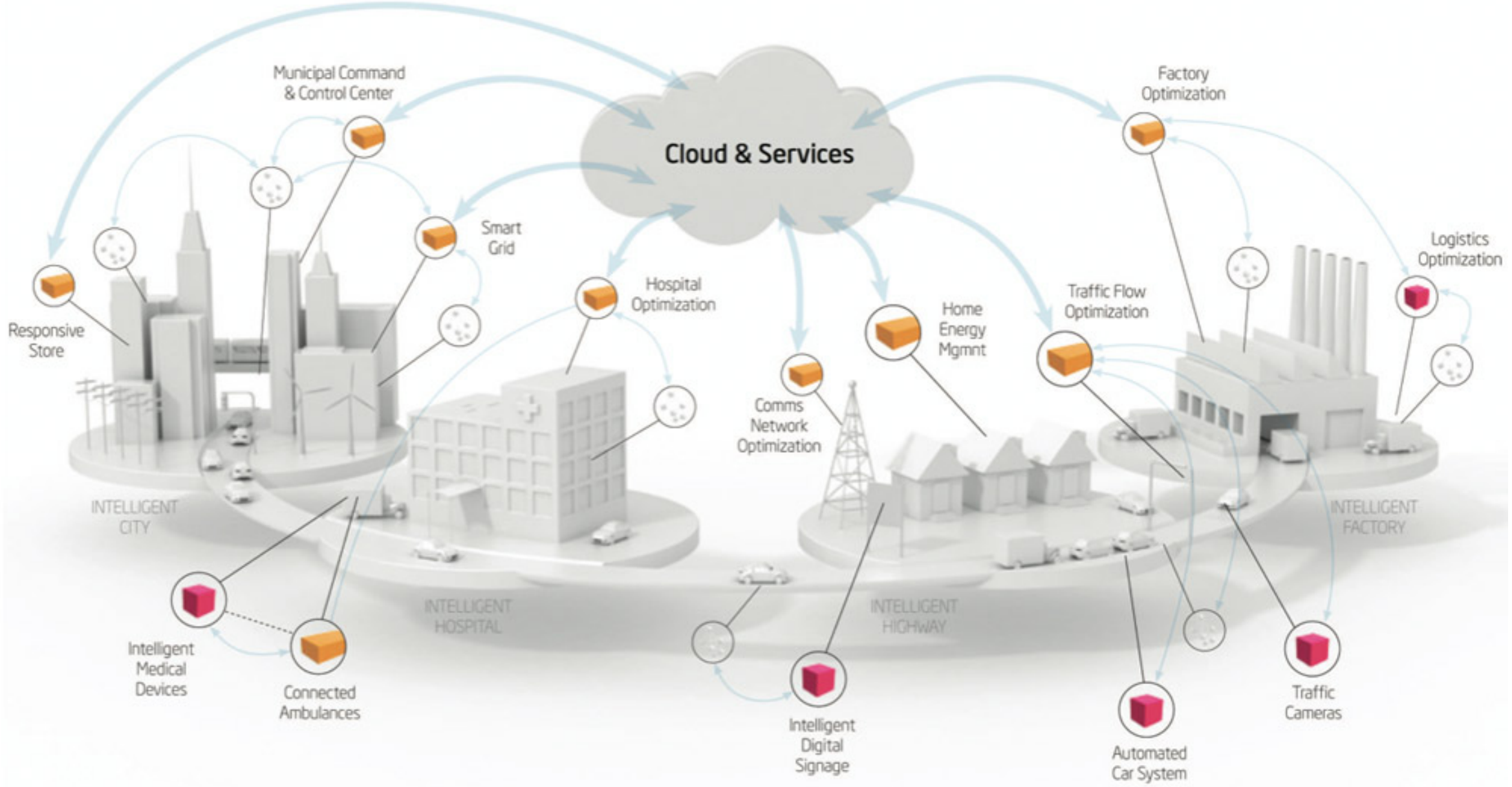

How to reset my password
Aug 10, 2016 23:52
(change) • Via support@wiseiodemo.zendesk.com
Public reply
Internal note
Hi ,
I've reset your password and have sent the link to create a new password to the email we have on file for your account, which is .
Let me know if there's anything else I can do to help!
Cheers,
Search macros...
Append
Replace
1 Account::Reset password
Hi ,
I've reset your password and have sent the link to create a new password to the email we have on file for your account, which is .
Let me know if there's anything else I can do to help!
Cheers,
2 Account::Update Profile
Documentation
Triage
VB
NEWS
EVENTS
RESEARCH
Search
BIG DATA
EXCLUSIVE
Wise.io launches automated support replies for Salesforce and Zendesk
JORDAN NOVET @JORDANNOVET JANUARY 27, 2016 4:00 AM

brinkar



Wise.io

brinkar


Cloud & Services
Municipal Command & Control Center
Factory Optimization
Smart Grid
Responsive Store
Hospital Optimization
Home Energy Mgmnt
Traffic Flow Optimization
Logistics Optimization
Comms Network Optimization
INTELLIGENT CITY
INTELLIGENT HOSPITAL
INTELLIGENT HIGHWAY
INTELLIGENT FACTORY
Intelligent Medical Devices
Connected Ambulances
Intelligent Digital Signage
Automated Car System
Traffic Cameras

brinkar

Wise.io

brinkar

Everyone ready…?

Wise.io

brinkar

ONE DOES NOT SIMPLY
DEPLOY ML INTO PRODUCTION

What to optimize for when building ML systems?

Wise.io

brinkar

Accuracy?

Implementation cost?

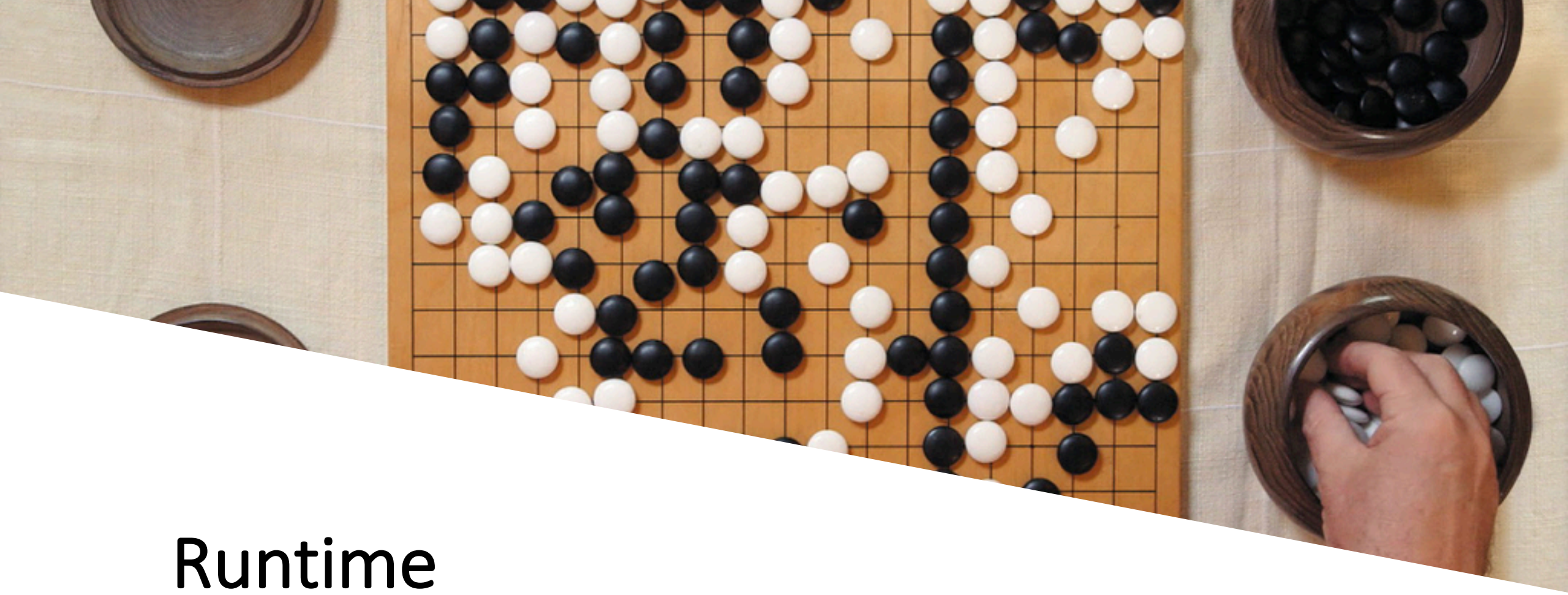
On the NETFLIX Prize

"We evaluated some of the new methods offline but the **additional accuracy gains** that we measured **did not seem to justify the engineering effort** needed to bring them into a **production environment.**"

Xavier Amatriain and Justin Basilico (April 2012)

Wise.io

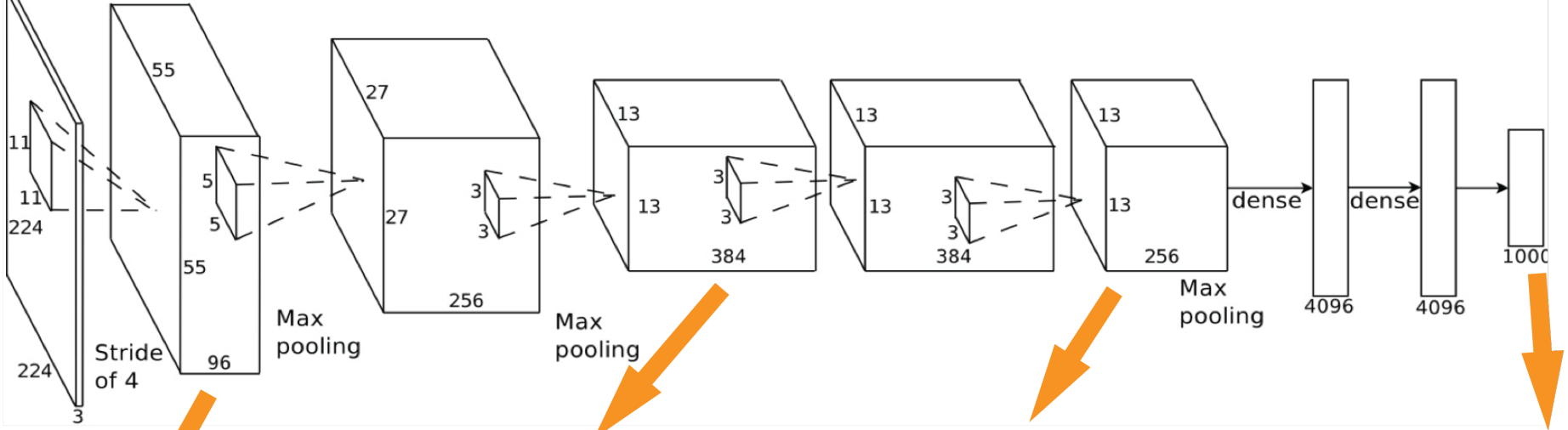
brinkar

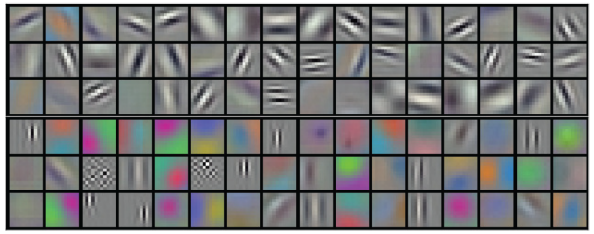


Runtime cost?

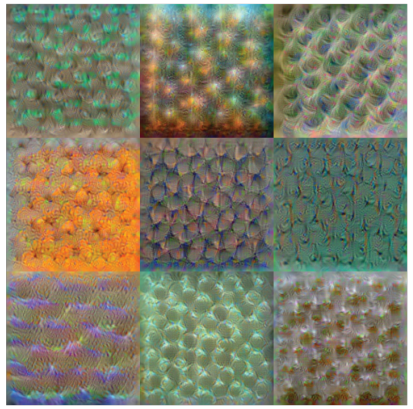
AlphaGo

1920 CPUs
280 GPUs





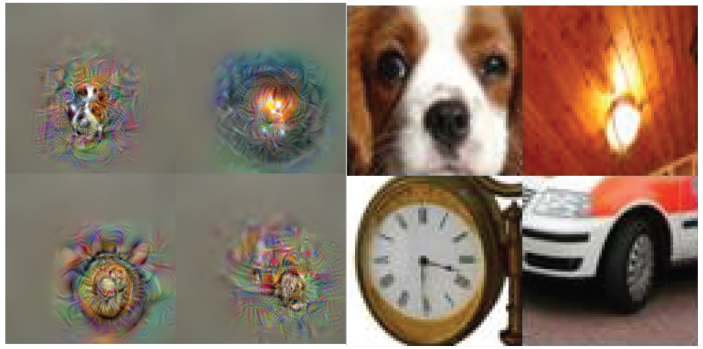
Conv 1: Edge+Blob



Conv 3: Texture

Numerical

Data-driven



Conv 5: Object Parts

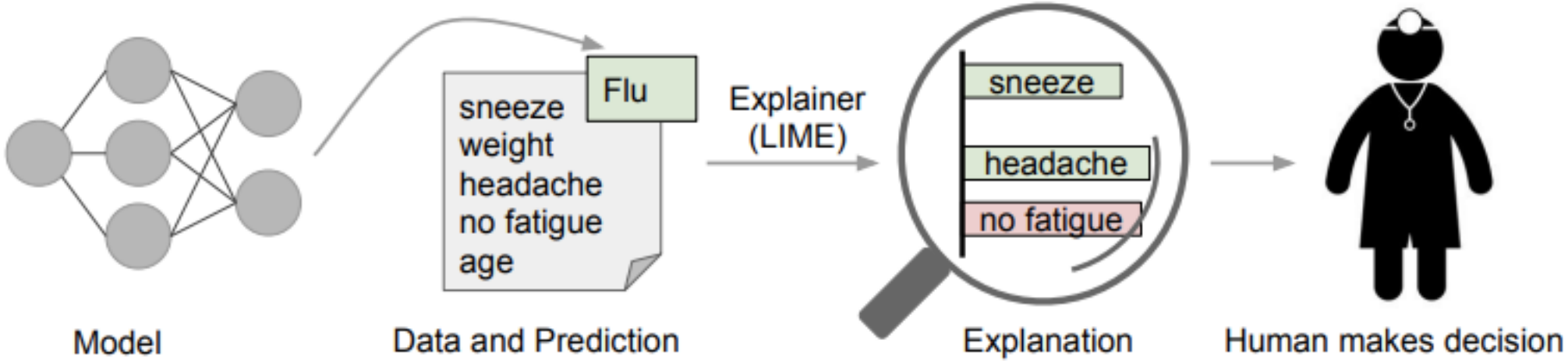
Fc8: Object Classes

Interpretability?

Wise.io

brinkar

LOCAL INTERPRETABLE MODEL-AGNOSTIC EXPLANATIONS **(LIME**



arxiv.org/abs/1602.04938

Automation
vs augmentation?

What to optimize for when building ML systems?

- Accuracy?
- Implementation cost?
- Runtime cost?
- Interpretability?
- Automation vs augmentation?

Wise.io

brinkar

A 5-step ML meta-algorithm

1. Identify and define the problem
2. Collect and understand the data
3. Build and deploy a simple model that works end-to-end
4. Iterate to optimize an deploy improved models (inner loop)
5. Monitor and back-propagate changes to problem parameters (outer loop)

Wise.io

brinkar

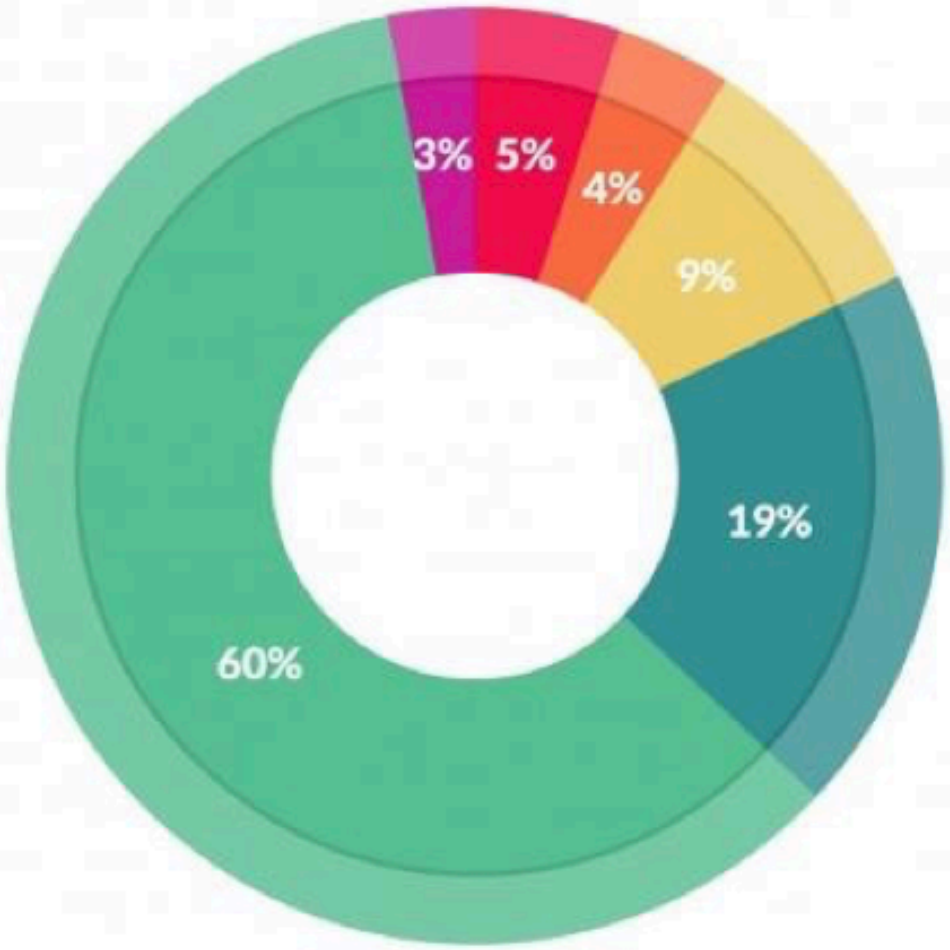
1. Identify and define the problem
5. Monitor and back-propagate changes to problem parameters

Predicted Class	True Outcome : Patients have Disease A	
	Positive (Patients have disease A)	Negative (Patients do not have disease A)
Positive (Patients have disease A)	True Positives	False Positives (Patients wrongly identified to have disease A)
Negative (Patients do not have disease A)	False Negatives (Patients have been left out from treatment for Disease)	True Negatives

Wise.io

brinkar

2. Collect and understand the data



What data scientists spend the most time doing

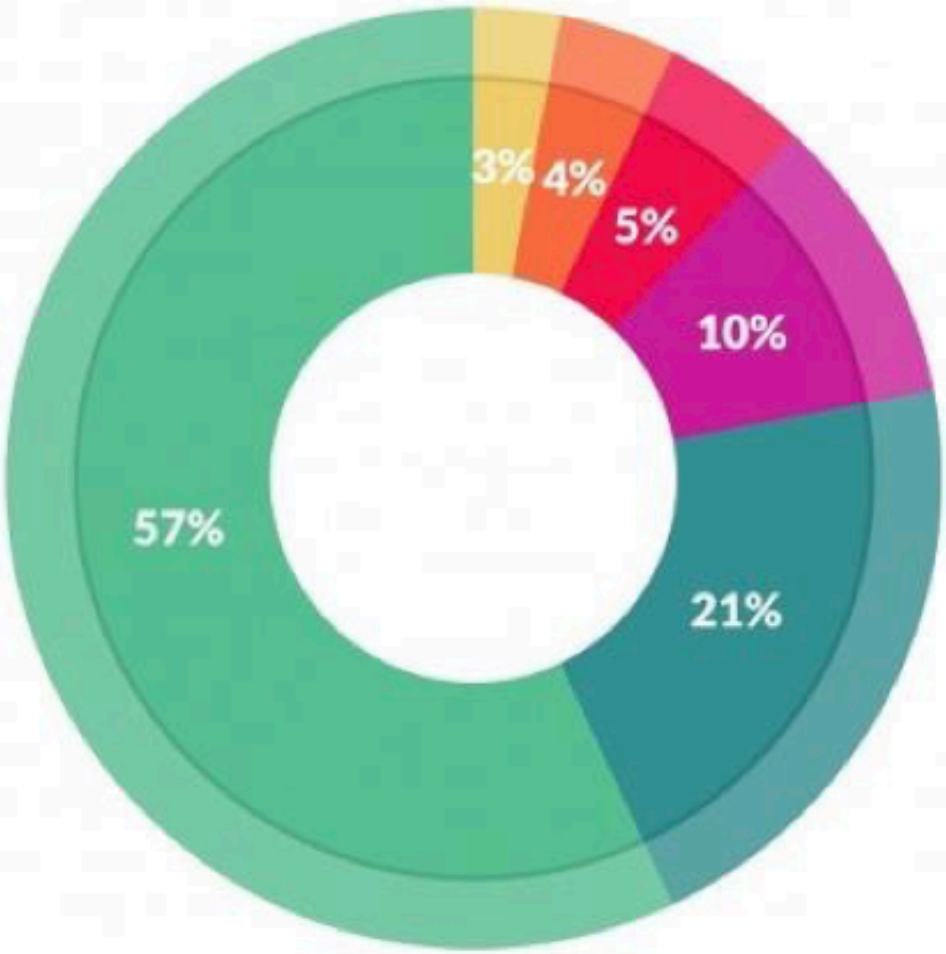
- *Building training sets: 3%*
- *Cleaning and organizing data: 60%*
- *Collecting data sets; 19%*
- *Mining data for patterns: 9%*
- *Refining algorithms: 4%*
- *Other: 5%*

Forbes

Wise.io

brinkar

2. Collect and understand the data



What's the least enjoyable part of data science?

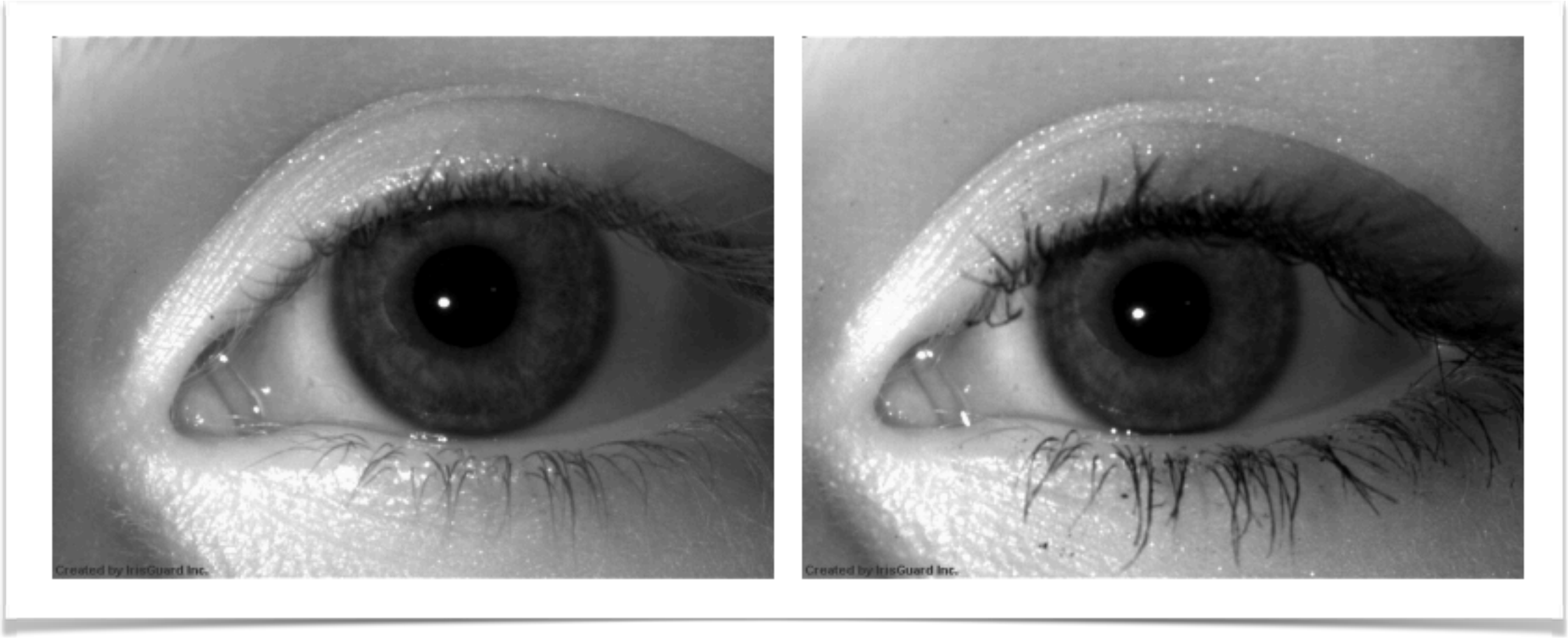
- *Building training sets: 10%*
- *Cleaning and organizing data: 57%*
- *Collecting data sets: 21%*
- *Mining data for patterns: 3%*
- *Refining algorithms: 4%*
- *Other: 5%*

Forbes

Wise.io

brinkar

2. Collect and understand the data



Wise.io

brinkar

3. Deploy simple model end-to-end
4. Iterate to optimize an deploy improved models

- Simplest possible model that solves the problem
- End-to-end production deployment: automated deployment, testing, logging, monitoring, feedback

Wise.io

brinkar

Getting closer...

Wise.io

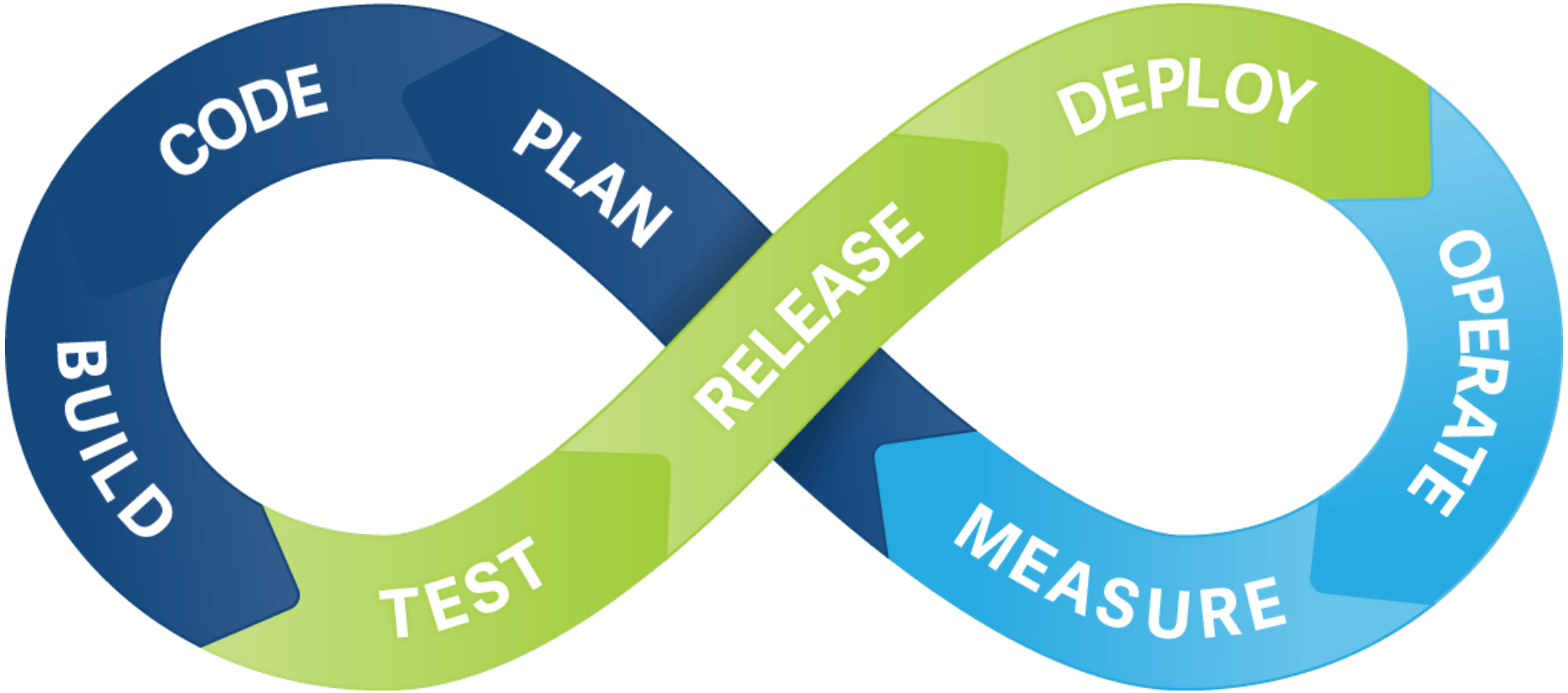
brinkar

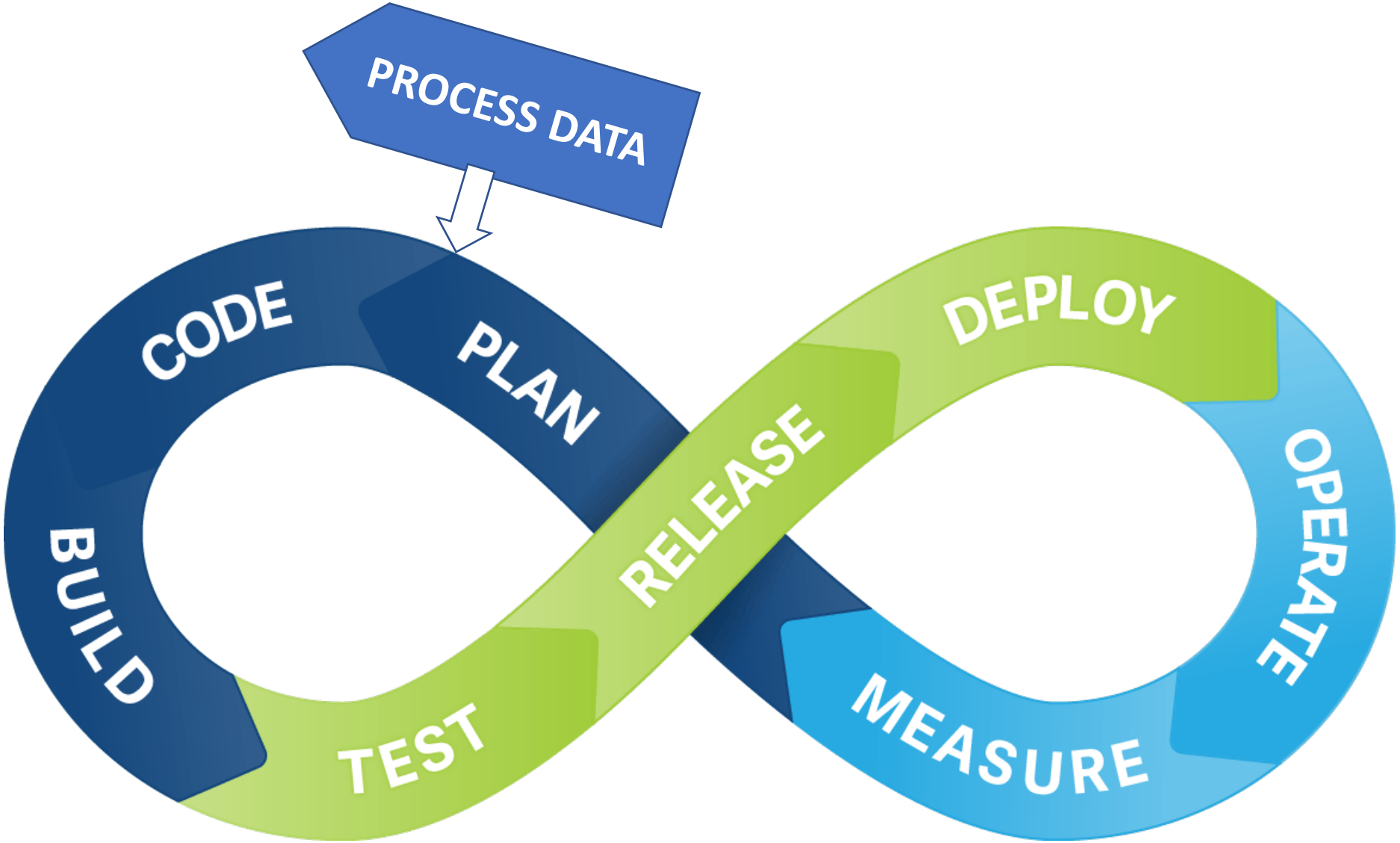
Machine learning infrastructure

Wise.io

brinkar

Continuous integration





Wise.io

brinkar

3 reason for using containers in machine learning...

Wise.io

brinkar

1 Packaging

python - numpy version error when importing tensorflow - Stack Overflow
https://stackoverflow.com/questions/.../numpy-version-error-when-importing-tensorfl...
Oct 7, 2016 - From the **error**, it looks like you're running python 2.7 from usr/local/bin . There is a mismatch **problem** between your **numpy version** and tensorflow ...

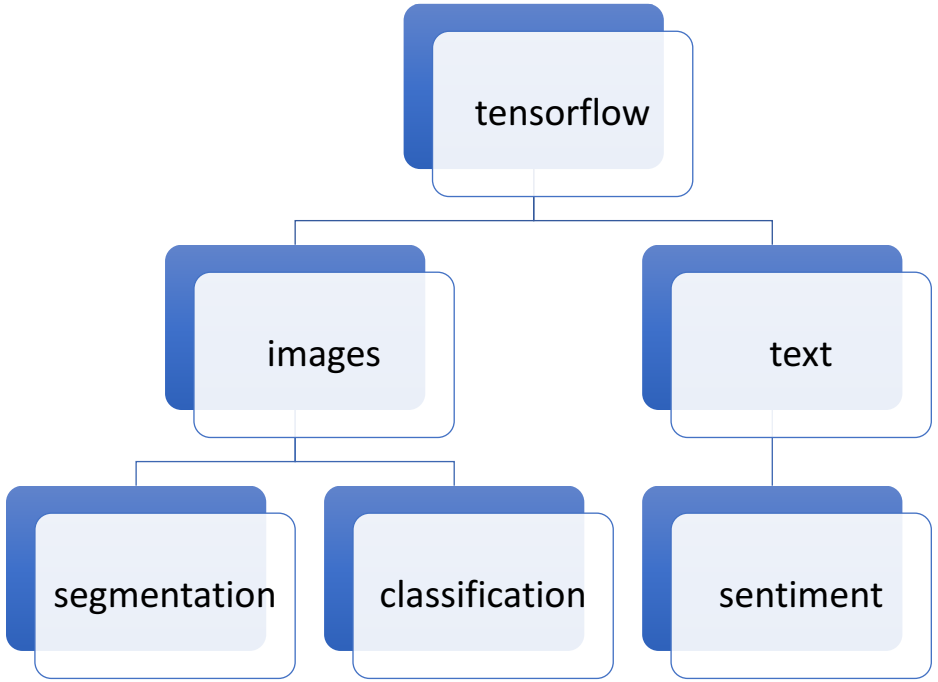
ImportError: numpy.core.multiarray failed to import · Issue #559 ...
https://github.com/tensorflow/tensorflow/issues/559
Dec 19, 2015 - When I ran into this, it was because I had another **version** of **numpy** .... chenghuige referenced this **issue** in tensorflow/serving on May 12, 2016.

Wise.io

brinkar

2 Inheritance

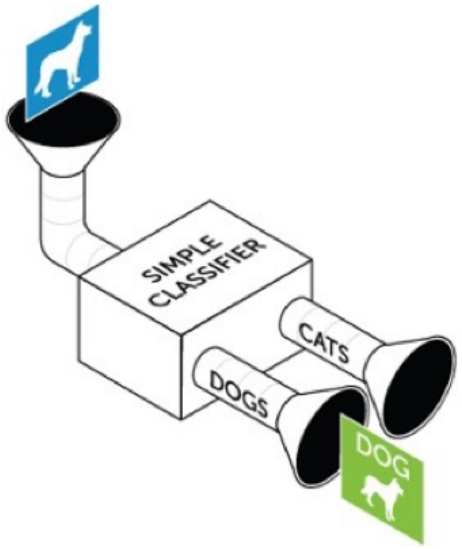
```
FROM tensorflow/tensorflow:latest-gpu

# Your specialized modeling pipeline


FROM my-special-pipeline

# Your even more specialized pipeline
```

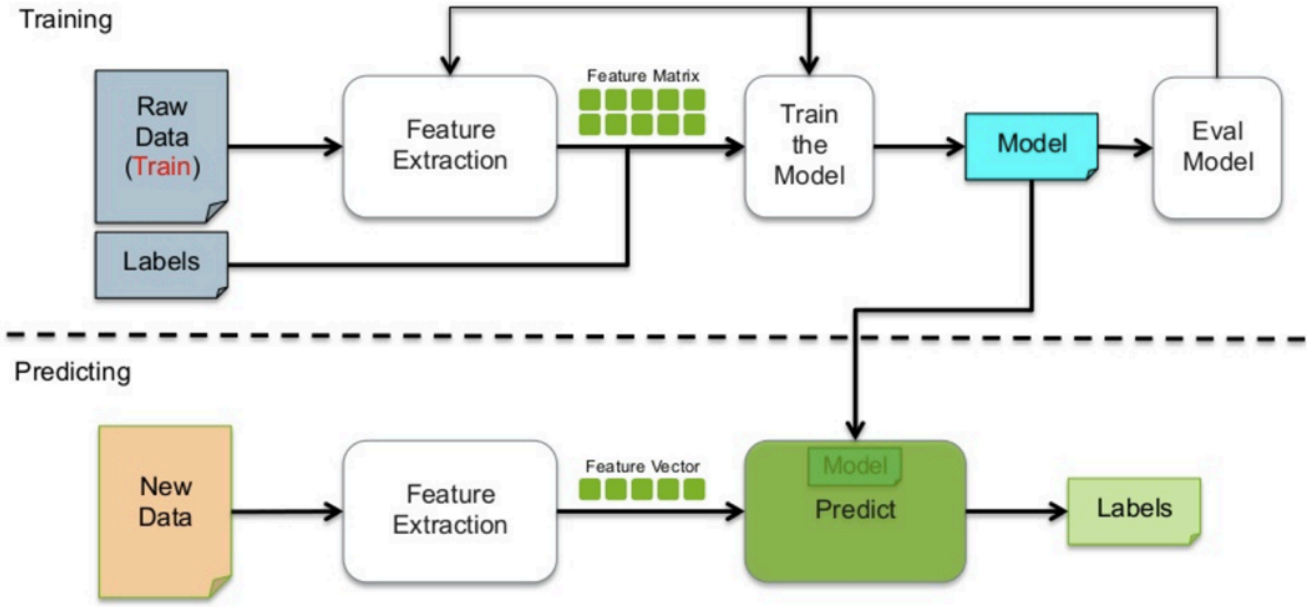


Wise.io

brinkar

Immutability





“Data should go through the exact same pipeline when making predictions, as when the model was built.”

-- Good ML practitioner

Wise.io

brinkar

Use great frameworks and services...

Wise.io

brinkar

Open source ML frameworks

RISELab

Clipper

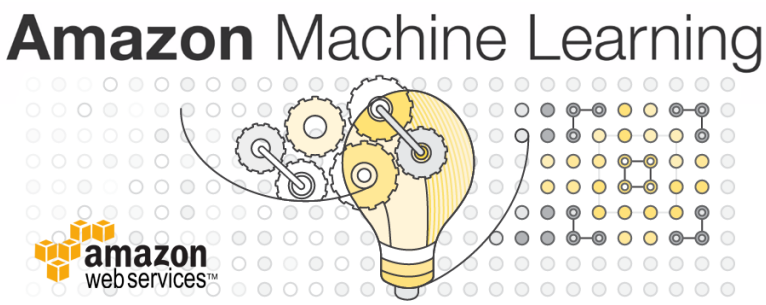
- **Deploy models** trained in your choice of framework to Clipper with a few lines of code by using an existing model container or writing your own
- Easily **update or add models** to running applications
- Use adversarial bandit algorithms to **dynamically select best model** for prediction at serving time
- Set latency service level objectives for **reliable query latencies**
- Run each model in a separate Docker container for simple **cluster management and resource allocation**
- Deploy **models running on CPUs, GPUs**, or both in the same application

Wise.io

brinkar

Hosted ML services

Google Cloud Platform



Wise.io

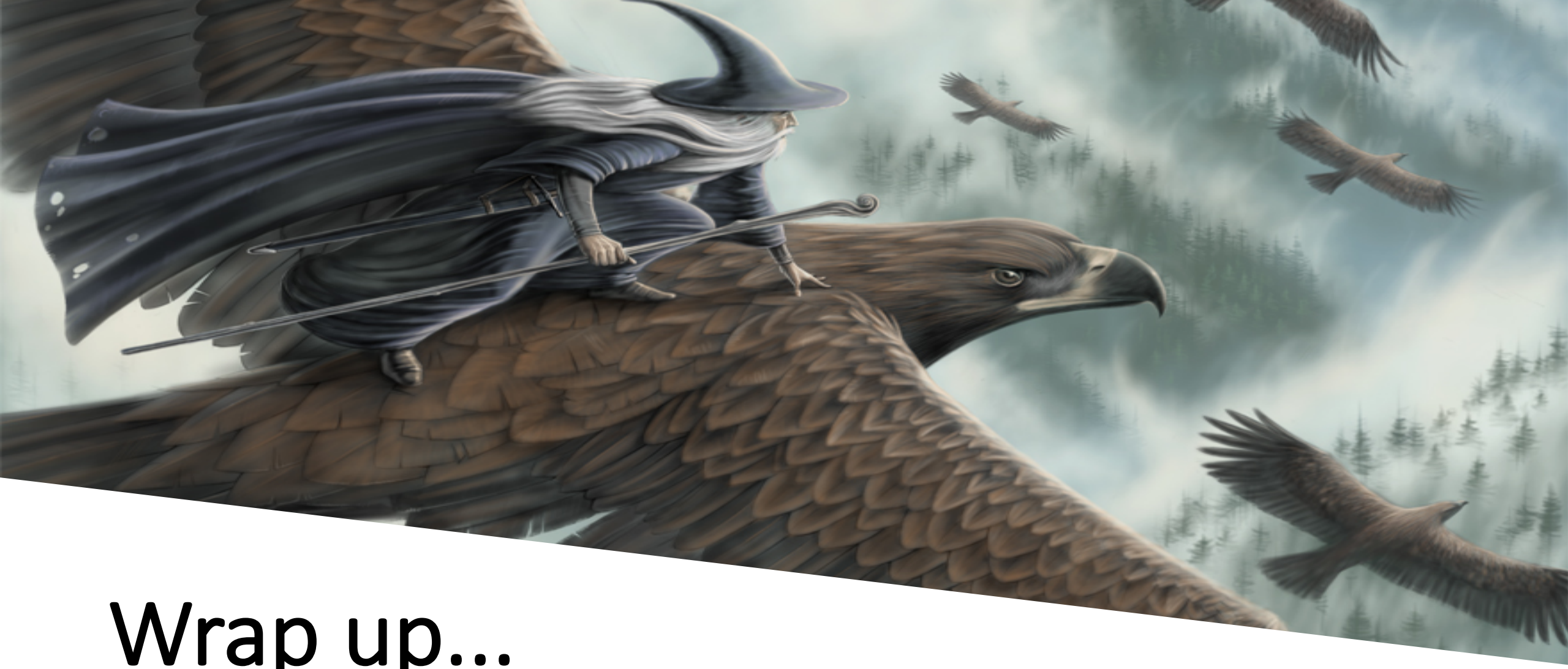
brinkar

Machine learning infrastructure

- Continuous integration
- Containerization
- Utilize ML platforms

Wise.io

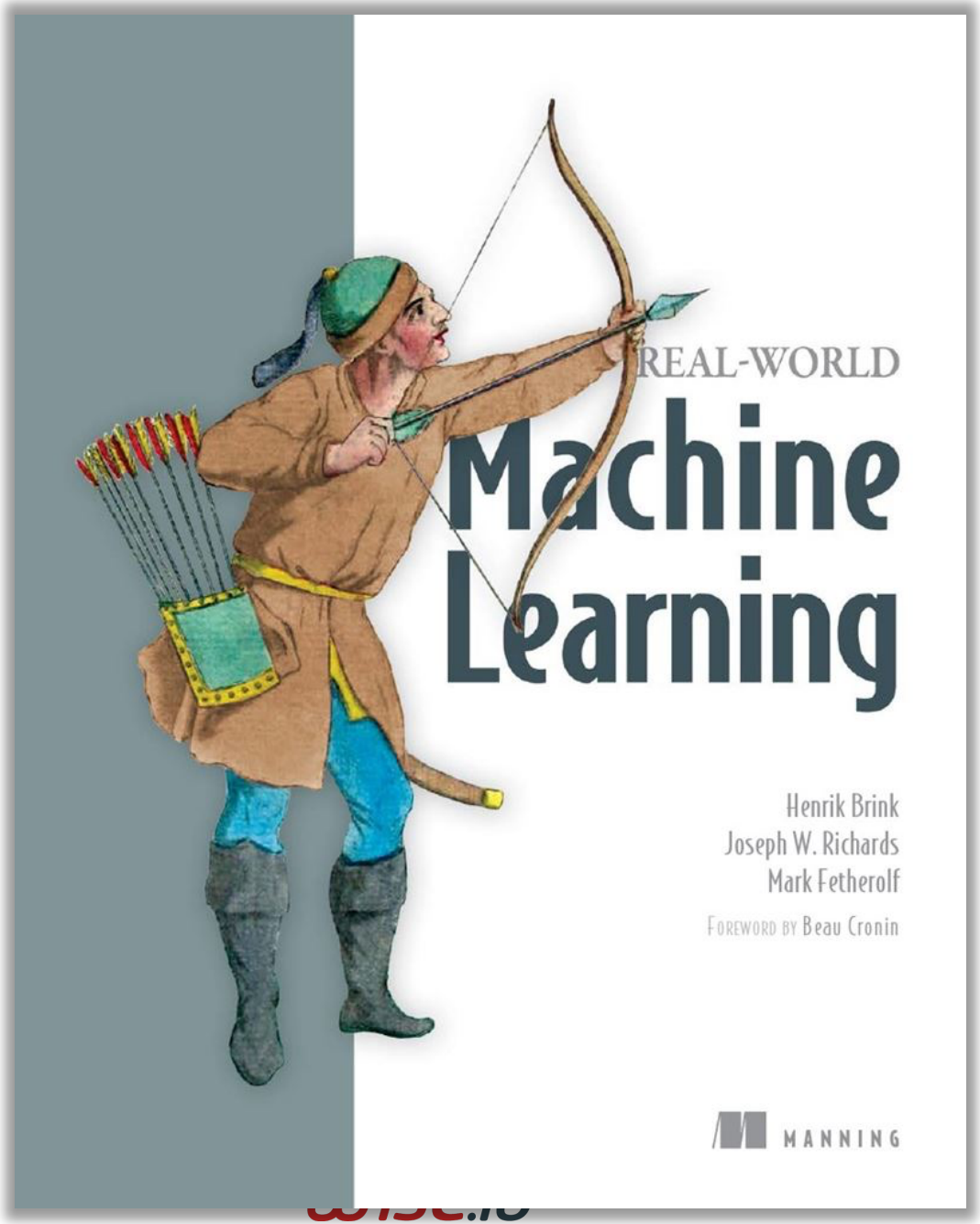
brinkar



Wrap up...

Real-World Machine Learning

manning.com/brink



brinkar

nordic.ai

meetup.com/datacph

brinkar

Questions?

Wise.io

brinkar