Next Generation AI: Towards Widespread Enterprise Adoption

dean@deanwampler.com
@deanwampler
Domino Data Lab

Last update: March 25, 2021
Download this and other talks from polyglotprogramming.com/talks.

DOMINO

dominodatalab.com

DOMINO

Products Solutions Customers Resources Partners Company Get a Demo Try Now

Let your data science team use the tools they love.

And bring them together in an enterprise-strength platform, that enables them to spend more time solving critical business problems.

Learn More

BAYER LOCKHEED MARTIN Allstate. Bristol Myers Squibb DELL

Read Our Customer Stories »

dominodatalab.com

System-of-Record for Enterprise Data Science Teams

Accelerate Research

Get self-serve access to the latest tools and scalable compute. Reuse past work and iterate more efficiently.

Learn More »

Centralize Infrastructure

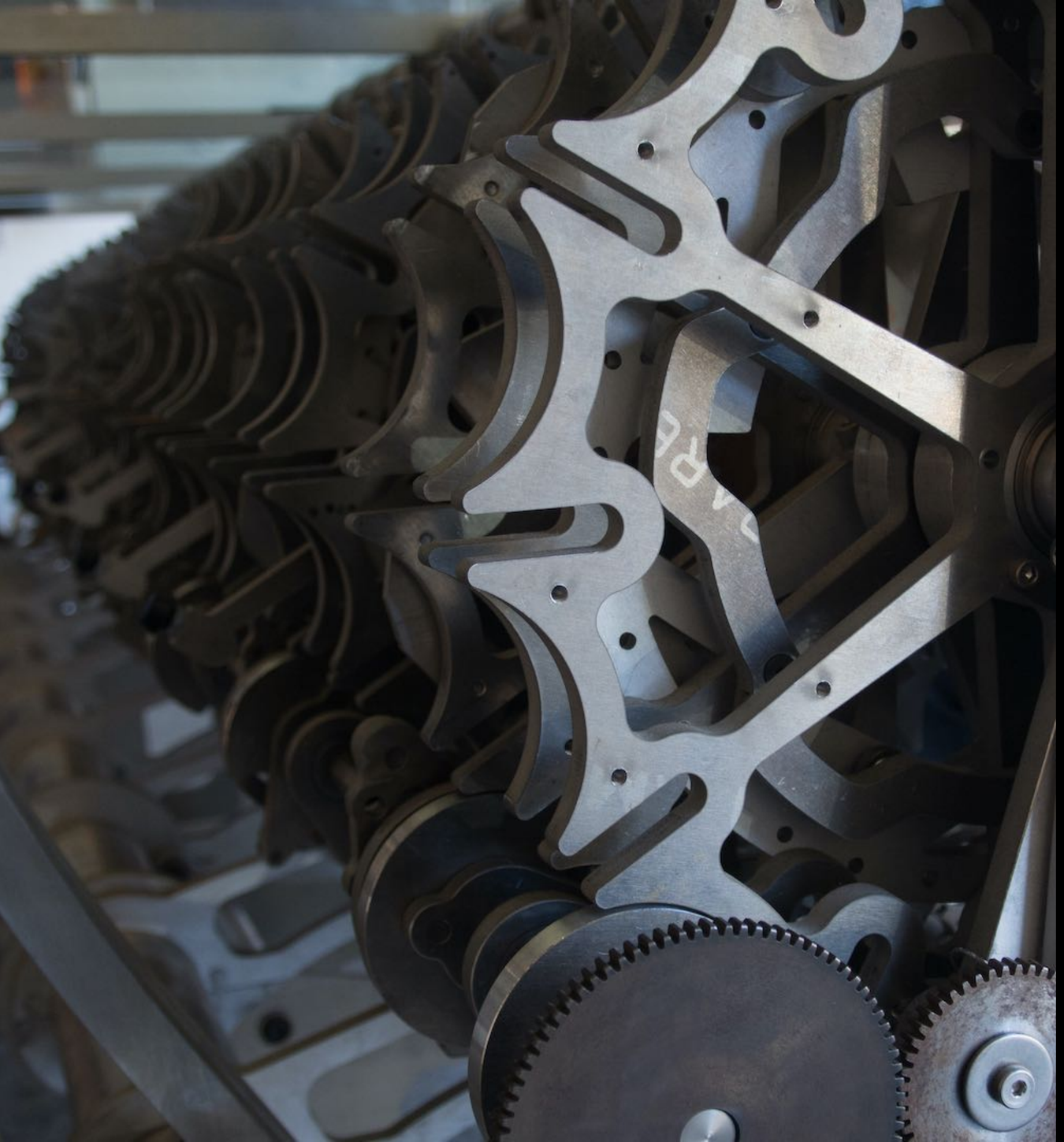
Manage the availability of powerful data science resources in a secure and governed system-of-record.

Learn More »

Deploy and Monitor Models

Expedite model consumption with apps, APIs, and more – and ensure their

DOMINO



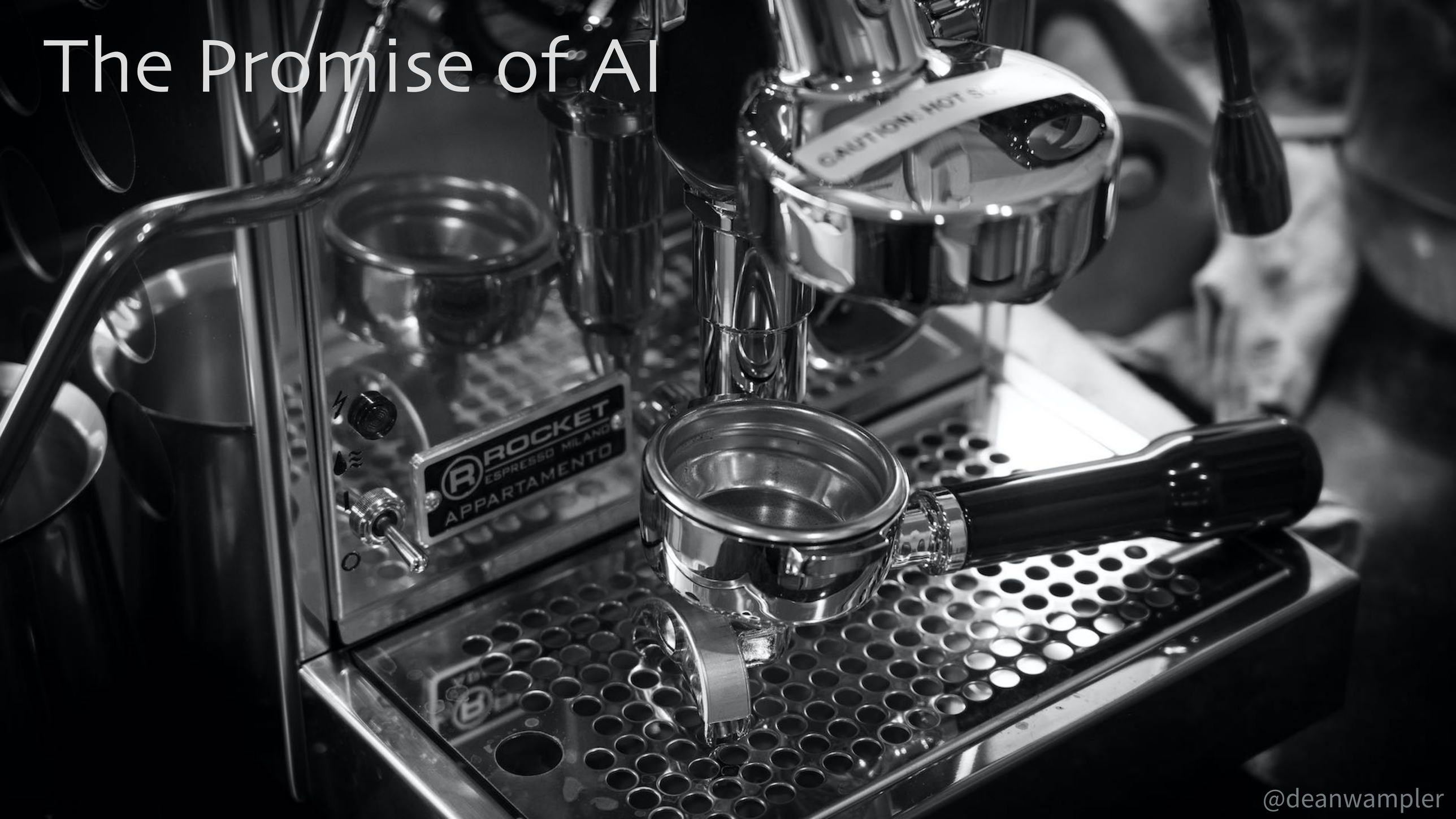
Outline

- The Promise of AI
- AI in the Enterprise
  - The Past
  - The Present
  - The Future
- Conclusions

Outline

- The Promise of AI
- AI in the Enterprise
  - The Past
  - The Present
  - The Future
- Conclusions

The Promise of AI



The Promise of AI

- Natural Language Processing
- Reinforcement Learning
- New applications of Deep Learning
- What Our Phones Are Telling Us…

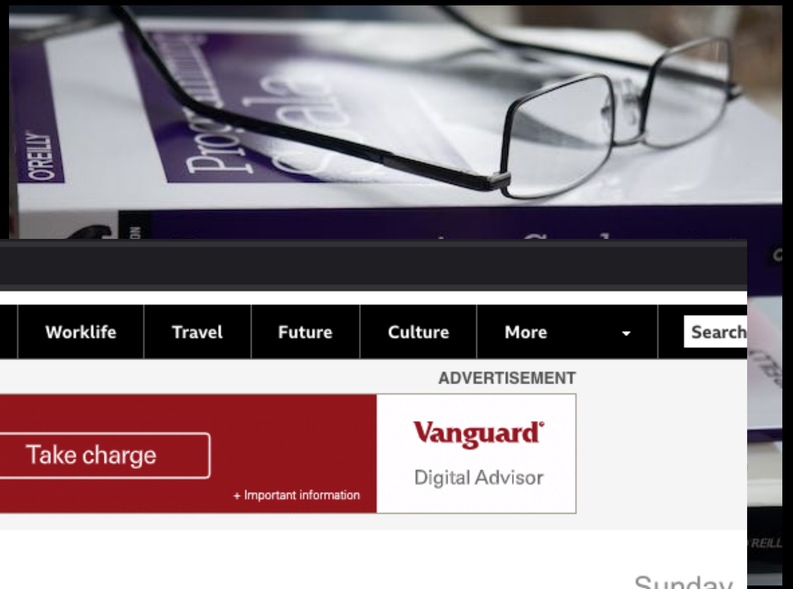
Natural Language Processing

Applications

- Summarization
- Dialogues
- Naturalistic text to speech
- Translation
- Sentiment Analysis
- Fraud & Veracity Analysis
- Question Answering & Search

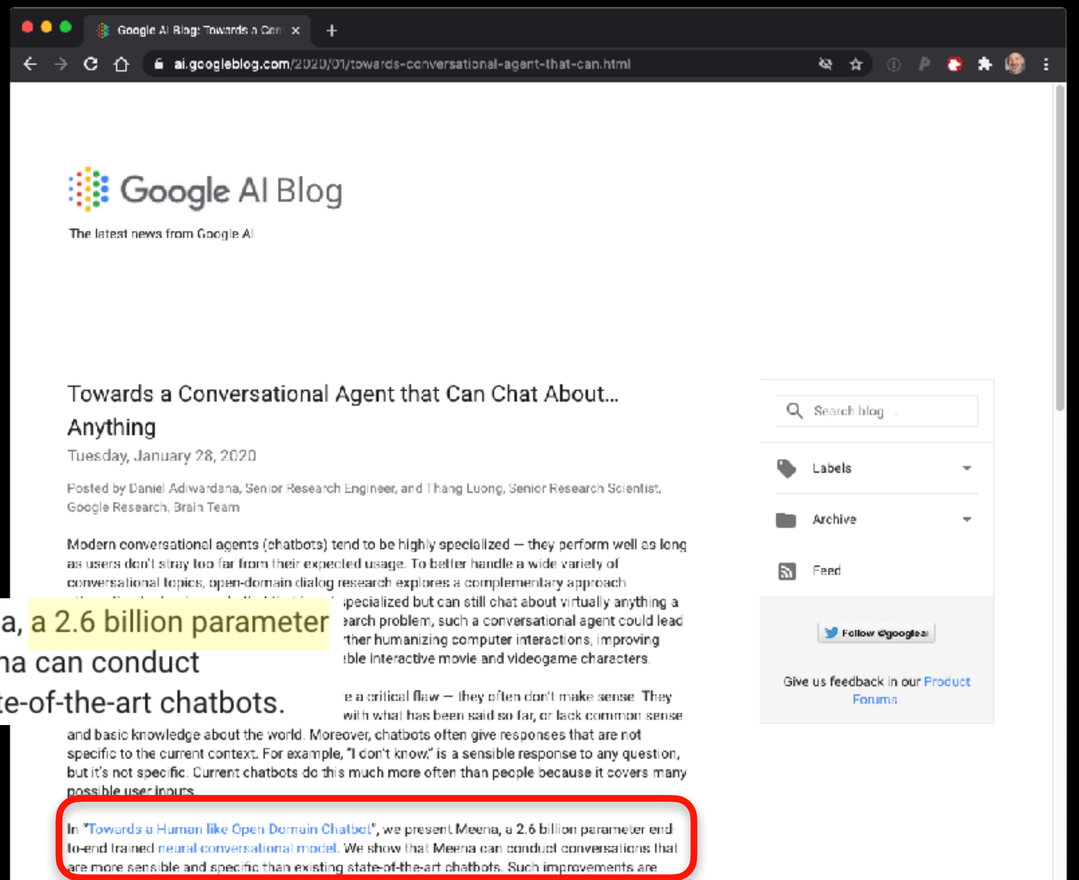
Summarization

- Legal documents
- Research papers
- News
- …



Dialogs

- Chatbots
- Human-computer dialogs



In "Towards a Human-like Open-Domain Chatbot", we present Meena, a 2.6 billion parameter end-to-end trained neural conversational model. We show that Meena can conduct conversations that are more sensible and specific than existing state-of-the-art chatbots.

Naturalistic text to speech

- Needed for dialog generation

Translation

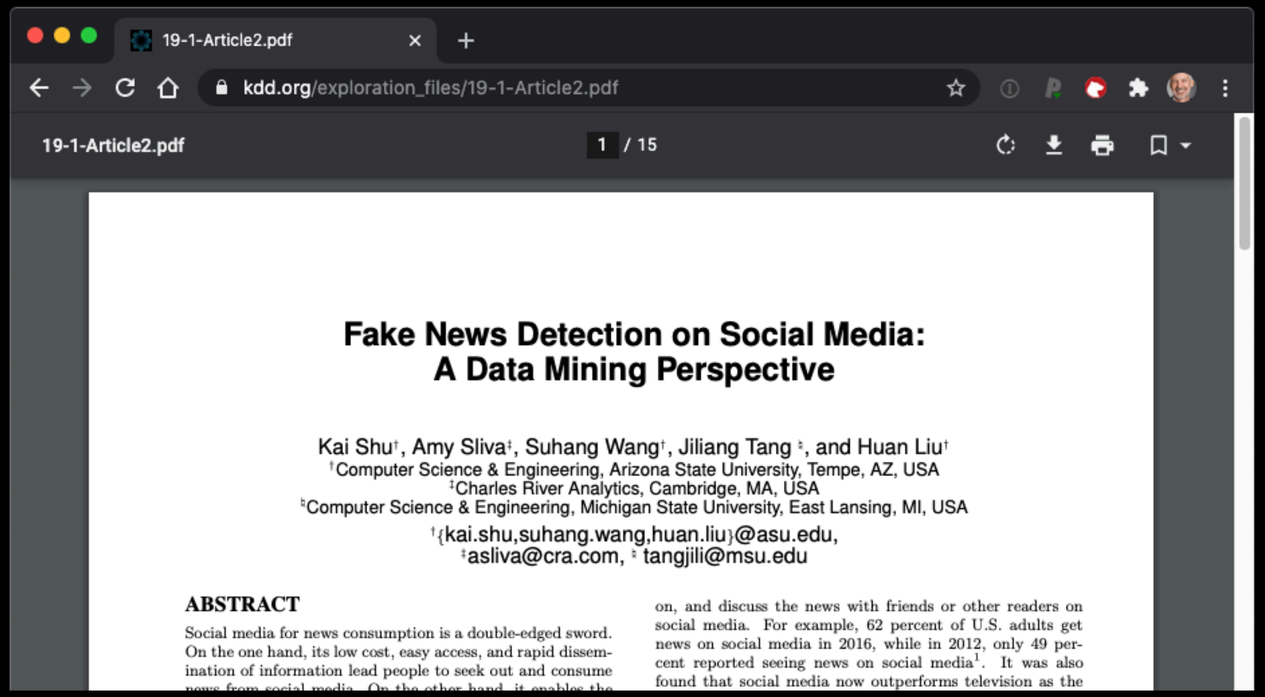
- Domain-specific languages
  - Medicine
  - Air traffic control
  - …
- “Rare” languages

Sentiment Analysis

- Customer support
- Social media
- Public relations

Fraud & Veracity Analysis

- "Fake news"
- Better SPAM, Phishing, etc. detection and mitigation.



Question Answering & Search

- Customer support
- More advanced, targeted search results
  - Support natural language queries
  - Search legal docs, research papers, patents, …

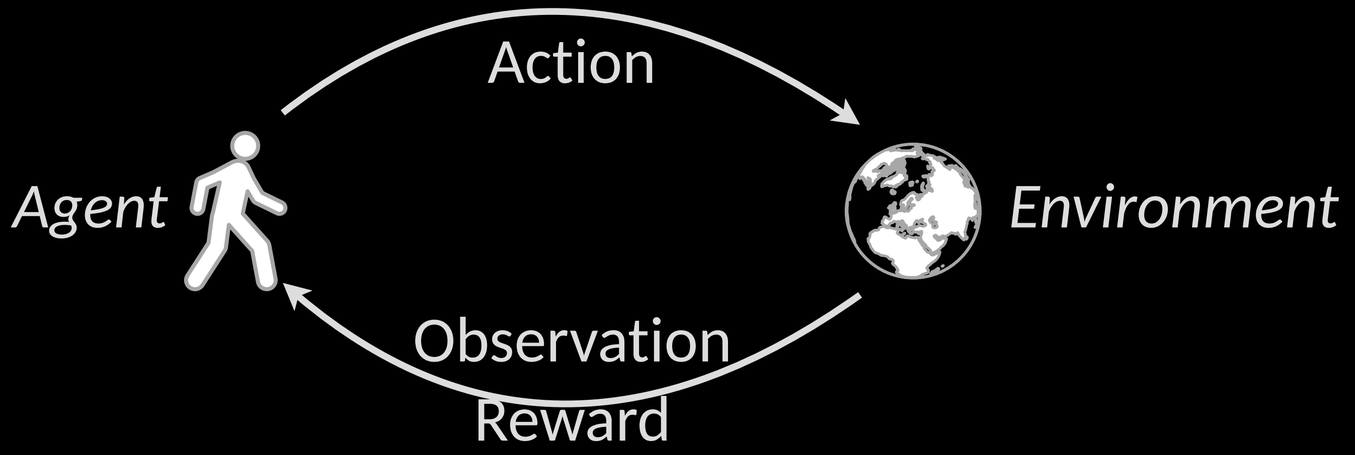
Images and Videos…

- Many of these same techniques and applications apply to image and video applications, too.

Reinforcement Learning

What Is RL?

- An agent observes an environment, takes a sequence of actions
- Goal: maximize the cumulative reward

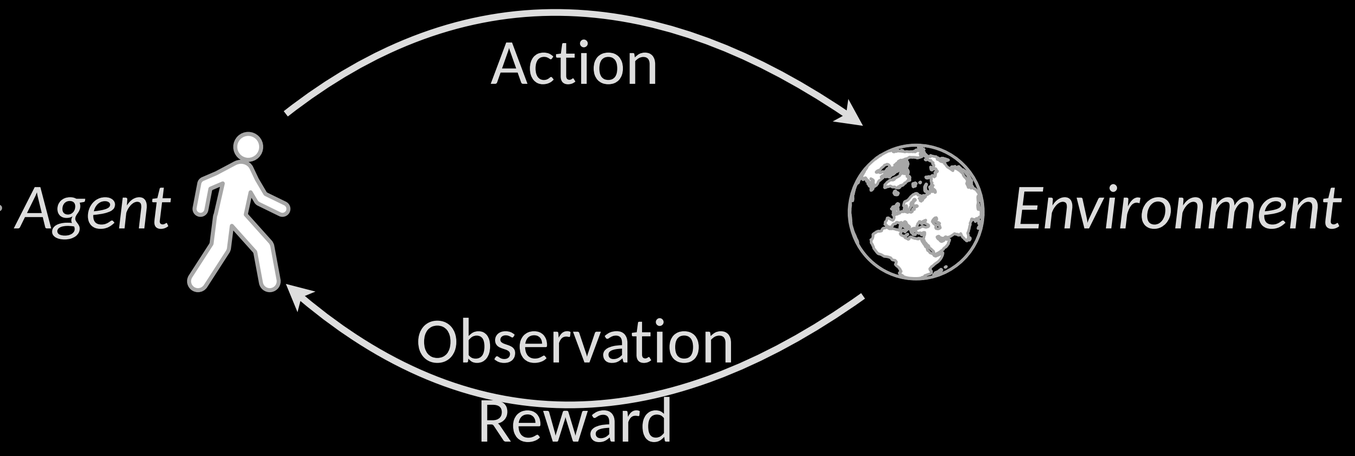


What Is RL?

- An agent observes an environment, takes a sequence of actions
- Goal: maximize the cumulative reward

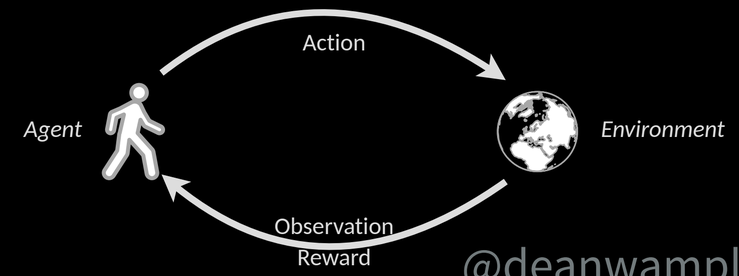
Episodes. RL is ideal for problems with sequential events.

Learn an optimal policy. Cumulative reward reinforces learning.



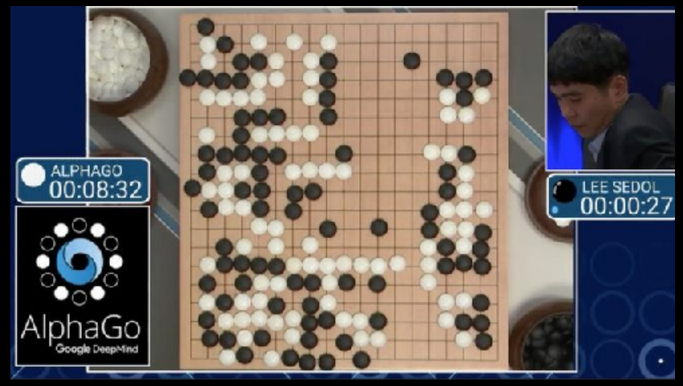
Applications

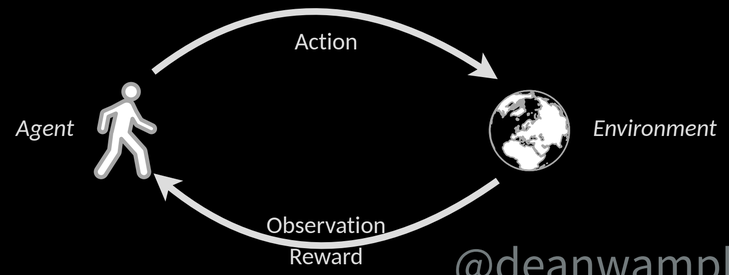
- Games
- Robots & Autonomous Vehicles
- Process Modeling & Automation
- System Optimization
- Advertising & Recommendation
- Markets



Games

- World's best expert game play in:

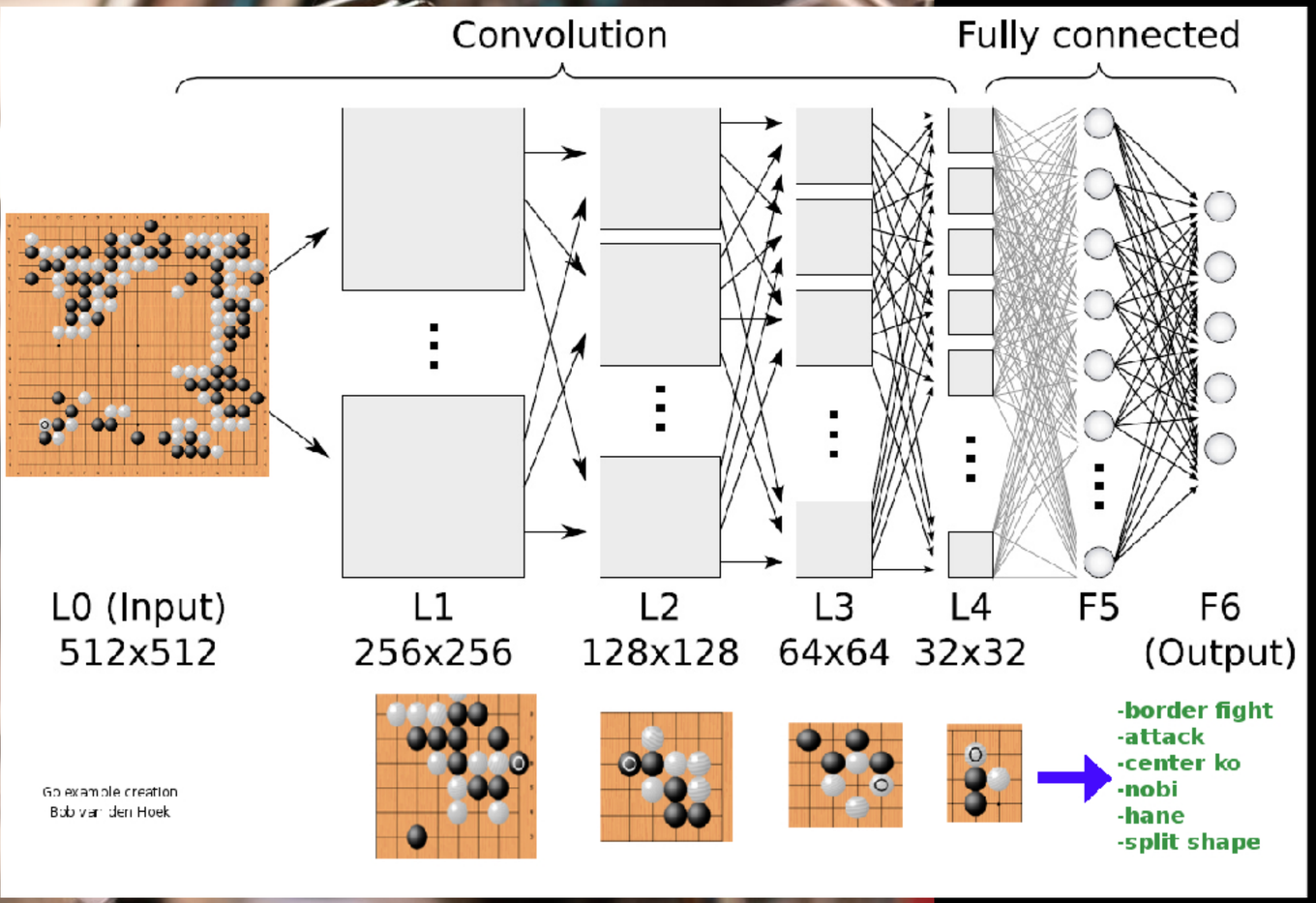


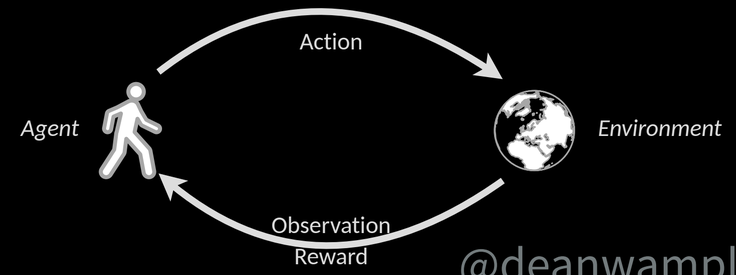


https://www.geekwire.com/2016/alphago-ai-program-wins-1-million-prize-go-showdown-champion-lee-sedol/
https://towardsdatascience.com/tutorial-double-deep-q-learning-with-dueling-network-architectures-4c1b3fb7f756

Games

- AlphaGo
  - Observations: board state
  - Actions: place stones
  - Rewards:
    - 1 if you win
    - 0 otherwise

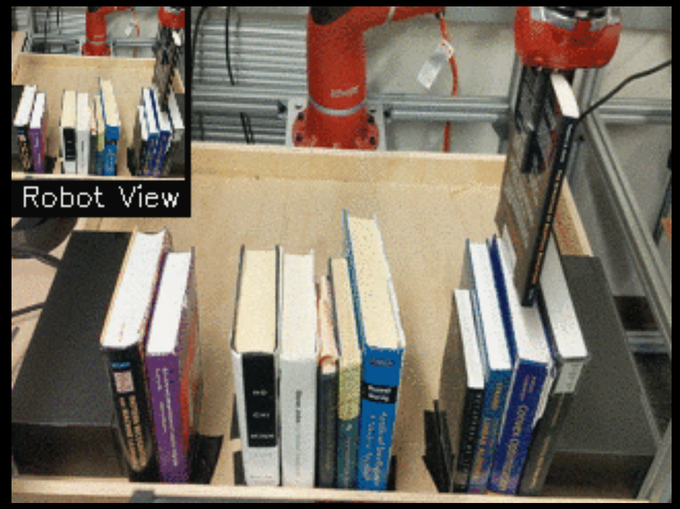


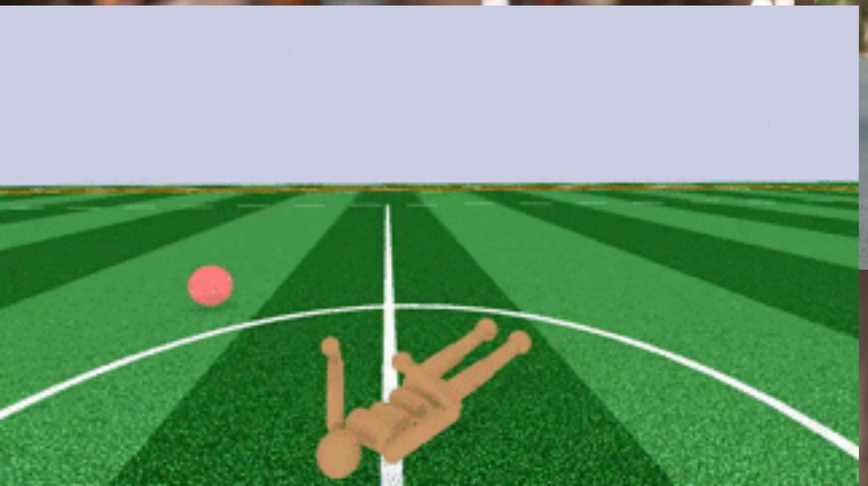


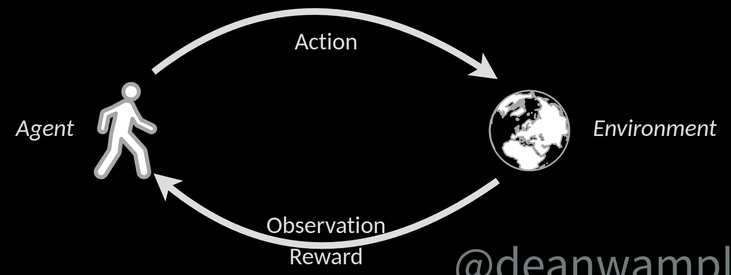
nature.com/articles/nature16961

Robotics & Autonomous Vehicles

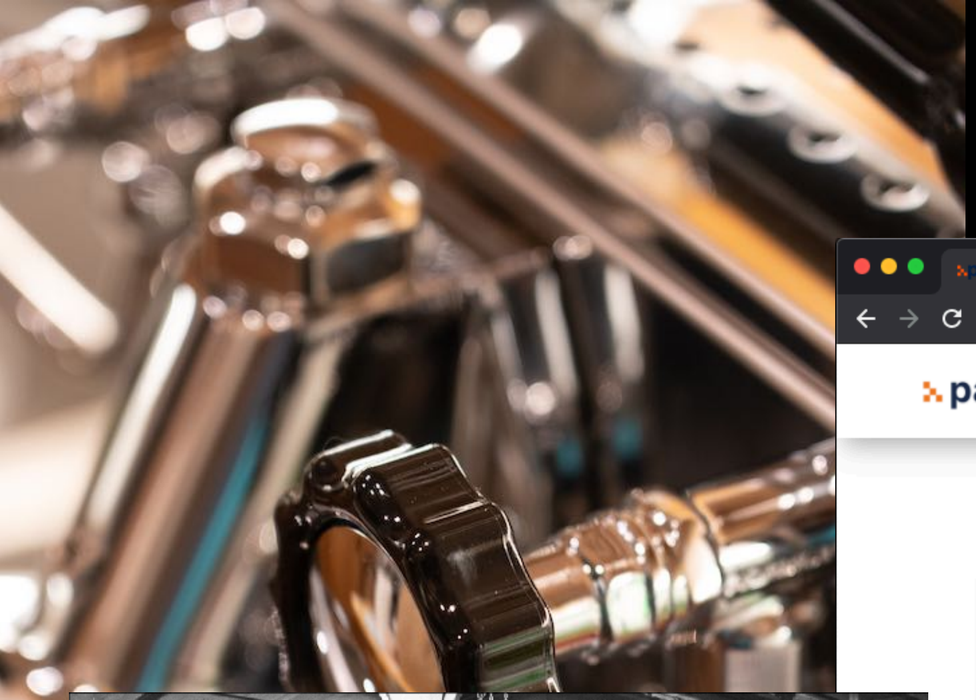
- Start with simulators, work up to real machines.

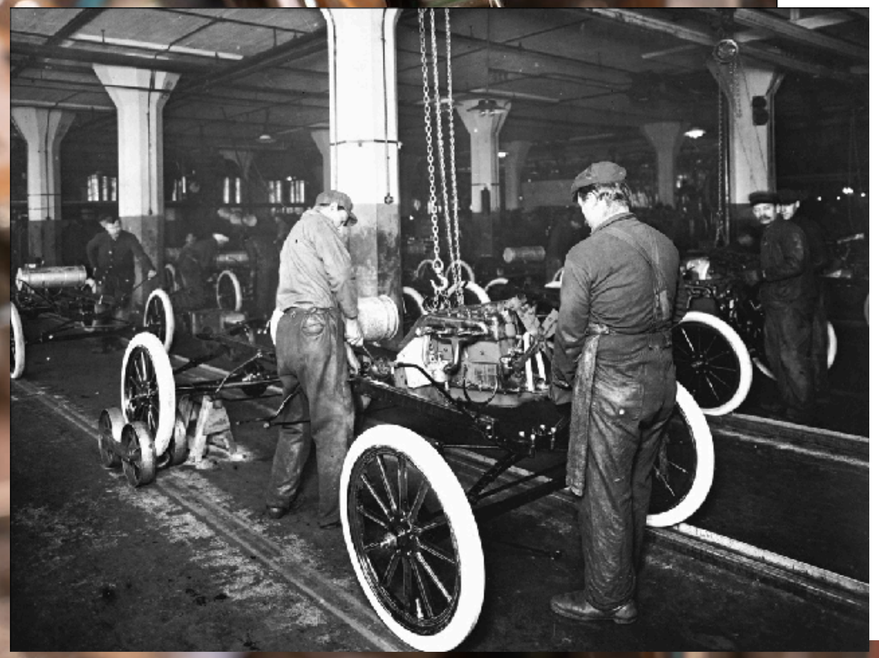






Process Modeling & Automation





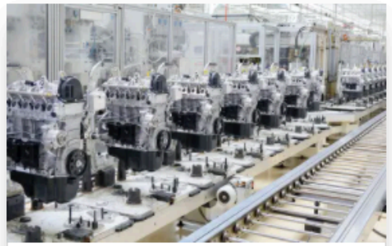
Simulation Optimization | Add

pathmind.com

pathmind

Products | Services | Industries | Resources | Company | SIGN IN | REQUEST DEMO

Recent Updates



Engineering Group: Manufacturing Optimization with AI Minimizes Factory Flow Bottlenecks

Oct 30, 2020 | Customer Success

Summary Engineering Group, a global engineering firm and technology consultancy with a strong practice in simulation, worked with Pathmind to apply reinforcement learning to intelligently route heavy industrial parts over a complex assembly line in...

Engineering Group: Using AI to Maximize Factory Output with Better Order Sequencing

Oct 29, 2020 | Customer Success

Summary Engineering Group, a global engineering firm and technology consultancy with a strong practice in simulation, worked with Pathmind to apply reinforcement learning to maximize factory output by making smarter decisions about order...

Princeton Consultants: Using AI to Maximize Efficiency of Machine Scheduling

Oct 13, 2020 | Customer Success

Summary Princeton Consultants, a simulation consulting firm, serves a manufacturing client with a hard machine scheduling problem. Its optimizer had difficulty scheduling machines for new types of items that needed to be processed; it was not able...

ment

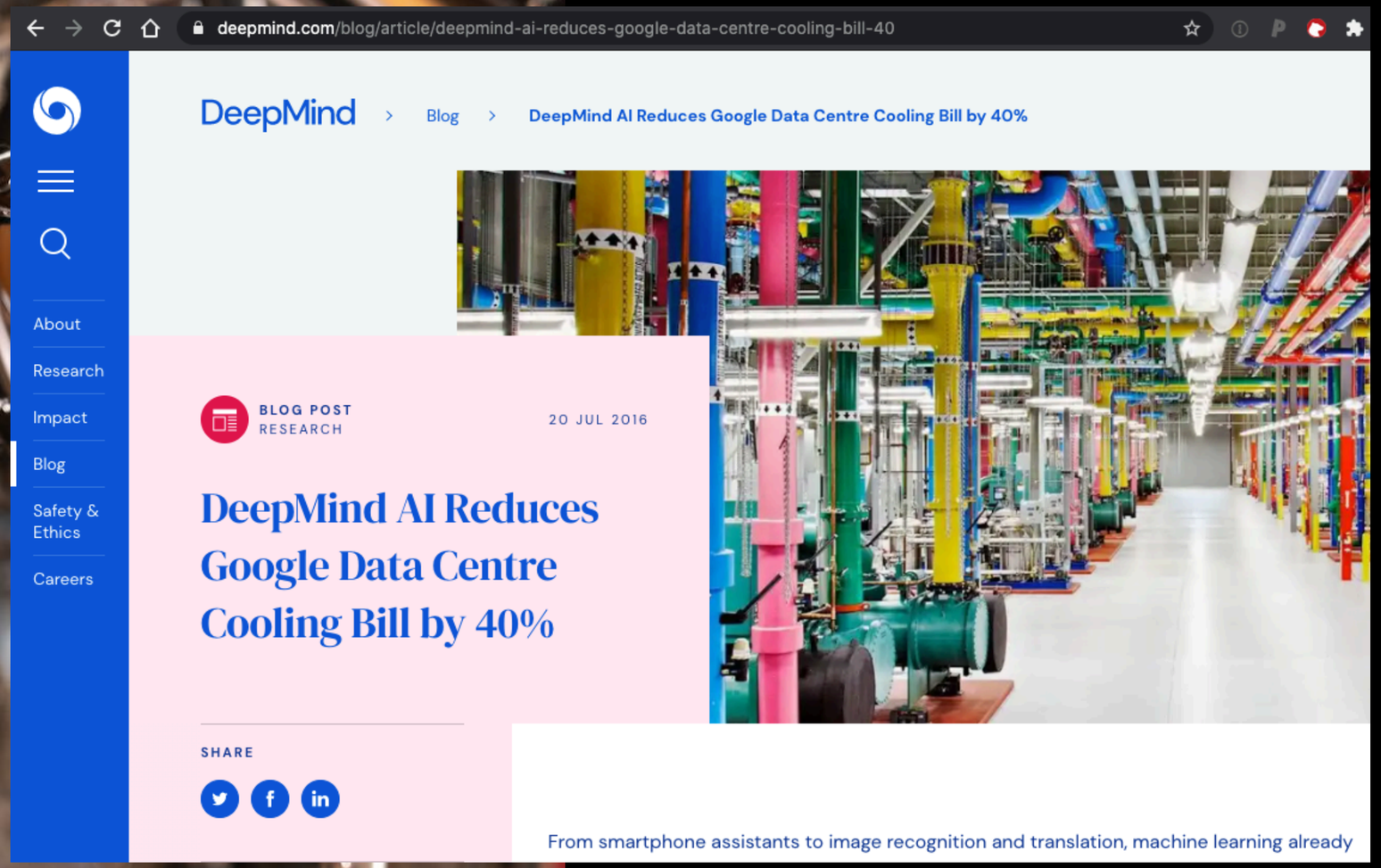
Observation

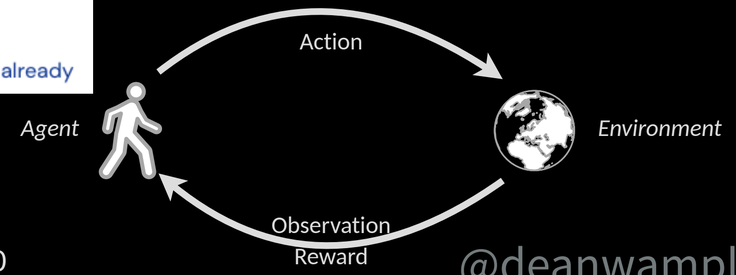
Reward

https://media.ford.com/content/fordmedia/fna/us/en/features/game-changer--100th-anniversary-of-the-moving-assembly-line.html

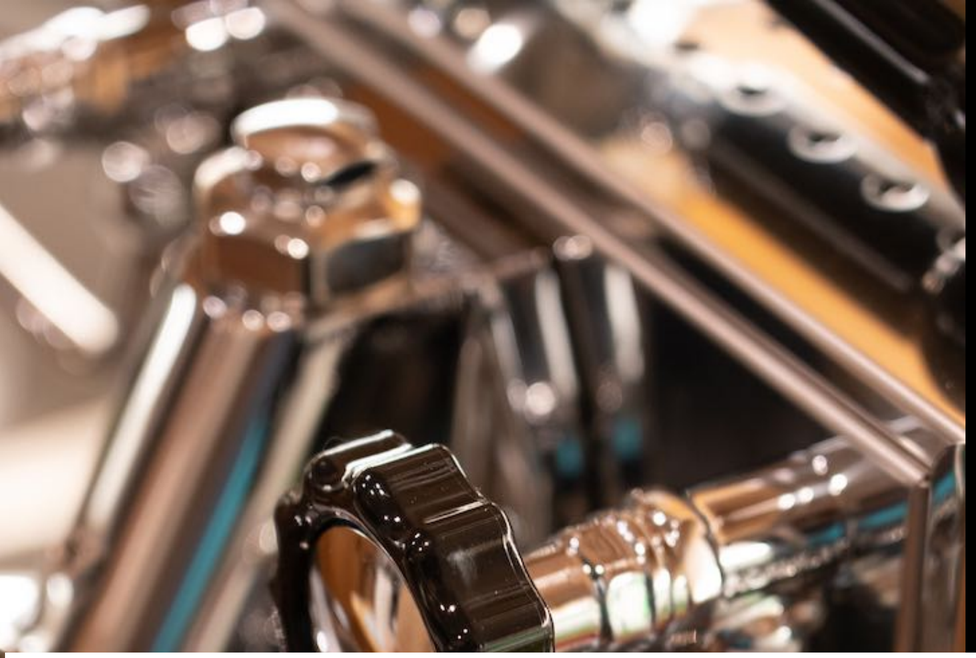
@deanwampler

System Optimization



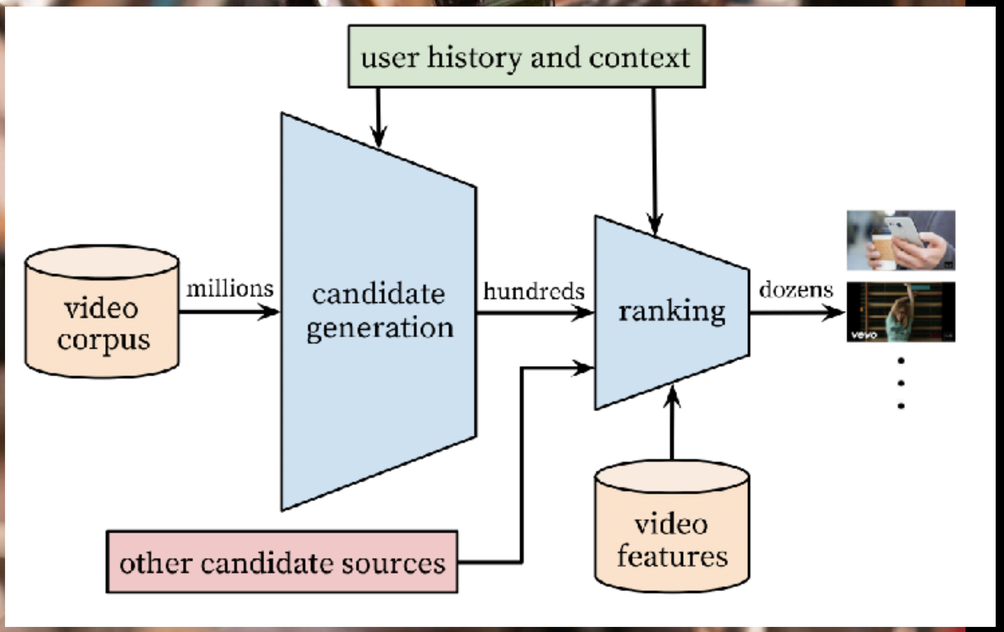


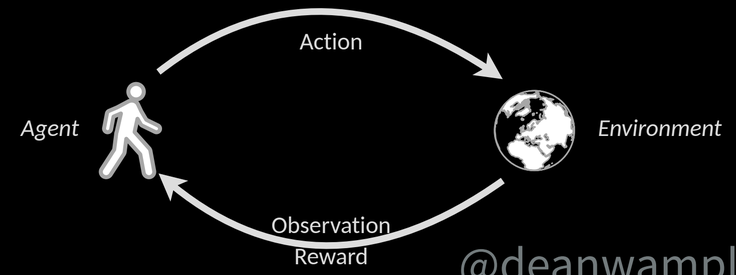
https://deepmind.com/blog/article/deepmind-ai-reduces-google-data-centre-cooling-bill-40



Advertising & Recommendation

- A "mature" problem, yet RL is providing a new approach.
  - Better modeling of evolving preferences.
  - Better scalability than collaborative filtering, etc.

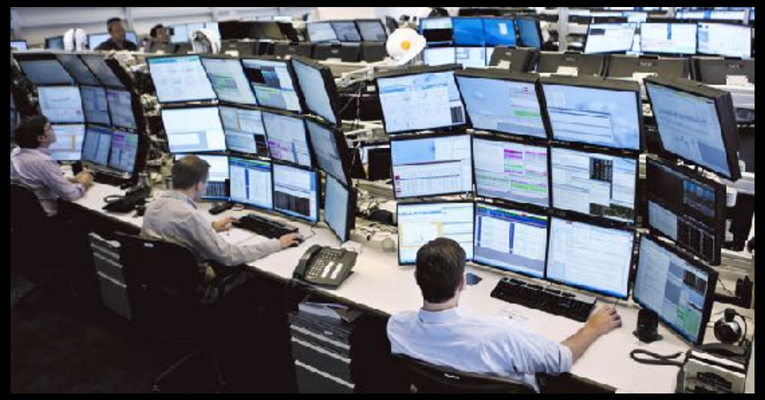


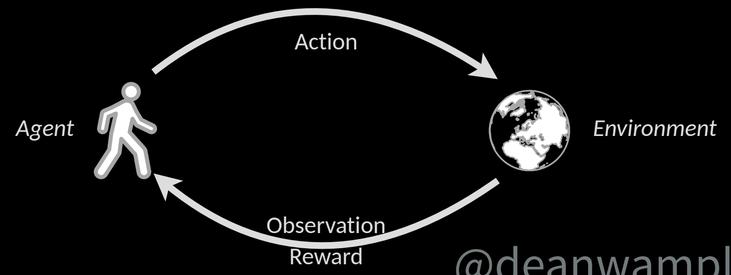


https://research.google/pubs/pub45530/

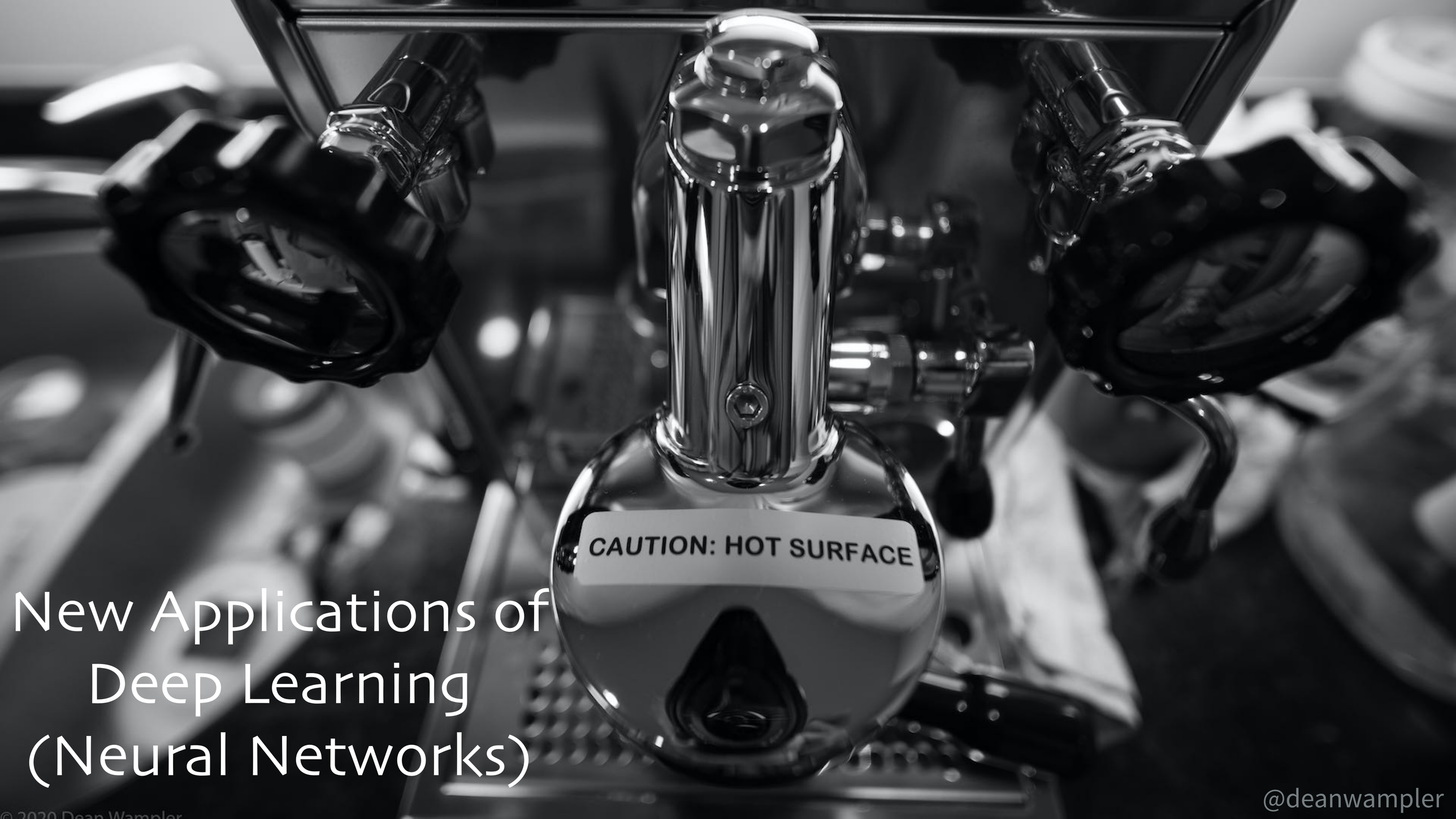
Markets

- Inherently time-ordered
- Lots of different "signals"
- Contextual, multi-armed bandits





http://tradinghub.co/watch-list-for-mar-26th-2015/



New Applications of Deep Learning (Neural Networks)

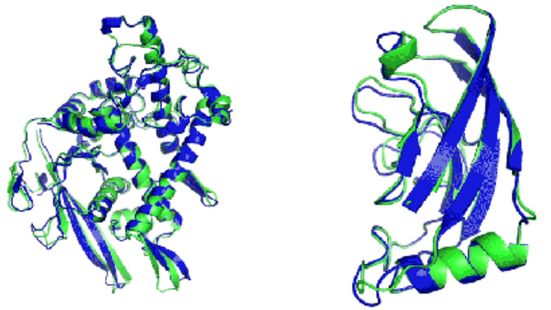
Biology, Medicine

About
Research
Impact
Blog

BLOG POST
RESEARCH

30 NOV 2020

AlphaFold: a solution to a 50-year-old grand challenge in biology



T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)

T1049 / 6y4f
93.3 GDT
(adhesin tip)

Experimental result
Computational prediction

deepmind.com

Biology, Medicine

nature reviews cancer

Perspective | Published: 17 May 2018

OPINION

Artificial intelligence in radiology

Ahmed Hosny, Chintan Parmar, John Quackenbush, Lawrence H. Schwartz & Hugo J. W. L. Aerts ✉

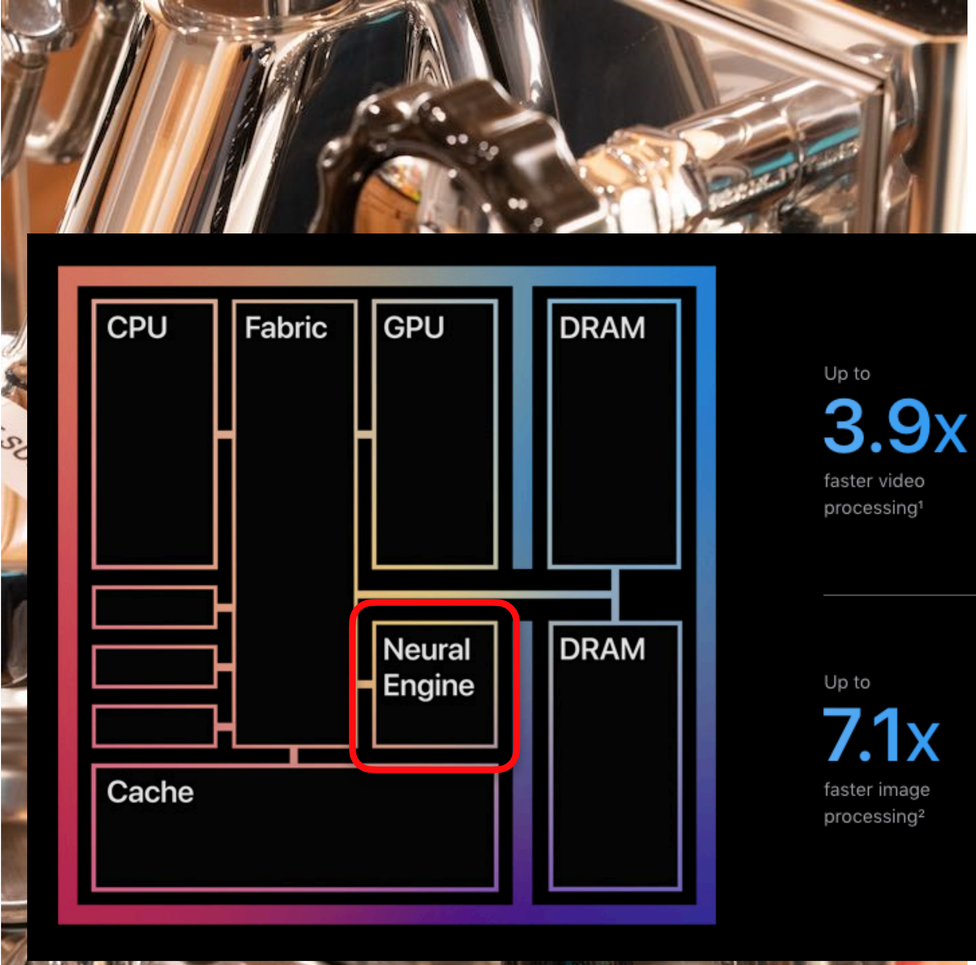
Nature Reviews Cancer **18**, 500–510(2018) | Cite this article

Abstract

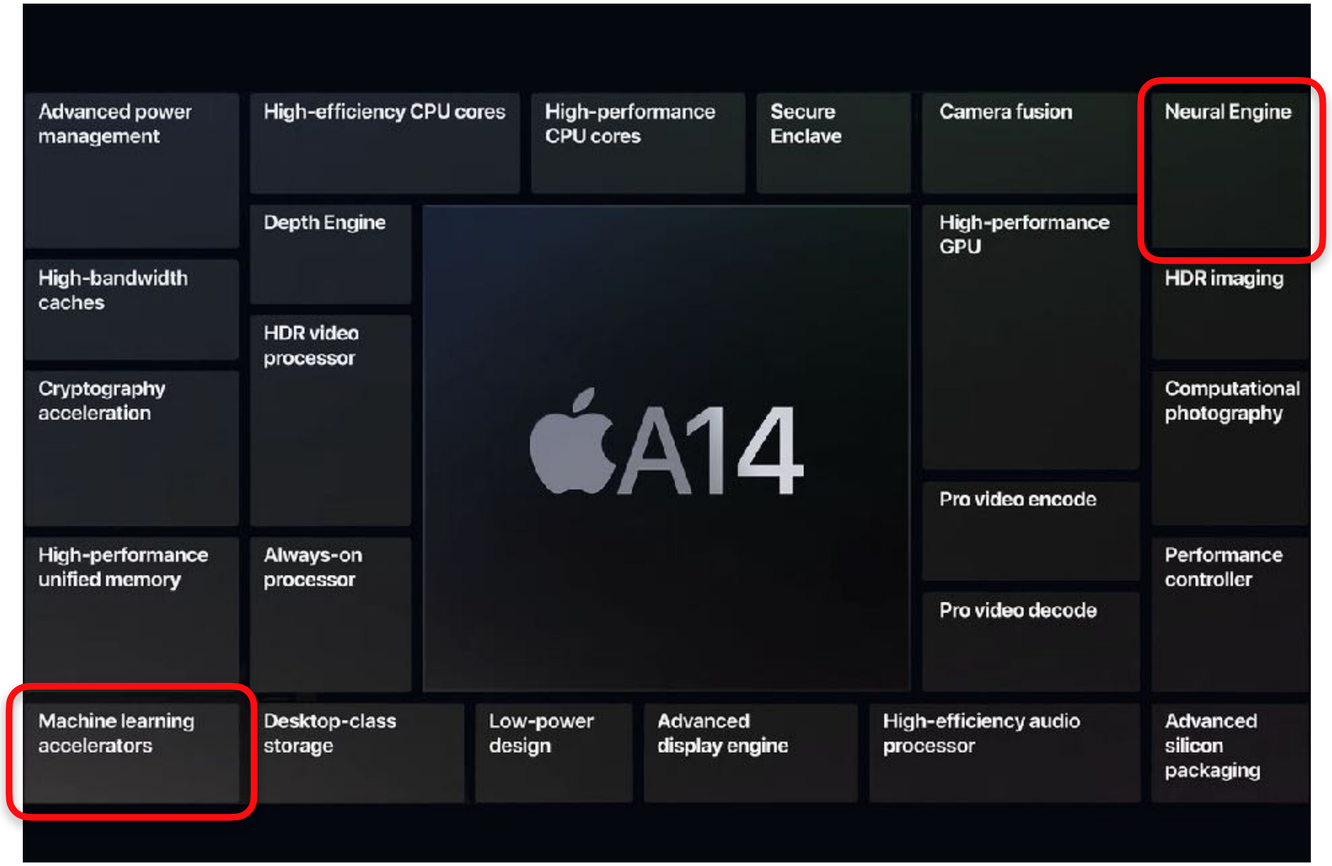
Artificial intelligence (AI) algorithms, particularly deep learning, have demonstrated remarkable progress in image-recognition tasks. Methods ranging from convolutional neural networks to variational autoencoders have found myriad applications in the medical image analysis field, propelling it forward at a rapid pace. Historically, in radiology practice, trained physicians visually assessed medical images for the detection


What Our Phones
Are Telling Us...
@deanwampler

Apple Silicon



M1 diagram



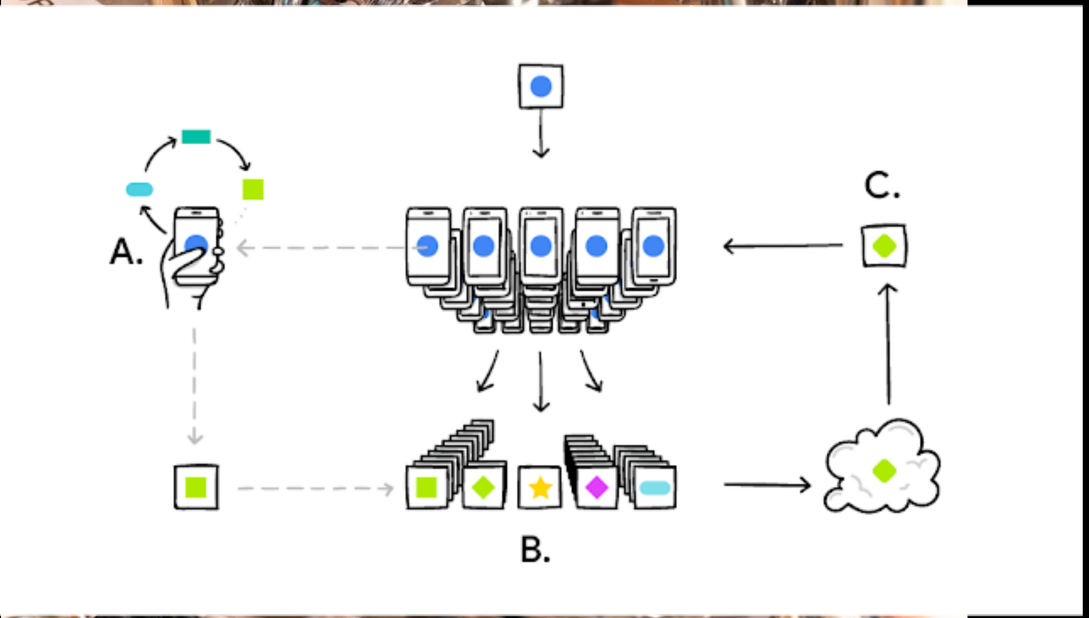
apple.com

Applications that Exploit ML/AI

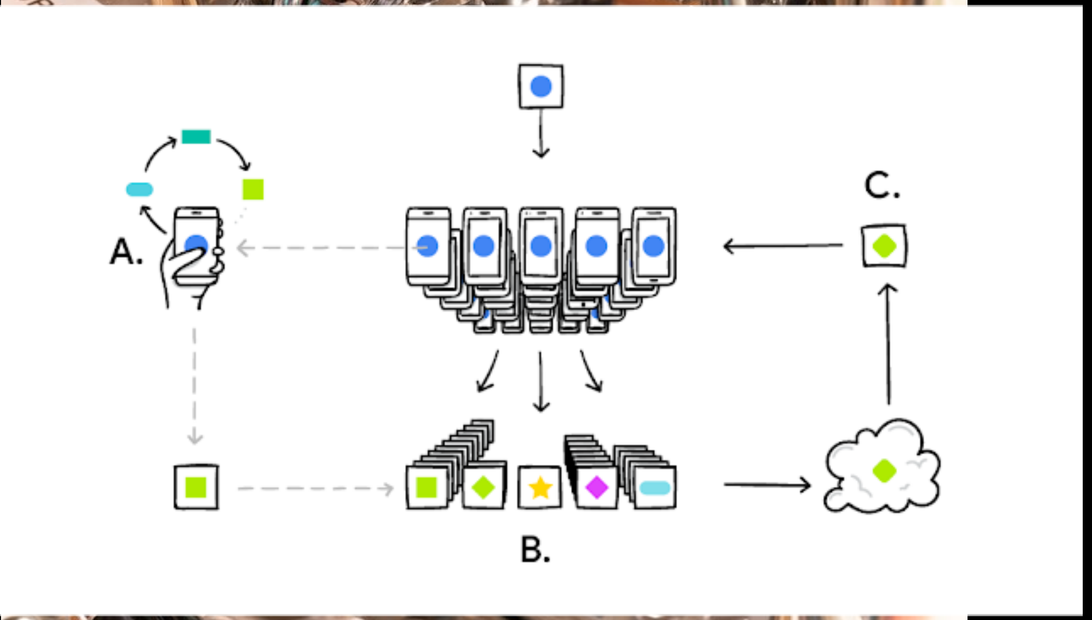
- Unlocking: finger and face ID
- Predictive typing
- Voice assist - Siri
- Health monitoring
- Security (beyond passwords…)
- Recommendations
- …
- Probably most apps will use it in one way or another, eventually!

Technologies that Make this Possible

- Federated Learning
  A. Your local usage trains a model
  B. Model updates from many users are aggregated to form a consensus update
  C. Updated model propagated to all users.
  D. Repeat…



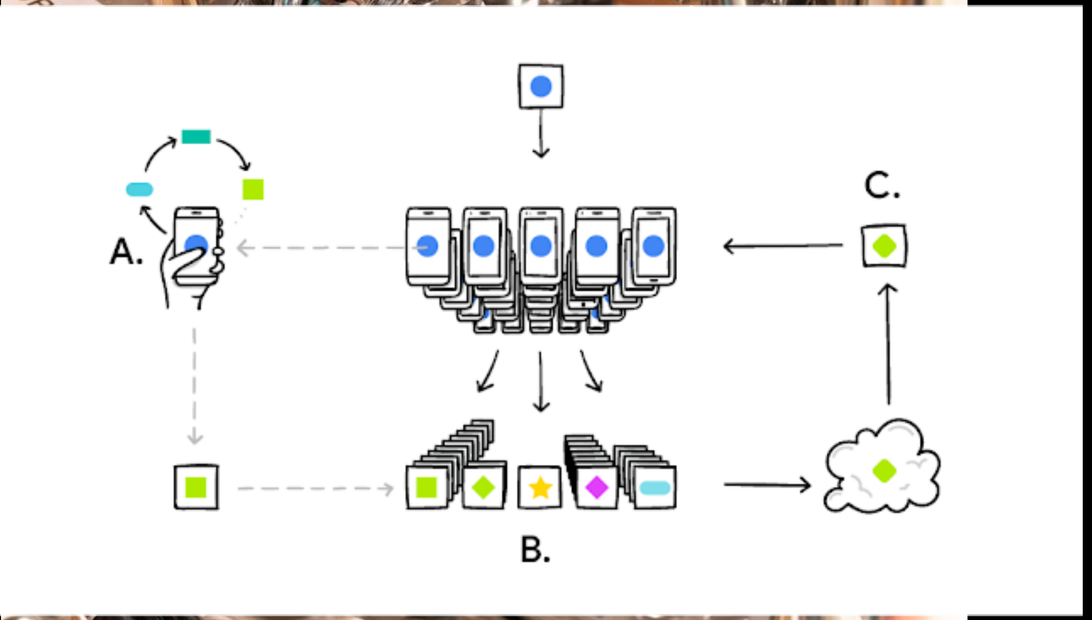
https://ai.googleblog.com/2017/04/federated-learning-collaborative.html



Technologies that Make this Possible

- Federated Learning Advantages
  - Your private data stays local
  - Local model is fine tuned for you
  - Less central data storage required
  - Central processing is minimized
    - Instead, our phones do most of the training

https://ai.googleblog.com/2017/04/federated-learning-collaborative.html



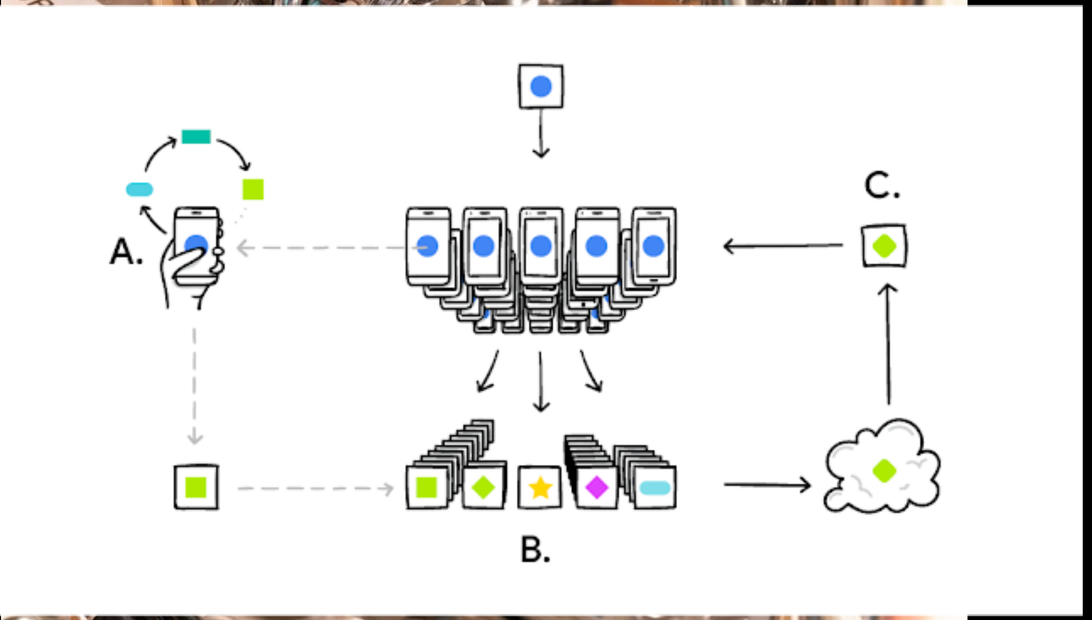
Technologies that Make this Possible

- Differential Privacy
  - "Differential" - If I run a query without your record, then with your record, what can I learn about you from the difference??
  - Introduce "noise" into the data so that:
    - Private data is obscured
    - Introduced error is bounded

https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

Enterprise Applications?

- What services would your customers reject now, but accept if you offered the services using Federated Learning & Differential Privacy??

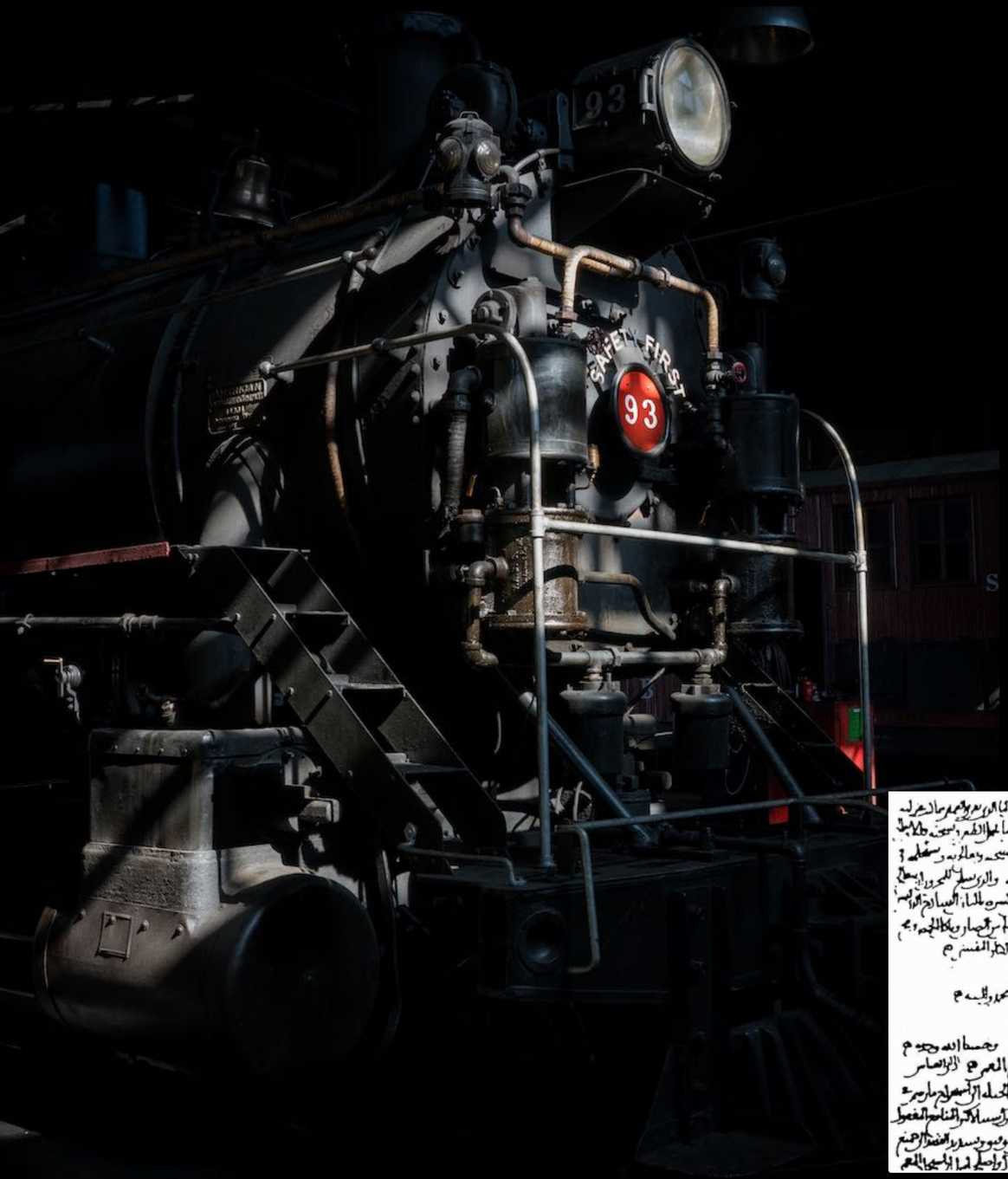


https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

Classic techniques that still deliver lots of value

Outline

- The Promise of AI
- AI in the Enterprise
  - The Past
  - The Present
  - The Future
- Conclusions



Statistical Inference

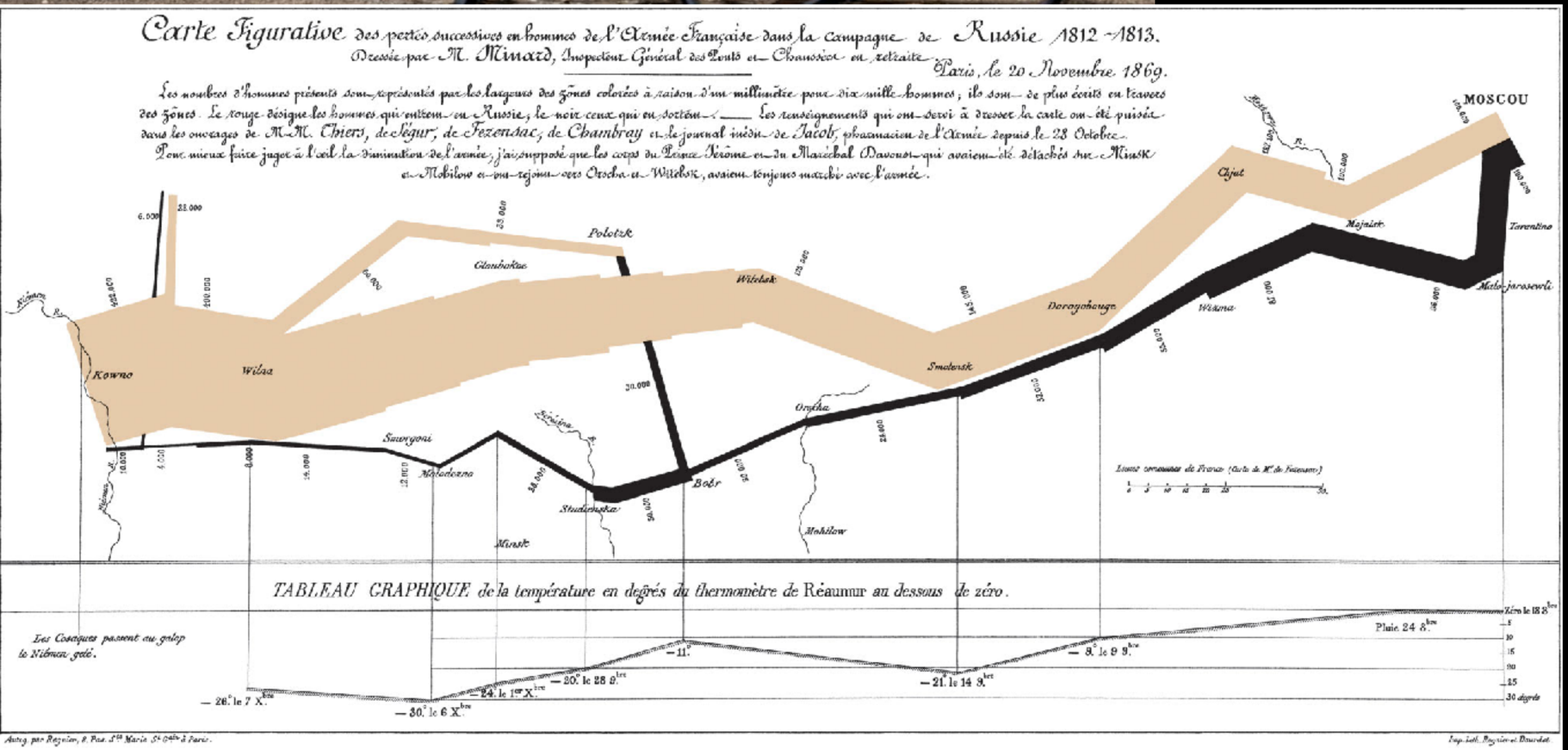
- Al Kindi (801-873):
  - On Deciphering Cryptographic Messages
  - Creator of cryptanalysis
  - Earliest known use of statistical inference

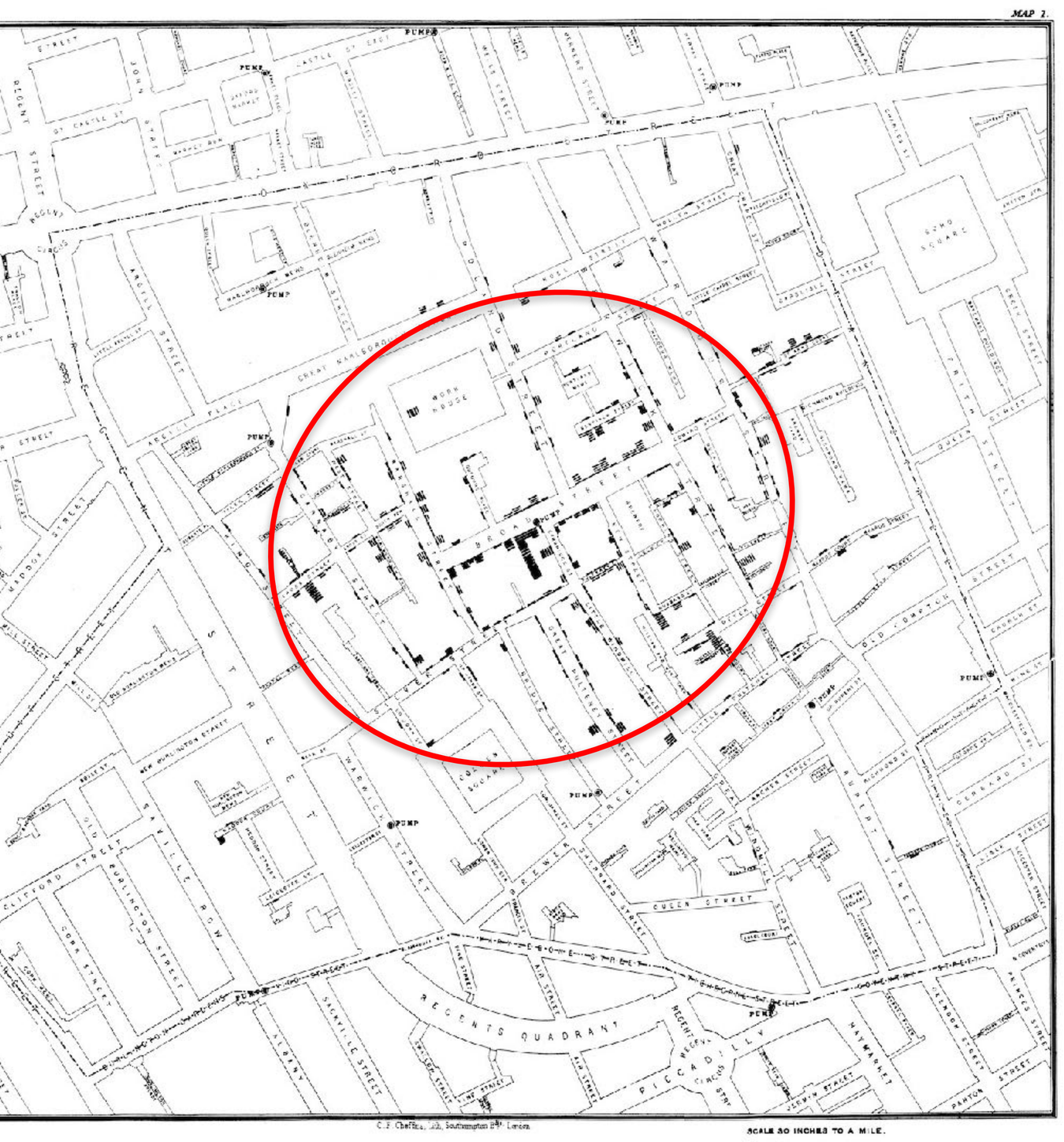
https://en.wikipedia.org/wiki/Al-Kindi

https://datavizblog.com/2013/05/26/dataviz-history-charles-minards-flow-map-of-napoleons-russian-campaign-of-1812-part-5/

Visualization

- Charles Minard's visualization of Napoleon's Russia Campaign (drawn 1861)





Visualization

- "On the Mode of Communication of Cholera", by John Snow (1854)

https://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak

Neural Nets

- 1943 - McCulloch and Pitts - single layer
- ...
- Le Cun, et al. (1989-1990)

Handwritten Zip Code Recognition with Multilayer Networks

Y. Le Cun, O. Matan, B. Boser, J. S. Denker, D. Henderson,
R. E. Howard, W. Hubbard, L. D. Jackel and H. S. Baird
AT&T Bell Laboratories, Holmdel, N. J. 07733

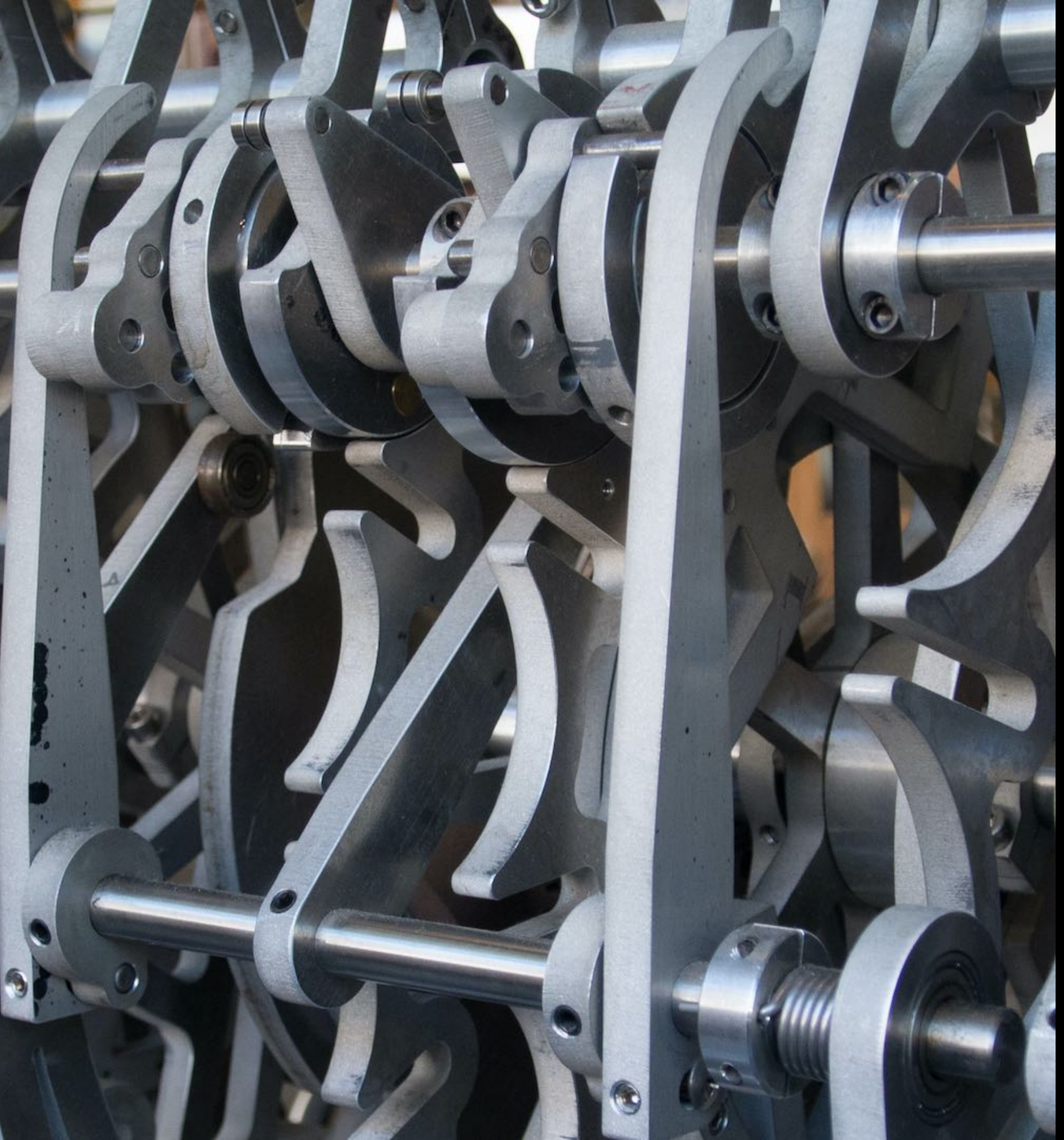
A zip code

Abstract

We present an application of back-

only be obtained by designing a network architecture that contains a certain amount of *a priori* knowledge about the problem. The basic design

http://yann.lecun.com/exdb/publis/pdf/lecun-90e.pdf



Outline

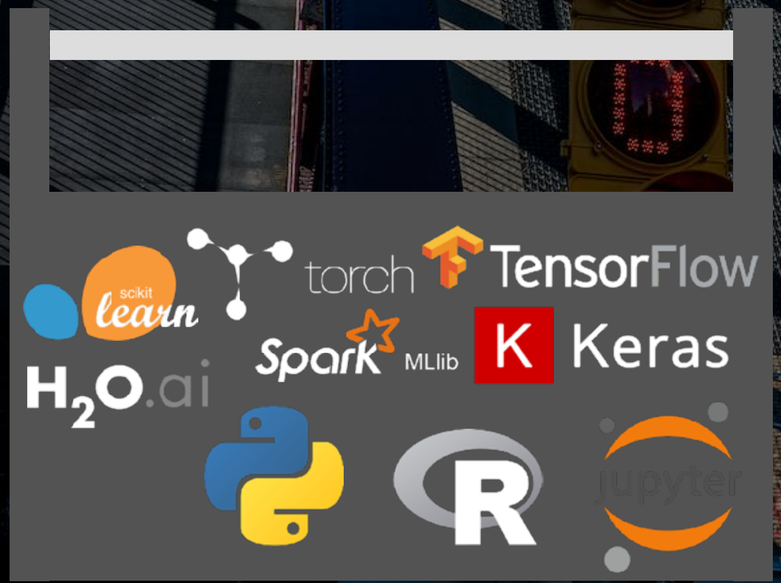
- The Promise of AI
- AI in the Enterprise
  - The Past
  - The Present
  - The Future
- Conclusions


All the current capabilities of the
Promise of AI section are
available now, but they are hard
to build and use.
@deanwampler

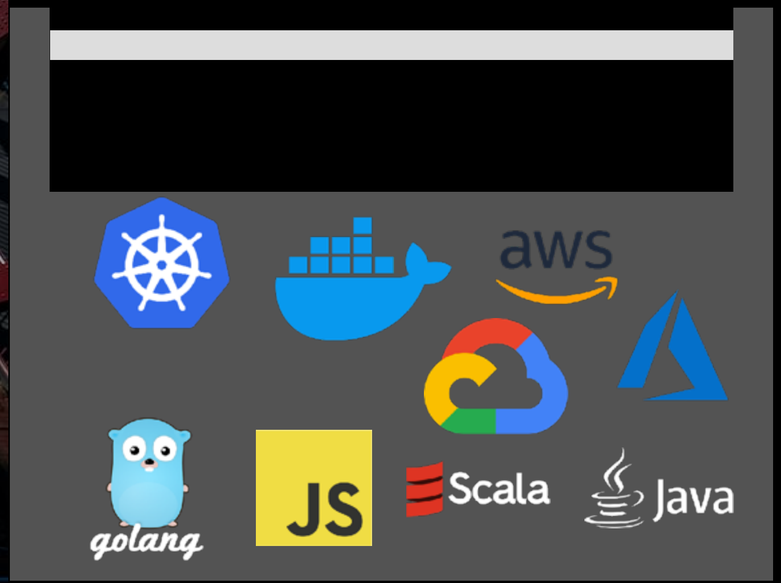
Data Science vs. Data Engineering

Data Science vs. Data Engineering

- A cultural and technical divide



Data Science toolbox



Software Engineering toolbox

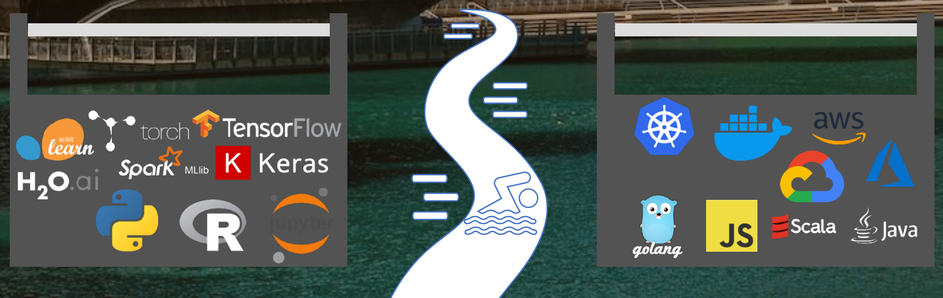
Data Scientists

- Comfortable with uncertainty
- Less process oriented
  - Iterative, experimental

Data Engineers

- Uncomfortable with uncertainty
- Process oriented
  - Agile Manifesto
  - … which does not mention data!

https://derwen.ai/s/6fqt



Bridging the Divide

Data Scientists

- Embrace control and repeatability

Data Engineers

- DevOps ⇒ ModelOps

Bridging the Divide

Data Scientists

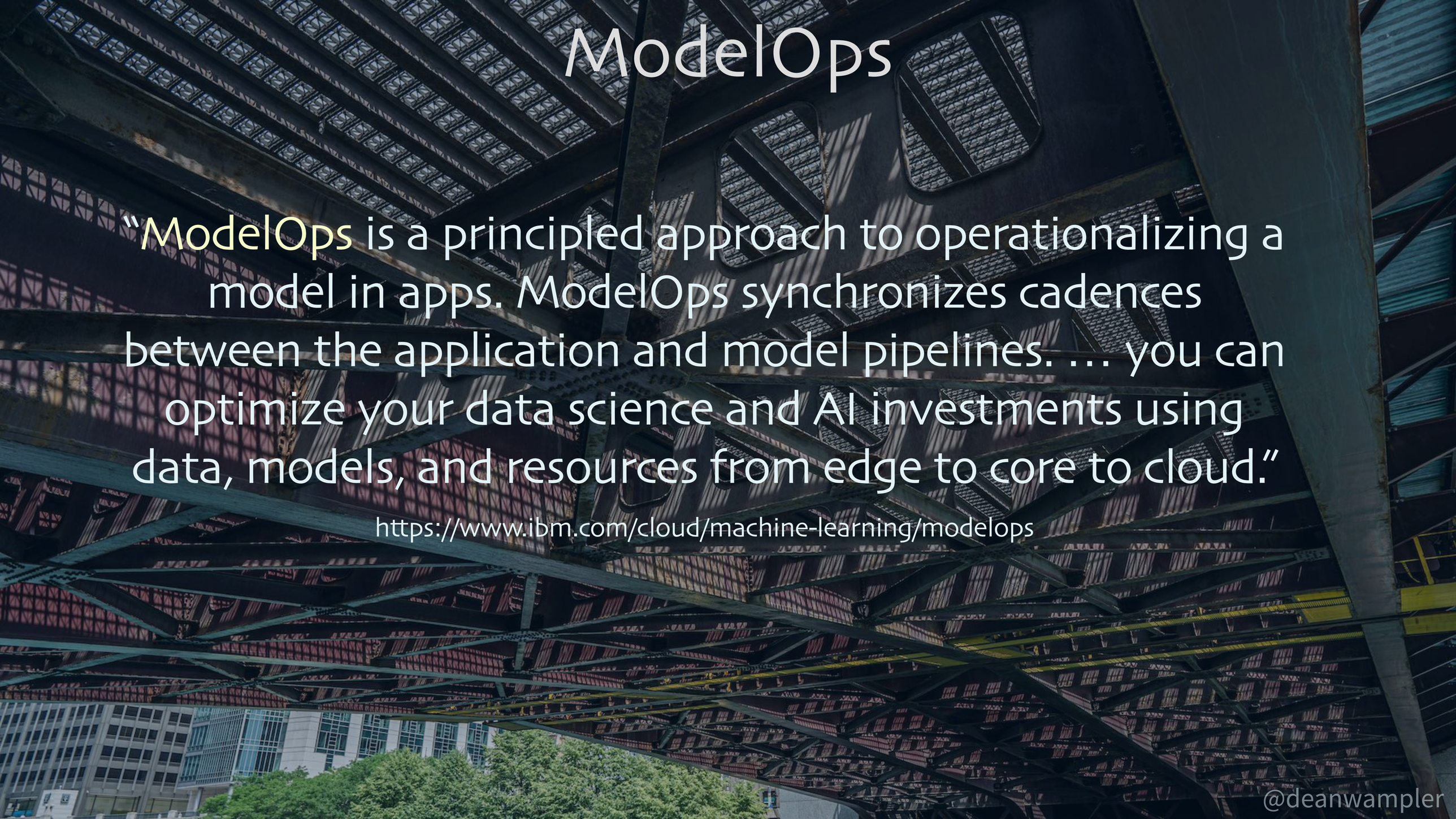
- Embrace control and repeatability

Data Engineers

- DevOps ⇒ ModelOps

Model: An algorithm that makes a prediction or recommendation or prescribes some action based on a probabilistic assessment. Data scientists make models.

https://www.dominodatalab.com/blog/model-management-and-the-era-of-the-model-driven-business/



ModelOps

"ModelOps is a principled approach to operationalizing a model in apps. ModelOps synchronizes cadences between the application and model pipelines. … you can optimize your data science and AI investments using data, models, and resources from edge to core to cloud."

https://www.ibm.com/cloud/machine-learning/modelops

ModelOps

And if you look at the most successful companies in the world, you'll find models at the heart of their business driving that success.

- Example: Netflix recommendation model
  - Drives subscriber engagement, retention, and operational efficiency.
  - Their recommendation model is worth more than $1B per year (2016).

https://www.dominodatalab.com/blog/model-management-and-the-era-of-the-model-driven-business/

ModelOps

And if you look at the most successful companies in the world, you'll find models at the heart of their business driving that success.

- Example: Coca-Cola
  - Optimizes orange juice production, …
- Example: Stitch Fix and Trunk Club
  - Clothing recommendations for customers

https://www.dominodatalab.com/blog/model-management-and-the-era-of-the-model-driven-business/

ModelOps

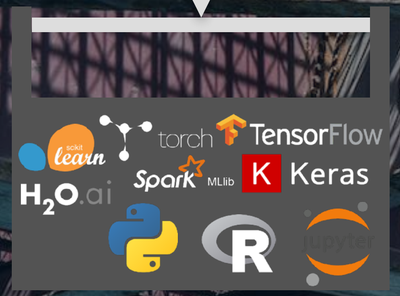
And if you look at the most successful companies in the world, you'll find models at the heart of their business driving that success.

- Example: Insurance companies
  - Actuarial models (very old technique…)
  - Now using models to make automated damage estimates from accident photos, reducing dependence on claims adjusters.

https://www.dominodatalab.com/blog/model-management-and-the-era-of-the-model-driven-business/

ModelOps

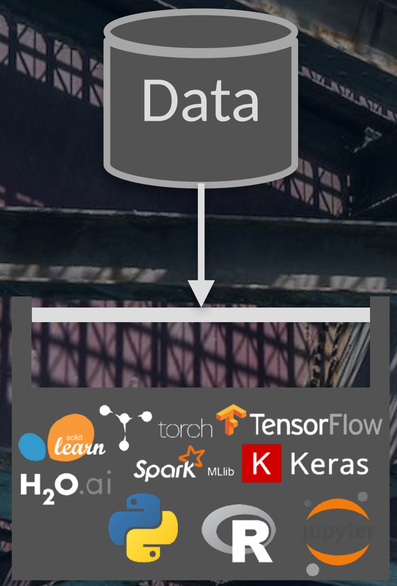
Data



Model Development

Research new models

ModelOps



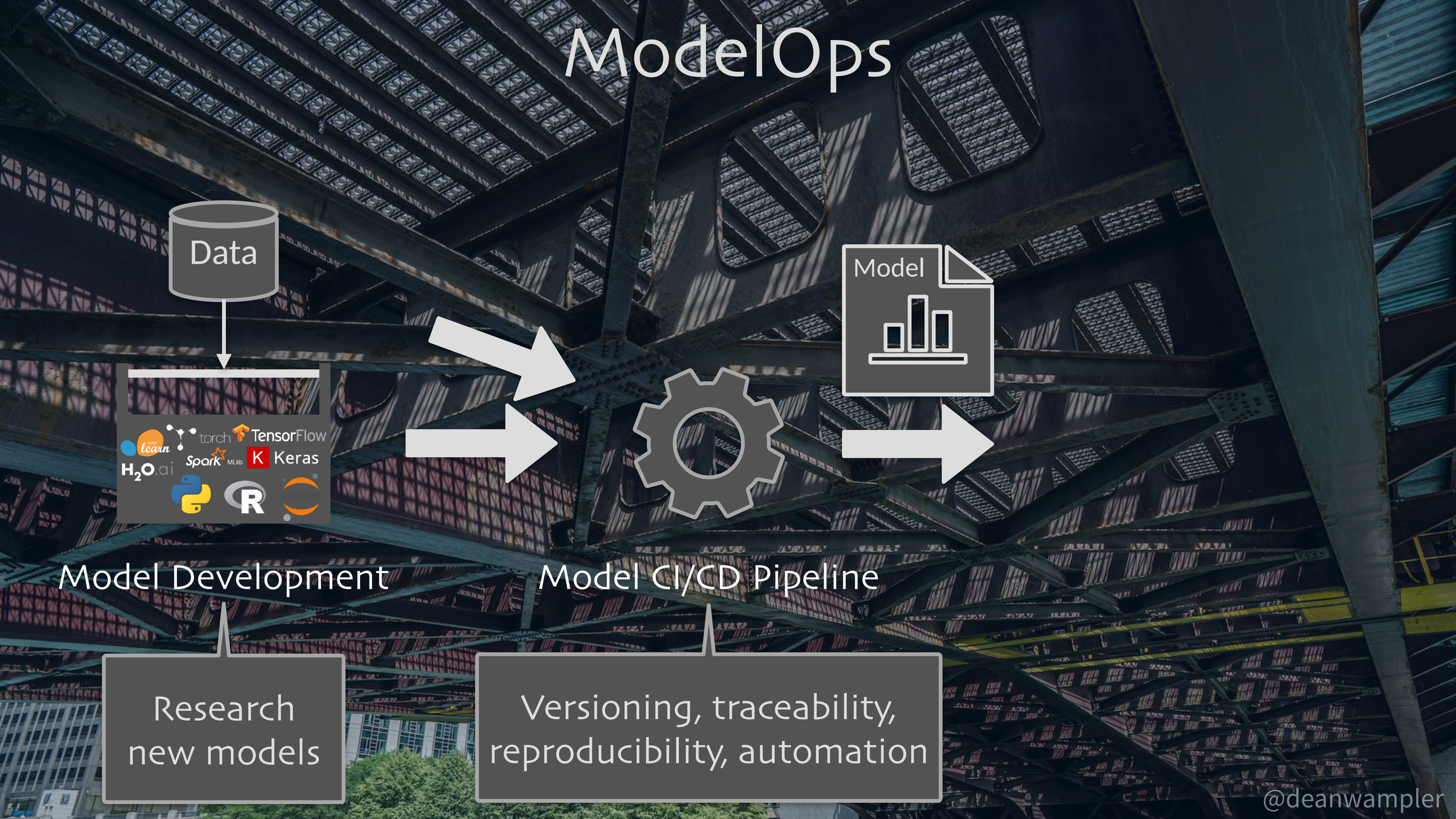
Model Development

Research new models

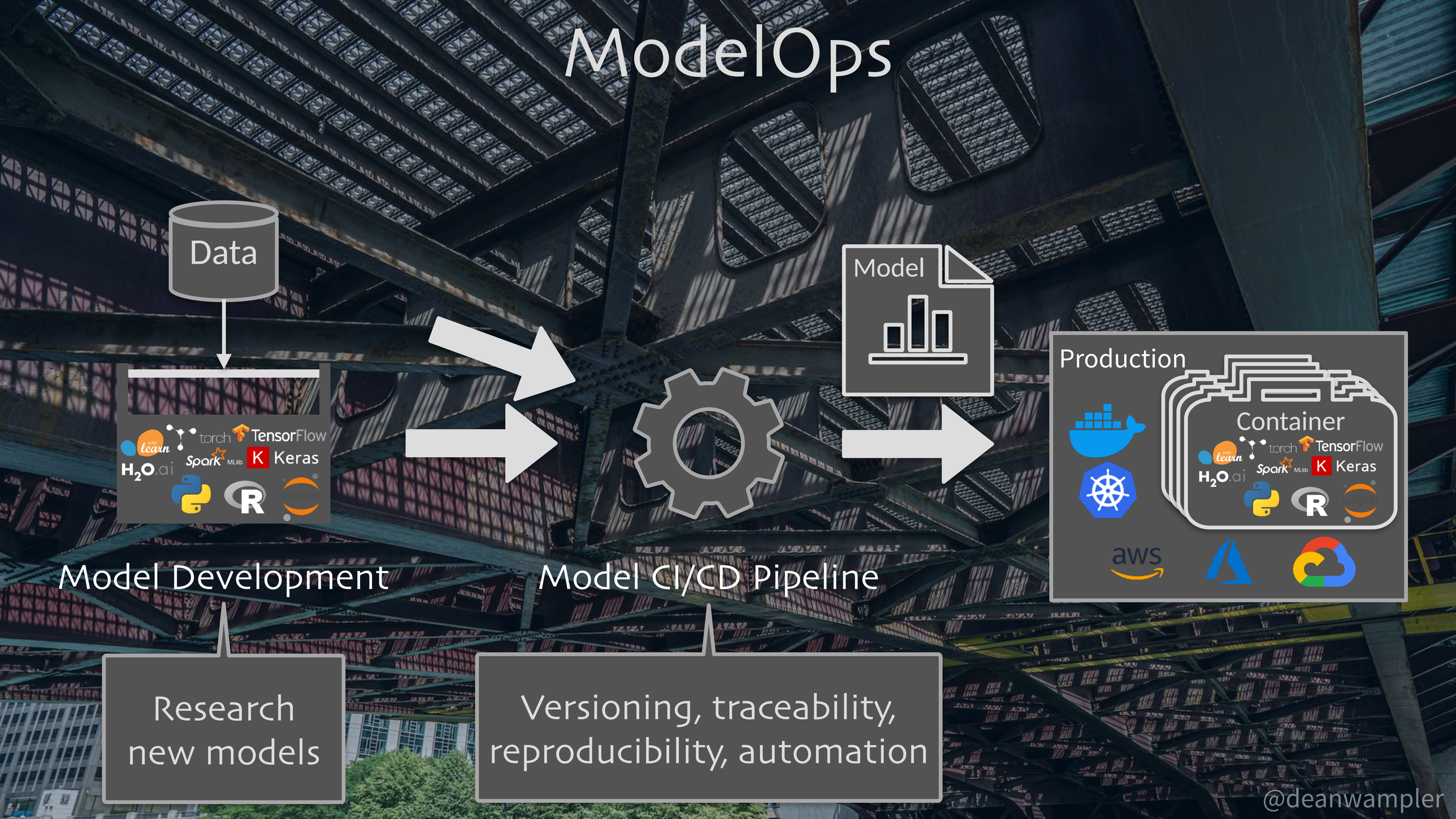
Model CI/CD Pipeline

Versioning, traceability, reproducibility, automation

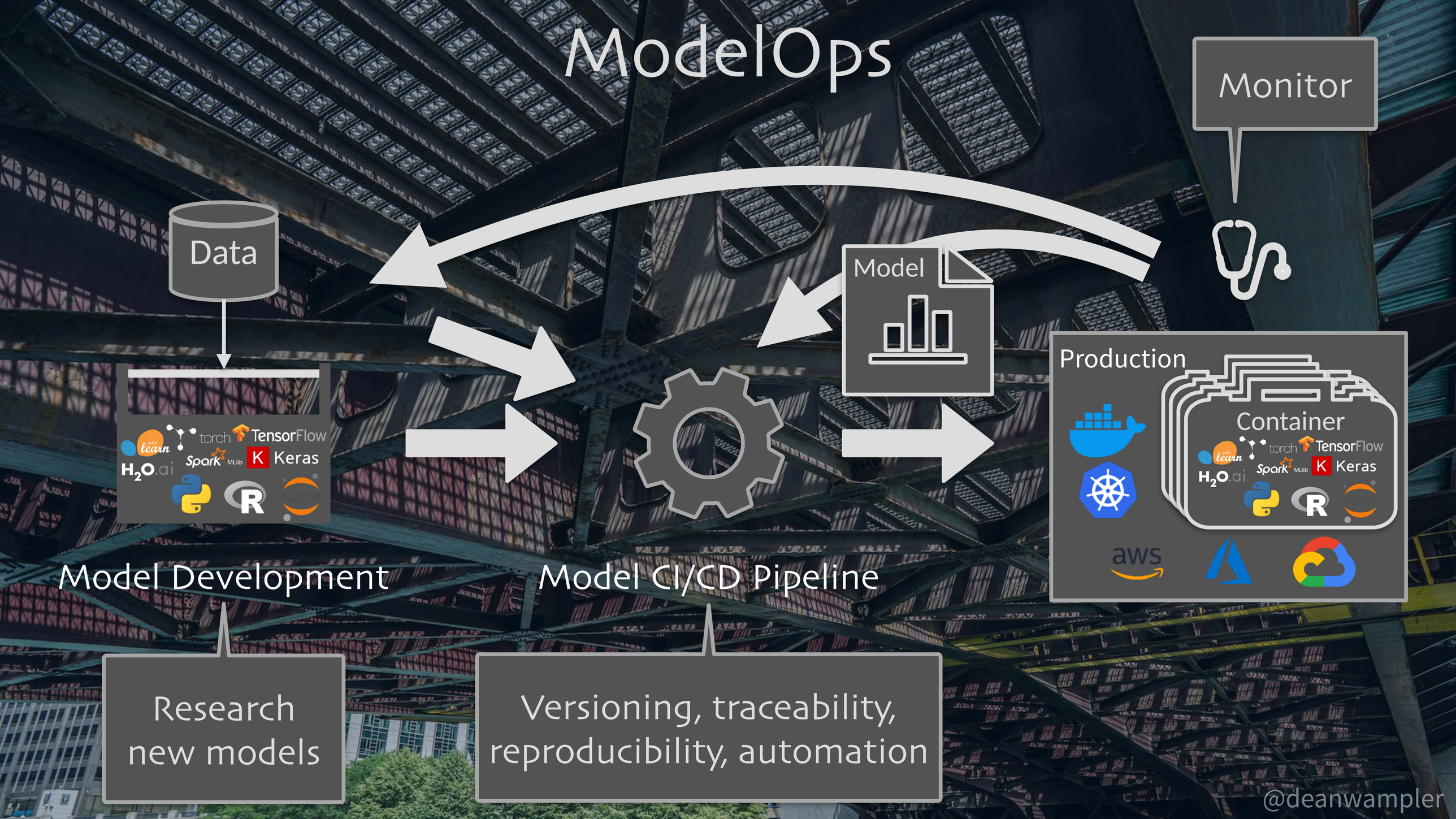
ModelOps


Data
Model
torch
TensorFlow
Spark MLlib
K Keras
R
Model Development
Model CI/CD Pipeline
Research new models
Versioning, traceability, reproducibility, automation

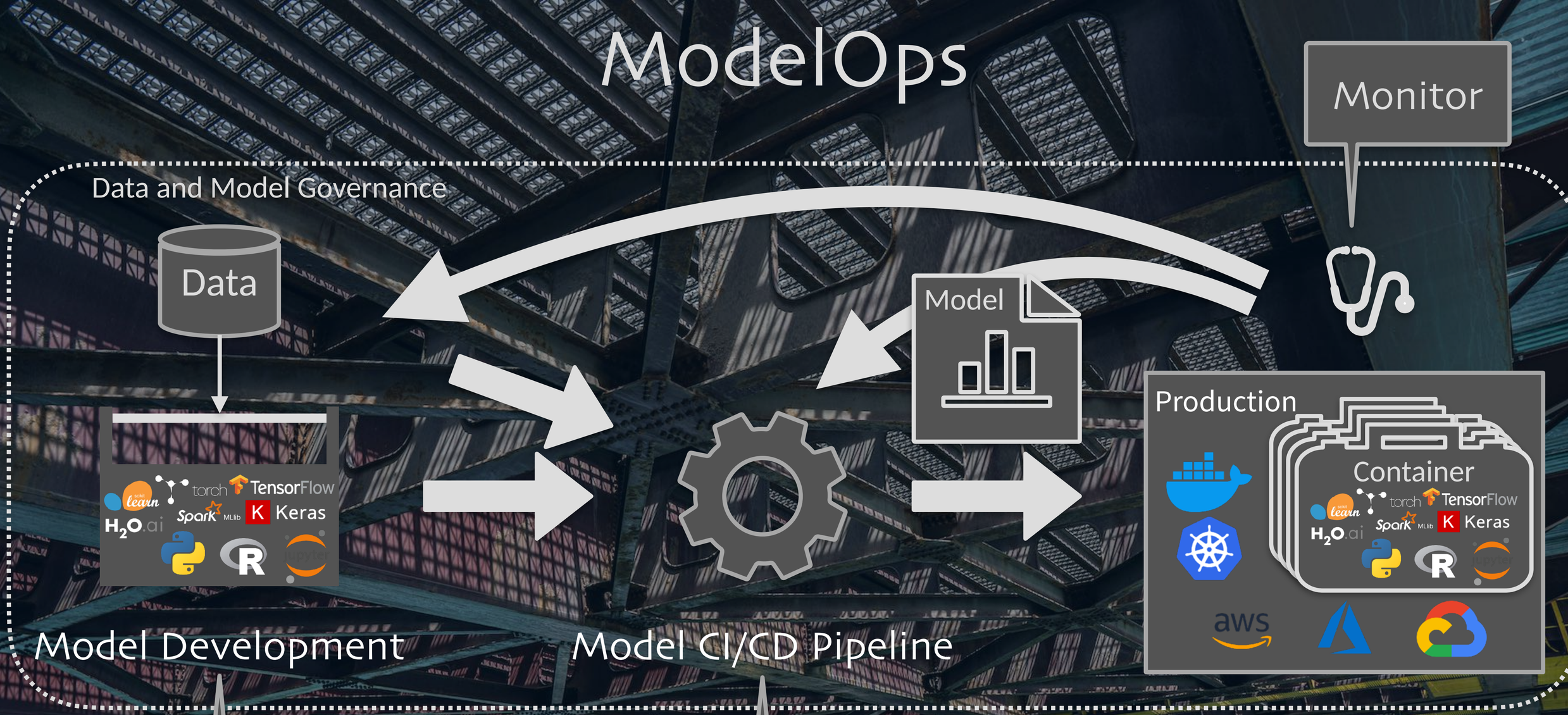
ModelOps


Data
Model
Production
Container
Model Development
Model CI/CD Pipeline
Research new models
Versioning, traceability, reproducibility, automation

ModelOps



ModelOps



ModelOps

Monitor

This is shown as a batch process, but expect these processes to evolve into streaming pipelines, with continuous training.

Mode

ntainer

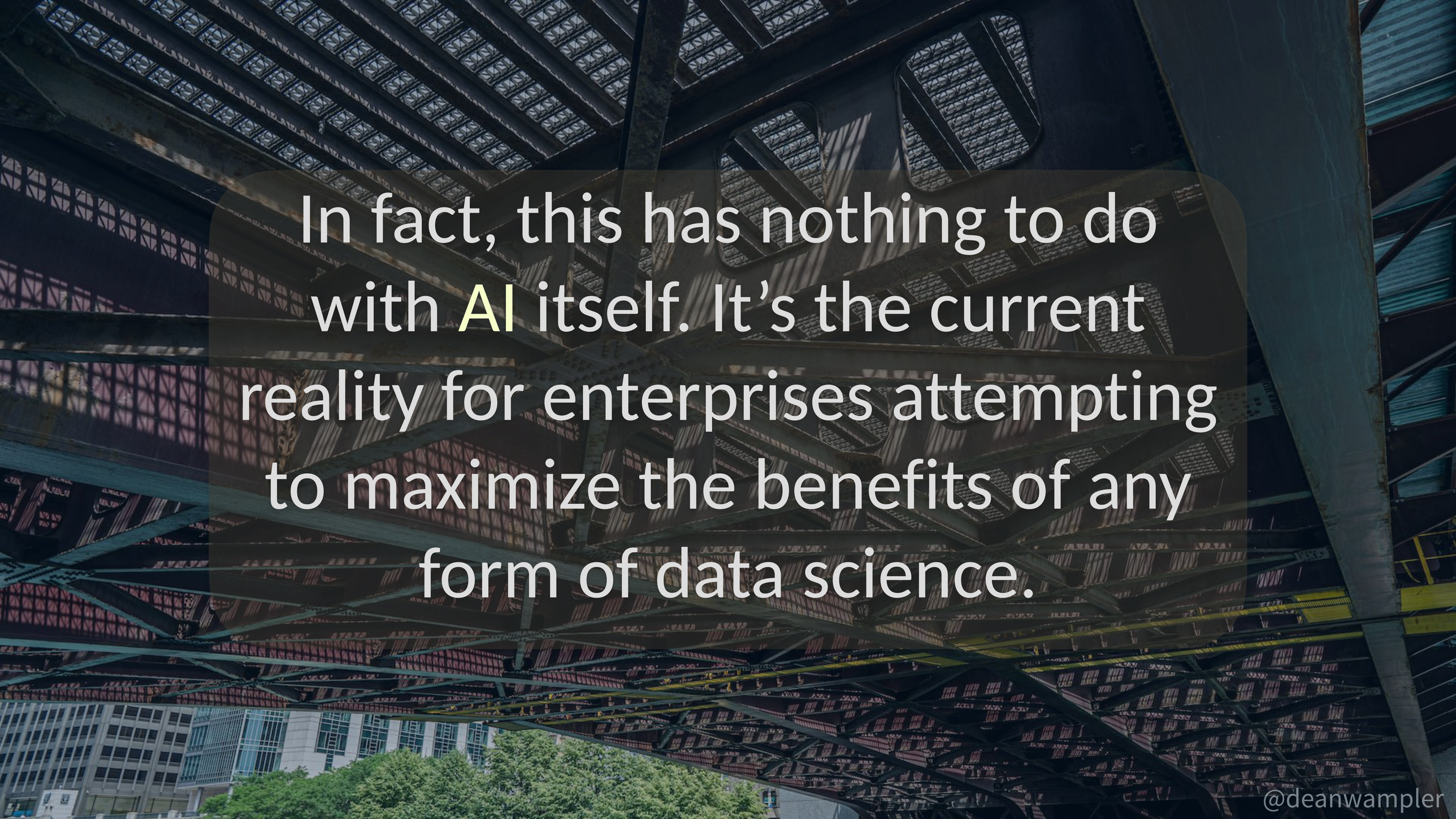
torch TensorFlow

MLlib K Keras

R

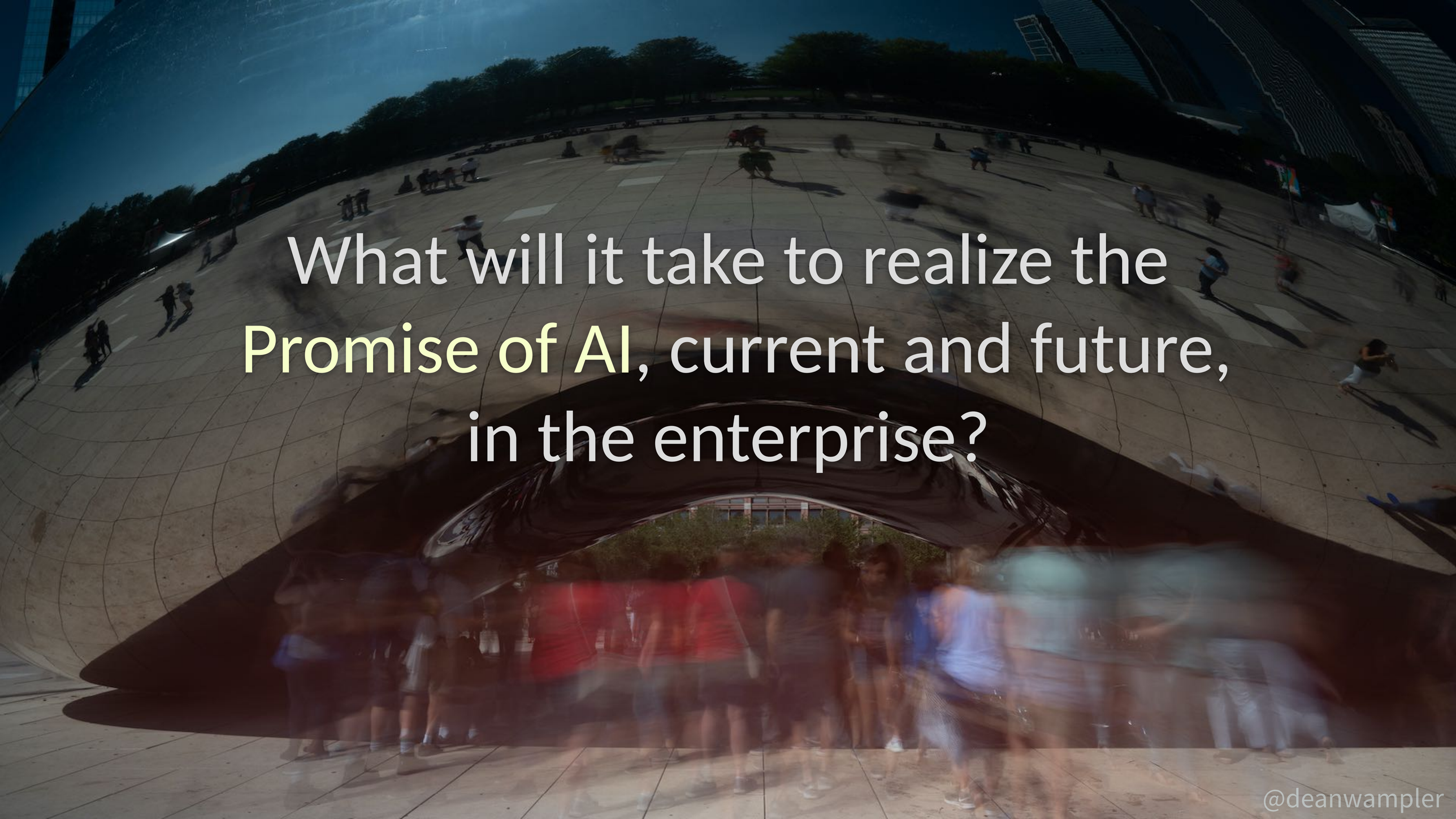
new models

reproducibility, automation


In fact, this has nothing to do with AI itself. It’s the current reality for enterprises attempting to maximize the benefits of any form of data science.
@deanwampler

Outline

- The Promise of AI
- AI in the Enterprise
  - The Past
  - The Present
  - The Future
- Conclusions


What will it take to realize the
Promise of AI, current and future,
in the enterprise?
@deanwampler

AI in the Enterprise

- Fully adopting:
  - Natural Language Processing
  - Reinforcement Learning
- Ubiquitous AI in Applications

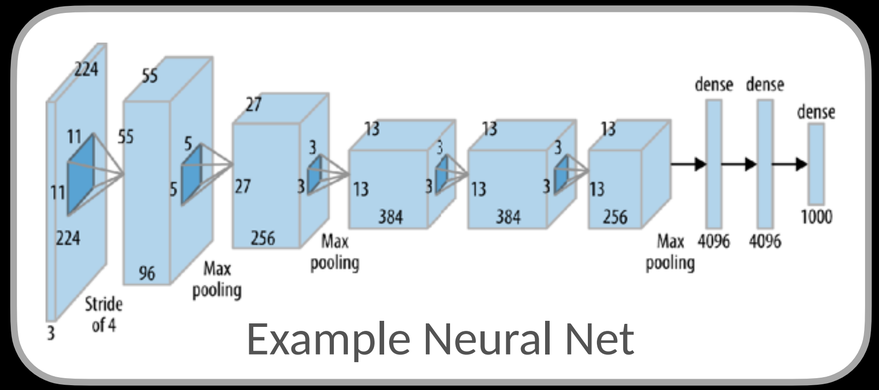
Along the way...

- Infrastructure Changes
  - Cloud
  - Scaling computation
  - Diff. Privacy & Fed. Learning
- Software Development

"The largest version GPT-3 175B or 'GPT-3' has 175 B Parameters, 96 attention layers and 3.2 M batch size."

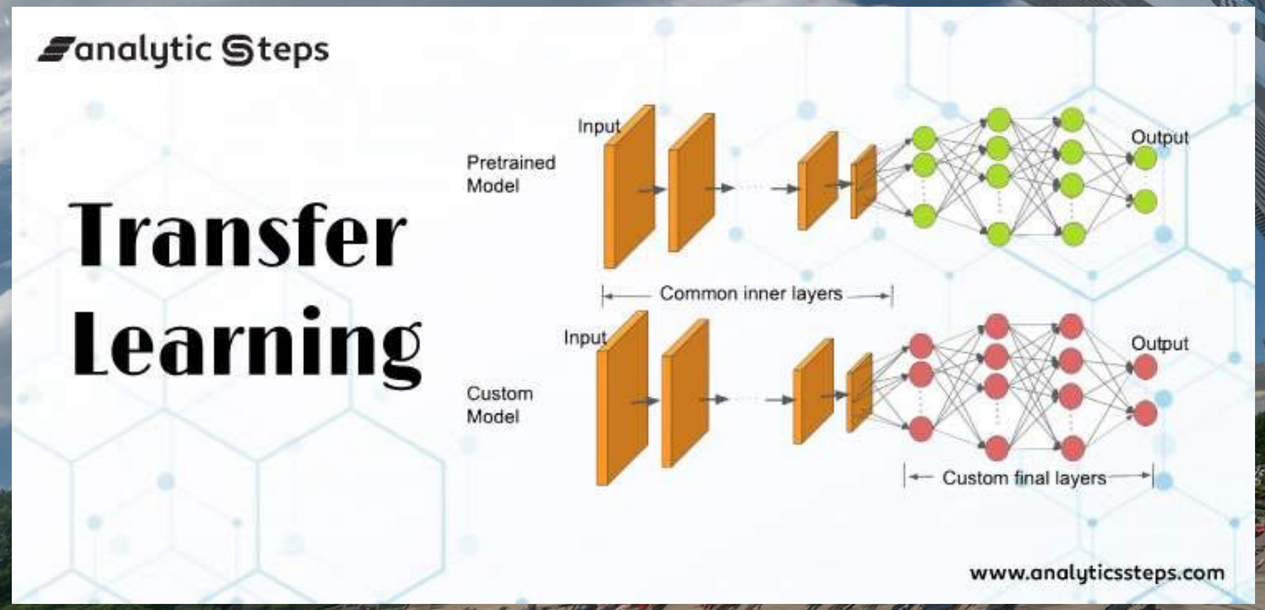
Transfer Learning

- NLP: The world's largest neural networks



Example Neural Net

https://in.springboard.com/blog/openai-gpt-3/



Transfer Learning

- Fortunately, you can start with a trained model and further refine it for your problem.

https://analyticssteps.com/blogs/how-transfer-learning-done-neural-networks-and-convolutional-neural-networks

Reinforcement Learning

- While “classic” RL uses a simulator, you can also train on historical (“offline”) data.
  - Use when a good simulator doesn’t exist or is too hard to create.

https://arxiv.org/abs/2005.01643

Infrastructure

- Model training, especially NNs, is very expensive.
  - Burst to the cloud
  - Or have lots of in-house compute available!

Infrastructure

- A hybrid-cloud model balances:
  - Security & regulatory benefits of on-premise cluster
  - Burst of resources when you need them.

Infrastructure

- But, don't forget the cost of moving data between on-premise clusters and the cloud, as well as between clouds!

Infrastructure

- Leverage federated learning and differential privacy.
  - Offload some computation!
  - Meet data privacy objectives.

https://openmined.org/

Software Development Impacts

- Ubiquitous AI requires:
  - Heterogeneous tools
  - Batch and stream data processing
  - Statistical & probabilistic thinking

Software Development Impacts

- Ubiquitous AI requires:
  - Heterogeneous tools
  - Batch and streaming data processing
  - Statistical & probabilistic thinking

Software Development Impacts

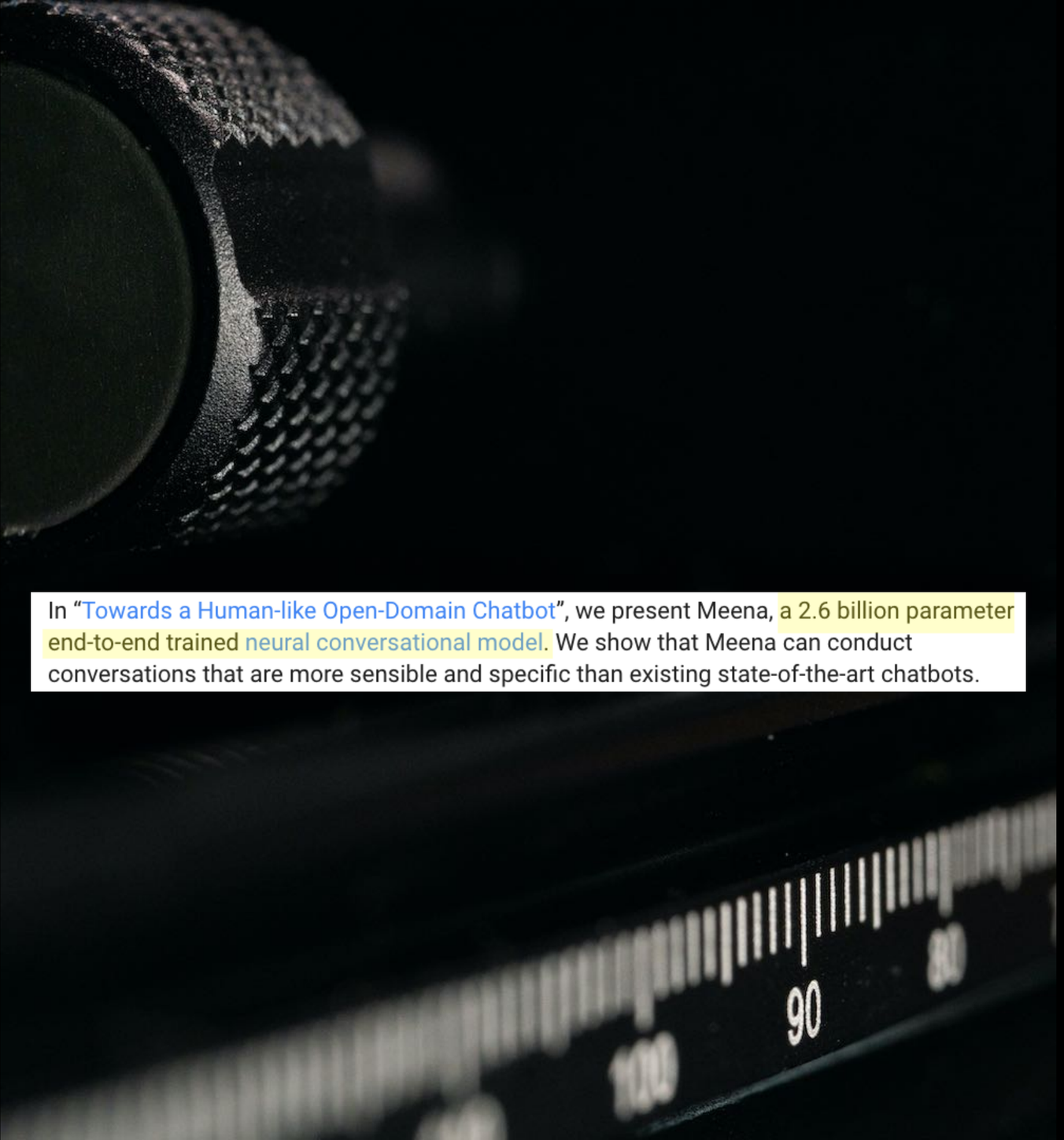
- Ubiquitous AI requires:
  - Heterogeneous tools
  - Batch and streaming data processing
  - Statistical & probabilistic thinking

Software Development Impacts

- Ubiquitous AI requires:
  - Heterogeneous tools
  - Batch and streaming data processing
  - Statistical & probabilistic thinking

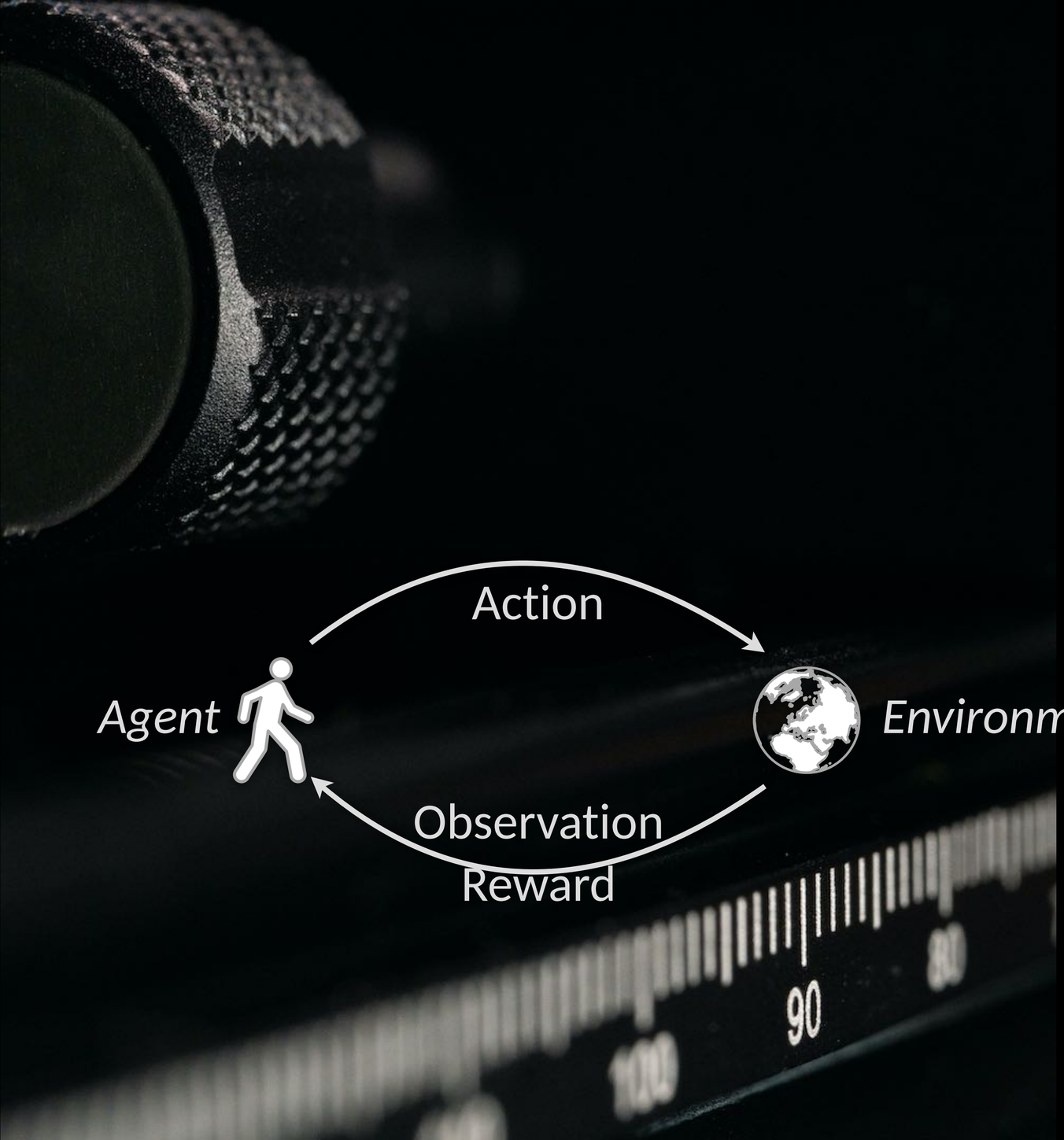
Outline

- The Promise of AI
- AI in the Enterprise
  - The Past
  - The Present
  - The Future
- Conclusions



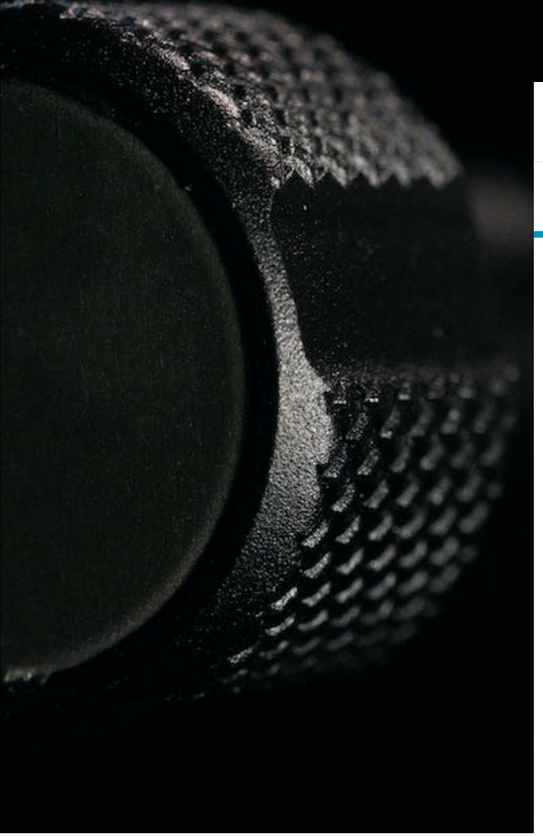
AI’s Promise

- Natural Language Processing has become very capable, with wide applications



AI's Promise

- Reinforcement Learning is being applied to many enterprise problems where sequential activity is central.



nature reviews cancer

View all

Explore our content ⌄ Journal information ⌄

nature › nature reviews cancer › perspectives › article

Perspective | Published: 17 May 2018

OPINION

Artificial intelligence in radiology

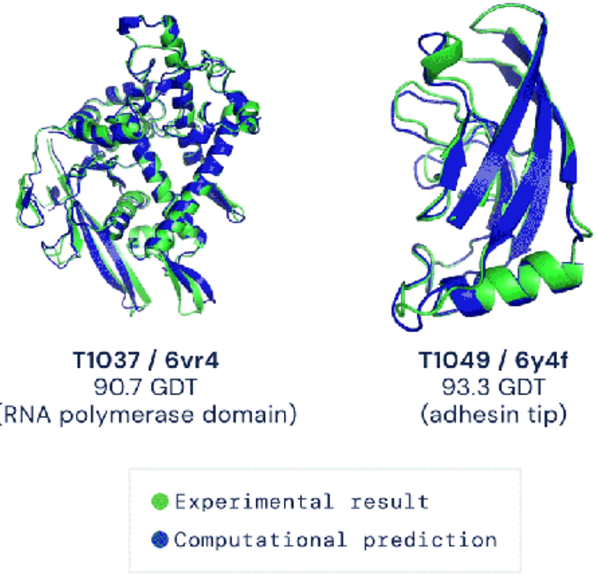
Ahmed Hosny, Chintan Parmar, John Quackenbush, Lawrence H. Schwartz & Hugo J. W. L. Aerts ✉

Nature Reviews Cancer **18**, 500–510(2018) | Cite this article

15k Accesses | **317** Citations | **311** Altmetric | Metrics

Abstract

Artificial intelligence (AI) algorithms, particularly deep learning, have
ognition tasks. Methods
riational autoencoders
age analysis field,
in radiology practice,
ges for the detection

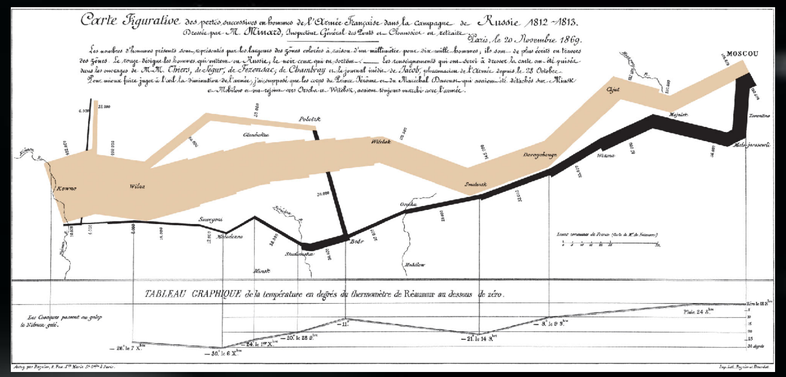


AI's Promise

- New sciences and industries are benefiting from AI

AI's Promise

- Mobile phones are showing us how AI is enabling new system features and enhancing capabilities in applications

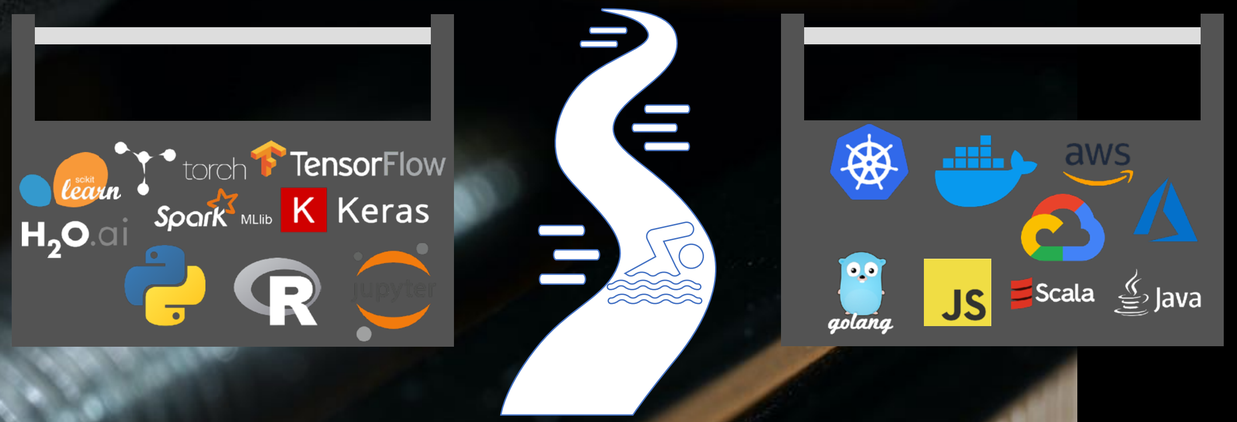


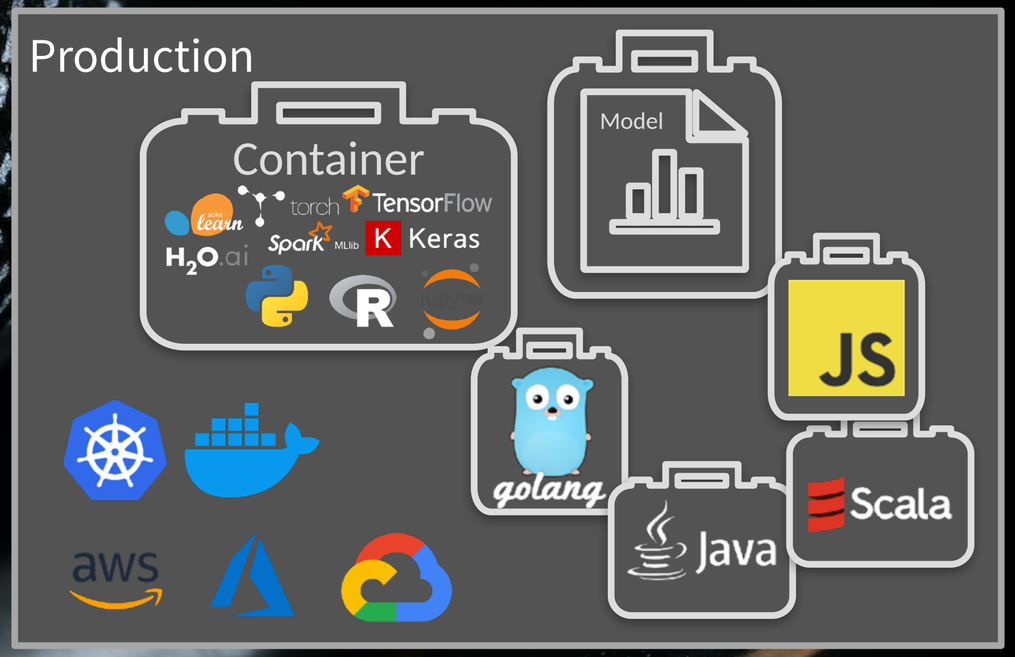
The Past

- Traditional data science tools still provide important benefits:
  - Proven Maturity
  - Explainability
  - Cheap to use!

The Present

- We have to bridge the divide between data science and data engineering now.
  - Or AI won't be an option.





The Future

- To fully benefit, we need to embrace:
  - Scalable compute
  - Hybrid cloud
  - Kubernetes & containers
  - New SW design and implementation tools and techniques

Questions?

DOMINO

dean@deanwampler.com
@deanwampler
polyglotprogramming.com/talks