The World of Artificial Intelligence

Conversation around AI has increased materially in past 5 years

“AI is the ‘runtime’ that is going to shape all what we do”

“AI can make humans more productive than ever imagined”

“We are now solving problems with machine learning that were in the realm of science fiction for the last several decades”

Comments on AI/ML by Indian Companies

"Multiple Machine Learning based credit models developed, 2000 attributes considered, upto 40%+ lift on GINI over generic bureau models"

AXIS BANK

"AI Doctor Assistant"

Apollo HOSPITALS

"AI based pre-claim assessment"

ICICI PRUDENTIAL LIFE INSURANCE

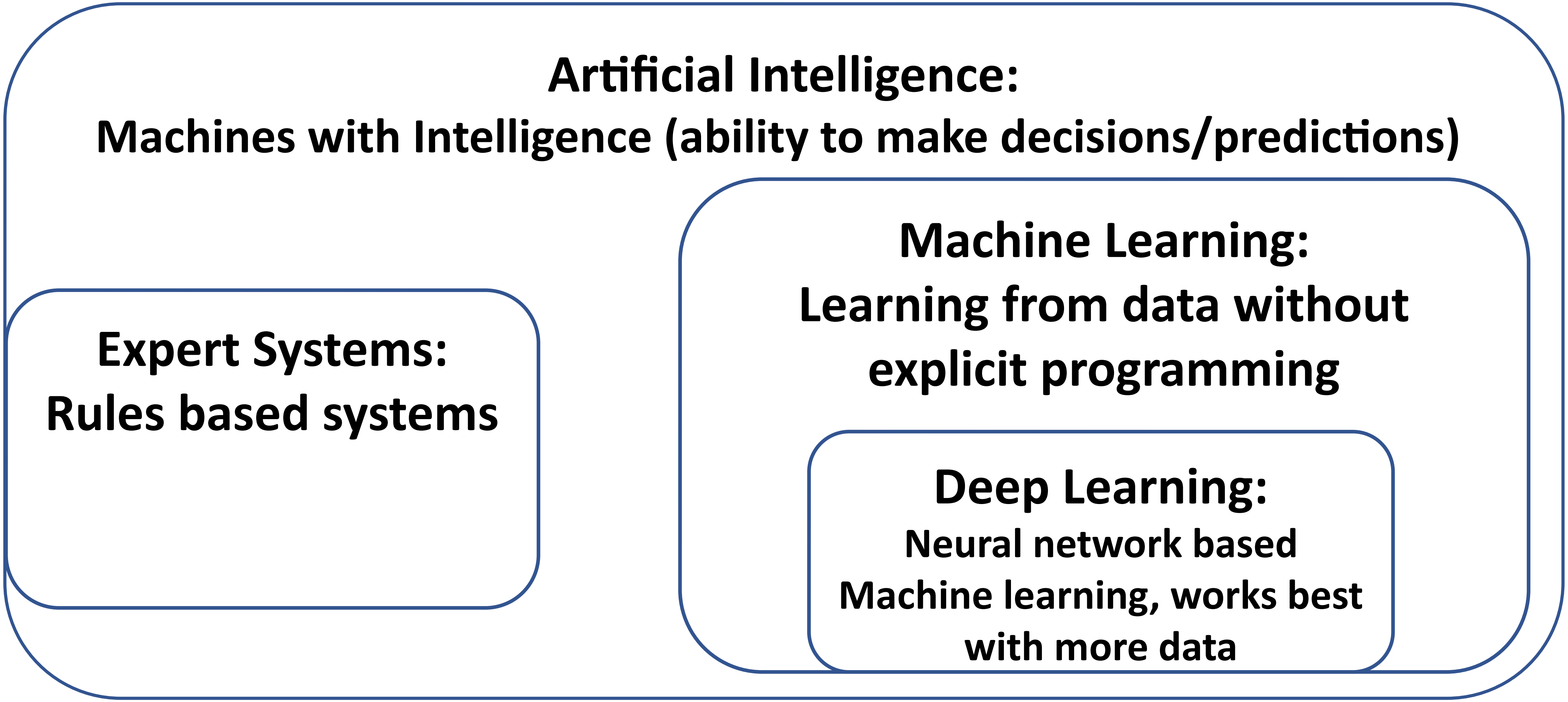
"AI led demand and commodity forecasting; ML powered intelligent planning for distributors"

Hindustan Unilever Limited

Source: Company Presentation / Transcripts

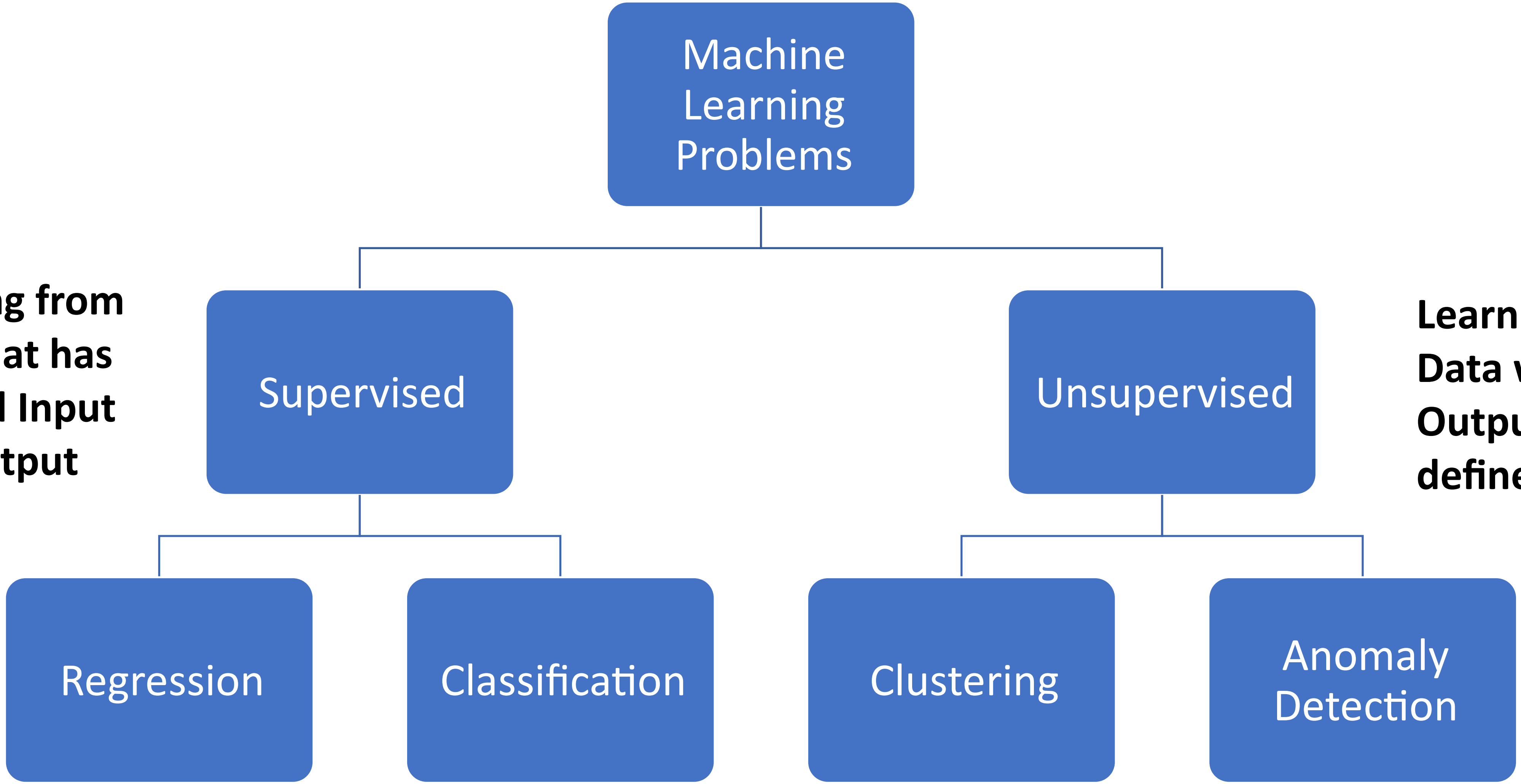
What is AI/ML?

AI vs. ML


Artificial Intelligence:
Machines with Intelligence (ability to make decisions/predictions)
Expert Systems:
Rules based systems
Machine Learning:
Learning from data without explicit programming
Deep Learning:
Neural network based Machine learning, works best with more data

Types of Machine Learning Problems

Learning from Data that has defined Input and Output



Learning from Data where Output isn’t defined

Regression: Predicting Continuous Outputs

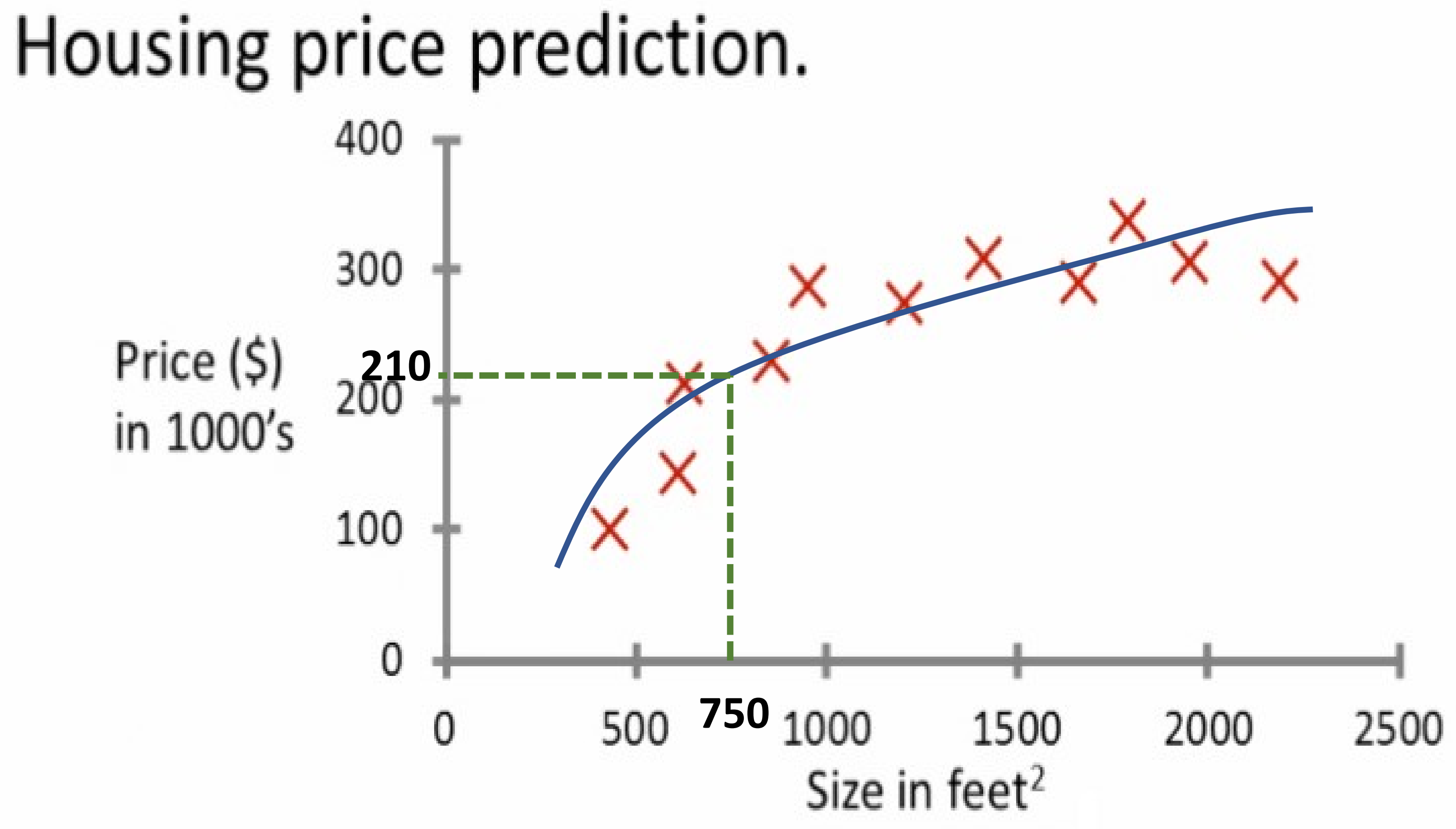
Example: House Price Prediction based on House Size

House Size (Input)	House Price ($ '000) (Output)
470	100
600	130
630	210
…	…
…	…
…	…

Source: Andrew Ng Slides

Regression: Predicting Continuous Outputs
Example: House Price Prediction based on House Size

House Size (Input)	House Price ($ '000) (Output)
470	100
600	130
630	210
…	…
…	…
…	…

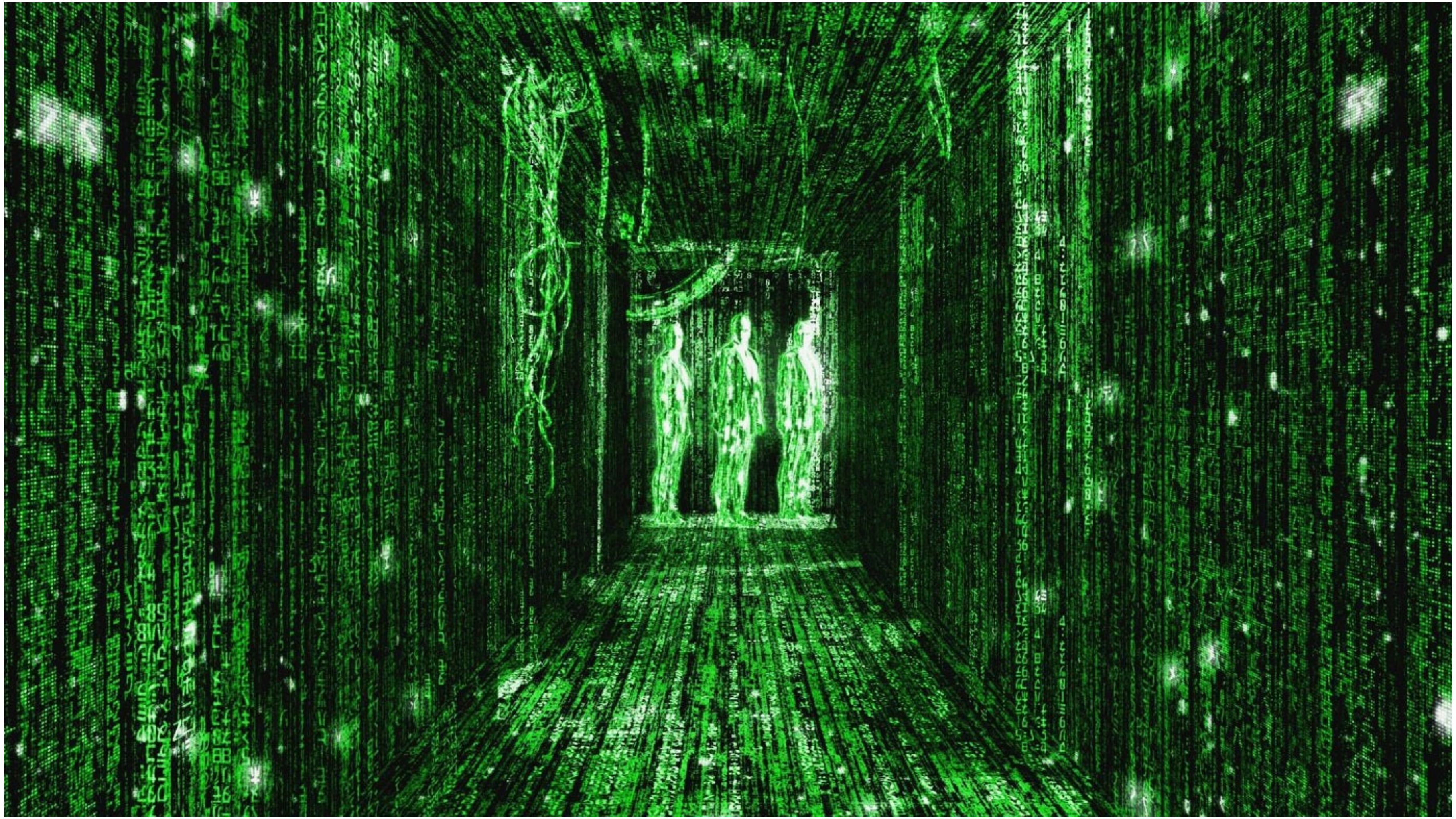


What is the price of a house with size 750 sq ft?

Source: Andrew Ng Slides

Classification: Prediction the Class/Category

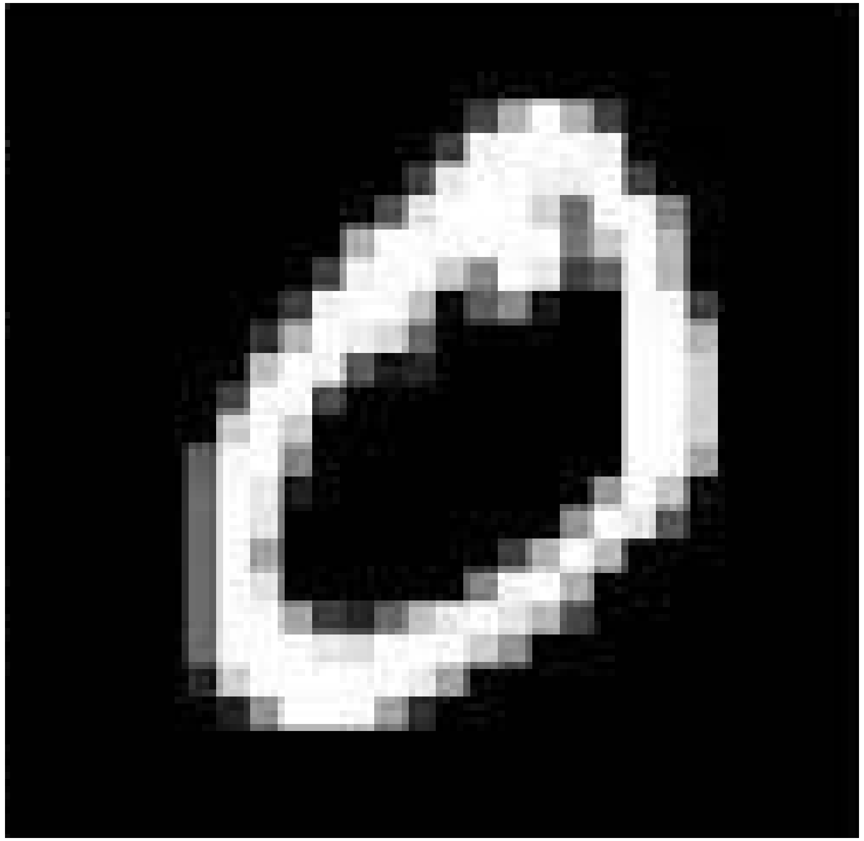
How machines see images?

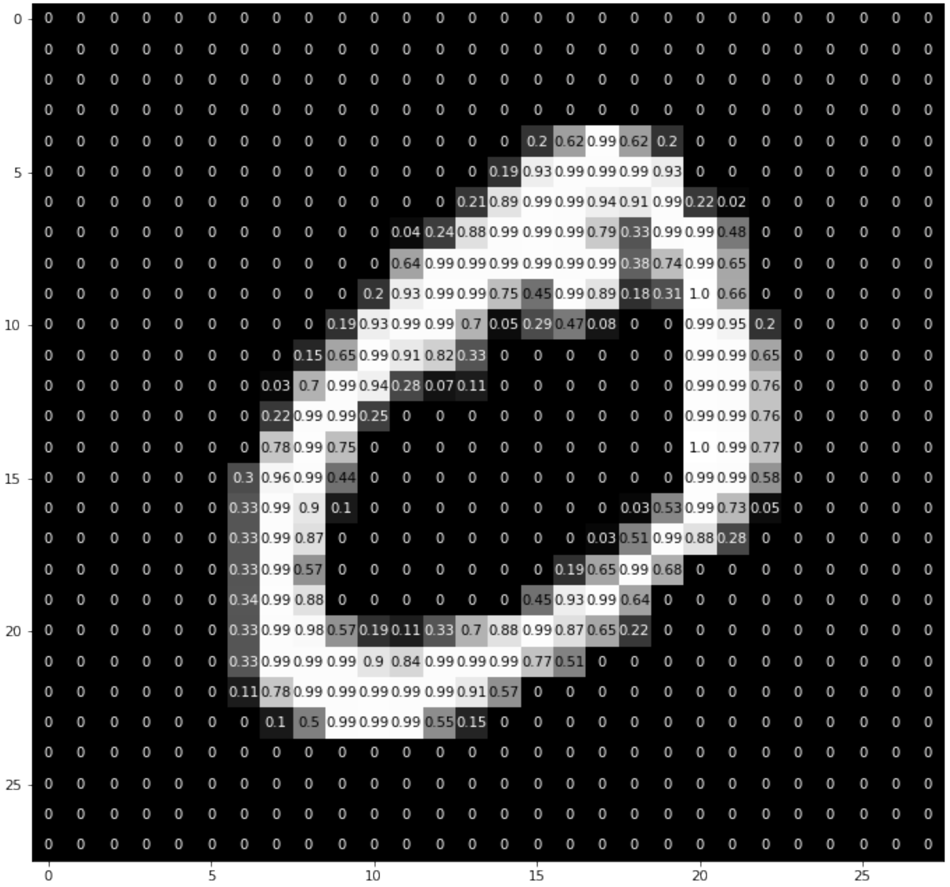


Classification: Prediction the Class/Category

Example: Image Classification – Identify Digits

28



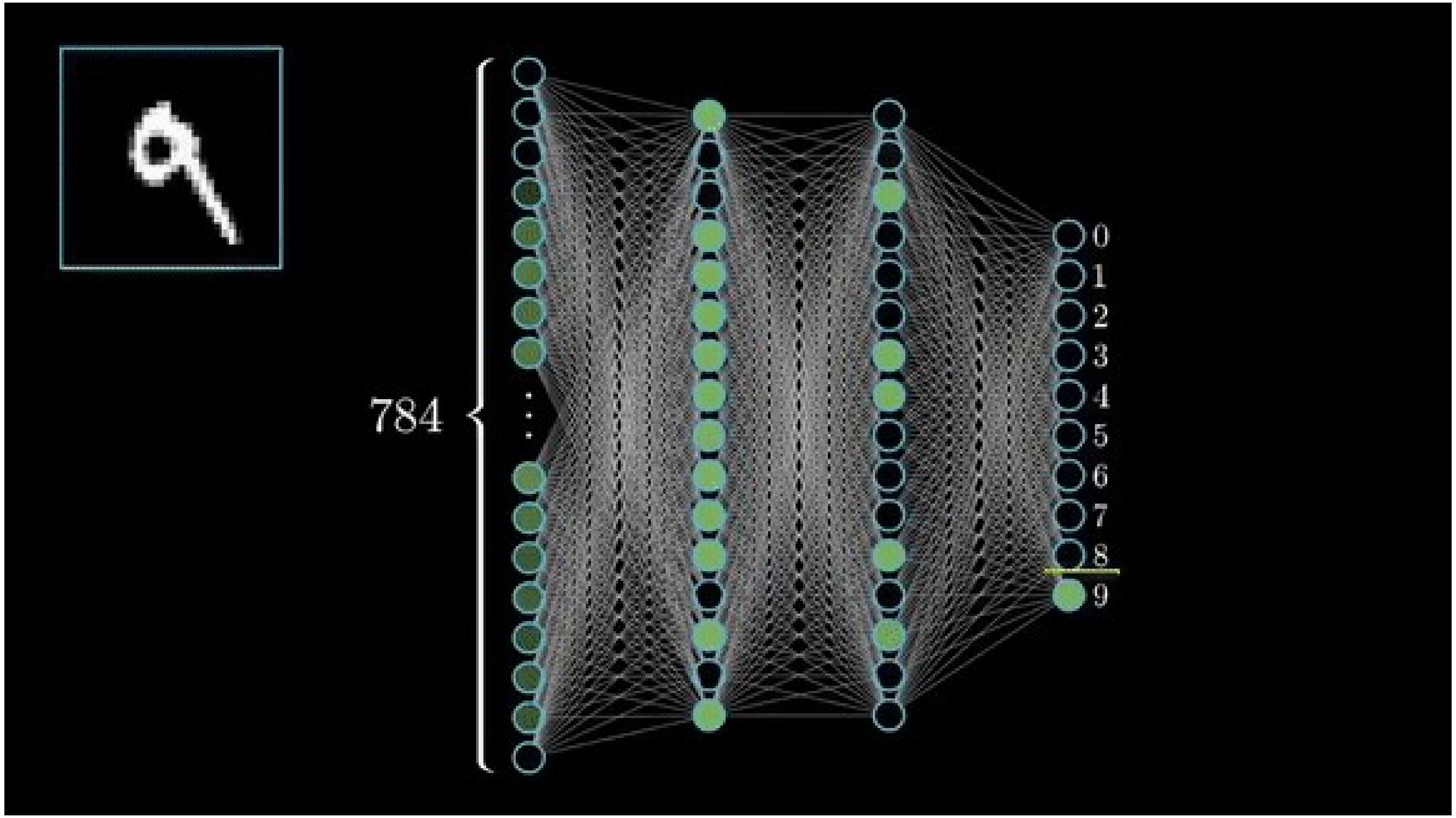


28

28x28 = 784
Pixels/Numbers

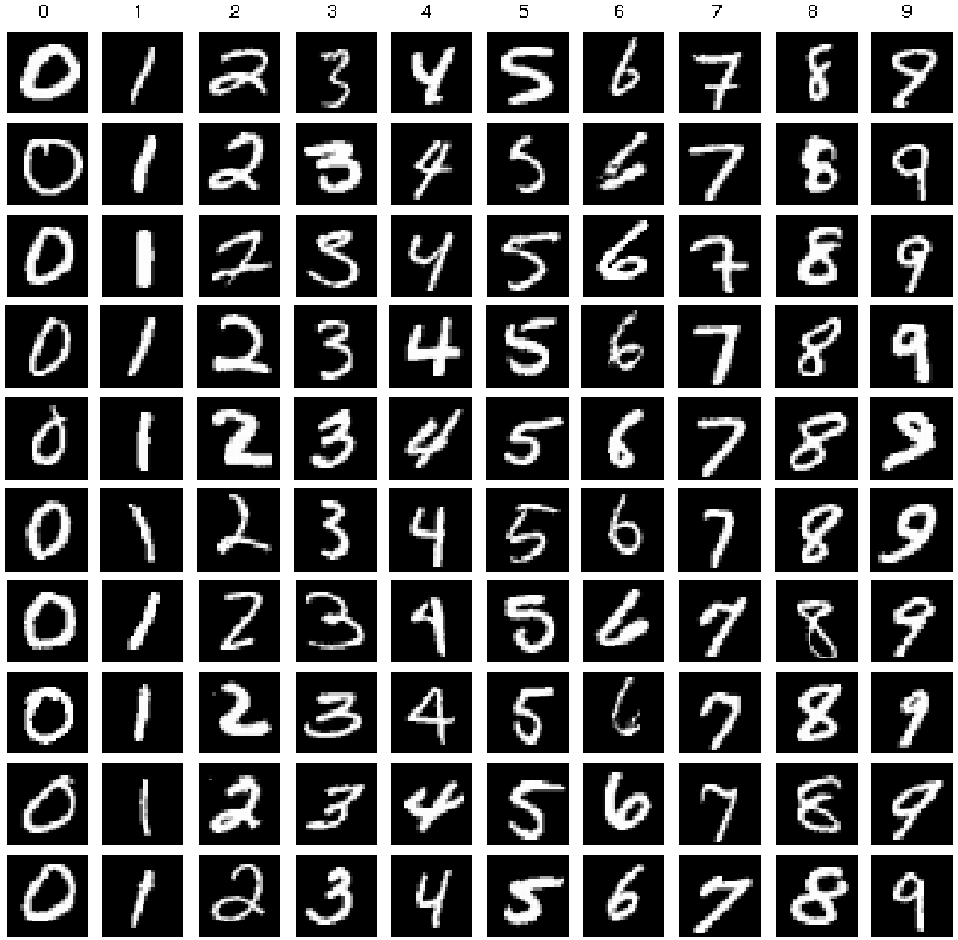
Classification: Prediction the Class/Category

Example: Image Classification – Identify Digits



Classification: Prediction the Class/Category
Example: Image Classification – Identify Digits

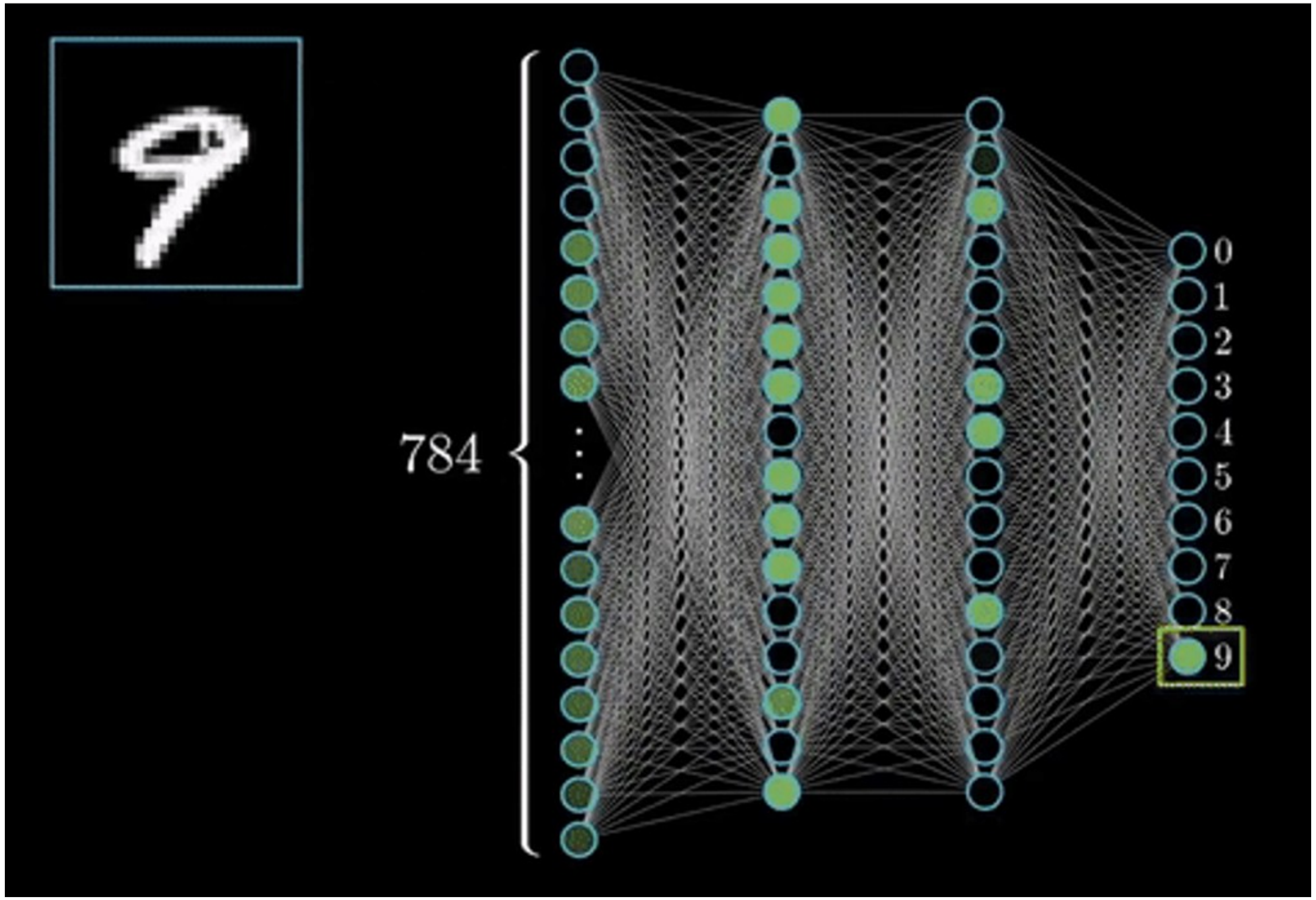
More Training Examples → Better Accuracy



Source: MNIST Database

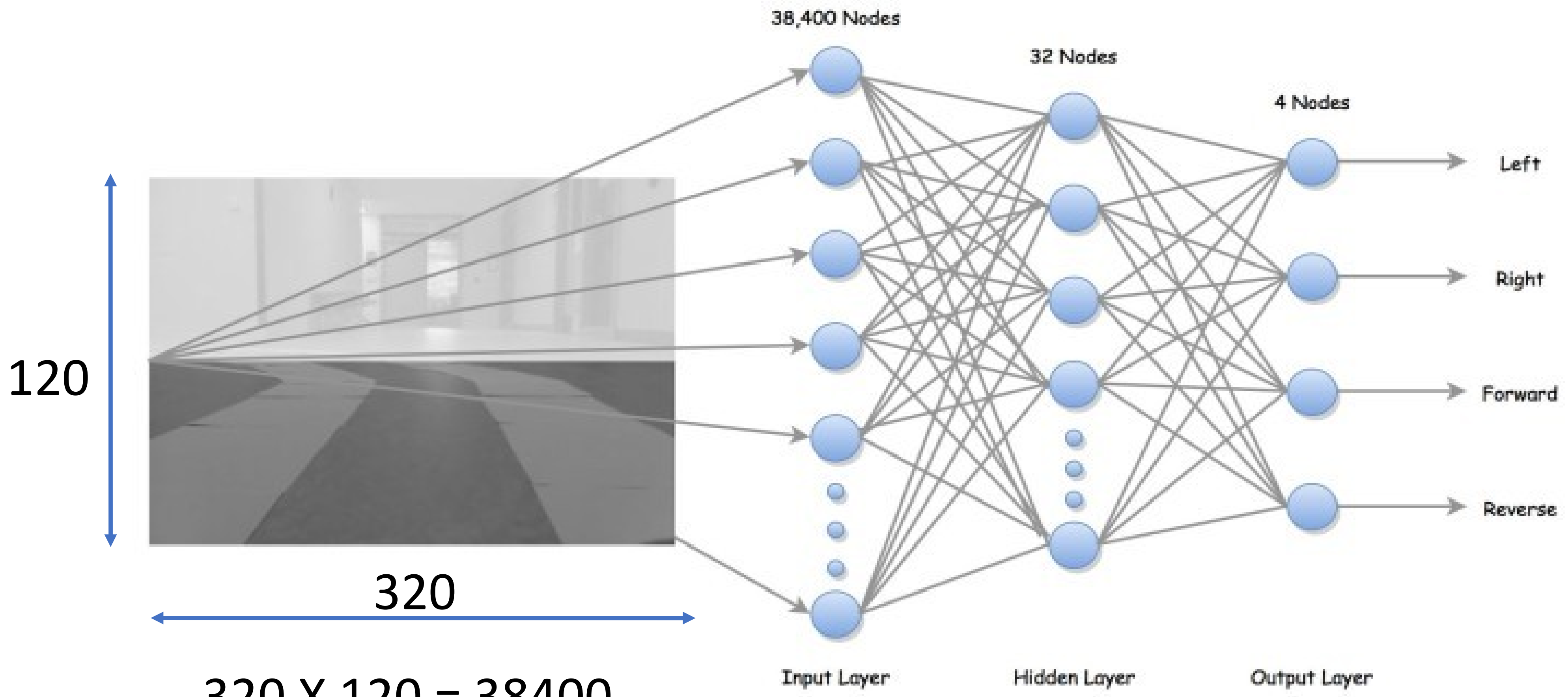
Classification: Prediction the Class/Category

Example: Image Classification – Identify Digits



Digit	Prob
0	5%
1	1%
2	1%
3	3%
4	1%
5	5%
6	1%
7	10%
8	15%
9	58%

Classification: Prediction the Class/Category
Self driving car can be framed as a Classification Problem



320 X 120 = 38400
Pixels/Numbers

Classification: Prediction the Class/Category

Self driving car can be framed as a Classification Problem

Source: https://zhengludwig.wordpress.com/projects/self-driving-rc-car/

Unsupervised Learning: Clustering
Example: Customer Segmentation

Age (Input)	Dollars Spent (Input)
25	100
20	85
35	200
40	175
55	75
60	80
…	…

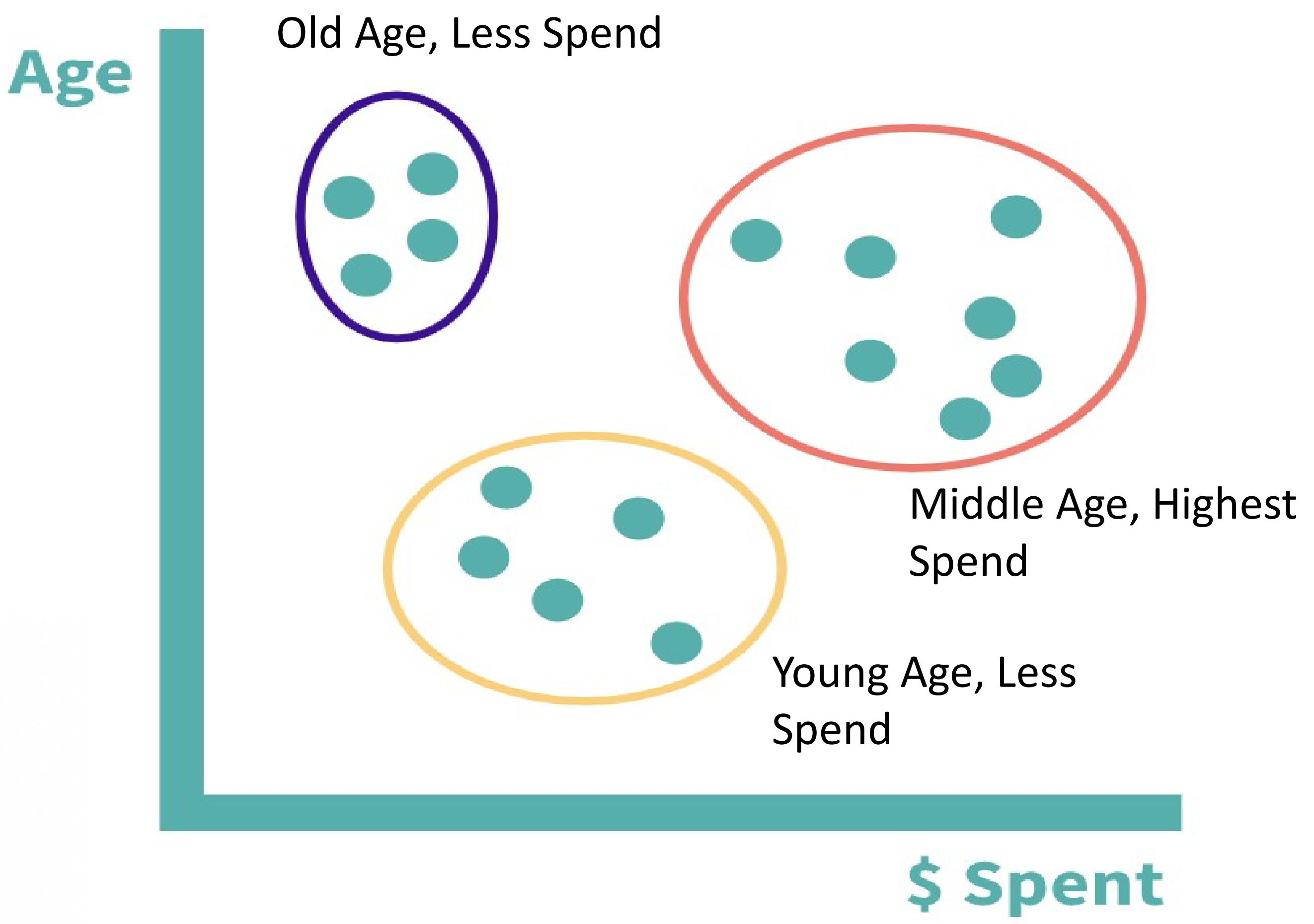
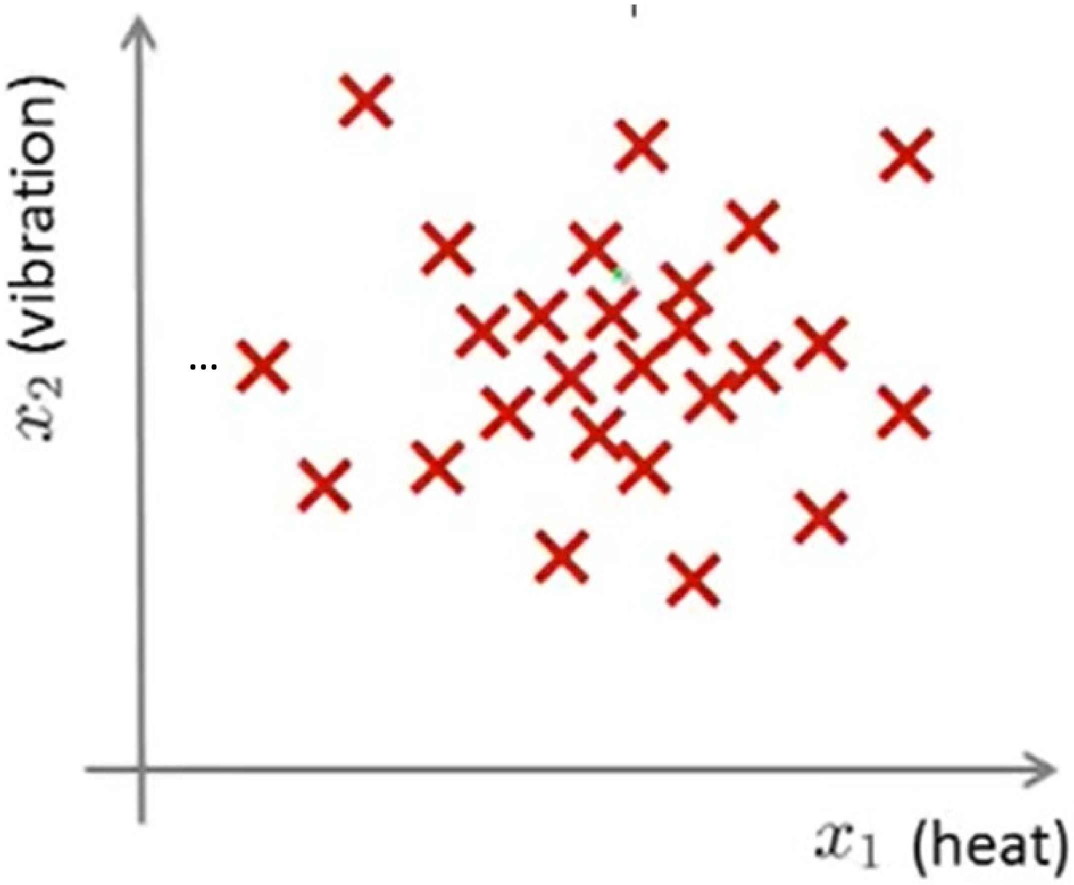


Image Source: Dataiku

Unsupervised Learning: Anomaly detection
Example: Fault Detection in Machine Parts

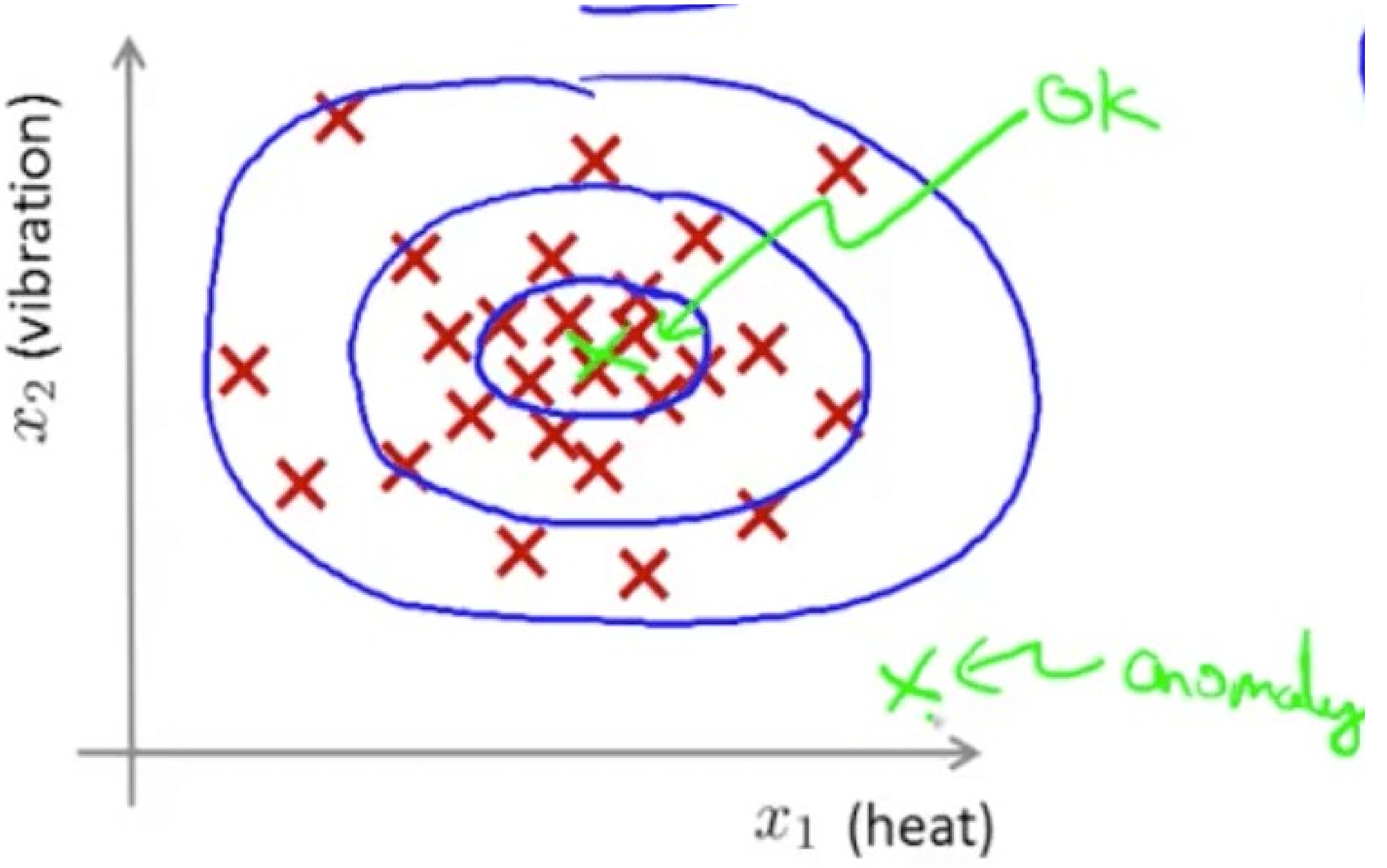
Heat (Input)	Vibration (Input)
…	…
…	…
…	…
…	…
…	…



Unsupervised Learning: Anomaly detection

Example: Fault Detection in Machine Parts

Heat (Input)	Vibration (Input)
…	…
…	…
…	…
…	…
…	…



History of Artificial Intelligence

AI as a field has existed since 1950s

Thinking Machine, MIT Documentary : 1961

AI: Story of Spring (Boom) and Winter (Bust)

Spring
1956 - 1974

- First attempts at machine translation, First neural networks written
- Major funding source: Defence
- ELIZA chatbot

1974 - 1980

- Many Applications fail to materialize. Research highlights limitations of existing algo
- Defence funding reduced dramatically

1980 – 1987

- Expert Systems become popular
- Market for LISP machines boomed
- Funding increased

1987 – 2000

- Limitations of Expert Systems become apparent
- Specialised hardware market (LISP Machines) collapsed
- Defence funding reduced again

2000 - Present

AI Spring: Research Breakthroughs, Increasing applications & Funding. But also Overpromise and Hype

The New York Times

NEW NAVY DEVICE LEARNS BY DOING; Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

Give this article

July 8, 1958

AI Winter: Reset of over inflated expectations, Interest & Funding Dries Up

"At its low point, some computer scientists and software engineers avoided the term artificial intelligence for fear of being viewed as wild-eyed dreamers."

- *New York Times Article :* ***2005***

A Brief History of AI

A BRIEF HISTORY OF AI: HOW TO PREVENT ANOTHER WINTER (A CRITICAL REVIEW)

Amirhosein Toosi
Department of Integrative Oncology
BC Cancer Research Institute
Vancouver, BC, V5Z 1L3
atoosi@bccrc.ca

Andrea Bottino
Department of Computer and Control Eng.
Polytechnic University of Turin
Turin, Italy 10129
andrea.bottino@polito.it

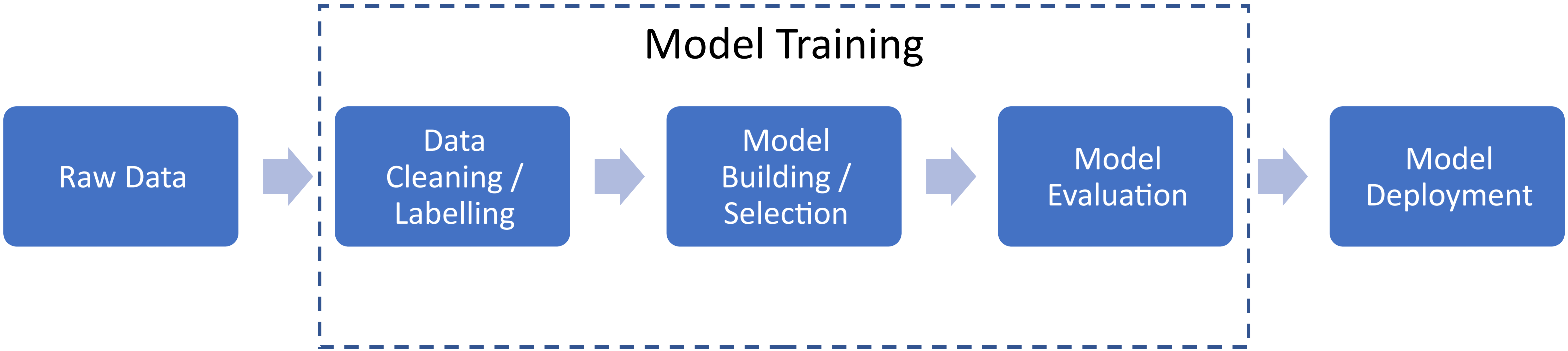
Babak Saboury
Department of Radiology and Imaging Sciences
National Institutes of Health
Bethesda, MD 20892
babak.saboury@nih.gov

Eliot Siegel
Department of Diagnostic Radiology and Nuclear Medicine
University of Maryland School of Medicine
Baltimore, MD 21201
esiegel@umaryland.edu

Arman Rahmim
Departments of Radiology and Physics
University of British Columbia
Vancouver, BC
arman.rahmim@ubc.ca

What is driving the current AI Spring?

Machine Learning Process
Greatly Simplified



- More Complex Algorithm → More Data & More Compute
- More Data → More storage & More Compute

What is driving the current AI spring?

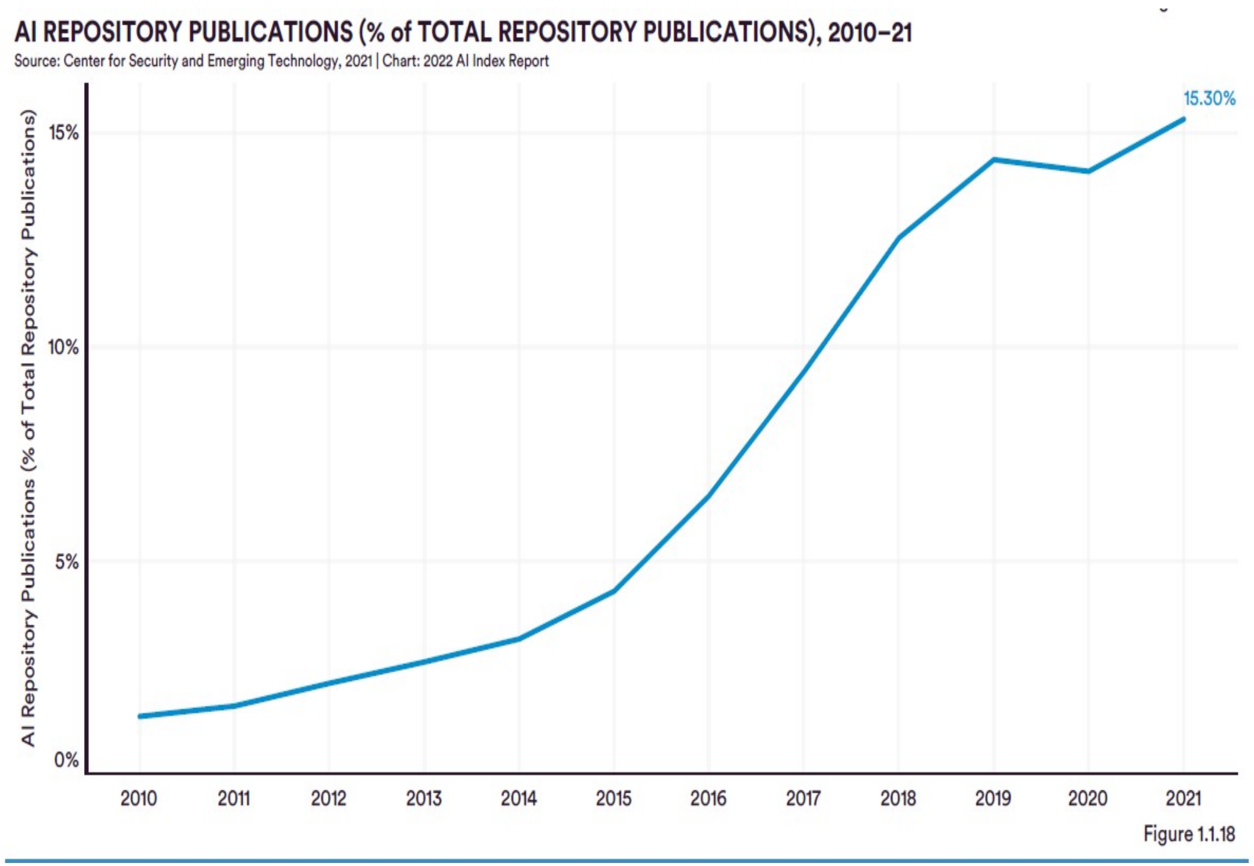
Research
Breakthrough
s

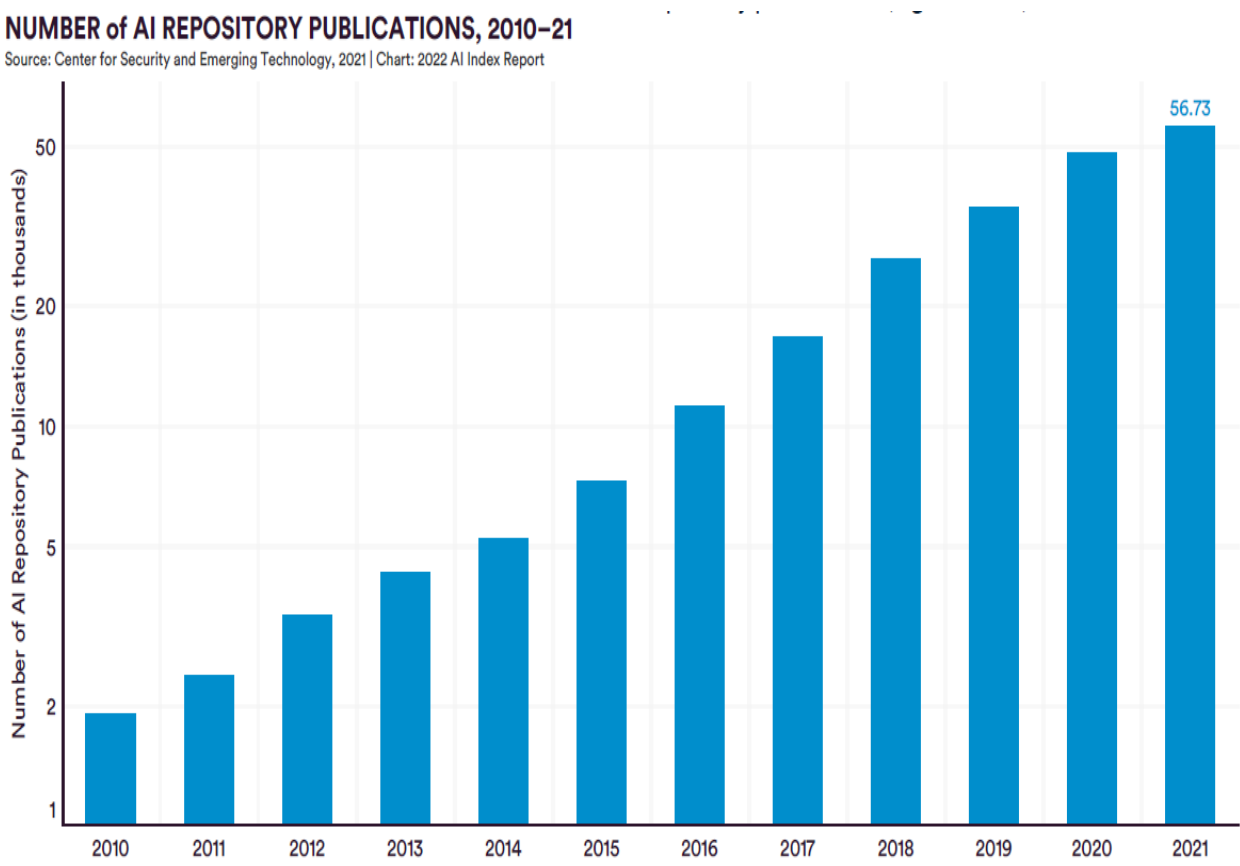
Cloud:
Accessible
computing
and Storage

Increasing
Data
Availability

Increasing
Computing
Power

Pace of Research in AI has been incredible





Source: State of AI 2021 Report

Increasing availability of Clean and Labelled Data

IMAGENET

14 Million Labelled Images

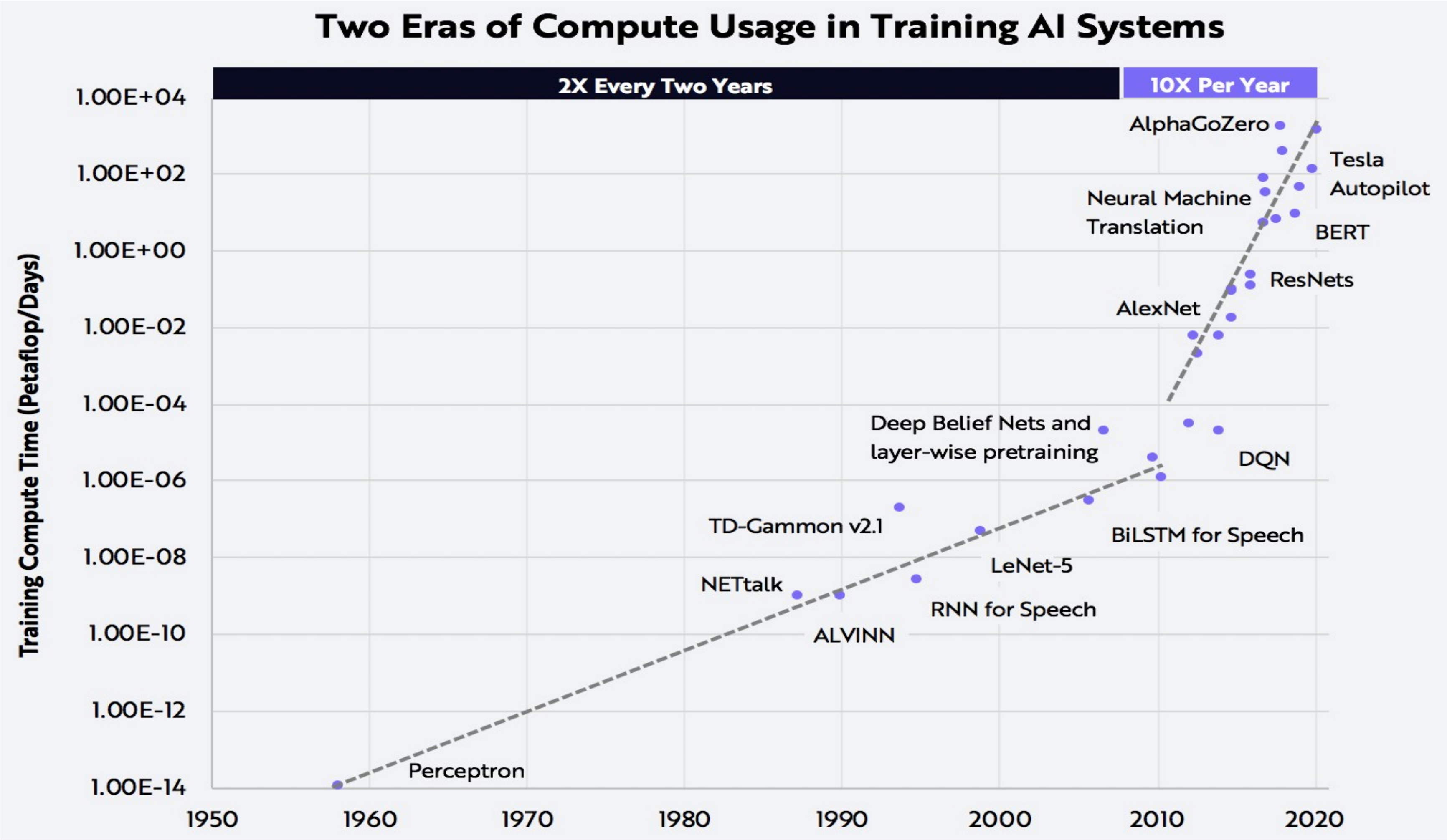
Google Open Images: 9M

Amazon Review Data:
233M Reviews with detailed Metadata

Malware Database

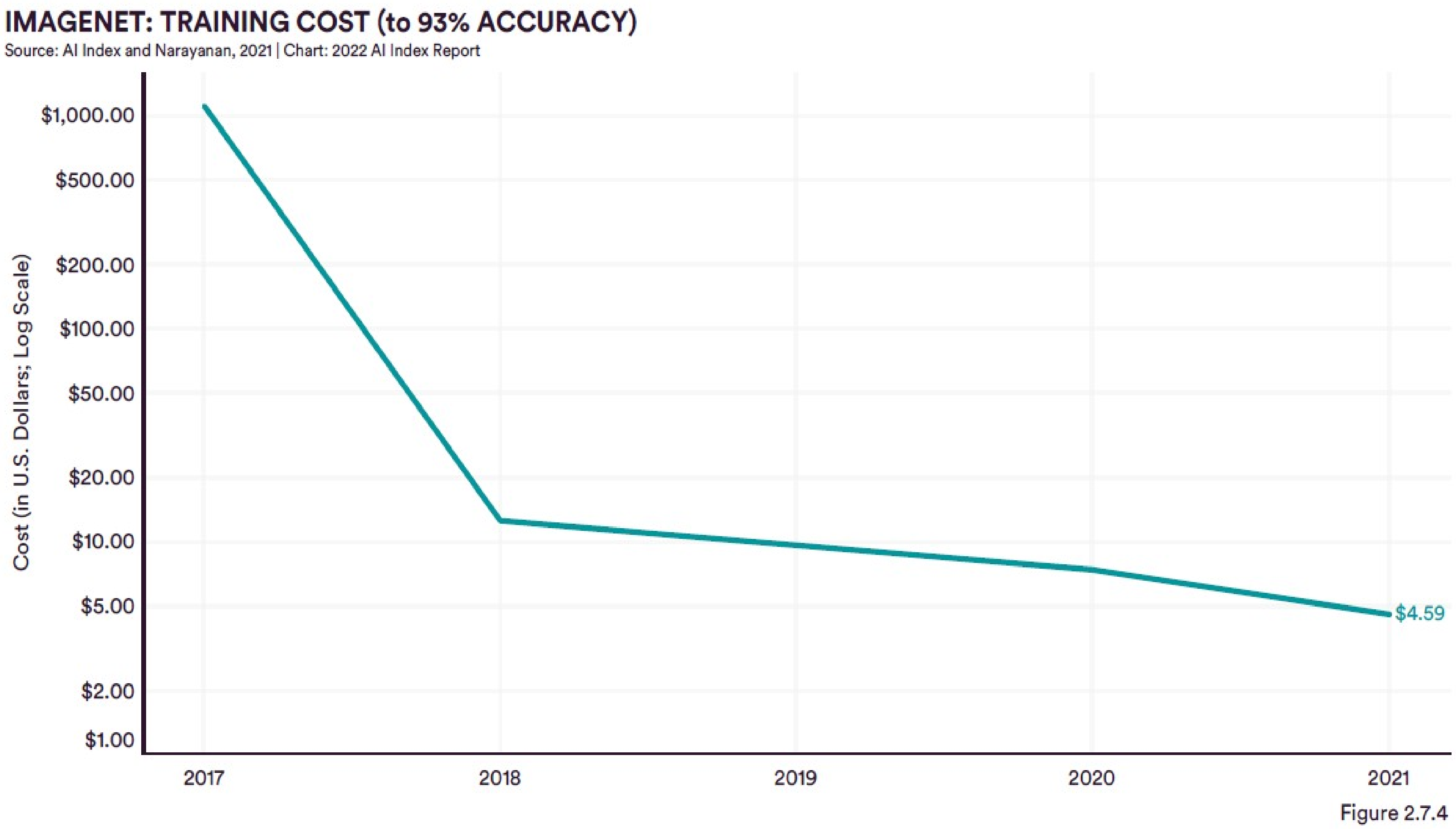
SOREL 20M

Increasing compute capability



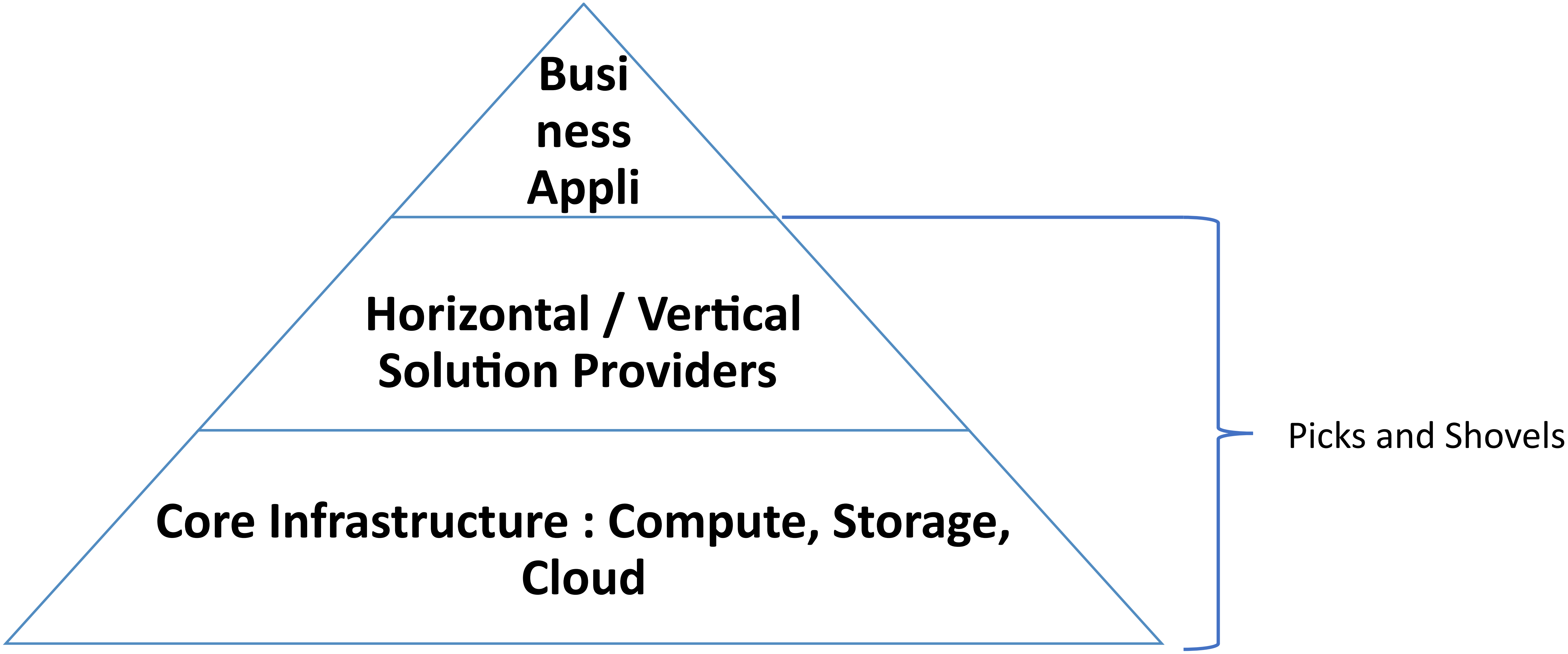
Source: OpenAI, ARK Invest

Like for like compute cost has reduced



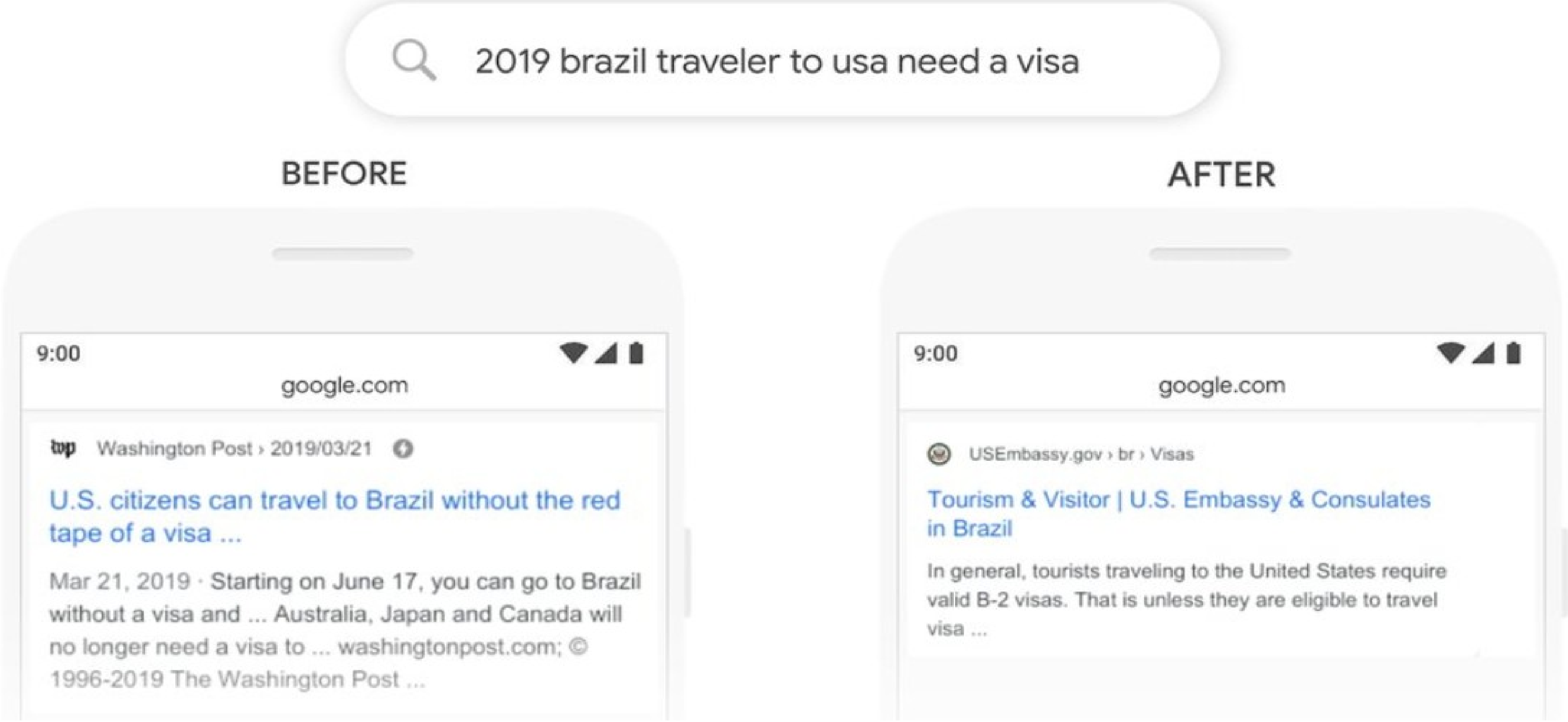
AI Index Report 2022

AI Ecosystem Pyramid



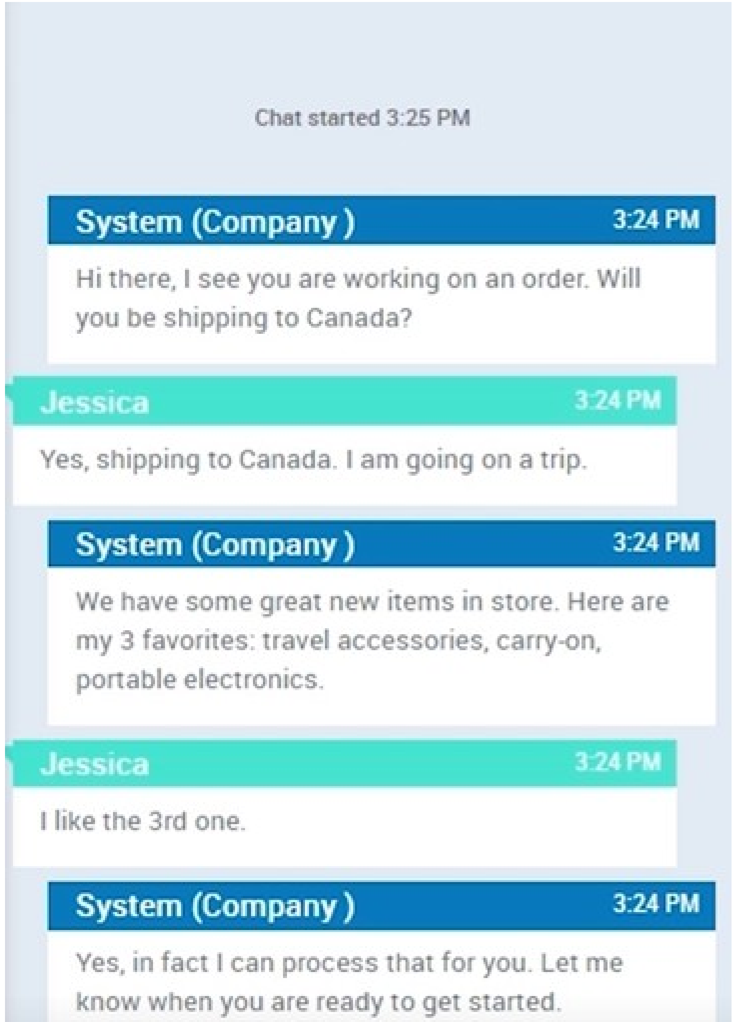
AI Ecosystem: AI / ML Applications & Challenges

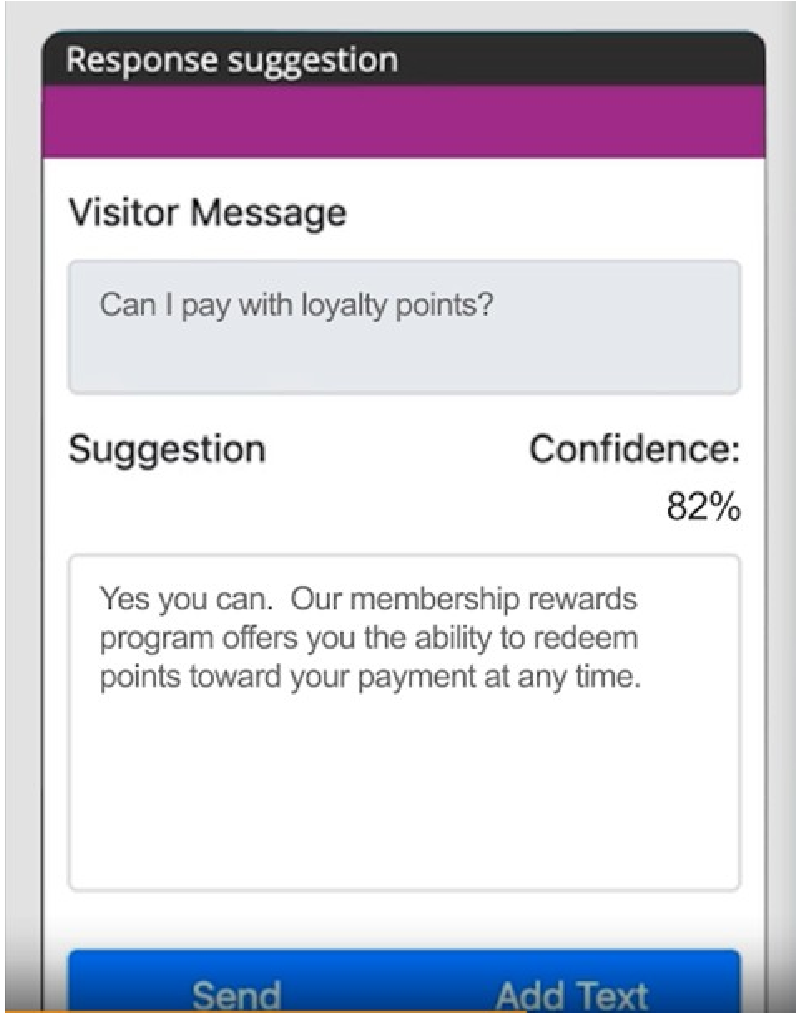
Natural Language Processing(NLP)
Search: Understanding the meaning of Text



Source: Google Techblog

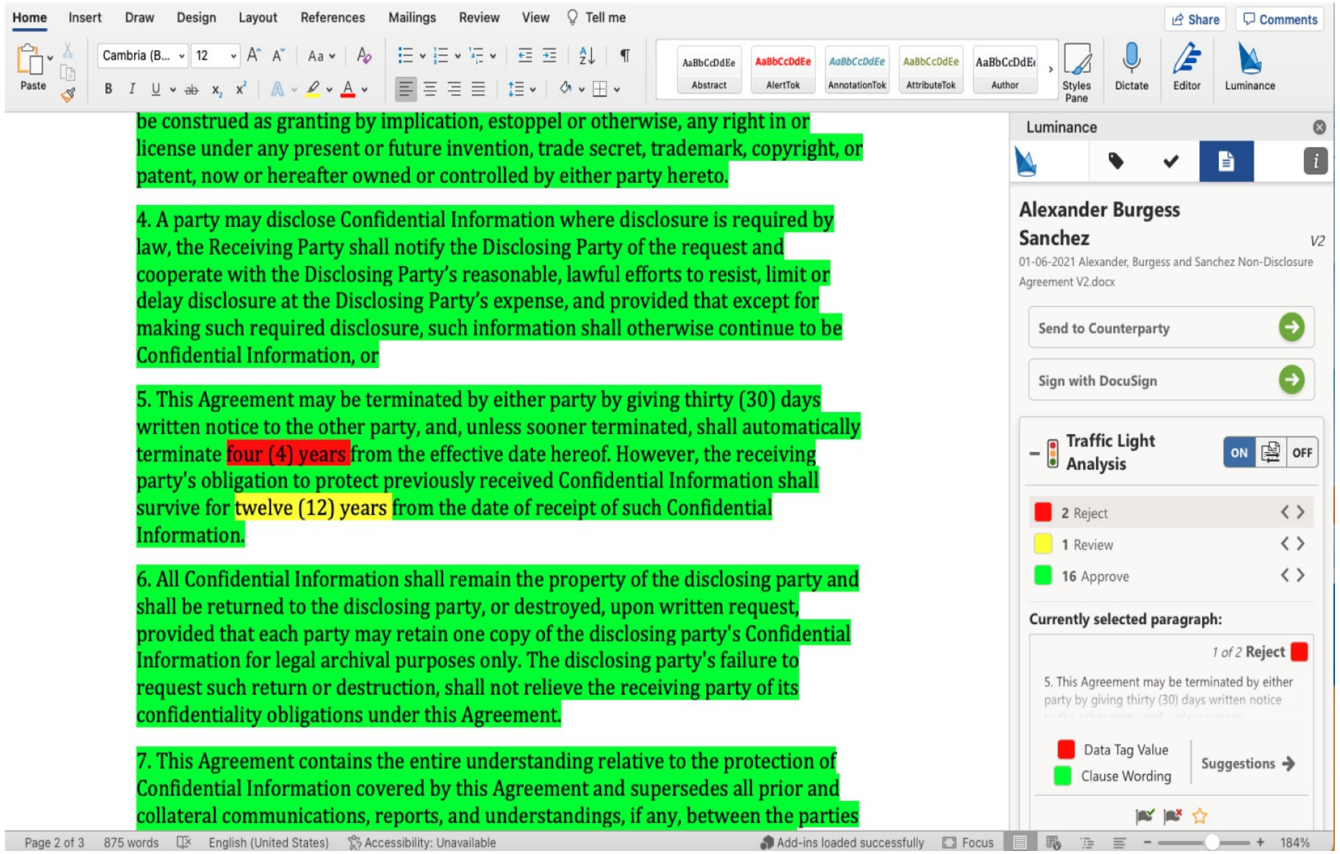
Natural Language Processing(NLP)
Chatbots





Source: NICE.AI

Natural Language Processing(NLP)
Legal Document Review



TRADITIONAL SAMPLING:

3 associates, 3 weeks, 10% review. **Cost: $88,800 (associate salary)**

REVIEW WITH LUMINANCE:

2 associates, 3 weeks, 10% review. **Cost: $22,200 (associate salary) + c.$32,000 (technology) = $54,200**

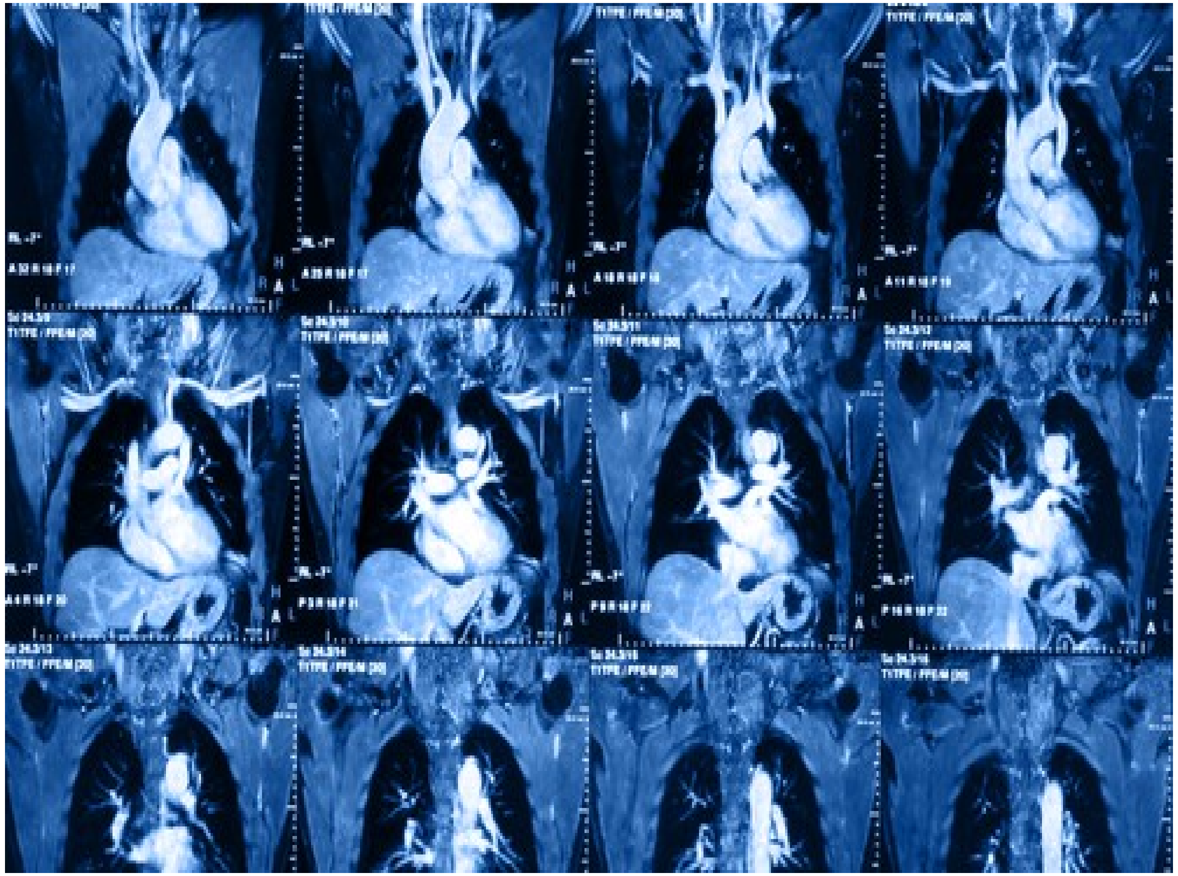
Cost saving for a sampled document review = 40%

Time saving for a complete document review = 18 associate weeks

Source: Luminance

Computer Vision
Image/Object Recognition

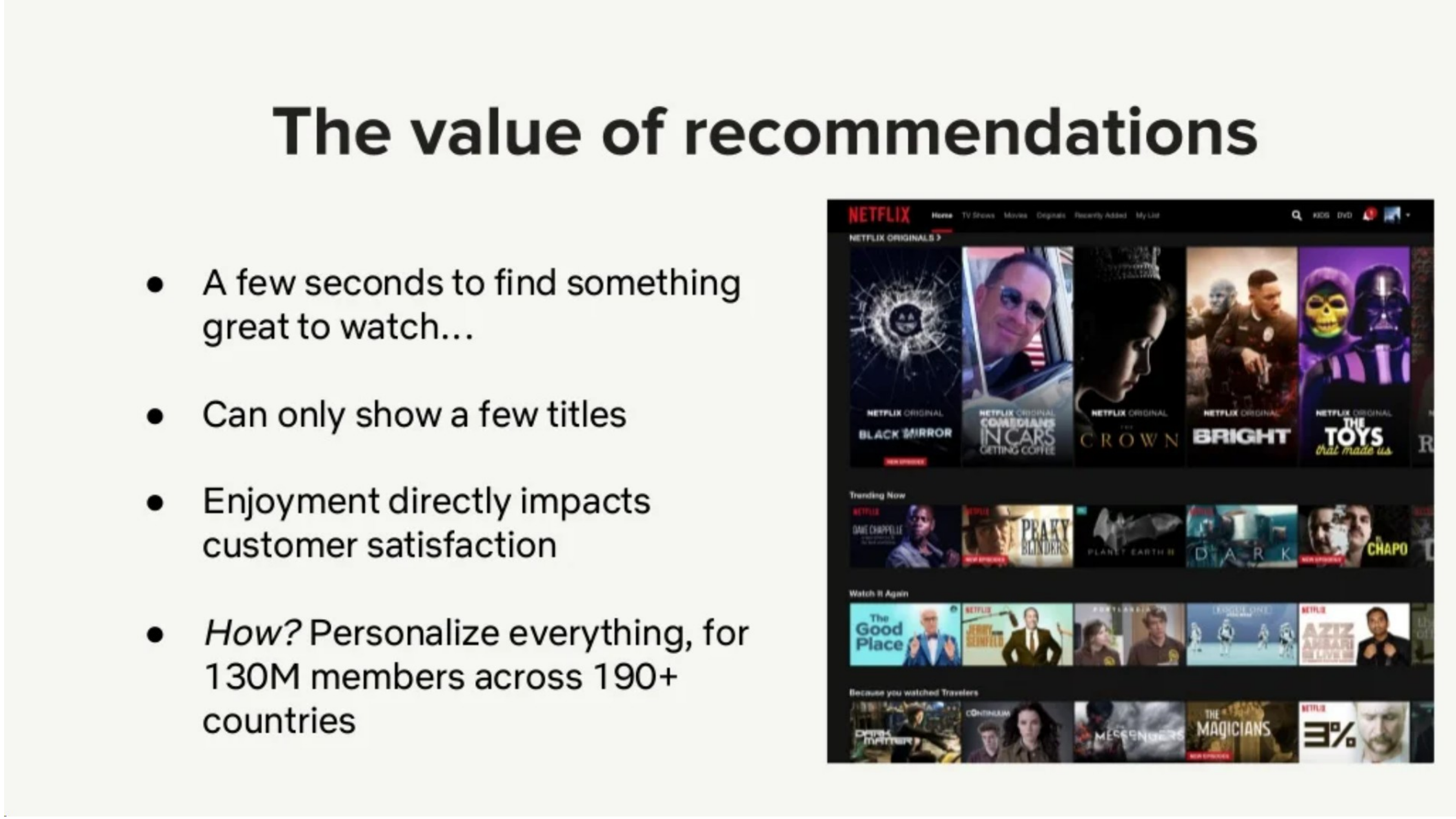
Nanox AI: Evaluating CT Scans



AES: Windmill Damage Detection

Source: Nanox AI, AES

Recommendation Systems
What will a customer like?



Source: Netflix Presentation

Recommendation Systems
What will a customer like?

WSJ | VIDEO

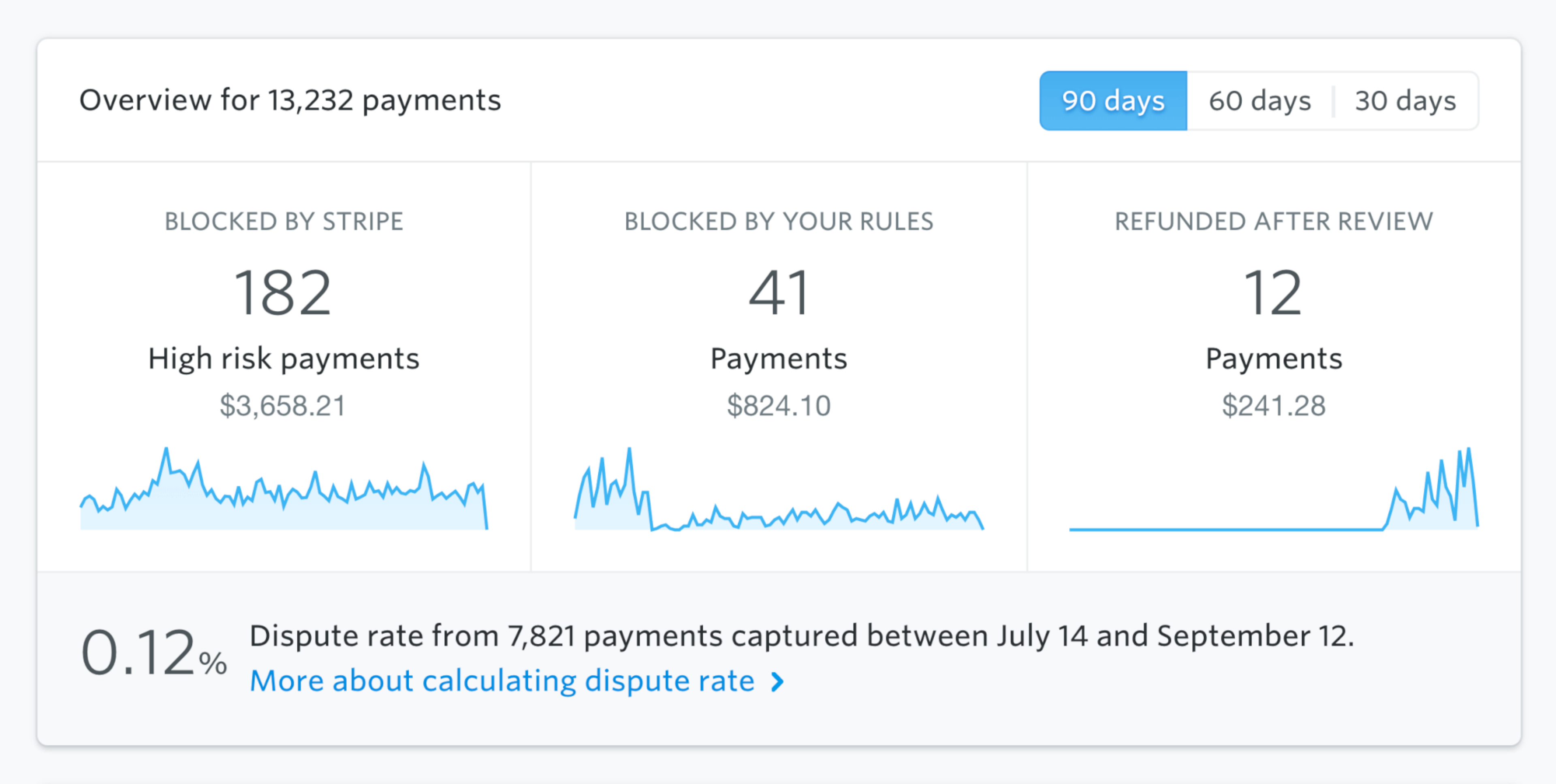
Investigation: How TikTok's Algorithm Figures Out Your Deepest Desires

The Wall Street Journal created dozens of automated accounts that watched hundreds of thousands of videos to reveal how the social network knows you so well

By Wall Street Journal

Jul 21, 2021 5:30 pm

Anomaly / Fraud Detection
Stripe: Radar



Analytics
Credit Quality Evaluation

Fundbox

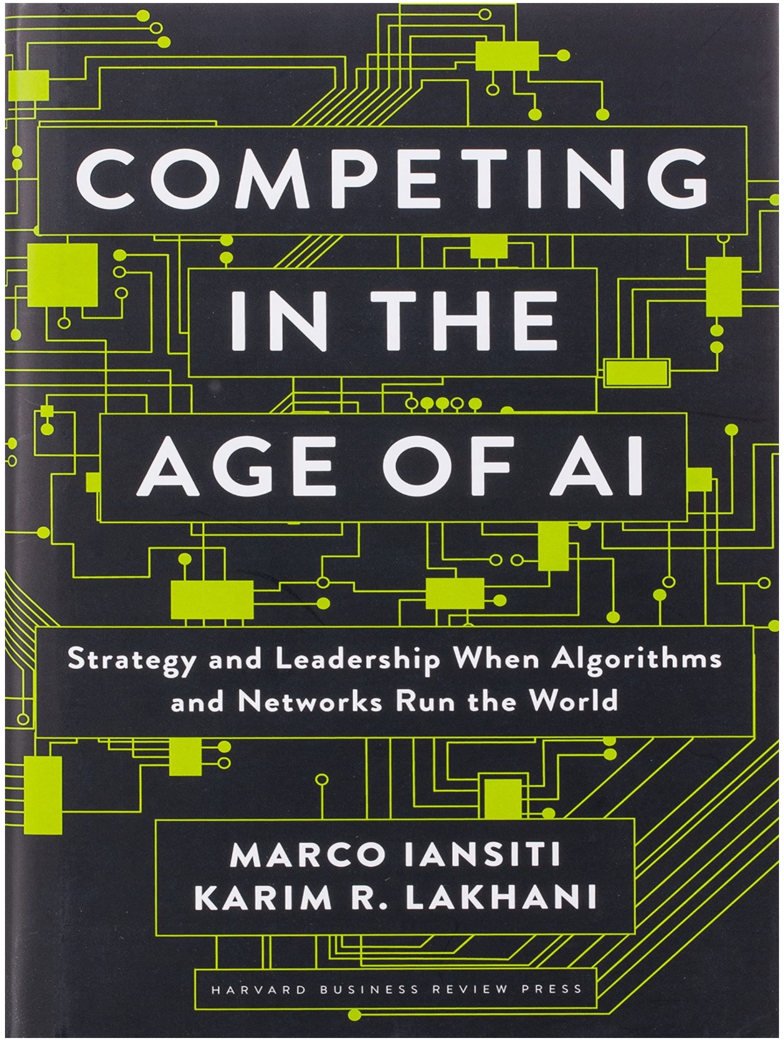
- When a small business works with Fundbox, it connects a “transactional system” i.e. **accounting software**, their **invoicing system**, or even its **bank account**.

- Fundbox immediately pulls business performance from this data source and our machine learning algorithms quickly assess business risk so that we can make a credit decision.

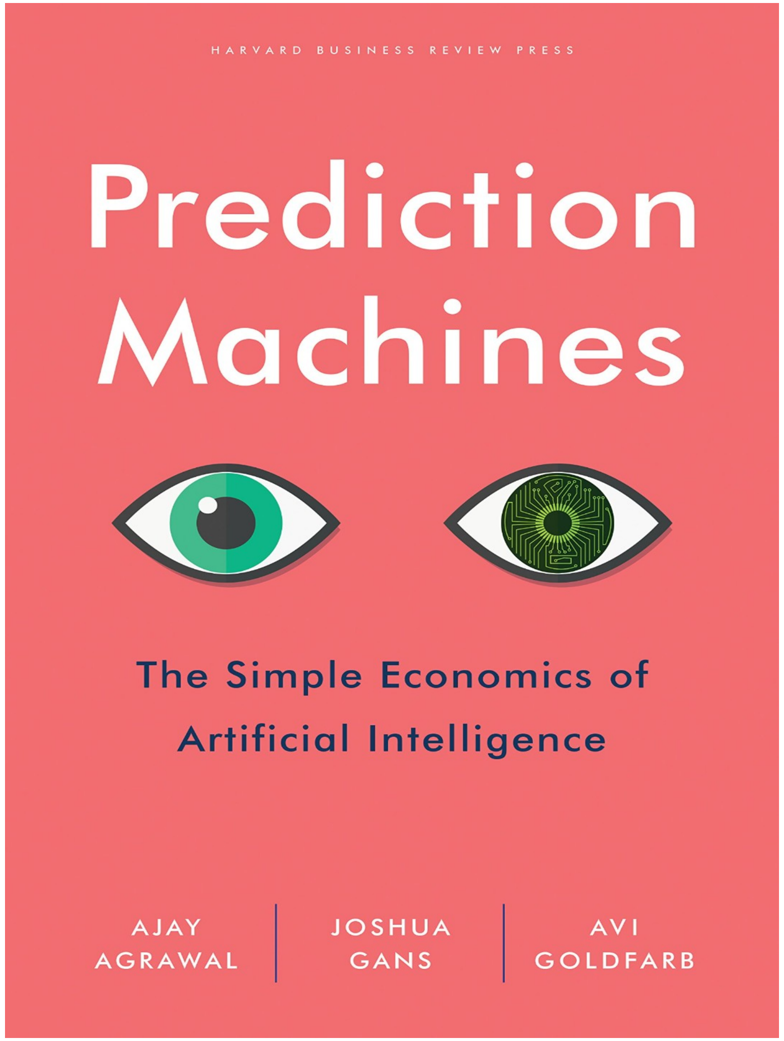
- The entire process typically takes a few minutes or less.

Source: Fundbox, Prashant Fuloria Interview

Examples on how AI is changing business



Broader Economic view of AI



- AI is making Predictions / Pattern Matching cheap
- Makes Predictions easily available in turn reduces uncertainty in some business activities
- Changing division of labor between man and machine

Challenges to AI/ML Application
60-80% of Projects don’t get deployed

ENTERPRISE & CLOUD

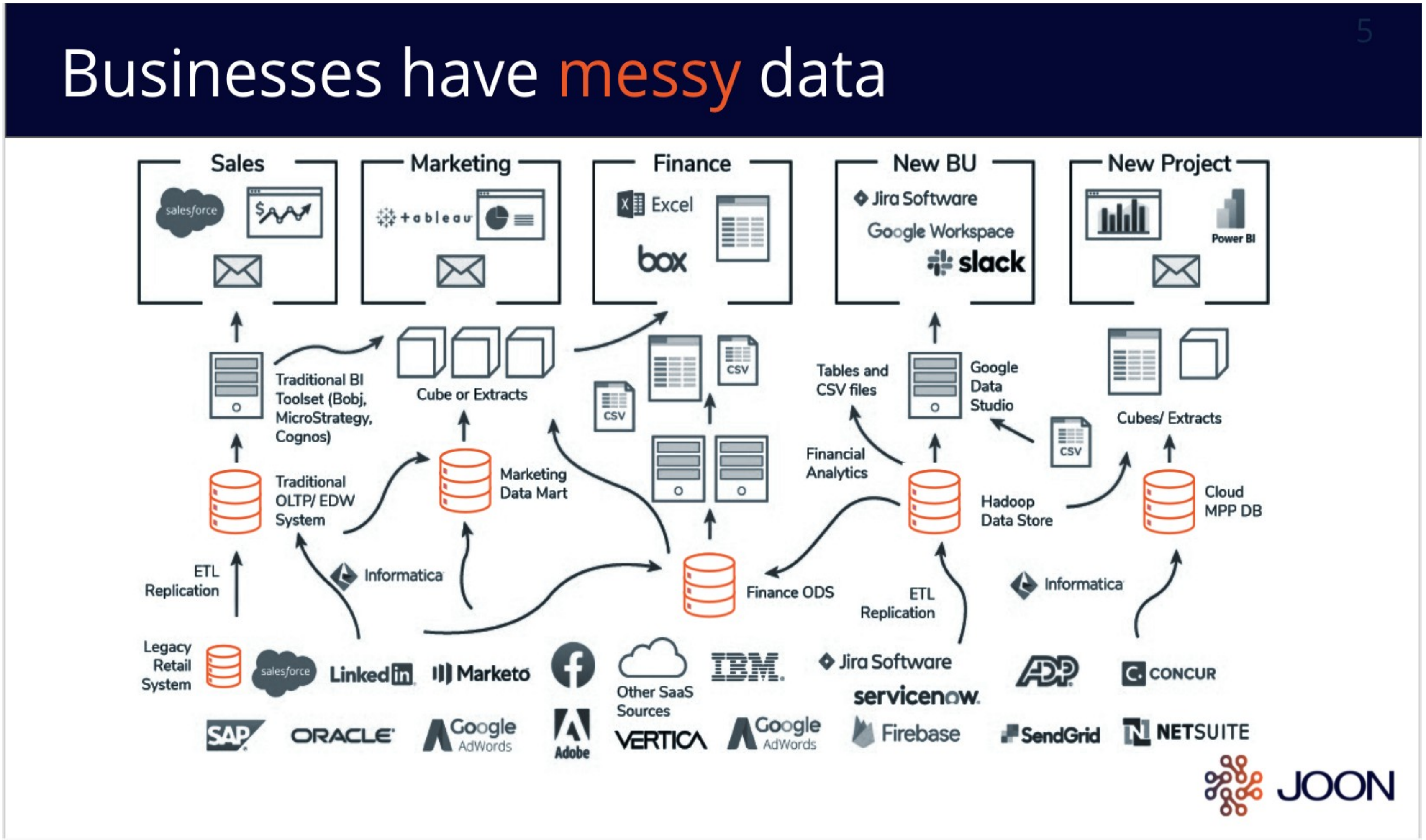
AI Stats News: Only 14.6% Of Firms Have Deployed AI Capabilities In Production

Gil Press Senior Contributor

I write about technology, entrepreneurs and innovation.

Challenges to AI/ML Application

Data Challenge



- Data Silos
- Multiple Data stores
- Storing structured and unstructured data
- Data keeps changing
- Little to no control on external Data Sources

Challenges to AI/ML Application
Data Challenge

Atlanta Firebird System

- Data joining a difficult problem.
- Data collected from 12 datasets
- Spatial information had different formats, sometimes contain minor differences such as different spellings.
- Turned out to be a massive effort that consumed a lot of time.

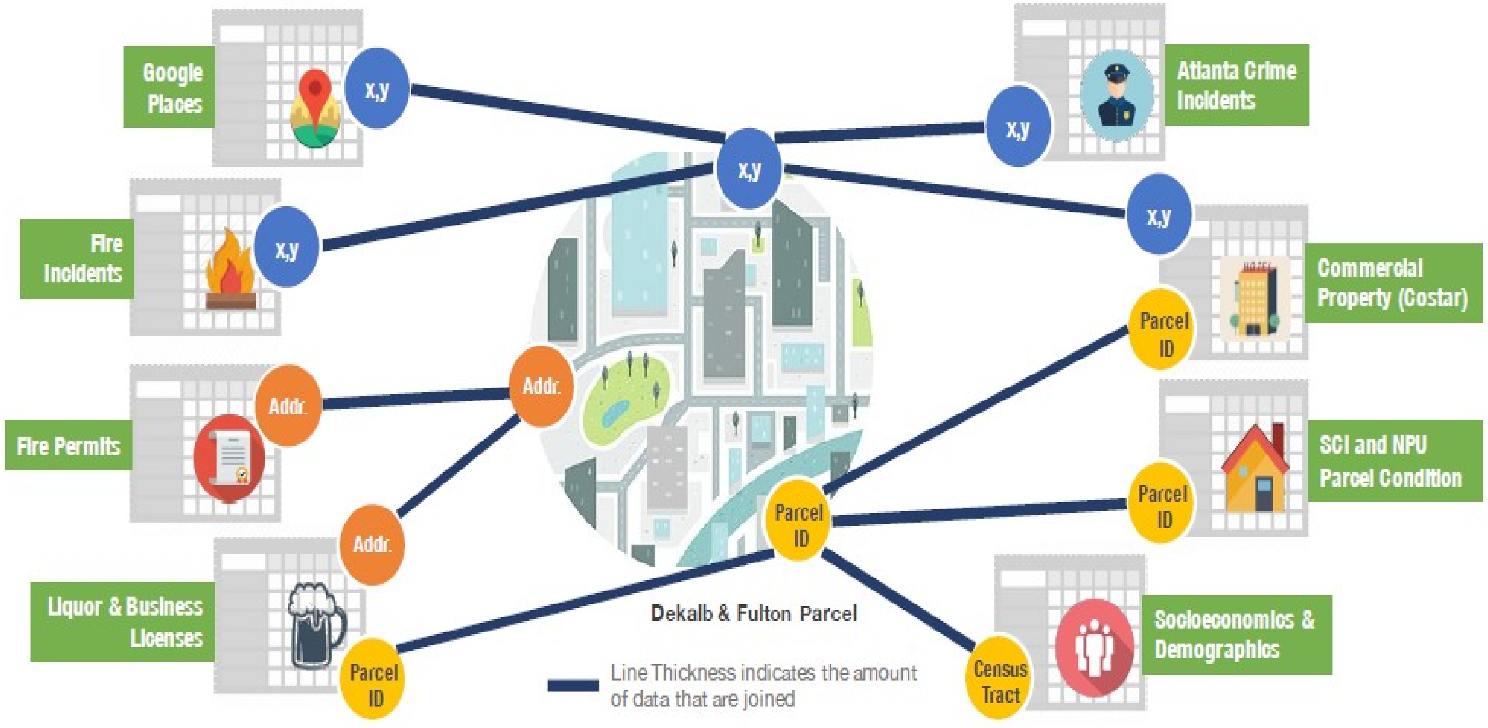
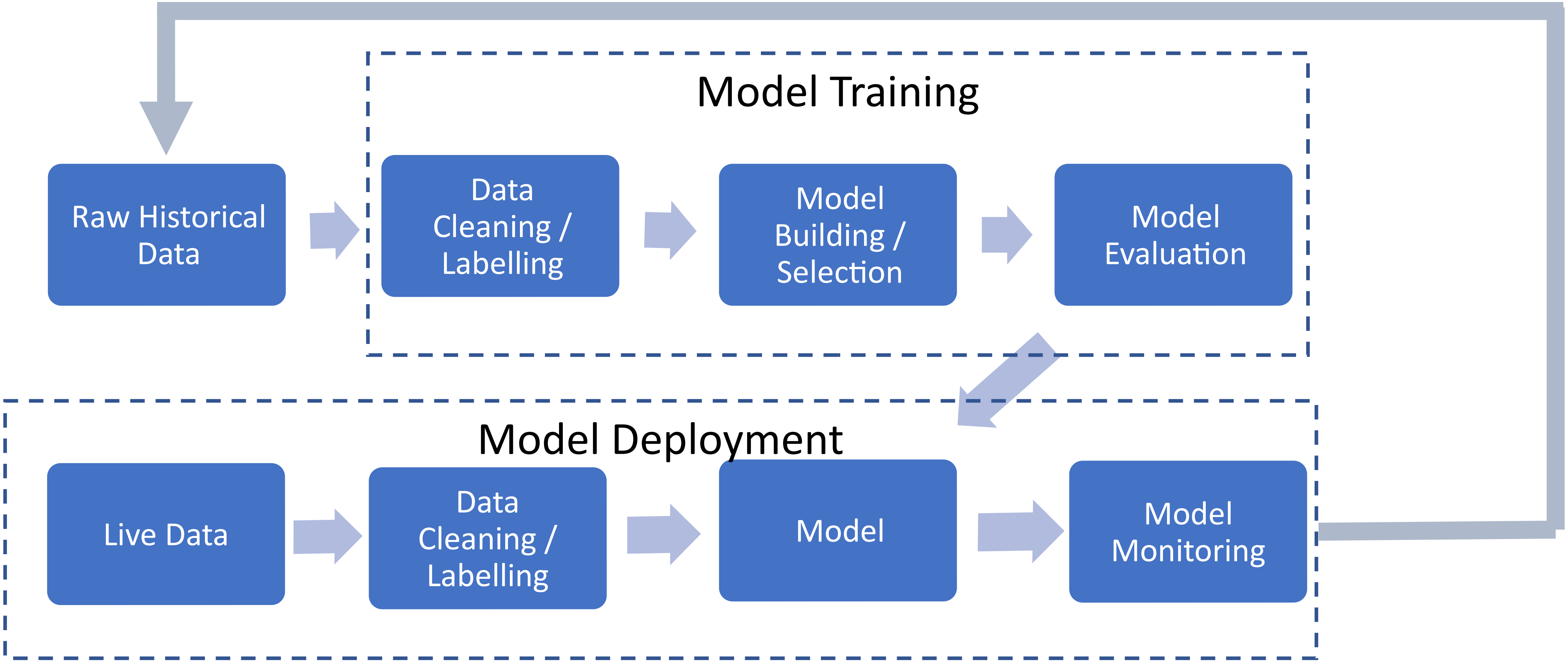


Figure 2: Joining eight datasets using three spatial information types (geocode, address, parcel ID).

Source: Firebird – Prediction Fire Risk and Prioritizing Fire Inspections in Atlanta

MLOps Challenge
Machine Learning Process Elaborated



MLOps Challenge

AI/ML needs lot of supporting Human and Tech infrastructure

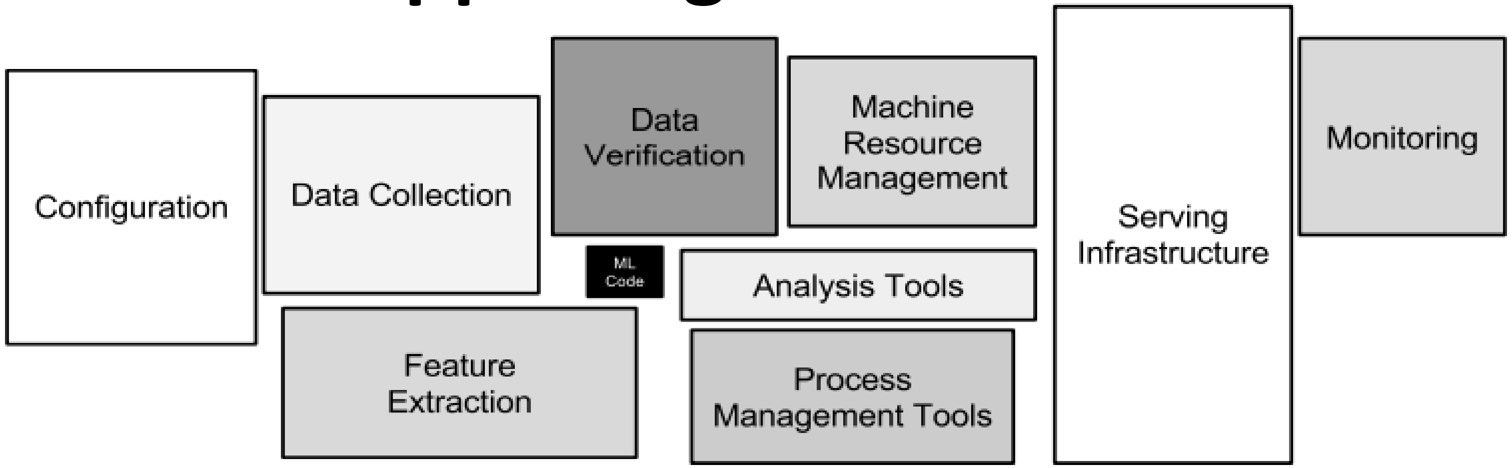


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

- Deployed Model should have minimal manual intervention. Needs software engineering
- Data Scientists are (mostly) not Software Engineers
- Deployed model needs to monitored for performance deterioration
- Tools used by Data Scientists for Model Training may not work for Model Deployment
- Other considerations: Data Governance and Privacy

Source: Google - Hidden Technical Debt in Machine Learning Systems

Data and MLOps as Competitive Advantage

“The Looper platform currently hosts 700 AI models and generates 4 million of AI outputs per second.”
- Meta (Facebook)

“Once we figure out where all the data is, we assemble data catalogs for all the different data sources. We take the data and mash it up into lakes, so we can build ML models.”
“We can now build AI and ML models on top of everything. We can search across the entirety of our data sets and do analysis on them. This is a critical departure from the previous operating model in IT where many apps and services were siloed with little sharing, and many versions of similar capabilities.”
- Kurt DelBene and Ludo Hauduc, Microsoft

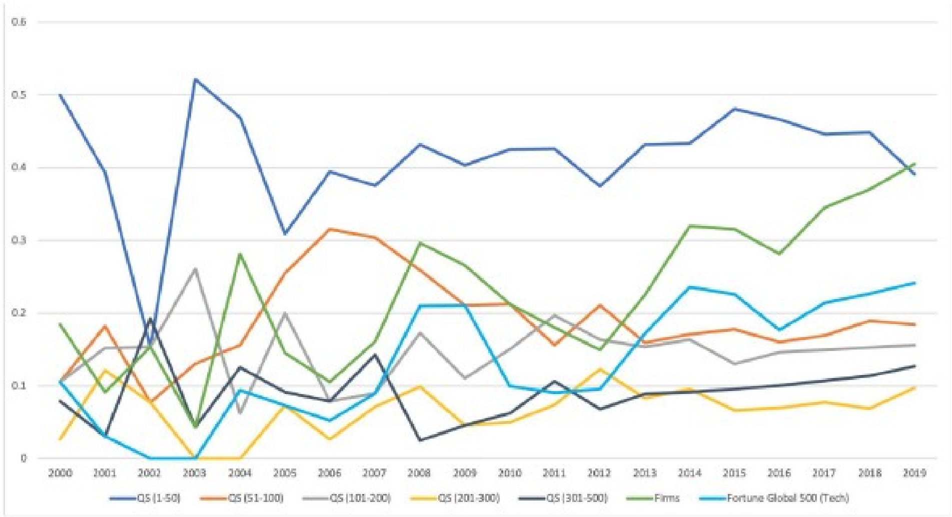
Source: Meta Techblog, Competing in Age of AI

Data and MLOps as Competitive Advantage

Elites work with elites: a compute divide drives the “de-democratization” of AI research

- Since 2012, large technology companies have increasingly published either on their own or in collaboration primarily with elite universities as opposed to mid-tier and lower-tier universities. Counterfactual analysis suggests a causal divergence between large technology companies and non-elite universities that is driven by access to computing power as a form of de-democratisation. This results in a small set of actors creating a majority of the high-impact research output.

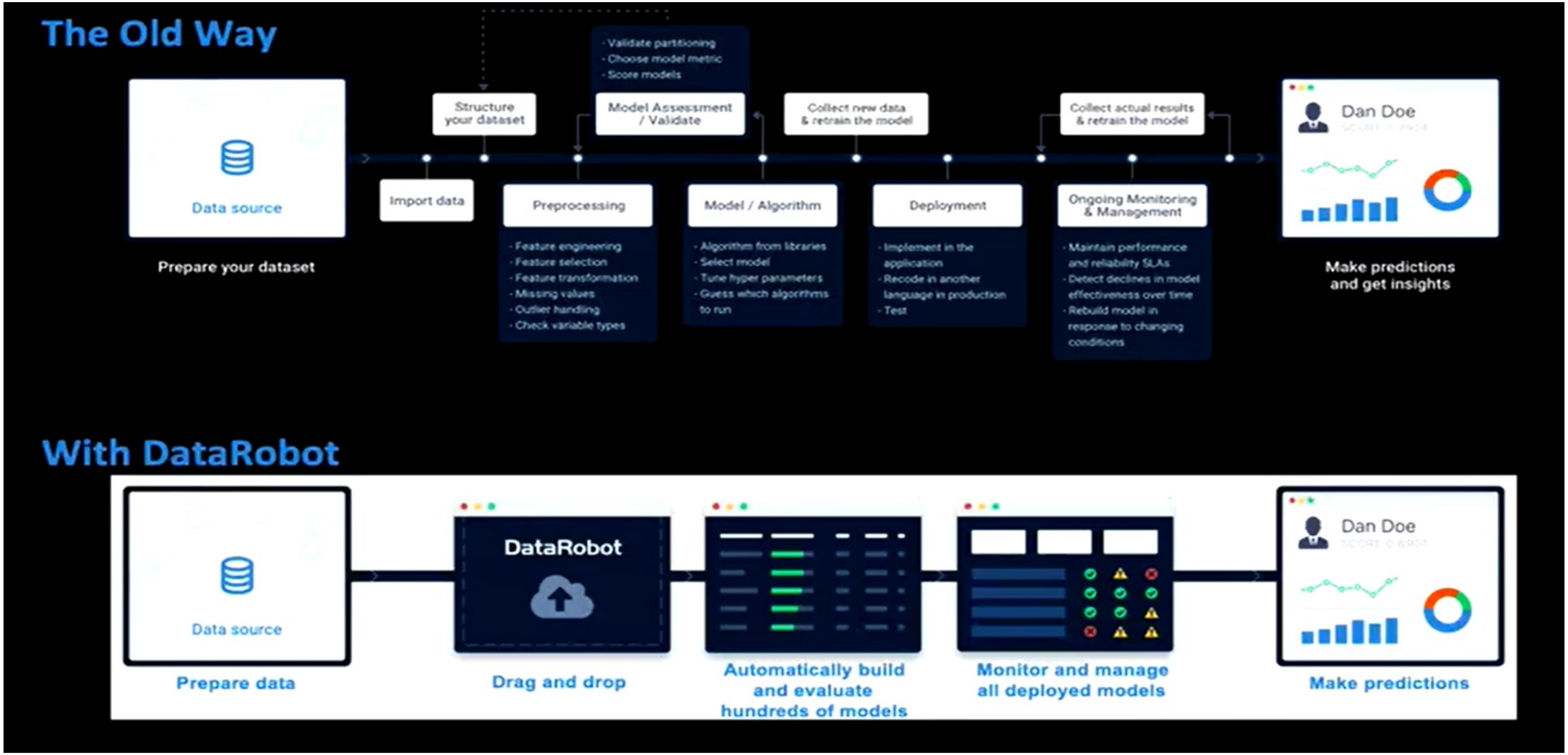
Figure 7: Share of papers within deep learning research



Note: This figure illustrates the share of papers that have at least one co-author from that specific group (e.g., firms, universities) within the deep learning papers.

AI Ecosystem: Picks and Shovels

Machine Learning as a Service (MLaaS)
Making ML accessible



Source: DataRobot

ML as a Service (MLaaS)

Accessible, Lower Cost and Reduced Time to Market

DataRobot

Val : $6 .3 Bn
Revenue: $210 Mn

H2O.ai

Val : $1 .7 Bn
Revenue: $45 Mn

aws machine learning

dataiku

Val : $4 .6 Bn
Revenue: $150 Mn

C3.ai

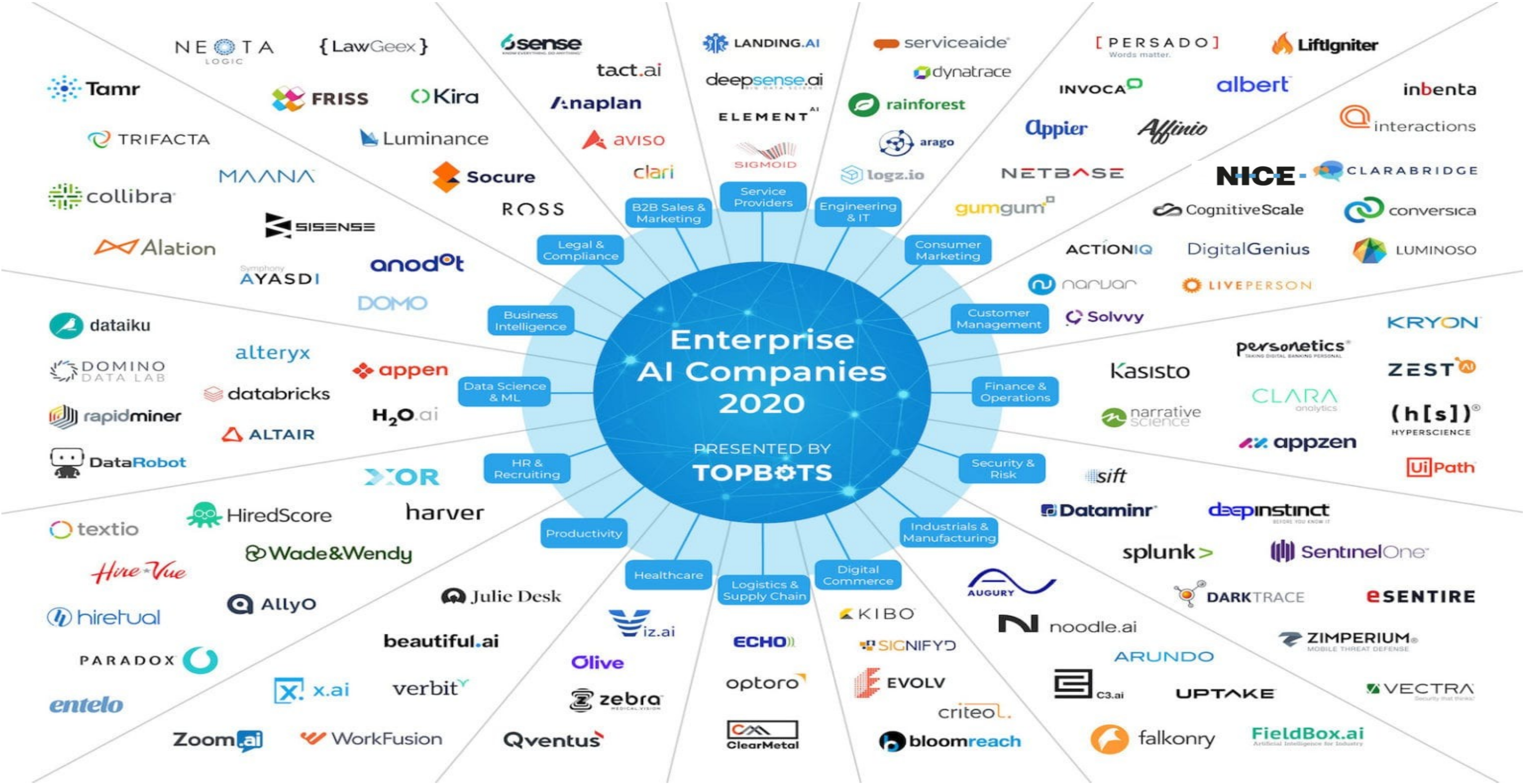
Val : $1 .6 Bn
Revenue: $232 Mn

Azure Machine Learning

vertex.ai

Watson Machine Learning

ML based Solution Providers



Data Providers

TGS

Bloomberg

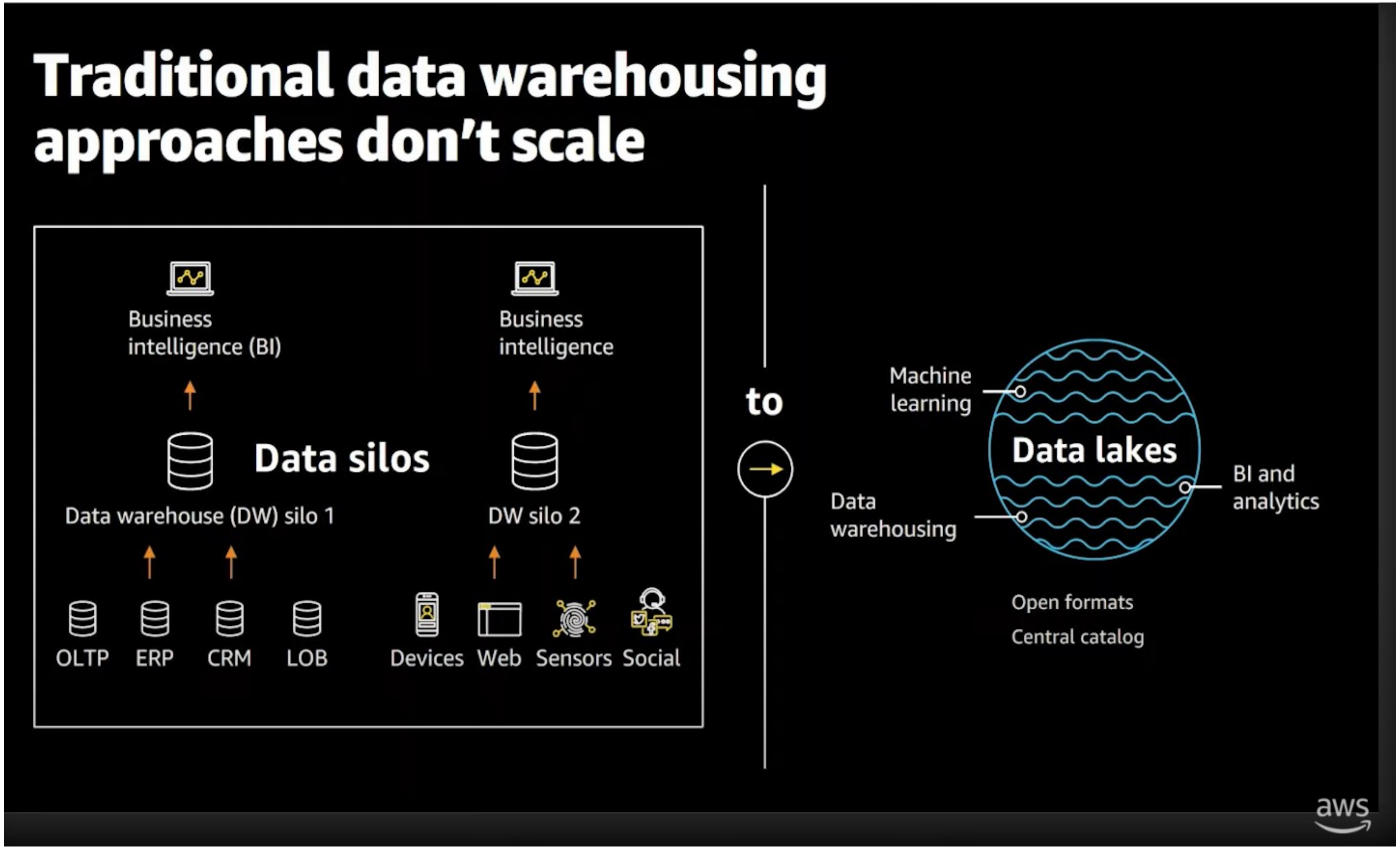
S&P Global

IQVIA

FACTSET

Datawarehouse and DataLakes

Storing and Processing Massive Data on the Cloud



- Data Lake: One central Database which support all data formats
- Cloud Based : No specialised hardware or software. Much less maintenance headache
- Separate Cost for Compute and Storage

Source: Amazon AWS Presentation

Datawarehouse and DataLakes
Storing and Processing Massive Data on the Cloud

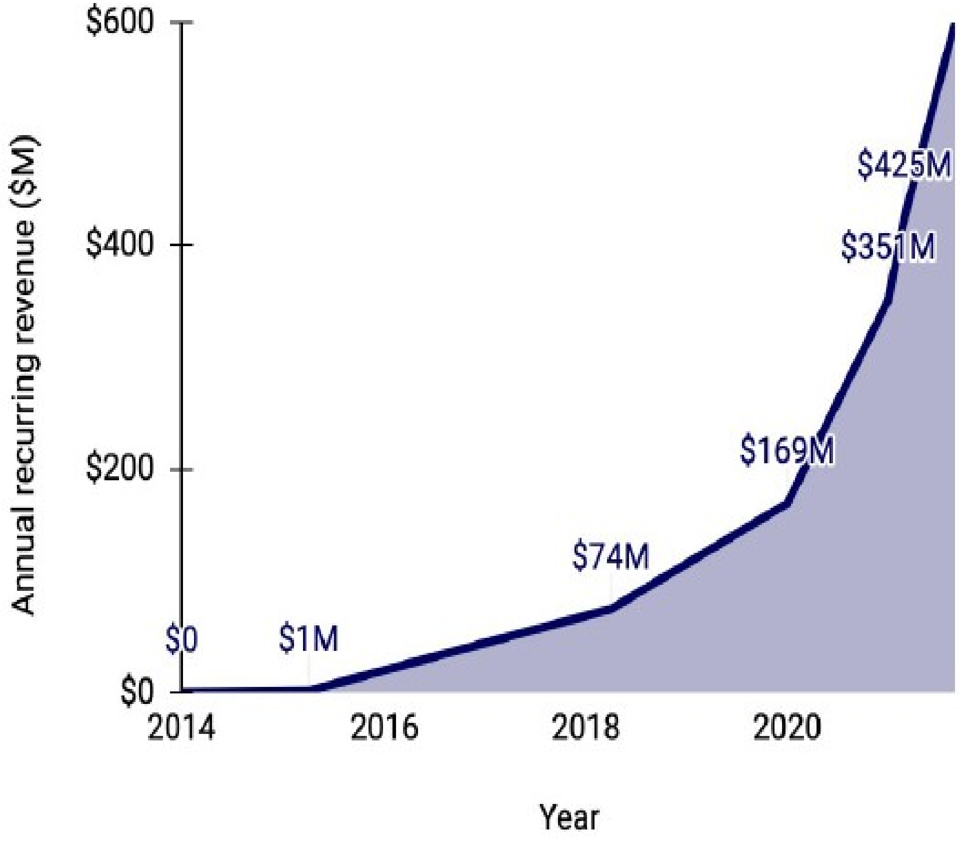
snowflake

databrick

amazon REDSHIFT

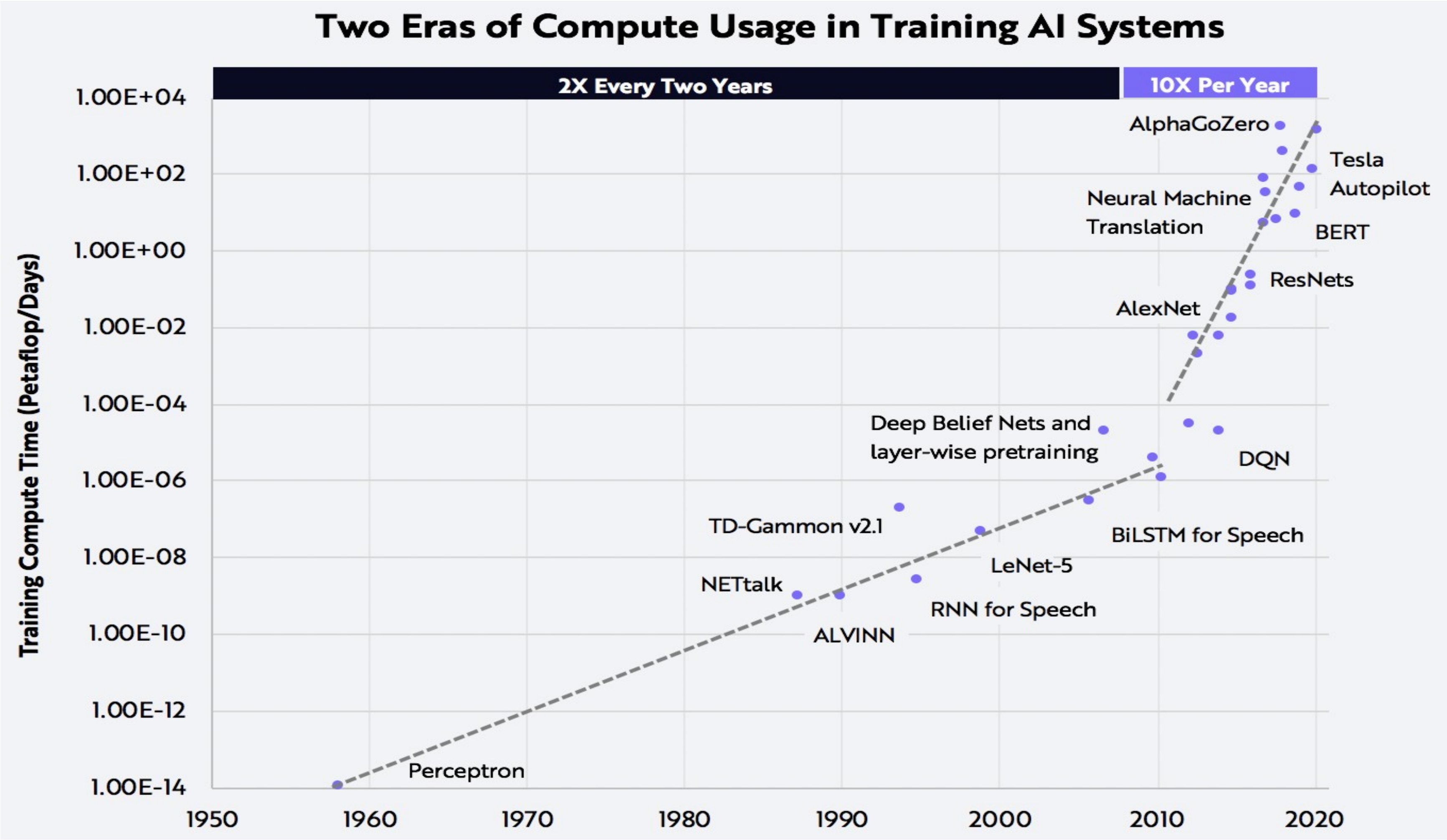
Azure Synapse Analytics

Google BigQuery



databricks

Providing Compute Power



Providing Compute Power
Rise of GPU computing

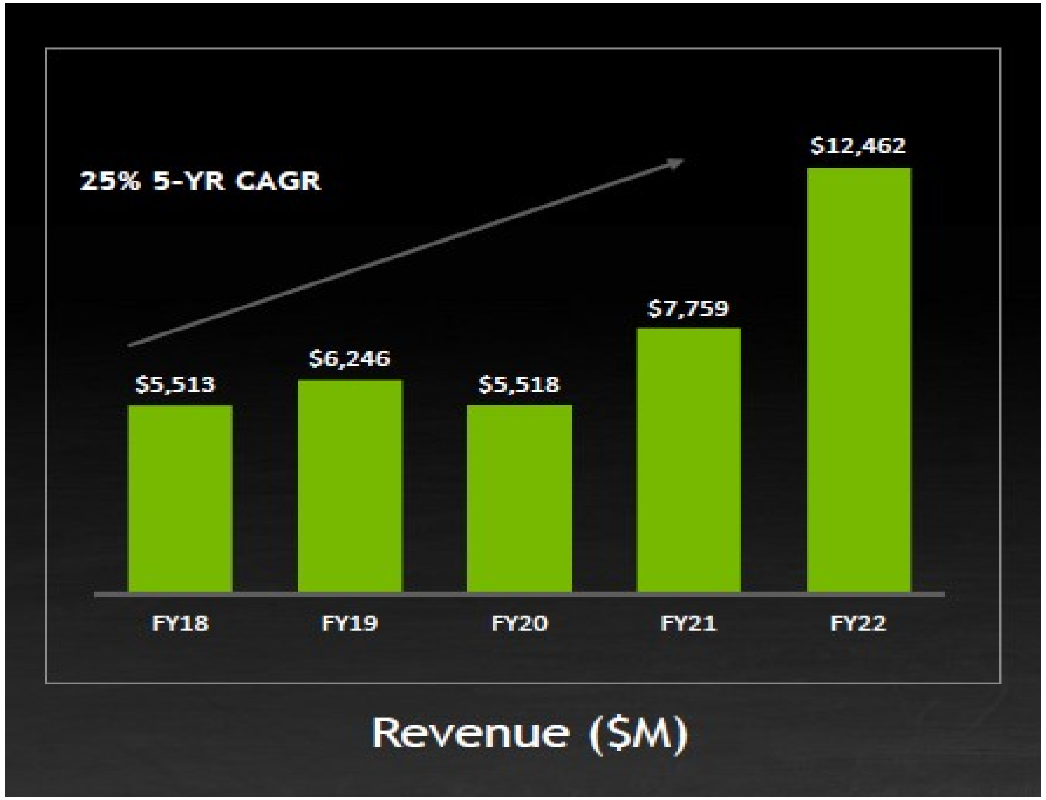
"GPUs are so incredibly powerful. Programs that previously ran on supercomputers, we're now realizing we can rewrite to run on GPUs at a fraction of the price."

- David Anderson, Computer scientist at Berkeley

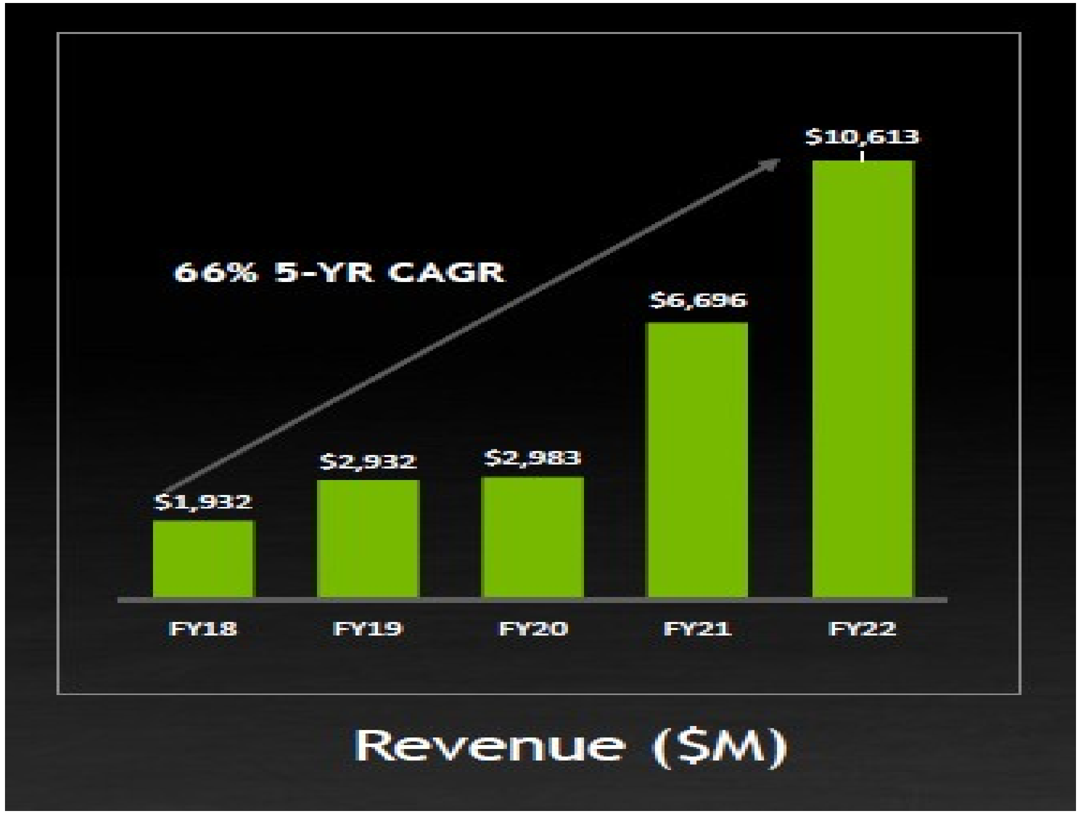
Providing Compute Power

Perspective on how ML has revolutionised GPU demand

Nvidia: Gaming Revenues



Nvidia: Data Centre Revenues



Source: Nvidia Investor Presentation

Providing Compute Power
Customised Chips for AI/ML

MARVELL™

GRAPHCORE

BROADCOM®

cerebras

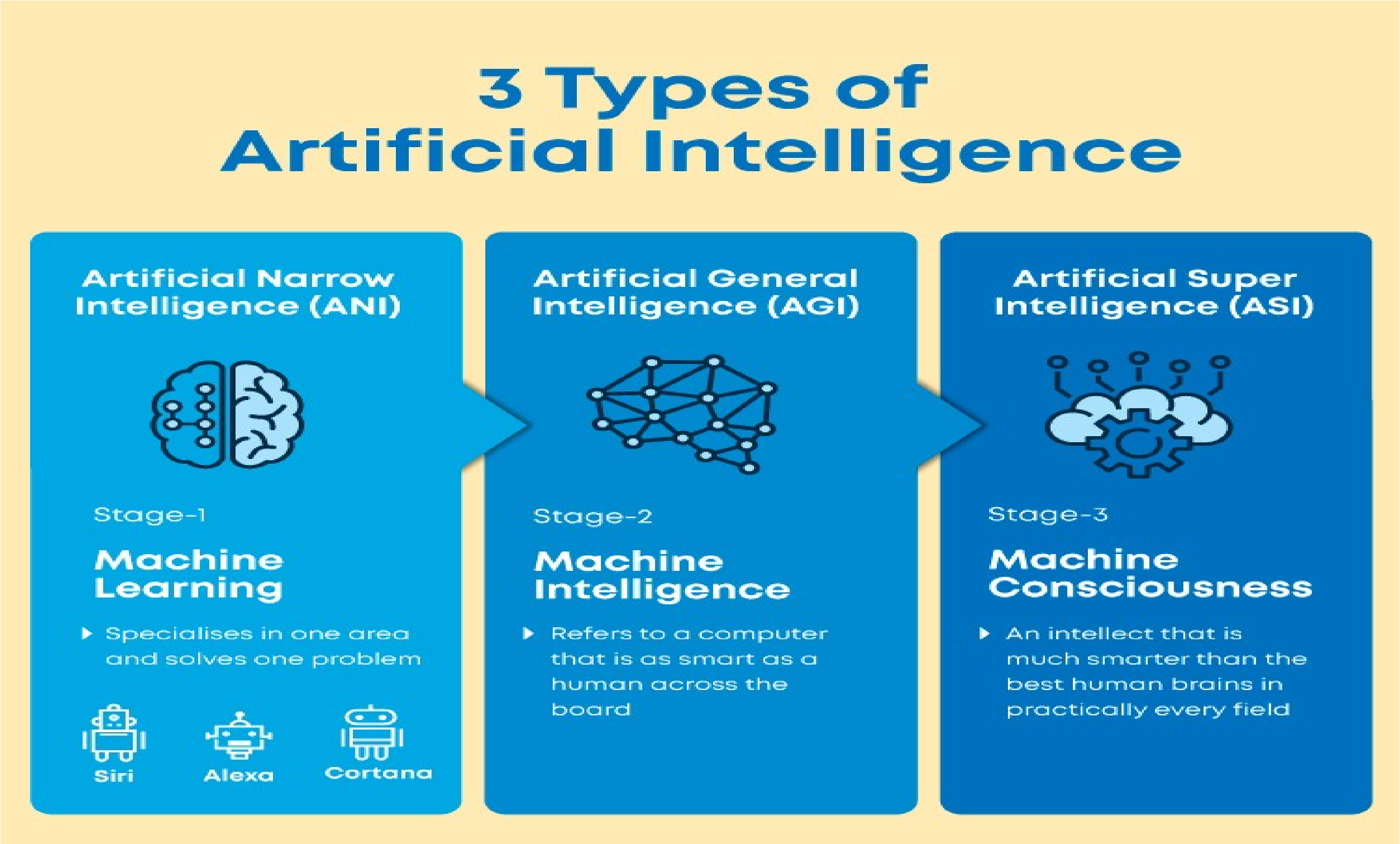
"We now believe that Broadcom's cloud/hyperscale ASIC revenues have recently passed $1B in annualized run-rate revenues and have more than doubled Y/Y since 2017 when the team was driving $50M in annualized revenues from the cloud titan (Google) " - Harlan Sur, JP Morgan

Cloud
The backbone of everything

Limitations

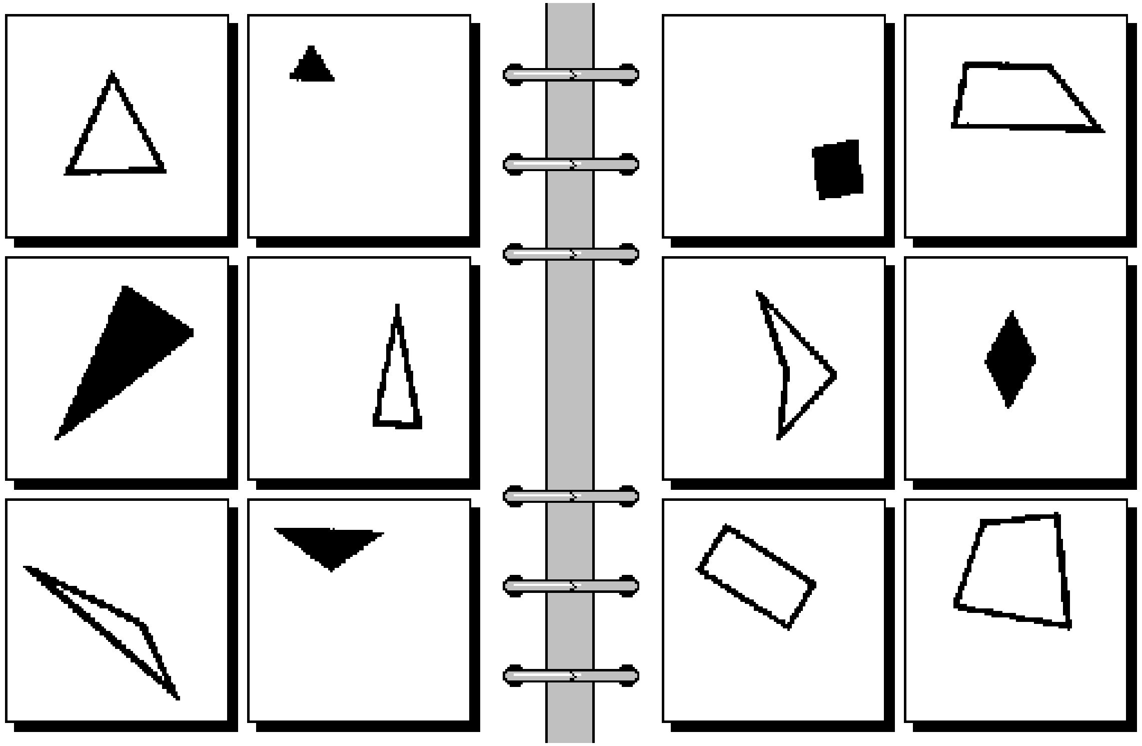
Current AI is ‘Narrow’ at best

“We are far away from human-level AI” – Yoshua Bengio



Current AI is 'Narrow' at best
Common Sense is not common (at least for machines)

Bongard's Problems



"A teenager learning to drive does not have to try to run off a cliff to see what happens. Whereas the AI system will have to run off the cliff to figure out that it is a bad idea; probably do it for a few thousand times before it realizes how not to do it"

"Current AI Systems are missing this common sense language which is why they are data hungry…and rigid…"

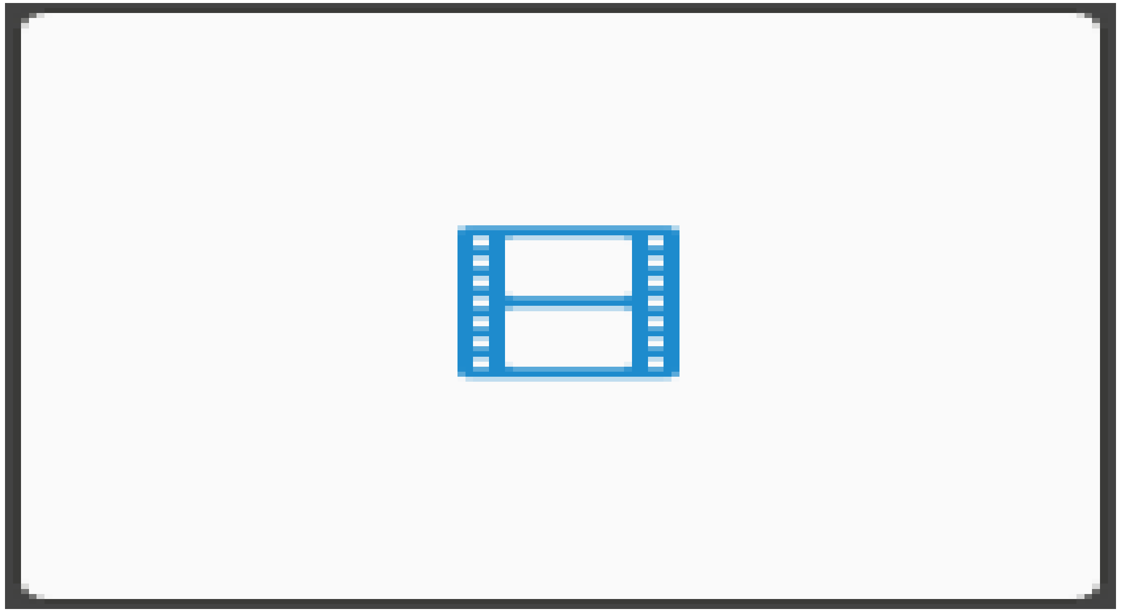
Yann LeCun

Real world is different from Training Data

NDTV LIVE TV LATEST COVID INDIA OPINION VIDEO CITIES WORLD OFFBEAT TRENDS

AI Camera Ruins Football Game By Mistaking Referee's Bald Head For Ball

Many complained that they missed their team's goals because the camera "kept thinking the Lino bald head was the ball."



Real world is different from Training Data

Zillow

How it Started?

Zestimate Algorithm: Invasion of AI on Real Estates

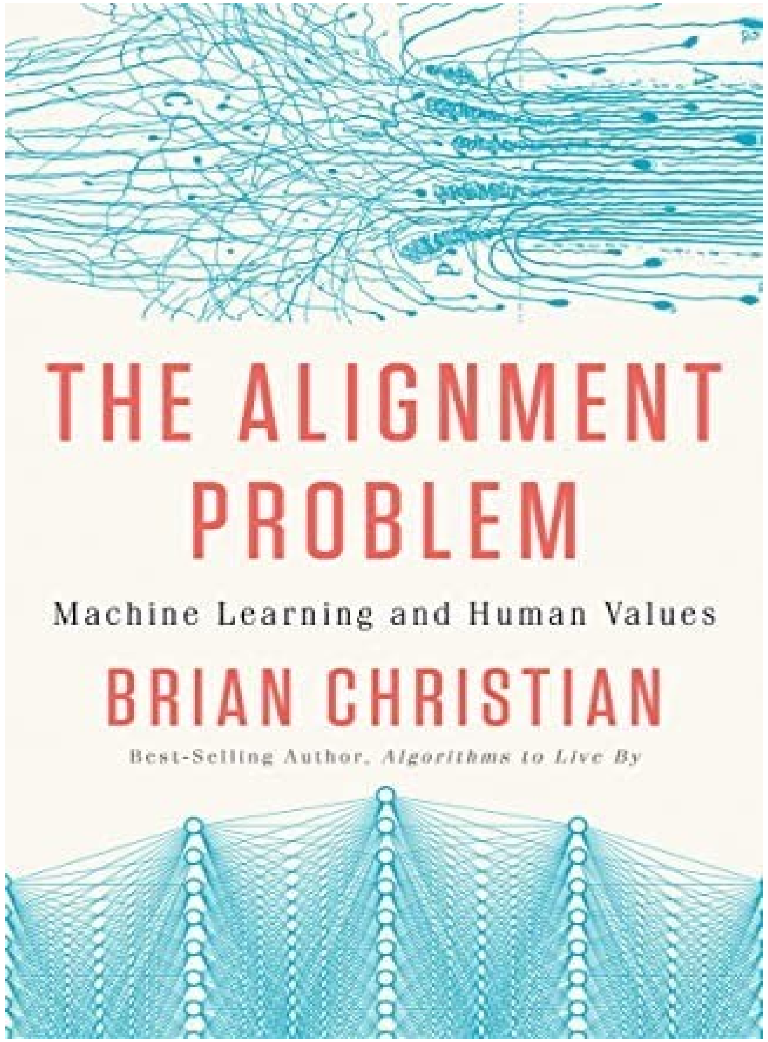
Incorporating artificial intelligence into several stages of the mortgage process.

How it's Going?

The $500mm+ Debacle at Zillow Offers – What Went Wrong with the AI Models?

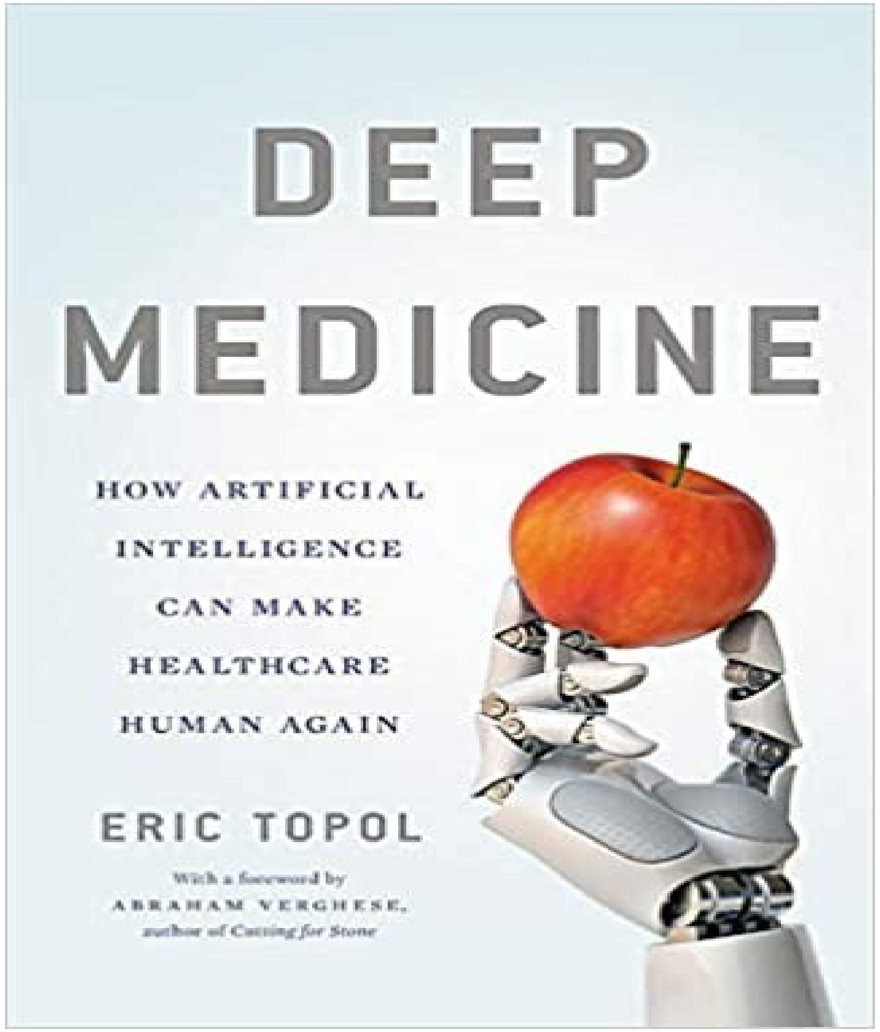
AI is (largely) Blackbox

- Data Inefficient
  - Not all domains have so much data
  - Research Areas : Zero Shot Learning , Few Shot Learning
- Blackbox
  - Explainabilitiy / Trust is a challenge.
  - Regulatory implications
- Ethical / Fairness Concerns
  - Training Data can have hidden bias



What happens to us?
Humans and Machines

- Humans and AI make different kind of errors
  - We aren't good at Statistical/Probabilistic Reasoning
  - We suffer from Behavioral biases
  - AI suffers from lack of Common Sense / Data Inefficiency / Bad Data
- AI can free humans to focus on more productive work
  - Spreadsheets didn't eliminate accountants. Rather made them more valuable



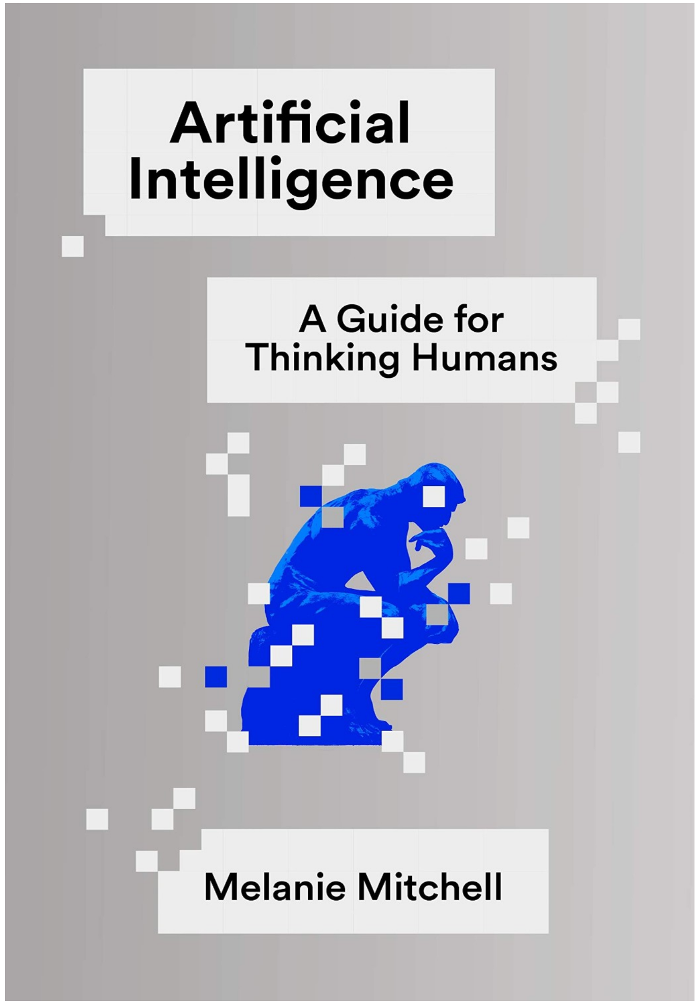
To Conclude

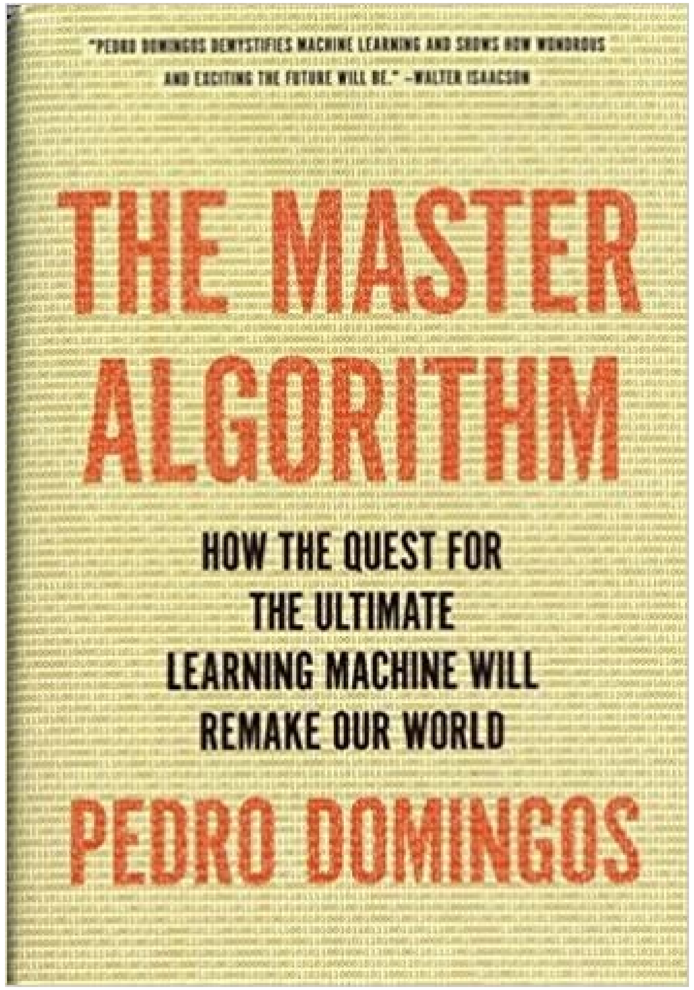
AI isn't magic but it can be great productivity tool

"The fundamental error was in chasing perfection and moonshots instead of the achievable markers of technological progress … one small step at a time. Implementing augmented intelligence into our day-to-day life does not need to be revolutionary, and it does not need to solve all of our problems at once to still be progress. But by taking the new tools and lessons of augmented intelligence, we can gradually improve every aspect of our lives.. ."

- Gary Kasparov

Appendix


Artificial Intelligence
A Guide for Thinking Humans
Melanie Mitchell


"PEDRO DOMINGOS DEMYSTIFIES MACHINE LEARNING AND SHOWS HOW WONDROUS AND EXCITING THE FUTURE WILL BE." —WALTER ISAACSON
THE MASTER ALGORITHM
HOW THE QUEST FOR THE ULTIMATE LEARNING MACHINE WILL REMAKE OUR WORLD
PEDRO DOMINGOS

Compute used to train deep learning models has increased 300,000x in six years

- Gumgum
- Stripe
- Diffbot

- “Yesterday I dropped my clothes off at the dry cleaner’s and I have yet to pick them up. Where are my clothes?”
  - **I have a lot of clothes**
- You are having a small dinner party. You want to serve dinner in the living room. The dining room table is wider than the doorway, so to get it into the living room, you will have to **remove the door. You have a table saw, so you cut the door in half and remove the top half.**
- **“**A teenager can learn to drive with 15-20 hours of practice, whereas millions of hours of training in different environments is not adequate for cars to drive themselves with the same degree of reliability, he said”
- “LeCun said the laws of physics do not change when you move from North America to Britain. He said this allows the teenager to drive and not have to try to run off a cliff to see what happens. Whereas the AI system will have to run off the cliff to figure out that it is a bad idea; probably do it for a few thousand times before it realises how not to do it. So, that is what we are missing, said LeCun”