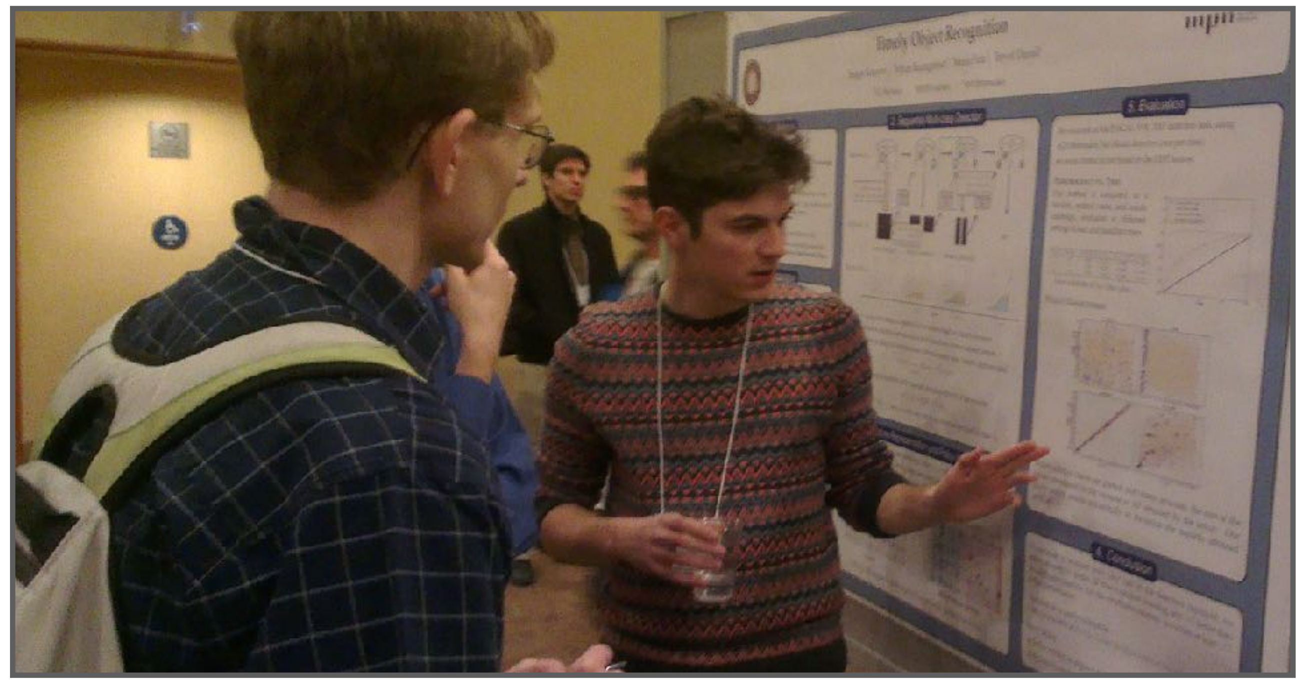
AI Masterclass

Sergey Karayev

ASU GSV 2019

About Me

- Head of AI for STEM at **Turnitin** [2018-]
  - 35M students served at 150+ countries. 1B paper submissions.
- Co-founder of **Gradescope** [2014-2018]
  - From TA side project to 80M answers graded by over 10K instructors.
- Co-organizer of **Full Stack Deep Learning** program
- PhD Computer Science at **UC Berkeley** [2009-2014]
  - Research in computer vision, deep learning



My Goals For This Presentation

- Reach shared terminology
- Situate ourselves in history and possibility
- Understand AI product development

Outline

- Introduction
  - **History and terminology**
  - What's possible, what's on the horizon, what's unknown
- Developing AI Products
  - Picking the problem
  - Data Flywheel
  - Most AI code is not AI
  - Roles and Hiring
- Example
- Q & A

Terminology

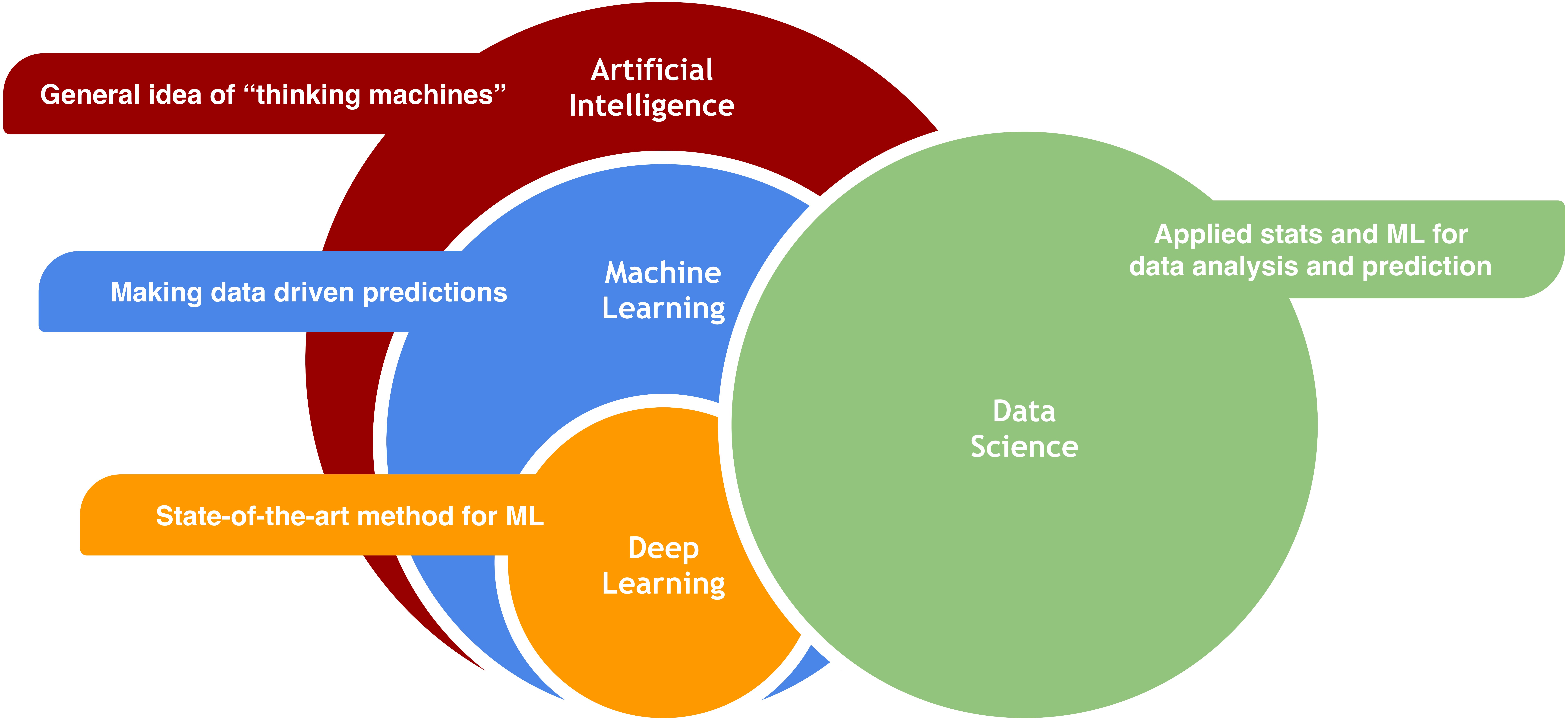


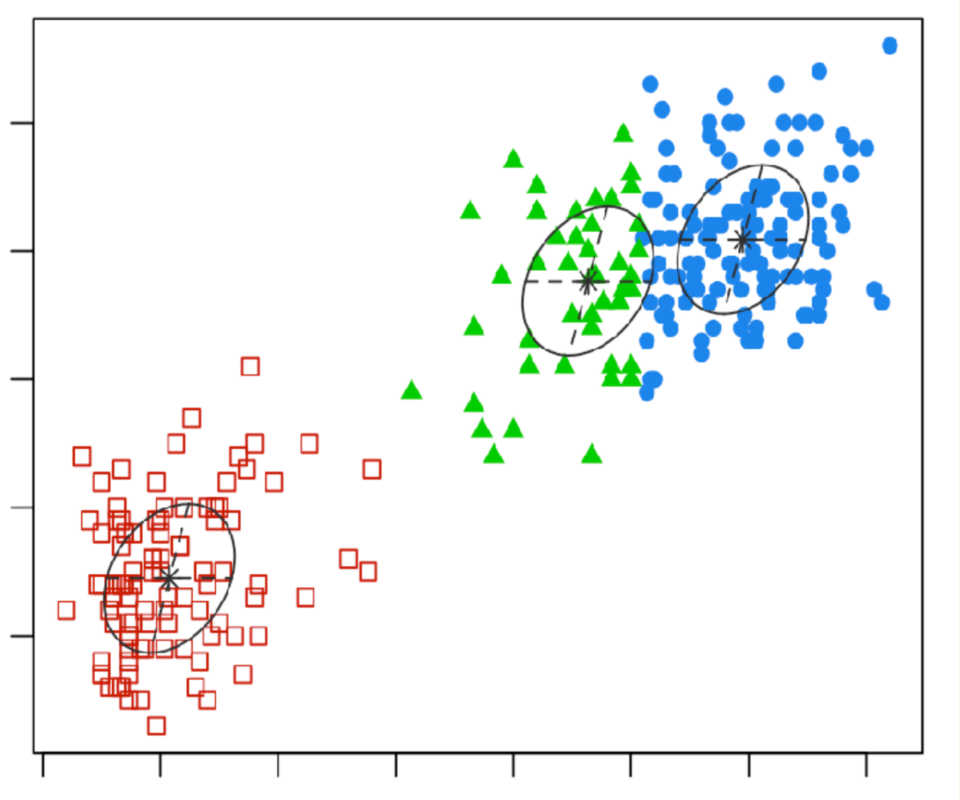
Diagram courtesy of Eric Wang

Types of Learning Problems

Unsupervised Learning

- ➔ Unlabeled data X
- ➔ Learn X
- ➔ Generate fakes, insights

"This product does what it is supposed to. I always keep three of these in my kitchen just in case ever I need a replacement cord."



Supervised Learning

- ➔ Labeled data X and Y
- ➔ Learn X -> Y
- ➔ Make Predictions

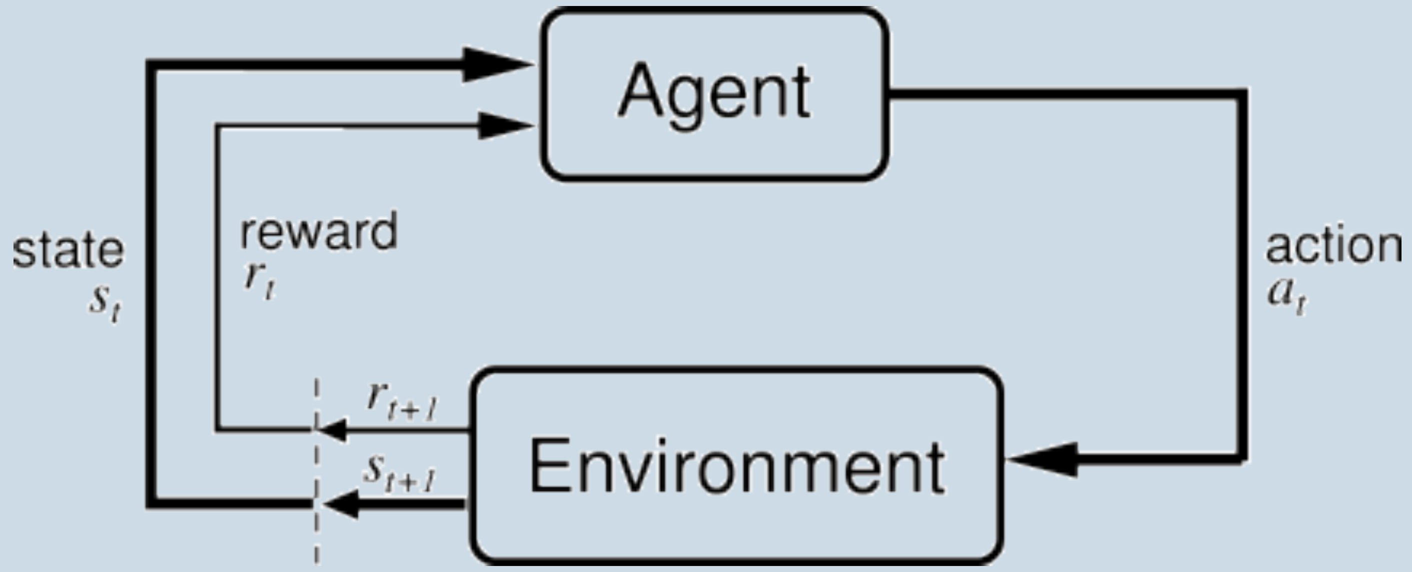
→ *cat*

→ *"Hey Siri"*

Commercially Viable Today

Reinforcement Learning

- ➔ Learn how to take Actions in an Environment

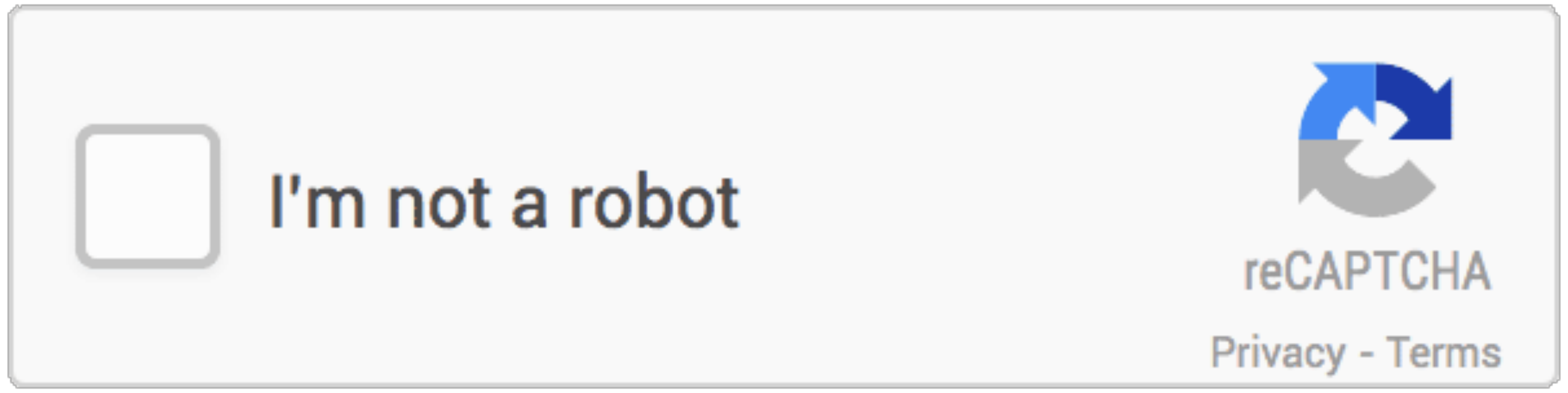


Up Next

Diagram courtesy of Shayne Miel

1950: Turing Test

- Alan Turing suggests test for determining human-level AI
    - *Can the AI convince a human interlocutor that it is another human?*
- **Today**
    - Intelligence still defined functionally
    - No good argument against possibility of superhuman intelligence



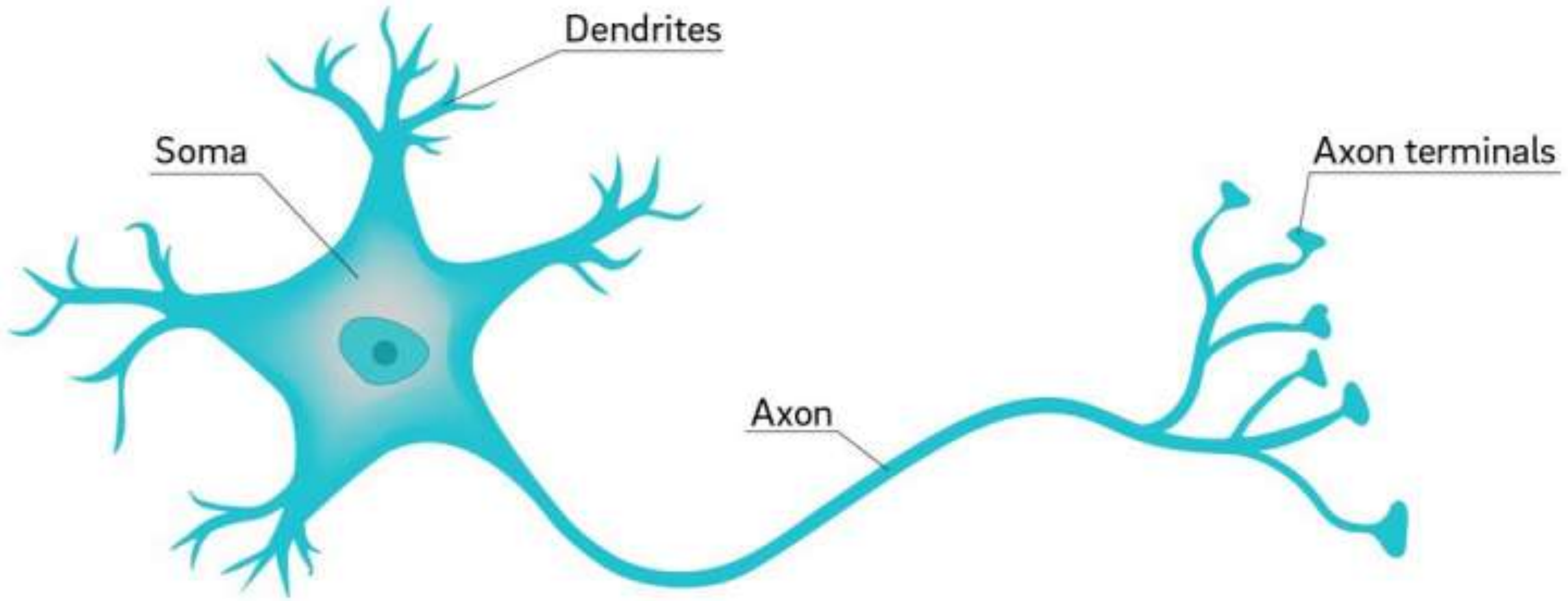
https://en.wikipedi

https://theness.com/neurologicablog/index.php/ai-and-the-chinese-room-argument/

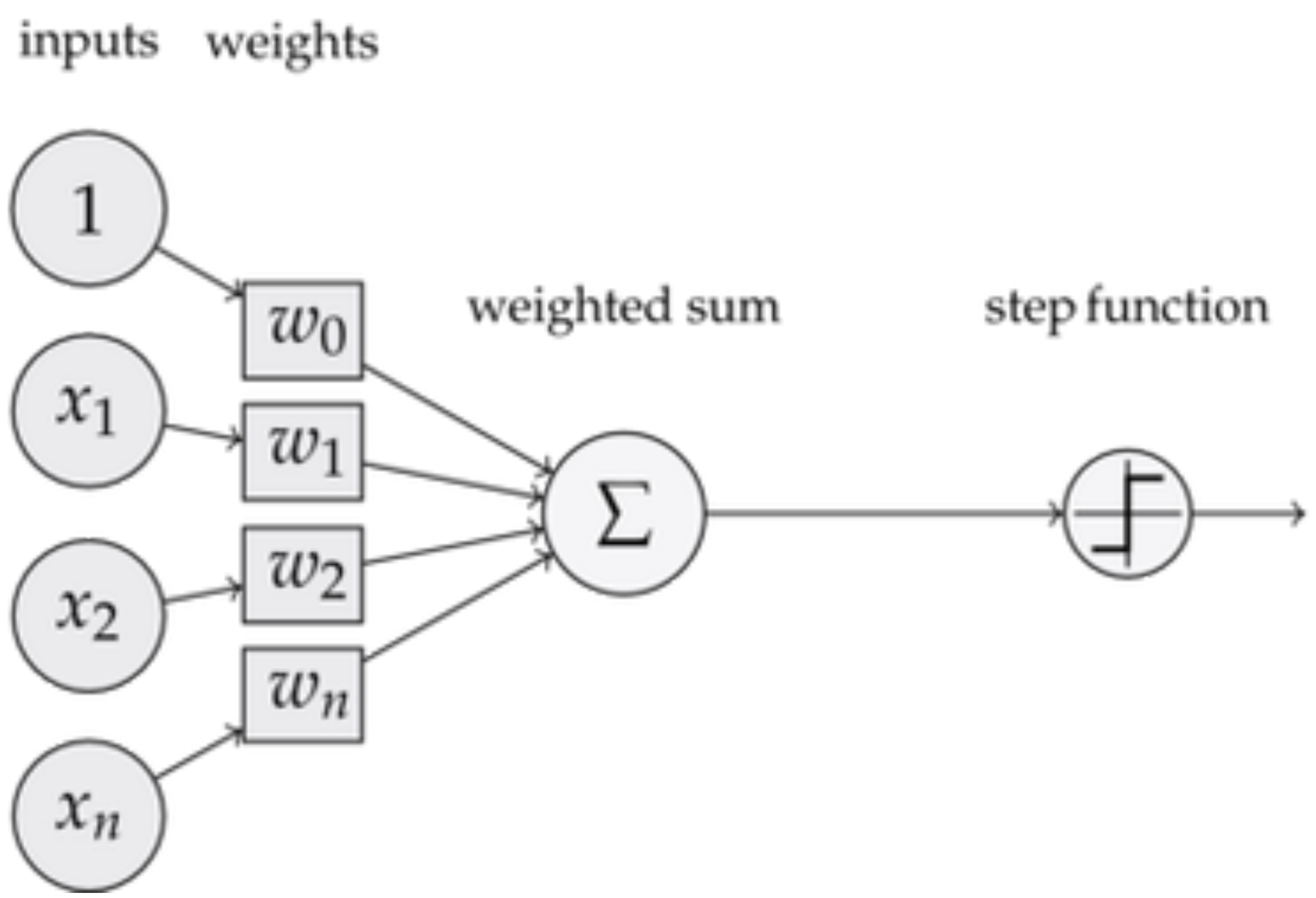
1957: The Perceptron

- Frank Rosenblatt builds machines inspired by spiking neurons
- Aggregate inputs, take weighted sum, apply non-linearity
- **Today:**
  - The building block of deep learning
  - AI still informed by neuroscience

https://blogs.umass.edu/comphon/2017/06/15/did-frank-rosenblatt-invent-deep-learning-in-1962/



https://medicalxpress.com/news/2018-07-neuron-axons-spindly-theyre-optimizing.html



https://www.jessicayung.com/explaining-tensorflow-code-for-a-multilayer-perceptron/

1960s-1970s: first summer and winter

- At first: lots of funding, great expectations
- *"In from 3-8 years we will have a machine with the general intelligence of an average human being."* - Marvin Minsky (1970)
- 1973: "In no part of the field have the discoveries made so far produced the major impact that was then promised" - Lighthill Report (cut most funding in UK)

https://www.computerhistory.org/internethistory/1960s/

https://www.youtube.com/watch?

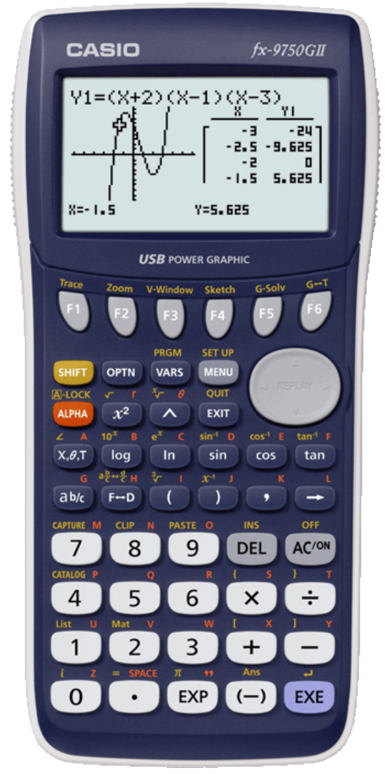
"Our first foray into AI was a program that did a credible job of solving problems in college calculus. Armed with that success, we tackled high school algebra; we found, to our surprise, that it was much harder. An exploration of the child's world of blocks proved insurmountable, except under the most rigidly constrained circumstances.

It finally dawned on us that the overwhelming majority of what we call intelligence is developed by the end of the first year of life."

- Marvin Minsky, 1977

https://unsplash.com/photos/x4jRmkuDImo

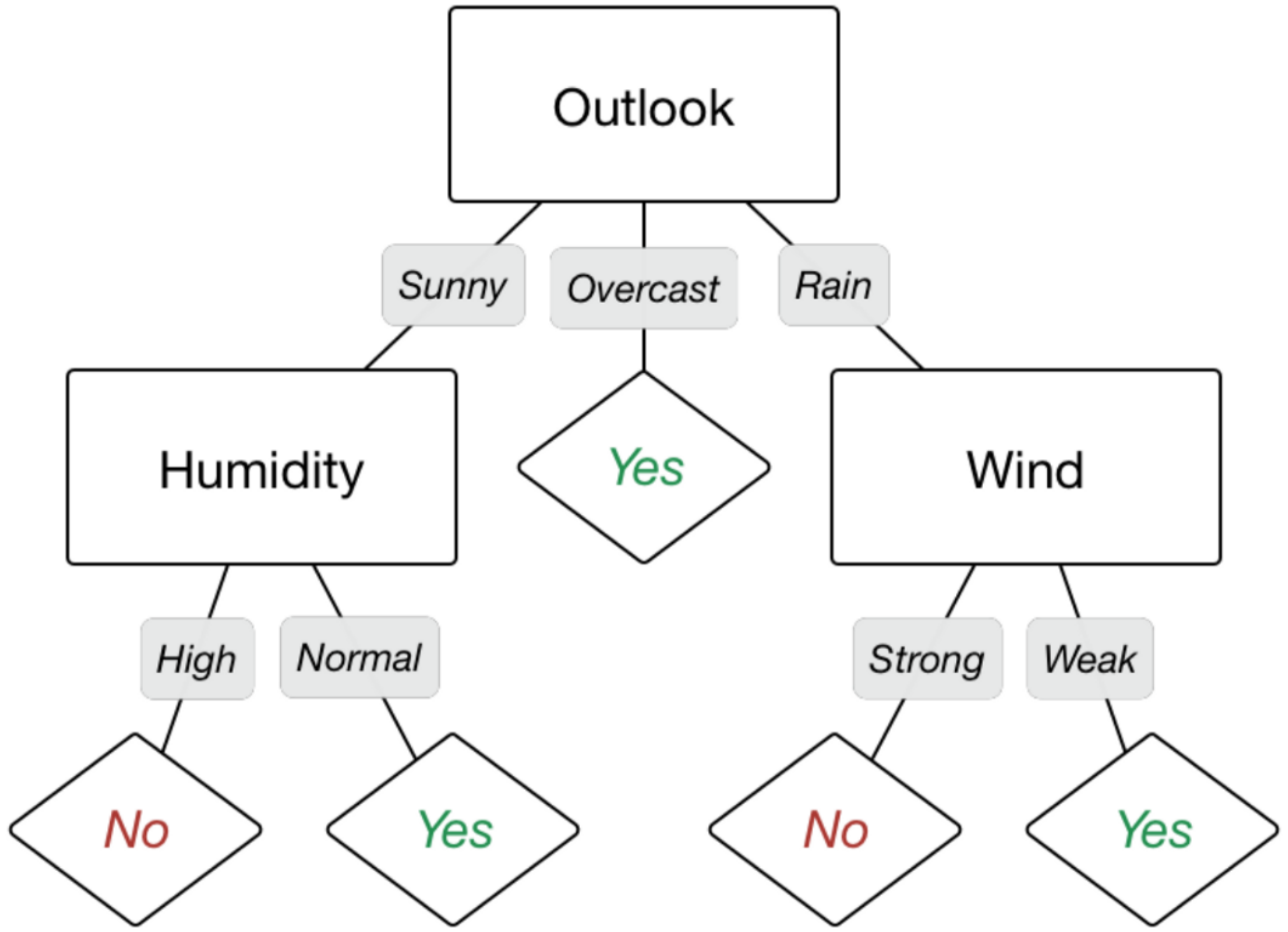
vs



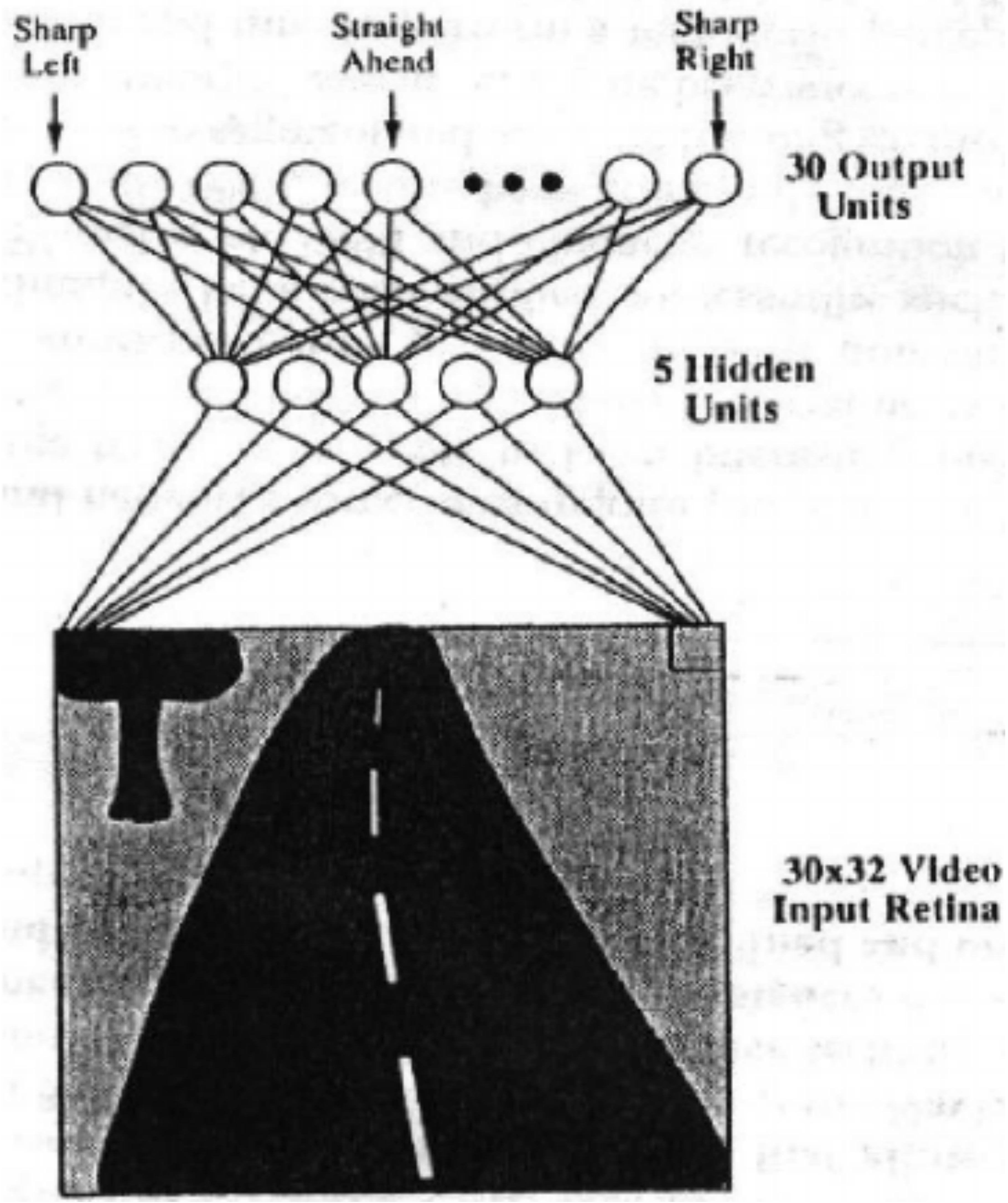
https://www.casio.co.uk

1980s and 1990s

- Rise of "expert systems"
- Use of logic languages like PROLOG
- Another rise of neural networks
- Another AI Winter



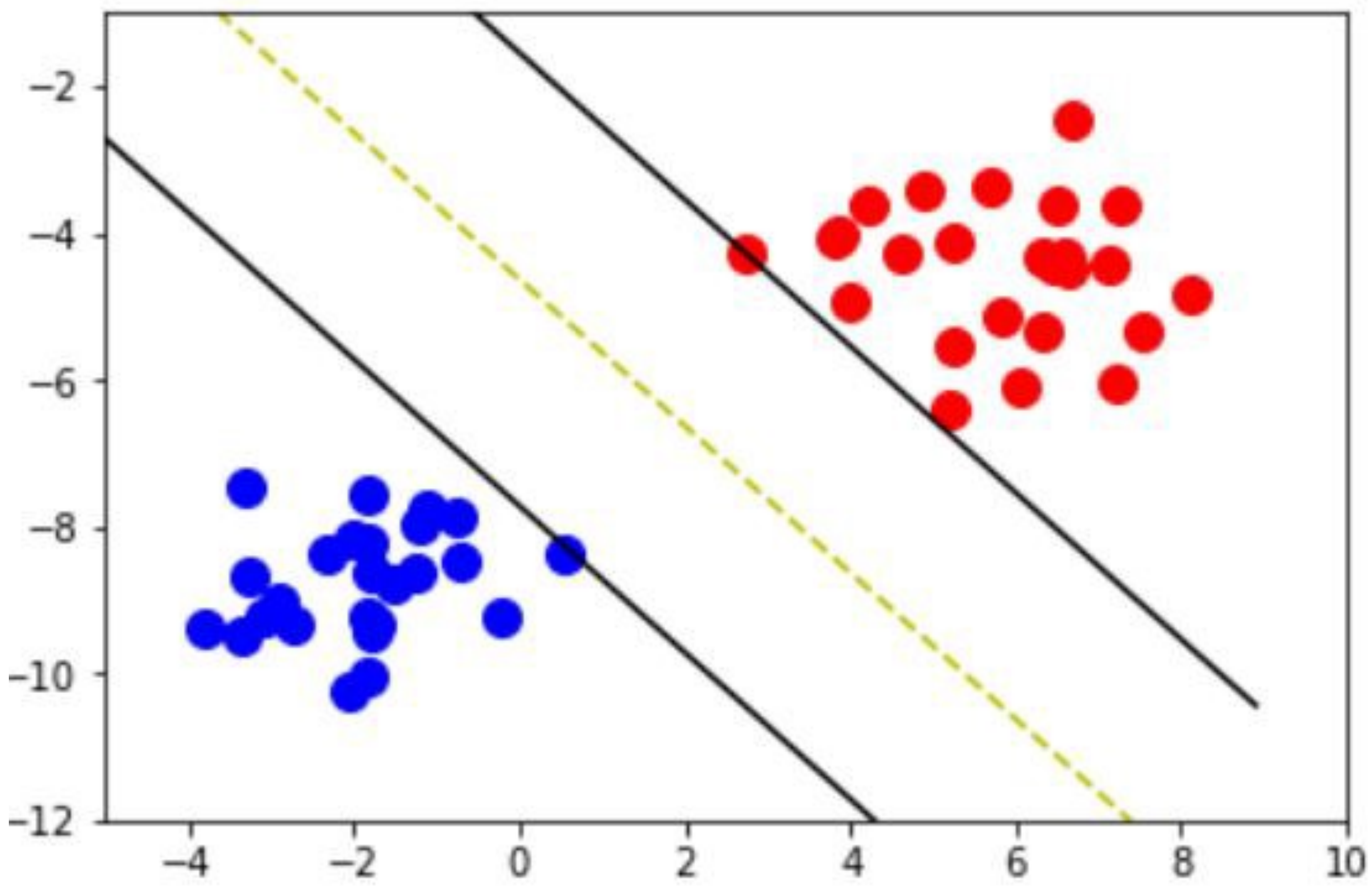
https://becominghuman.ai/decision-trees-in-machine-learning-f362b296594a



https://www.researchgate.net/publication/3193351_Vision_for_Mobile_Robot_Navigation_A_Surve

2000s: Machine Learning

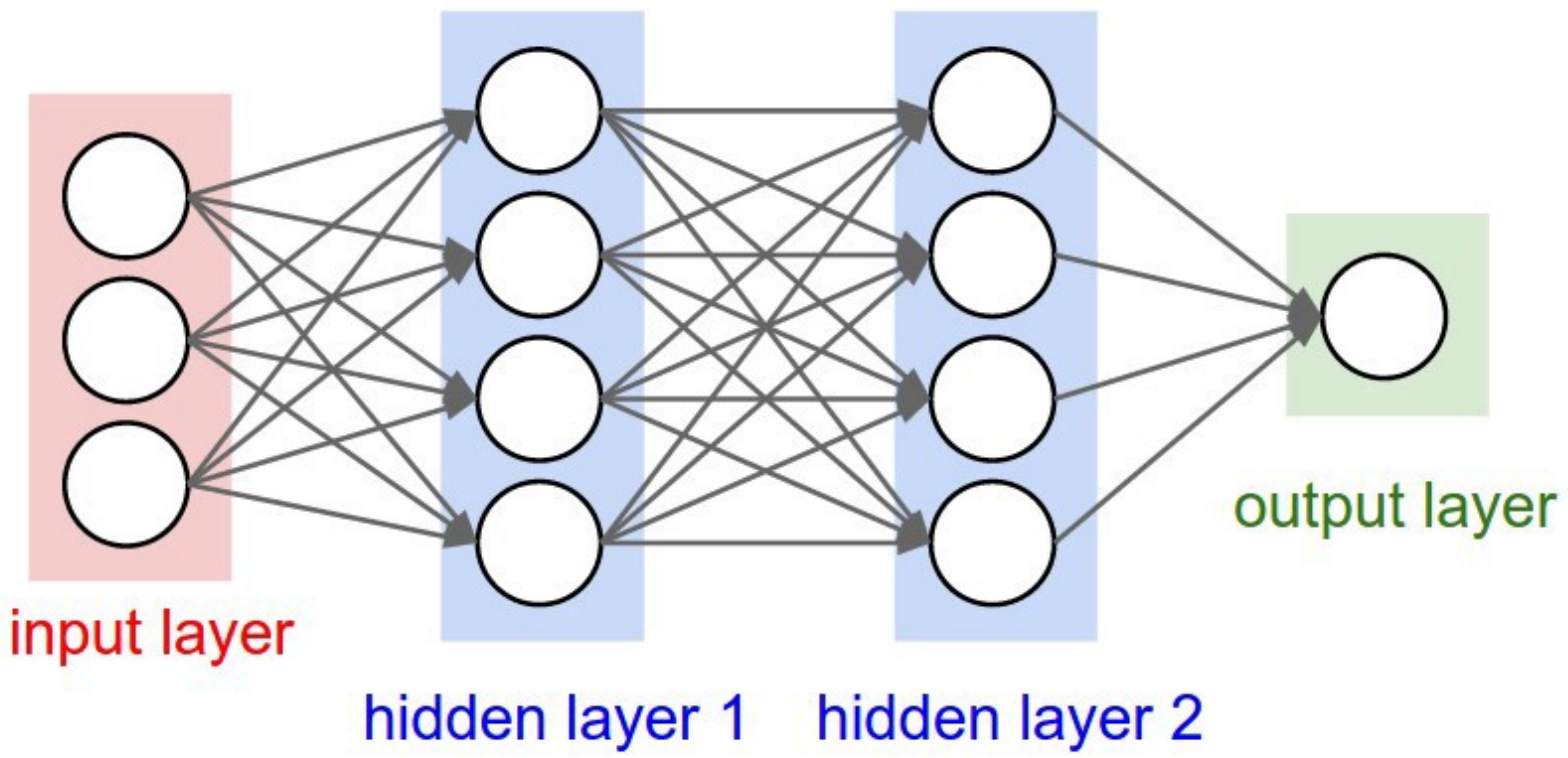
- Informed by stats, increasingly big data
- Main techniques: SVMs
- Hand-designed data representations
- **Today**
  - Full understanding that data trumps algorithms



https://medium.com/deep-math-machine-learning-ai/chapter-3-support-vector-machine-with-

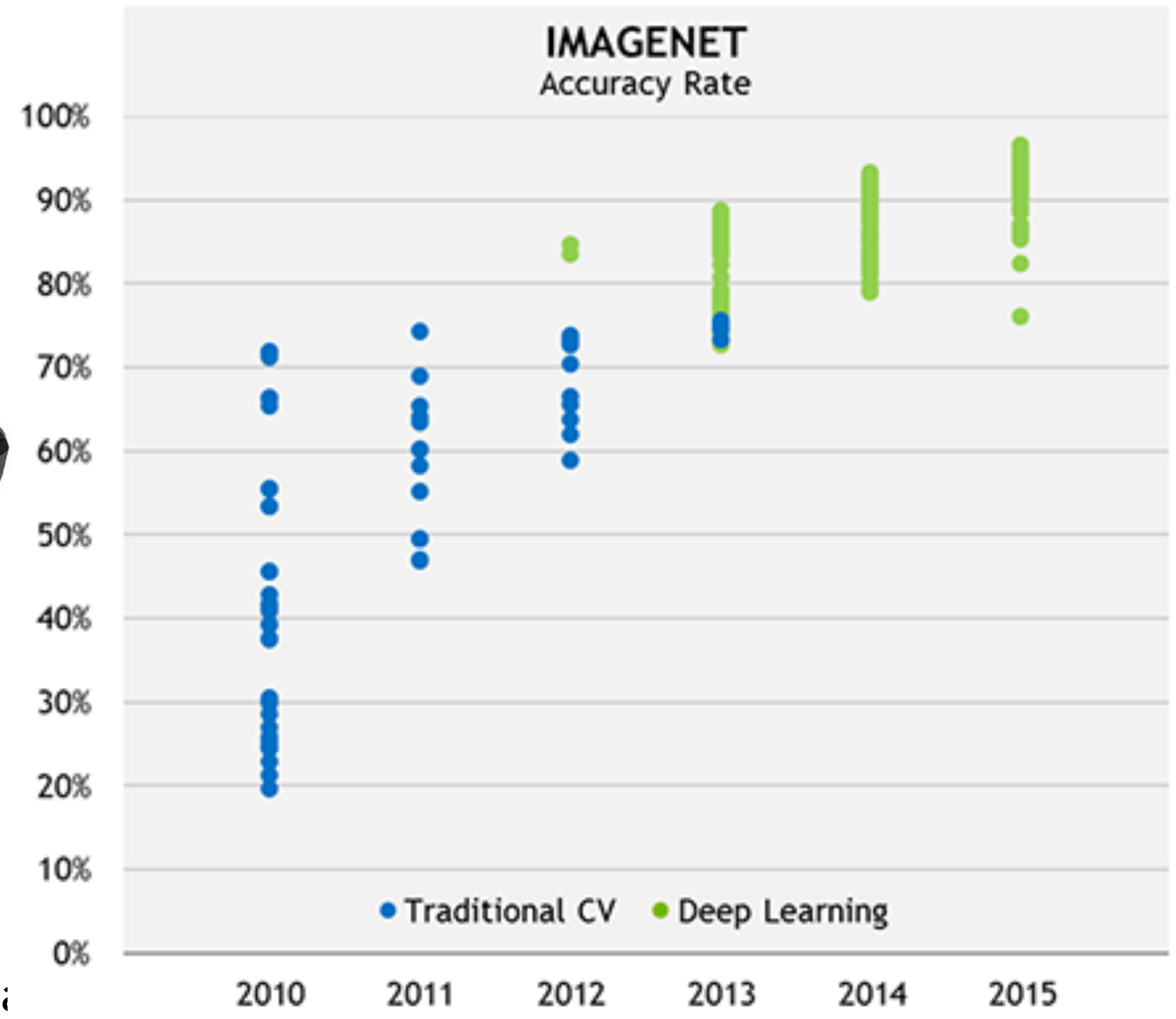
2010s: Deep Learning

- Many perceptrons arranged in many hidden layers
- **Better than hand-designed data representations**
- Combination of algorithms (backpropagation) and increased compute power (GPUs)
- Major breakthrough on challenge datasets in vision, language, robotics
- Significant commercial applications

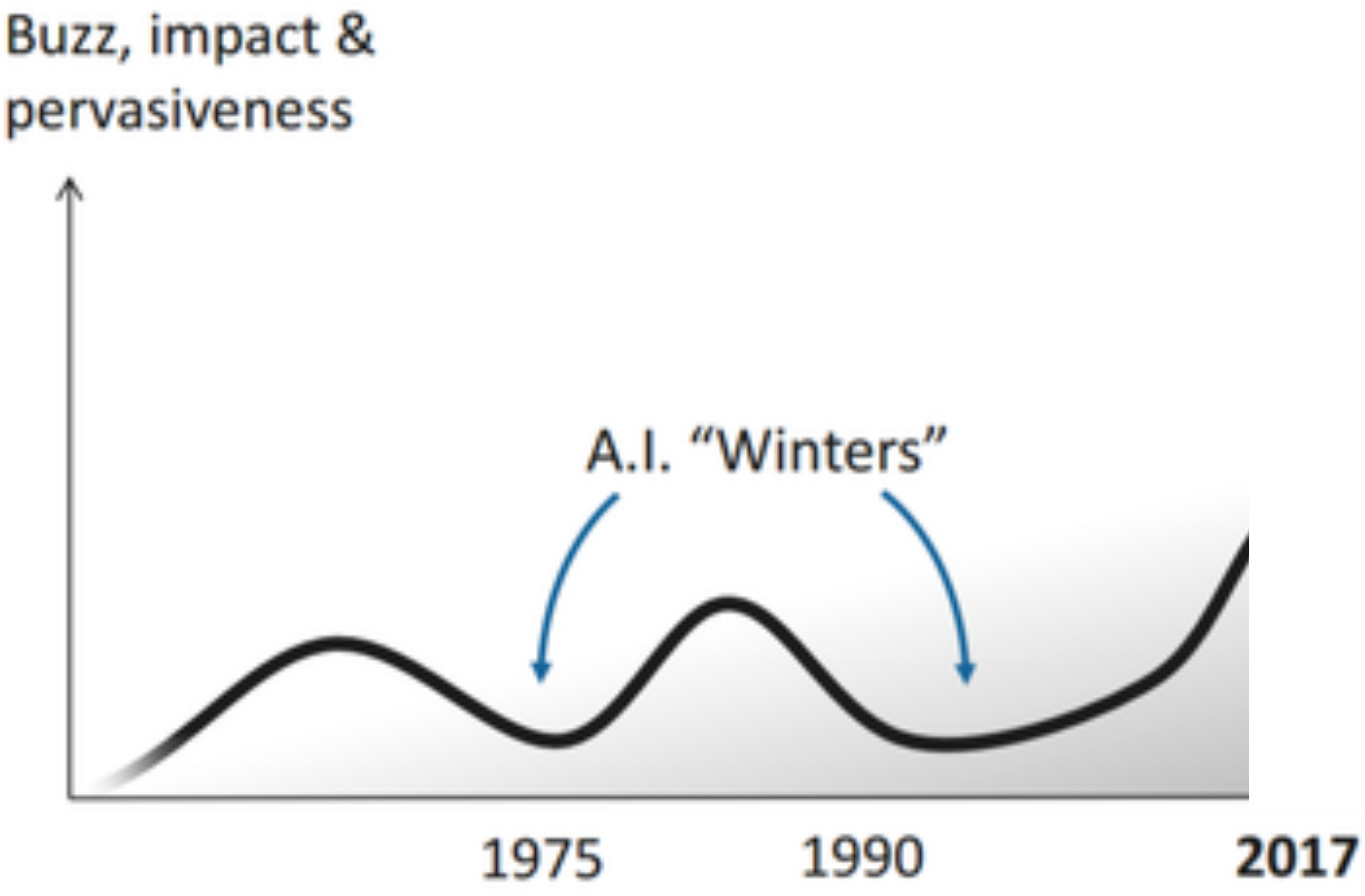


https://www.digitaltrends.com/cool-tech/what-is-an-artificial-neural-network/

https://nvidia.com

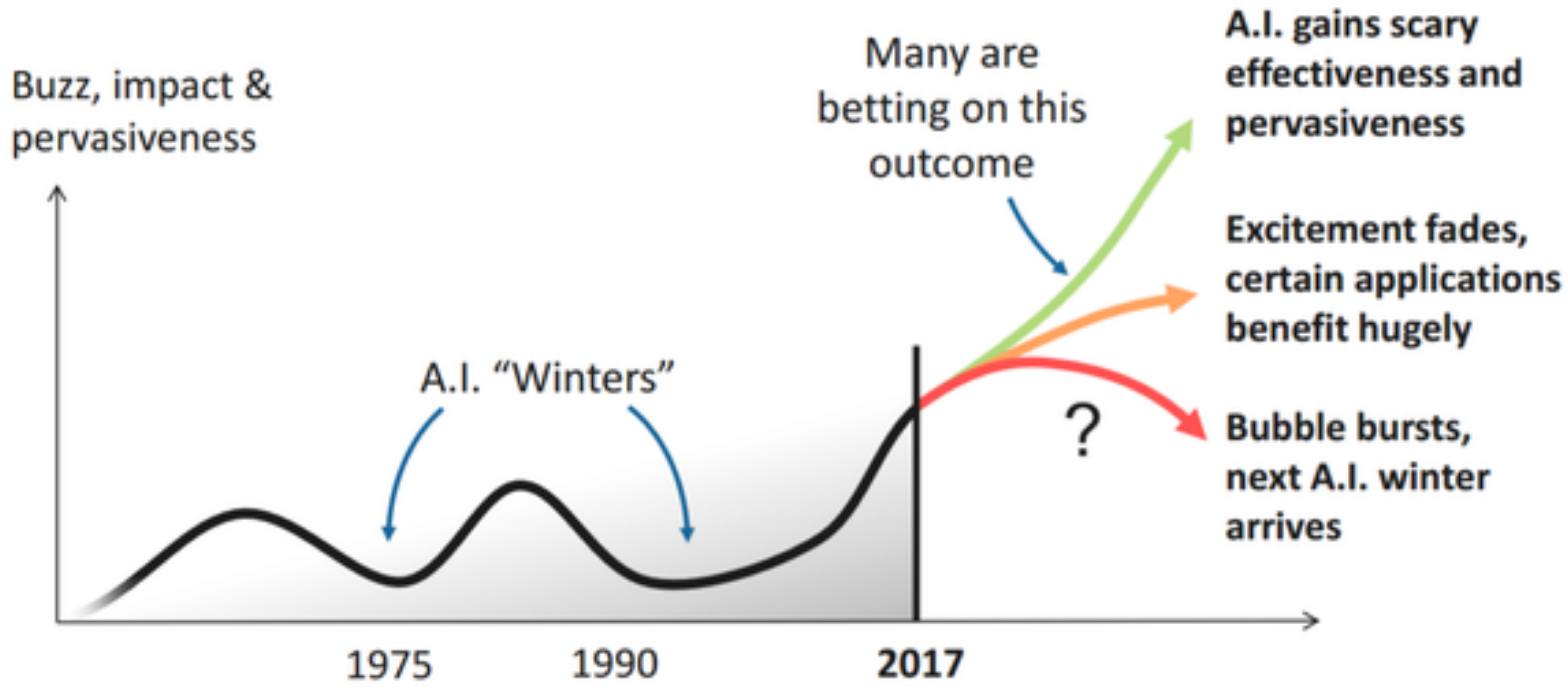


2020s: ?



https://www.cambridgewireless.co.uk/media/uploads/resources/AI%20Group/AIMobility-11.05.17-Cambridge_Consultants-Monty_Barlow.pdf

2020s: ?



https://www.cambridgewireless.co.uk/media/uploads/resources/AI%20Group/AIMobility-11.05.17-Cambridge_Consultants-Monty_Barlow.pdf

Outline

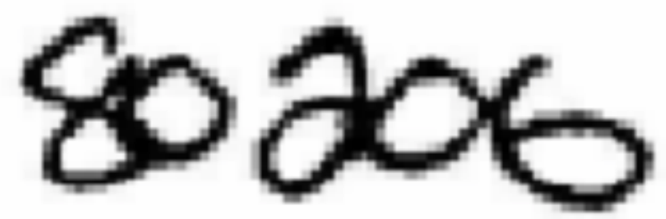
- Introduction
  - History and terminology
  - **What's possible, what's on the horizon, what's unknown**
- Developing AI Products
  - Picking the problem
  - Data Flywheel
  - Most AI code is not AI
  - Roles and Hiring
- Example
- Q & A

AI is a Moving Target

Problem	Before computer solution	After computer solution
Advanced mathematical computation	The most difficult task for humans, must be most difficult for machines	Just an algorithm, no intelligence at all!
Chess	A proving ground for human ingenuity and insight	Obviously computers are able to crank through more possible moves!
Speech	Only humans are able to produce and undestand vocal language	Just translating between soundwave and text, no understanding!
Humor	We can't even define what makes something funny, so how will a computer generate it?	?

What was possible yesterday

- Recognize handwritten zip codes (1989)
- Self-drive car across the US (1995)
- Beat humans at chess (1997)
- Vacuum the floor (2002)



https://www.researchgate.net/figure/An-Illustrative-example-for-USPS-Zip-Code-Data-Sample-from-17_fig1_320250462

https://jalopnik.com/they-drove-cross-country-in-an-autonomous-minivan-witho-1696330141

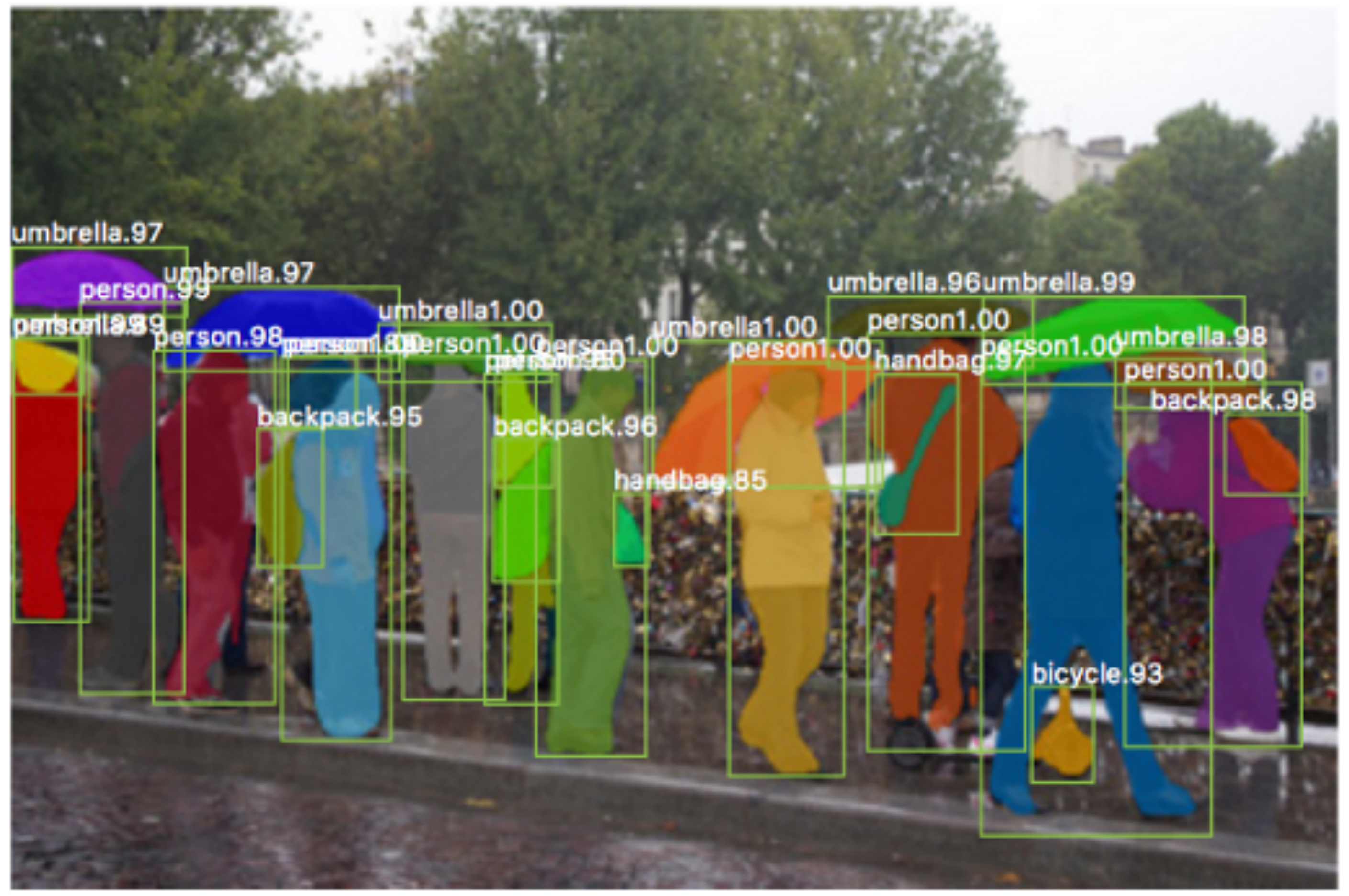
http://mentalfloss.com/article/503178/brief-history-deep-blue-ibms-chess-computer

What's possible today

- Car that can self-drive on well-mapped roads in good conditions
- Beat humans at Go
- Recognize general handwriting
- Recognize things and people in photos
- Translate text from one language to another
- Transcribe speech to text and back
- Generate realistic human faces
- Stilted customer support

https://www.wired.com/story/waymo-self-driving-cars-california/

https://thispersondoesnotexist.com/



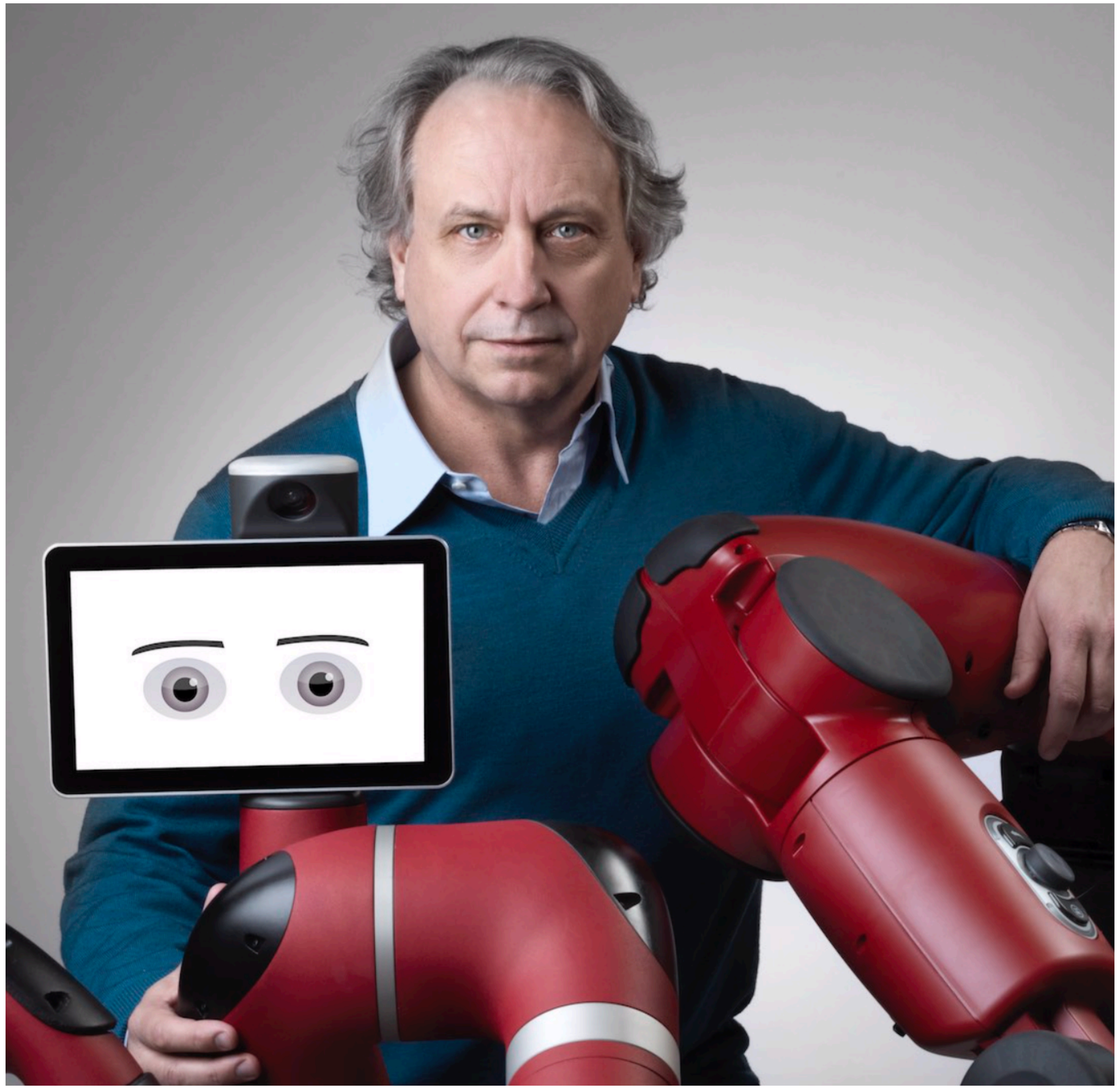
https://arxiv.org/abs/1703.06870

What's likely within 5 years

- Cars that can pick up passengers in restricted environments
- Beat humans at complicated video game
- Generate realistic videos
- Generate realistic news articles
- Generate realistic music
- Customer support good enough to fool humans

What's unknown

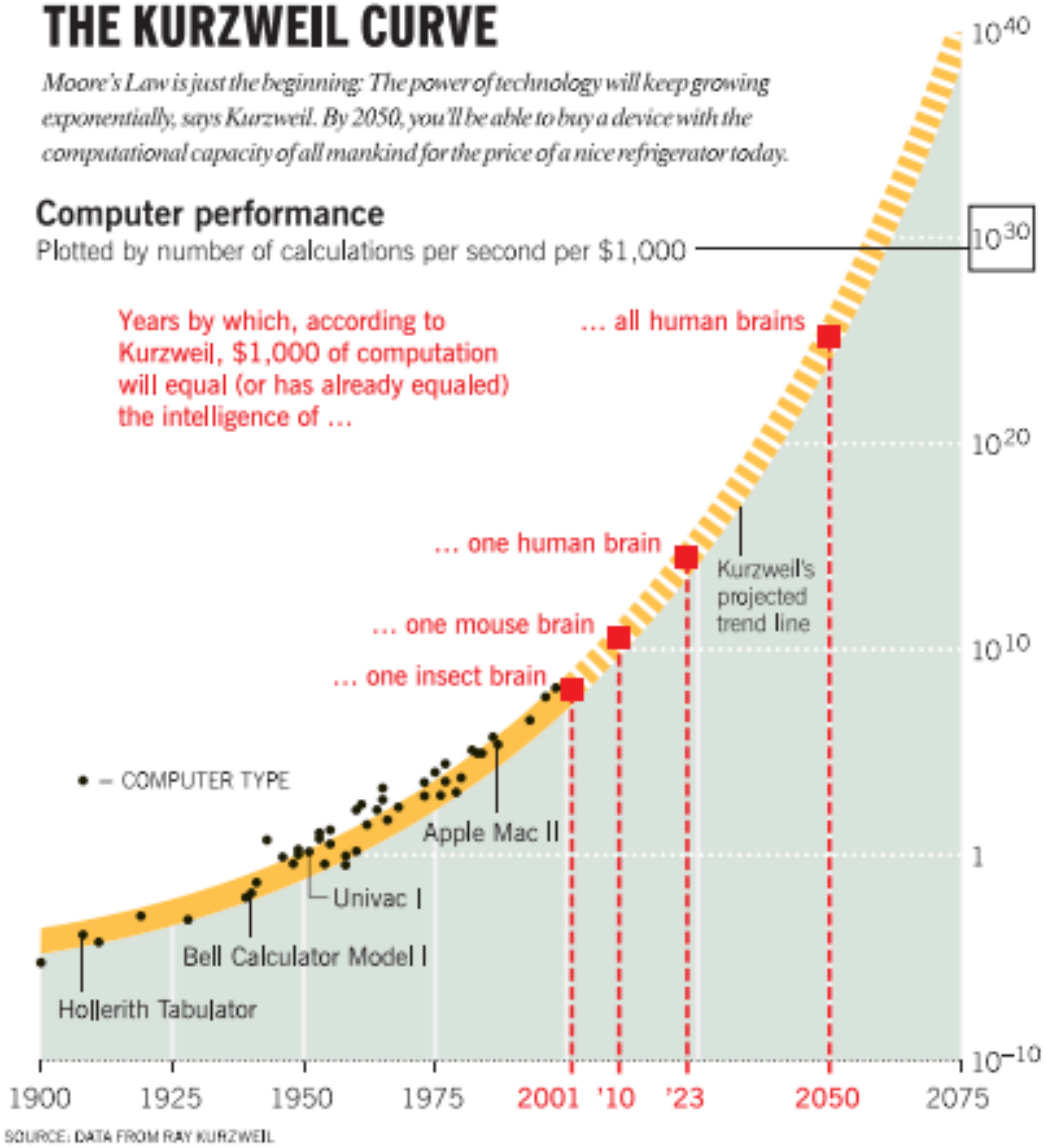
- Cars that can self-drive and pick up passengers on all roads at all times
- A robot that is perceived to be as smart as a dog
- An intelligent tutor in all subjects
- *"Almost all innovations in Robotics and AI take far, far, longer to get to be really widely deployed than people in the field and outside the field imagine."* - Rodney Brooks
- **Artificial General Intelligence (AGI) and Artificial Superhuman Intelligence (ASI)**



https://rodneybrooks.com/the-seven-deadly-sins-of-predicting-the-future-of-ai/

The Singularity

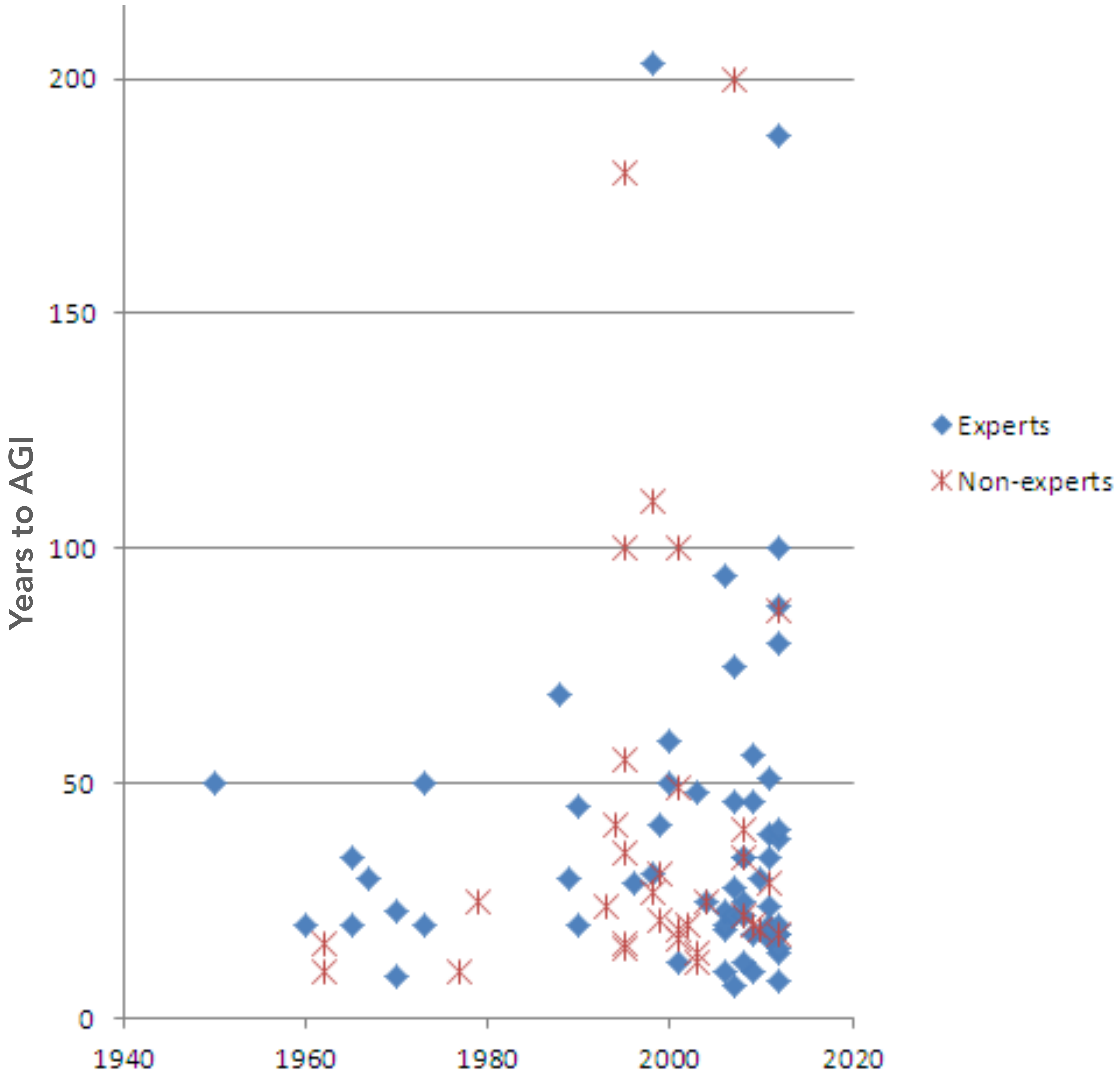
- Idea that technological progress is exponential, and will enter essentially "vertical" growth
    - Humans ------------> AGI -> ASI
- Some notable people are alarmed
- **Let's not worry just yet**



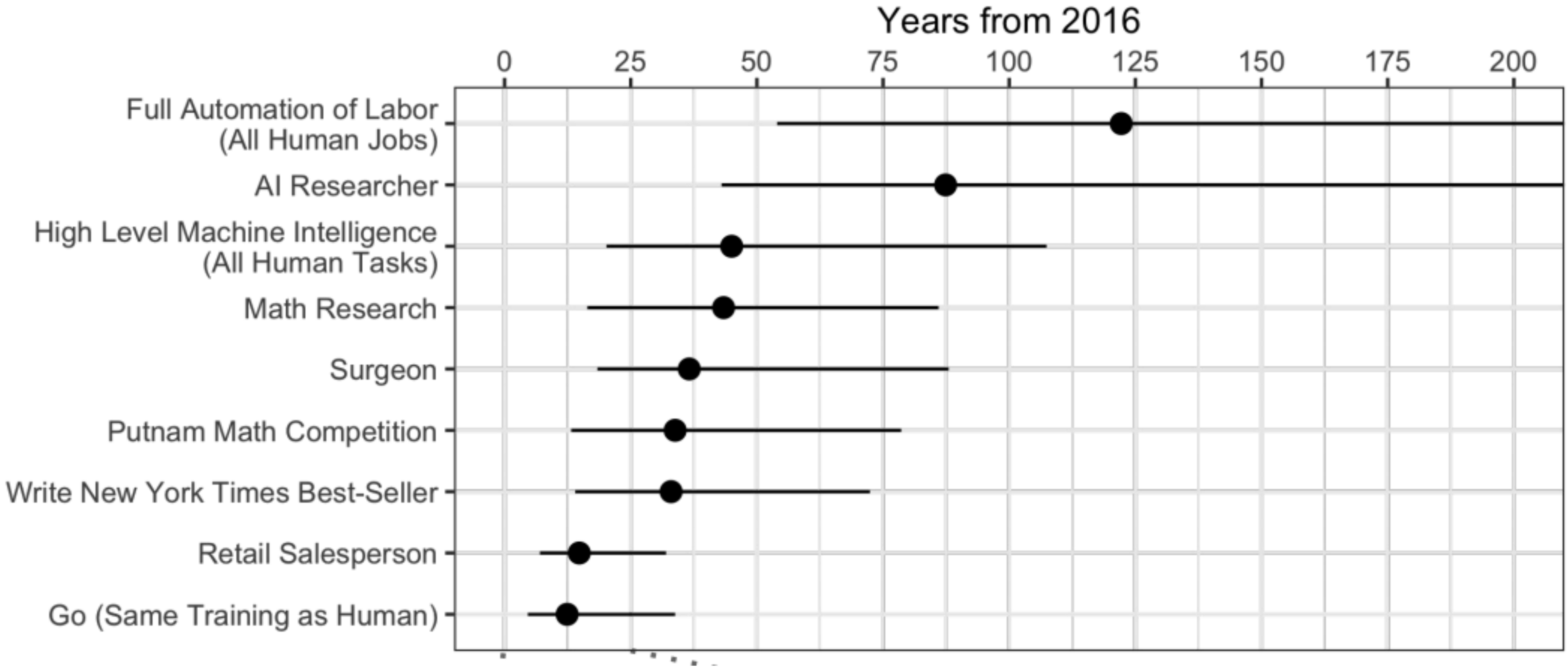
https://www.washingtonpost.com/news/morning-mix/wp/2014/11/18/why-elon-musk-is-scared-of-killer-robots

Historical Predictions

People have been predicting that human-level AI is 25 years away for a while.



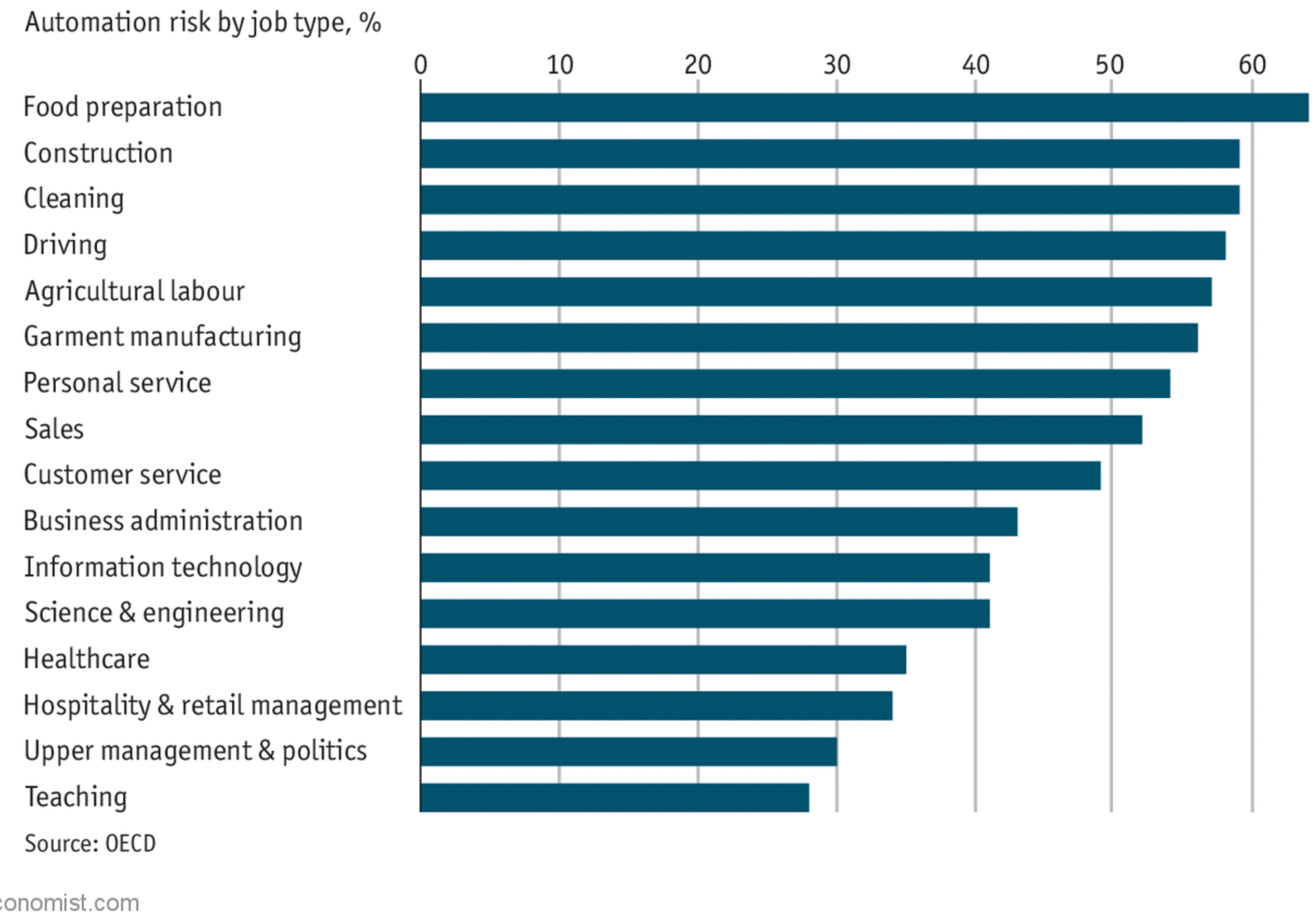
Disappearing Lines of Work



When Will AI Exceed Human Performance?
Evidence from AI Experts
https://arxiv.org/pdf/1705.08807.pdf

Disappearing Lines of Work

- **Let's worry right now!**
- Long-term benefit, but short-term upheaval.
- What to do?
  - **10x more effective education**
  - Stronger support systems
  - Radical societal re-org? 😯



https://www.transhuman-party.org/transhumanism-and-marxism

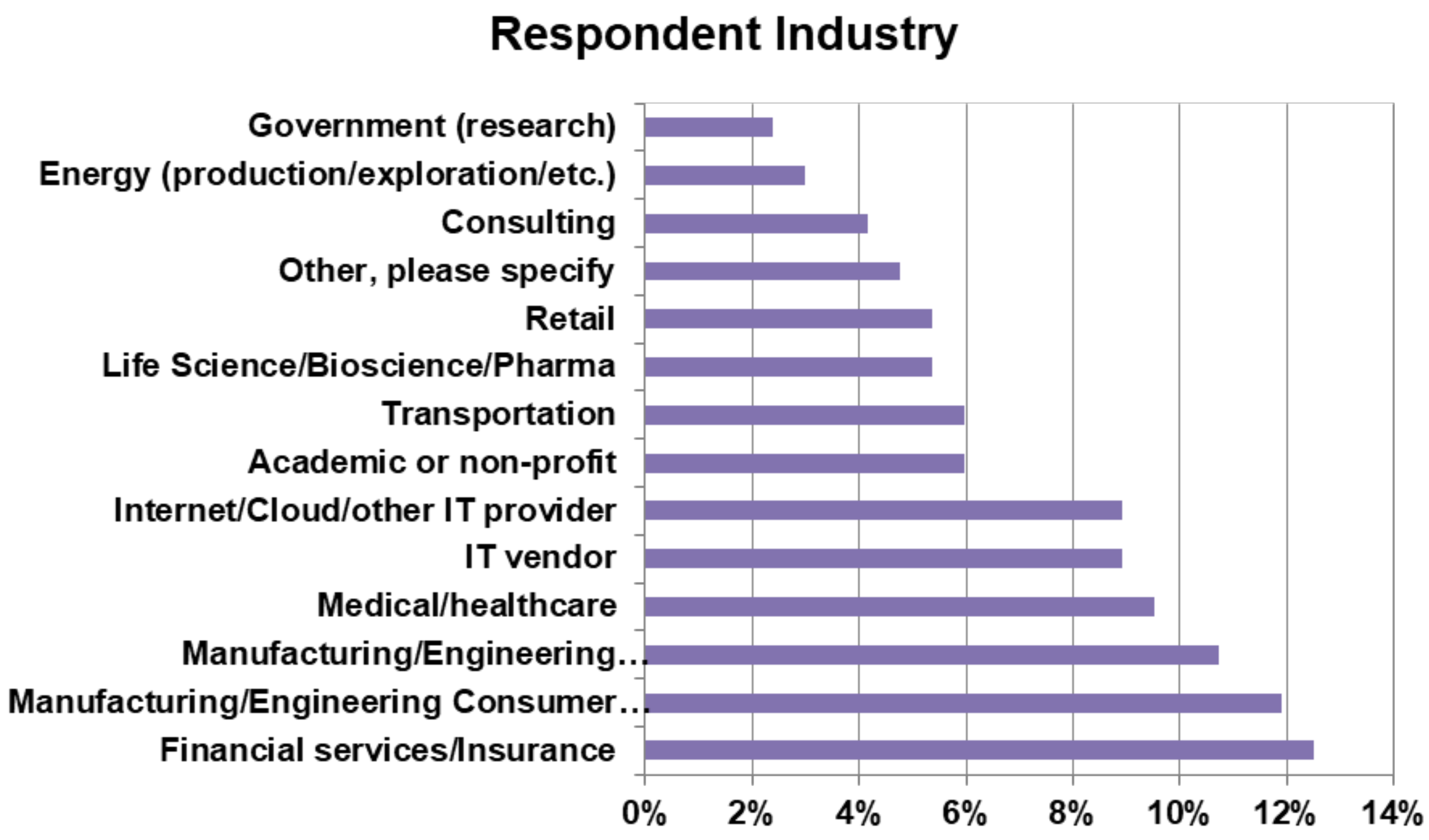
https://www.economist.com/graphic-detail/2018/04/24/a-study-finds-nearly-half-of-jobs-are-vulnerable-to-automation

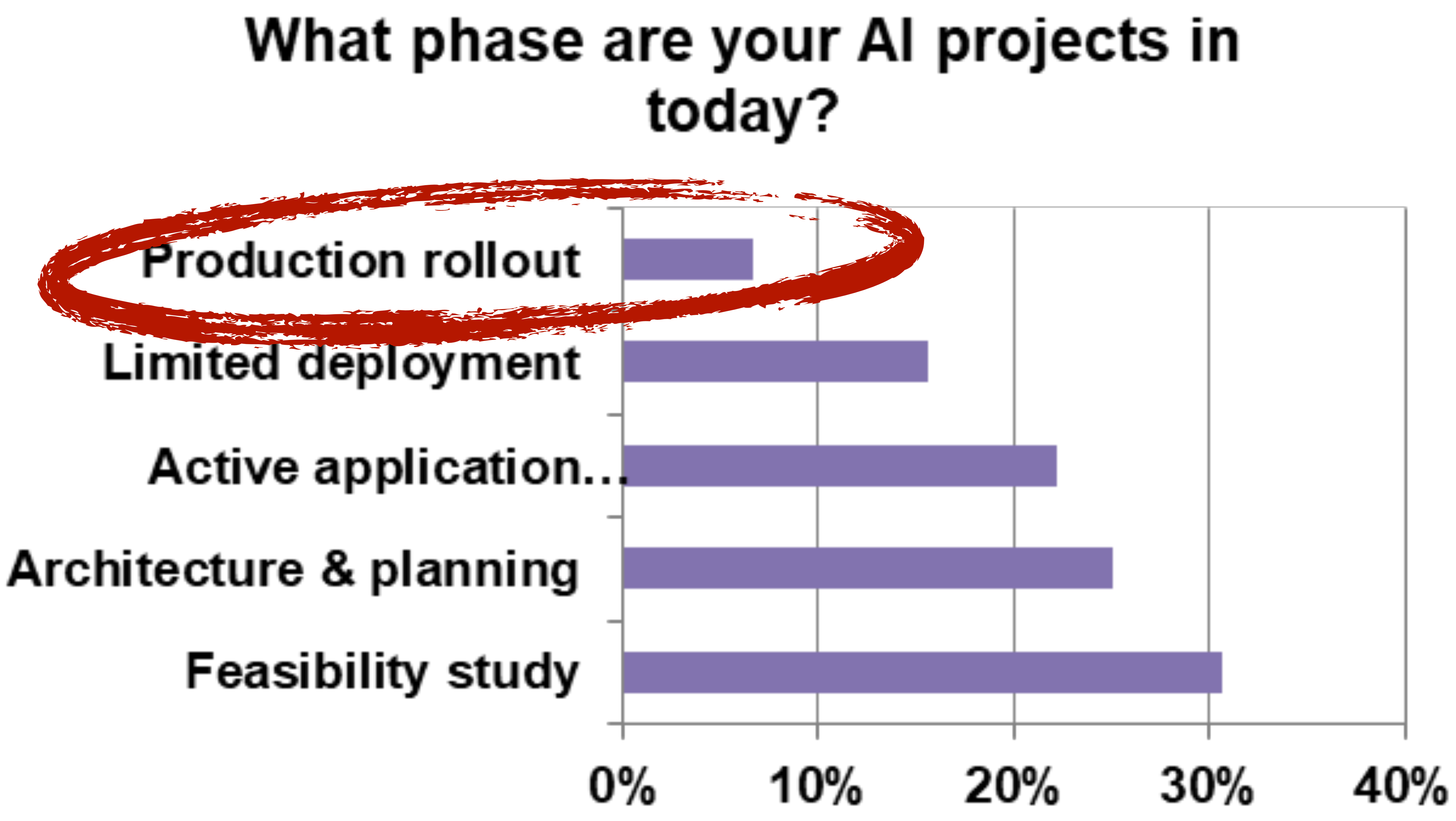
Outline

- Introduction
  - History and terminology
  - What's possible, what's on the horizon, what's unknown
- Developing AI Products
  - **Picking the problem**
  - Data Flywheel
  - Most AI code is not AI
  - Roles and Hiring
- Example
- Q & A

It's Early Days!

2018 survey of 300+ ML-involved developers



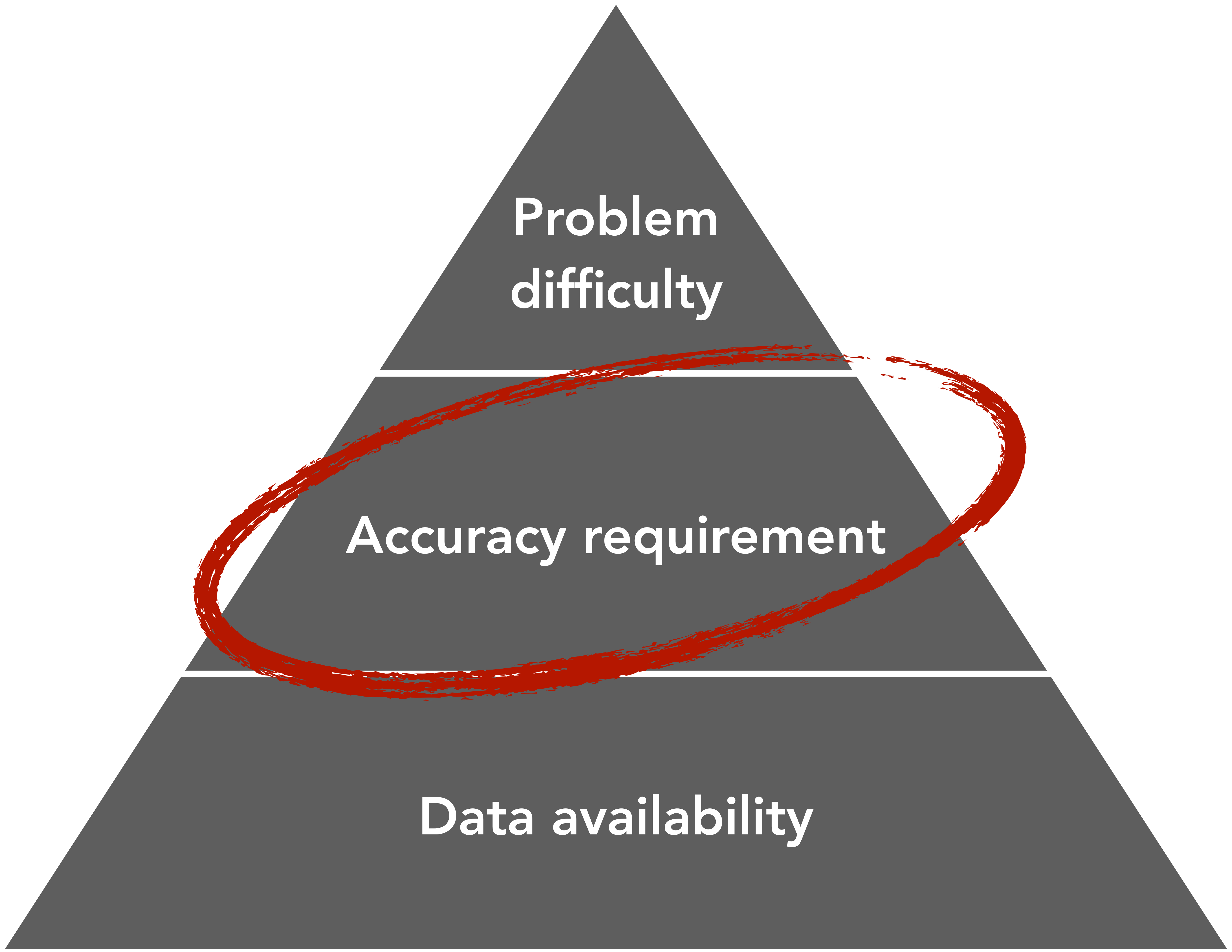


https://www.nextplatform.com/2018/04/24/lagging-in-ai-dont-worry-its-still-early/

Evaluating a problem

1. Is automated prediction valuable?
2. Is the required level of accuracy feasible?
3. Will it move the needle for your business?
4. Will you be able to build a data moat?

Cost drivers



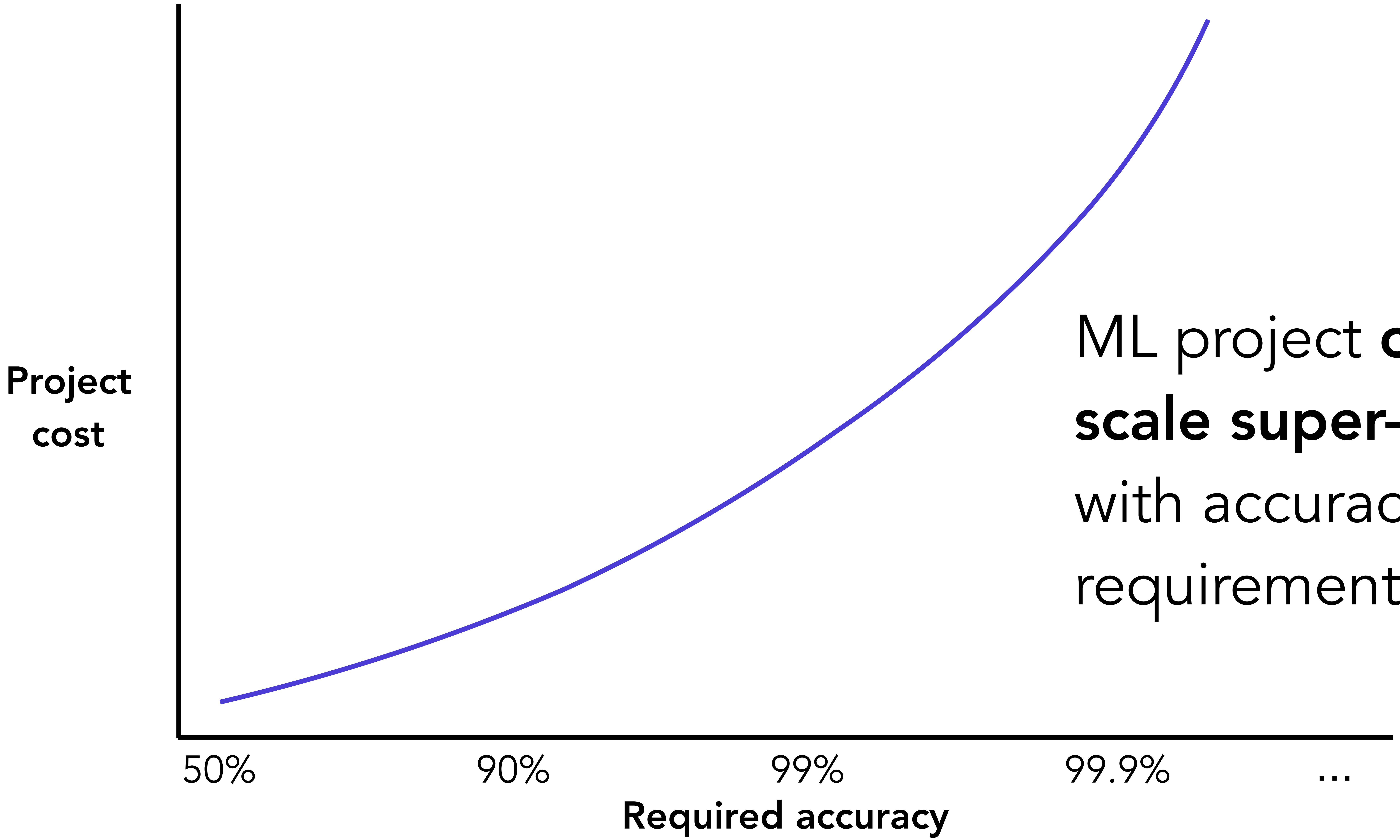
Main considerations

- Good published work on similar problems? (newer problems mean more risk & more technical effort)

- How costly are wrong predictions?
- How frequently does the system need to be right to be useful?

- How hard is it to acquire enough data?
- How expensive is data labeling?
- Unique advantage?

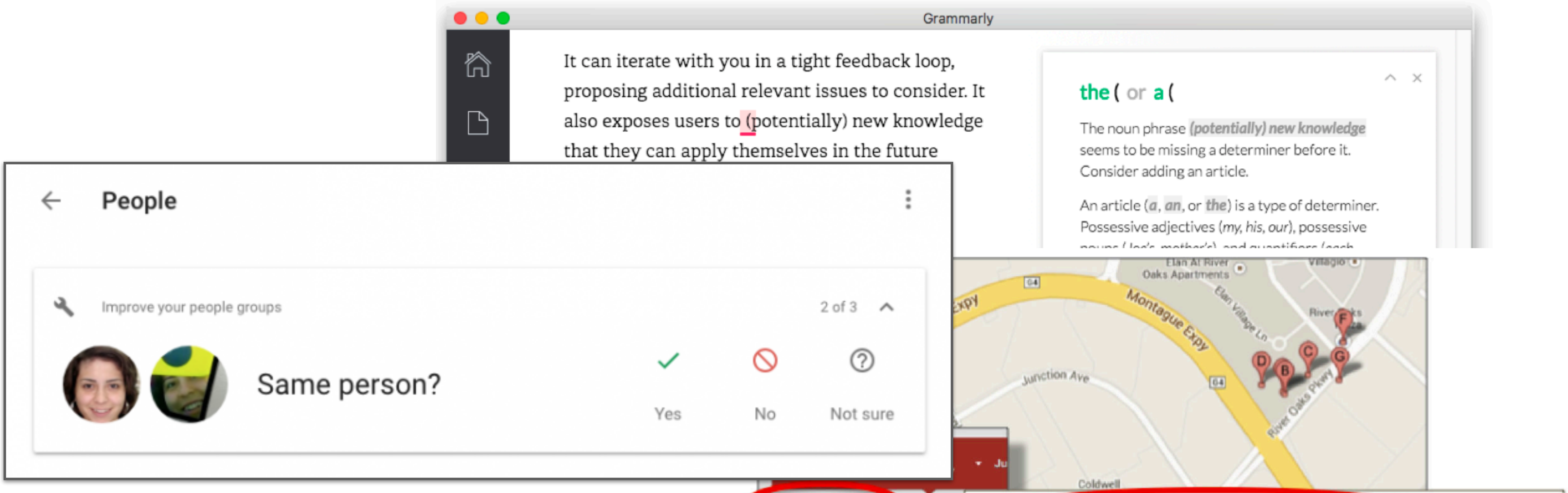
Diagram courtesy of Josh Tobin



ML project **costs scale super-linearly** with accuracy requirement

Diagram courtesy of Josh Tobin

Product Design Can Reduce Accuracy Requirements



See "Designing Collaborative AI" (Ben Reinhardt and Belmer Negrillo):
https://medium.com/@Ben_Reinhardt/designing-collaborative-ai-5c1e8dbc8810

Outline

- Introduction
    - History and terminology
    - What's possible, what's on the horizon, what's unknown
- Developing AI Products
    - Picking the problem
    - **Data Flywheel**
    - Most AI code is not AI
    - Roles and Hiring
- Example
- Q & A

Defensible AI means proprietary dataset

PYTORCH

Data is cheap

Labels are expensive

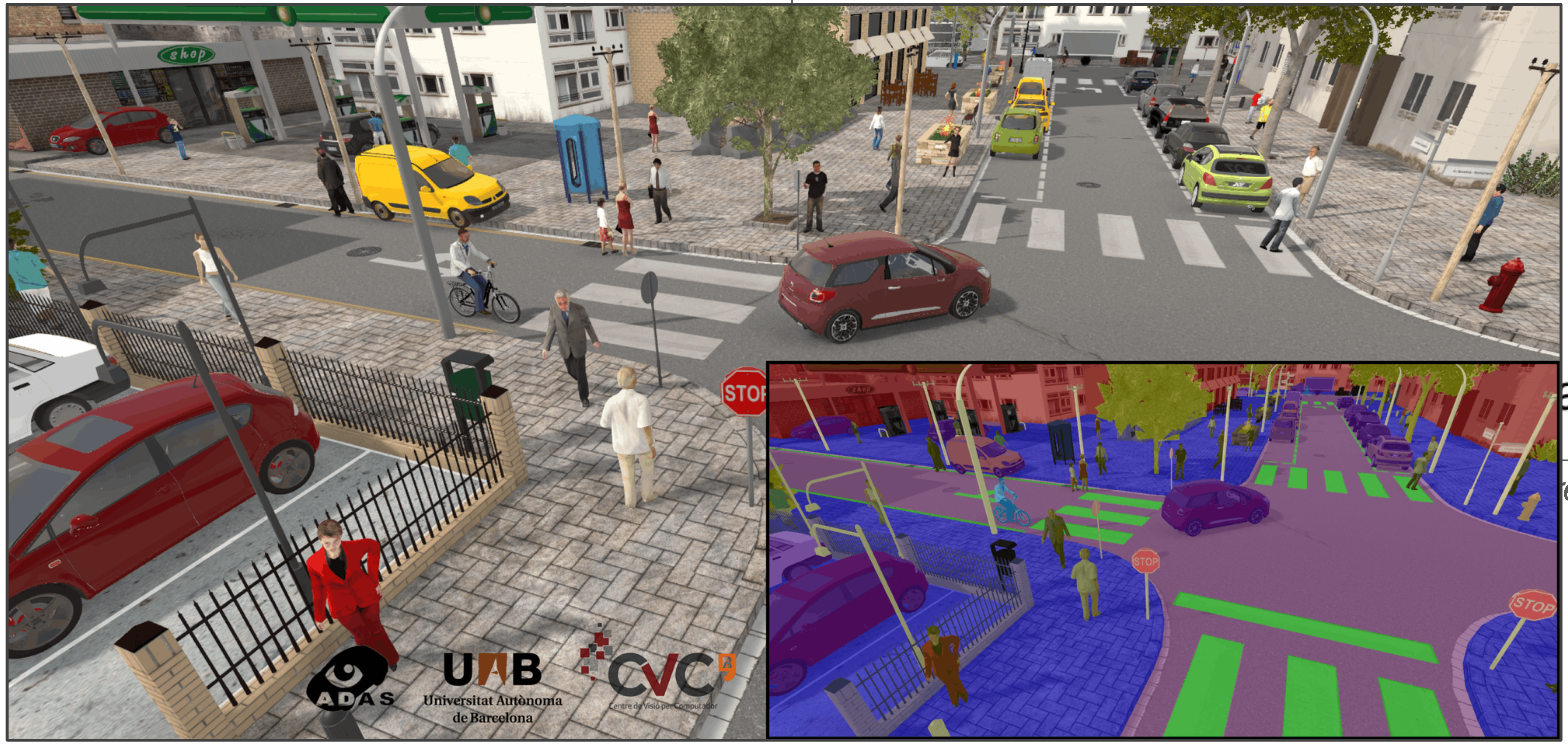
Usually: spend time and $

https://cdn-sv1.deepsense.ai/wp-content/uploads/2017/04/sample_image_from_the_training_set.jpg

Underrated: Synthetic Data

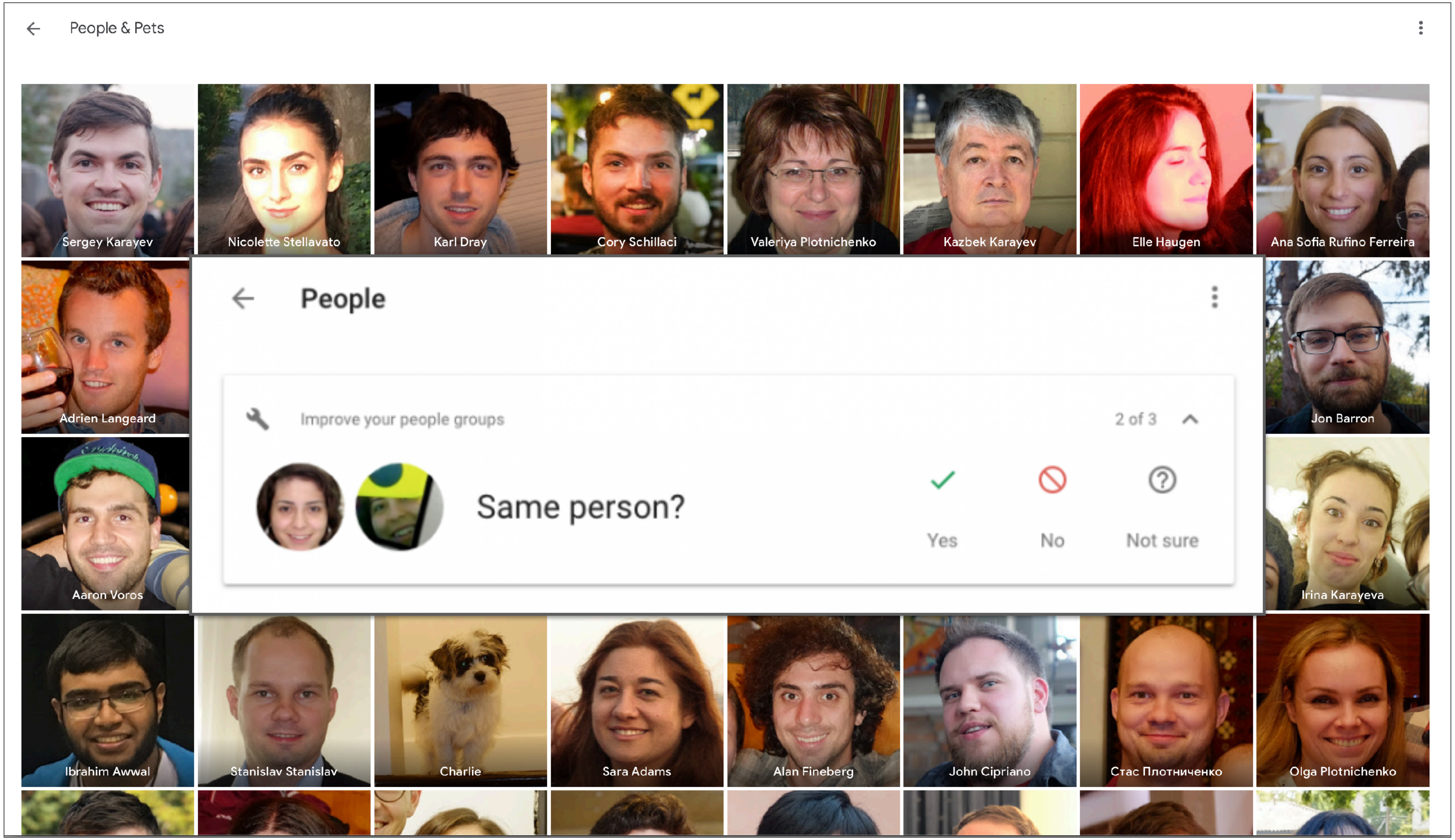
ally generated word images

eating-a-modern-ocr-pipeline-using-computer-vision-and-deep-learning/



https://newatlas.com/synthia-dataset-self-driving-cars/43895/

Ideal: Data Flywheel



“We believe a new, more powerful, business model has evolved from its software predecessor. These companies structure their business processes to put continuously learning models, built on “closed loop” data, at the center of what they do.

When built right, they create a reinforcing cycle: Their products get better, allowing them to collect more data, which allows them to build better models, making their products better, and onward.

If software ate the world, models will run it."

- Wall Street Journal, Jan. 21, 2019, “Models Will Run the World”

Not Flywheel-Driven

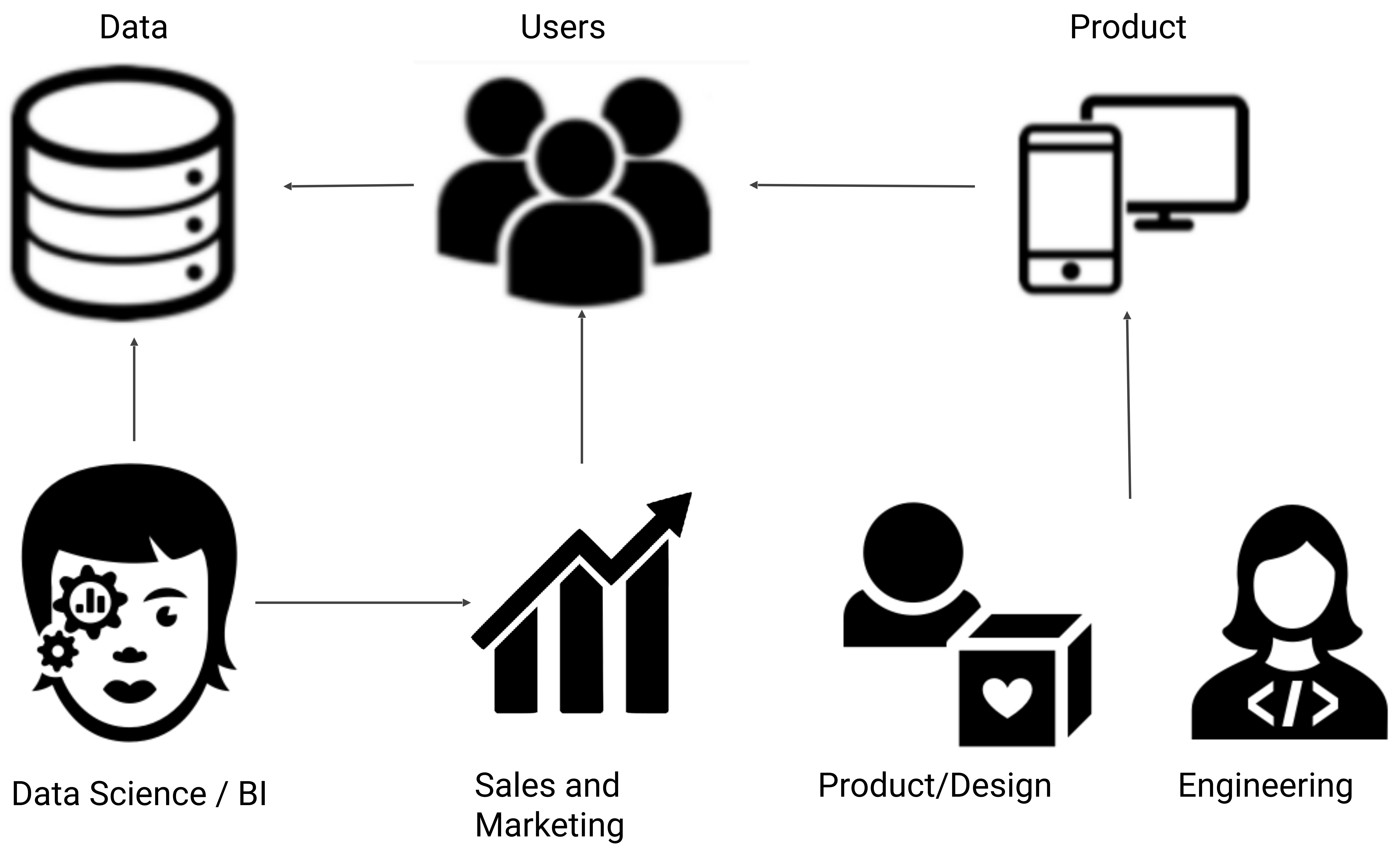


Diagram courtesy of Eric Wang

Flywheel-driven

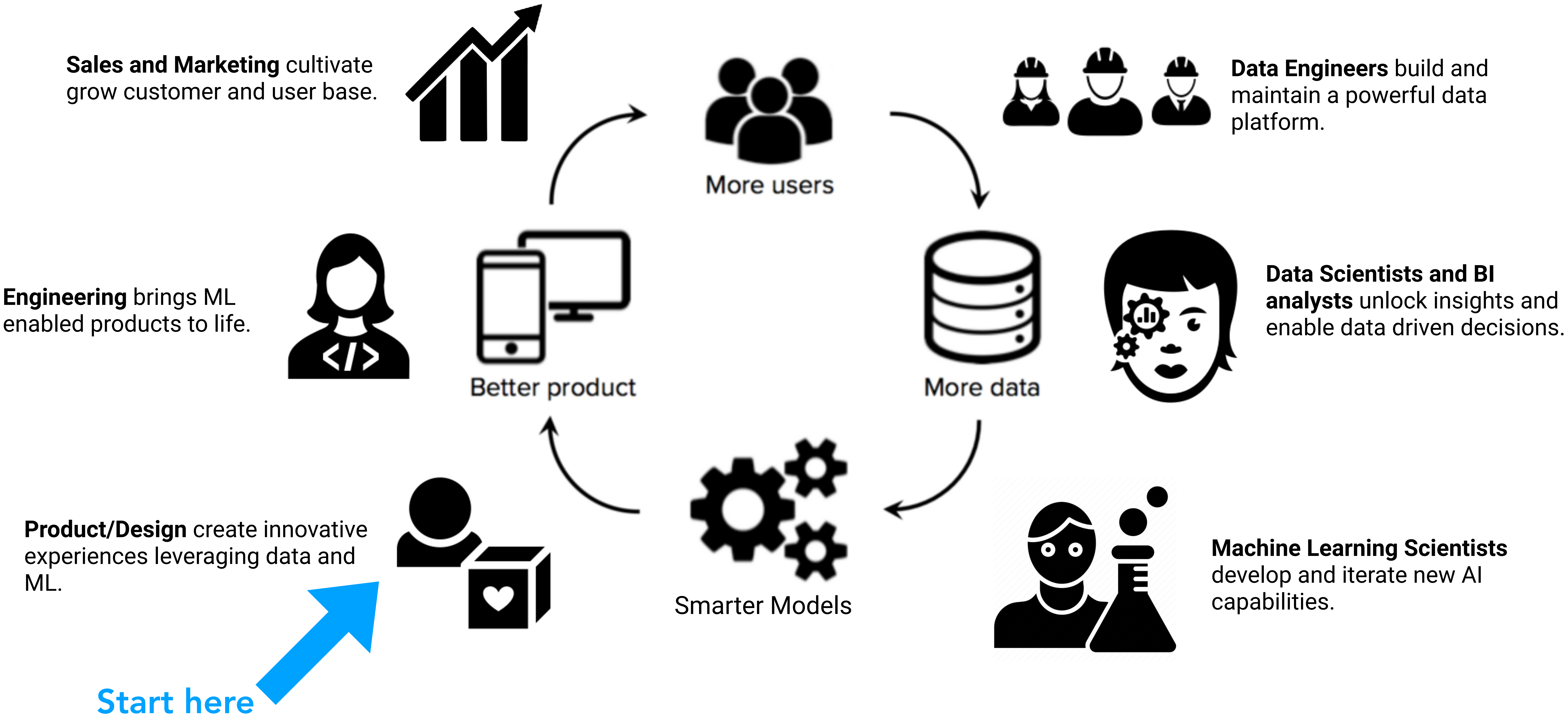
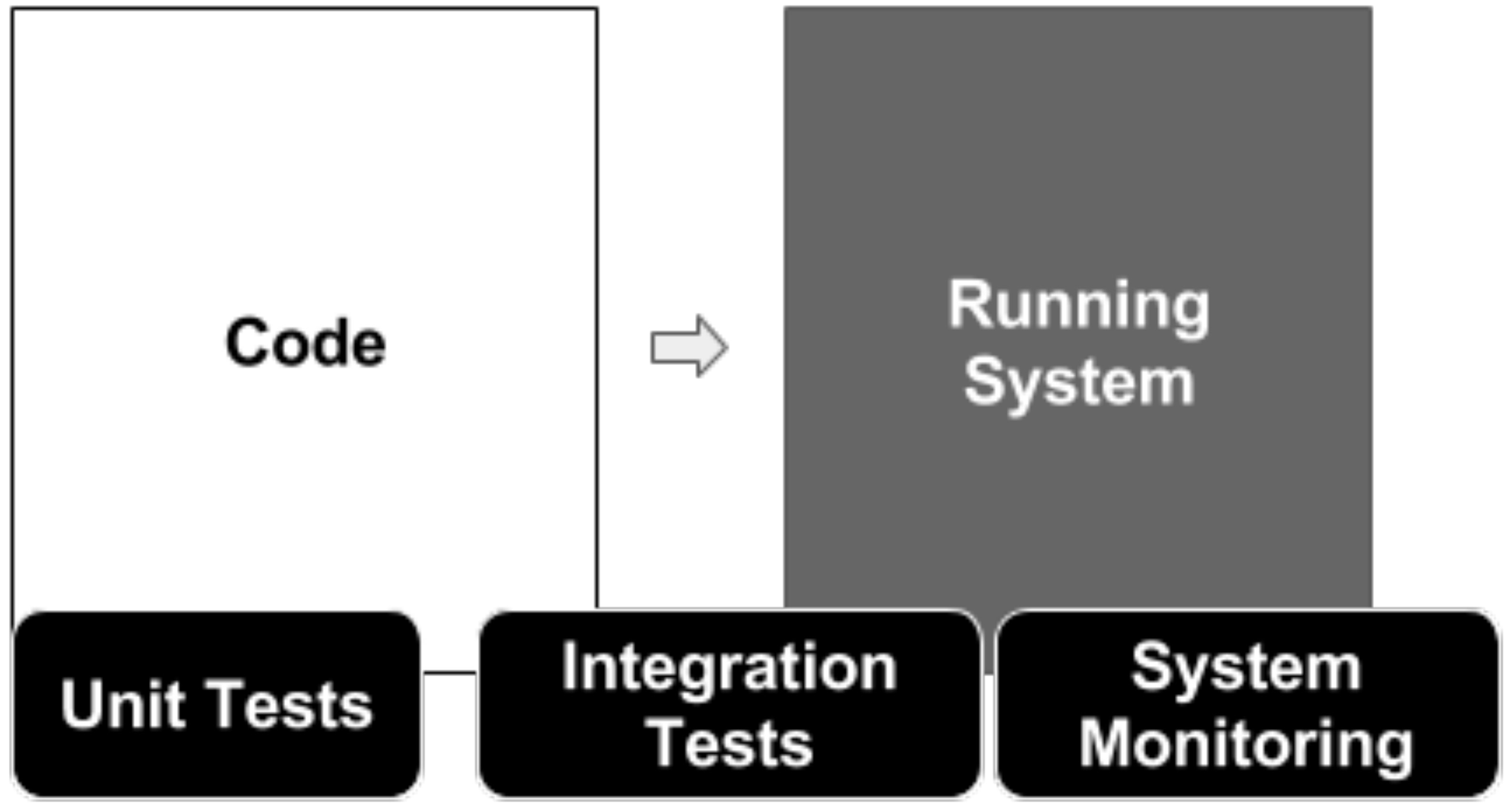


Diagram courtesy of Eric Wang

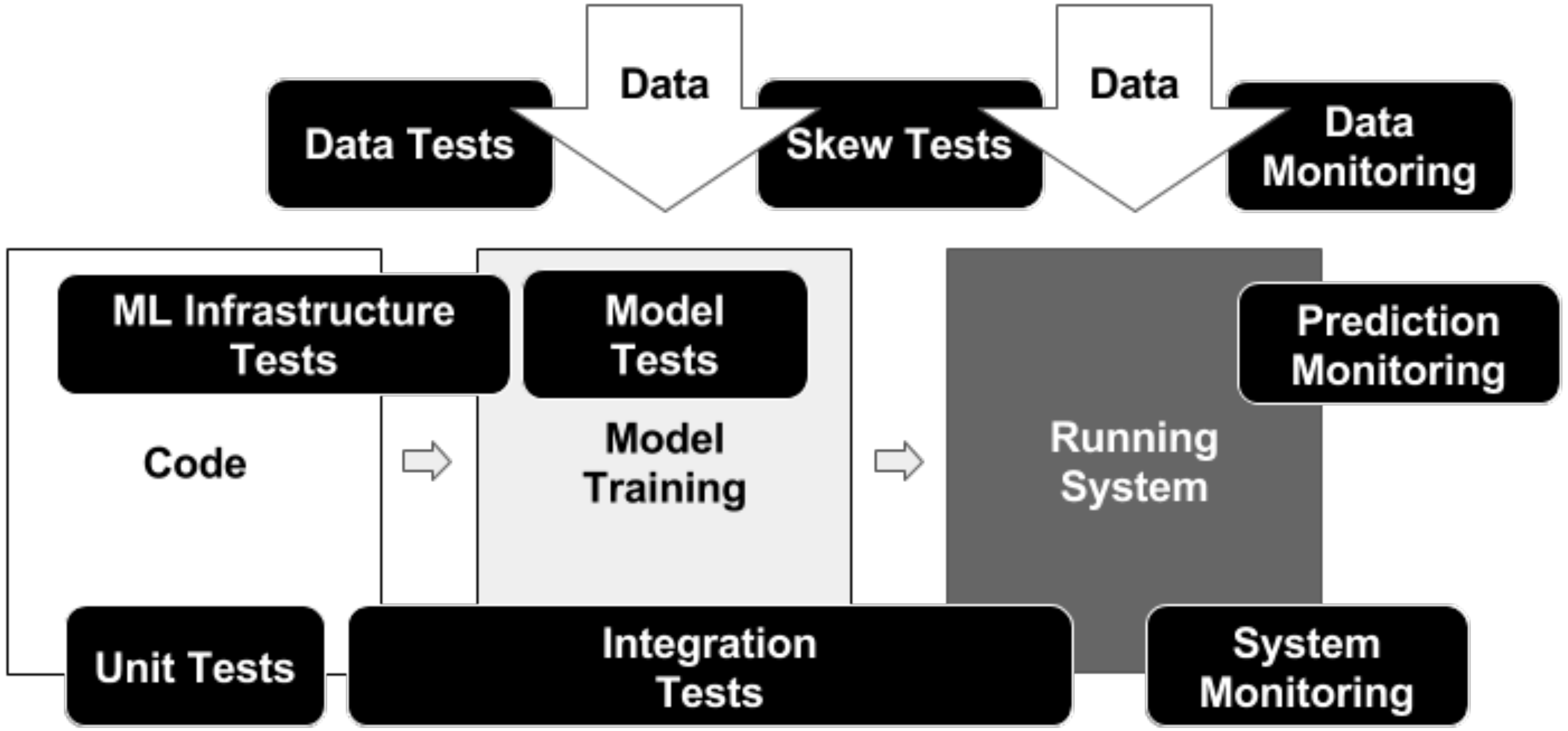
Outline

- Introduction
  - History and terminology
  - What's possible, what's on the horizon, what's unknown
- Developing AI Products
  - Picking the problem
  - Data Flywheel
  - **Most AI code is not AI**
  - Roles and Hiring
- Example
- Q & A

Traditional Software



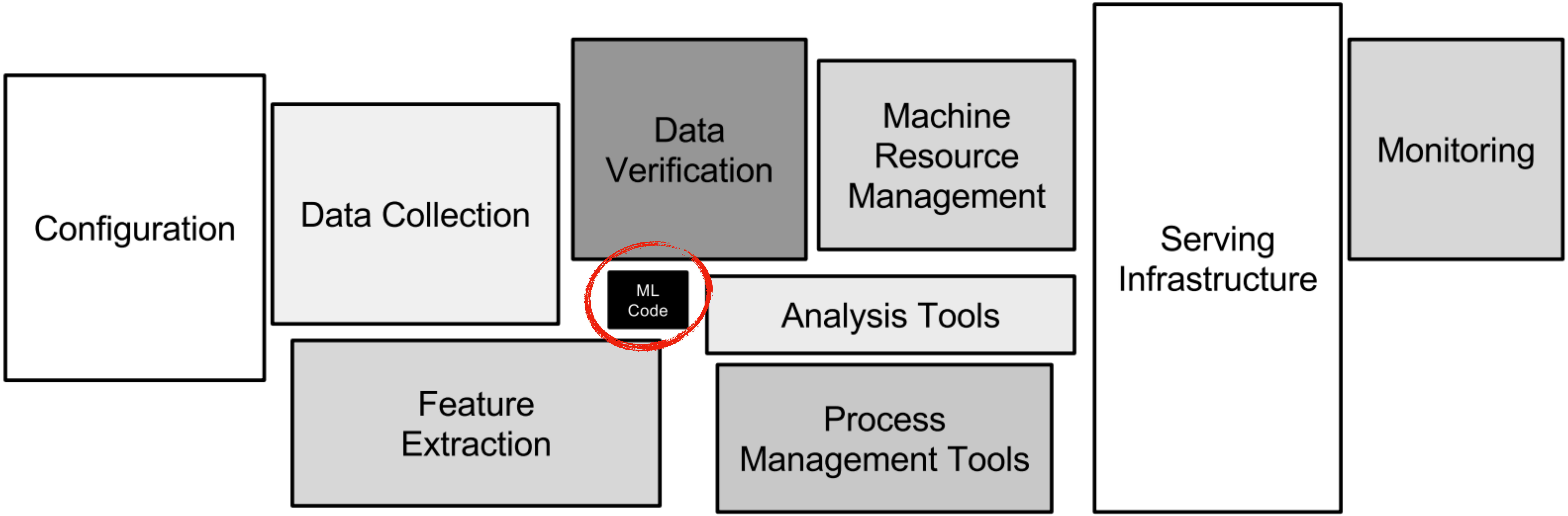
Machine Learning Software



Machine Learning: The High-Interest Credit Card of Technical Debt

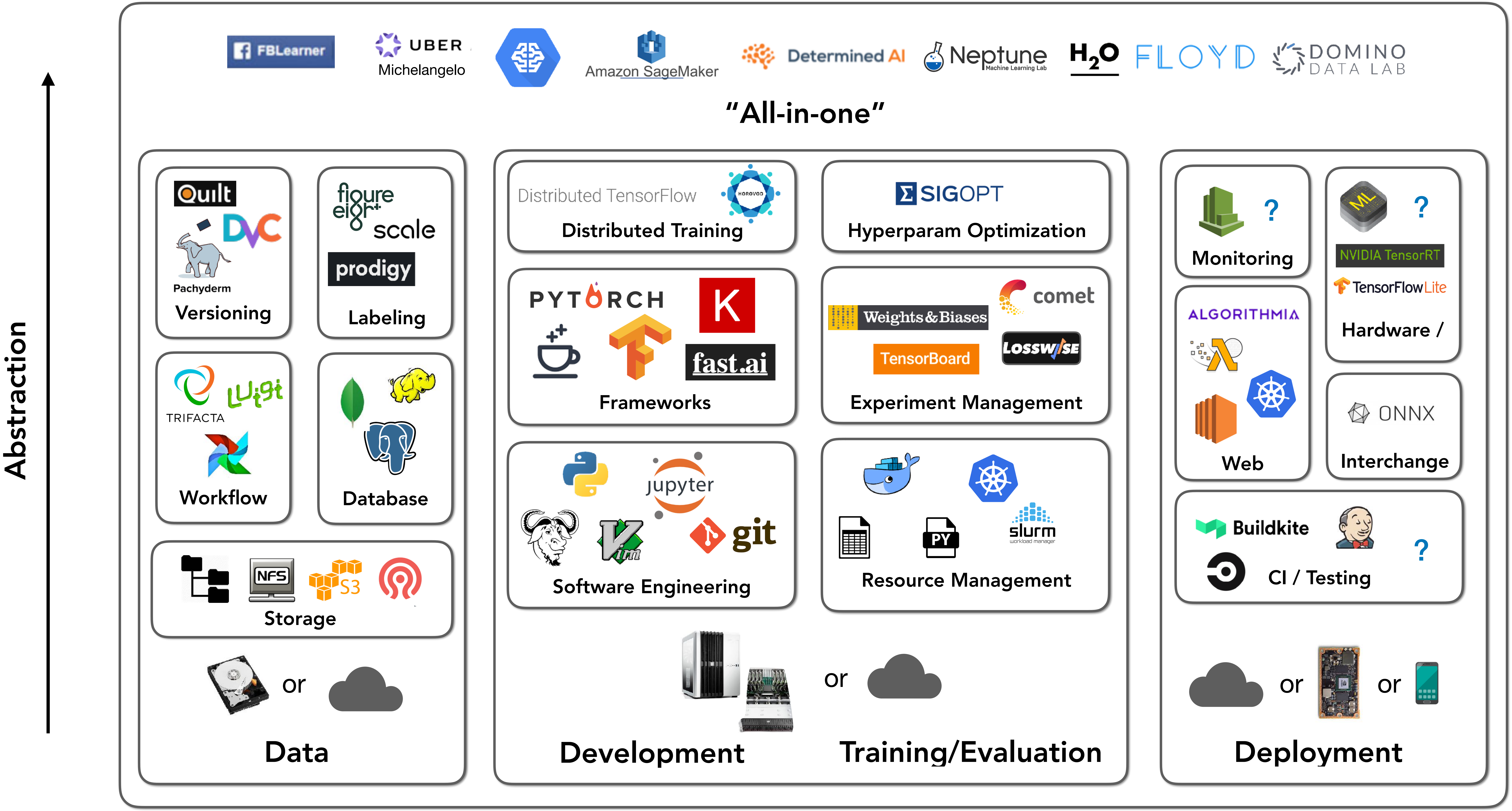
D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov,
Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young
{dsculley,gholt,dgg,edavydov}@google.com
{toddphillips,ebner,vchaudhary,mwyoung}@google.com
Google, Inc

SE4ML: Software Engineering for Machine Learning (NIPS 2014 Workshop)



Machine Learning: The High-Interest Credit Card of Technical Debt

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov,
Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young

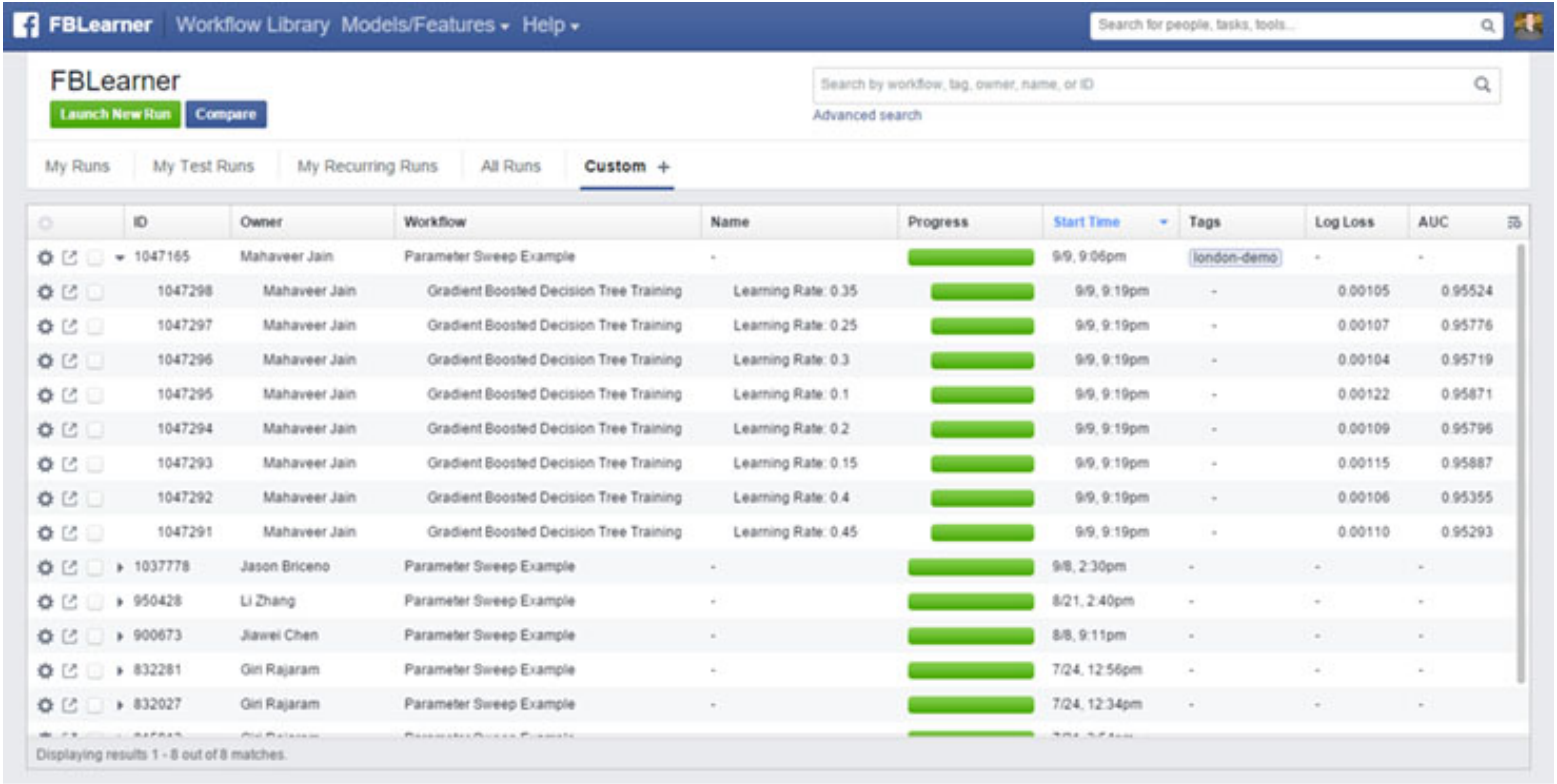


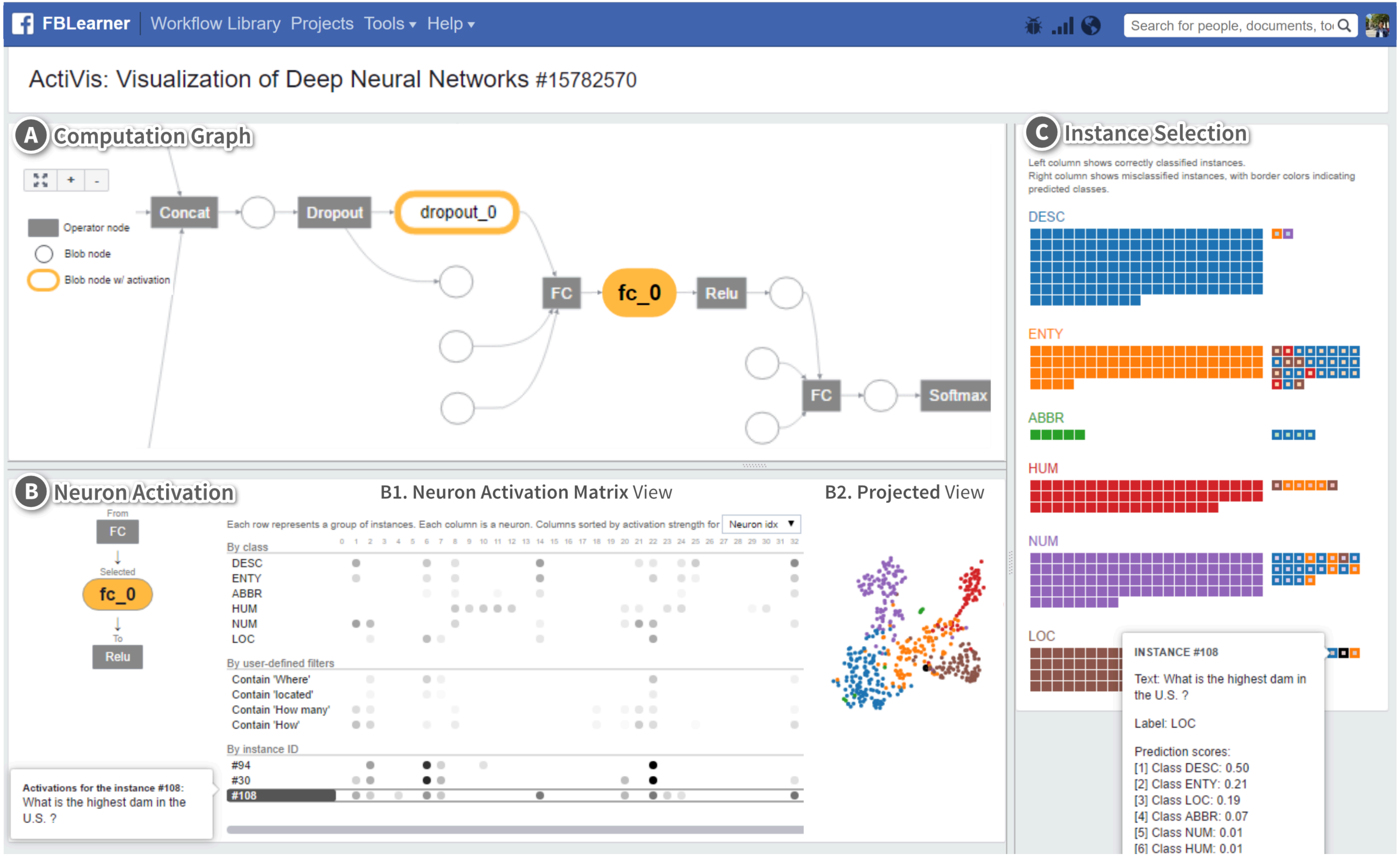
ASU GSV 2019 - AI Masterclass - Sergey Karayev

AI Development Workflow

POSTED ON MAY 9, 2016 TO AI RESEARCH, APPLIED MACHINE LEARNING, CORE DATA

Introducing FBLearner Flow: Facebook's AI backbone


FBLearner
Workflow Library
Models/Features
Help
Search for people, tasks, tools...
FBLearner
Launch New Run
Compare
Search by workflow, tag, owner, name, or ID
Advanced search
My Runs
My Test Runs
My Recurring Runs
All Runs
Custom +
ID | Owner | Workflow | Name | Progress | Start Time | Tags | Log Loss | AUC
1047165 | Mahaveer Jain | Parameter Sweep Example | - | | 9/9, 9:06pm | london-demo | - | -
1047298 | Mahaveer Jain | Gradient Boosted Decision Tree Training | Learning Rate: 0.35 | | 9/9, 9:19pm | - | 0.00105 | 0.95524
1047297 | Mahaveer Jain | Gradient Boosted Decision Tree Training | Learning Rate: 0.25 | | 9/9, 9:19pm | - | 0.00107 | 0.95776
1047296 | Mahaveer Jain | Gradient Boosted Decision Tree Training | Learning Rate: 0.3 | | 9/9, 9:19pm | - | 0.00104 | 0.95719
1047295 | Mahaveer Jain | Gradient Boosted Decision Tree Training | Learning Rate: 0.1 | | 9/9, 9:19pm | - | 0.00122 | 0.95871
1047294 | Mahaveer Jain | Gradient Boosted Decision Tree Training | Learning Rate: 0.2 | | 9/9, 9:19pm | - | 0.00109 | 0.95796
1047293 | Mahaveer Jain | Gradient Boosted Decision Tree Training | Learning Rate: 0.15 | | 9/9, 9:19pm | - | 0.00115 | 0.95887
1047292 | Mahaveer Jain | Gradient Boosted Decision Tree Training | Learning Rate: 0.4 | | 9/9, 9:19pm | - | 0.00106 | 0.95355
1047291 | Mahaveer Jain | Gradient Boosted Decision Tree Training | Learning Rate: 0.45 | | 9/9, 9:19pm | - | 0.00110 | 0.95293
1037778 | Jason Briceno | Parameter Sweep Example | - | | 9/8, 2:30pm | - | - | -
950428 | Li Zhang | Parameter Sweep Example | - | | 8/21, 2:40pm | - | - | -
900673 | Jiawei Chen | Parameter Sweep Example | - | | 8/8, 9:11pm | - | - | -
832281 | Giri Rajaram | Parameter Sweep Example | - | | 7/24, 12:56pm | - | - | -
832027 | Giri Rajaram | Parameter Sweep Example | - | | 7/24, 12:34pm | - | - | -
Displaying results 1 - 8 out of 8 matches.


FBLearner
Workflow Library
Projects
Tools
Help
Search for people, documents, to
ActiVis: Visualization of Deep Neural Networks #15782570
A Computation Graph
Operator node
Blob node
Blob node w/ activation
Concat
Dropout
dropout_0
FC
fc_0
Relu
FC
Softmax
C Instance Selection
Left column shows correctly classified instances.
Right column shows misclassified instances, with border colors indicating predicted classes.
DESC
ENTY
ABBR
HUM
NUM
LOC
B Neuron Activation
B1. Neuron Activation Matrix View
B2. Projected View
From
FC
Selected
fc_0
To
Relu
Each row represents a group of instances. Each column is a neuron. Columns sorted by activation strength for
Neuron idx
By class
DESC
ENTY
ABBR
HUM
NUM
LOC
By user-defined filters
Contain 'Where'
Contain 'located'
Contain 'How many'
Contain 'How'
By instance ID
#94
#30
#108
Activations for the instance #108:
What is the highest dam in the U.S. ?
INSTANCE #108
Text: What is the highest dam in the U.S. ?
Label: LOC
Prediction scores:
[1] Class DESC: 0.50
[2] Class ENTY: 0.21
[3] Class LOC: 0.19
[4] Class ABBR: 0.07
[5] Class NUM: 0.01
[6] Class HUM: 0.01

Outline

- Introduction
    - History and terminology
    - What's possible, what's on the horizon, what's unknown
- Developing AI Products
    - Picking the problem
    - Data Flywheel
    - Most AI code is not AI
    - **Roles and Hiring**
- Example
- Q & A

Breakdown of roles

Role	Job Function	Work product
Data engineer	Build data pipelines, aggregation, storage	Data systems
ML Engineer	Train & deploy production-grade prediction models	Overall prediction system
ML Researcher	Develop research-grade prediction models	New models
Data Scientist	Bad orgs: all of the above Good orgs: answer business questions with data	Reports

What skills are needed for the roles?

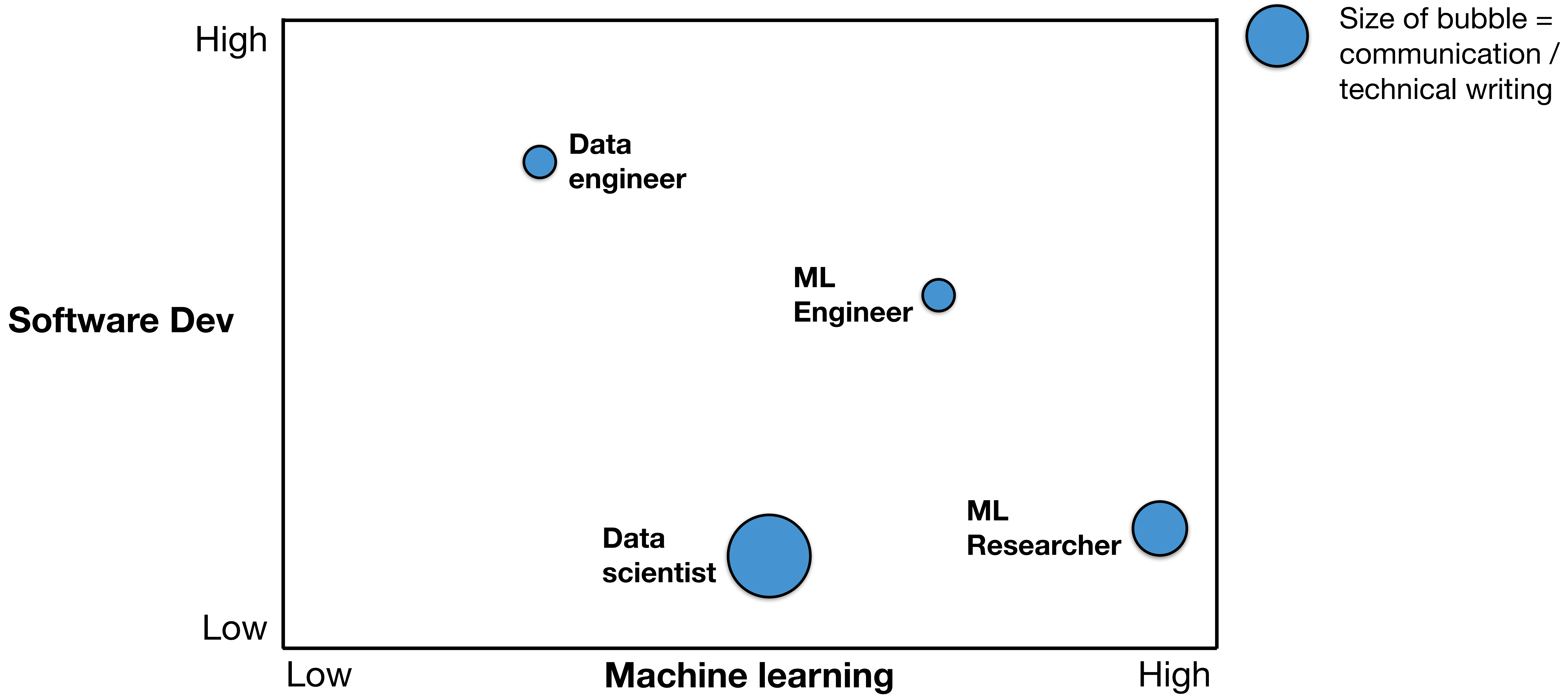


Diagram courtesy of Josh Tobin

What skills are needed for the roles?

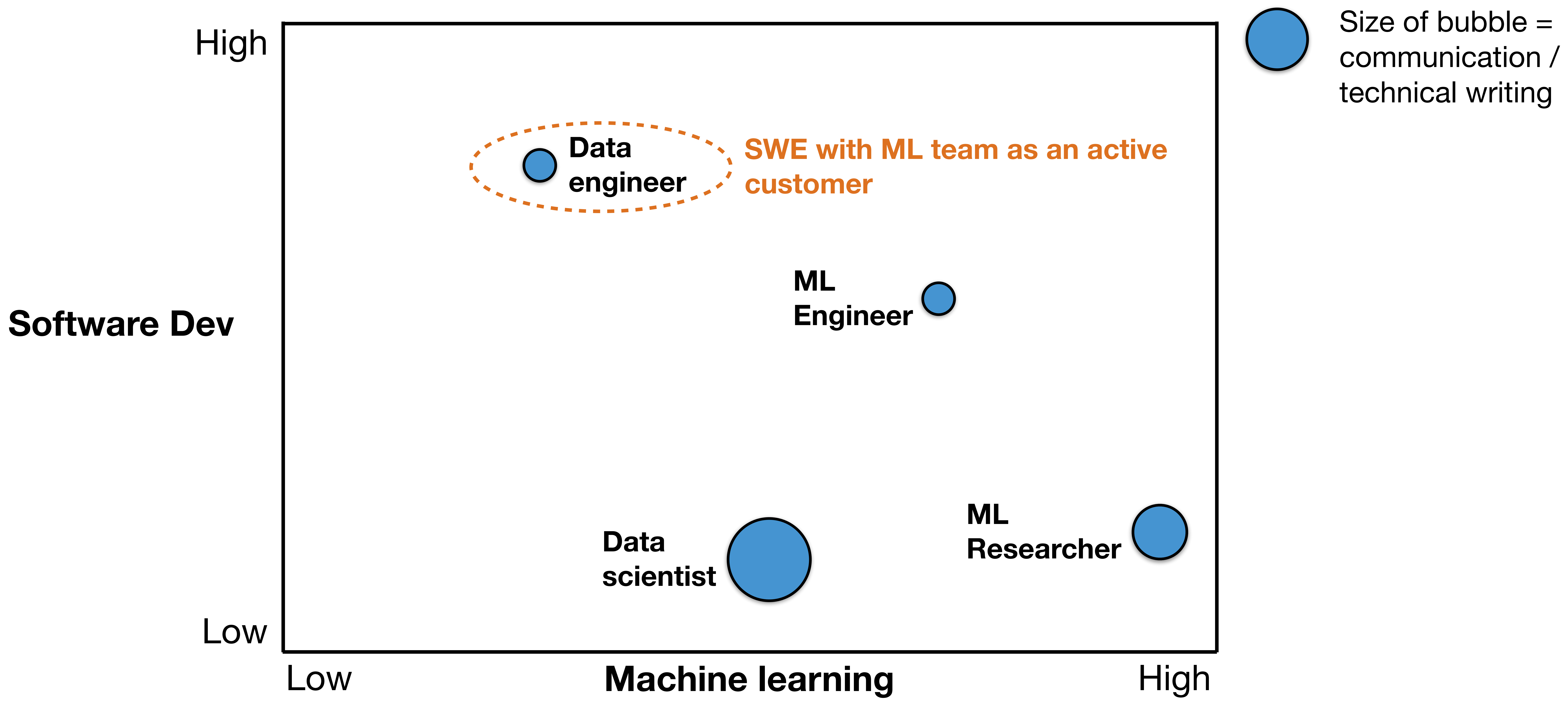


Diagram courtesy of Josh Tobin

What skills are needed for the roles?

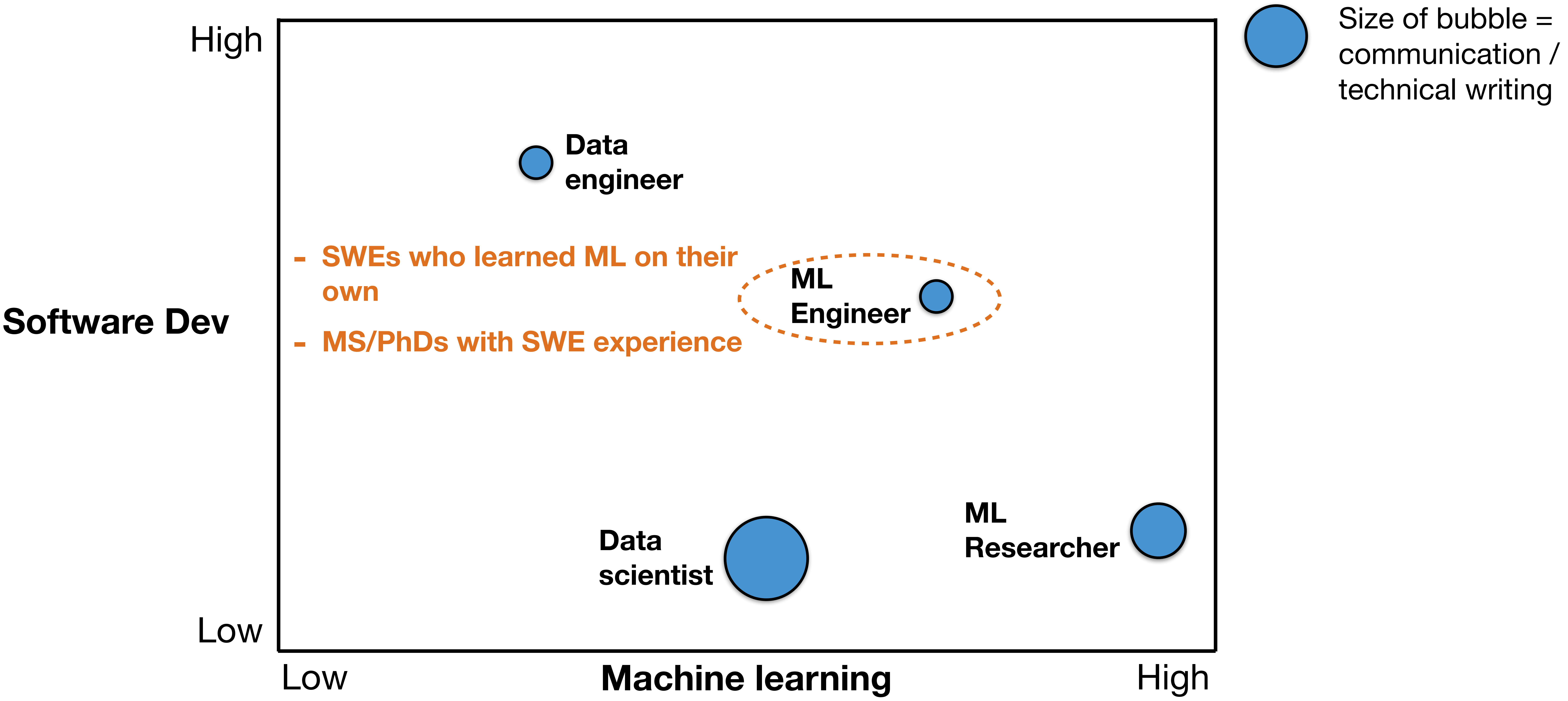


Diagram courtesy of Josh Tobin

What skills are needed for the roles?

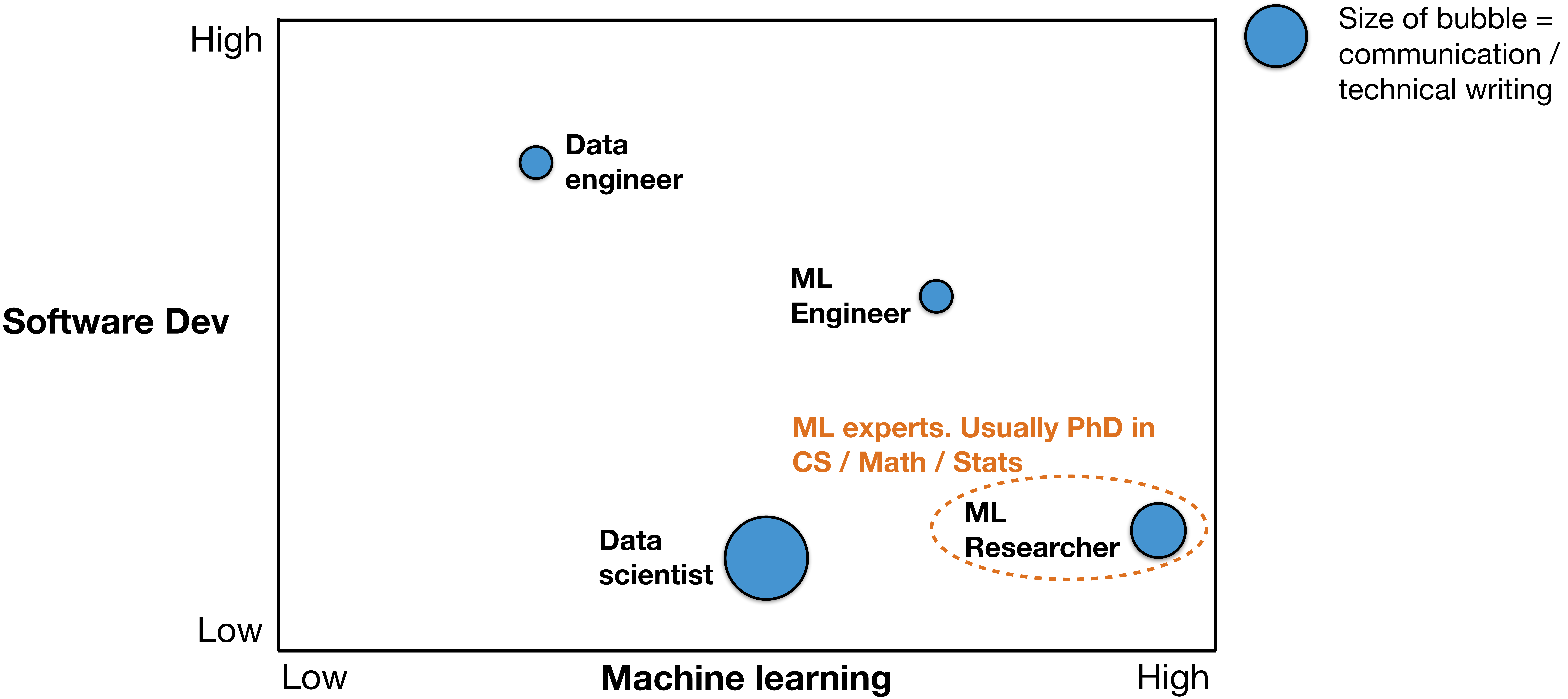


Diagram courtesy of Josh Tobin

What skills are needed for the roles?

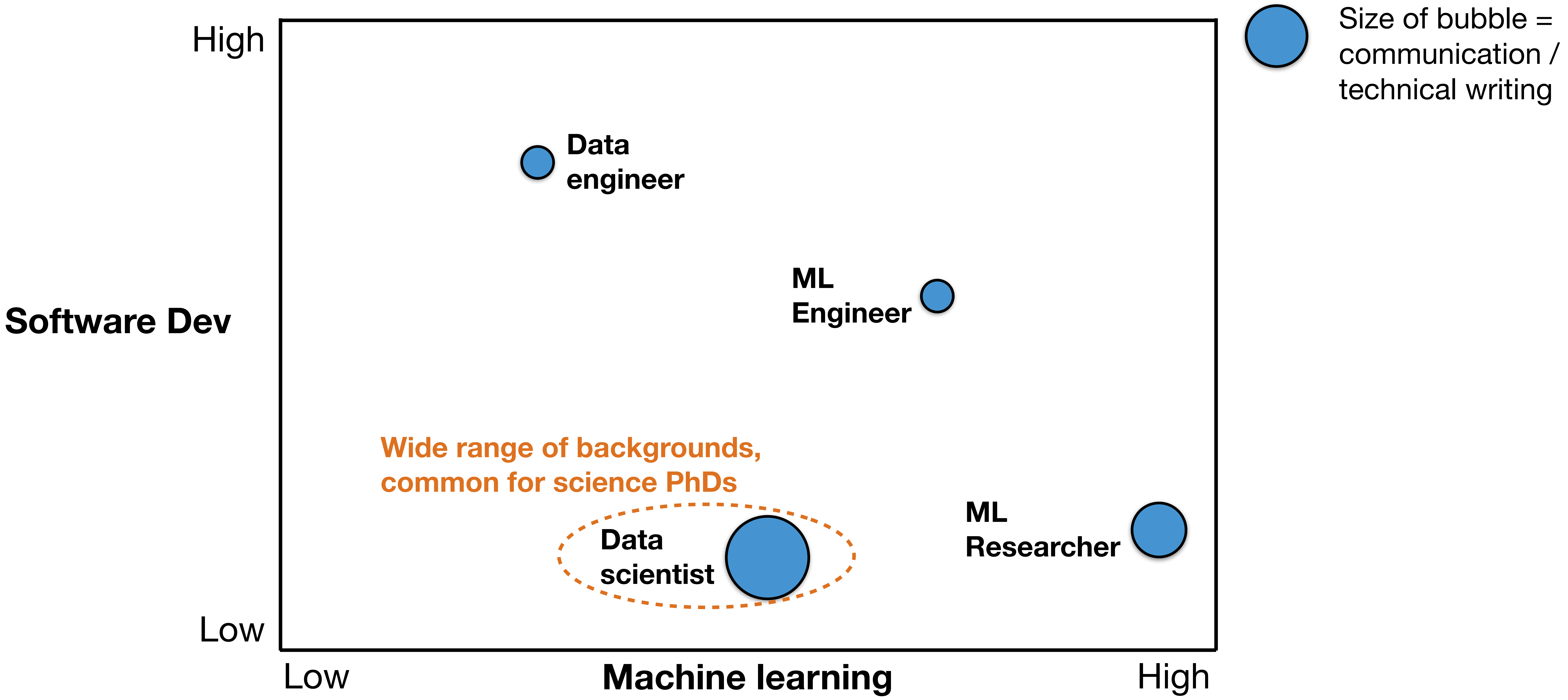


Diagram courtesy of Josh Tobin

Hiring

Role	Where to hire
Data engineer	People with "Big Data" experience, ops-focused SWEs
ML Engineer	SWEs that took ML courses, MS/PhDs interested in software dev
ML Researcher	Hardest one. Huge talent gap!
Data Scientist	Non-CS PhD with a data science project they can show.

Outline

- Introduction
  - History and terminology
  - What's possible, what's on the horizon, what's unknown
- Developing AI Products
  - Picking the problem
  - Data Flywheel
  - Most AI code is not AI
  - Roles and Hiring
- Example
- **Q & A**