FSDL 2022

Introduction to the Course

Josh Tobin

August 8, 2022

Agenda

00

Course Vision

Why ML-powered products and why this course

01

When to Use ML

Should you use ML? How do you know if you're ready?

02

Lifecycle

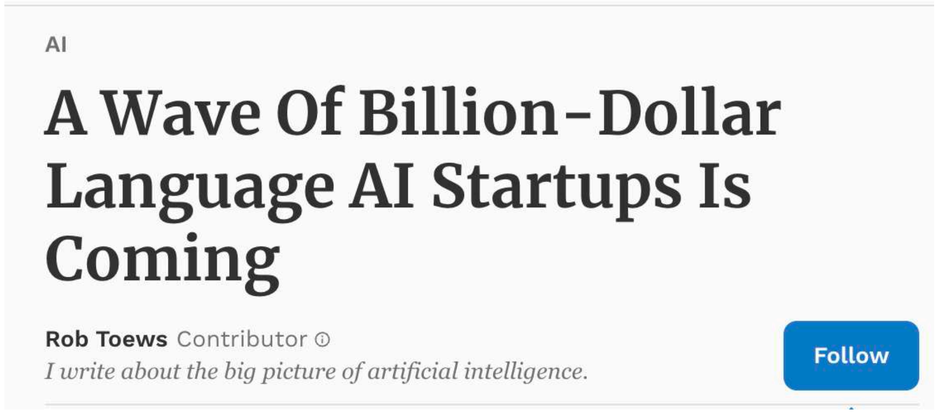
What are the steps to a ML project?

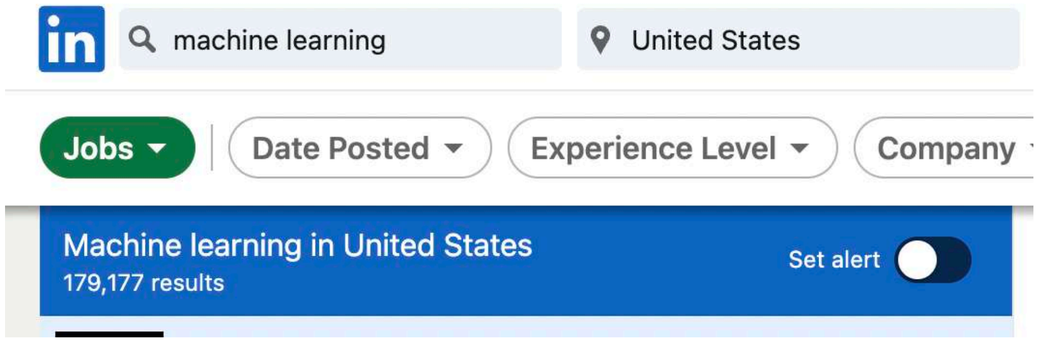
00

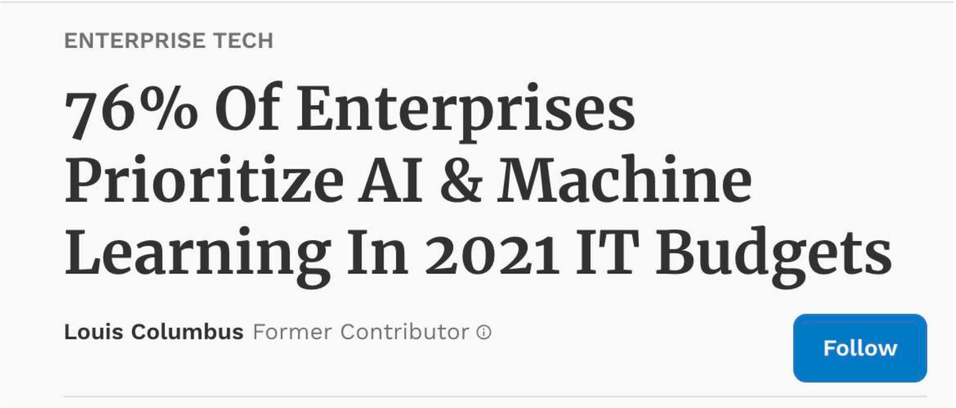
Course Vision

The course (and community) for people building *ML-powered products*

ML is becoming a mainstream technology



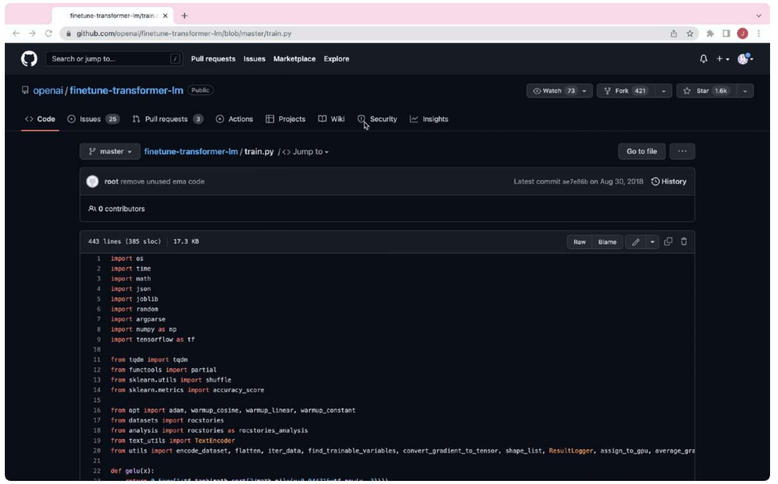




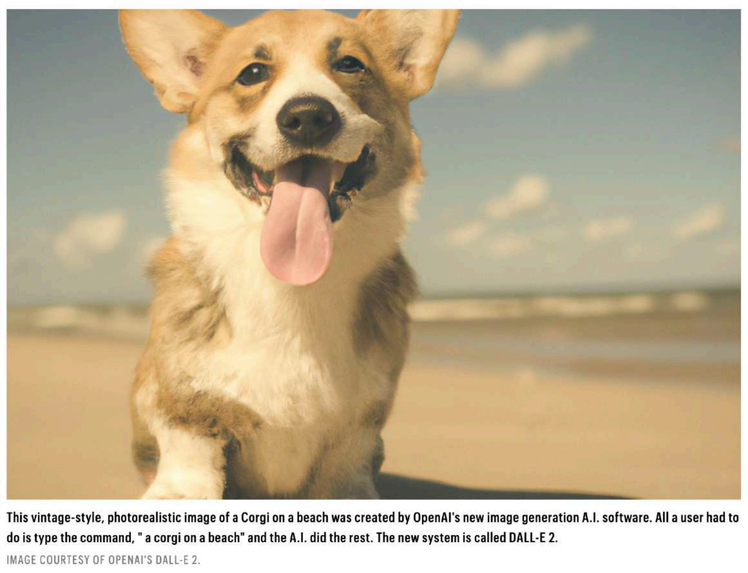
ML in 2018 (the first FSDL class)

Improving Language Understanding by Generative Pre-Training

Alec Radford OpenAI alec@openai.com
Karthik Narasimhan OpenAI karthikn@openai.com
Tim Salimans OpenAI tim@openai.com
Ilya Sutskever OpenAI ilyasu@openai.com

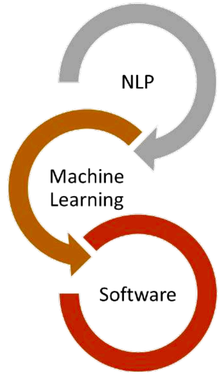


ML in 2022

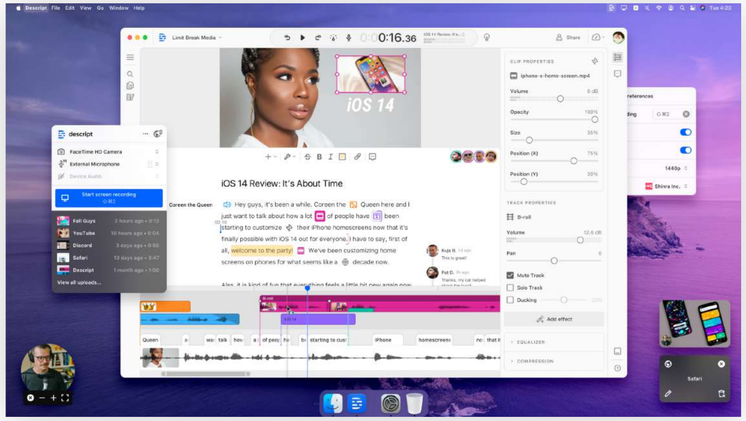


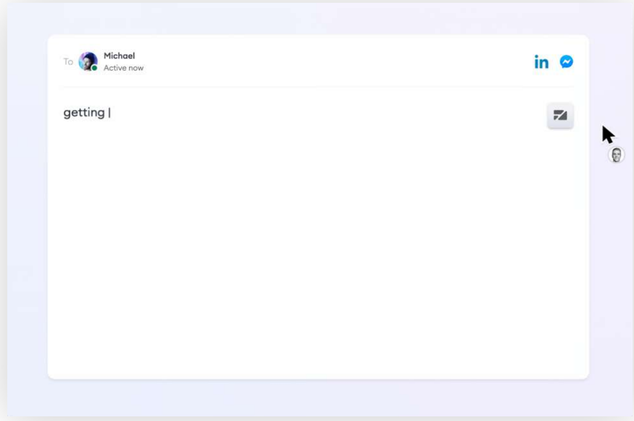
This vintage-style, photorealistic image of a Corgi on a beach was created by OpenAI's new image generation A.I. software. All a user had to do is type the command, " a corgi on a beach" and the A.I. did the rest. The new system is called DALL-E 2.

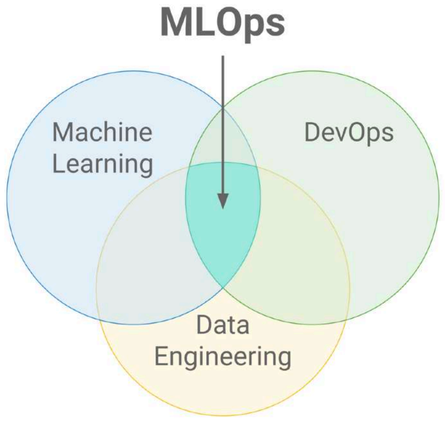
IMAGE COURTESY OF OPENAI'S DALL-E 2.



Software Is Eating the World, Machine Learning is Eating Software and NLP is Eating Machine Learning
Jesus Rodriguez[1]

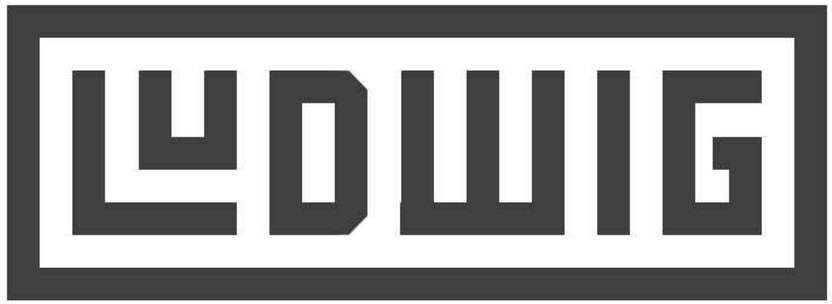


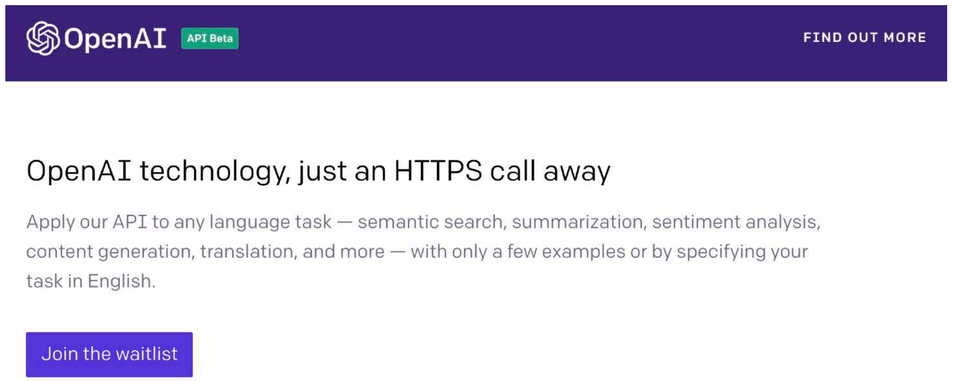




1. https://jrodthoughts.medium.com/software-is-eating-the-world-machine-learning-is-eating-software-and-nlp-is-eating-machine-bcb49d770938

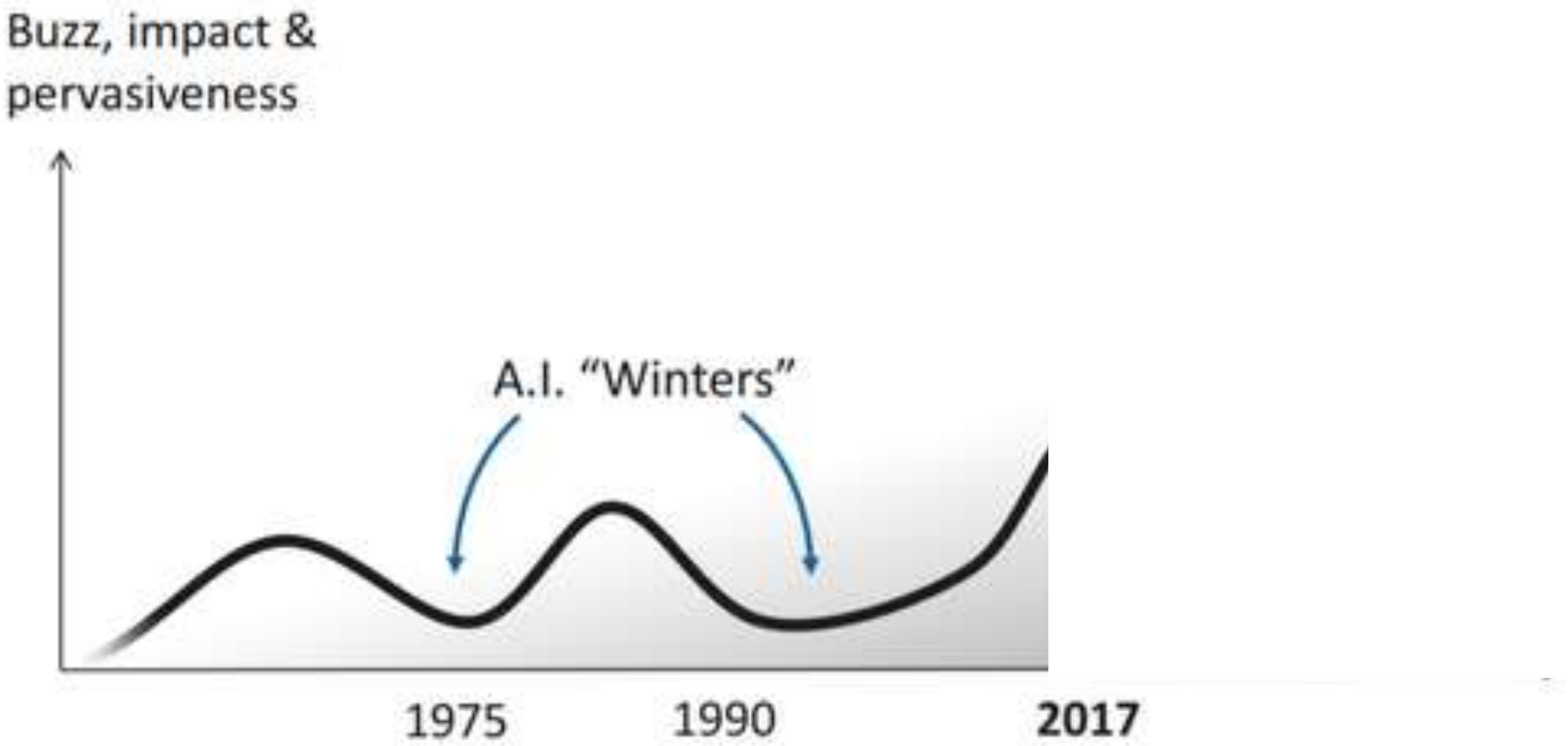
Why? Commoditization of model training





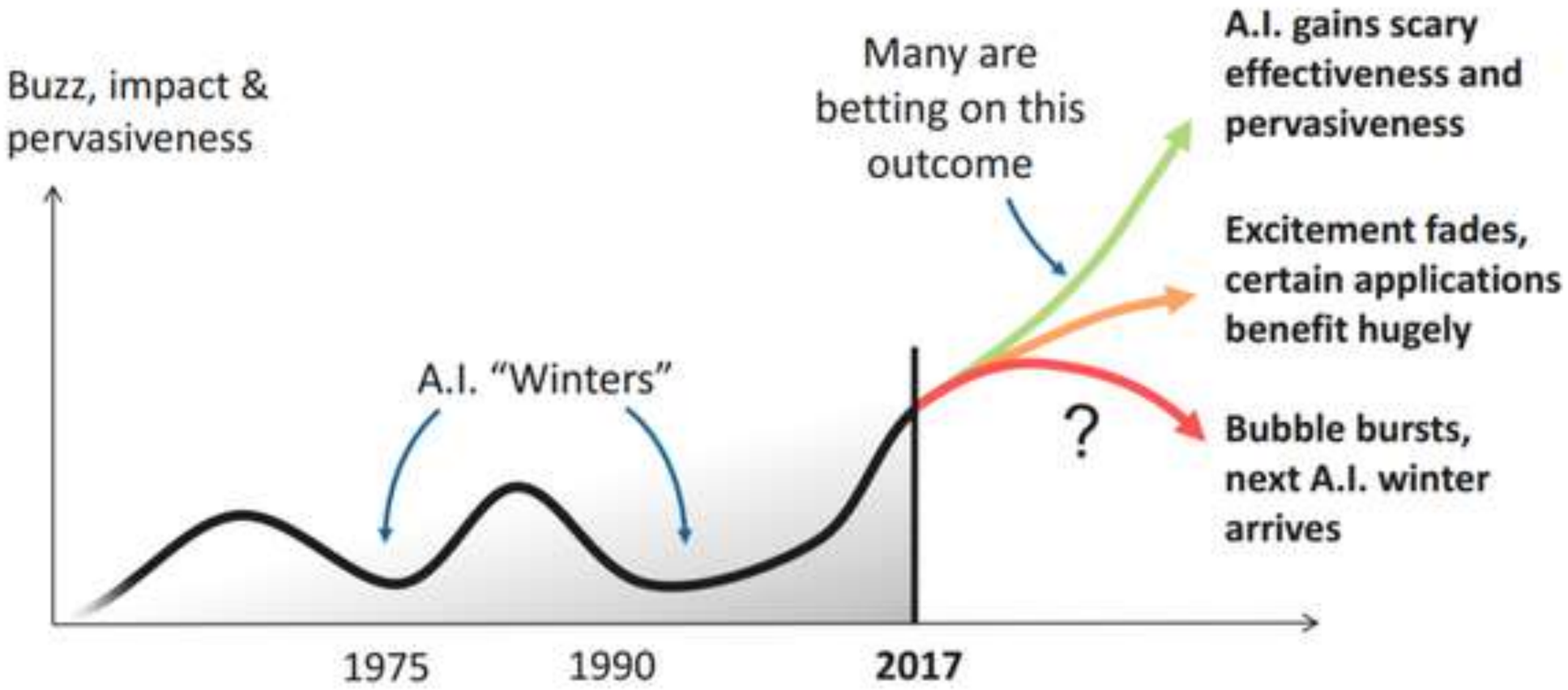

```
>>> from transformers import pipeline

>>> classifier = pipeline("sentiment-analysis")
```

2020s: ?



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

2020s: ?

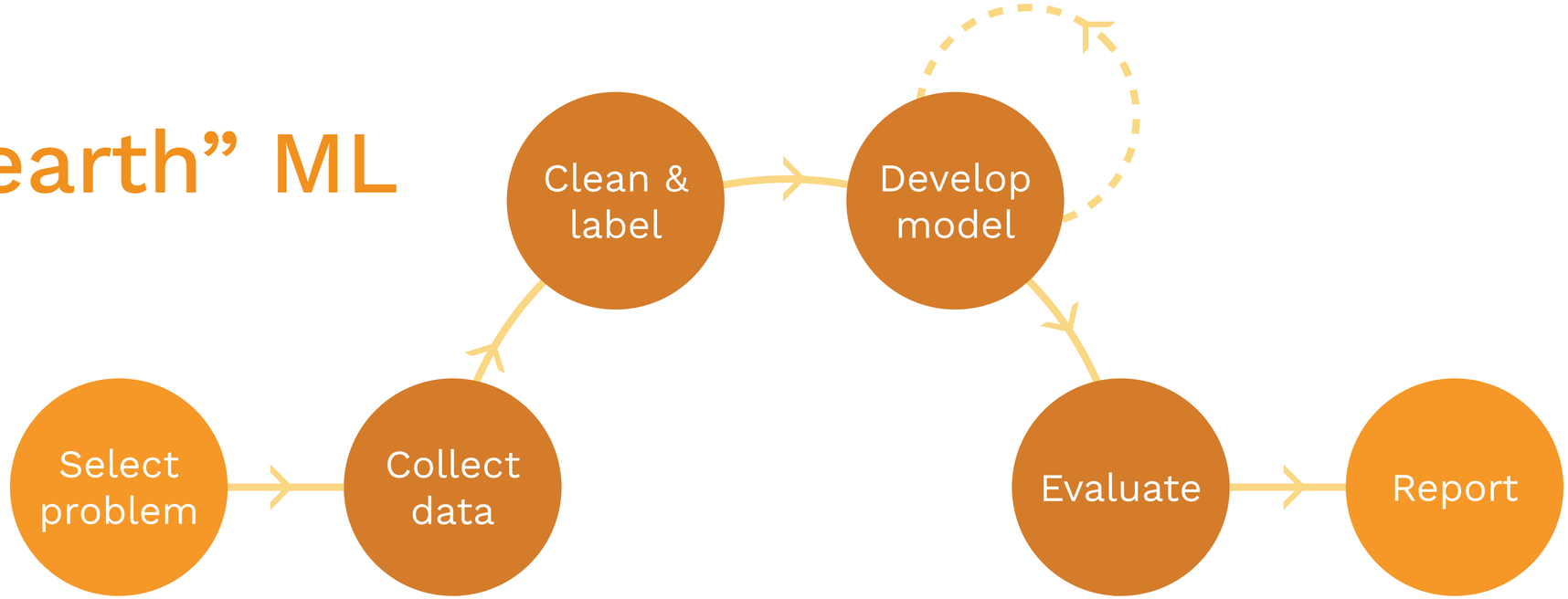


https://www.cambridgewireless.co.uk/media/uploads/resources/AI%20Group/AIMobility-11.05.17-Cambridge_Consultants-Monty_Barlow.pdf

Conjecture: we avoid an AI winter by translating research progress to real-world products

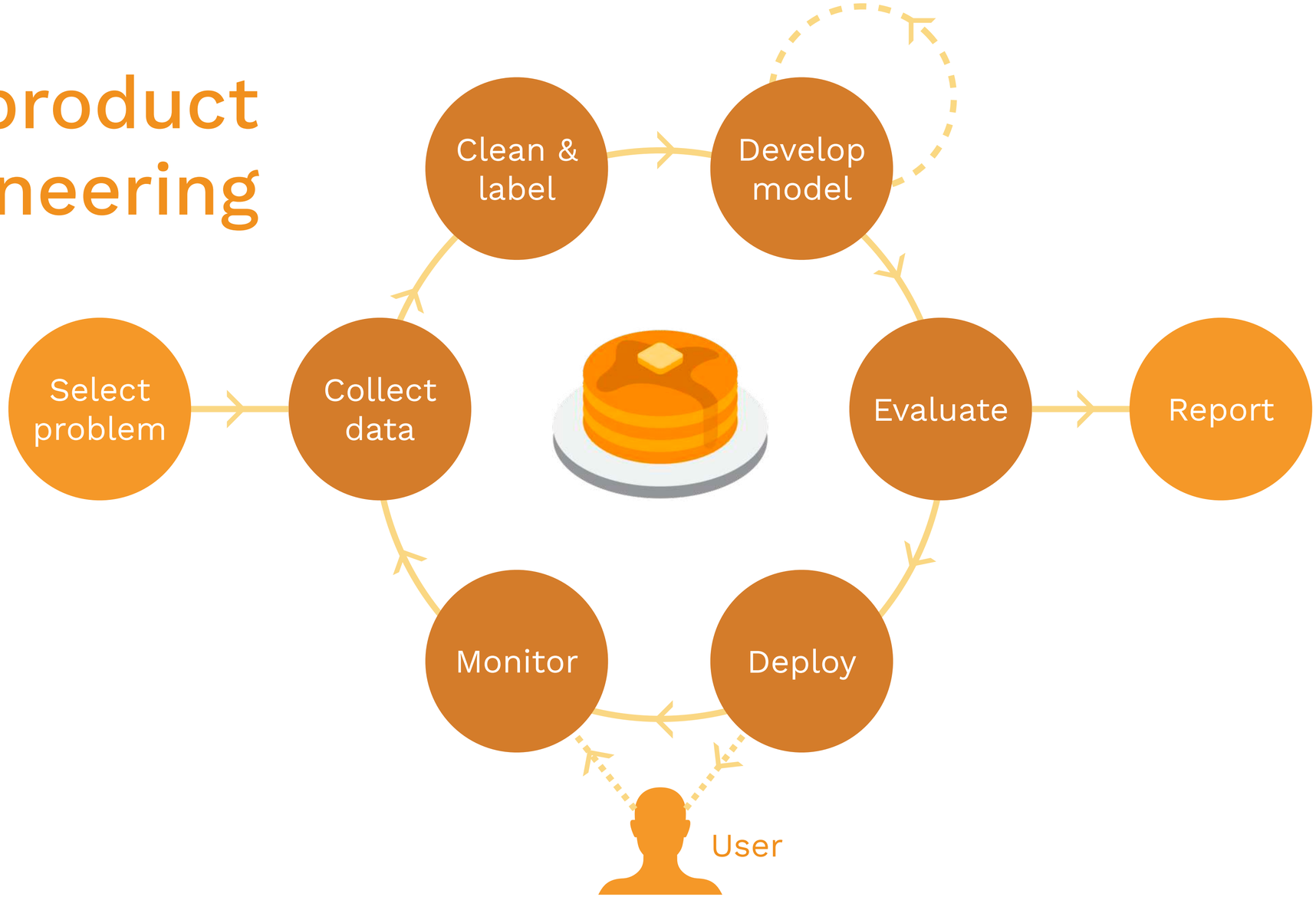
ML-powered products require a different process

“Flat-earth” ML

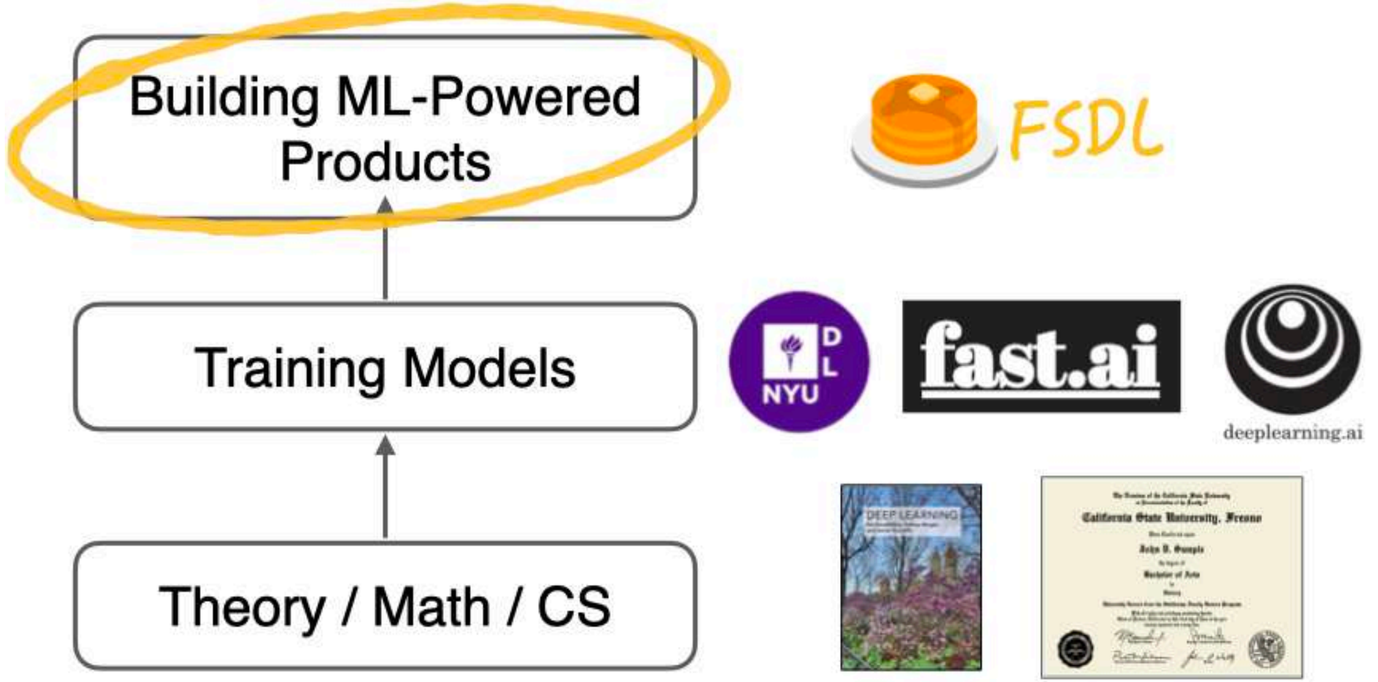


ML-powered products require a different process

ML product engineering



This course


Building ML-Powered Products
Training Models
Theory / Math / CS
FSDL
NYU
D
L
fast.ai
deeplearning.ai
DEEP LEARNING
California State University, Fresno

Our goals

- Build up generalist skills and an understanding of the components of a ML-powered product (and ML projects more generally)
- Teach you enough MLOps to get things done
- Share some best practices and explain the motivation behind them
- Learn some things that might help you with ML engineer job interviews
- Form a community to learn together and from each other

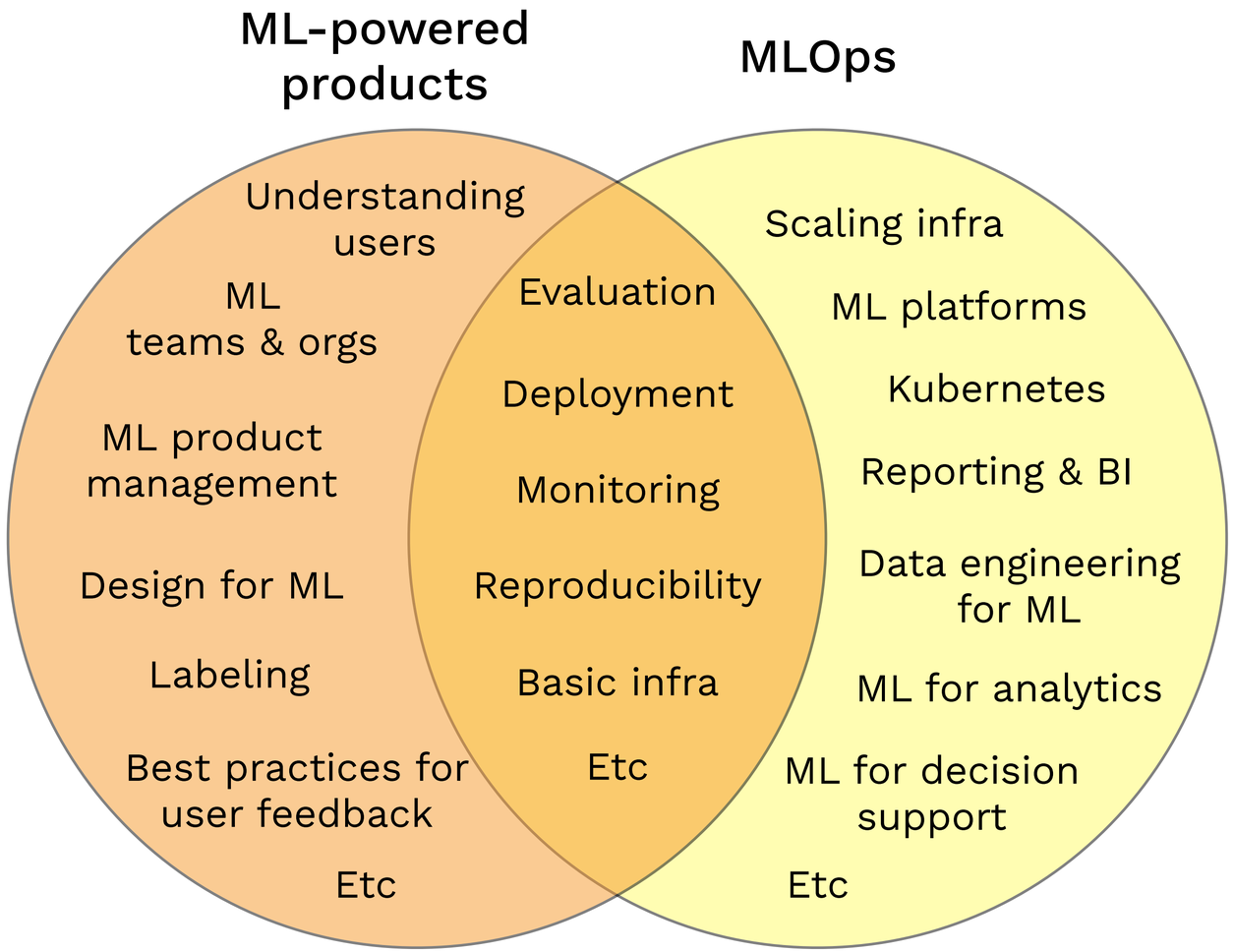
NOT our goals

- Teach you ML or SWE from scratch
- Cover the whole breadth of deep learning techniques
- Make you an expert in any single aspect of ML
- Do research in deep learning
- Cover the full spectrum of MLOps

How to refresh your prerequisites

- ML
  - Andrew Ng: https://www.coursera.org/collections/machine-learning
  - Google ML: https://developers.google.com/machine-learning/crash-course
- Software engineering
  - The Missing Semester: https://missing.csail.mit.edu/

ML-powered products vs MLOps



About us

Charles Frye teaches people on the internet. He worked in education and growth at Weights & Biases after getting a PhD in Neuroscience at UC Berkeley.

Sergey Karayev is Co-founder of Volition. He co-founded Gradescope after getting a PhD in Computer Vision at UC Berkeley.

Josh Tobin is Co-founder and CEO of **Gantry**. He worked as a Research Scientist at OpenAI and received a PhD in AI at UC Berkeley.

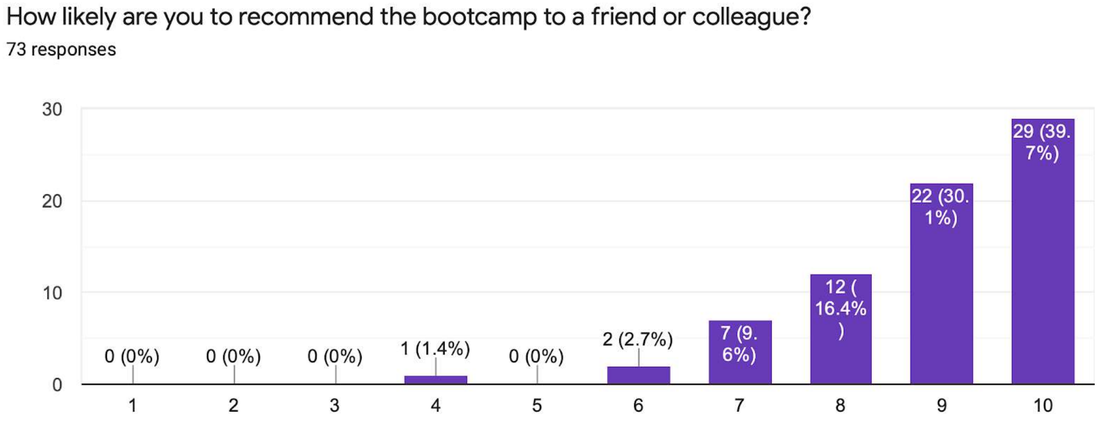
FSDL started as a Bootcamp

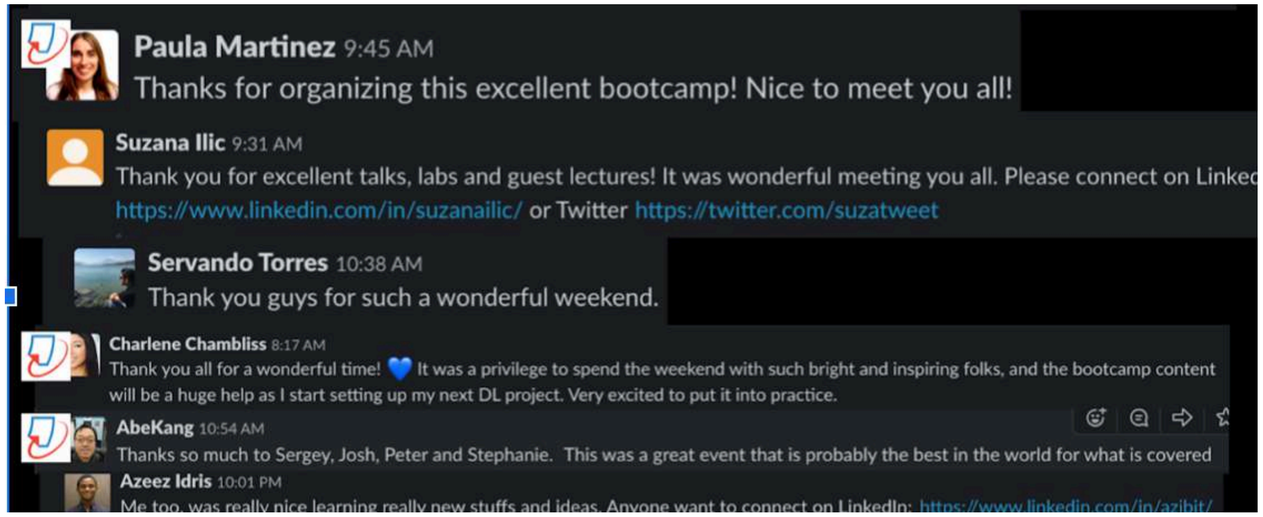
Aug 2018

Mar 2019

Nov 2019

We got good feedback, so we kept going





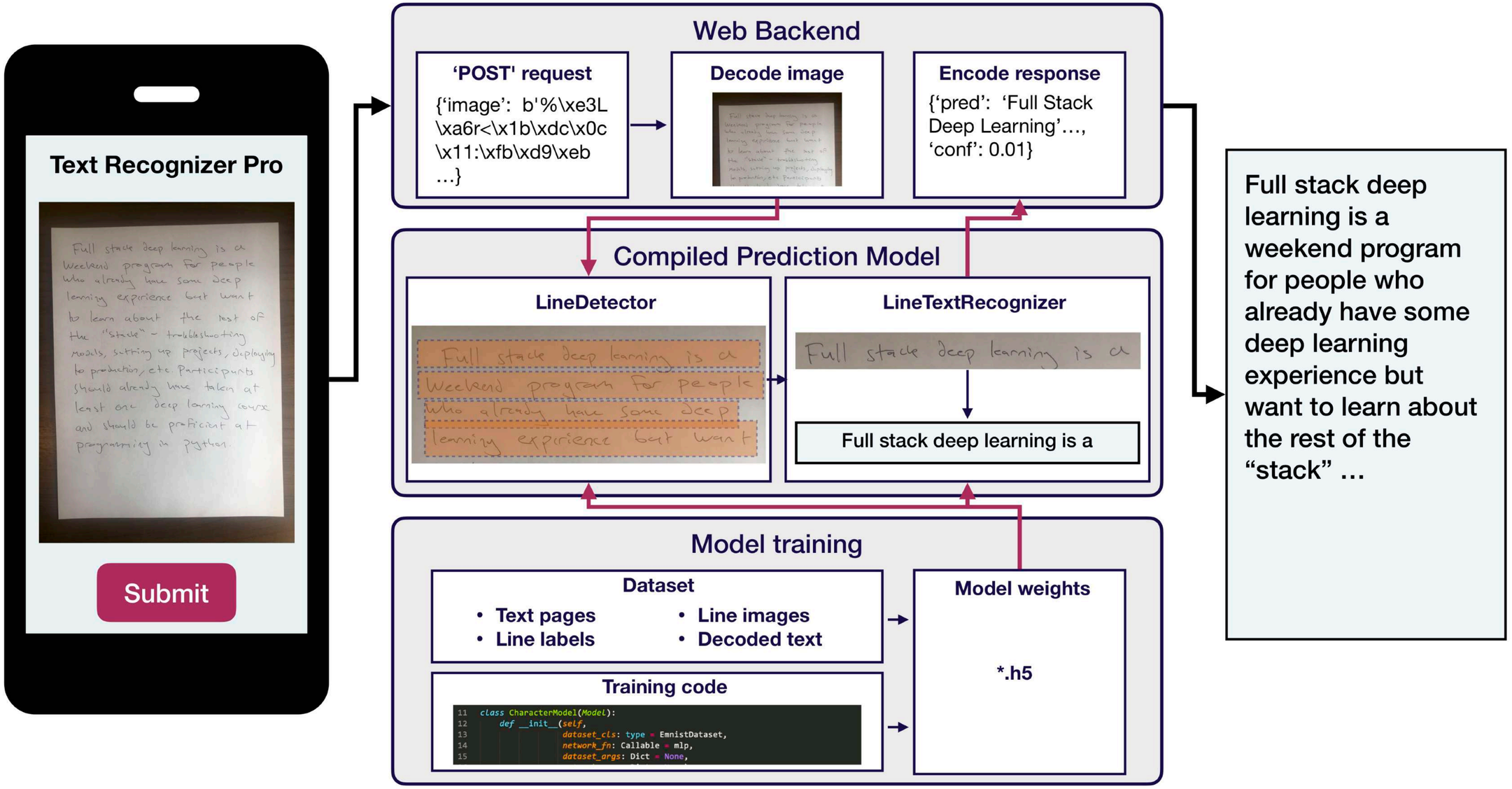
How we developed this course

- Personal experience and study
- Interviews with practitioners from these companies and more
- Posts, papers, product demos

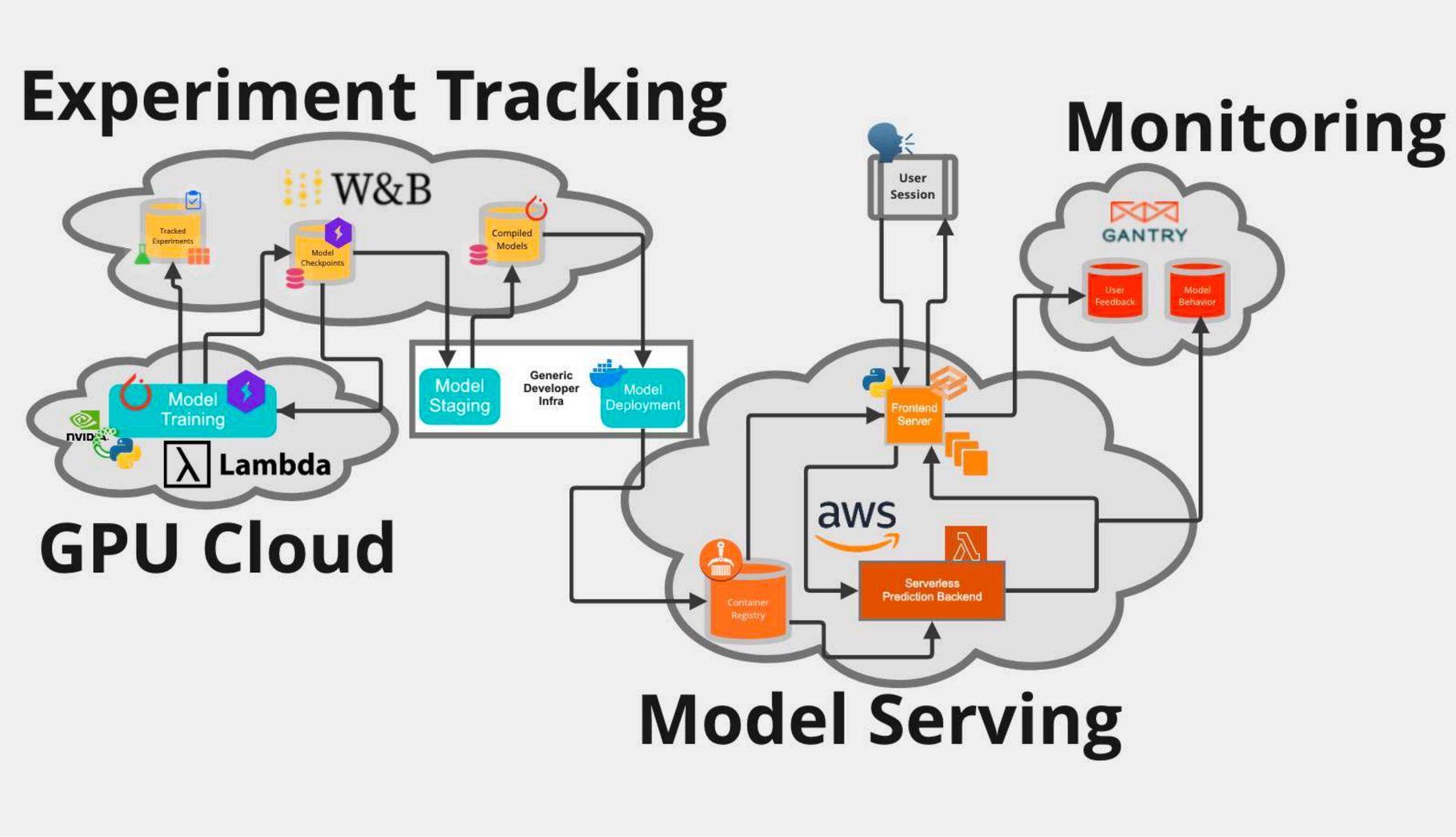
Some logistics

- Discord
- Course project
- Labs

Labs - the problem



Labs — architecture



Summary

- ML-powered products are going mainstream thanks to the democratization of modeling
- However, building great ones requires a different process vs building models
- FSDL is here to help!

01

When to use Machine Learning

Key points in this section

- ML introduces complexity
  - Don’t do it before you’re ready
  - Exhaust your other options first
  - BUT: you don’t need perfect infrastructure to start
- Prioritize projects you know are feasible and will have an impact

When to use ML at all

Lots of ML projects fail

- Commonly quoted statistic: 87%[1]
  - However, 73% of all statistics are made up on the spot
- Anecdotally, probably more like 25%

https://mtszkw.medium.com/why-do-87-of-data-science-projects-fail-and-are-we-sure-that-it-is-true-fe8b5ba1404c

Why?

- ML is still research — shouldn't aim for 100% success
- But, many are doomed to fail:
  - Technically infeasible or poorly scoped
  - Never make the leap to prod
  - Unclear success criteria
  - Works, but doesn't solve a big enough problem to be worth the complexity

The value of your project must outweigh its complexity

...and ML introduces a lot of complexity

- Erodes the boundaries between systems
- Relies on expensive data dependencies
- Commonly plagued by system design anti-patterns
- Subject to the instability of the external world

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

So before starting an ML project, ask yourself:

- Are we ready to use ML?
- Do we really need ML to solve this problem?
- Is it ethical?

So before starting an ML project, ask yourself:

- Are we ready to use ML?
  - Do we have a product?
- Do we really need ML to solve this problem?
- Is it ethical?

So before starting an ML project, ask yourself:

- Are we ready to use ML?
  - Do we have a product?
  - Are we collecting data and storing it in a sane way?
- Do we really need ML to solve this problem?
- Is it ethical?

So before starting an ML project, ask yourself:

- Are we ready to use ML?
  - Do we have a product?
  - Are we collecting data and storing it in a sane way?
  - Do we have the right people?
- Do we really need ML to solve this problem?
- Is it ethical?

So before starting an ML project, ask yourself:

- Are we ready to use ML?
- Do we really need ML to solve this problem?
  - Do we need to solve the problem?
- Is it ethical?

So before starting an ML project, ask yourself:

- Are we ready to use ML?
- Do we really need ML to solve this problem?
  - Do we need to solve the problem?
  - Have we tried using rules or simple stats?
- Is it ethical?

So before starting an ML project, ask yourself:

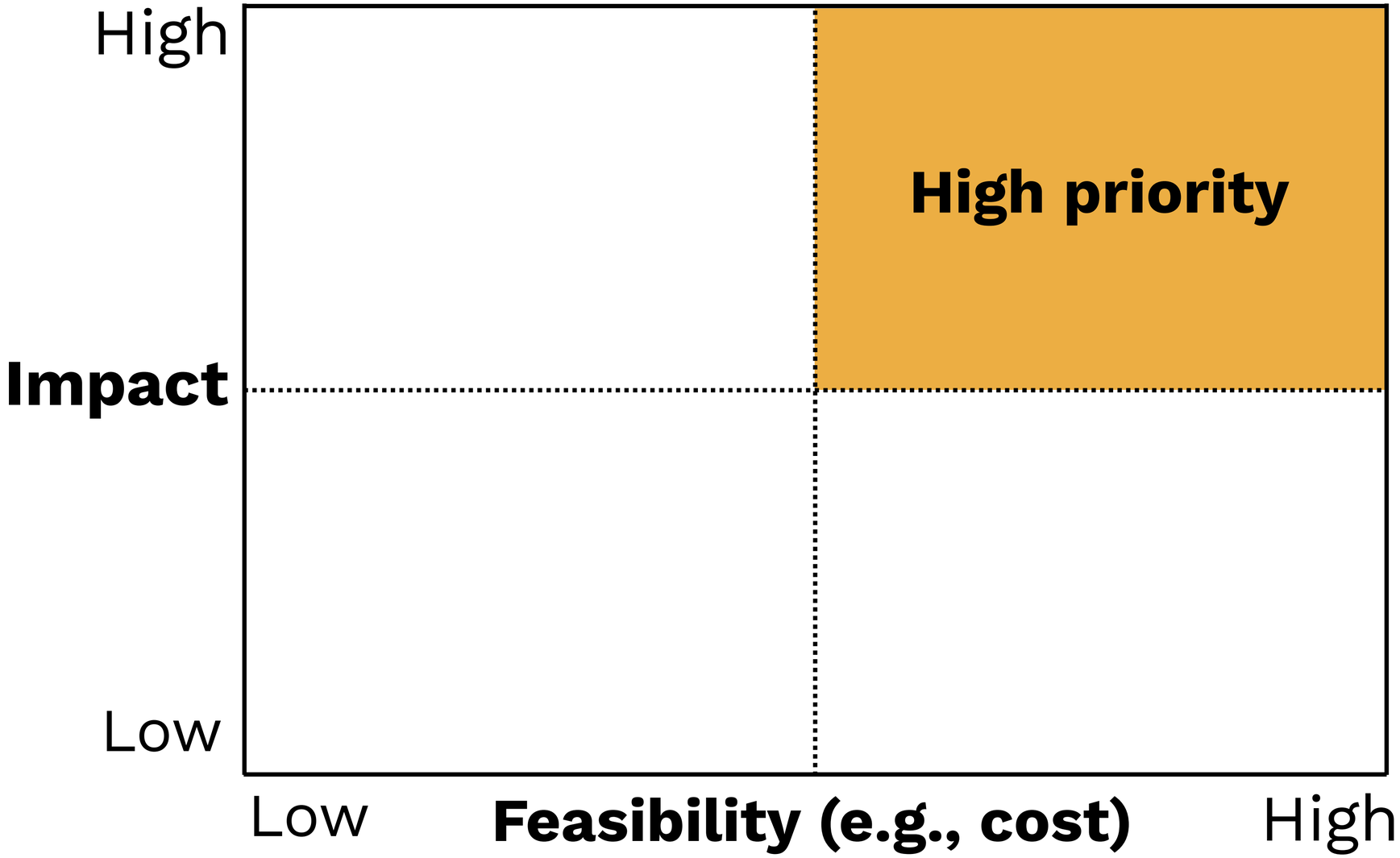
- Are we ready to use ML?
- Do we really need ML to solve this problem?
- Is it ethical?
  - Ethics lecture!

How to pick problems to solve with ML

TL/DR: High impact, low-cost

- High impact problems are likely to be those that address
  - Friction in your product
  - Complex parts of your pipeline
  - Places where cheap prediction is valuable
  - What other people in your industry are doing
- Low-cost projects are those with data available, and where bad predictions aren’t too harmful

A (general) prioritization framework



Mental models for high-impact ML projects

- Where can you take advantage of cheap prediction?
- Where is there friction in your product?
- Where can you automate complicated manual processes?
- What are other people doing?

What does ML make economically feasible?

The economics of AI
(Agrawal, Gans, Goldfarb)

- AI reduces cost of prediction
- Prediction is central for decision making
- Cheap prediction means
  - Prediction will be everywhere
  - Even in problems where it was too expensive before (e.g., for most people, hiring a driver)
- **Implication:** Look for projects where cheap prediction will have a huge business impact

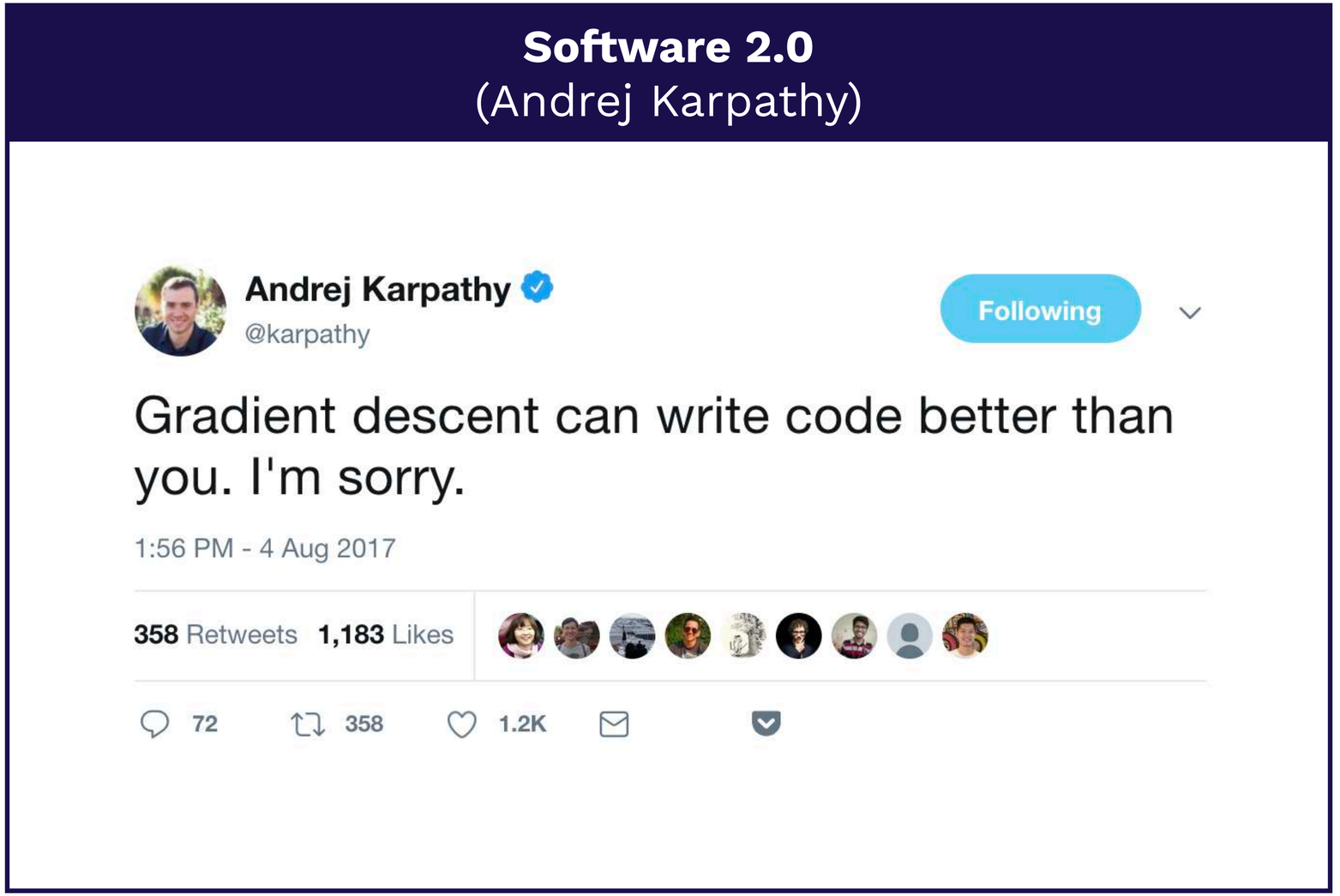
Prediction Machines: The Simple Economics of Artificial Intelligence (Agrawal, Gans, Goldfarb)

What does your product need?

“Discover Weekly removed the friction of chasing everything down yourself and instead brought the music to you in a neat little package every Monday morning.”

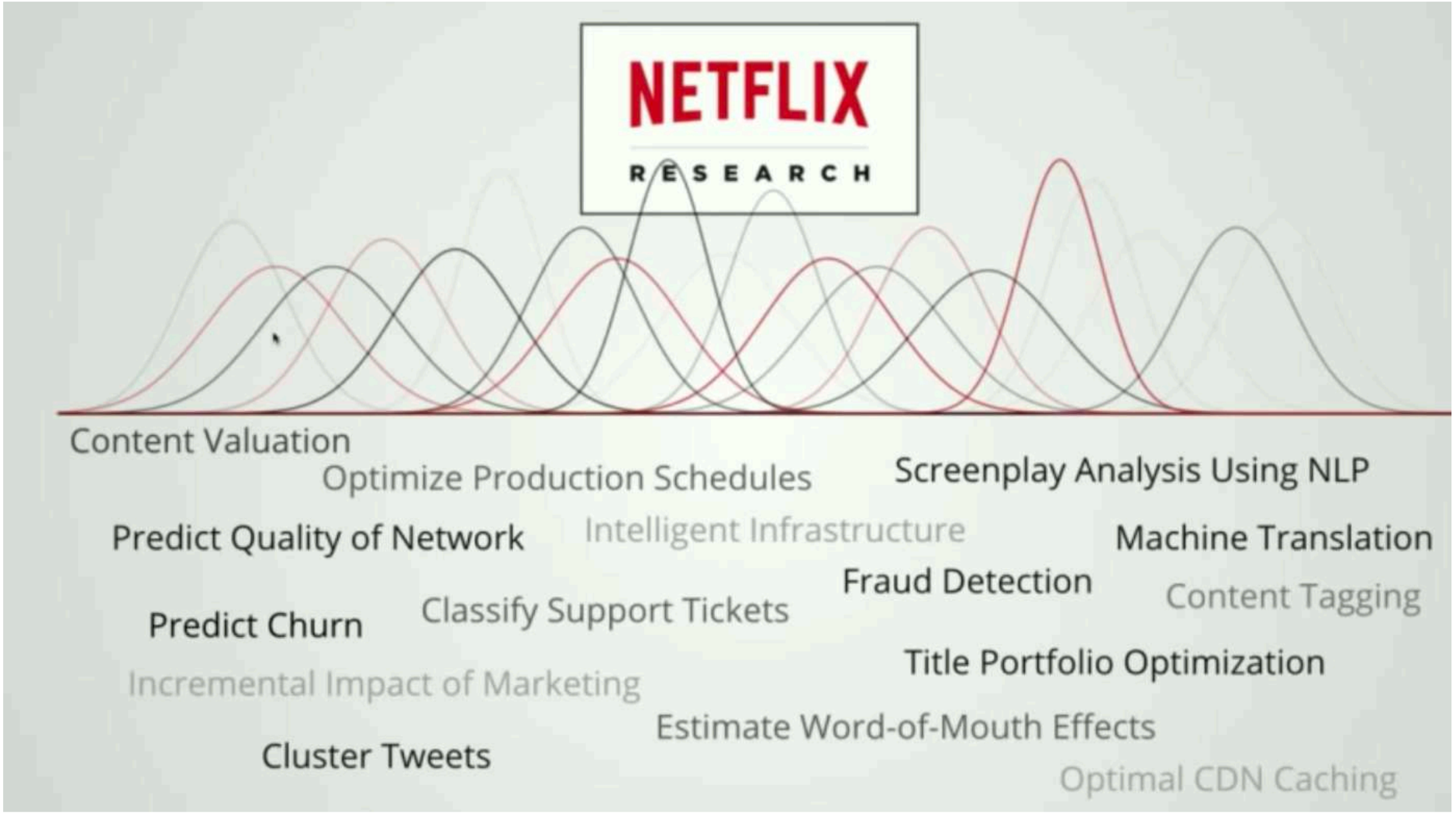
What is ML good at?

Software 2.0
(Andrej Karpathy)



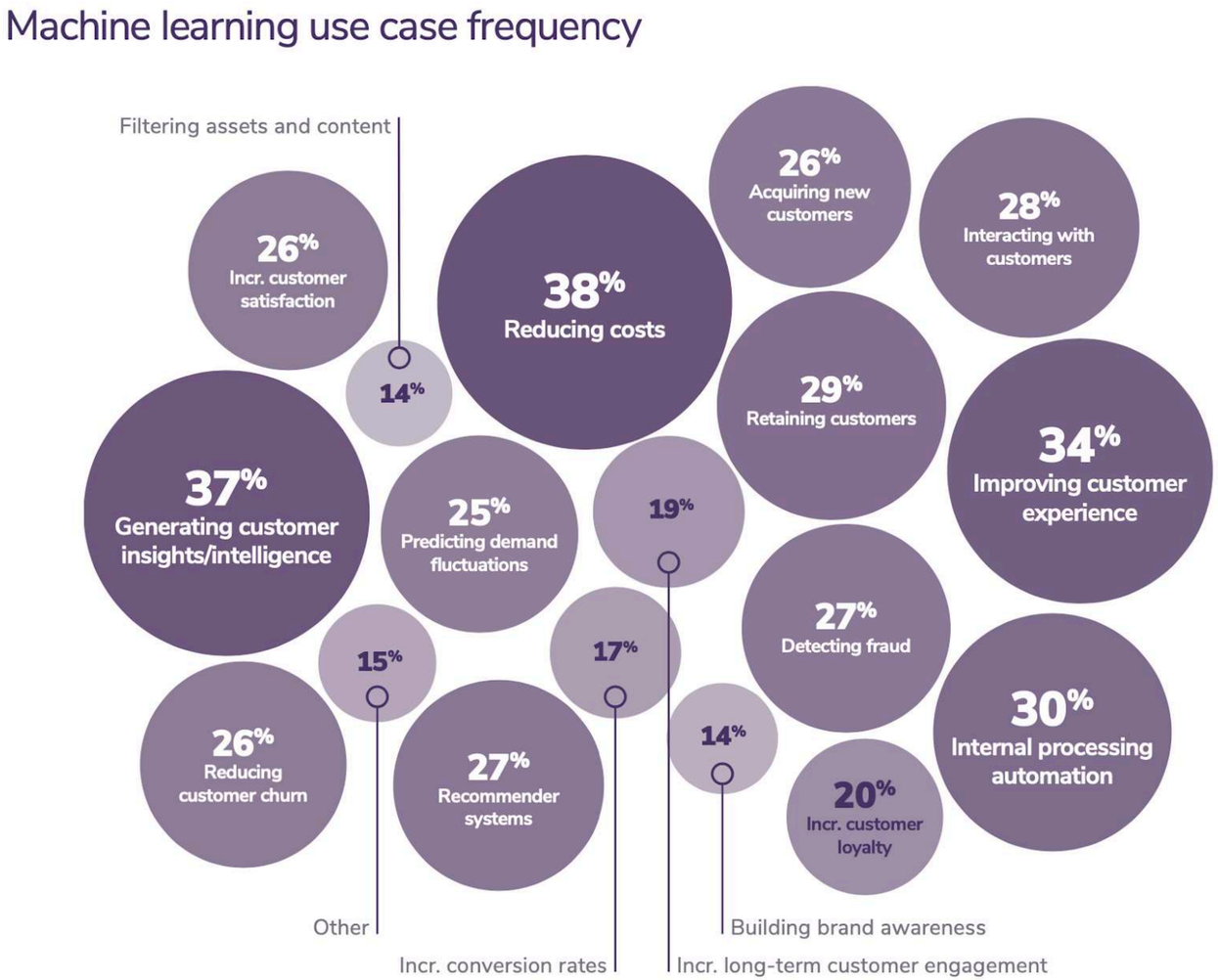
Software 2.0 (Andrej Karpathy): https://medium.com/@karpathy/software-2-0-a64152b37c35

What are other people doing?



Human-Centric Machine Learning Infrastructure @Netflix (Ville Tuulos, InfoQ 2019)

What are other people doing?



2020 state of enterprise machine learning (Algorithmia, 2020)

What are other people doing?

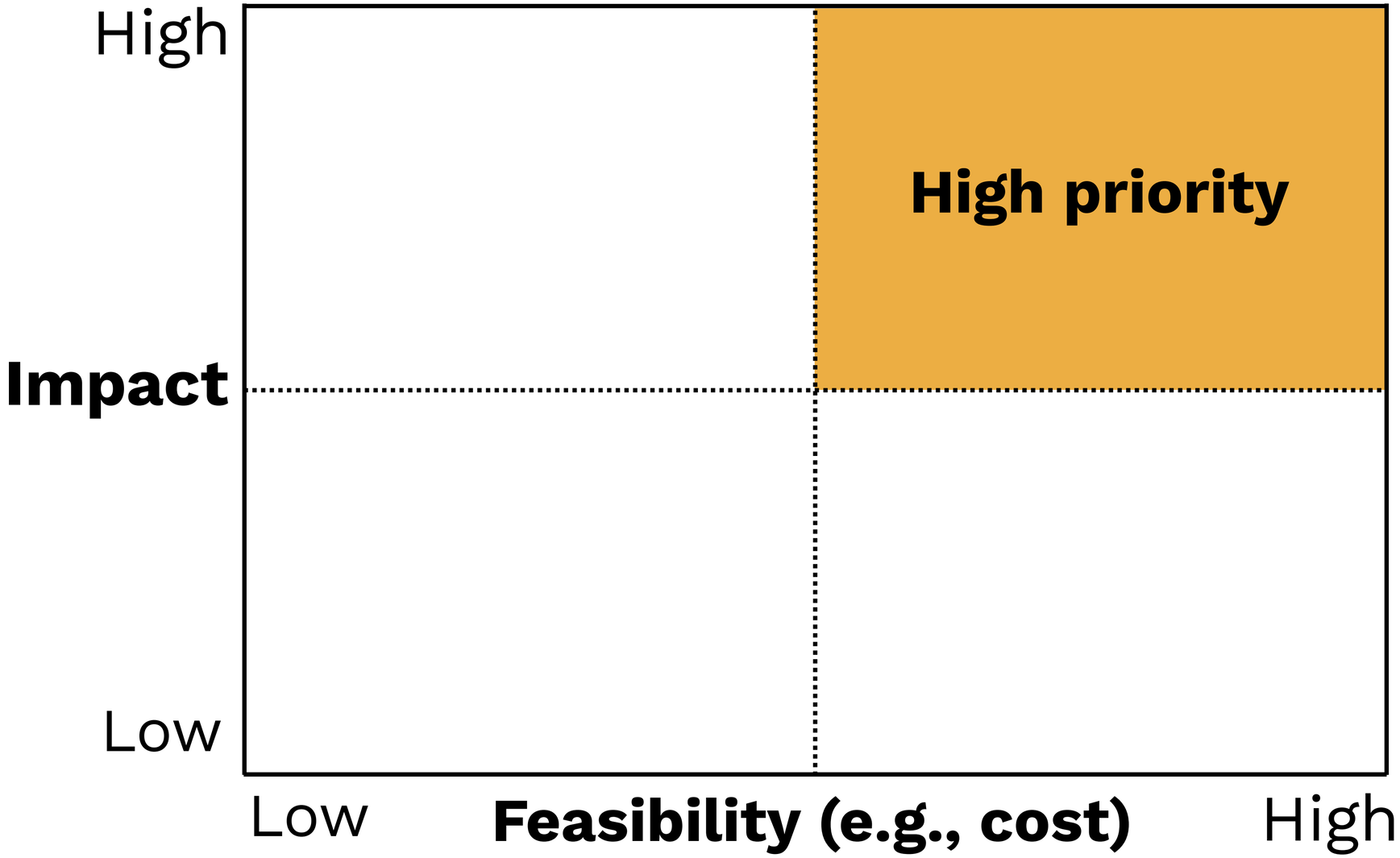
- Papers from Google, Facebook, Nvidia, Netflix, etc
- Blog posts from top earlier-stage companies (Uber, Lyft, Spotify, Stripe, etc)

Case studies

- Using Machine Learning to Predict Value of Homes On Airbnb (Robert Chang, Airbnb Engineering & Data Science, 2017)
- Using Machine Learning to Improve Streaming Quality at Netflix (Chaitanya Ekanadham, Netflix Technology Blog, 2018)
- 150 Successful Machine Learning Models: 6 Lessons Learned at Booking.com (Bernardi et al., KDD, 2019)Asdf
- How we grew from 0 to 4 million women on our fashion app, with a vertical machine learning approach (Gabriel Aldamiz, HackerNoon, 2018)
- Machine Learning-Powered Search Ranking of Airbnb Experiences (Mihajlo Grbovic, Airbnb Engineering & Data Science, 2019)
- From shallow to deep learning in fraud (Hao Yi Ong, Lyft Engineering, 2018)
- Space, Time and Groceries (Jeremy Stanley, Tech at Instacart, 2017)
- Creating a Modern OCR Pipeline Using Computer Vision and Deep Learning (Brad Neuberg, Dropbox Engineering, 2017)
- Scaling Machine Learning at Uber with Michelangelo (Jeremy Hermann and Mike Del Balso, Uber Engineering, 2019)
- Spotify’s Discover Weekly: How machine learning finds your new music (Umesh .A Bhat, 2017)

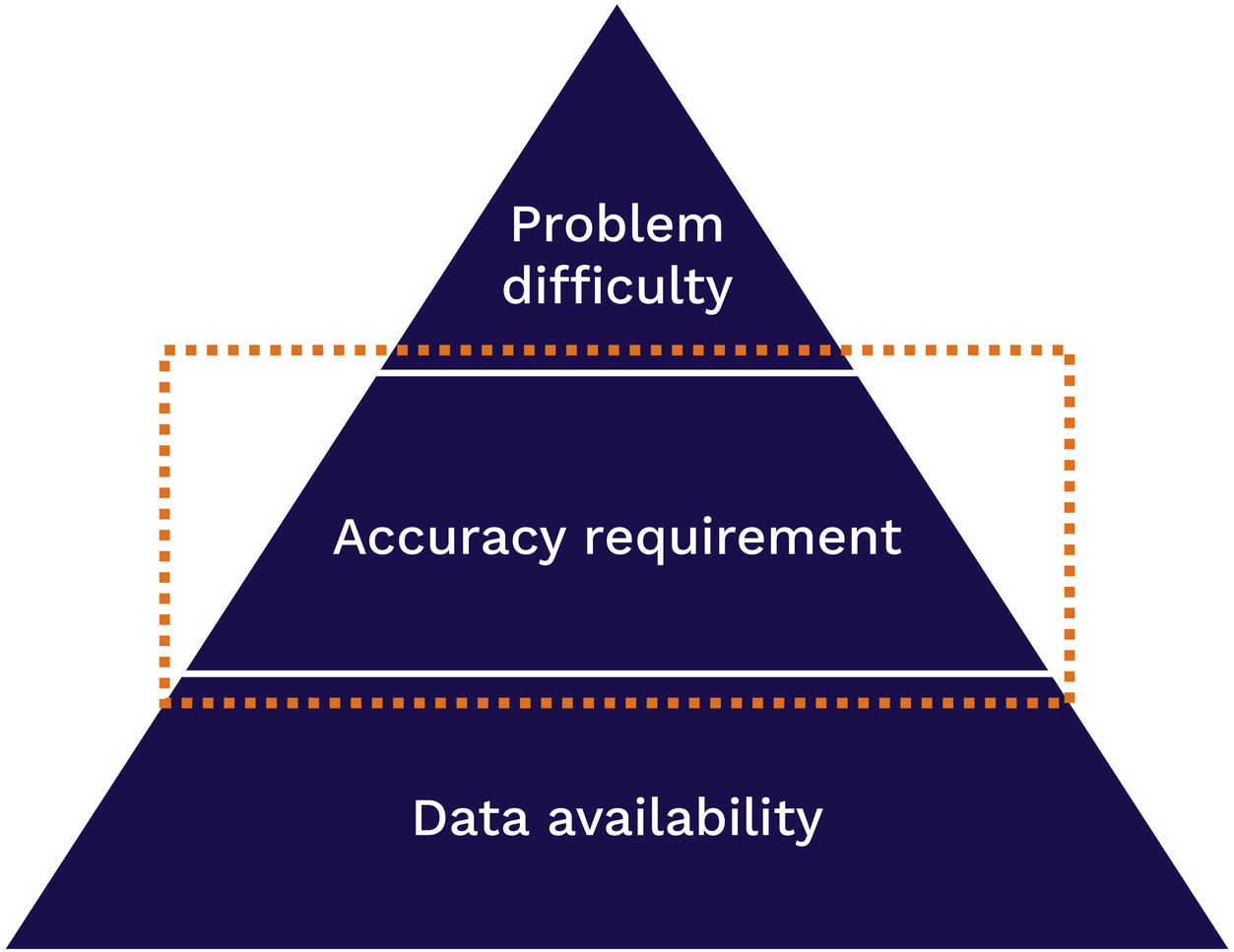
Credit for compiling this list: Chip Huyen (Machine Learning Systems Design Lecture 2 note)

A (general) prioritization framework



Assessing feasibility of ML projects

Cost drivers



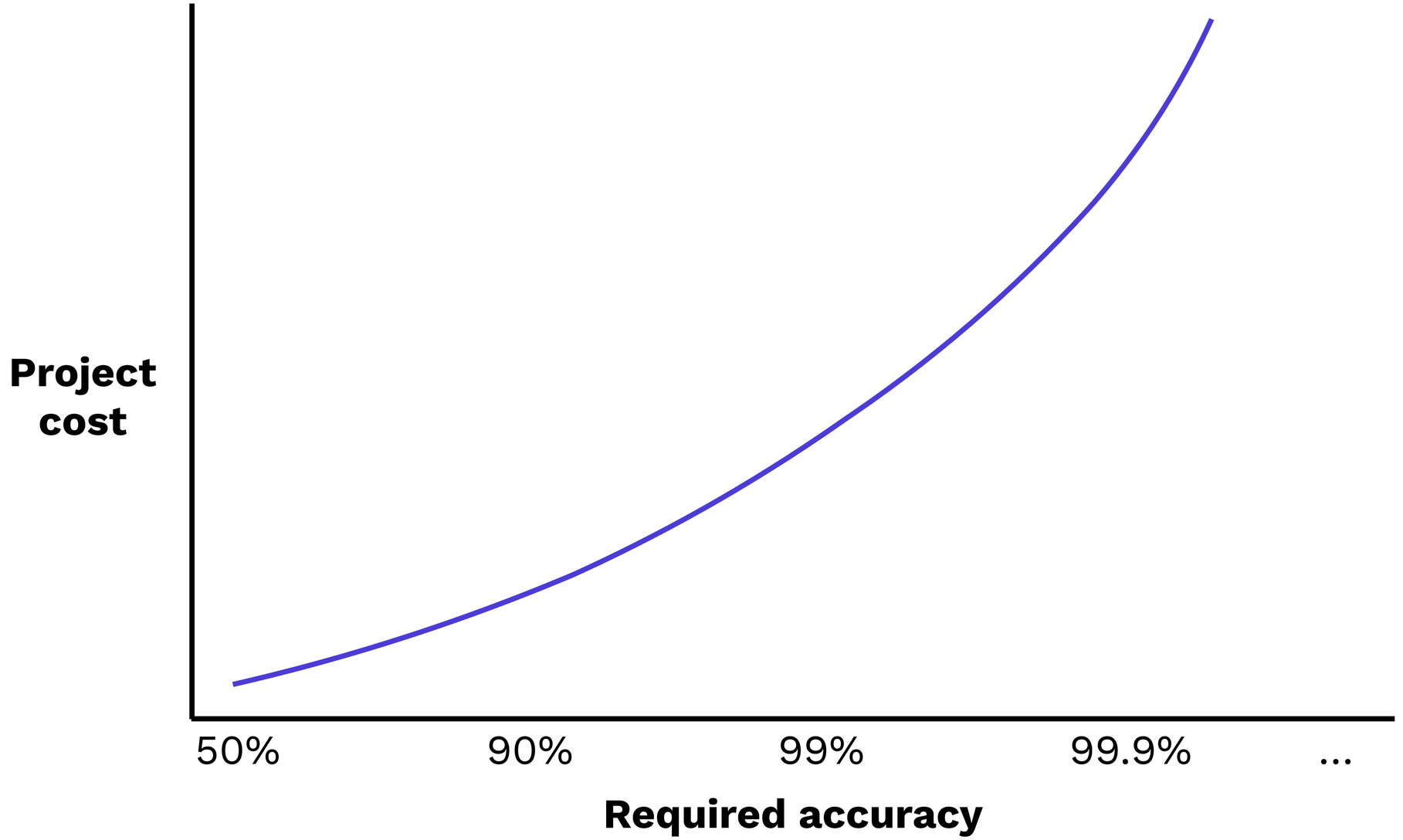
Main considerations

- Is the problem well-defined?
- Good published work on similar problems? (newer problems mean more risk & more technical effort)
- Compute requirements?
- Can a human do it?

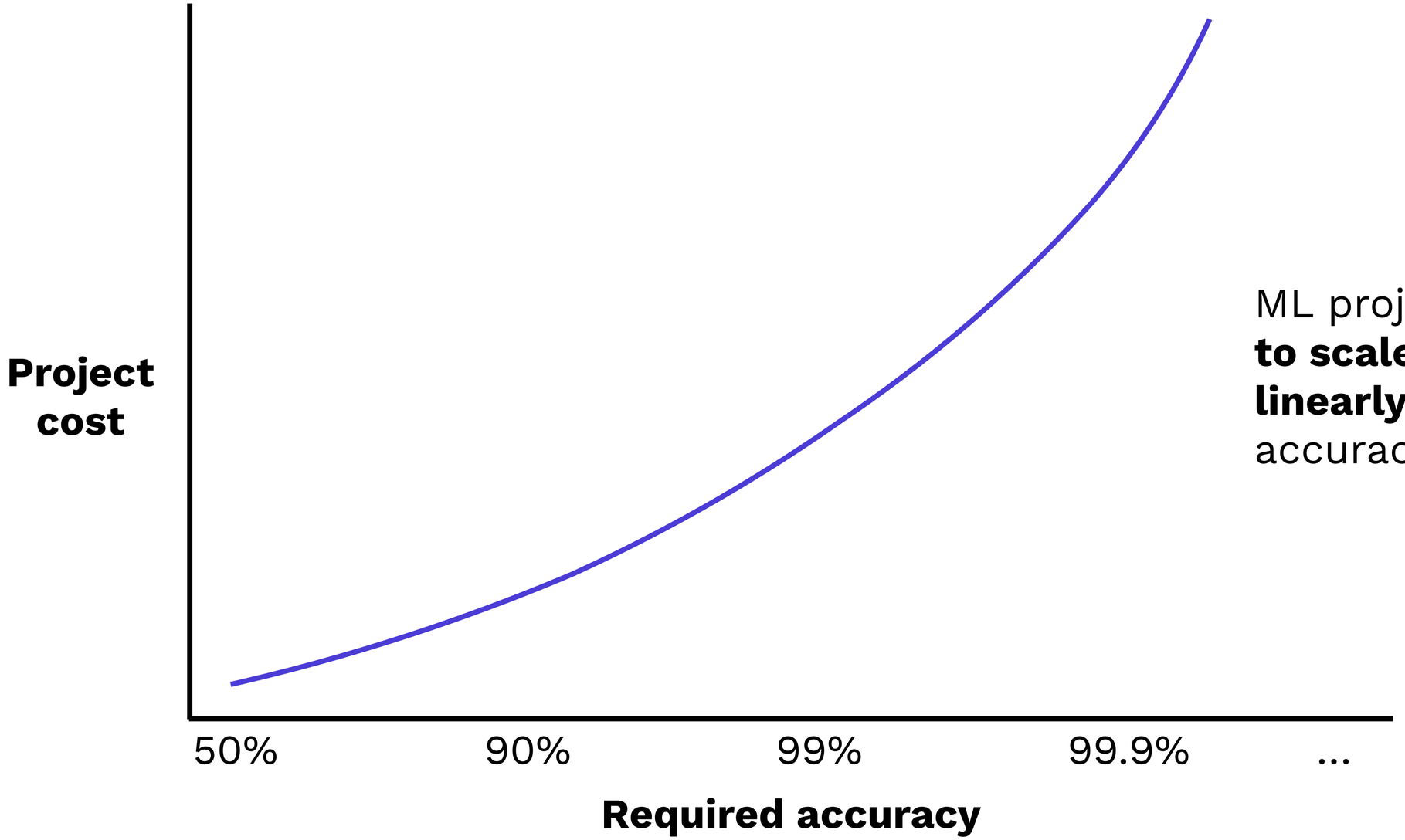
- How costly are wrong predictions?
- How frequently does the system need to be right to be useful?
- Ethical implications?

- How hard is it to acquire data?
- How expensive is data labeling?
- How much data will be needed?
- How stable is the data?
- Data security requirements?

Why are accuracy requirements so important?



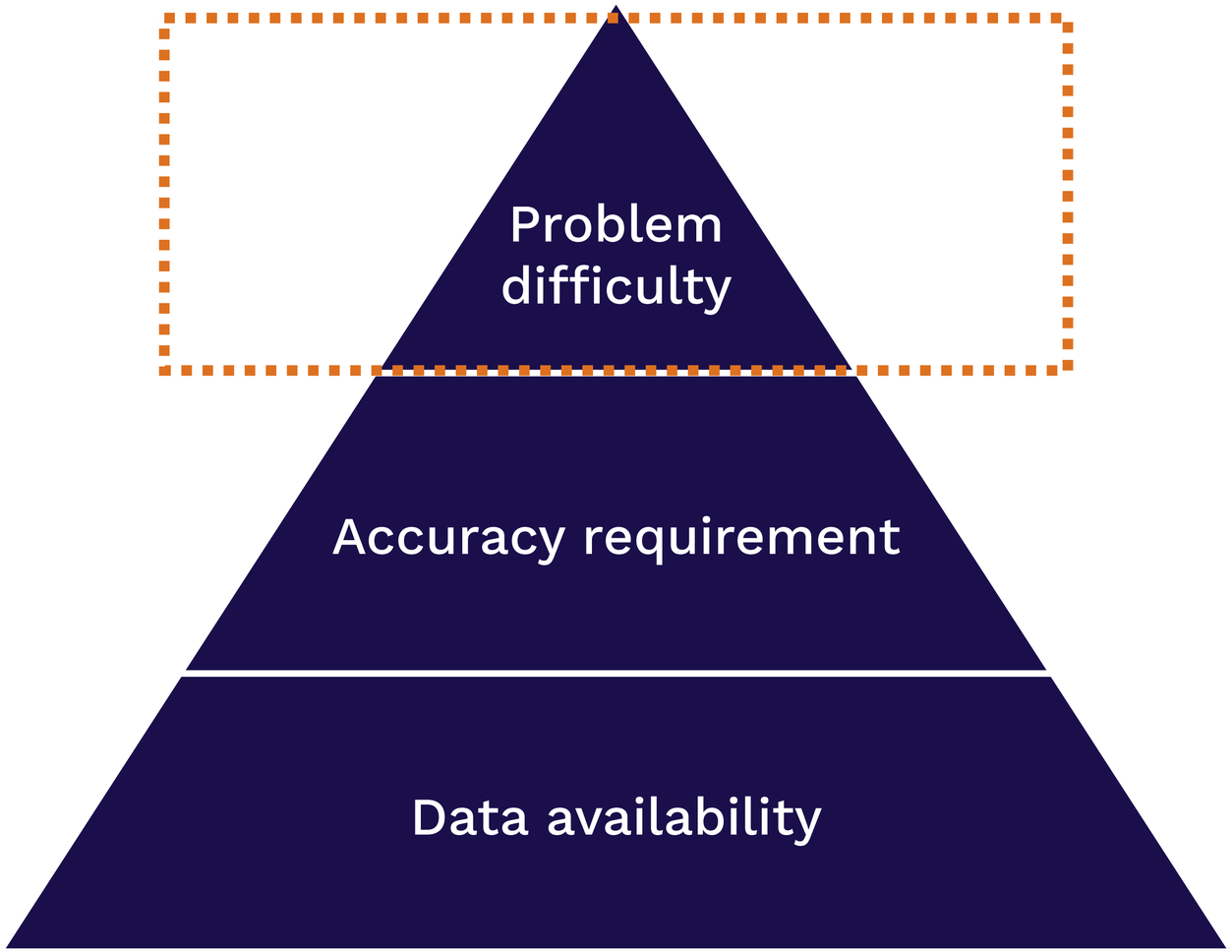
Why are accuracy requirements so important?



ML project **costs tend to scale super-linearly** in the accuracy requirement

Assessing feasibility of ML projects

Cost drivers



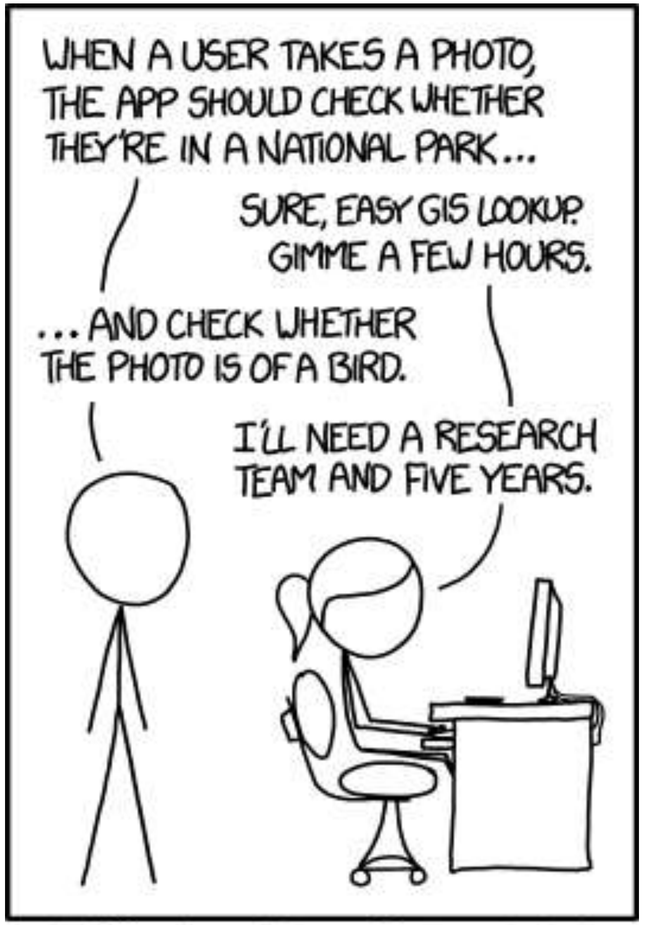
Main considerations

- Is the problem well-defined?
- Good published work on similar problems? (newer problems mean more risk & more technical effort)
- Compute requirements?
- Can a human do it?

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

- How hard is it to acquire data?
- How expensive is data labeling?
- How much data will be needed?
- How stable is the data?
- Data security requirements?

It’s hard to reason about what’s feasible in ML



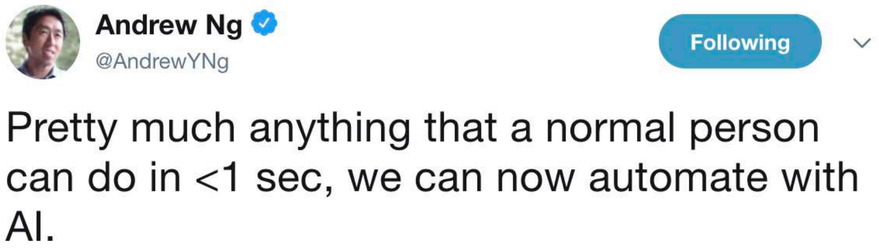
What's still hard in machine learning?

"It may be a hundred years before a computer beats humans at Go -- maybe even longer," said Dr. Piet Hut, an astrophysicist at the Institute for Advanced Study in Princeton, N.J., and a fan of the game. "If a reasonably intelligent person learned to play Go, in a few months he could beat all existing computer programs. You don't have to be a Kasparov."

New York Times, July 1997

What's still hard in machine learning?

What's still hard in machine learning?



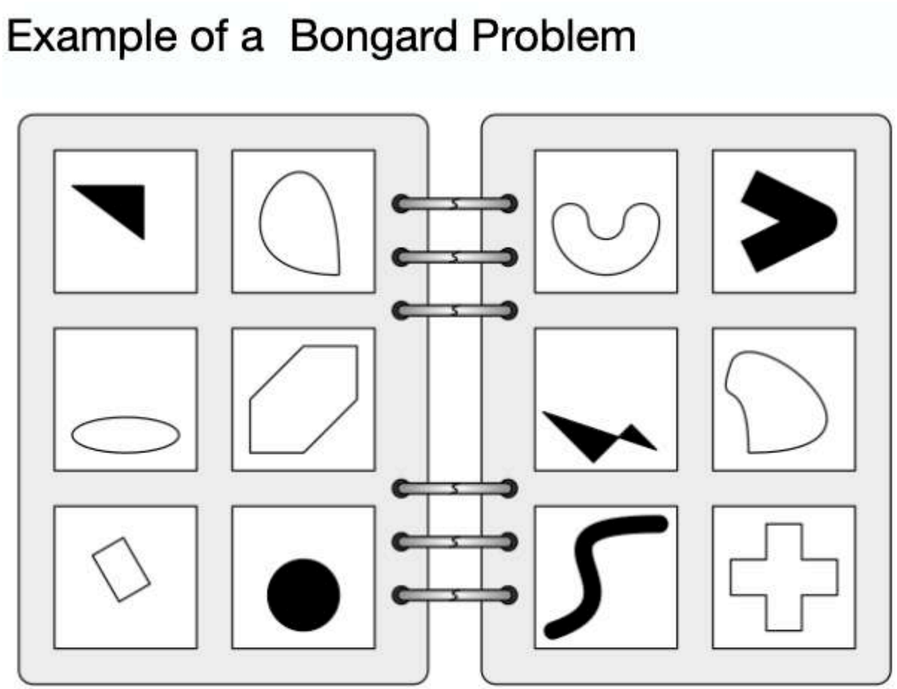
Examples	Counter-examples?
• Recognize content of images • Understand speech • Translate speech • Grasp objects • etc.	• ~~Understand humor / sarcasm~~ • In-hand robotic manipulation • Generalize to new scenarios • etc.

What's still hard in machine learning?

- ~~Unsupervised learning~~
- Reinforcement learning
- Showing promise in limited domains where tons of data and compute are available

What's still hard in supervised learning?

- Predicting video
- Real-world speech recognition
- Resisting adversarial examples
- Solving word puzzles
- Bongard problems
- ~~Summarizing text~~
- ~~Building 3D models~~
- ~~Answering questions~~
- ~~Doing math~~



What types of problems are hard?

	Instances	Examples
Output is complex	• High-dimensional output • Ambiguous output	• 3D reconstruction • Video prediction • Dialog systems • Open-ended recommender systems
Reliability is required	• High precision is required • Robustness is required	• Failing safely out-of-distribution • Robustness to adversarial attacks • High-precision pose estimation
Generalization is required	• Out of distribution data • Reasoning, planning, causality	• Self-driving: edge cases • Self-driving: control • Small data

How to run a ML feasibility assessment

A. Are you sure you need ML at all?

B. Put in the work up-front to define success criteria with all of the stakeholders

C. Consider the ethics of using ML

D. Do a literature review

E. Try to rapidly build a labeled benchmark dataset

F. Build a *minimum* viable model (e.g., manual rules)

G. Are you sure you need ML at all?

Not all ML projects should be planned the same way

Machine learning product archetypes

	Definition
Software 2.0	Taking something software does today and doing it better with ML
Human-in-the-loop	Helping humans do their jobs better by complementing them with ML-based tools
Autonomous systems	Taking something humans do today and automating it with ML

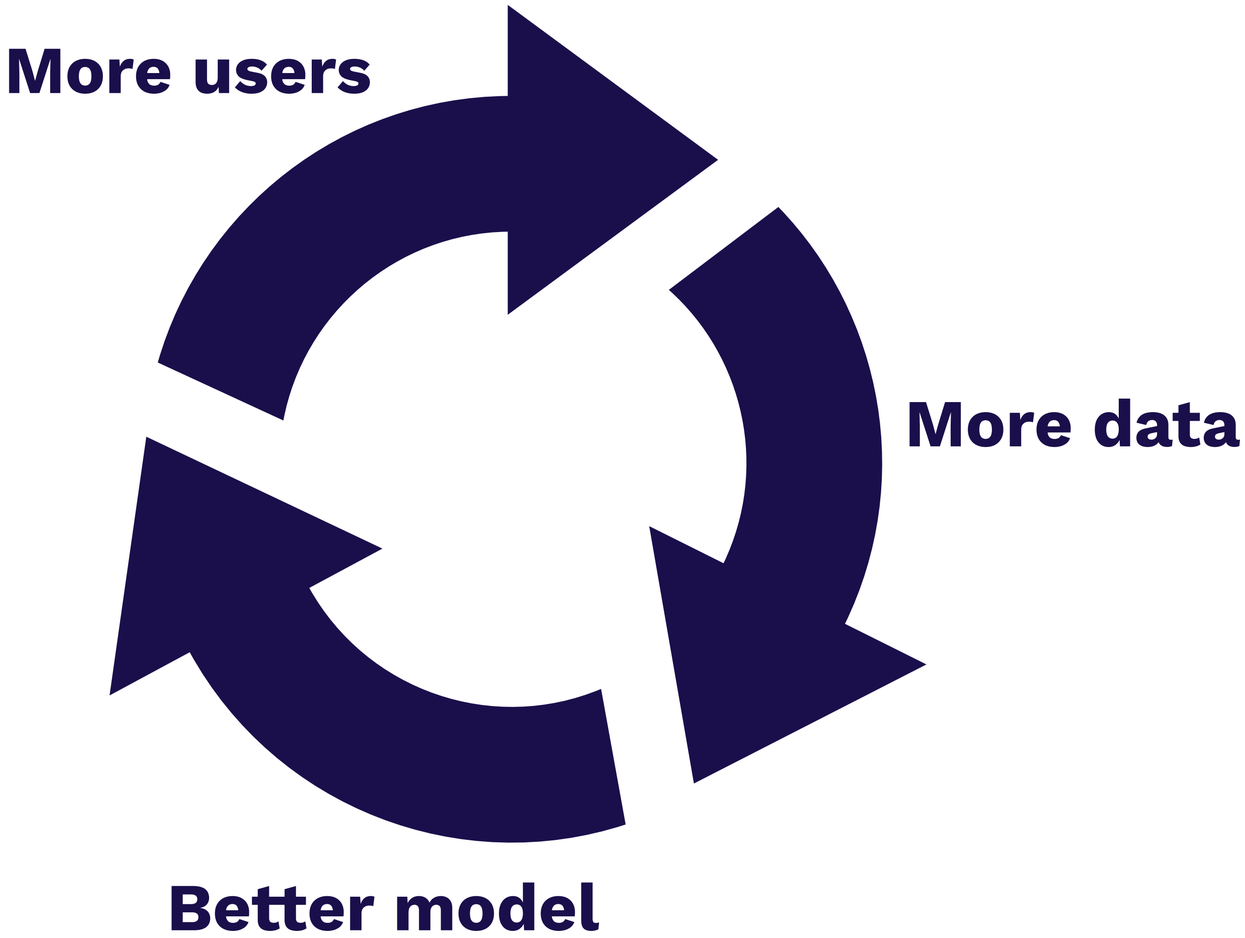
Machine learning product archetypes

	Examples
Software 2.0	• Improve code completion in an IDE • Build a customized recommendation system • Build a better video game AI
Human-in-the-loop	• Turn sketches into slides • Email auto-completion • Help a radiologist do their job faster
Autonomous systems	• Full self-driving • Automated customer support • Automated website design

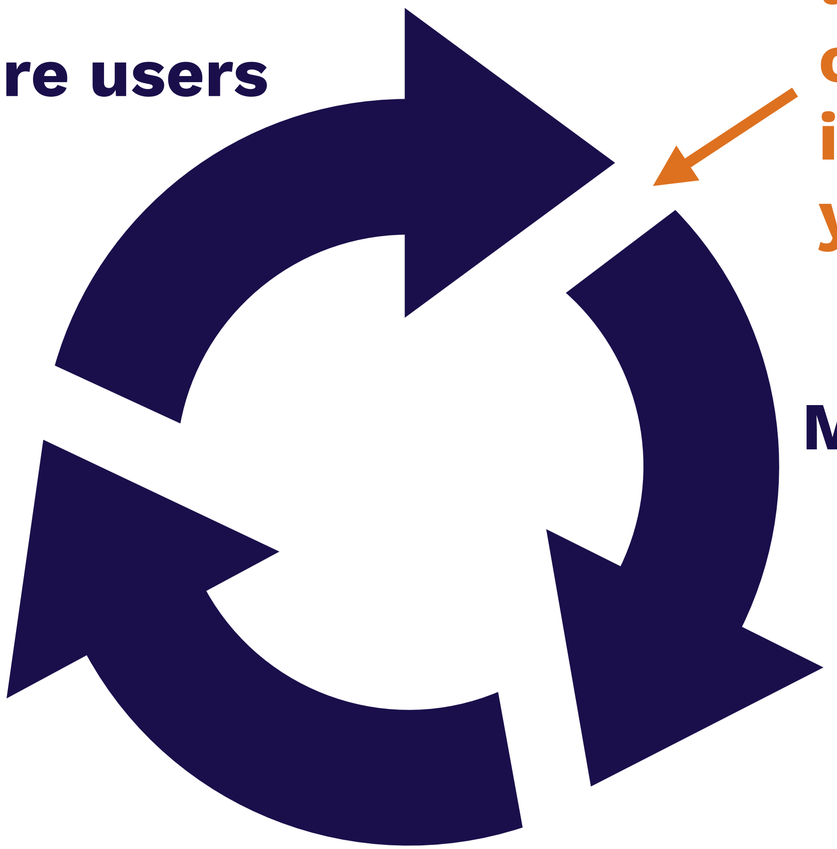
Machine learning product archetypes

	Key questions
Software 2.0	• Do your models truly improve performance? • Does performance improvement generate business value? • Do performance improvements lead to a data flywheel?
Human-in-the-loop	• How good does the system need to be to be useful? • How can you collect enough data to make it that good?
Autonomous systems	• What is an acceptable failure rate for the system? • How can you guarantee that it won't exceed that failure rate? • How inexpensively can you label data from the system?

Data flywheels

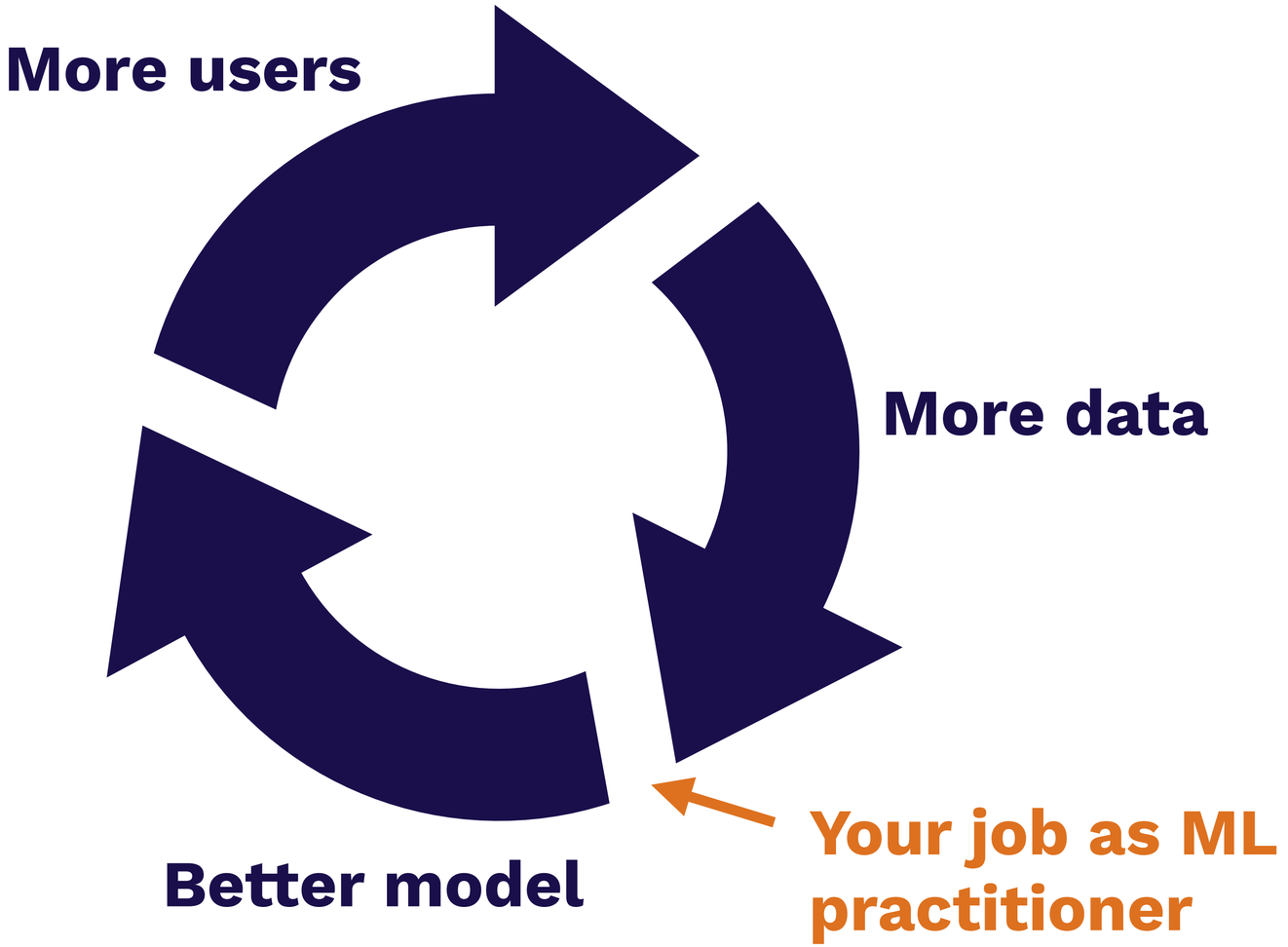


Data flywheels

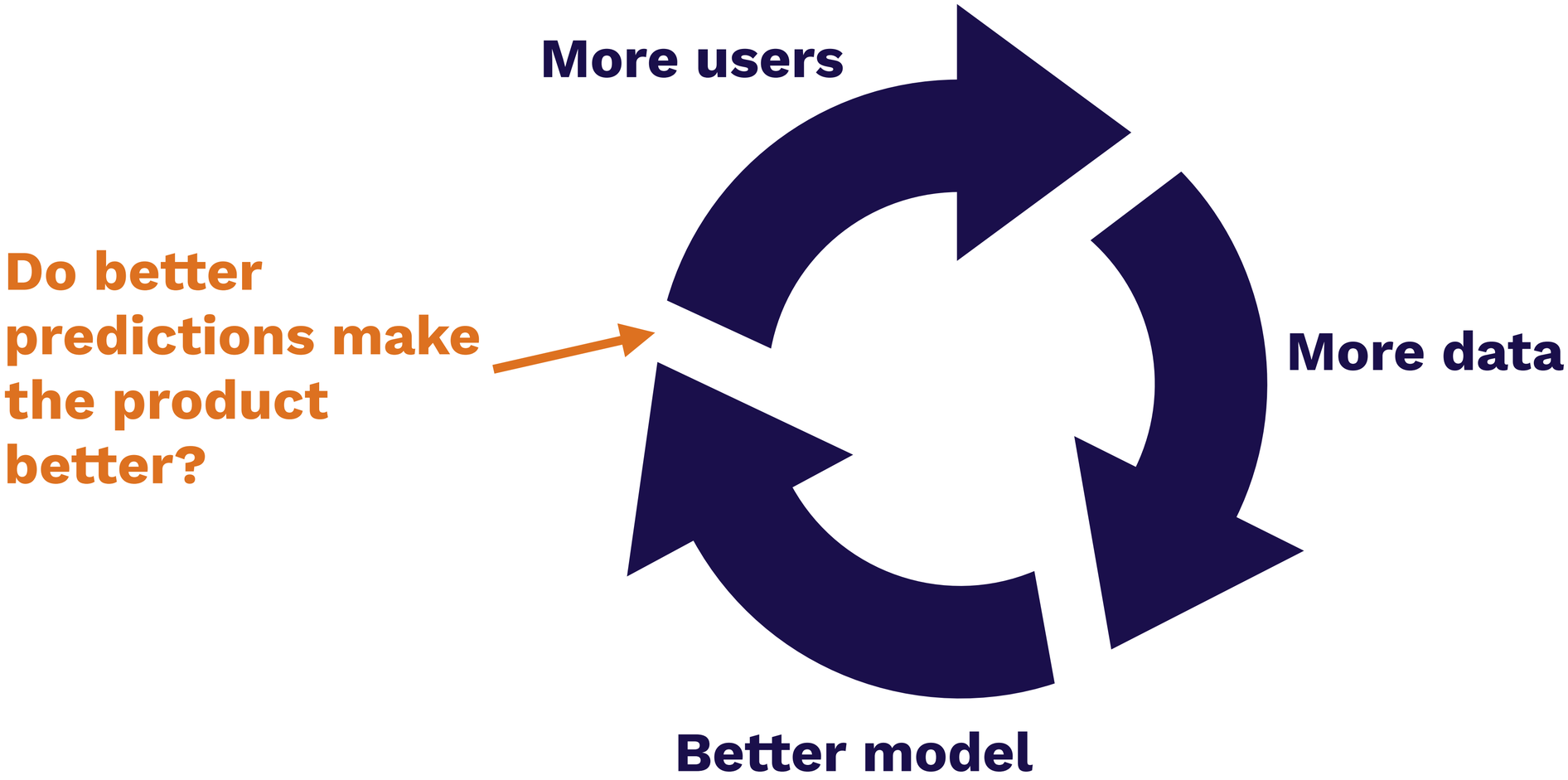


Do you have a data loop? Automatically collect data (and ideally labels) from your users

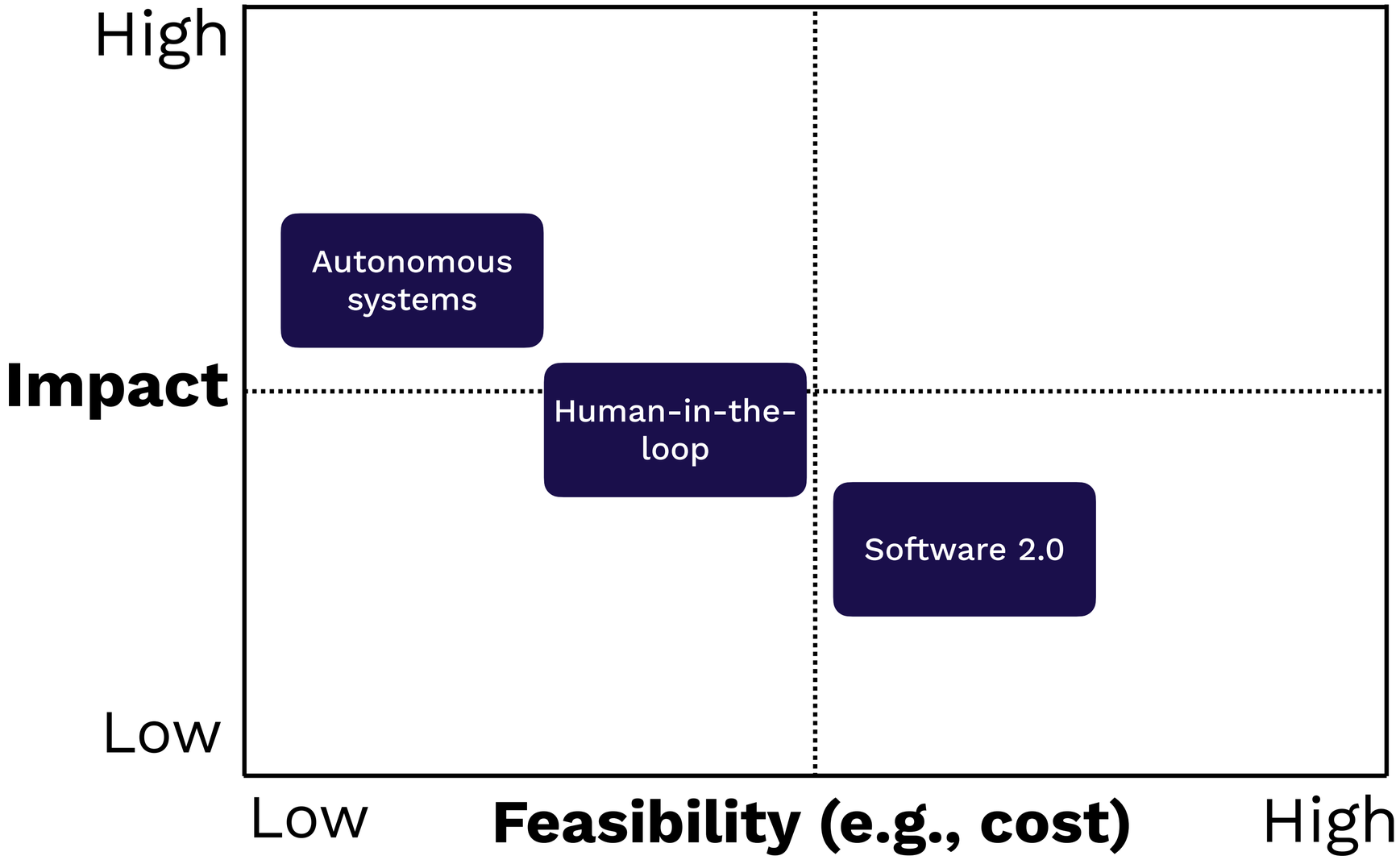
Data flywheels



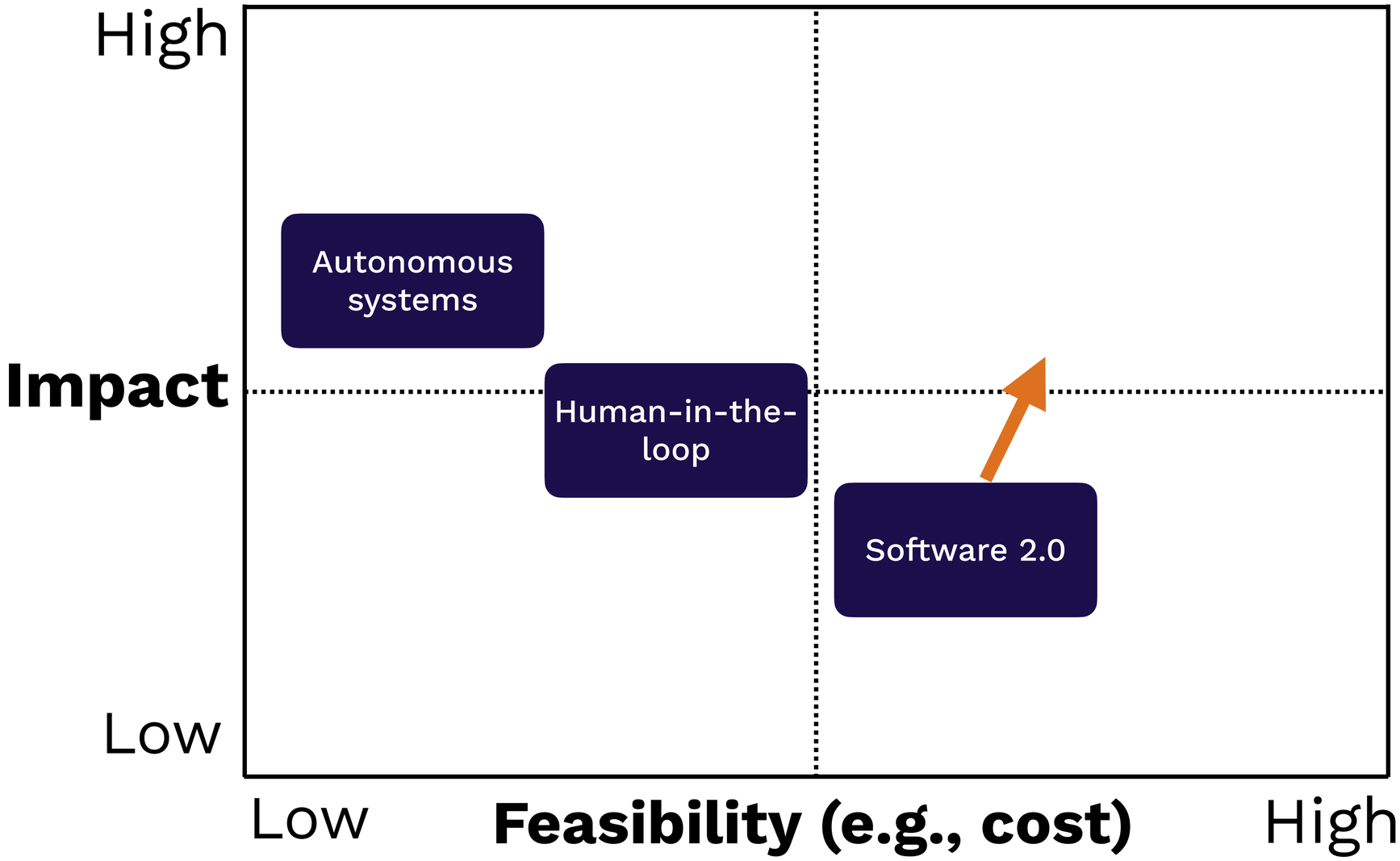
Data flywheels



Machine learning project archetypes

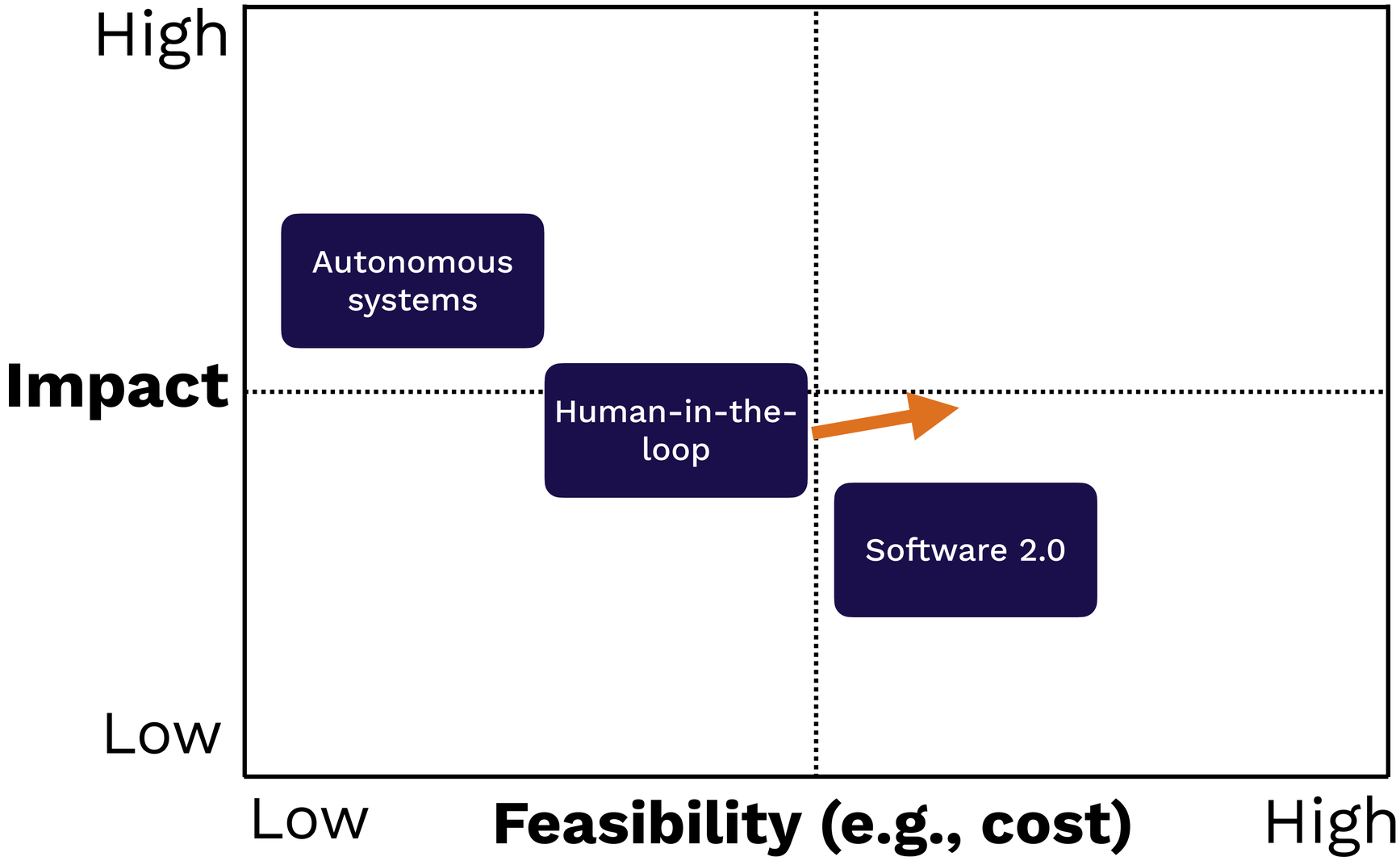


Machine learning product archetypes



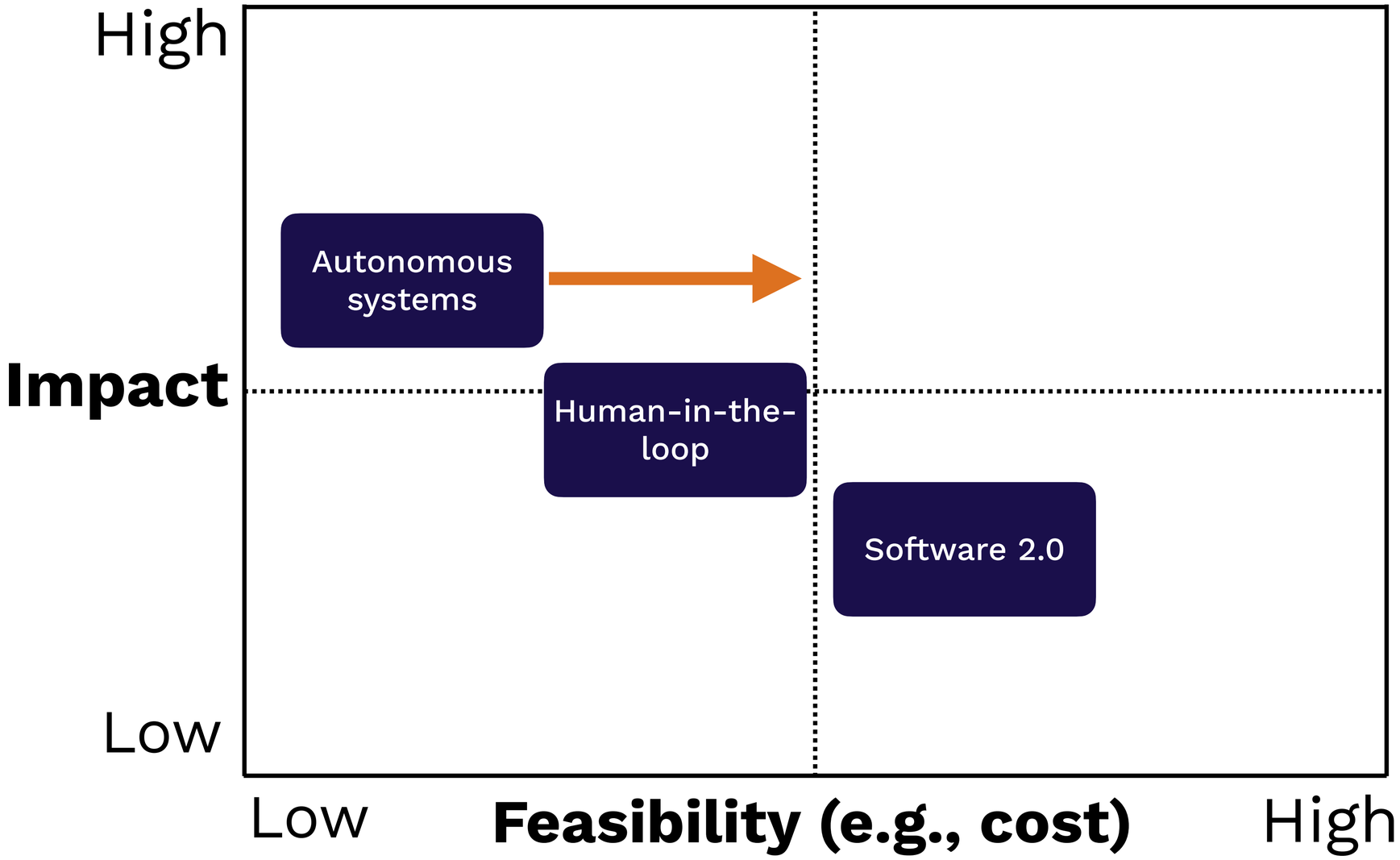
Implement a data loop that allows you to improve on this task and future ones

Machine learning product archetypes



Good product design. Release a ‘good enough’ version

Machine learning product archetypes



Add humans in the loop. Add guardrails and/or limit initial scope.

Despite all of this: just get started!

Avoiding tool fetishization

- You don’t need a perfect model to get started
- You don’t need perfect infrastructure, either
  - Just because Google or Uber does it, doesn’t mean you need to
  - For many use cases, just running your model every day and storing the predictions in a database is hard to beat
  - That’s why FSDL is a ML-powered product class, *not* an MLOps class

MLOps at reasonable scale

MLOps at Reasonable Scale, Jacopo Tagliabue, Ciro Greco, and Andrea Poloniolo. Stanford MLSys Seminar, https://www.youtube.com/watch?v=Ndxpo4PeEms

Summary

- ML adds complexity. Consider whether you really need it
- Make sure what you're working on is high impact, or else it might get killed

02

Lifecycle of a ML project

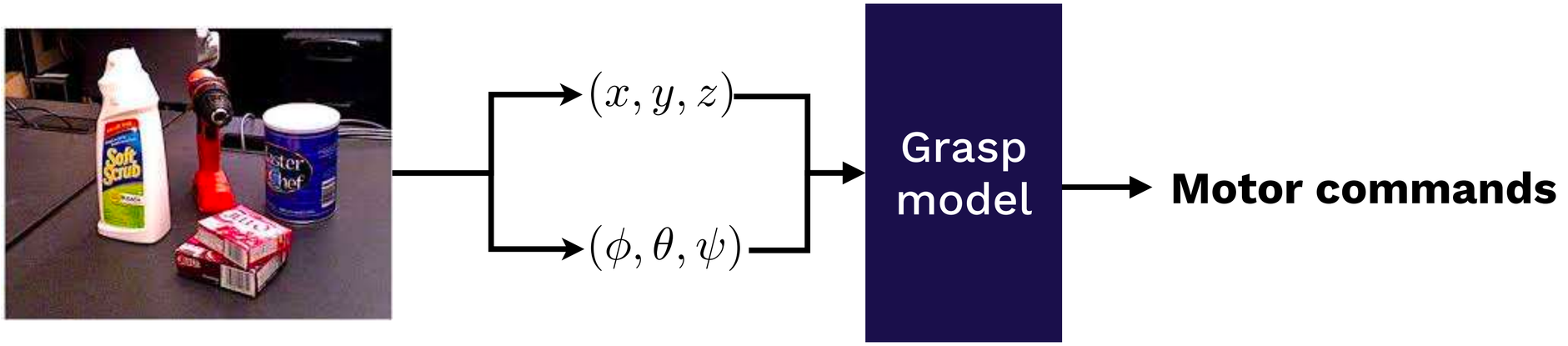
Running case study - pose estimation

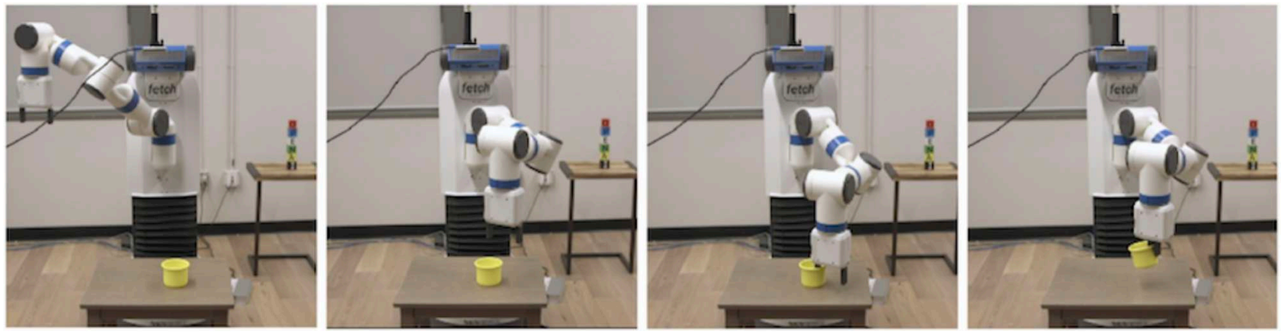
(x, y, z) **Position (L2 loss)**

(ϕ, θ, ψ) **Orientation (L2 loss)**

Xiang, Yu, et al. "PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes." arXiv preprint arXiv:1711.00199 (2017).

Full Stack Robotics works on grasping



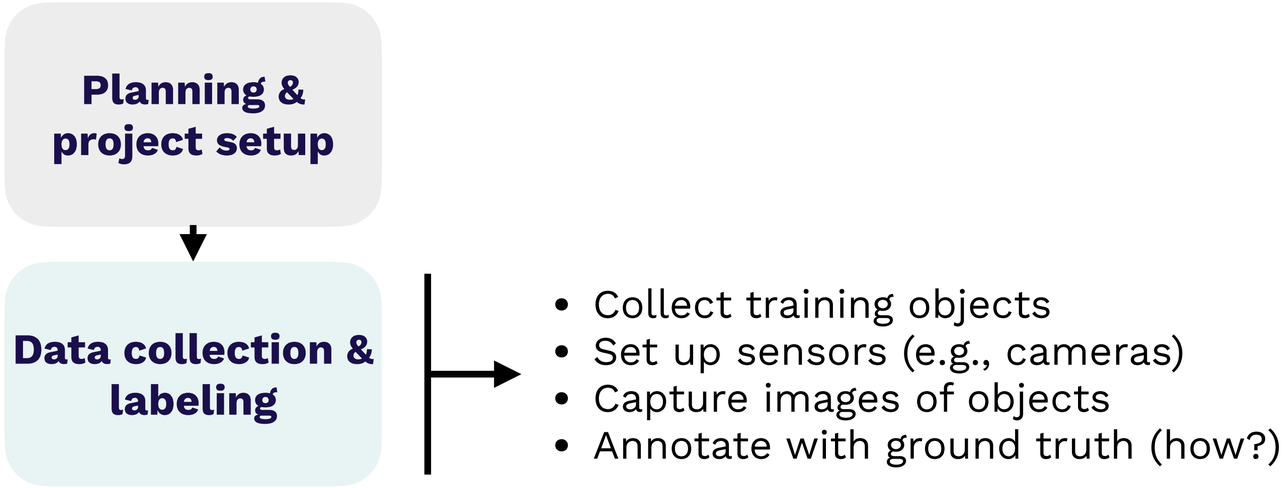


Lifecycle of a ML project

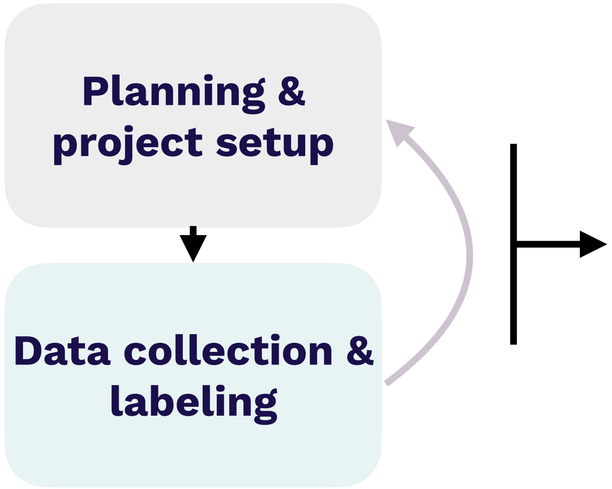
Planning & project setup

- Decide to work on pose estimation
- Determine requirements & goals
- Allocate resources
- Consider the ethical implications
- Etc.

Lifecycle of a ML project

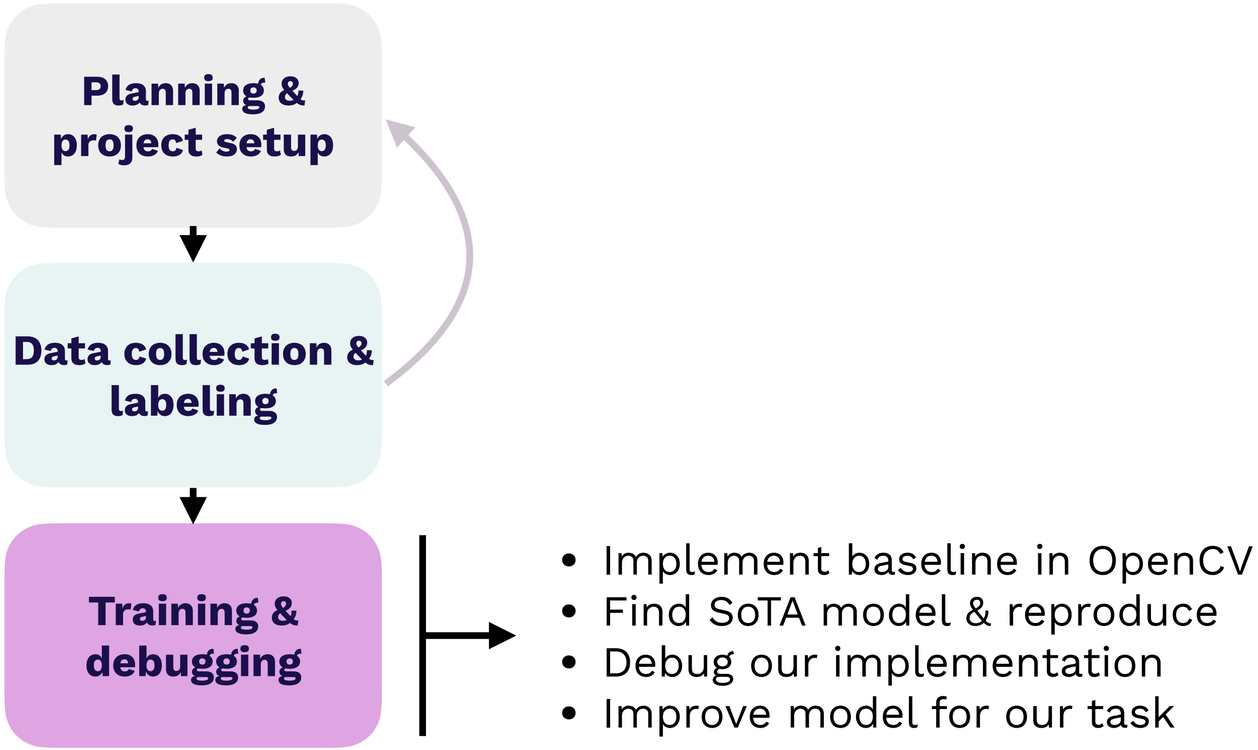


Lifecycle of a ML project

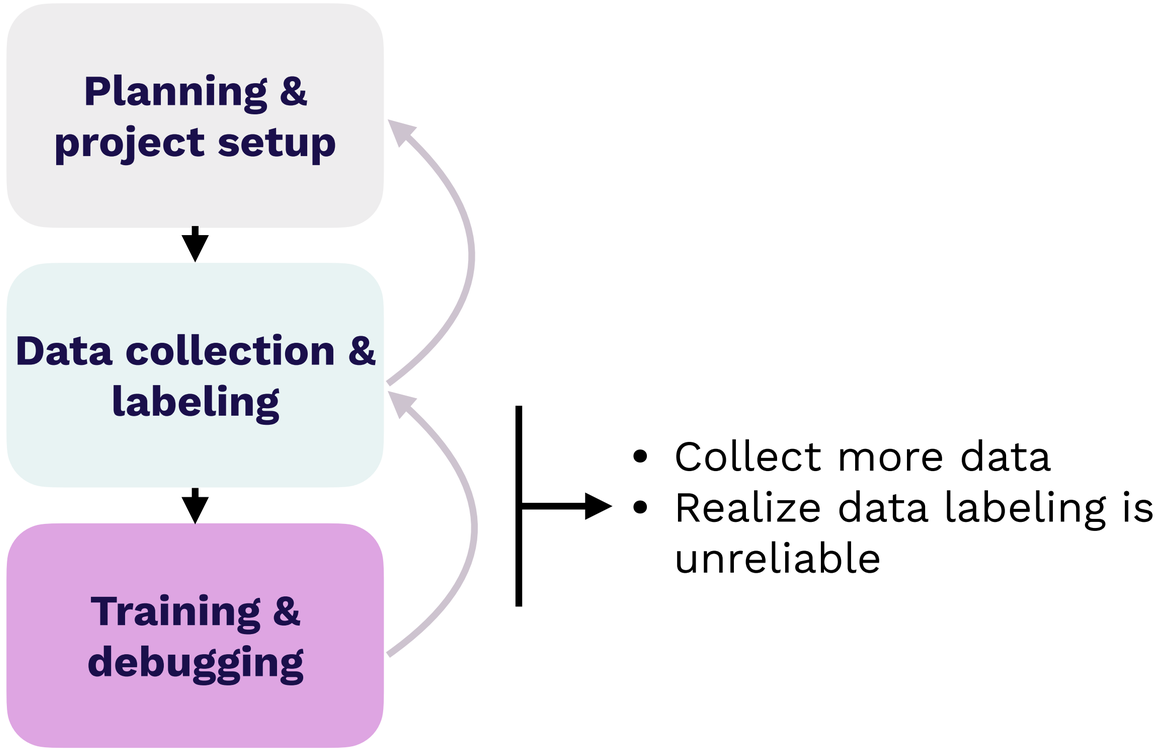


- Too hard to get data.
- Easier to label a different task (e.g., hard to annotate pose, easier to annotate per-pixel segmentation)

Lifecycle of a ML project

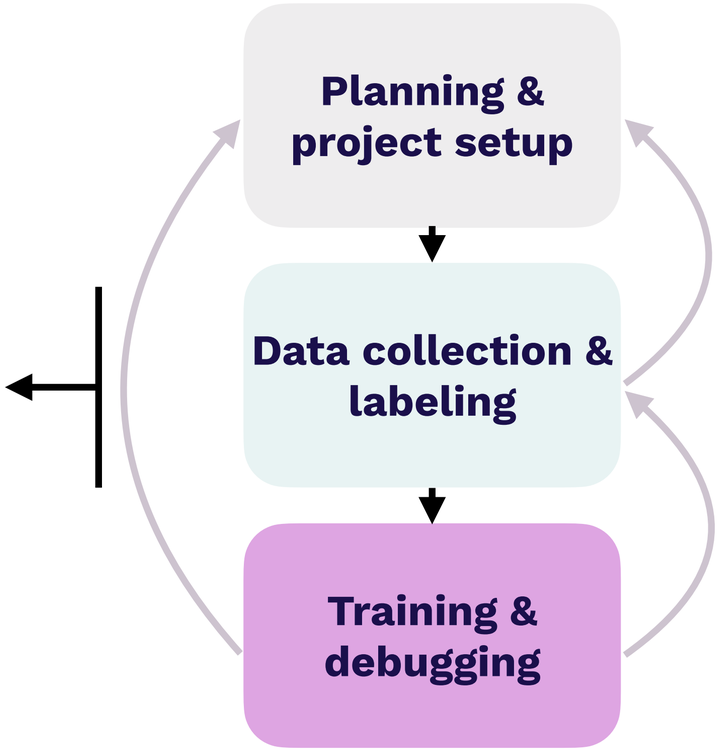


Lifecycle of a ML project

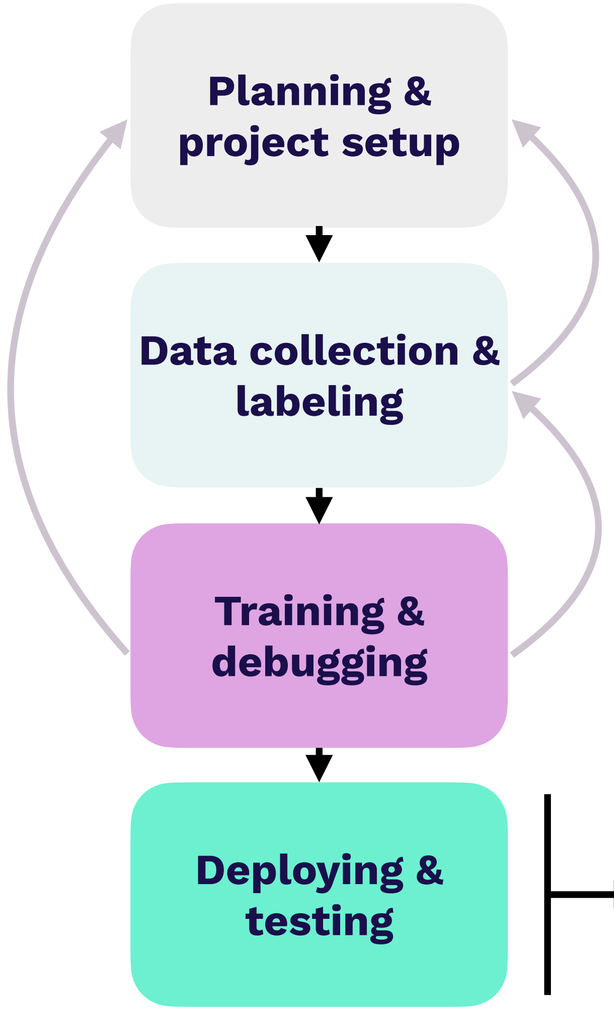


Lifecycle of a ML project

- Realize task is too hard
- Requirements trade off with each other - revisit which are most important

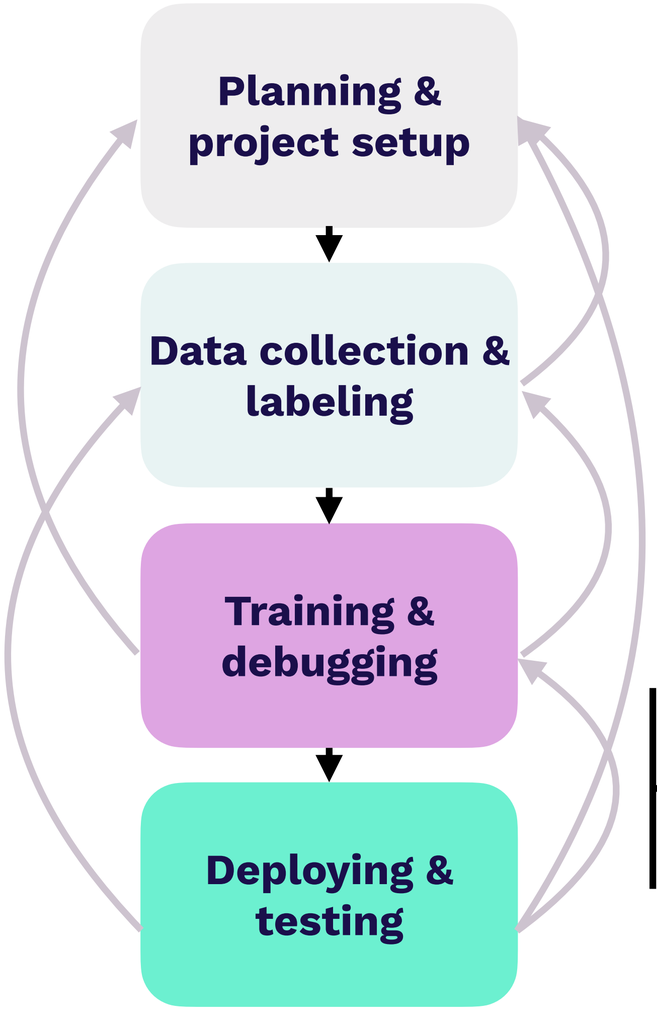


Lifecycle of a ML project



- Pilot in grasping system in the lab
- Write tests to prevent regressions and evaluate for biases
- Roll out in production

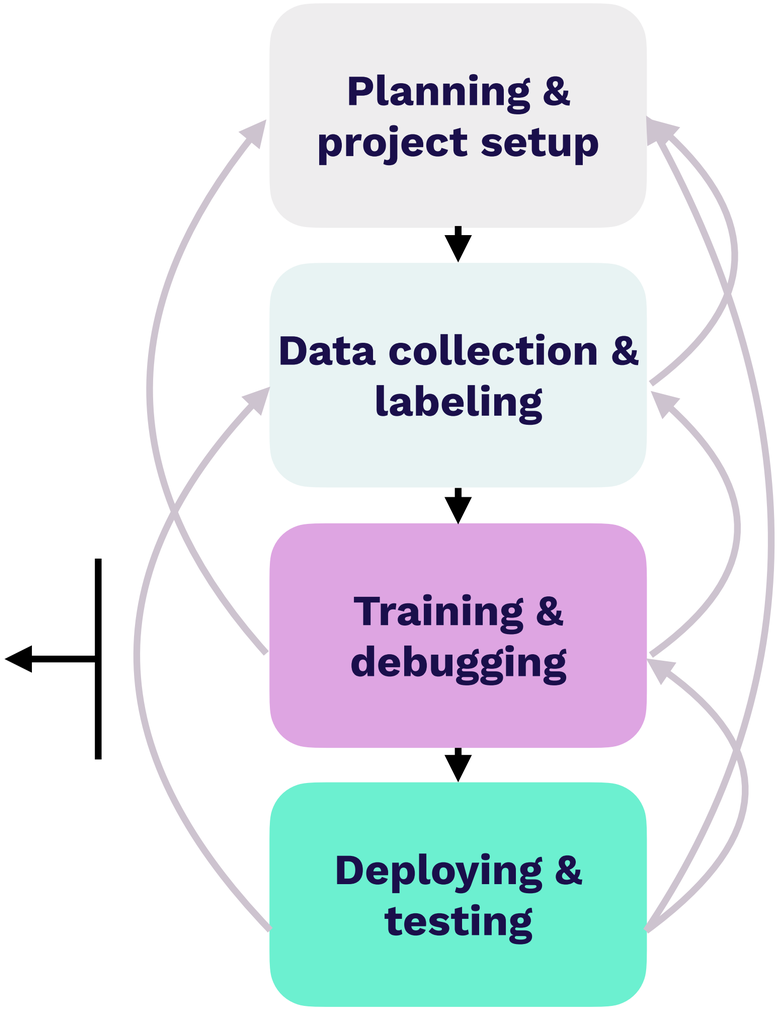
Lifecycle of a ML project



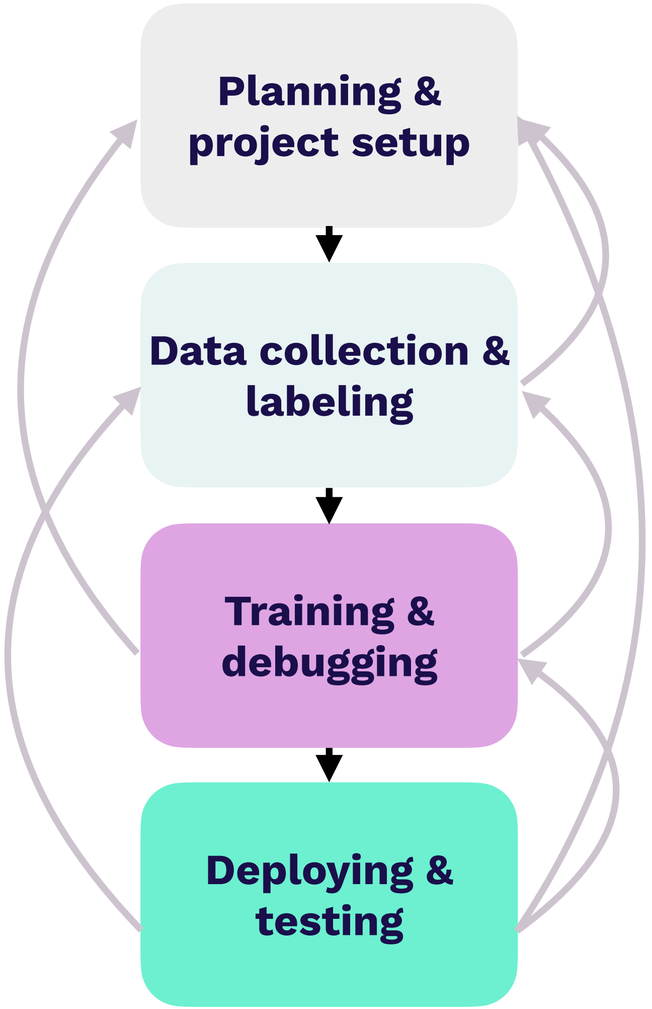
- Doesn’t work in the lab - keep improving accuracy of model

Lifecycle of a ML project

- Fix data mismatch between training data and data seen in deployment
- Collect more data
- Mine hard cases

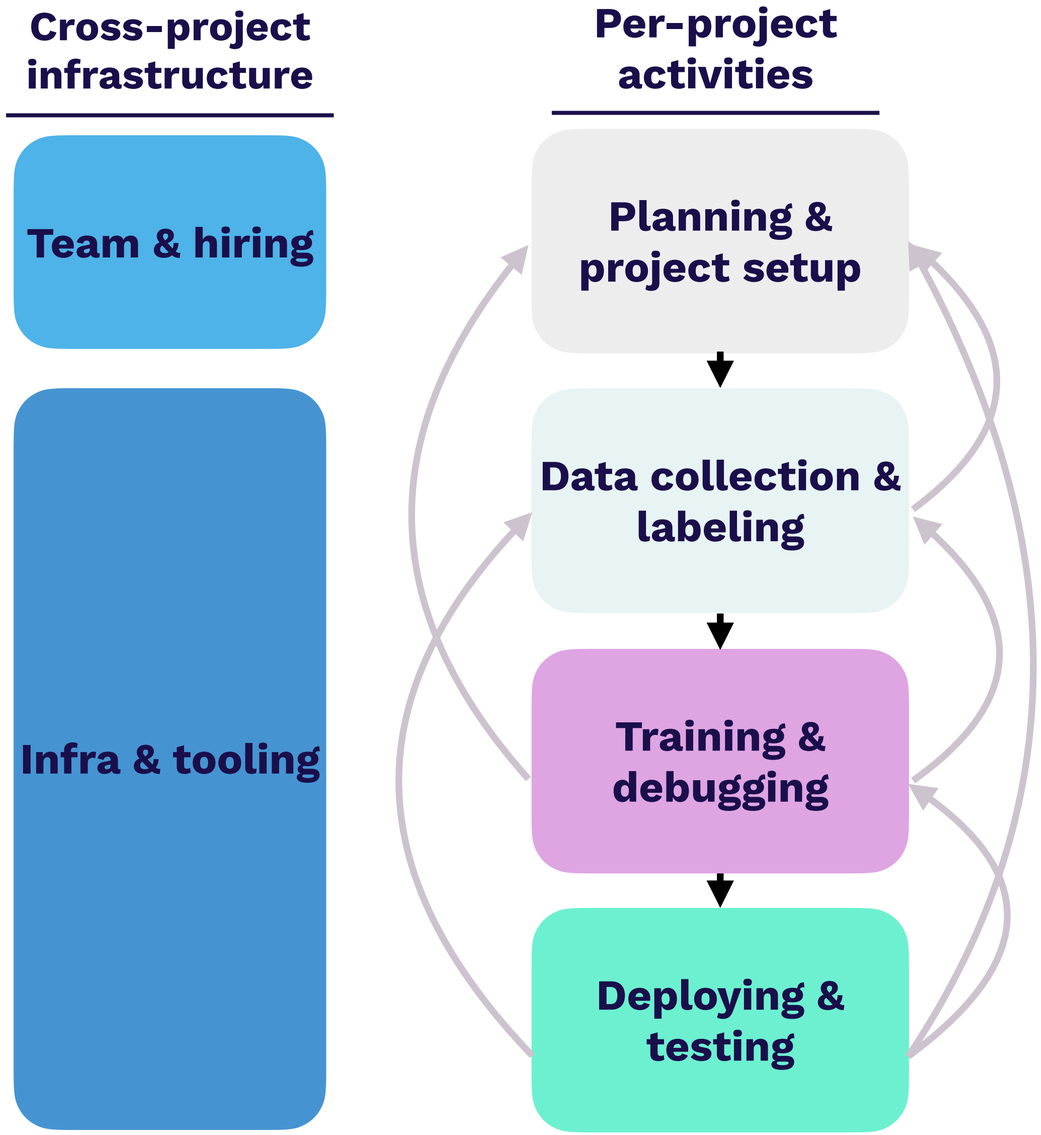


Lifecycle of a ML project



- The metric you picked doesn't actually drive downstream user behavior. Revisit the metric.
- Performance in the real world isn't great - revisit requirements (e.g., do we need to be faster or more accurate?

Lifecycle of a ML project



Wrapping up

- ML is complex, so use it because you need it and it will generate value. It’s not a cure-all
- In spite of this, you don’t need a perfect setup to get started
- Let’s walk through the project lifecycle and learn how to build ML-powered products together!