Full Stack AI Projects

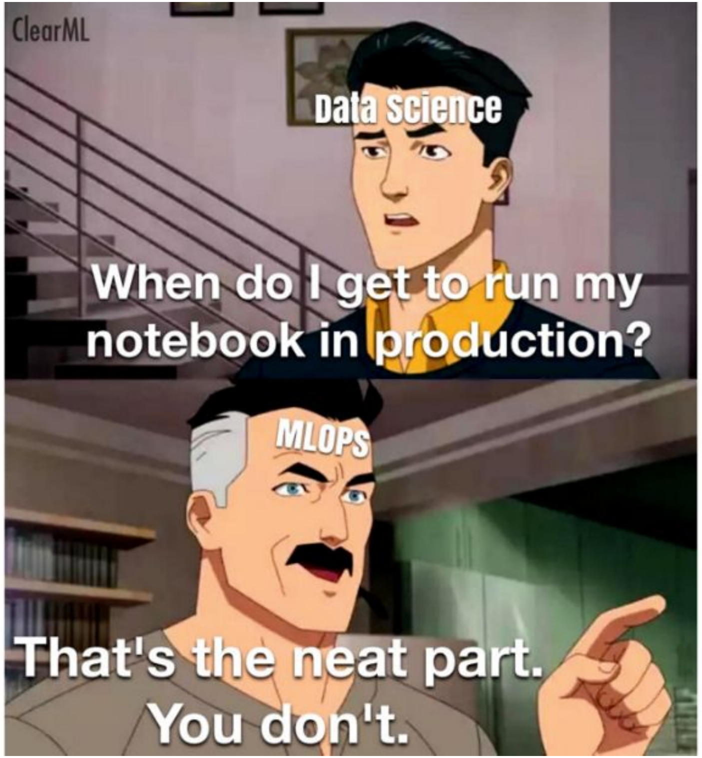
ML Teams

Henry Ruiz
GDE ML
@devharuiz
https://haruiz.github.io/

What we have discussed so far?

What is machine learning system design?

The process of **defining the <u>interface, algorithms, data, infrastructure, and hardware</u>** for a machine learning system to satisfy specified requirements.

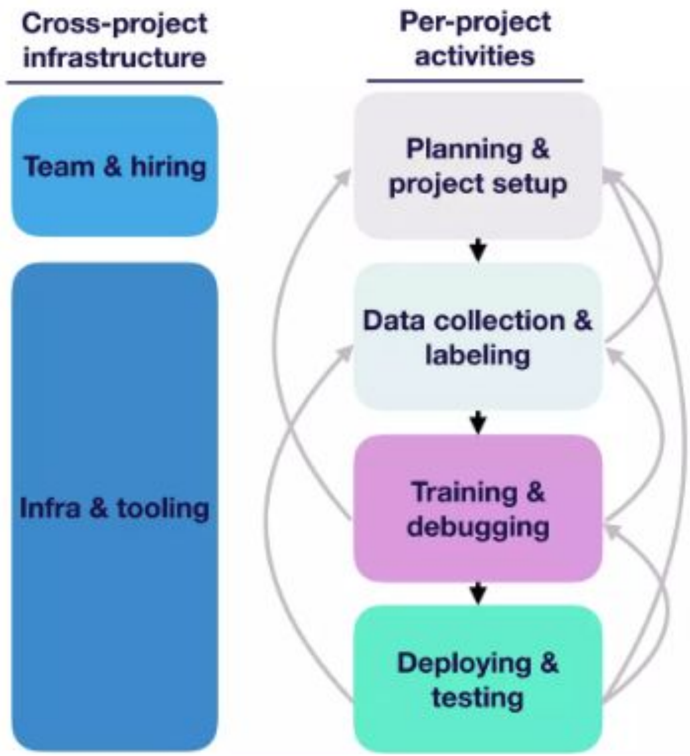


So, What is required to design and build an ML system?

What we have discussed so far?

Define a methodology

Like any other software solution, ML systems require a **well-structured methodology** to maximize the success rate of the implementation.



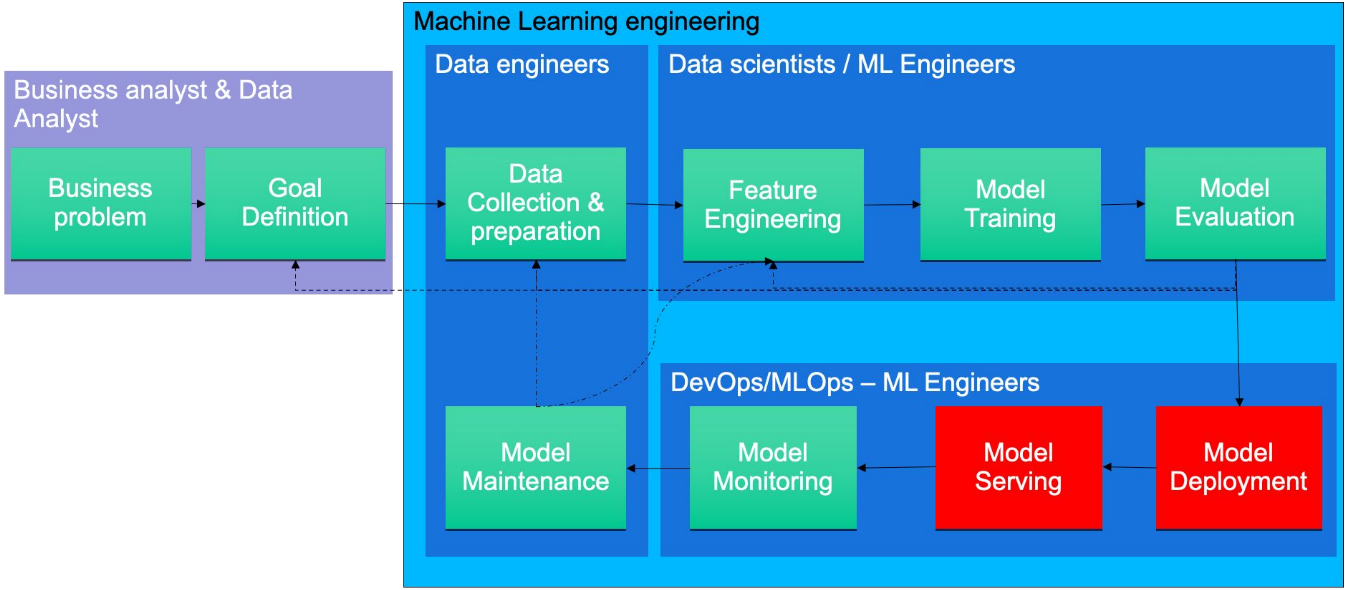
What we have discussed so far?

So, What is required to design and build an ML system?

<u>Set up a team</u>

Managing and leading ML and Data Science teams require unique skills.

ML teams have diverse roles.

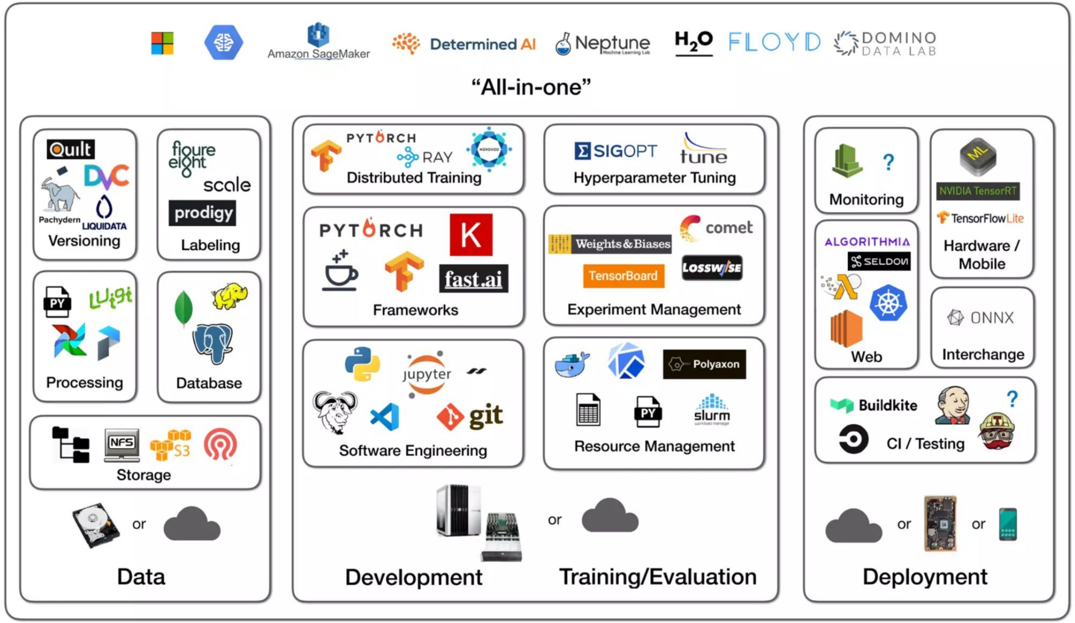


What we have discussed so far?

Define Development infrastructure & Tooling

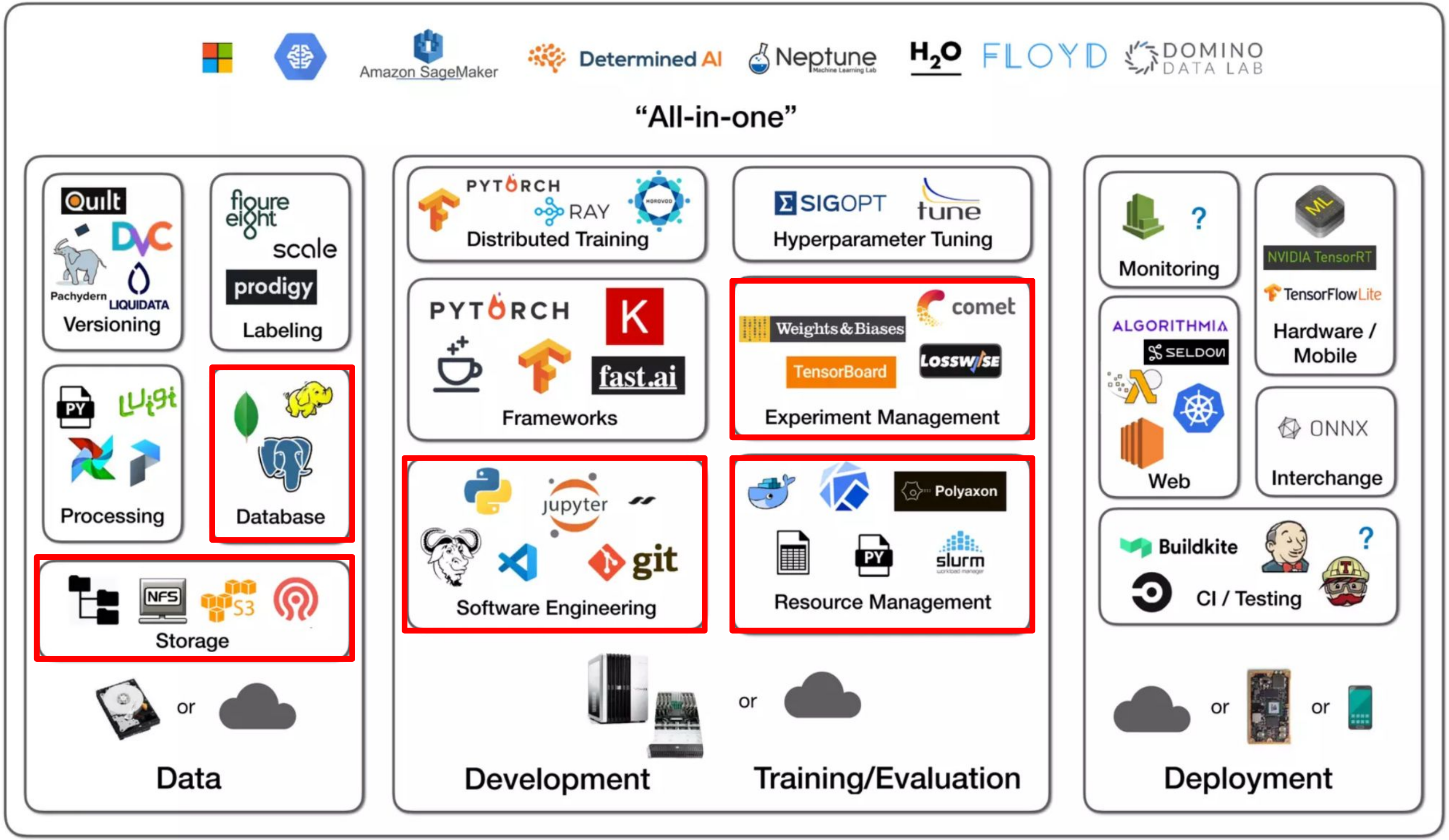
The number of **available tools to work with ML seems endless.**

Selecting the appropriate tools depends on:
the kind of problem, **type of solution, deployment scenario, capacity building**, team experience, cost, hardware and software infrastructure, etc.



Data management tools

Credits: https://fullstackdeeplearning.com/



Last class…

Data Management Tools

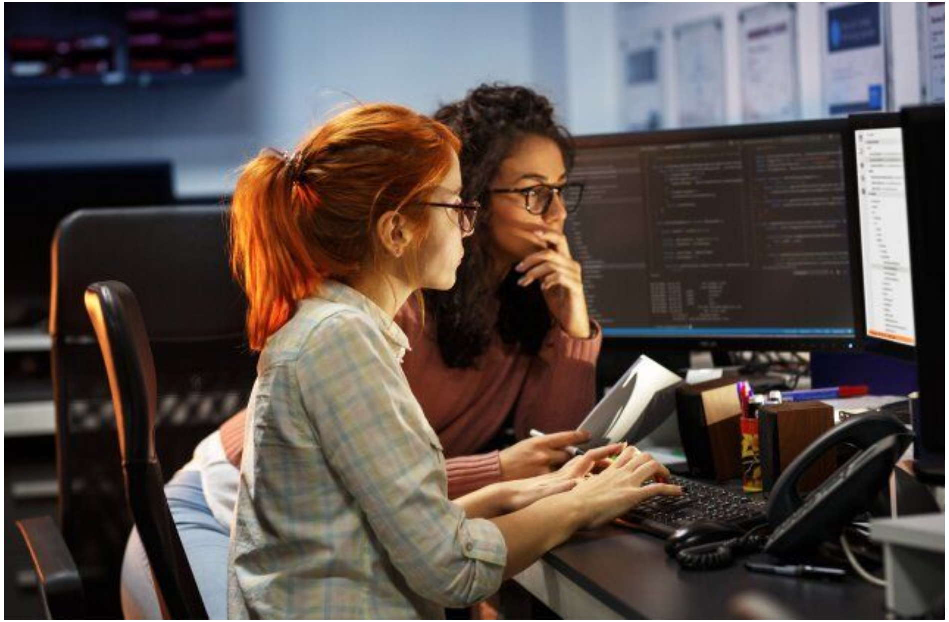
Data management is the **process of collecting, storing, organizing, maintaining, and utilizing data effectively and efficiently**.In ML, it is an important component, since data is the fuel of the algorithms.

Data refers *to* ***a collection of facts, information, or statistics that are often represented in a numerical or digital form.*** Data can be in various forms, such as text, images, audio, video, or any other format ***that can be stored and analyzed using digital technologies.***

Data is an important resource for decision-making in many fields, including business, science, healthcare, education, and government. With the advent of big data and advanced analytics technologies, the analysis and interpretation of data have become increasingly important for organizations ***to gain insights into patterns,*** **trends, and behaviors that can inform decision-making and drive innovation.**

Building any product is hard:

- You have to hire great people.
- You have to manage and develop those people.
- You have to manage your team's output and make sure your vectors are aligned.
- You have to make good long-term technical choices and manage technical debt.
- You have to manage expectations from leadership.
- You have to define and communicate requirements with stakeholders.



Credits: https://fullstackdeeplearning.com/course/2022/lecture-8-teams-and-pm/

Machine Learning (ML) adds complexity to that process:

- ML talent is expensive and scarce.
- ML teams have a diverse set of roles.
- Projects have unclear timelines and high uncertainty.
- The field is moving fast, and ML is the "**high-interest credit card of technical debt**."
- Leadership often doesn't understand ML.
- ML products fail in ways that are hard for laypeople to understand.

Credits: https://fullstackdeeplearning.com/course/2022/lecture-8-teams-and-pm/

ML Teams

Developing successful machine learning projects requires a variety of expertise, and hence, it is an extremely collaborative job.

Designing a problem statement needs the company to ask critical questions such as:

Why build an ML model?

How to get the required data?

How to design the systems infrastructure? and more.

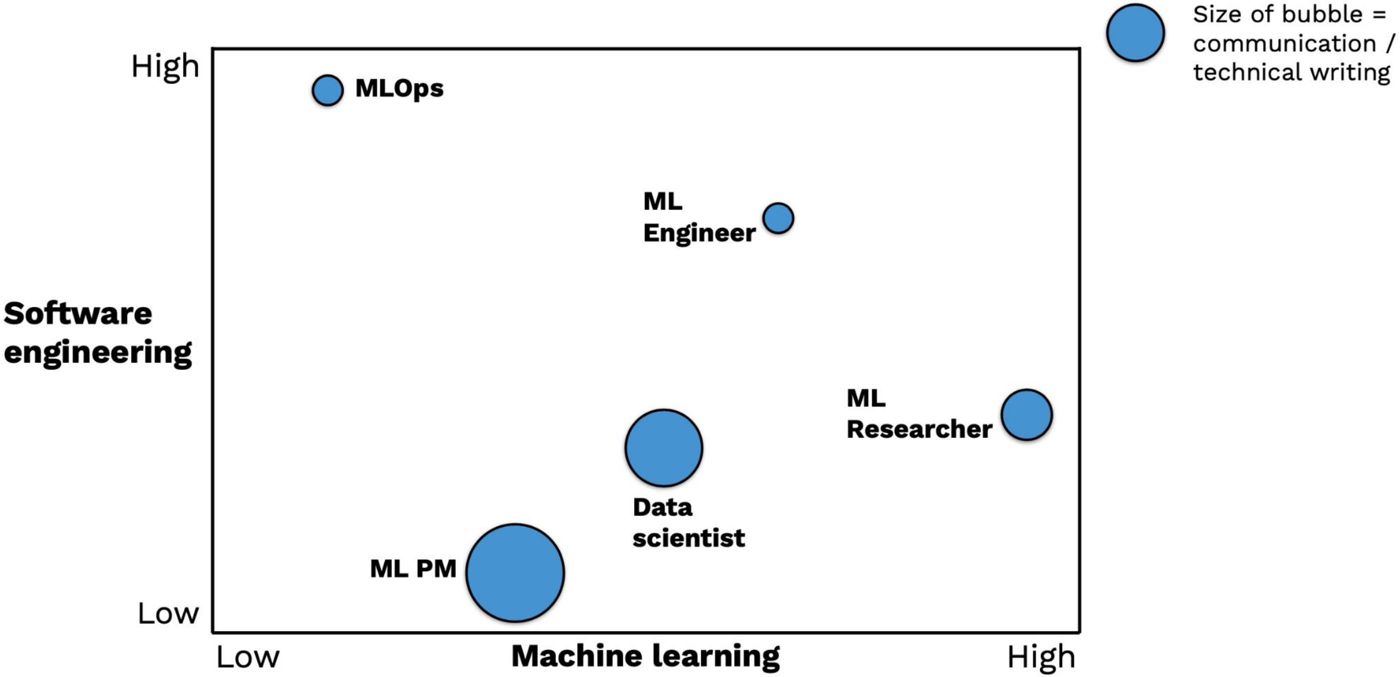
Common roles

Role	Job Function	Work product	Commonly used tools
ML product manager	Work with ML team, business, users, data owners to prioritize & execute projects	Design docs, wireframes, work plans	Jira, etc
MLOps / ML platform	Build the infrastructure to make models easier to deploy, more scalable, etc	ML infrastructure	AWS, Kafka, ML tooling vendors, etc.
ML engineer	Train, deploy, & maintain prediction models	Prediction system running on real data in production	Tensorflow, Docker
ML researcher	Train prediction models (often forward looking or not production-critical)	Prediction model & report describing it	Tensorflow, pytorch, Jupyter
Data scientist	Blanket term used to describe all of the above. In some orgs, means answering business questions using analytics	Prediction model or report	SQL, Excel, Jupyter, Pandas, SKLearn, Tensorflow

Credits: https://fullstackdeeplearning.com/course/2022/lecture-8-teams-and-pm/

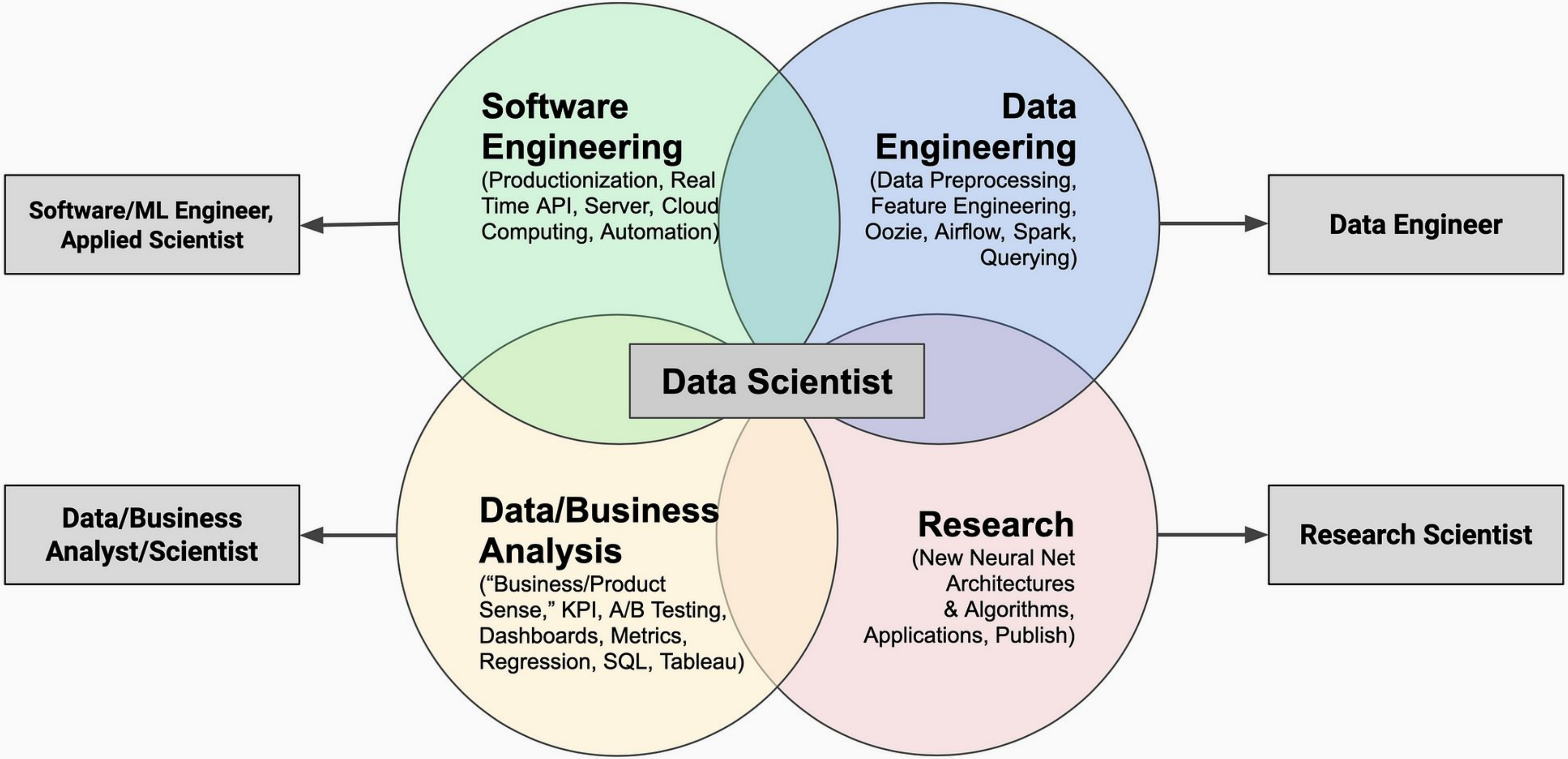
Skills required

What skills are needed for these roles? The chart below displays a nice visual - where the horizontal axis is the level of ML expertise and the size of the bubble is the level of communication and technical writing (the bigger, the better).



Skills required

- The **MLOps** is primarily a software engineering role, which often comes from a standard software engineering pipeline.
- The **ML Engineer** requires a rare mix of ML and Software Engineering skills. This person is either an engineer with significant self-teaching OR a science/engineering Ph.D. who works as a traditional software engineer after graduate school.
- The **ML Researcher** is an ML expert who usually has an MS or Ph.D. degree in Computer Science or Statistics or finishes an industrial fellowship program.
- The **ML Product Manager** is just like a traditional Product Manager but with a deep knowledge of the ML development process and mindset.
- The **Data Scientist** role constitutes a wide range of backgrounds, from undergraduate to Ph.D. students.


Software/ML Engineer, Applied Scientist
Software Engineering
(Productionization, Real Time API, Server, Cloud Computing, Automation)
Data Engineering
(Data Preprocessing, Feature Engineering, Oozie, Airflow, Spark, Querying)
Data Engineer
Data Scientist
Data/Business Analyst/Scientist
Data/Business Analysis
("Business/Product Sense," KPI, A/B Testing, Dashboards, Metrics, Regression, SQL, Tableau)
Research
(New Neural Net Architectures & Algorithms, Applications, Publish)
Research Scientist

Hiring

The AI Talent Gap

In 2018 (when we started FSDL), the AI talent gap was the main story. There were so few people who understood this technology, so the biggest block for organizations was that they couldn't find people who were good at ML.

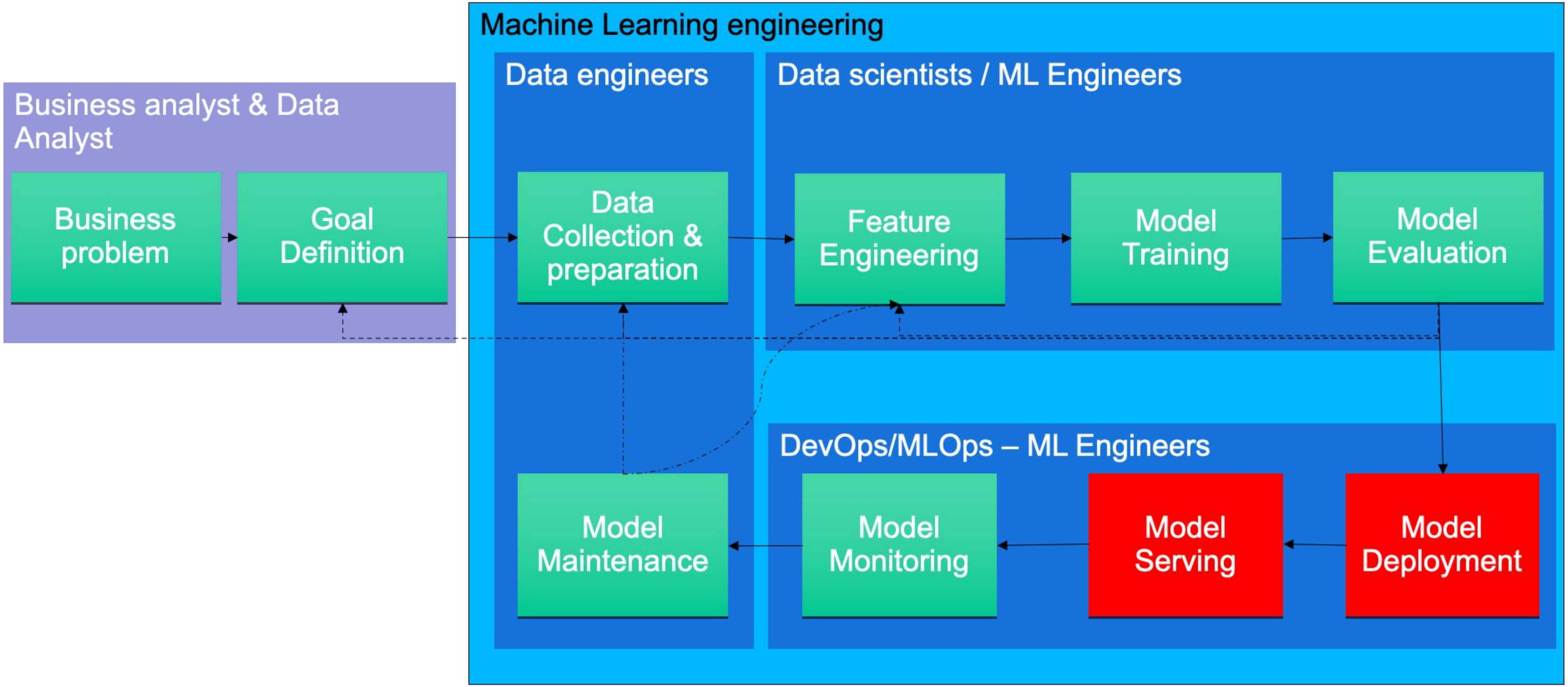
In 2022, the AI talent gap persists. But it tends to be less of a blocker than it used to be because we have had four years of folks switching careers into ML and software engineers emerging from undergraduate with at least a couple of ML classes under their belts.

The gap tends to be in folks that understand more than just the underlying technology but also have experience in seeing how ML fails and how to make ML successful when it's deployed. That's the reality of how difficult it is to hire ML folks today, especially those with **production experience**.

Job Description

- Job Description (Unicorn Machine Learning Engineer)
  - Duties
    - Keep up with the state of the art
    - Implement models from scratch
    - Deep understanding of mathematics & ability to come up with new models
    - Build tooling & infrastructure for the ML team
    - Build data pipelines for the ML team
    - Deploy & monitor models into production
  - Requirements
    - PhD
    - At least 4 years tensorflow experience
    - At least 4 years as a software engineer
    - Publications in top ML conference
    - Experience building large-scale distributed systems

ML Pipeline


Machine Learning engineering
Business analyst & Data Analyst
Business problem
Goal Definition
Data engineers
Data Collection & preparation
Data scientists / ML Engineers
Feature Engineering
Model Training
Model Evaluation
DevOps/MLOps – ML Engineers
Model Maintenance
Model Monitoring
Model Serving
Model Deployment

Hiring Process

While there's no perfect way to **hire ML engineers**, there's definitely a wrong way to hire them, with extensive job descriptions that demand only the best qualifications (seen above). Certainly, there are many good examples of this bad practice floating around.

- Rather than this unrealistic process, consider hiring for software engineering skills, an interest in ML, and a desire to learn. You can always train people in the art and science of ML, especially when they come with strong software engineering fundamentals.
- Another option is to consider adding junior talent, as many recent grads come out with good ML knowledge nowadays.
- Finally, and most importantly, be more specific about what you need the position and professional to do. It's impossible to find one person that can do everything from full-fledged DevOps to algorithm development.

Hiring Process

To **hire ML researchers**, here are our tips:

- Evaluate the quality of publications, over the quantity, with an eye toward the originality of the ideas, the execution, etc.
- Prioritize researchers that focus on important problems instead of trendy problems.
- Experience outside academia is also a positive, as these researchers may be able to transition to industry more effectively.
- Finally, keep an open mind about research talent and consider talented people without PhDs or from adjacent fields like physics, statistics, etc.

Where to find these profiles

To find quality candidates for these roles, here are some ideas for sourcing:

- Use standard sources like LinkedIn, recruiters, on-campus recruiting, etc.
- Monitor arXiv and top conferences and flag the first authors of papers you like.
- Look for good implementations of papers you like.
- Attend ML research conferences (NeurIPS, ICML, ICLR).

As you interview candidates for ML roles, try to **validate your hypotheses of their strengths while testing a minimum bar on weaker aspects**. For example, ensure ML researchers can think creatively about new ML problems while ensuring they meet a baseline for code quality. It's essential to test ML knowledge and software engineering skills for all industry professionals, though the relative strengths can vary.

https://huyenchip.com/ml-interviews-book/

Finding a Job

To find an ML job, you can take a look at the following sources:

- Standard sources such as LinkedIn, recruiters, on-campus recruiting, etc.
- ML research conferences (NeurIPS, ICLR, ICML).
- Apply directly (remember, there's a talent gap!).

Standing out for competitive roles can be tricky! Here are some tips (in increasing order of impressiveness) that you can apply to differentiate yourself:

1. Exhibit ML interest (e.g., conference attendance, online course certificates, etc.).
2. Build software engineering skills (e.g., at a well-known software company).
3. Show you have a broad knowledge of ML (e.g., write blog posts synthesizing a research area).
4. Demonstrate ability to get ML projects done (e.g., create side projects, re-implement papers).
5. Prove you can think creatively in ML (e.g., win Kaggle competitions, publish papers).