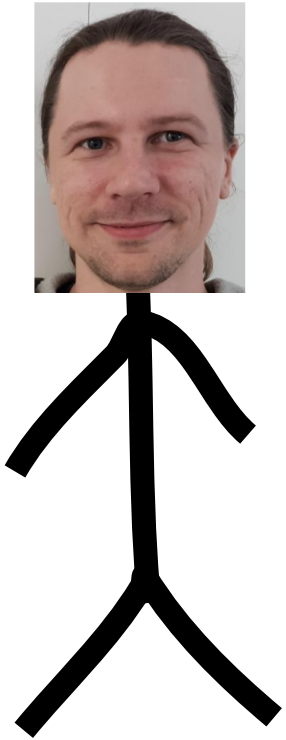
Full-stack data scientist

Alexey Grigorev
02.07.2020

https://www.goodfreephotos.com/vector-images/unicorn-with-rainbow-vector-clipart.png.php



- I’m Alexey
- Data Scientist

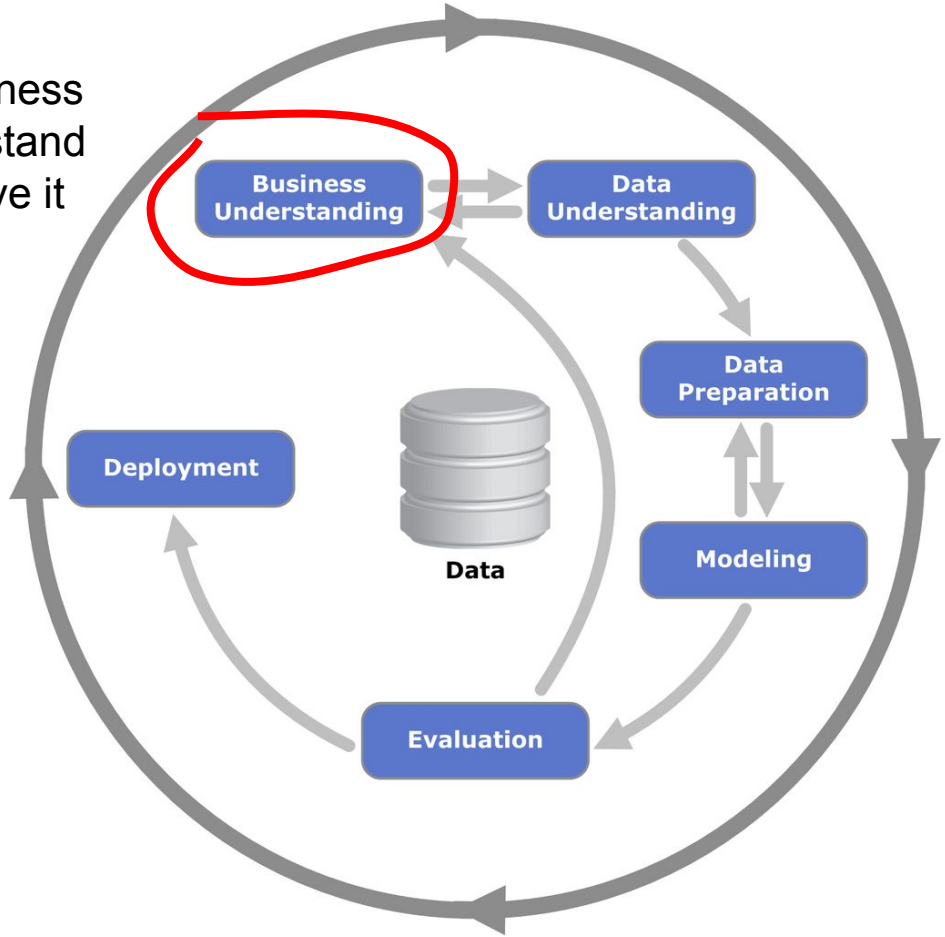
Plan

- Data science process
- Roles in the team
- Full-stack data scientist: Jack of all trades
- Becoming full-stack

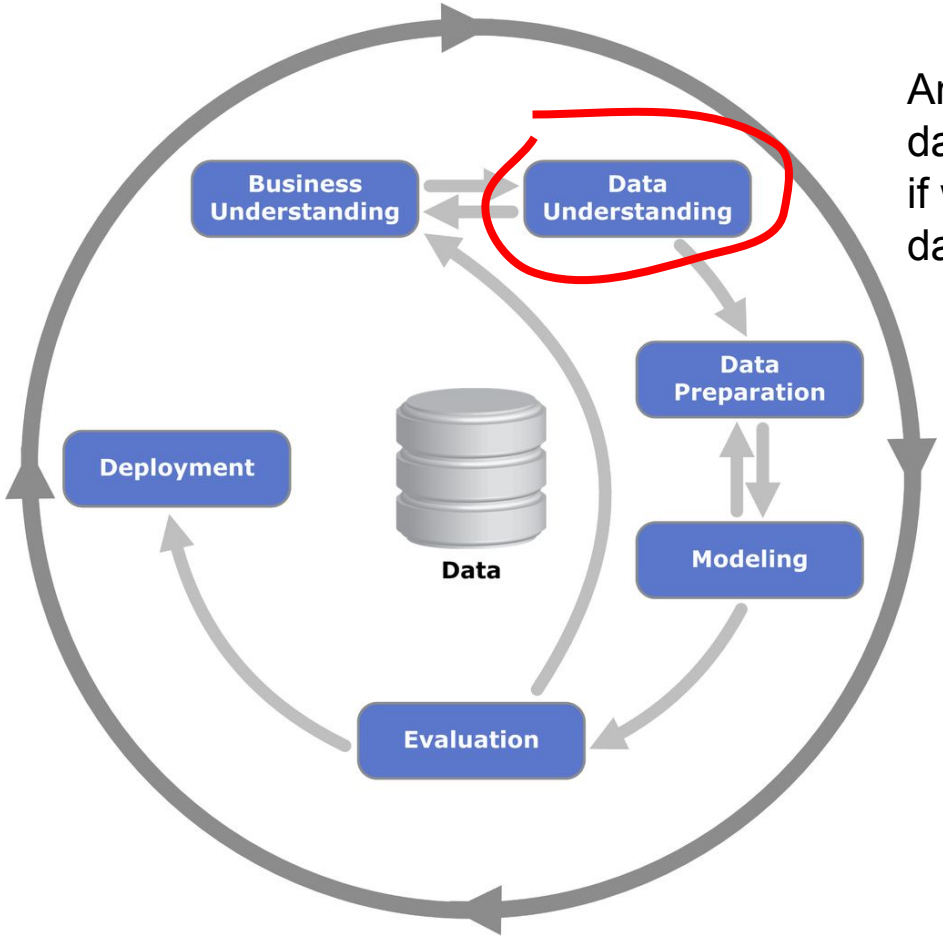
Data science

Picture: CRISP-DM

Identify the business problem, understand how we can solve it

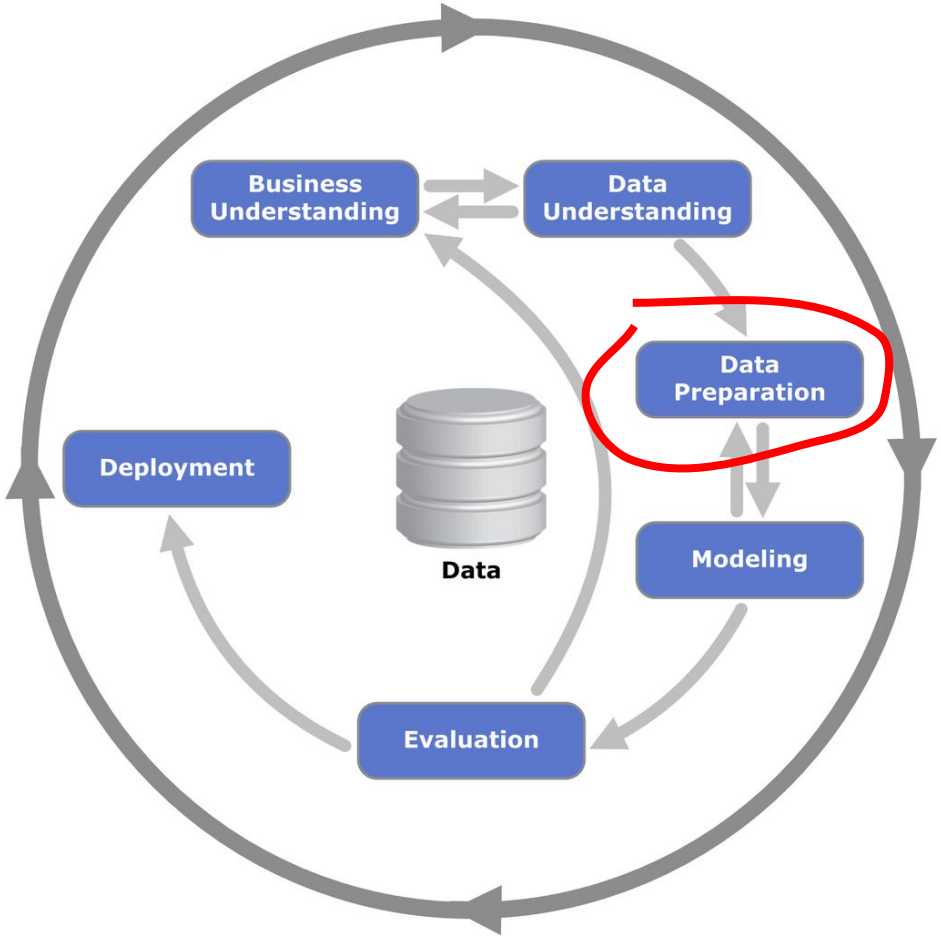


Picture: CRISP-DM



Analyze available data sources, decide if we need to get more data

Picture: CRISP-DM

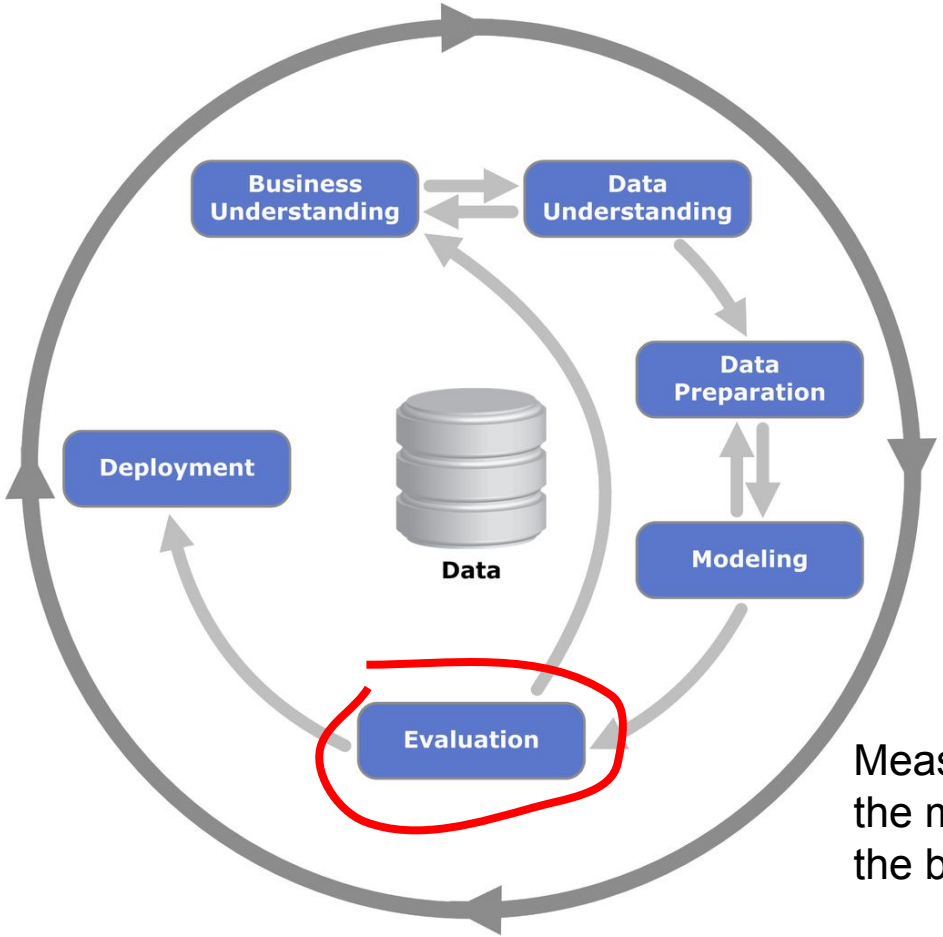


Transform the data so it can be put into a ML algorithm

Picture: CRISP-DM

Training the models: the actual machine learning happens here

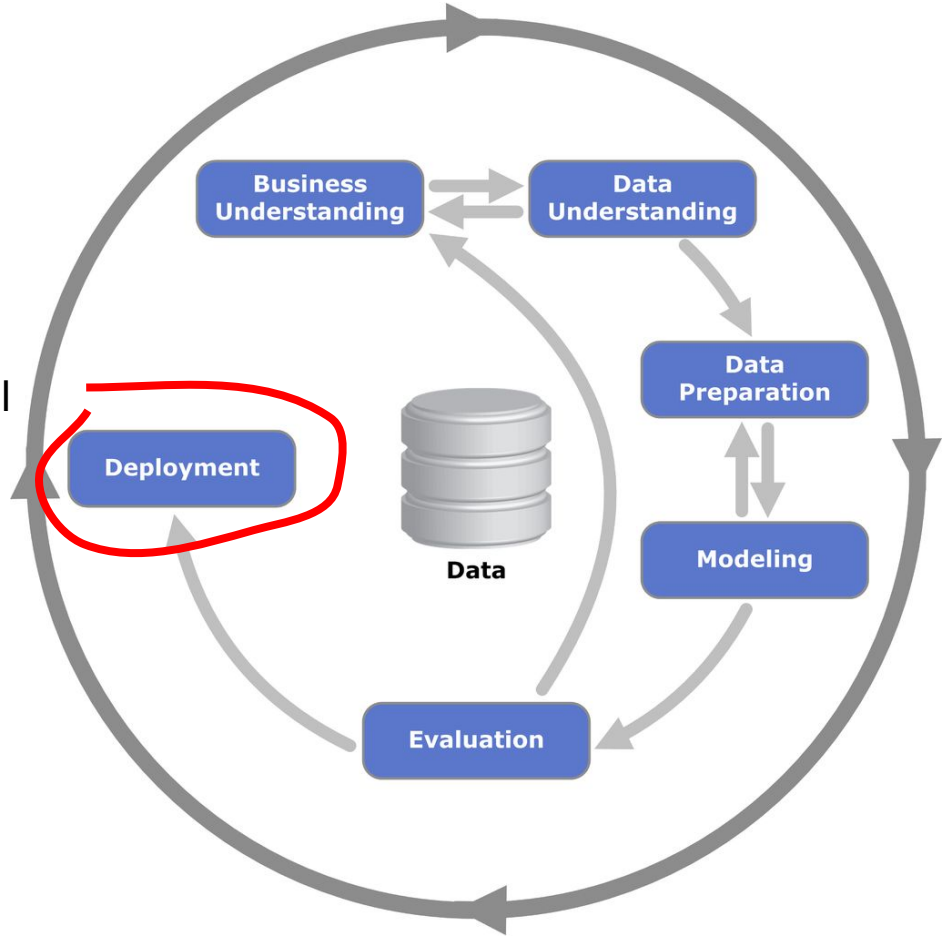
Picture: CRISP-DM



Measure how well the model solves the business problem

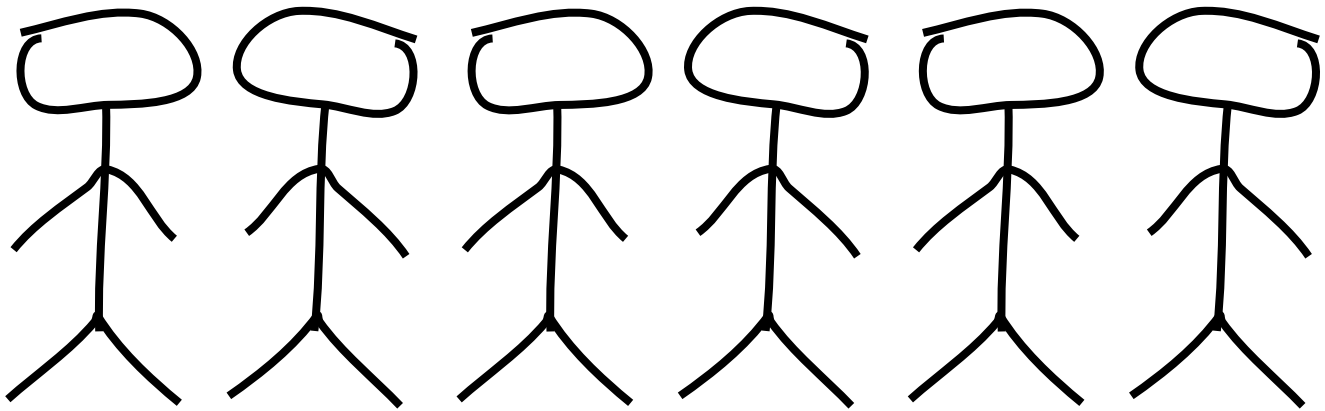
Picture: CRISP-DM

Deploy the model to production



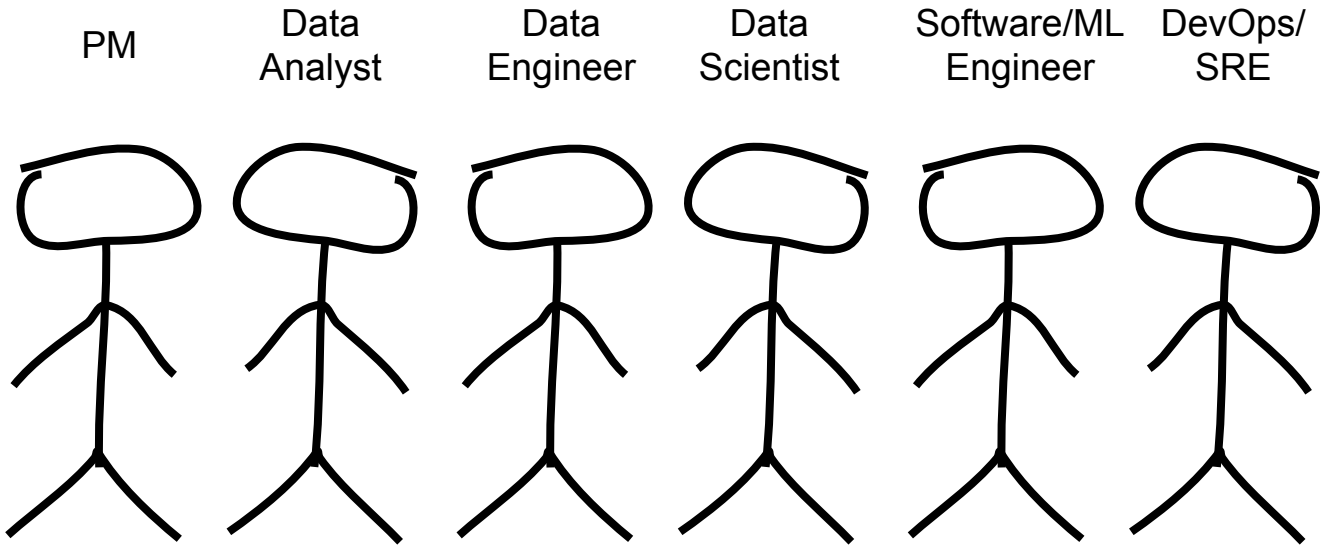
Picture: CRISP-DM

We need a team for that!



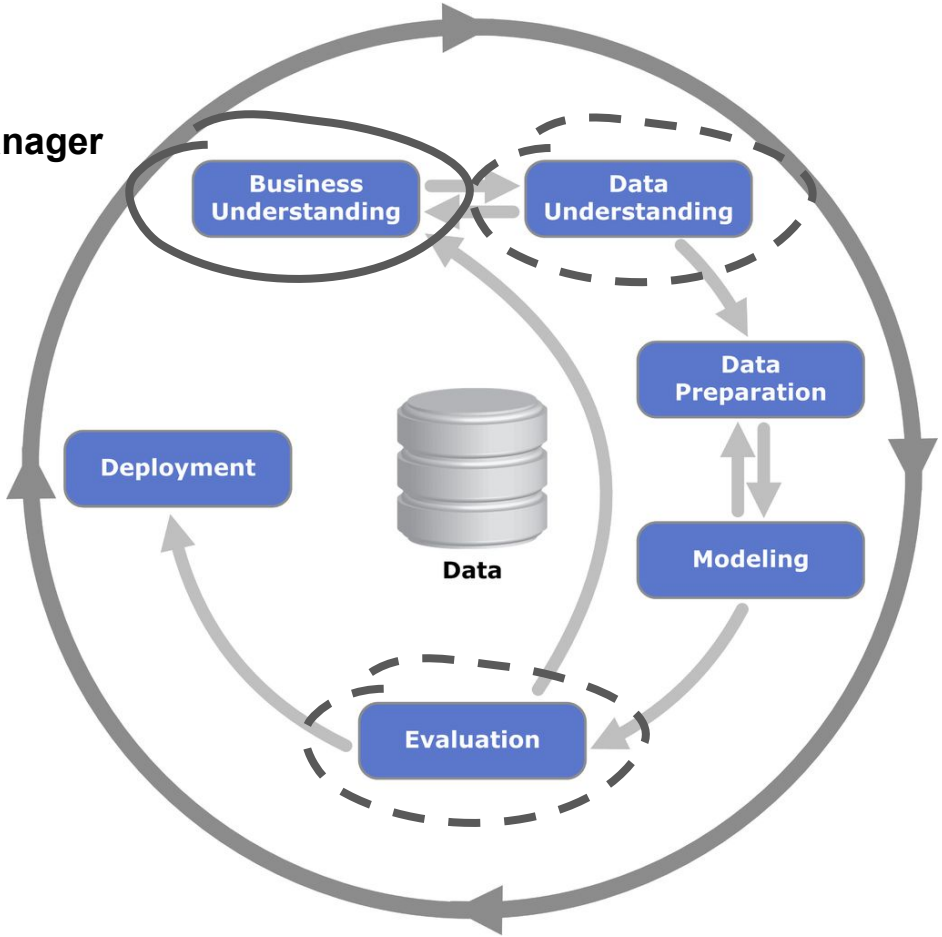
Plan

- Data science process
- **Roles in the team**
- Full-stack data scientist: Jack of all trades
- Becoming full-stack


PM
Data Analyst
Data Engineer
Data Scientist
Software/ML Engineer
DevOps/ SRE

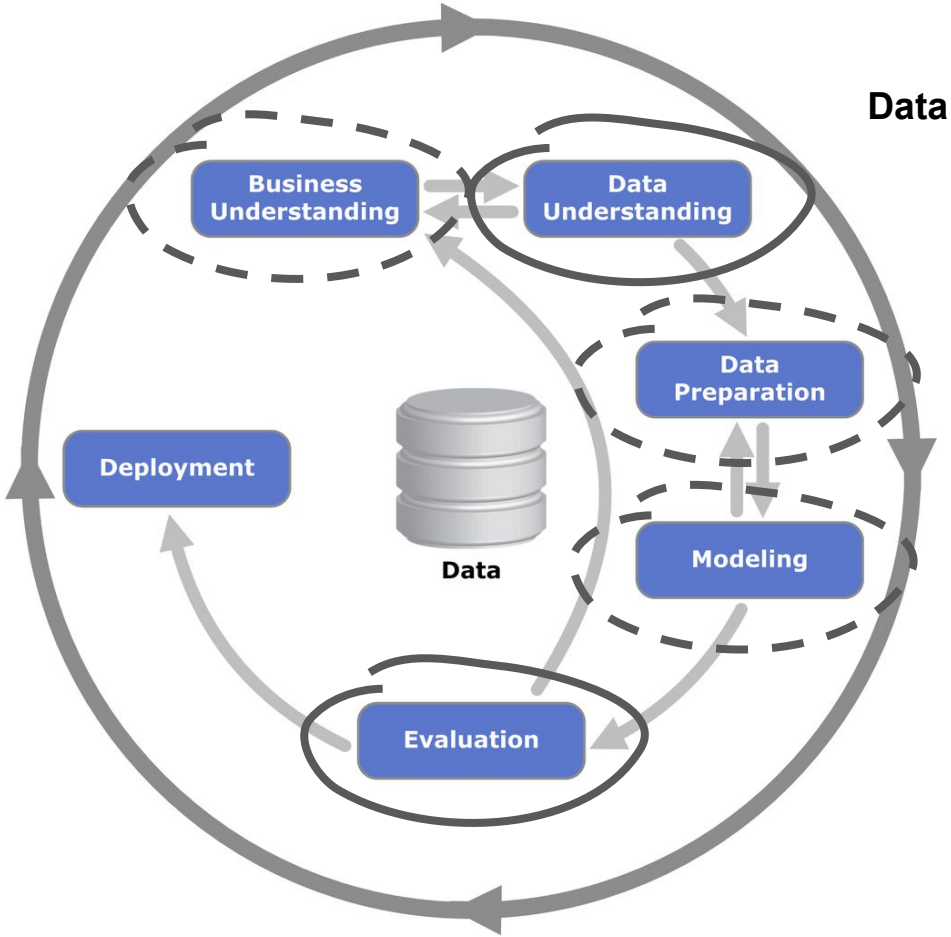

Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Data

Product Manager

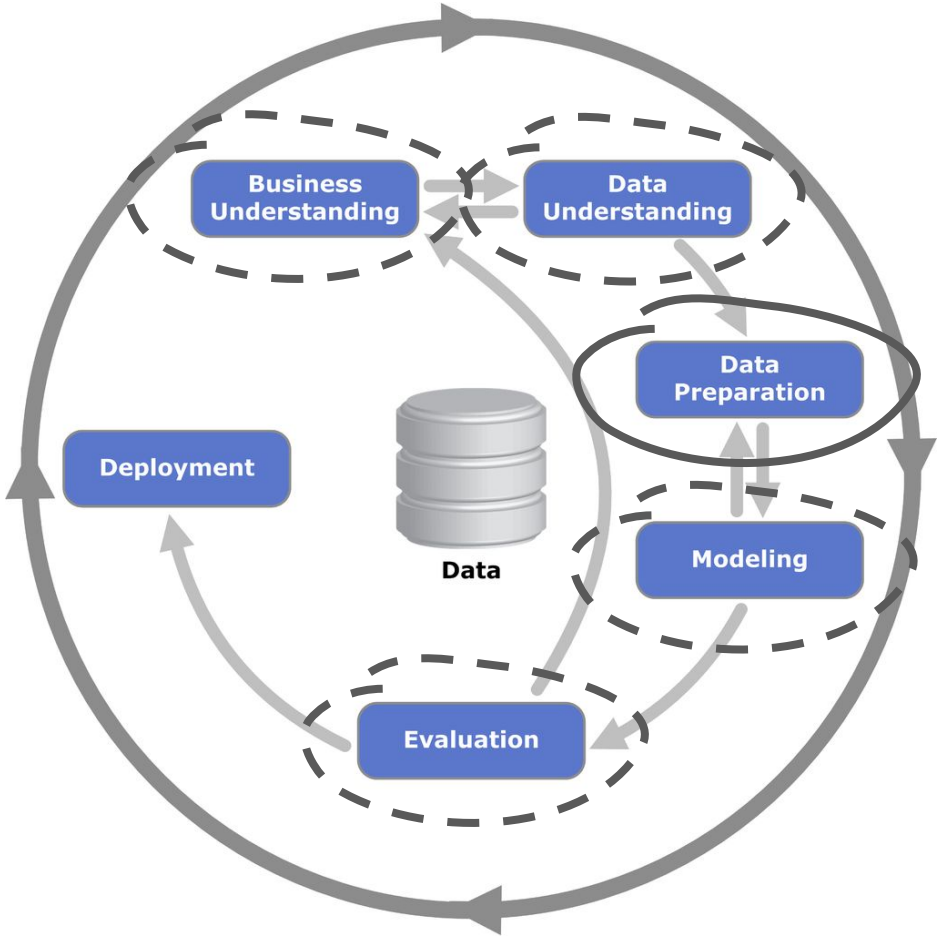


Picture: CRISP-DM

Data Analyst

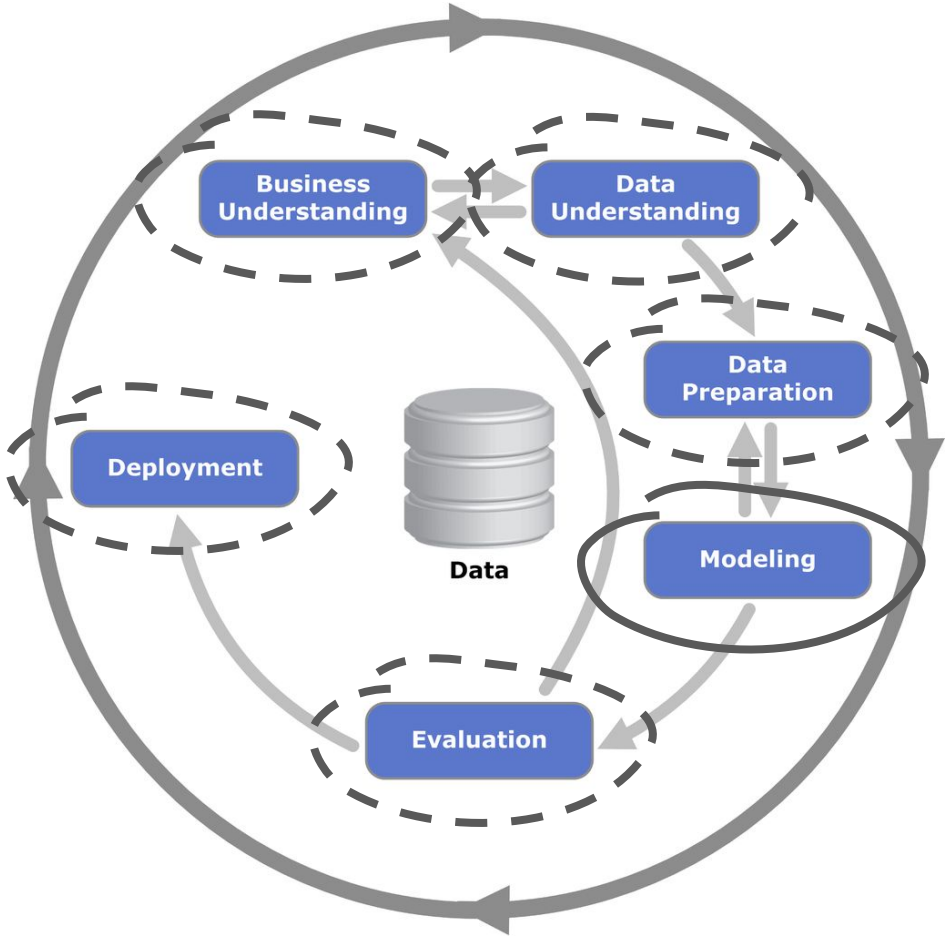


Picture: CRISP-DM



Data Engineer

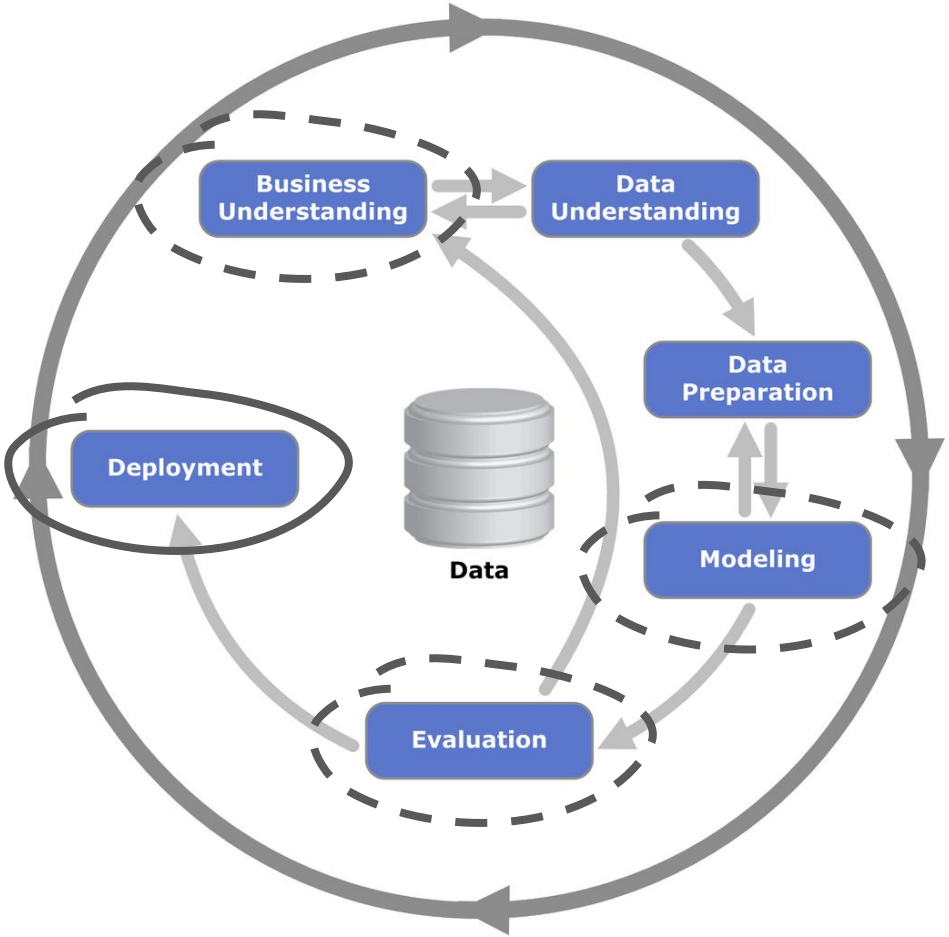
Picture: CRISP-DM



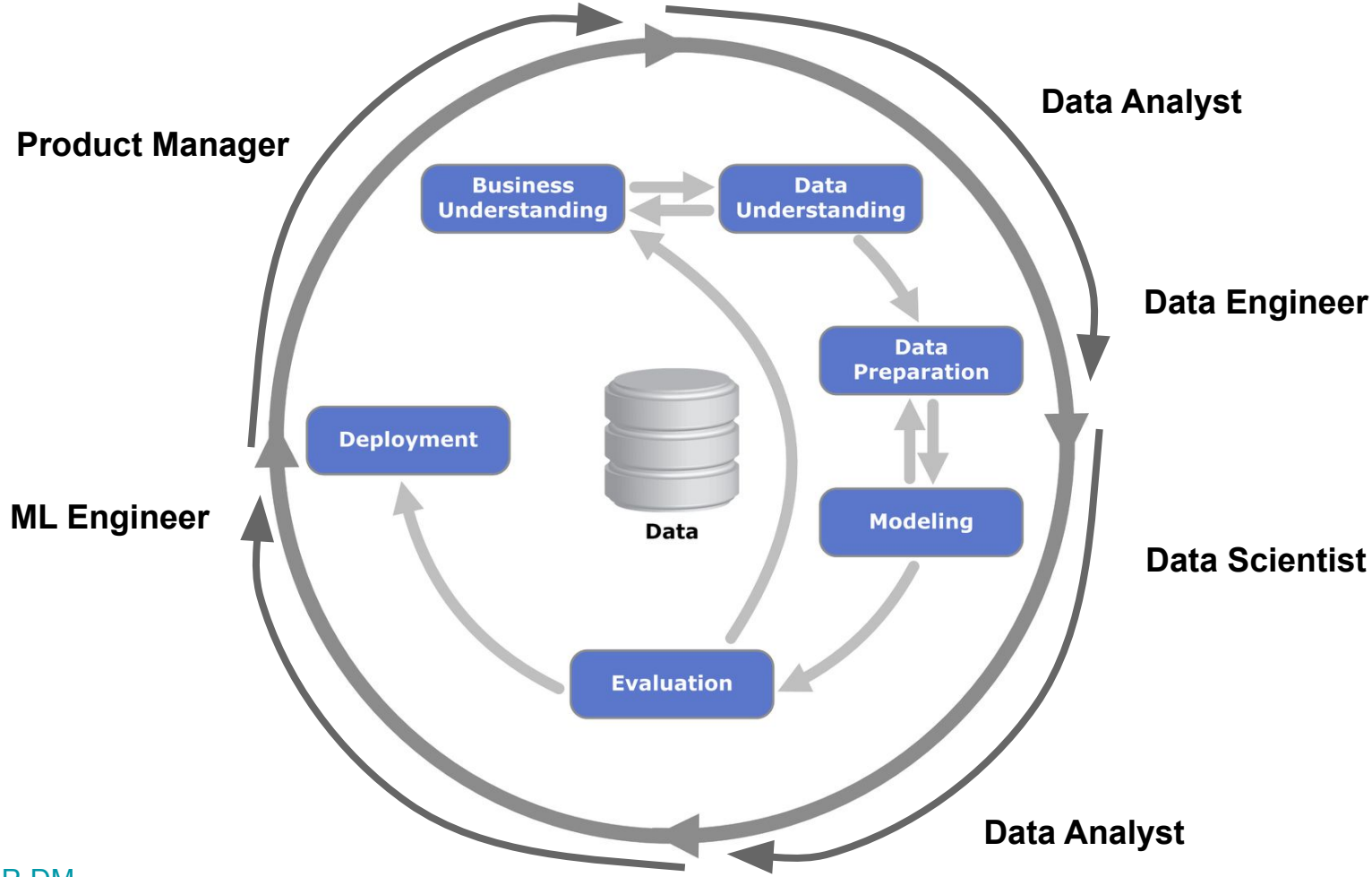
Data Scientist

Picture: CRISP-DM

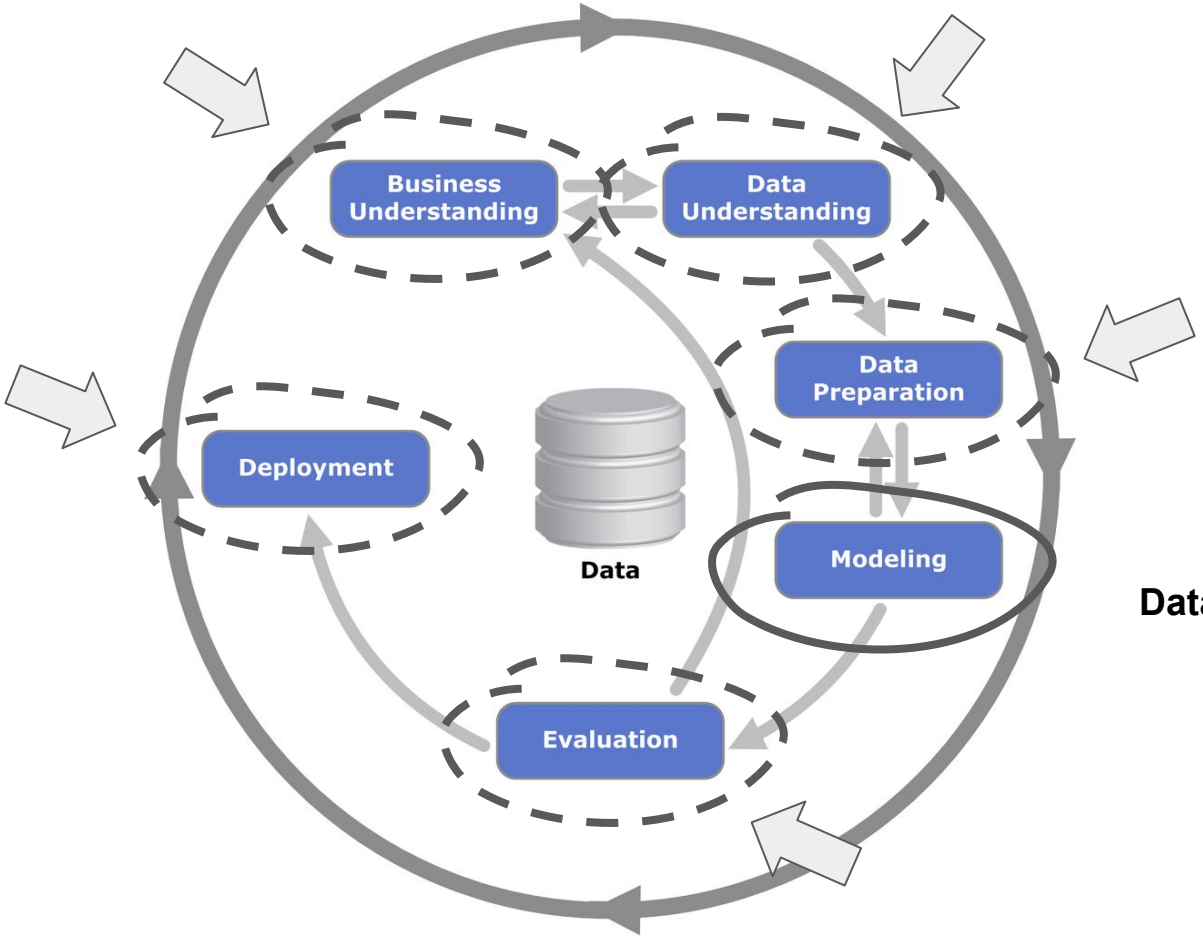
ML Engineer



Picture: CRISP-DM



Picture: CRISP-DM

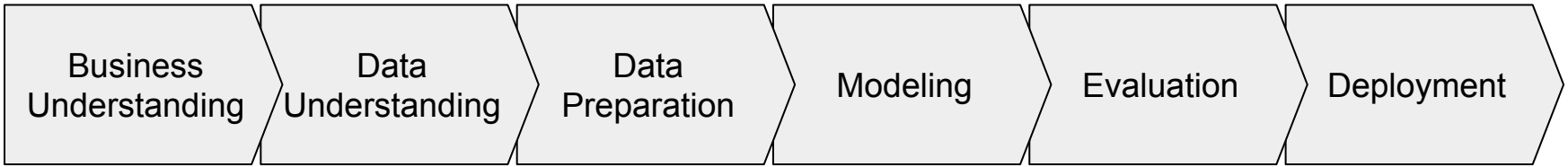


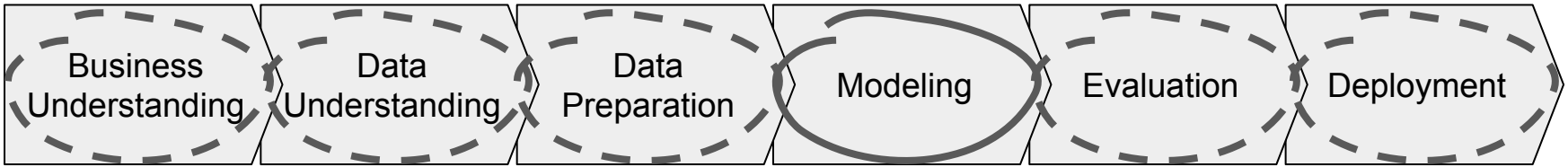
Data Scientist

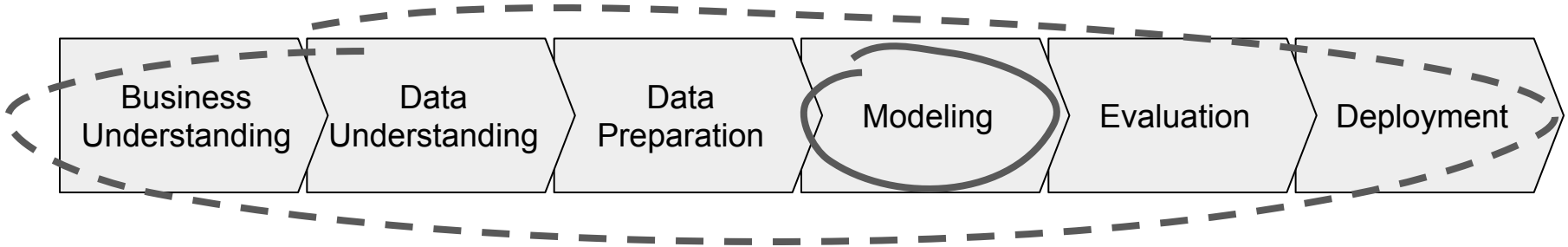
Picture: CRISP-DM

Plan

- Data science process
- Roles in the team
- **Full-stack data scientist: Jack of all trades**
- Becoming full-stack


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment

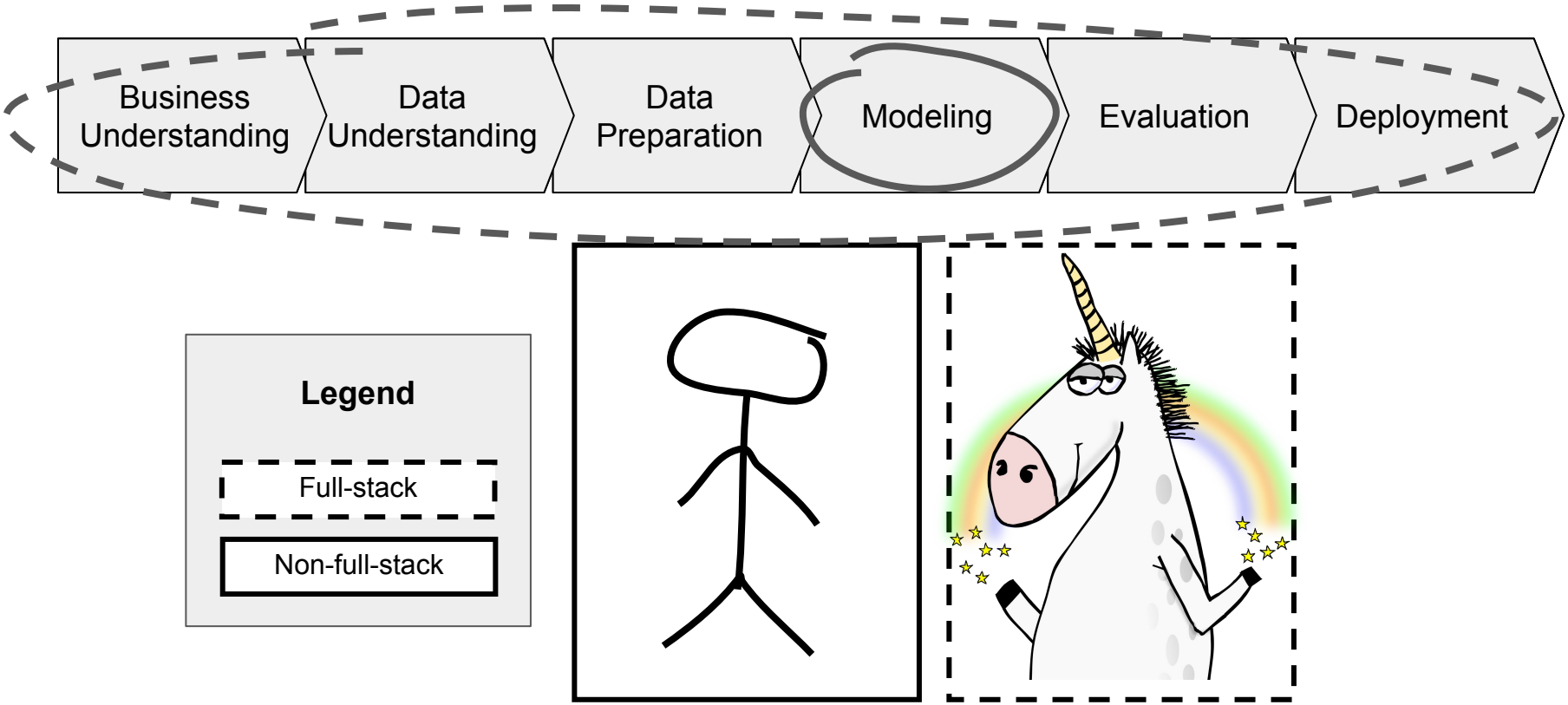

Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment

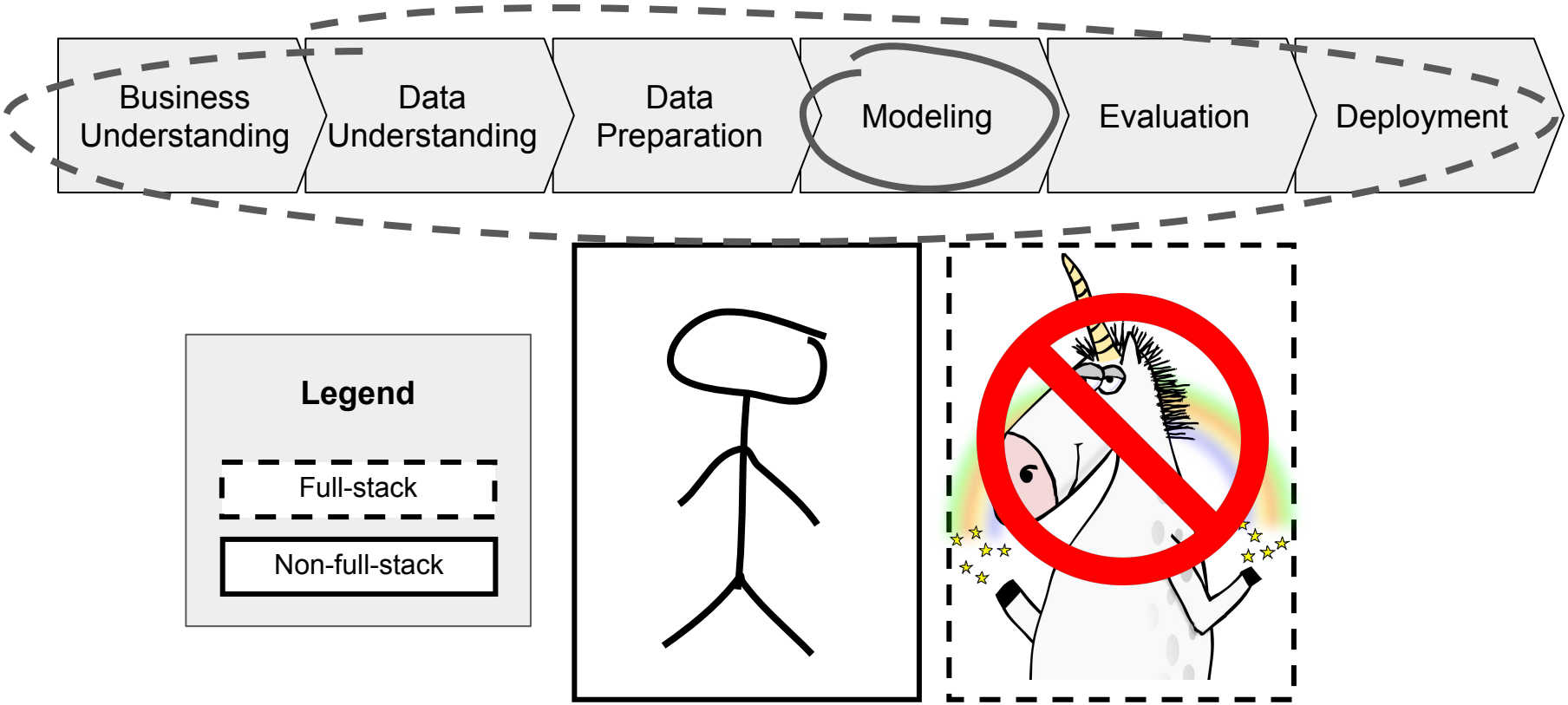


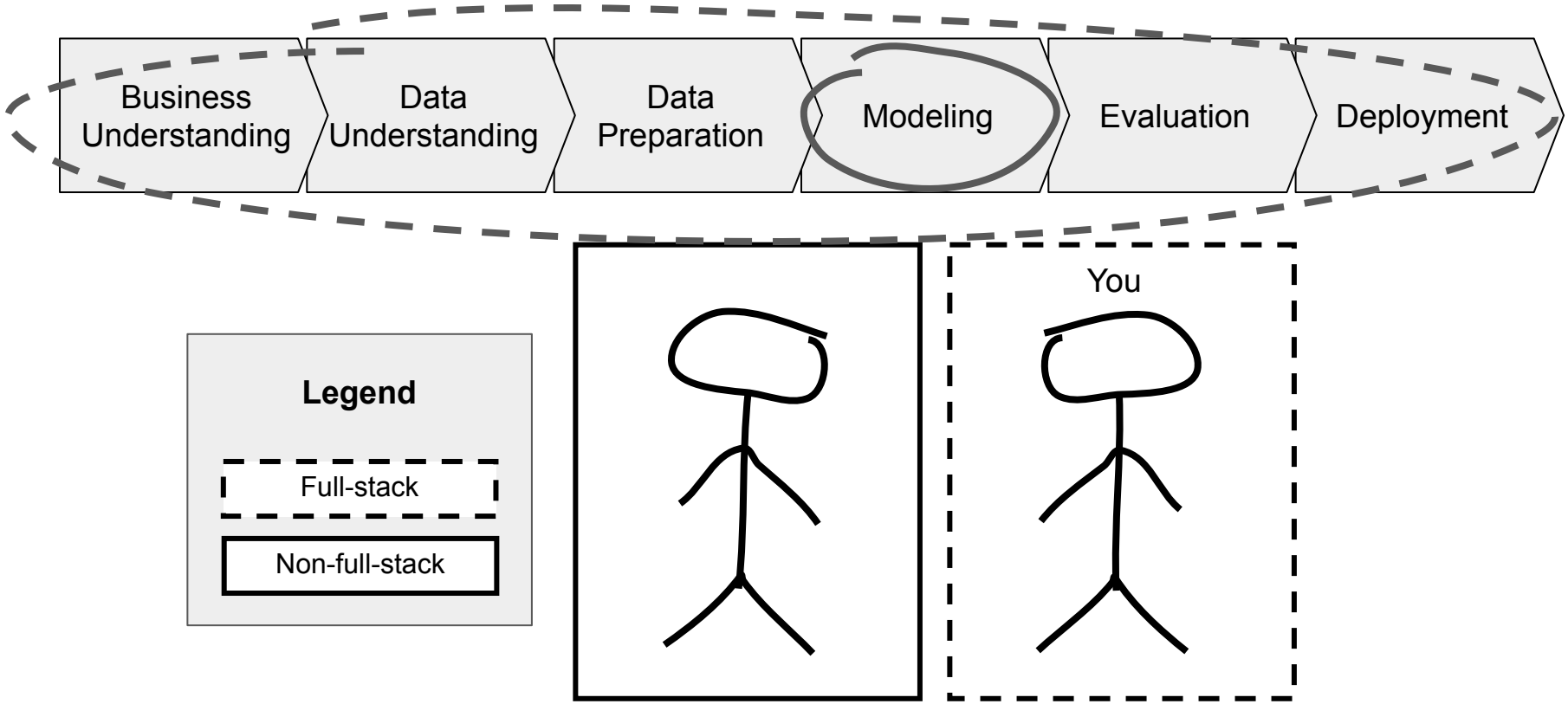
Legend

Full-stack

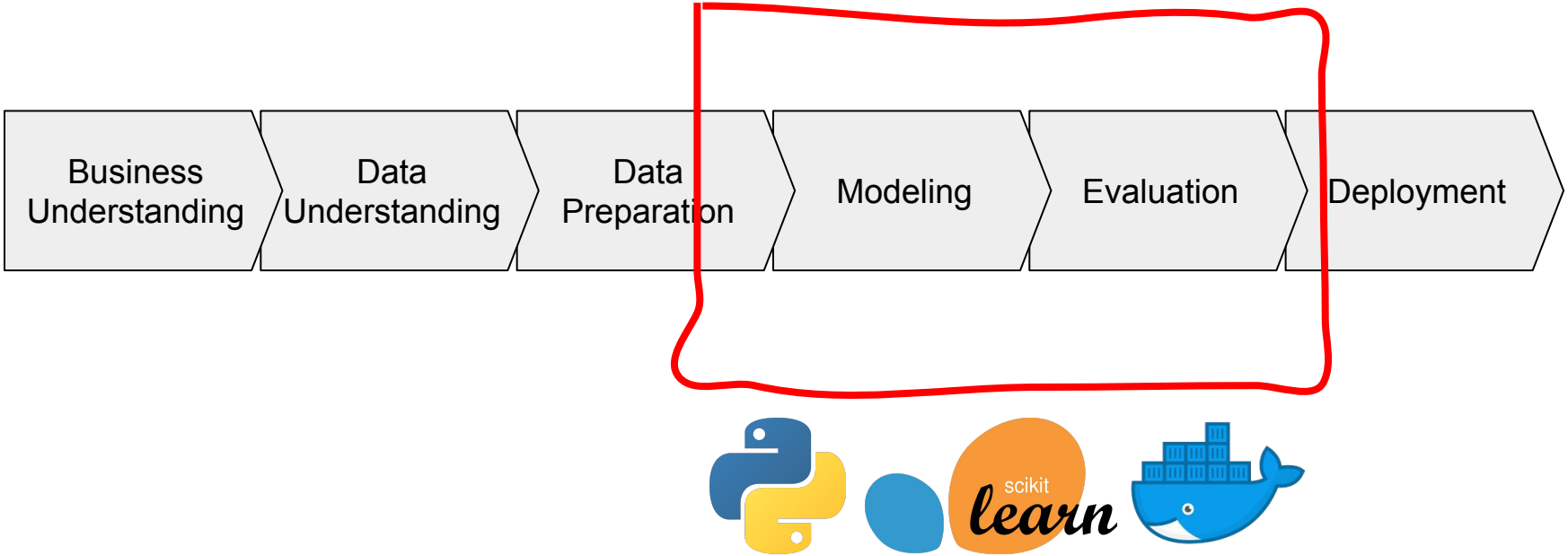
Non-full-stack


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Legend
Full-stack
Non-full-stack

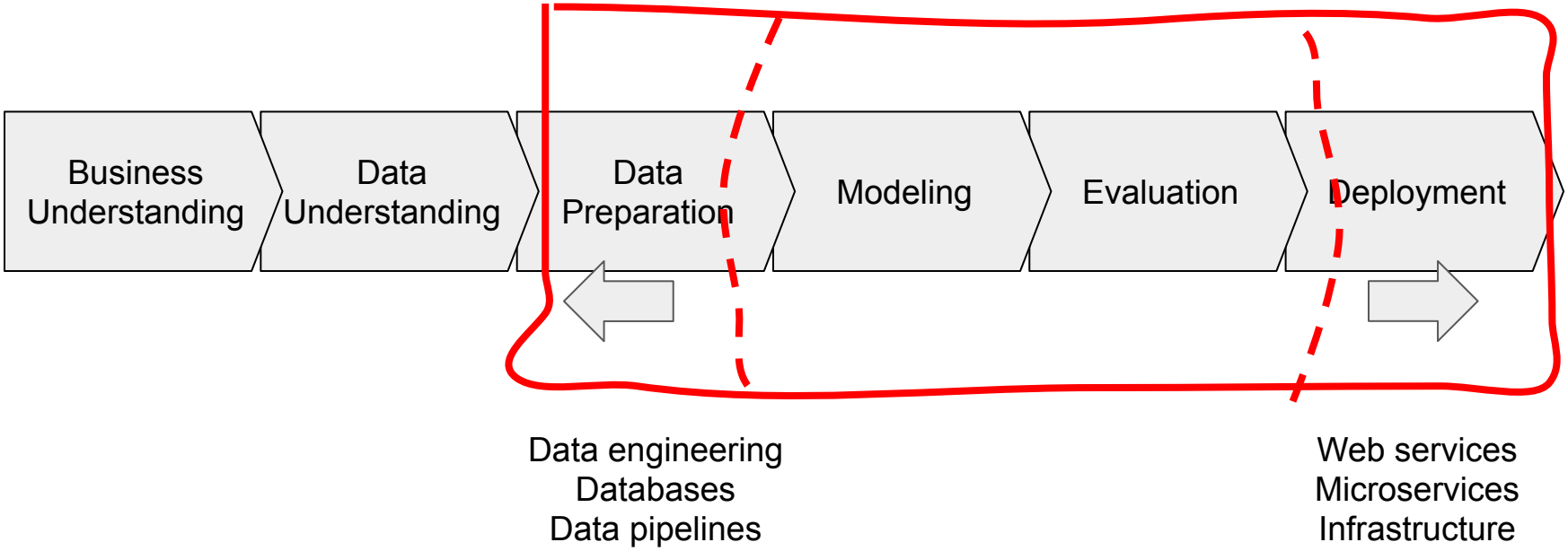

Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Legend
Full-stack
Non-full-stack


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Legend
Full-stack
Non-full-stack
You

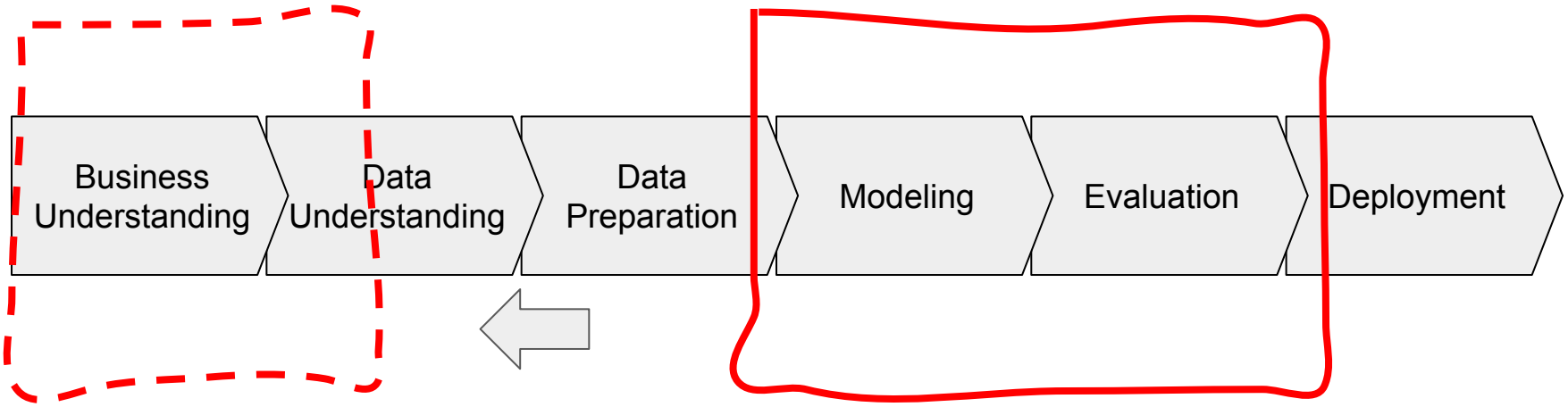
Core data science



Software engineering



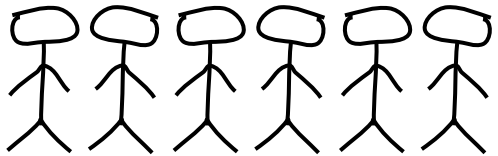
Product management



Product management
Domain expertise
Data analysis
Communication

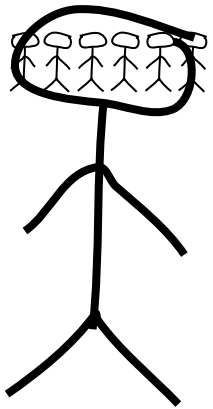
We need to learn

- Product management
- Data analysis
- Data engineering
- Backend engineering
- DevOps
- …



We need to learn

- Product management
- Data analysis
- Data engineering
- Backend engineering
- DevOps
- …

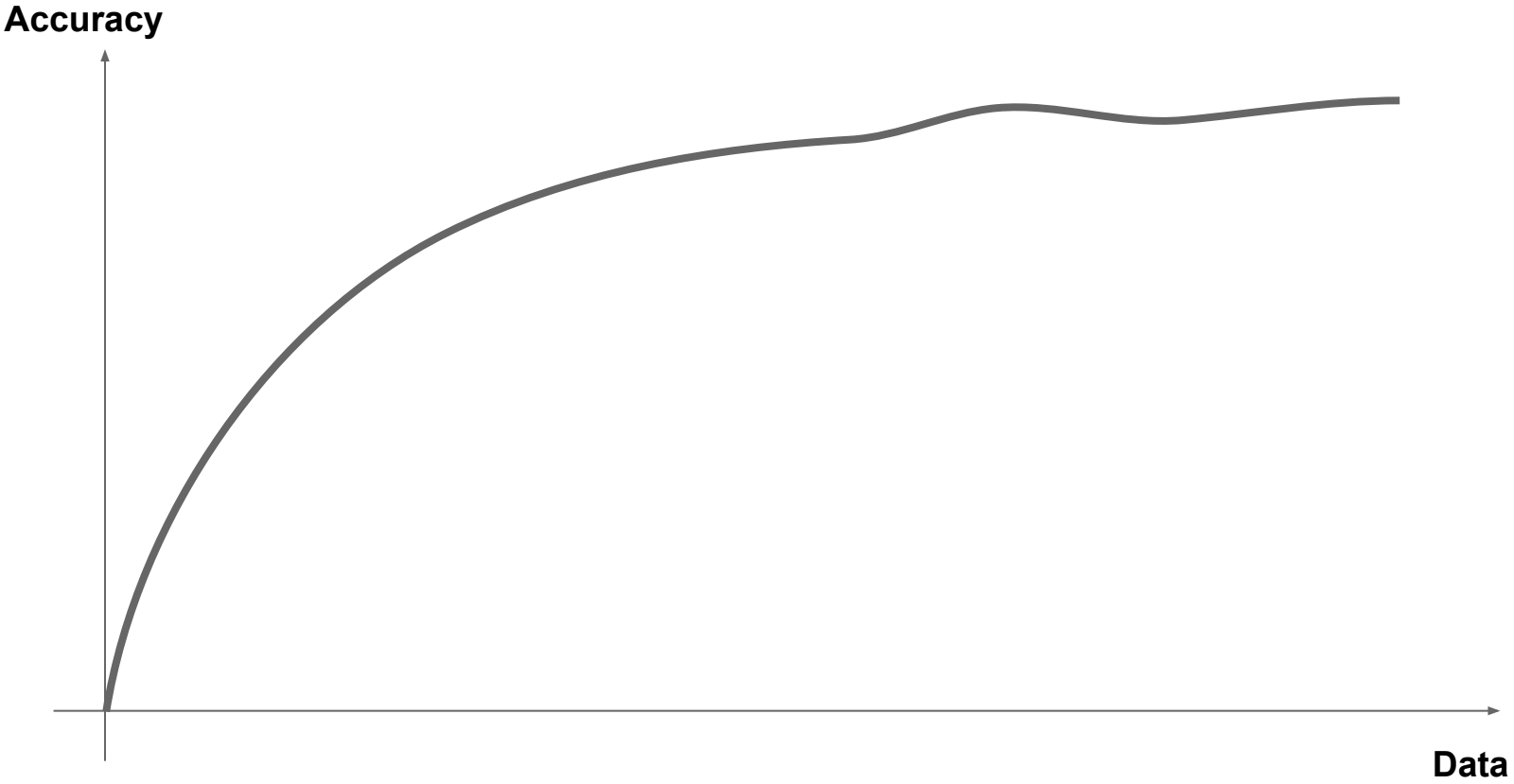


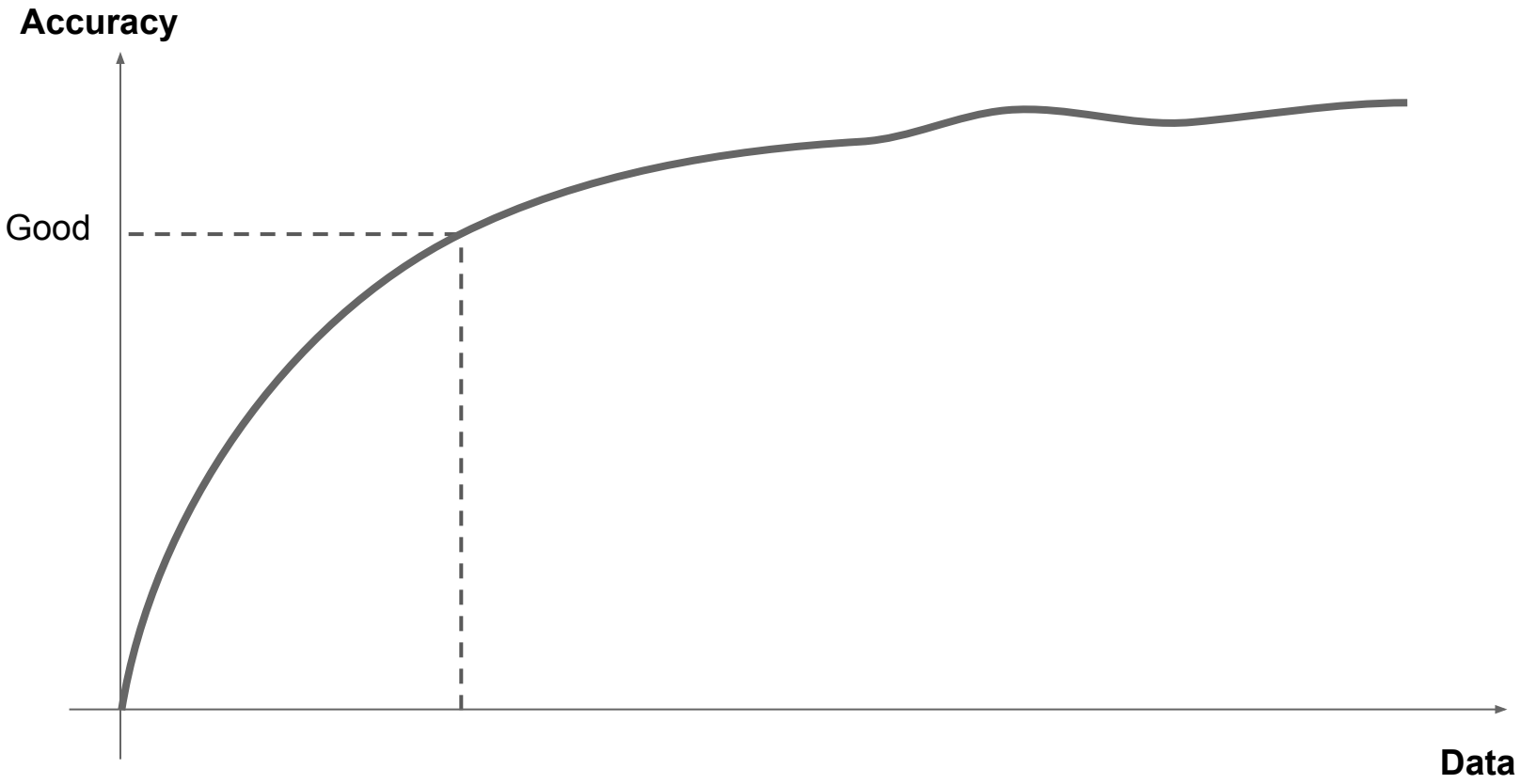
We need to learn

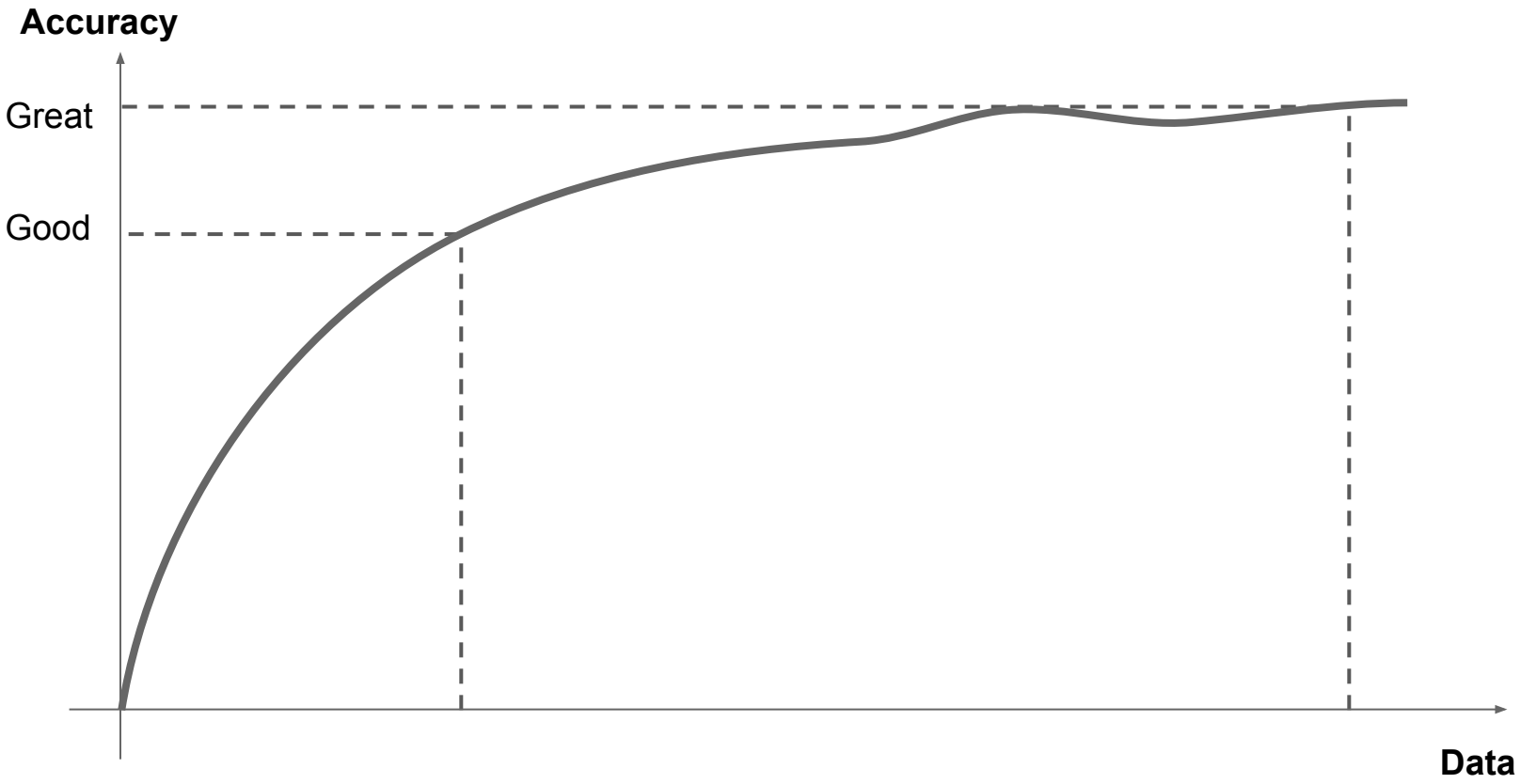
- Product management
- Data analysis
- Data engineering
- Backend engineering
- DevOps
- …

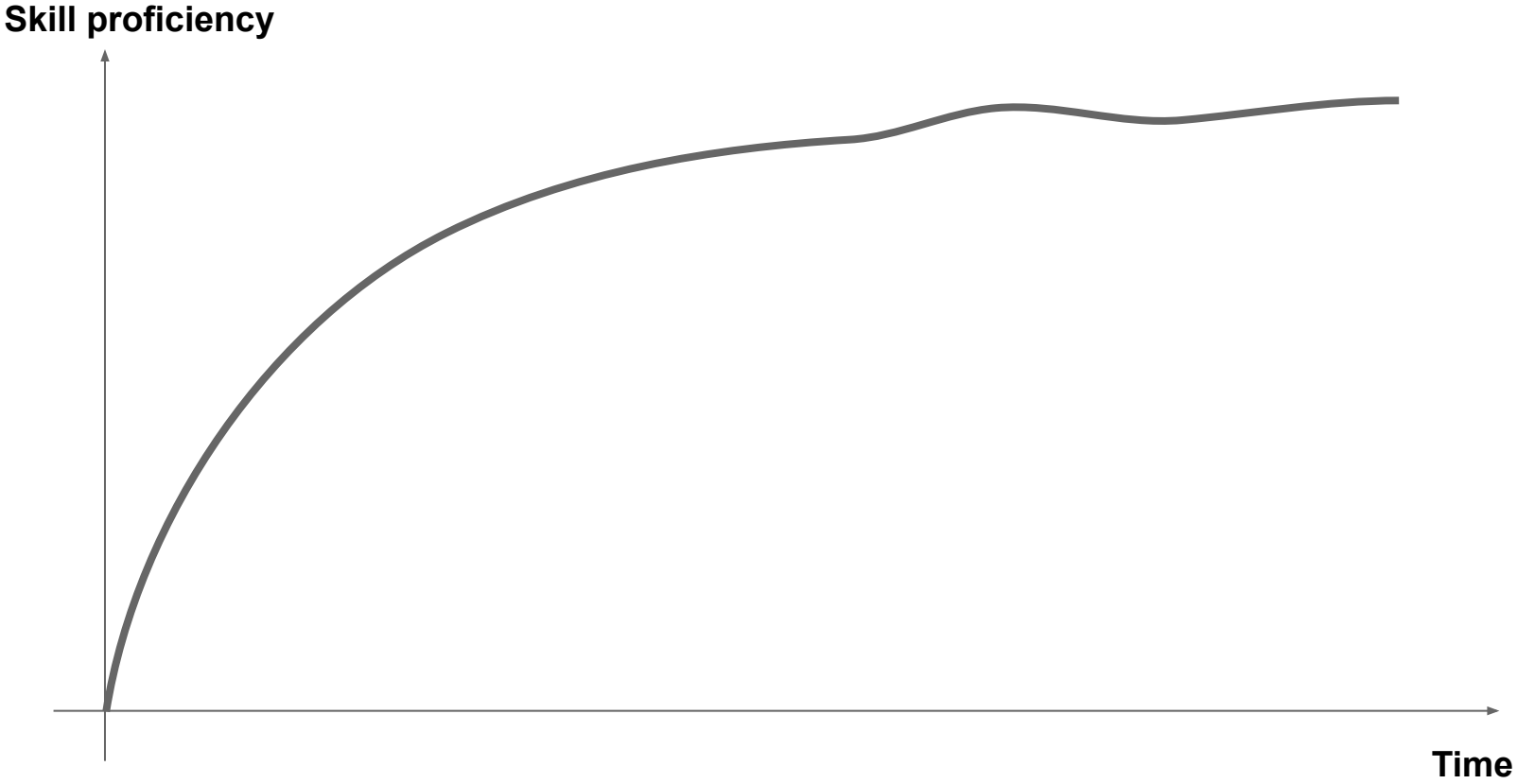
Plan

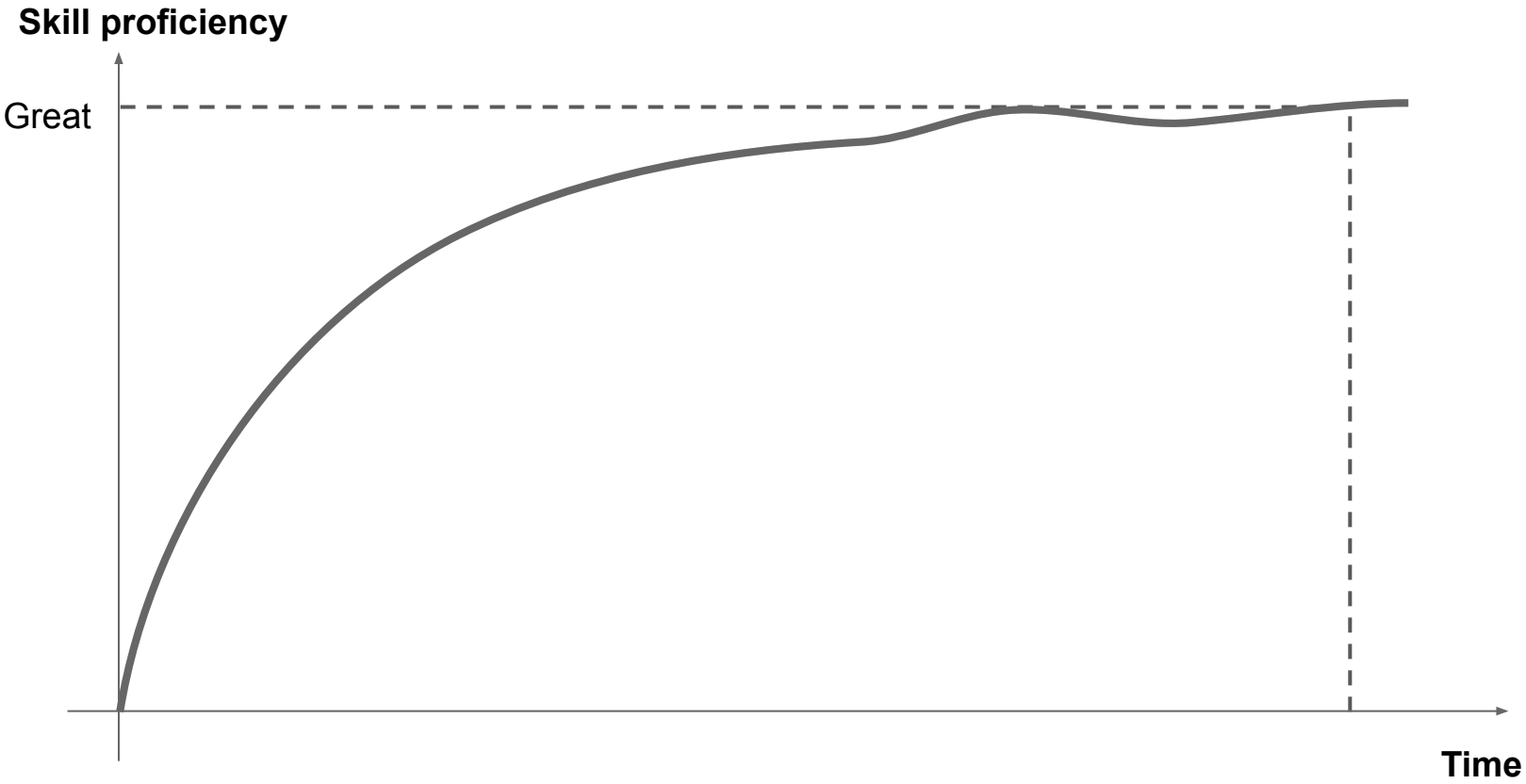
- Data science process
- Roles in the team
- Full-stack data scientist: Jack of all trades
- **Becoming full-stack**

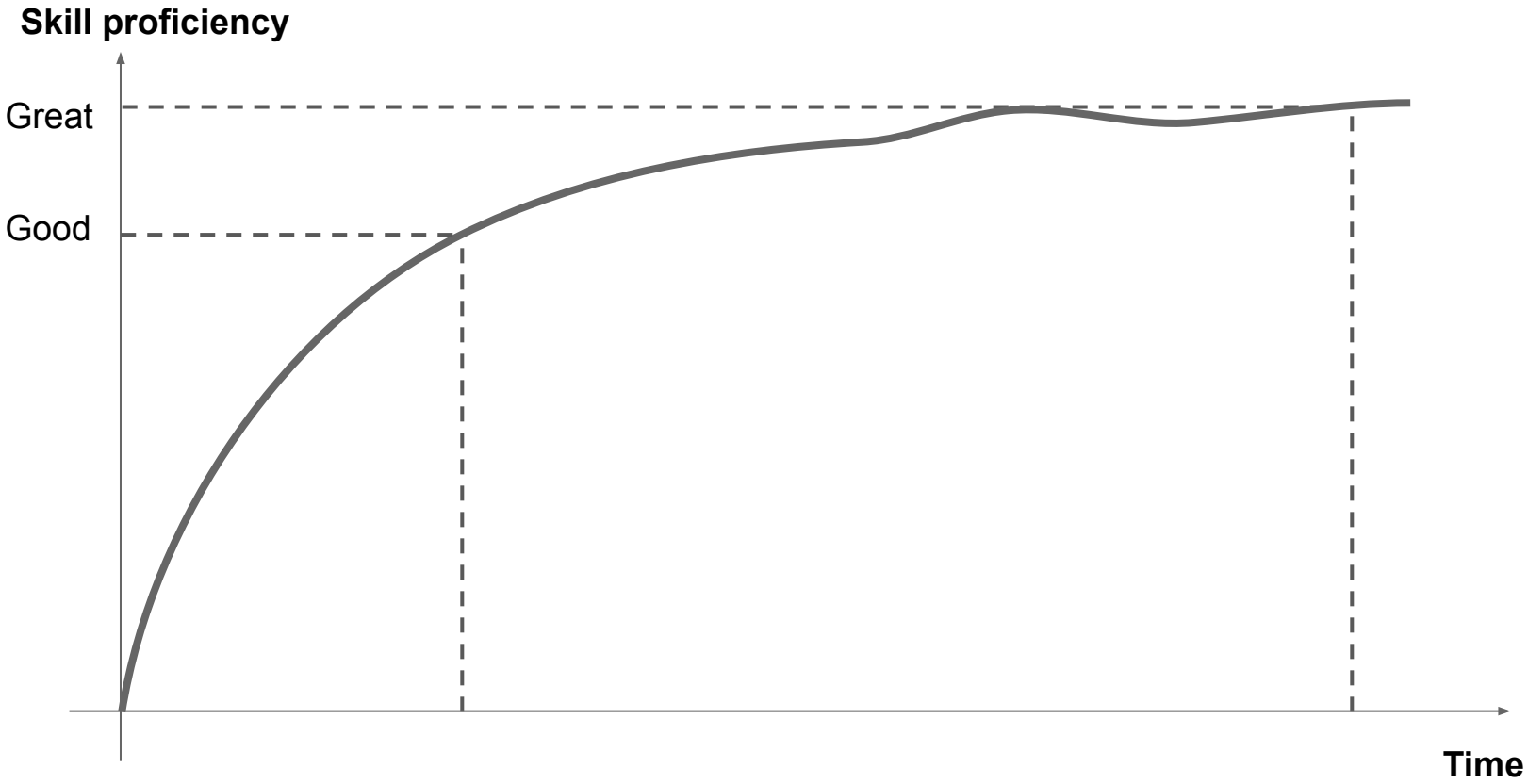

Accuracy
Data

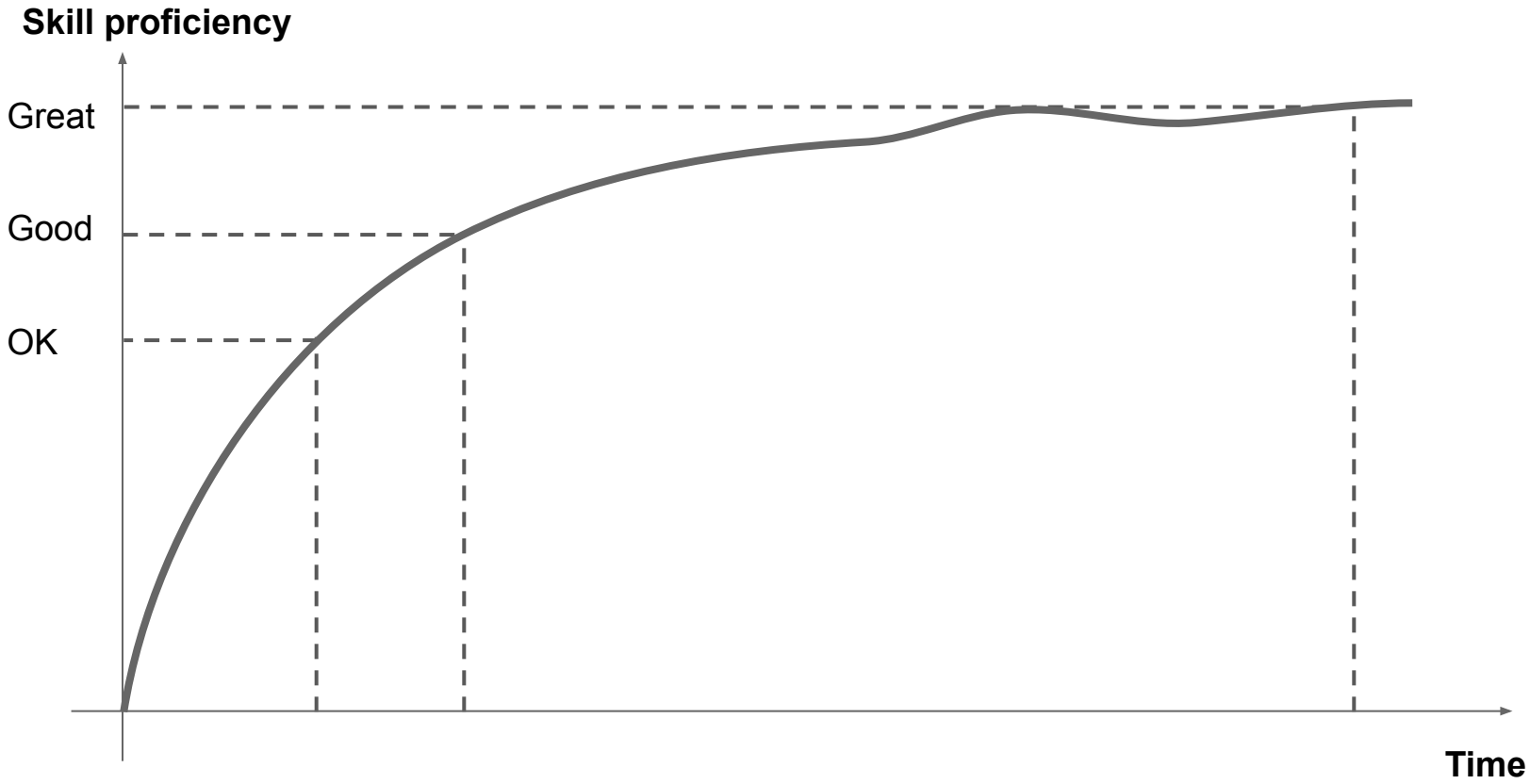

Accuracy
Good
Data


Accuracy
Great
Good
Data


Skill proficiency
Time


Skill proficiency
Great
Time

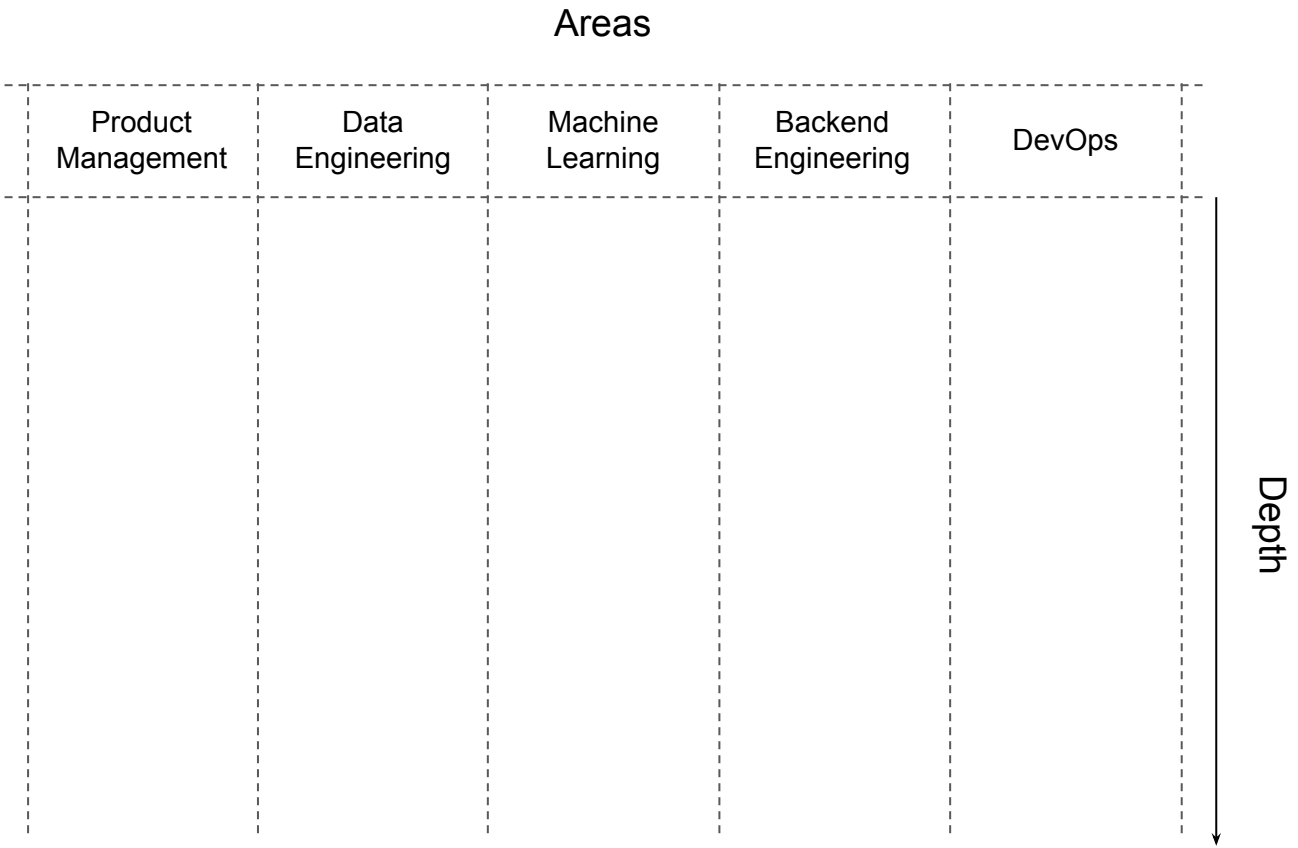

Skill proficiency
Great
Good
Time

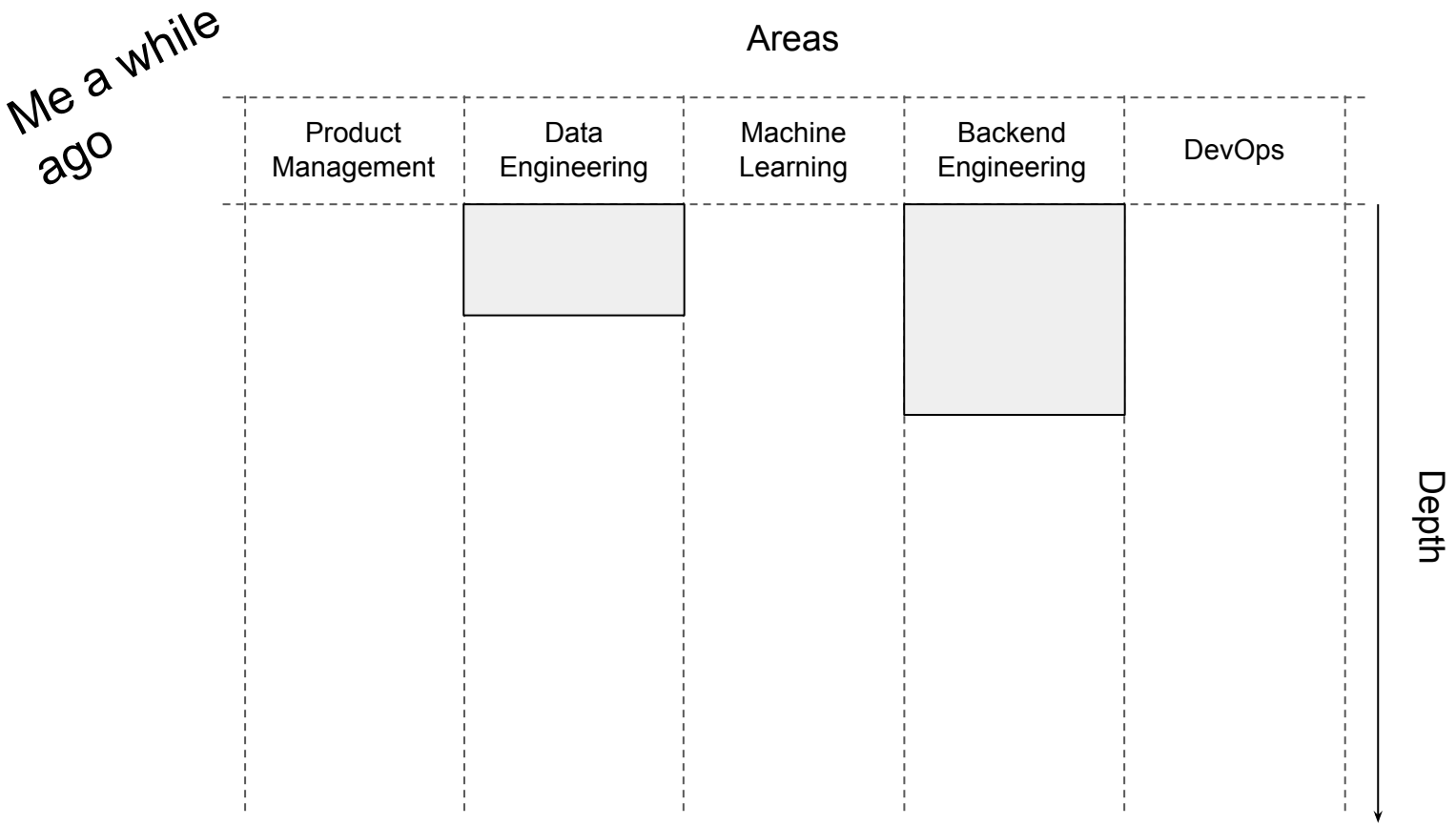

Skill proficiency
Great
Good
OK
Time

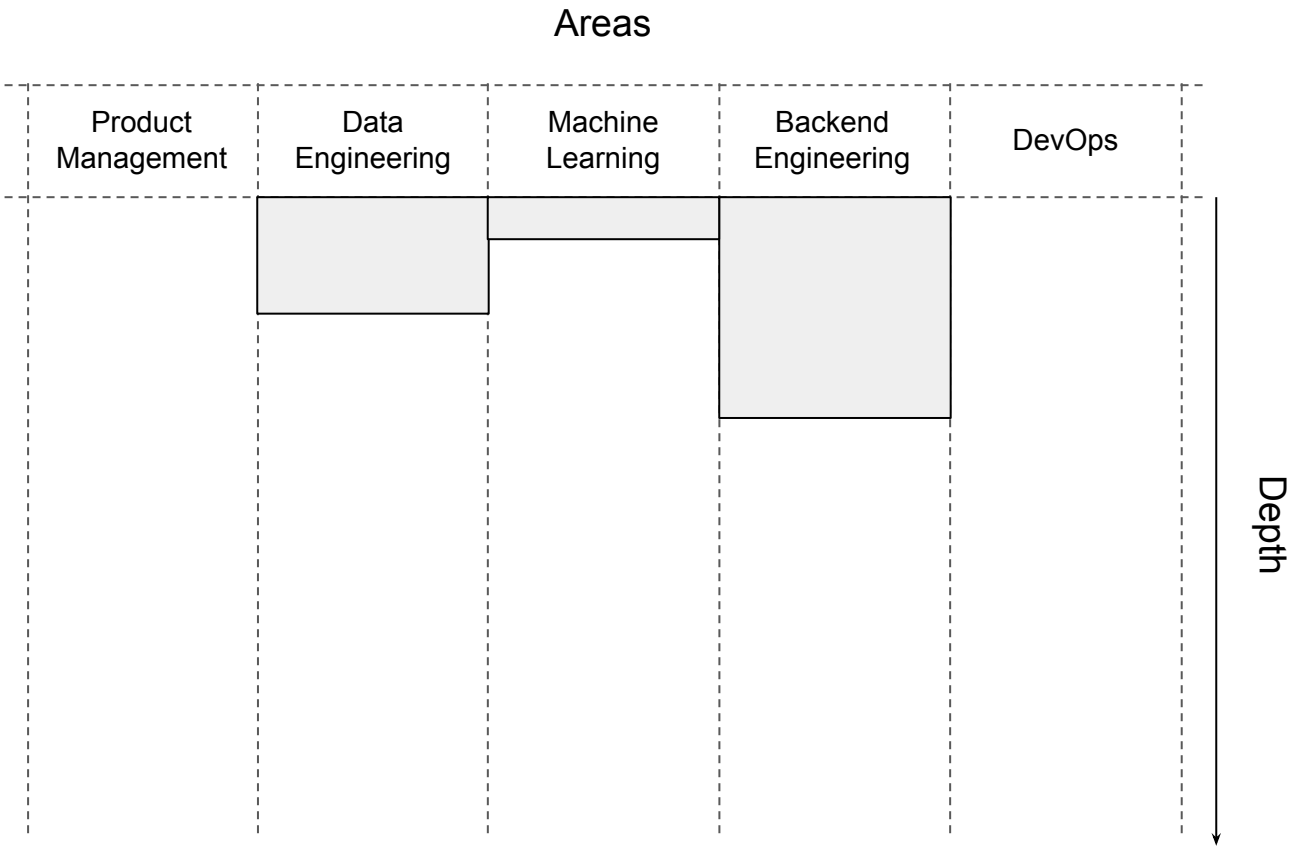
Skill proficiency

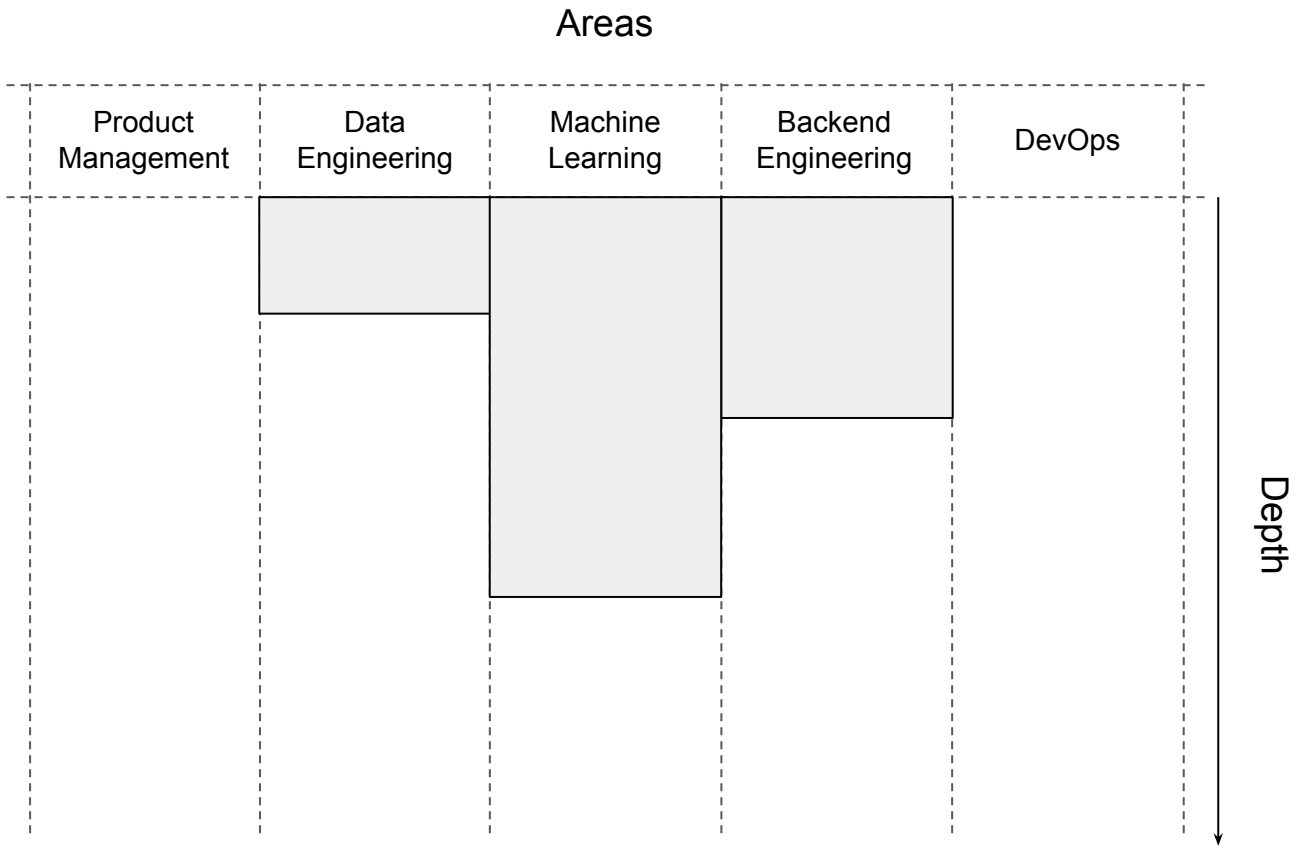
- OK — can use the skill, mostly independently
- Good — can use the skill independently
- Great — expert

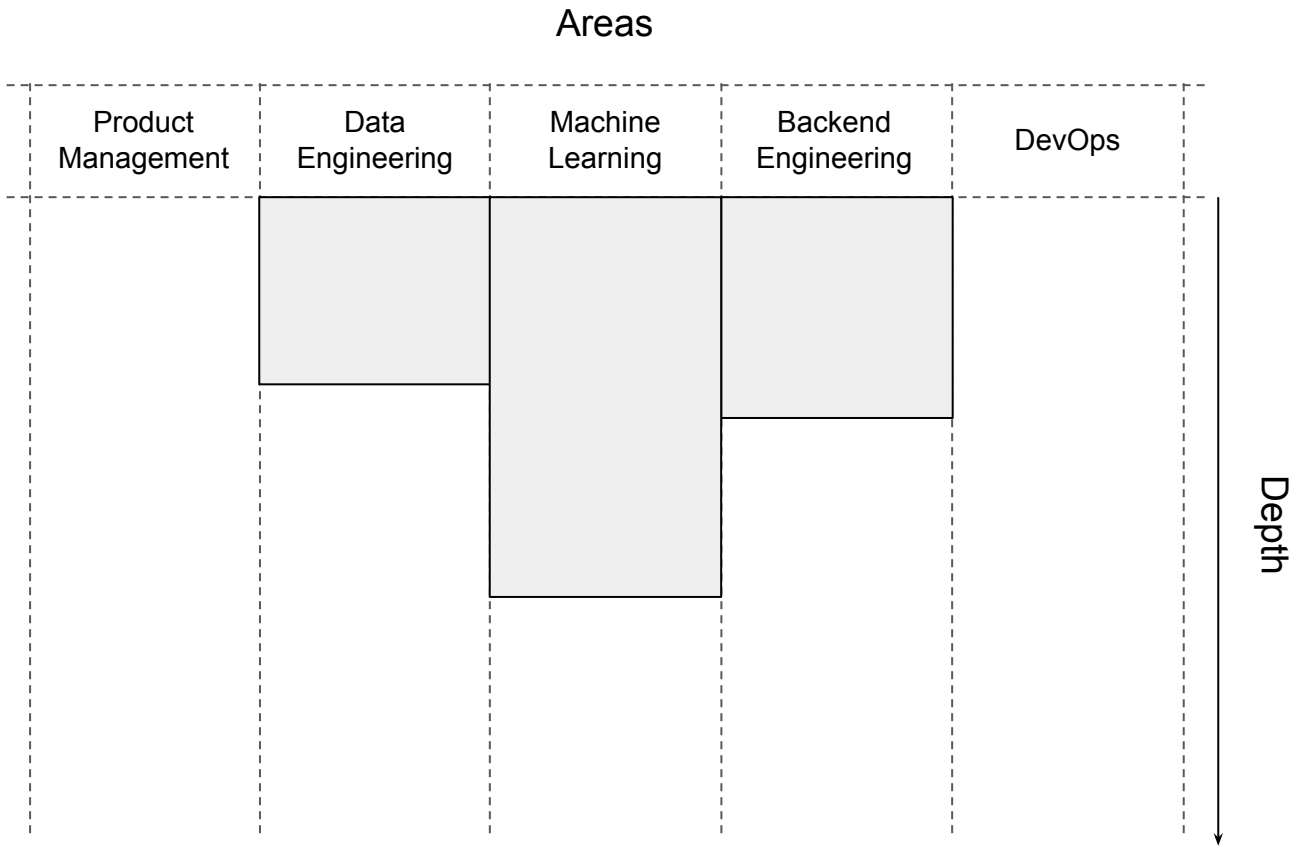
We don’t need to be
experts in everything!

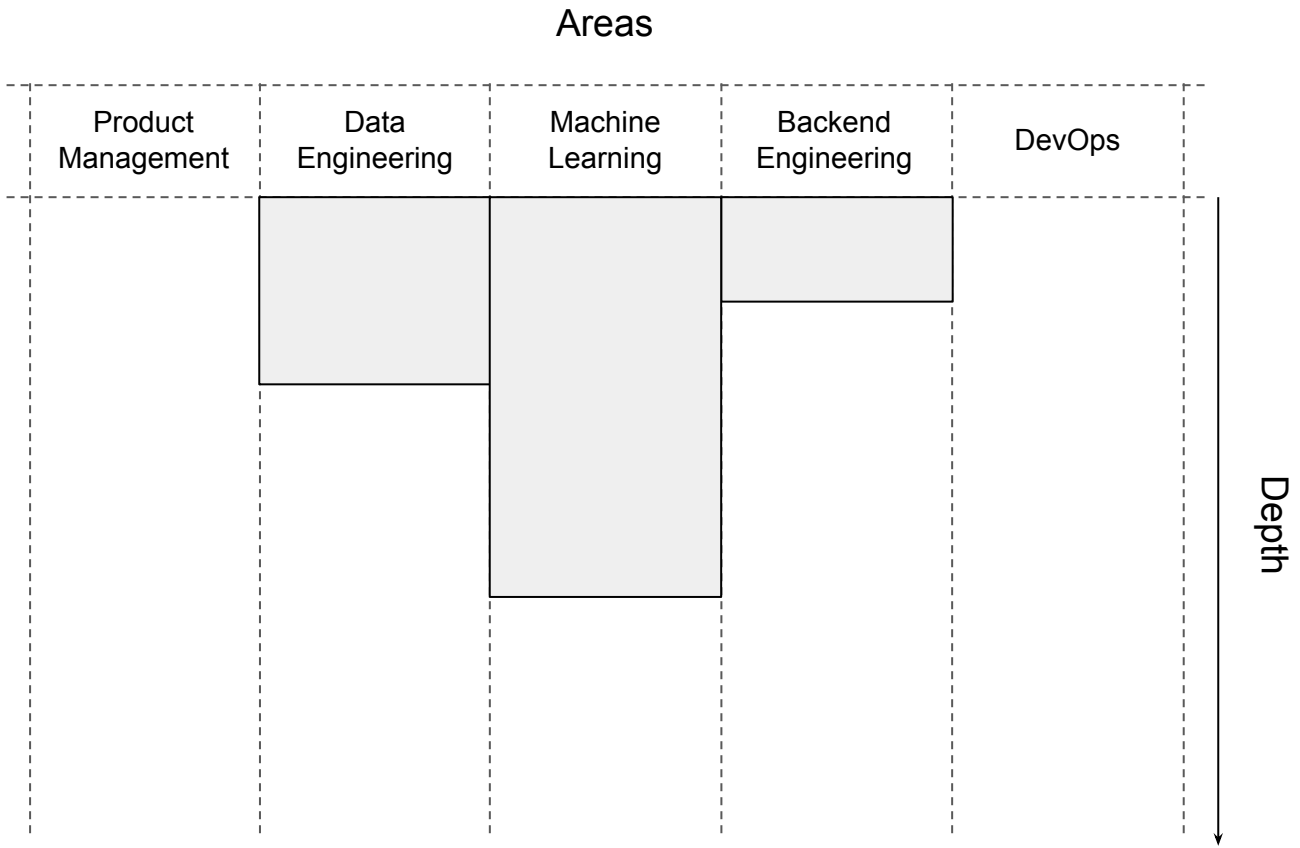

Areas
Product Management
Data Engineering
Machine Learning
Backend Engineering
DevOps
Depth

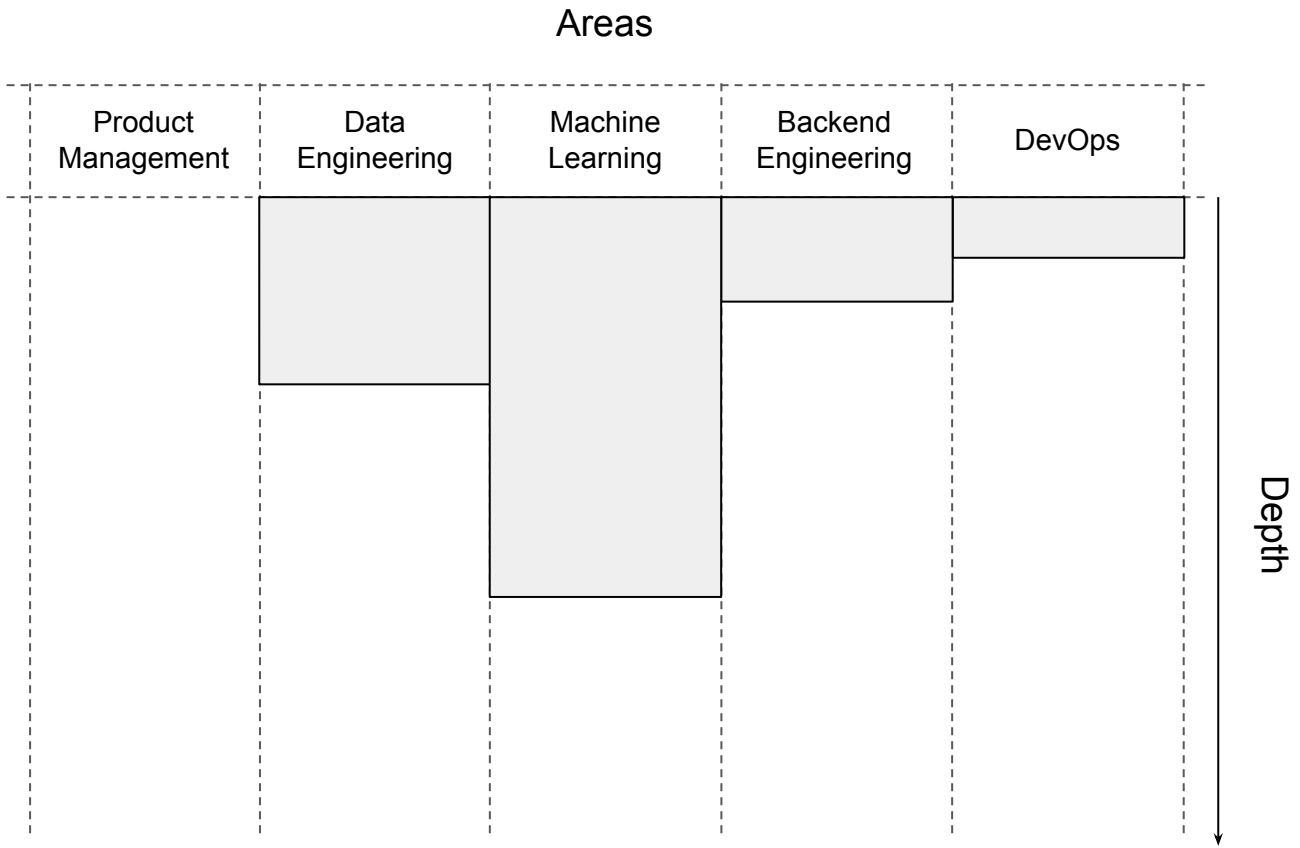

Me a while ago
Areas
Product Management
Data Engineering
Machine Learning
Backend Engineering
DevOps
Depth

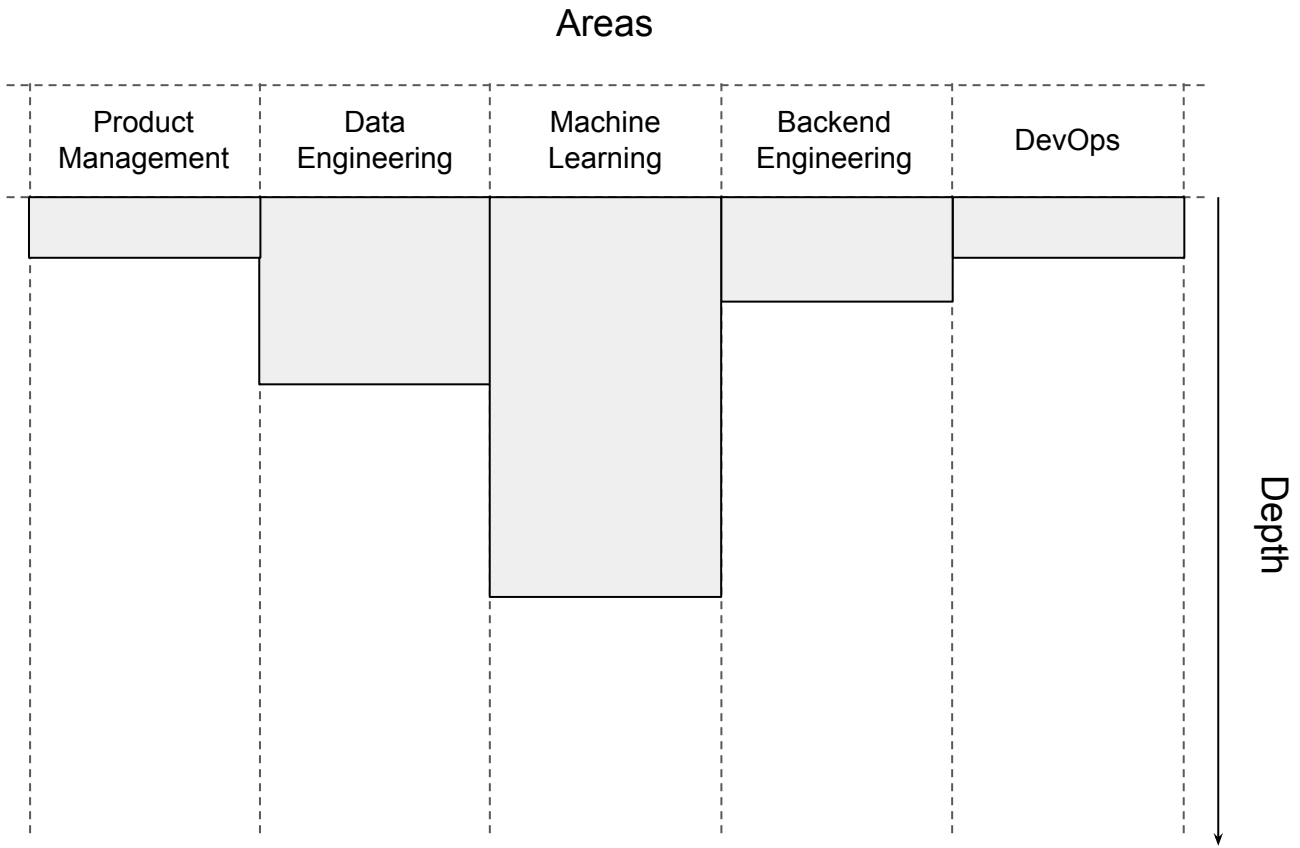

Areas
Product Management
Data Engineering
Machine Learning
Backend Engineering
DevOps
Depth


Areas
Product Management
Data Engineering
Machine Learning
Backend Engineering
DevOps
Depth

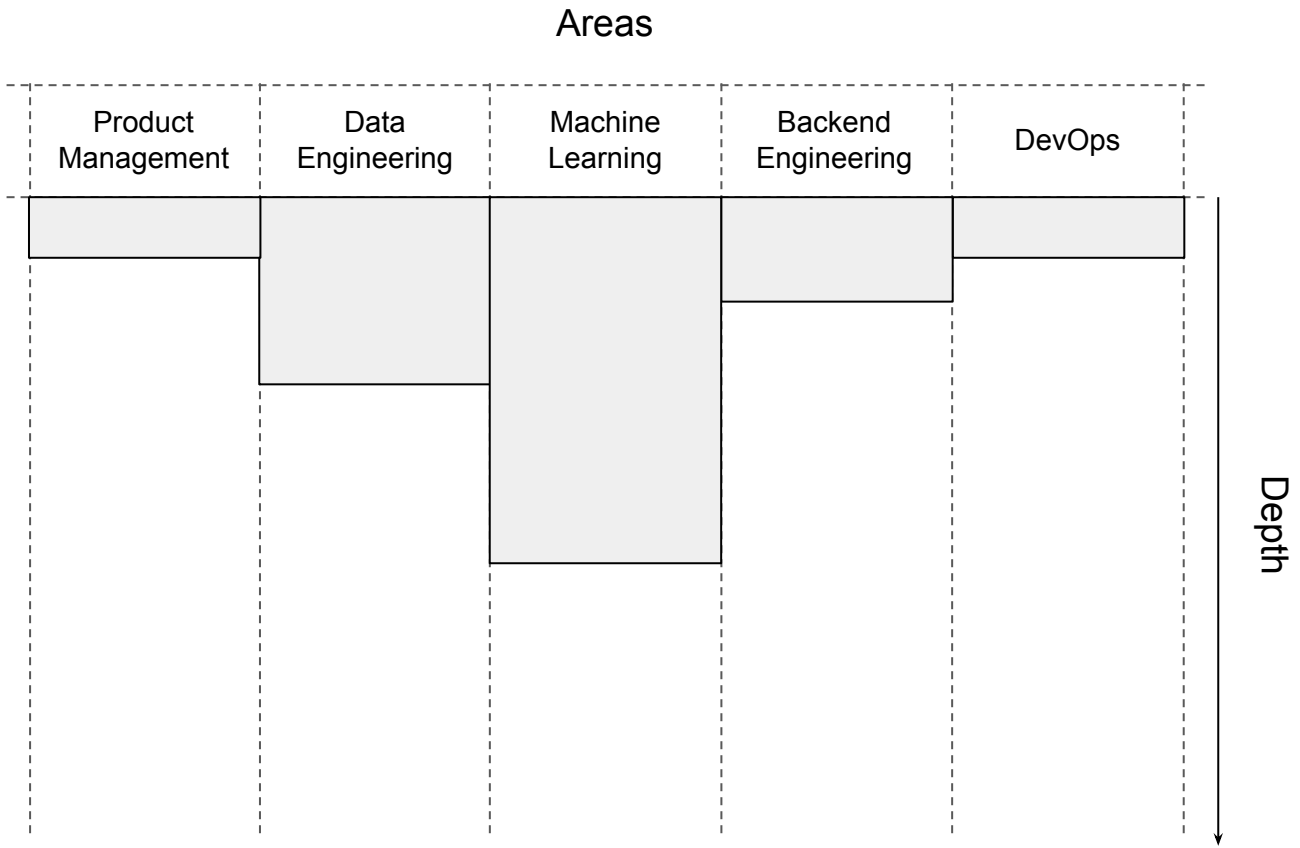

Areas
Product Management
Data Engineering
Machine Learning
Backend Engineering
DevOps
Depth


Areas
Product Management
Data Engineering
Machine Learning
Backend Engineering
DevOps
Depth

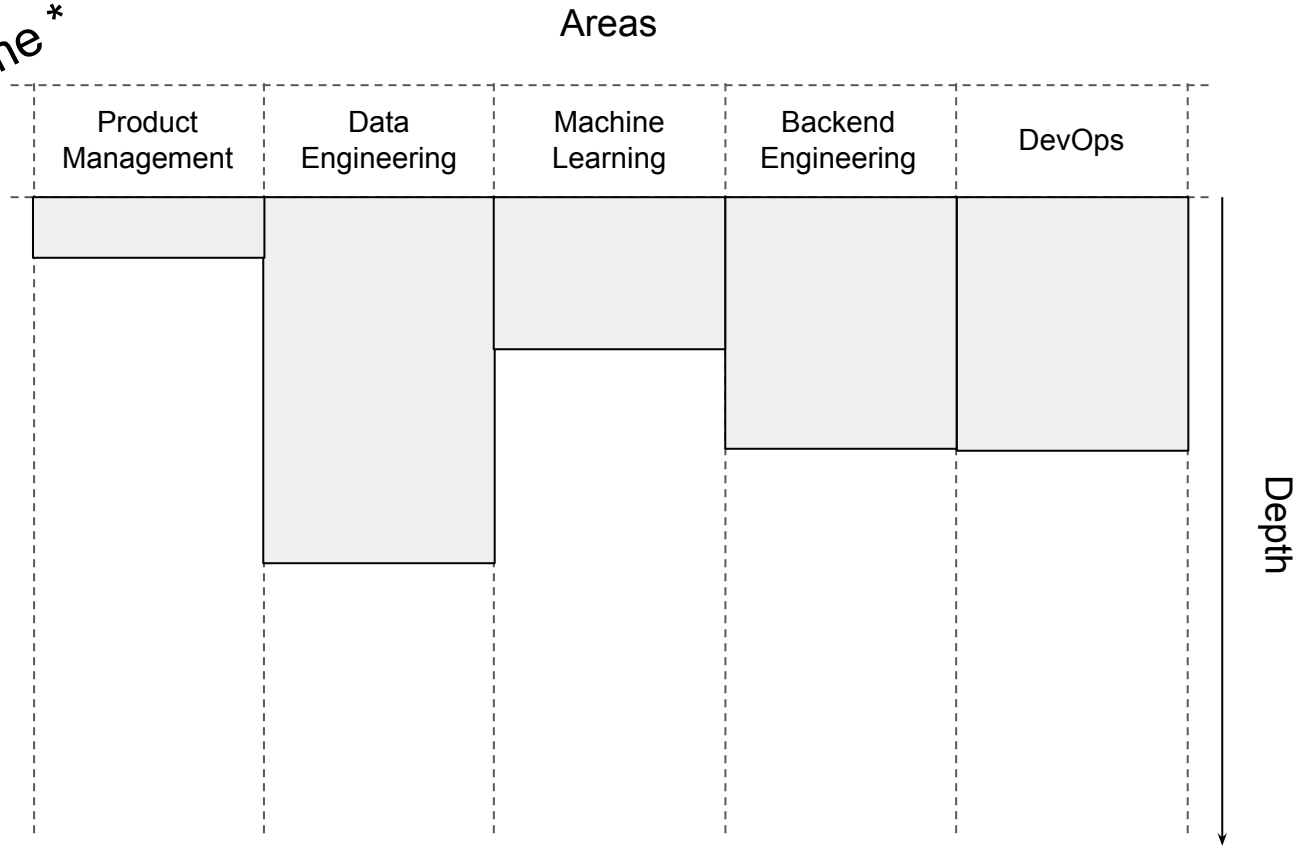

Areas
Product Management
Data Engineering
Machine Learning
Backend Engineering
DevOps
Depth


Areas
Product Management
Data Engineering
Machine Learning
Backend Engineering
DevOps
Depth

Me now

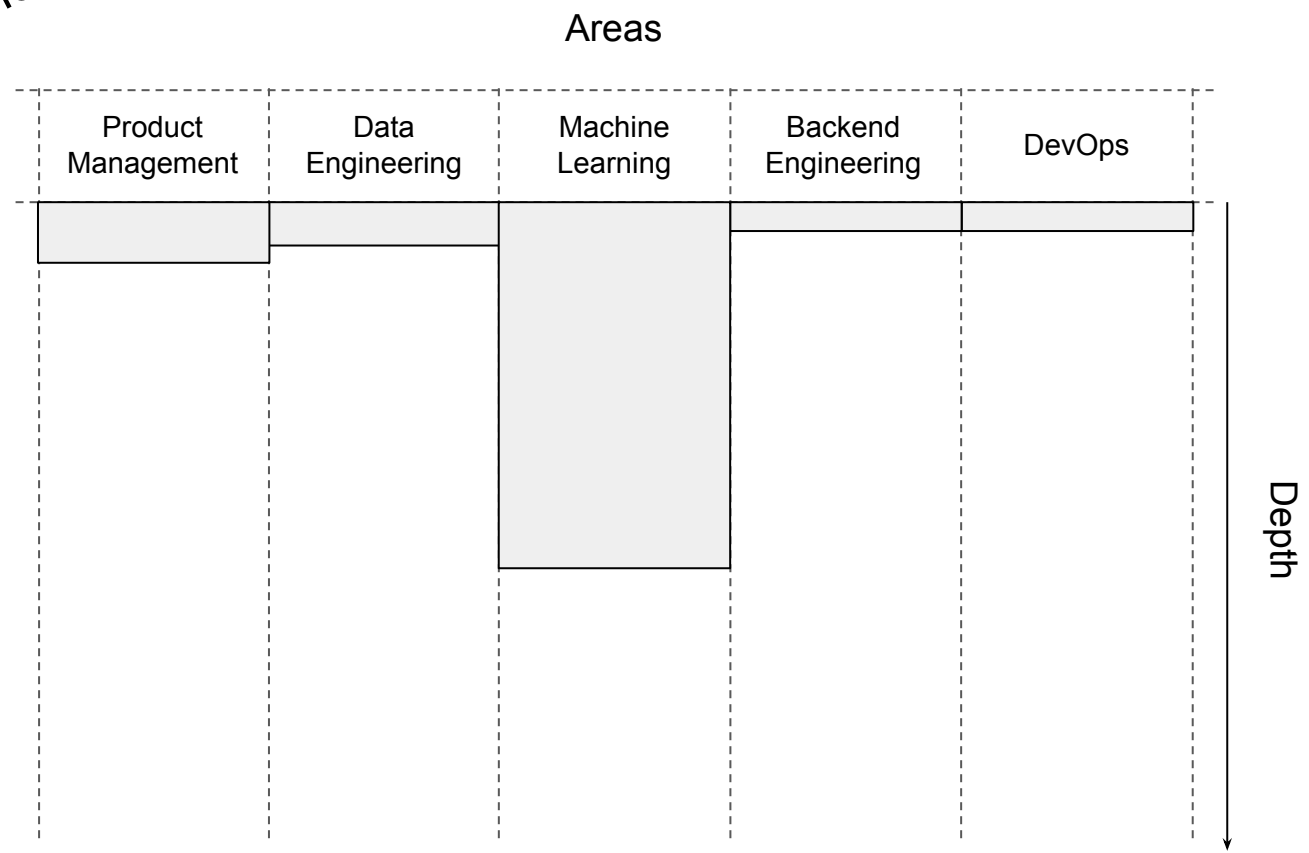

Areas
Product Management
Data Engineering
Machine Learning
Backend Engineering
DevOps
Depth

How other DS see me *

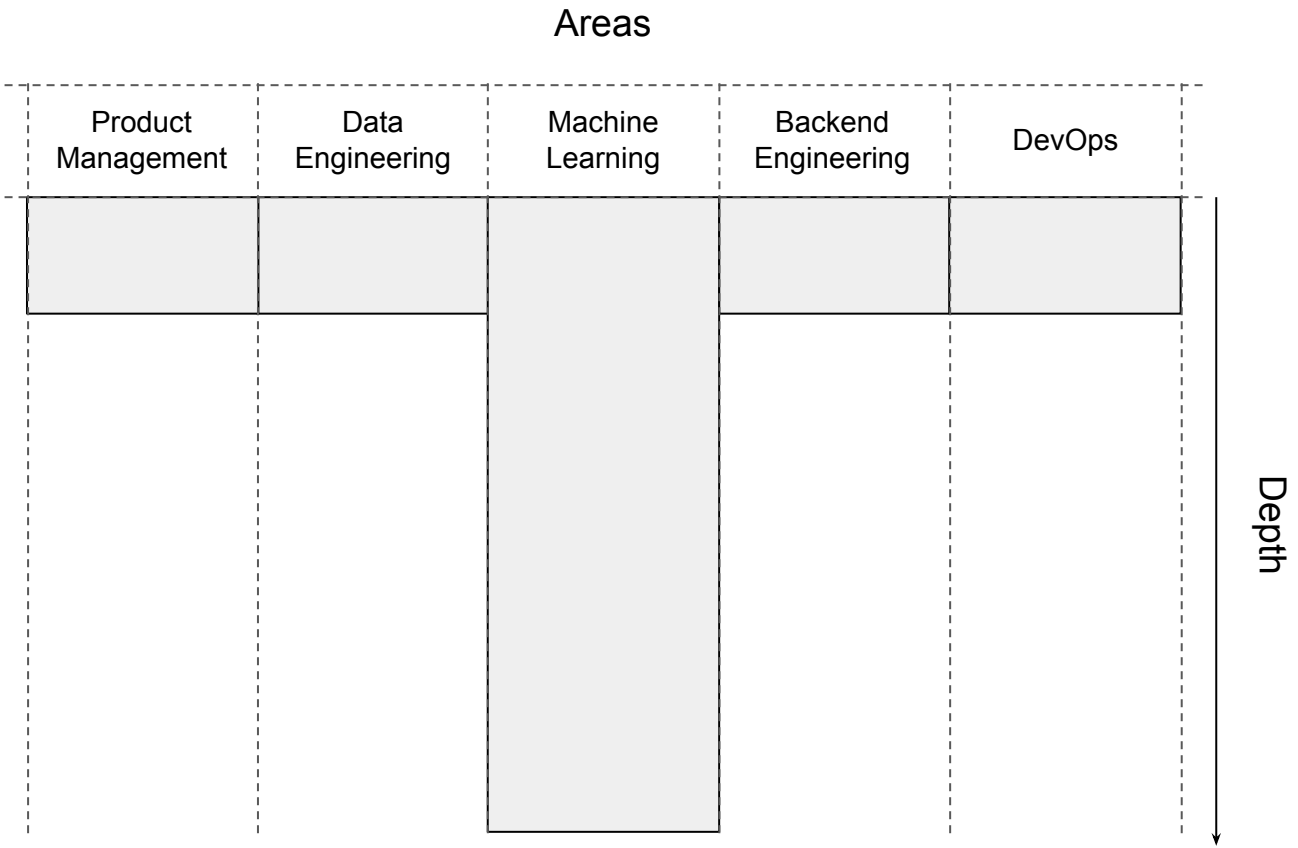


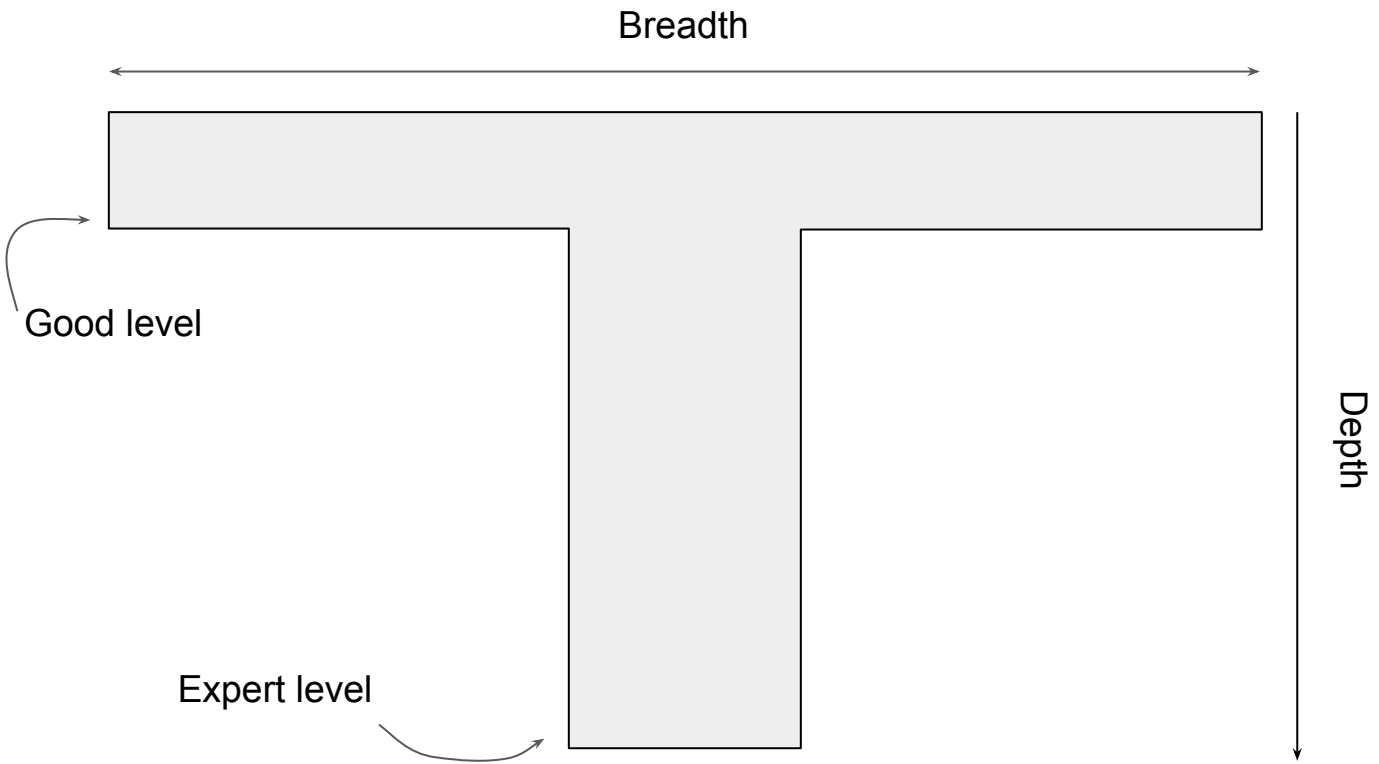
* maybe not

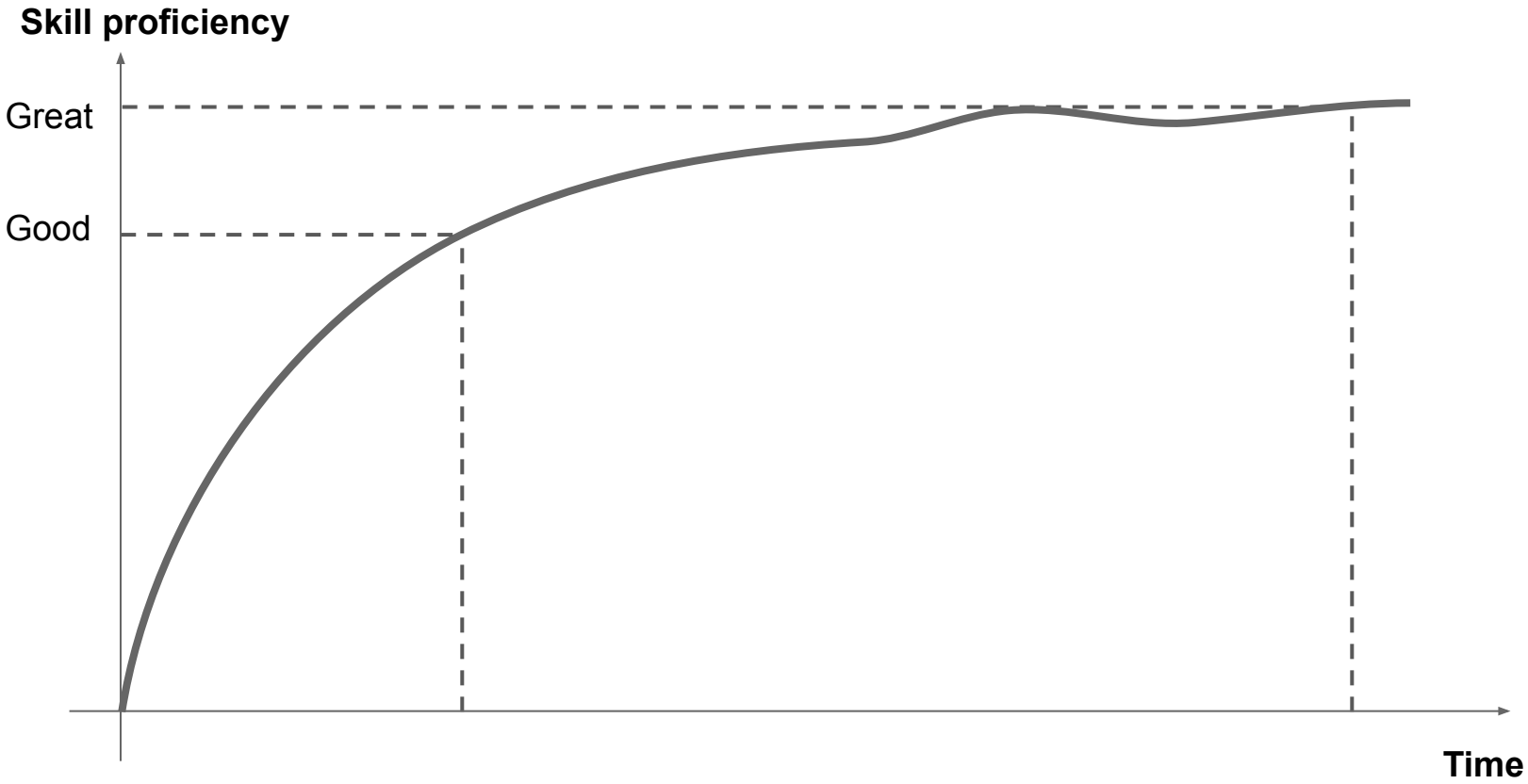
How engineers see me *



* maybe not


Areas
Product Management
Data Engineering
Machine Learning
Backend Engineering
DevOps
Depth


Breadth
Good level
Depth
Expert level


Skill proficiency
Great
Good
Time

80/20 rule

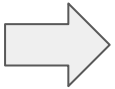
- Break down the role into core areas and skills
- Order the skills by importance
- Pick the most important ones
- Practice practice practice

Product management

- Strategy
- UX & Design
- Communication
- Planning
- Evaluation

Product management

- Strategy
- UX & Design
- Communication
- Planning
- Evaluation



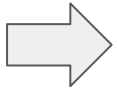
- Requirement gathering
- Prioritization
- Stakeholder management

Data engineering

- SQL Databases
- NoSQL Databases
- Stream processing
- Batch processing
- Data pipelines

Data engineering

- SQL Databases
- NoSQL Databases
- Stream processing
- Batch processing
- Data pipelines



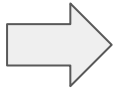
- MySQL
- AWS (S3, Kinesis)
- Spark
- Airflow

Backend engineering

- SQL & NoSQL Databases
- CS fundamentals
- Languages and frameworks
- Web services
- Best practices

Backend engineering

- SQL & NoSQL Databases
- CS fundamentals
- Languages and frameworks
- Web services
- Best practices



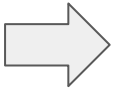
- Python
- Docker
- Tests
- Clean code

DevOps

- Infrastructure
- Automation
- Monitoring
- Reliability

DevOps

- Infrastructure
- Automation
- Monitoring
- Reliability



- AWS
- Kubernetes
- Terraform

Plan

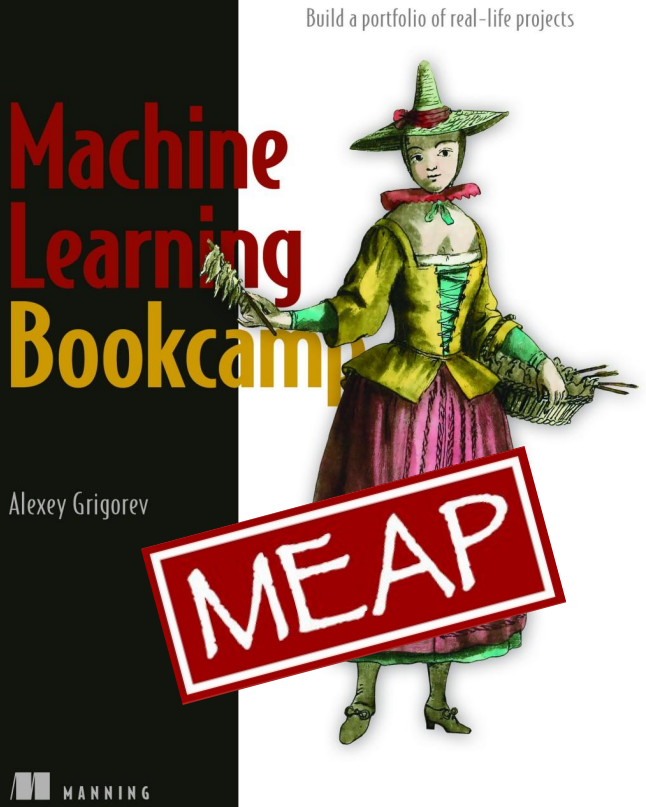
- Data science process (CRISP-DM)
- Roles in the team
- Full-stack data scientist: Jack of all trades
- Becoming full-stack

Summary

- Cover the whole ML project lifecycle
- Invest in software engineering
- It’s not only about technical skills
- Be a T-shaped professional
- Focus on what matters

OLX GROUP

olxgroup.com/careers



Machine Learning Bookcamp

- Learn Machine Learning by doing projects
- http://bit.ly/mlbookcamp
- Get 40% off with code “grigorevpc”

mlbookcamp.com

agrigorev @Al_Grigor
alexeygrigorev.com
mlbookcamp.com