UNIVERSITÀ
DEGLI STUDI
DI TRIESTE

Class 13: MLOps

Master Course:

Data-driven Systems Engineering (ML Operations)

440MI and 305SM

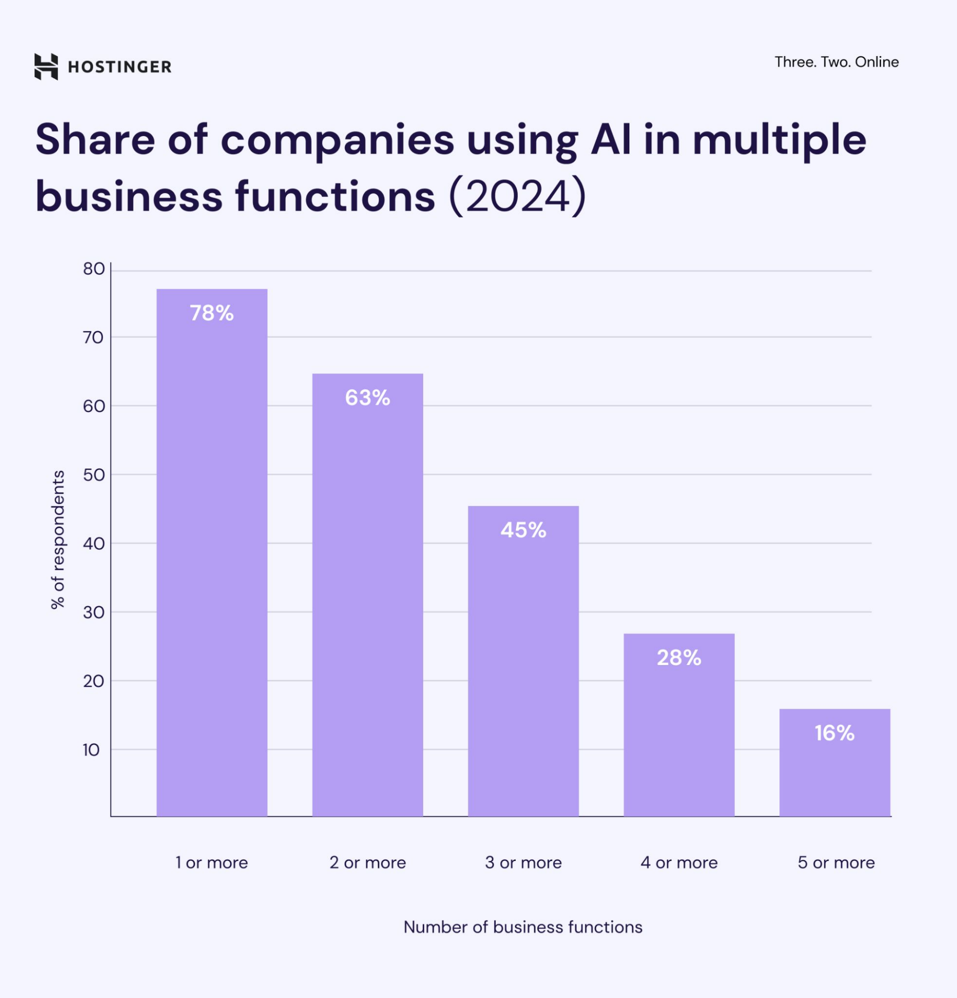
L16- Introduction to MLOps

Today's goal:

- AI Maturity
- Overview of MLOps
- People and Roles
- Preparing for Production

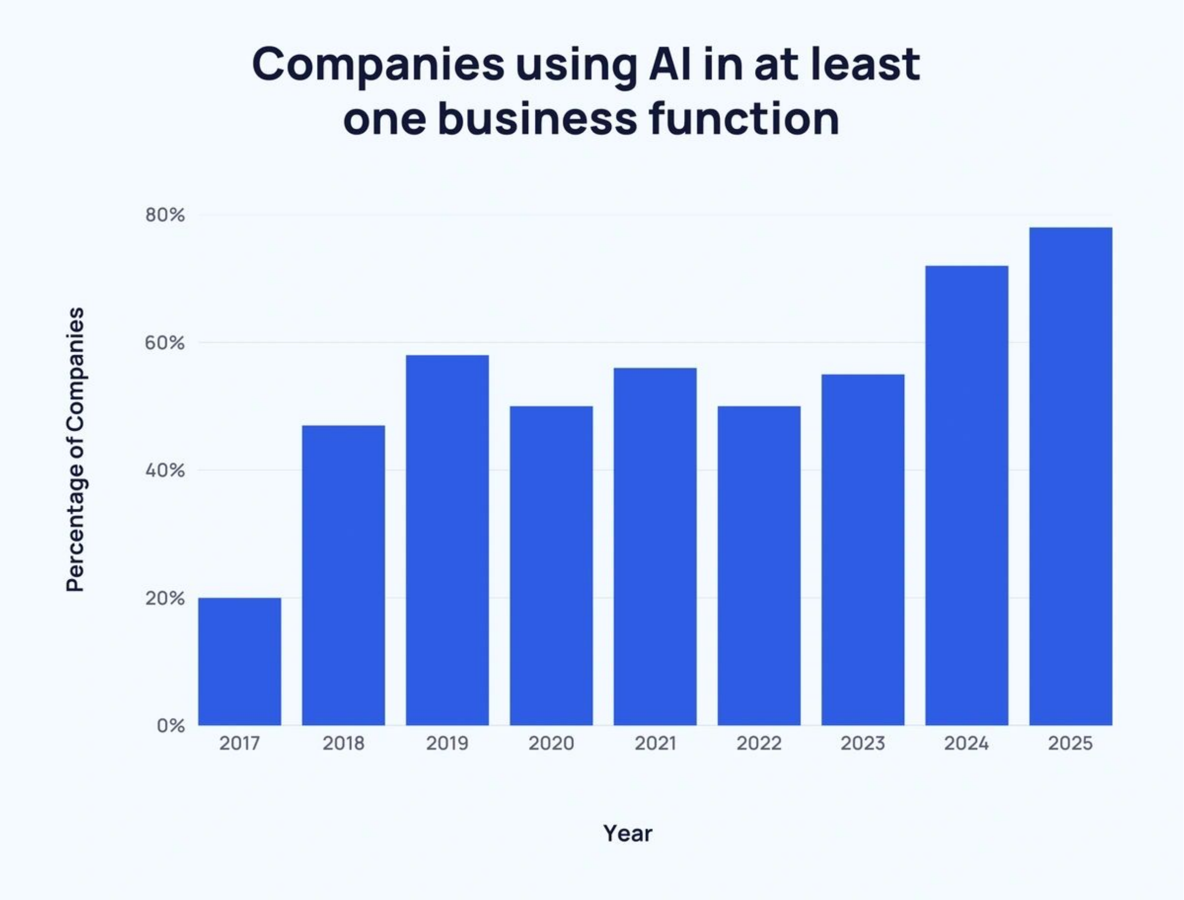
Preliminaries:

According to McKinsey & Company’s “State of AI 2025” survey, 88 % of organisations report using AI in at least one business function (up from 78 % a year ago).



Preliminaries:

From the Stanford Institute for Human-Centred Artificial Intelligence "AI Index 2025", in 2024 the share of companies using AI jumped to 78 %, up from 55 % in 2023.



Preliminaries:

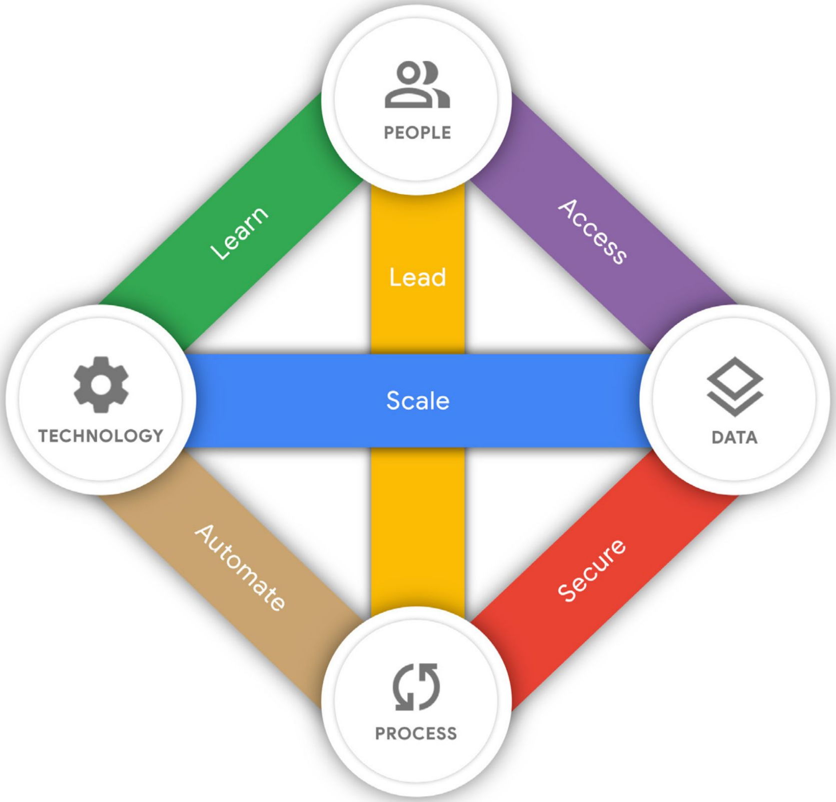
Why AI is so hard (or could be) to be adopted?

1. Excessive manual work and lack of standardization;
2. Inefficient handoffs between Data Science and IT/Engineering. Typical handoffs include:
   a. Data Engineering → Data Science
   Data pipeline, cleaned datasets, feature store access.
   1. Data Science → ML Engineering
   Model code, training scripts, evaluation metrics.
   2. Data Science → IT / DevOps
   Final model artifact, dependencies, environment configuration.
   3. ML Engineering → Production Ops
   Deployment pipeline, monitoring dashboards, rollback procedures.
3. Integration gaps;
4. Deployment, scaling, and versioning challenges;
5. Weak change-management processes and governance;

Preliminaries: AI Maturity

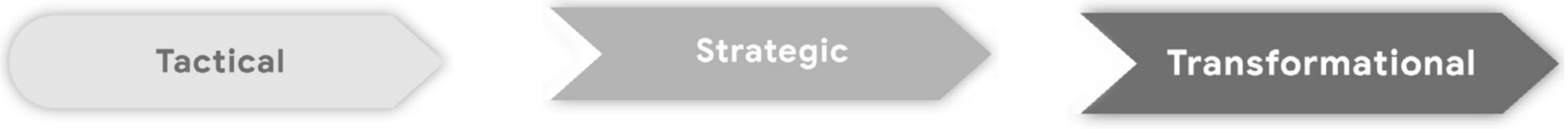
- How do you take advantage of the power inherent in AI, while avoiding any potential missteps?

AI Maturity Themes!



Preliminaries: AI Maturity

For each theme, those practices will fall into one of the following phases:

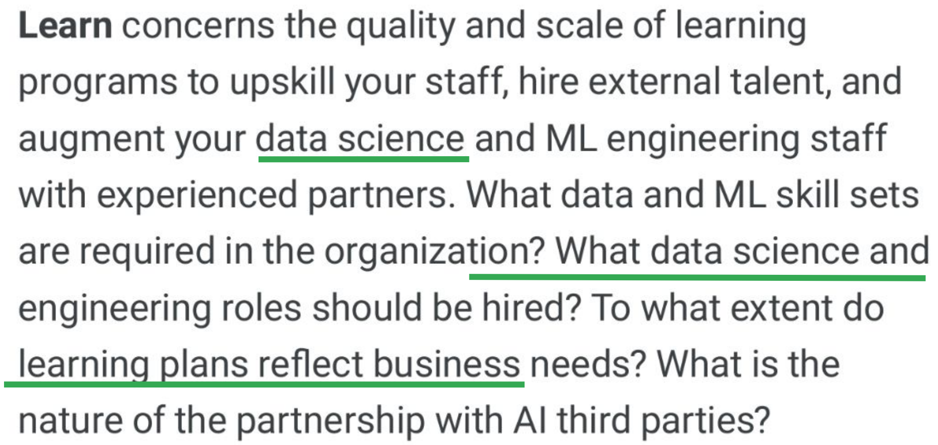


The focus is on **easy** adoption, minimal disruption, and **quick** achievements.

A broader vision governs AI adoption. Perception **starts to move** beyond the hype, **becoming** a pivotal accelerator for the **business**.

There is a **mechanism** in place for scaling and **promoting ML** capabilities across the organization.

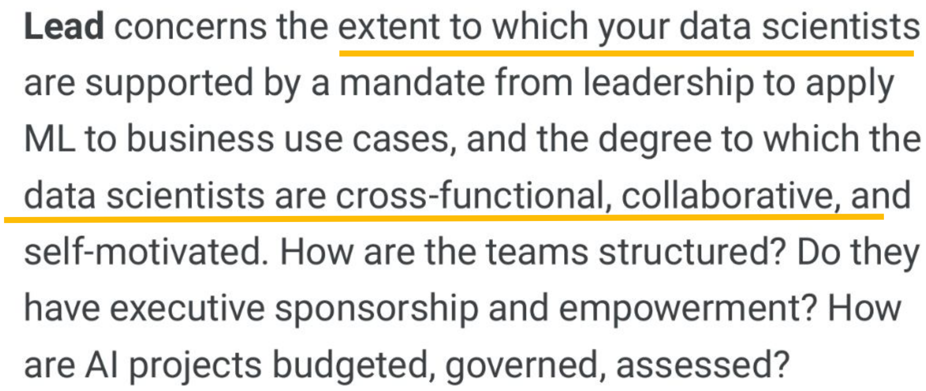
Preliminaries: AI Maturity



Learn concerns the quality and scale of learning programs to upskill your staff, hire external talent, and augment your data science and ML engineering staff with experienced partners. What data and ML skill sets are required in the organization? What data science and engineering roles should be hired? To what extent do learning plans reflect business needs? What is the nature of the partnership with AI third parties?

PEOPLE
Learn
TECHNOLOGY

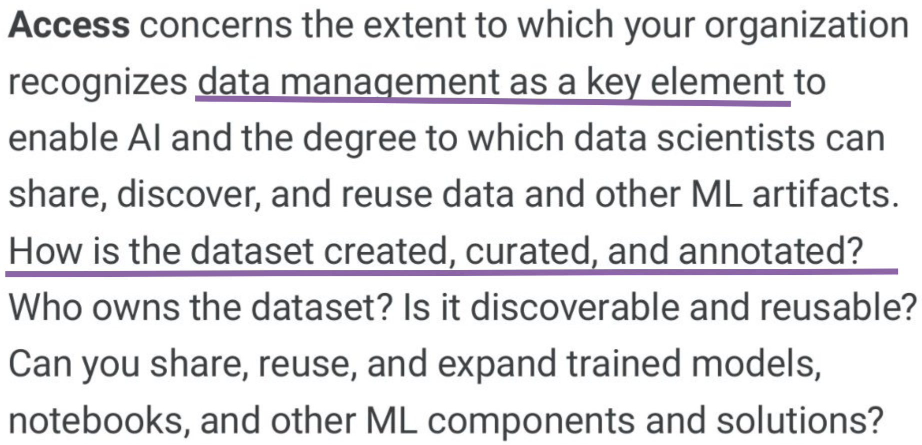
Preliminaries: AI Maturity



Lead concerns the extent to which your data scientists are supported by a mandate from leadership to apply ML to business use cases, and the degree to which the data scientists are cross-functional, collaborative, and self-motivated. How are the teams structured? Do they have executive sponsorship and empowerment? How are AI projects budgeted, governed, assessed?

PEOPLE
Lead
PROCESS

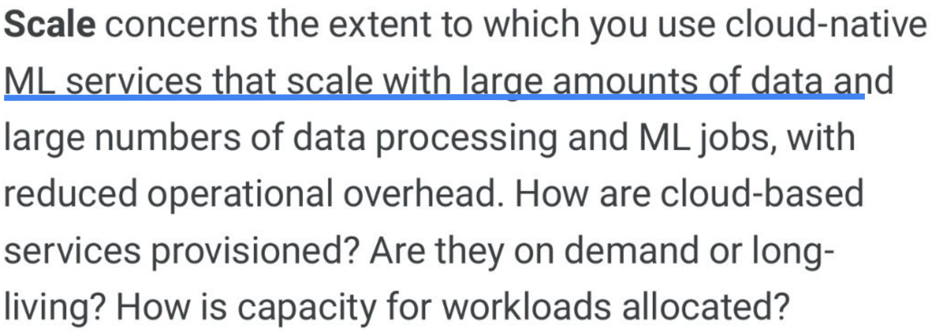
Preliminaries: AI Maturity



Access concerns the extent to which your organization recognizes data management as a key element to enable AI and the degree to which data scientists can share, discover, and reuse data and other ML artifacts. How is the dataset created, curated, and annotated? Who owns the dataset? Is it discoverable and reusable? Can you share, reuse, and expand trained models, notebooks, and other ML components and solutions?

PEOPLE
Access
DATA

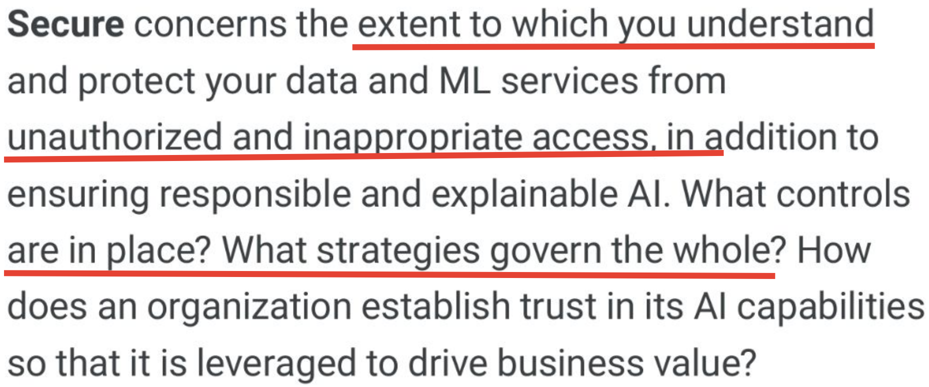
Preliminaries: AI Maturity



Scale concerns the extent to which you use cloud-native ML services that scale with large amounts of data and large numbers of data processing and ML jobs, with reduced operational overhead. How are cloud-based services provisioned? Are they on demand or long-living? How is capacity for workloads allocated?

DATA
Scale
TECHNOLOGY

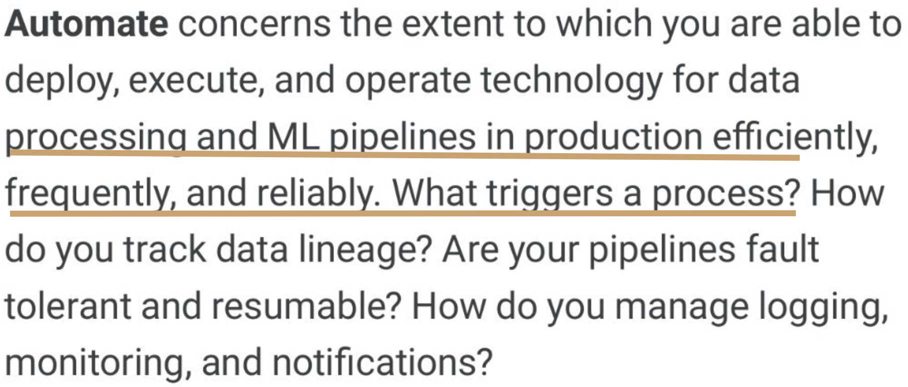
Preliminaries: AI Maturity



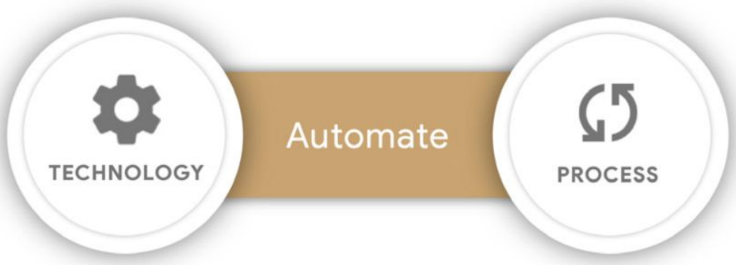
Secure concerns the extent to which you understand and protect your data and ML services from unauthorized and inappropriate access, in addition to ensuring responsible and explainable AI. What controls are in place? What strategies govern the whole? How does an organization establish trust in its AI capabilities so that it is leveraged to drive business value?

DATA
Secure
PROCESS

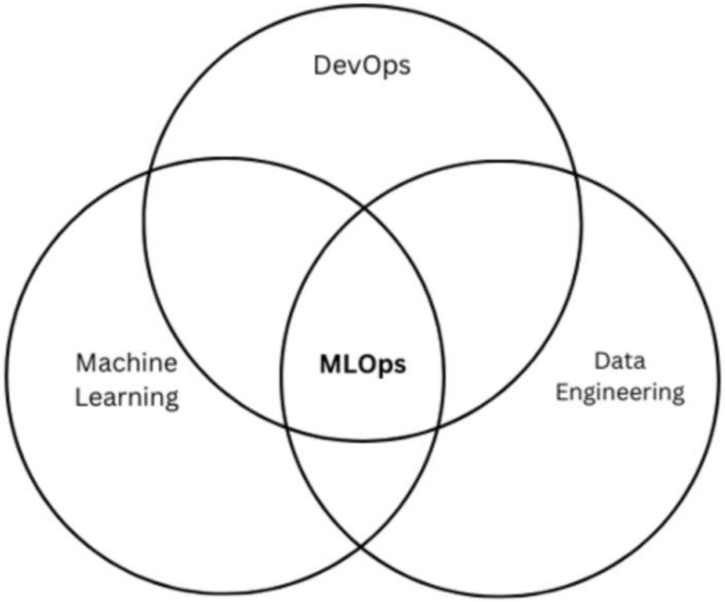
Preliminaries: AI Maturity



Automate concerns the extent to which you are able to deploy, execute, and operate technology for data processing and ML pipelines in production efficiently, frequently, and reliably. What triggers a process? How do you track data lineage? Are your pipelines fault tolerant and resumable? How do you manage logging, monitoring, and notifications?



Overview of MLOps



- MLOps is a **methodology** for ML engineering that unifies ML system development (the **ML element**) with ML system operations (the **Ops element**).

- It advocates **formalizing** and **automating** critical steps of ML system construction.

- MLOps **provides** a set of **standardized processes** and **technology** capabilities for building, deploying, and operationalizing ML systems rapidly and reliably.

Overview of MLOps

DevOps

When you deploy a web service, you care about:

- resilience;
- queries per second;
- load balancing;

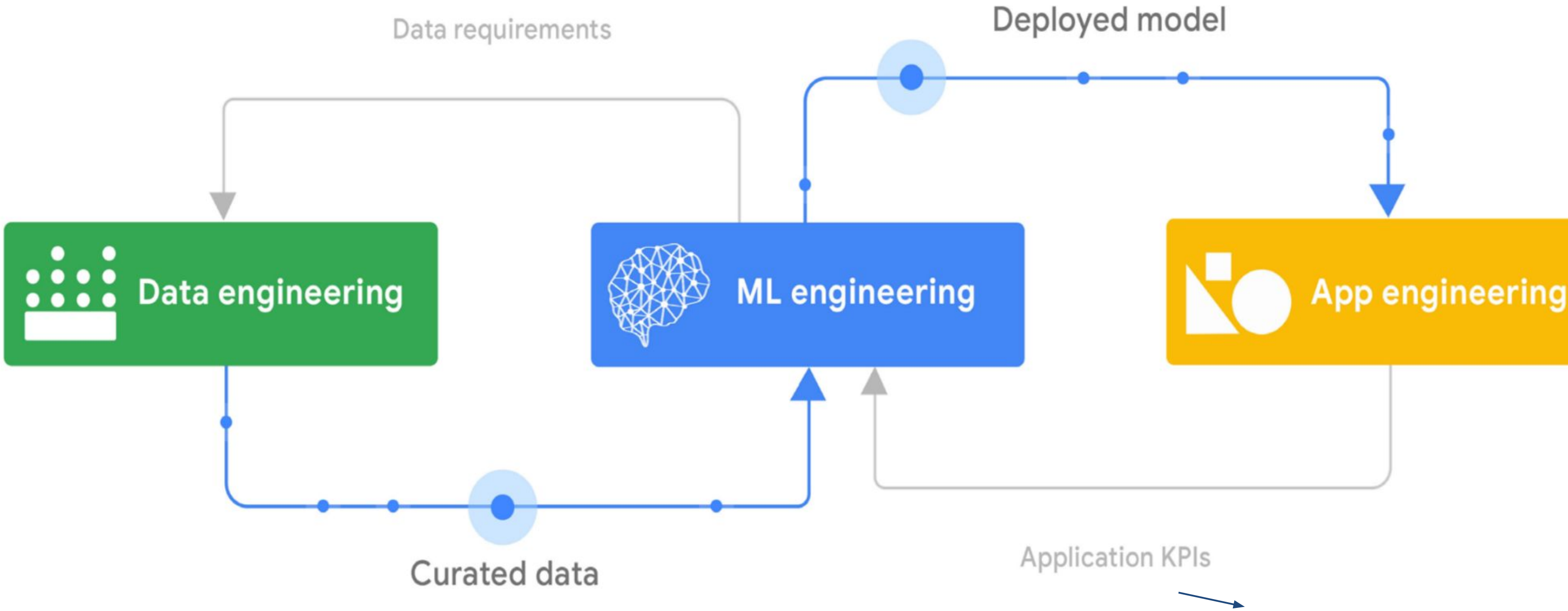
MLOps

When you deploy an ML model, you also need to:

- worry about changes in the data;
- changes in the model;
- users trying to “cheat” the system;

Overview of MLOps

Building ML-based System

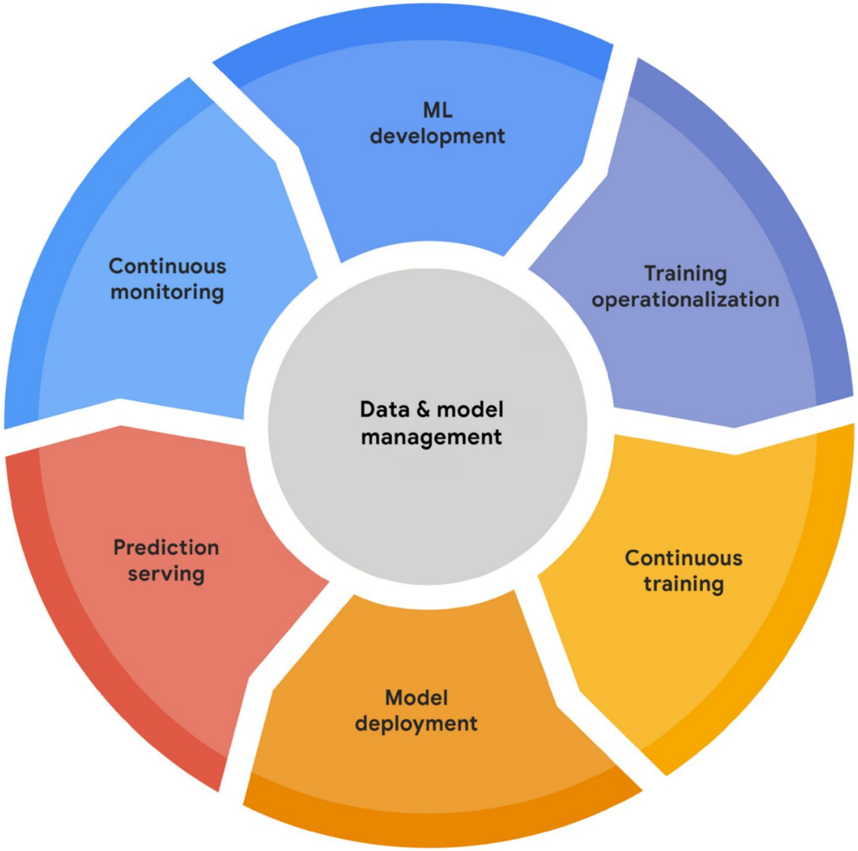


Key Performance Indicator

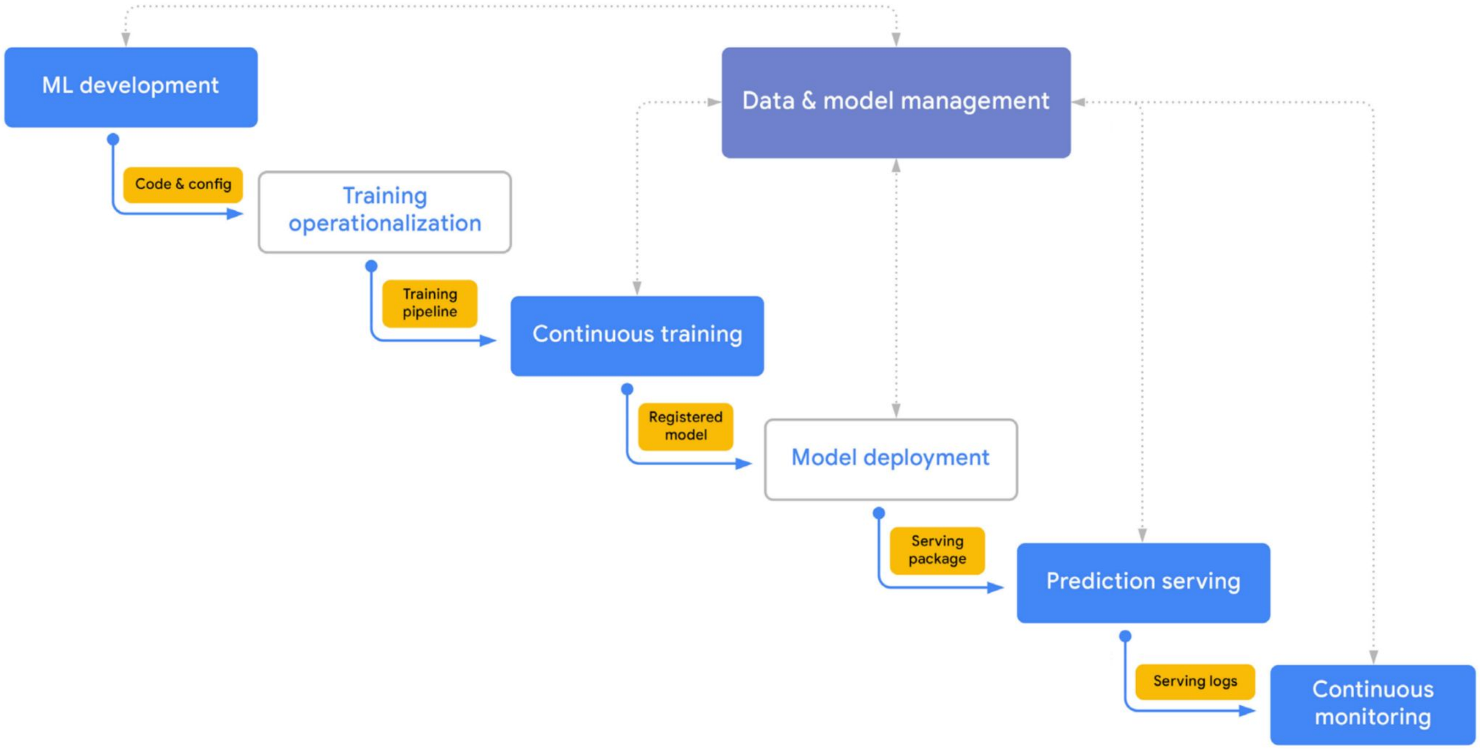
Overview of MLOps

The MLOps lifecycle:

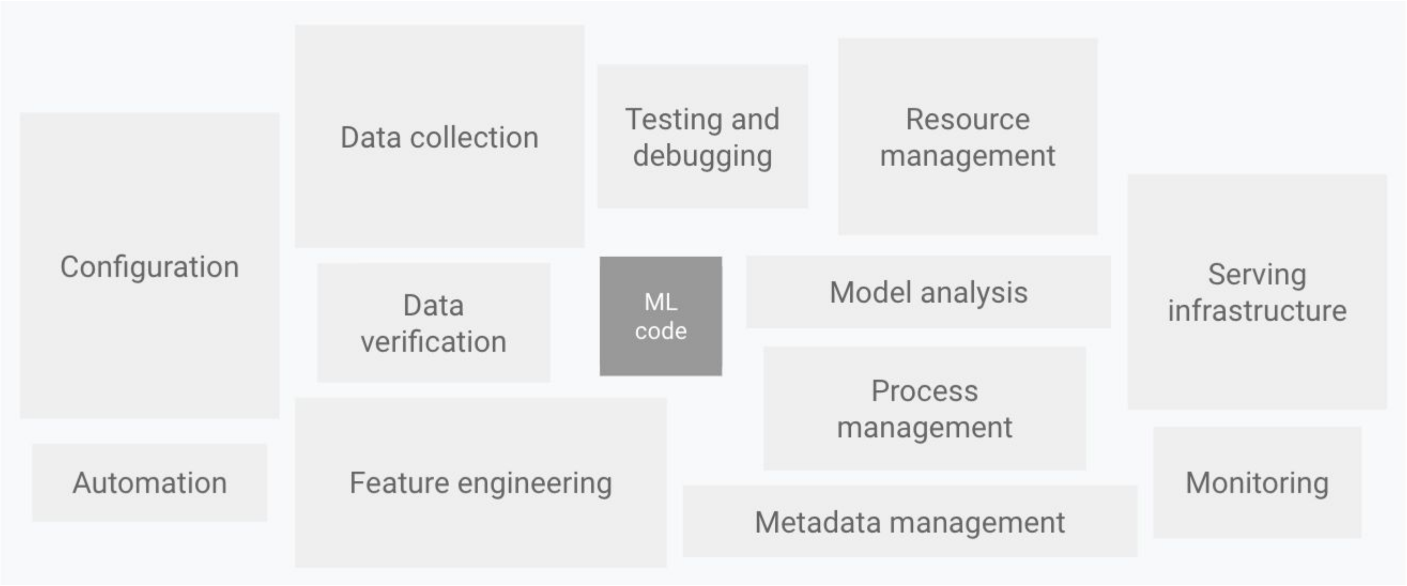
7 processes



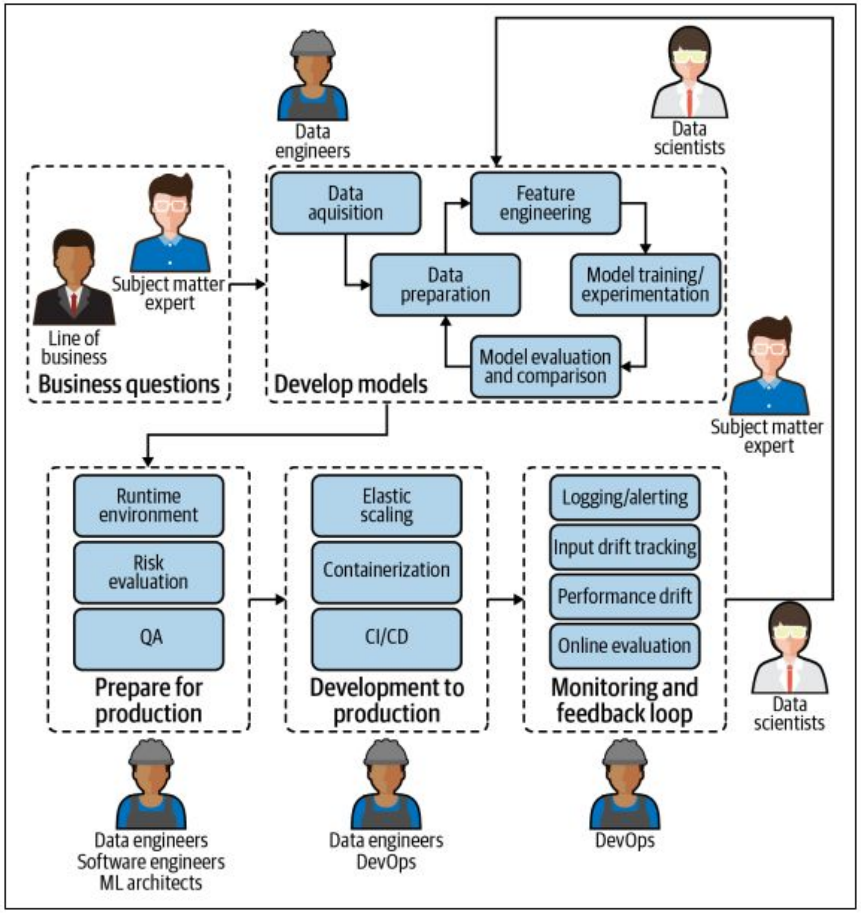
MLOps Workflow



MLOps Infrastructure



People and Roles:


Business questions

Line of business

Subject matter expert

Develop models

Data engineers

Data scientists

Data aquisition

Feature engineering

Data preparation

Model training/ experimentation

Model evaluation and comparison

Subject matter expert

Prepare for production

Runtime environment

Risk evaluation

QA

Data engineers
Software engineers
ML architects

Development to production

Elastic scaling

Containerization

CI/CD

Data engineers
DevOps

Monitoring and feedback loop

Logging/alerting

Input drift tracking

Performance drift

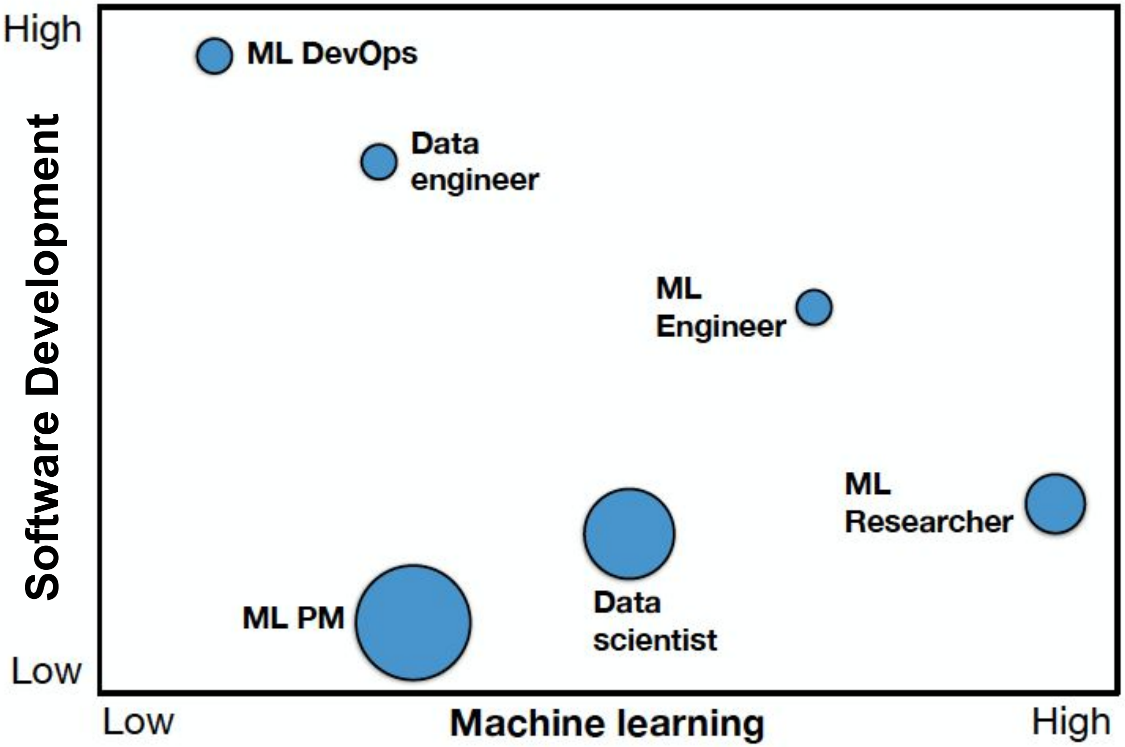
Online evaluation

DevOps

Data scientists

People and Roles:

AI (ML) and Dev Community!



People and Roles:

Subject matter experts:

Role:

- Provide business questions, **goals**, or **KPIs** around which ML models should be framed.
- **Continually evaluate** and ensure that model **performance** aligns with or resolves the initial need.

Requirements:

- Easy way to understand deployed model **performance in business terms**.
- Mechanism or feedback loop for model results.

People and Roles:

Data scientists:

Role:

- **Build models** that address the business question or needs brought by subjec matter experts.
- **Deliver operationalizable models** so that they can be properly used in the production environment and with production data.
- **Assess model quality** (of both original and tests) in tandem with subject mat experts to ensure they answer initial business questions or **needs.**

Requirements:

- **Automated model packaging** for quick and easy deployment to production.
- **Ability** to develop tests to **determine the quality of deployed models**
- **Ability to investigate data pipelines** of each model to make quick assessments and adjustments regardless of who originally built the model.

People and Roles:

Data engineers:

Role:

- **Optimize** the **retrieval and use of data** to power ML models.

Requirements:

- **Visibility into performance** of all deployed models.
- **Ability to see** the full details of individual data pipelines to address underlying **data plumbing issues.**

People and Roles:

Software engineers:

Role:

- **Integrate ML models** in the company's **applications and systems**.
- **Ensure that ML models work** seamlessly with other non-machine-learning-based applications.

Requirements:

- **Versioning and automatic tests.**
- The ability to work in parallel on the same application.

People and Roles:

DevOps:

Role:

- Conduct and build operational systems and **test for security**, **performance**, **availability**.
- **Continuous Integration/Continuous Delivery** (CI/CD) pipeline management.

Requirements:

- Seamless integration of MLOps into the larger DevOps strategy of the enterprise.
- **Seamless deployment pipeline**.

People and Roles:

Machine learning architects:

Role:

- **Ensure a scalable and flexible** environment for **ML model pipelines**, from design to development and monitoring.
- **Introduce new technologies** when appropriate that **improve ML model** performance in production

Requirements:

- **High-level overview of models** and their resources consumed.
- **Ability to drill down into data pipelines to assess** and adjust infrastructure needs.

People and Roles:

Model risk managers/ auditors:

Role:

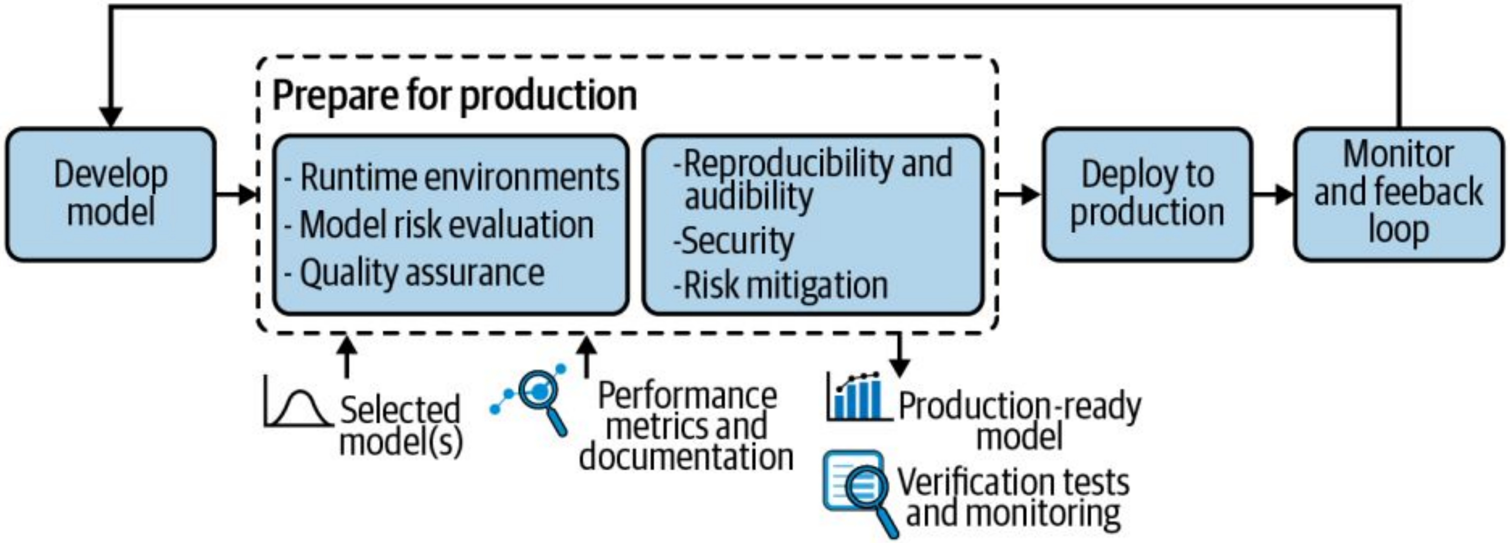
- **Minimize overall risk** to the company as a result of ML models in production.
- **Ensure compliance with internal and external requirements** before pushing ML models to production

Requirements:

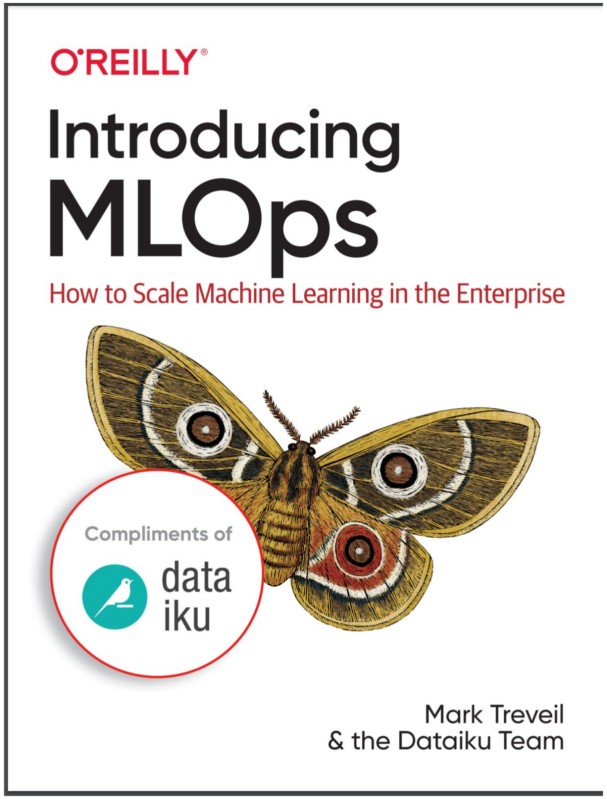
- Robust, likely automated, **reporting tools** on all models (currently or ever in production), including data lineage.

Preparing for Production

- Confirming that something **works in the laboratory** has never been a sure sign it will



Textbook



O'REILLY

Introducing MLOps

How to Scale Machine Learning in the Enterprise

Compliments of dataiku

Mark Treveil
& the Dataiku Team