—

MLOps

best practices of putting machine learning models to production

BORIS CERGOL, REGIONAL HEAD OF DATA

Agenda

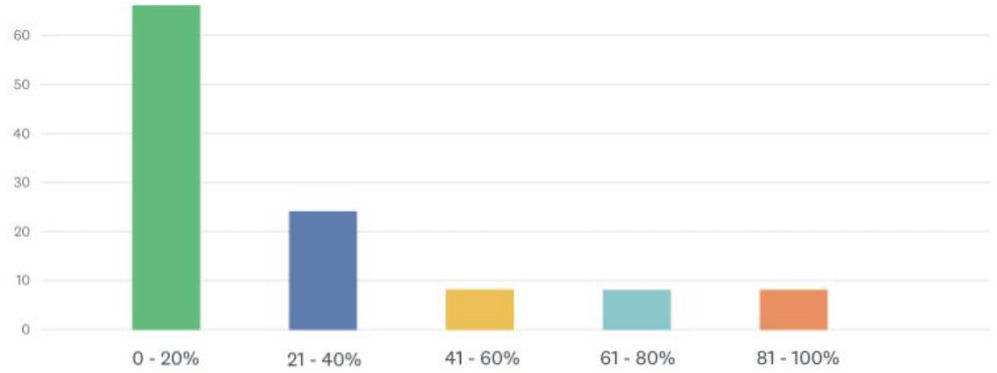
1. INTRODUCTION
2. MLOPS PRINCIPLES
3. MLOPS TOOLS
4. MLOPS WORKFLOWS

1

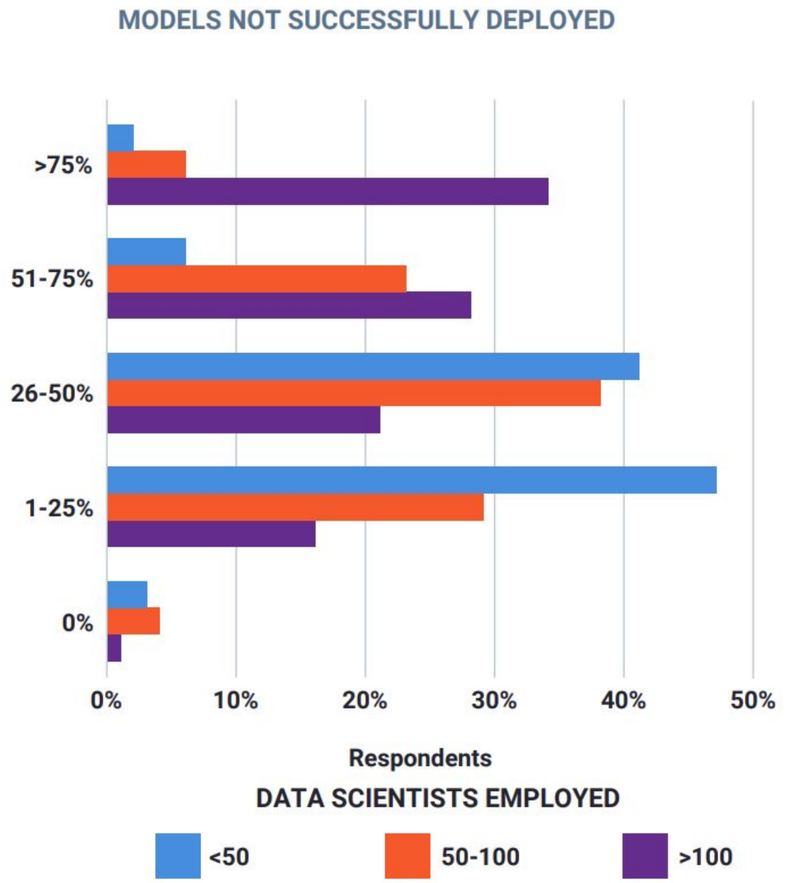
Introduction

Machine learning deployment gap

- Machine learning can solve many practical problems.
- Business value only generated if models are deployed as part of wider software system.
- Bridging the gap between building models and deploying them is challenging.



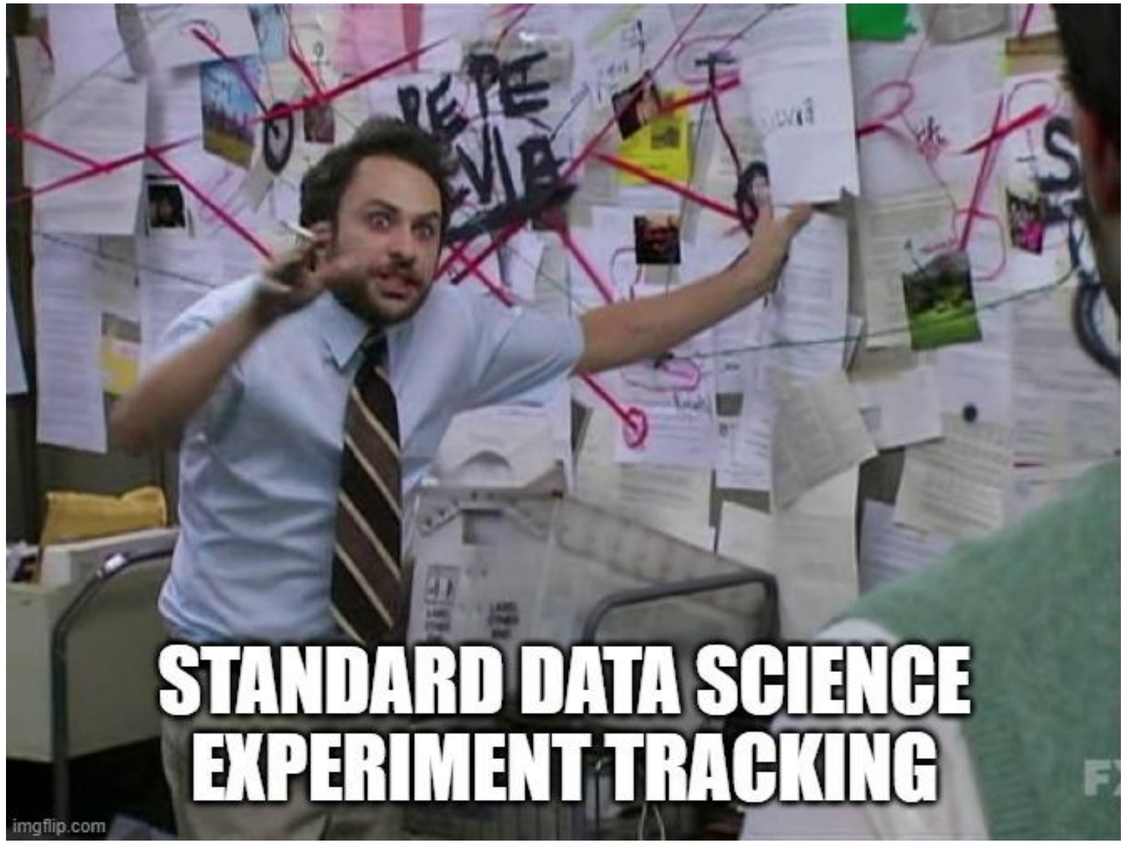
Percentage of ML models intended to be deployed
that were actually deployed,
Kdnuggets survey, 2021.



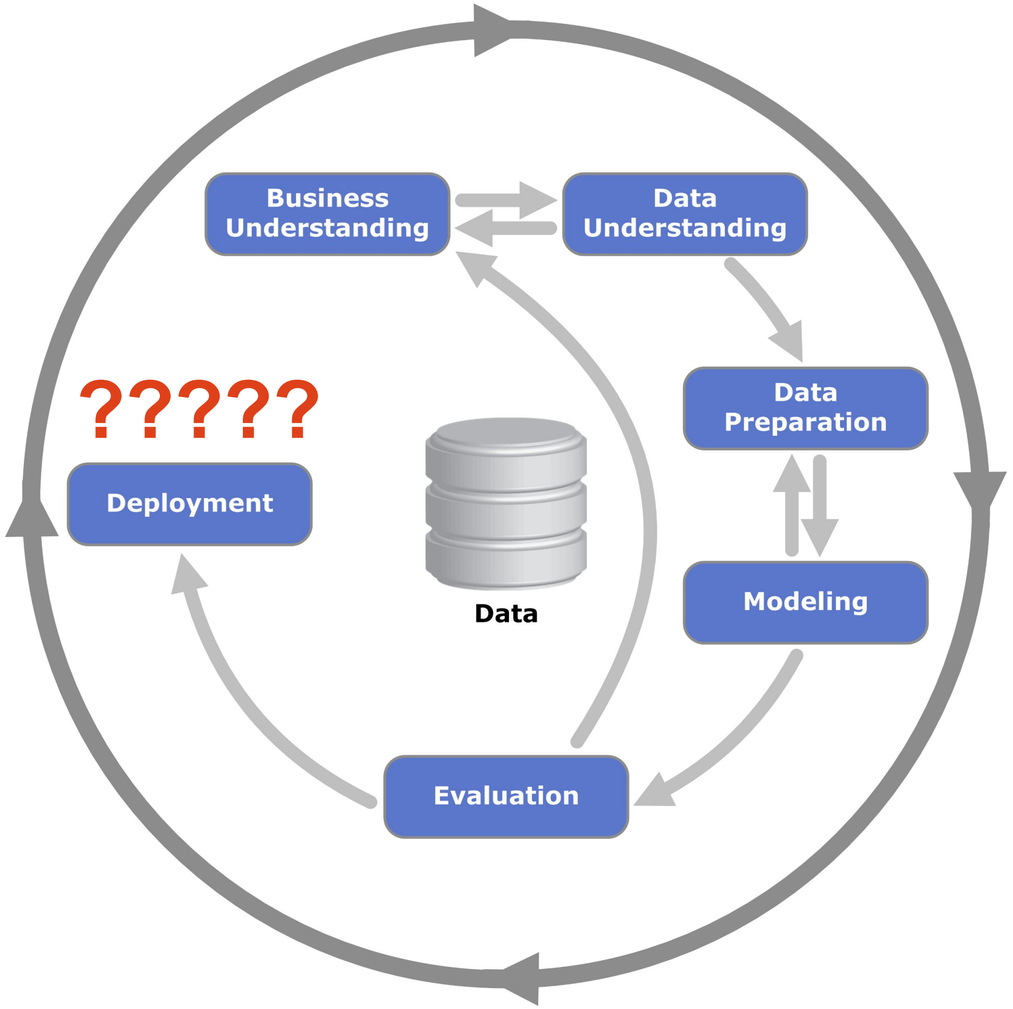
https://www.datarobot.com/resources/5-latest-trends-in-enterprise-machine-learning-2021/thank-you/

Problematic practices in data science

- A lot of repetitive manual work.
- Messy experiment tracking.
- Data, models and sometimes even code not versioned.
- Avoiding testing code and code reviews.
- Laser focus on improving a single metric.
- ...

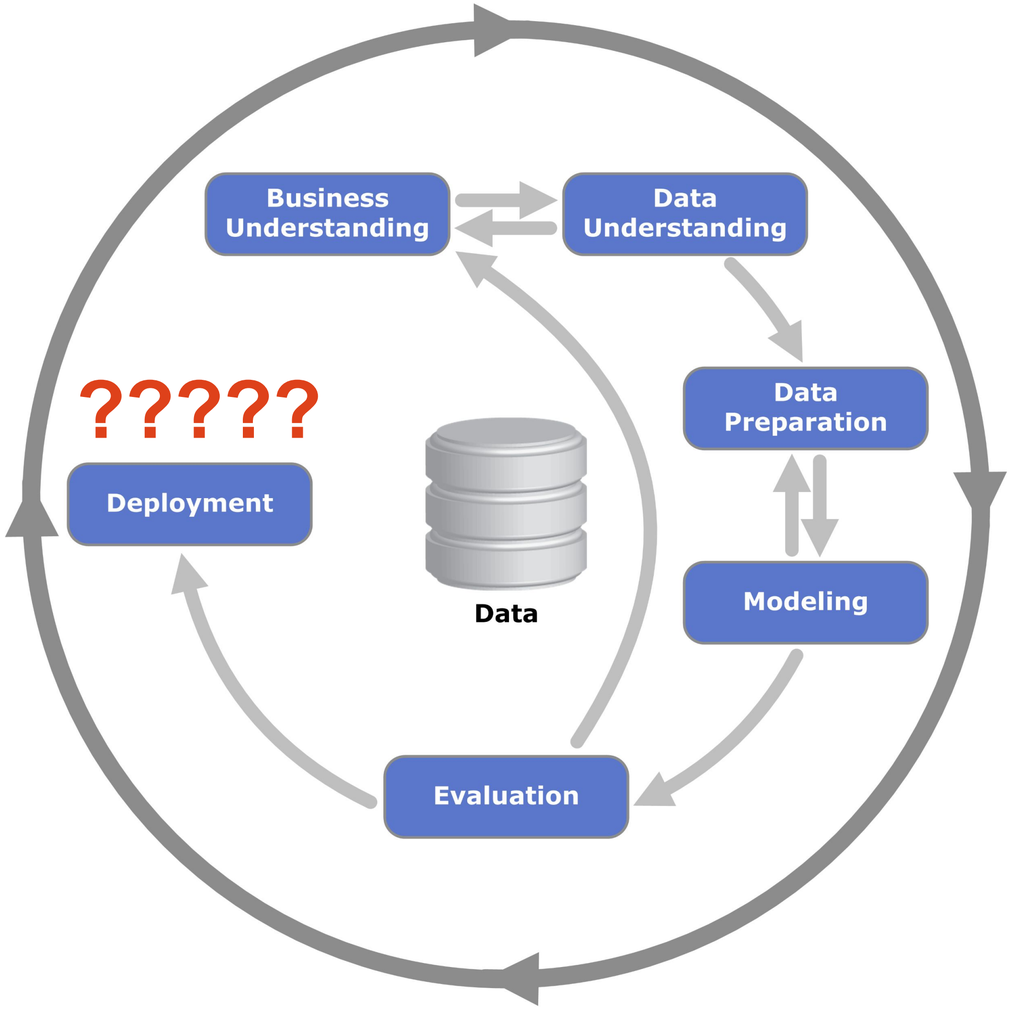


The root of all evil



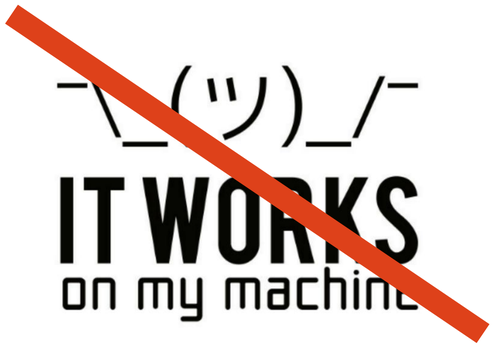
CRISP-DM – Cross-Industry Standard Process for Data Mining

The root of all evil

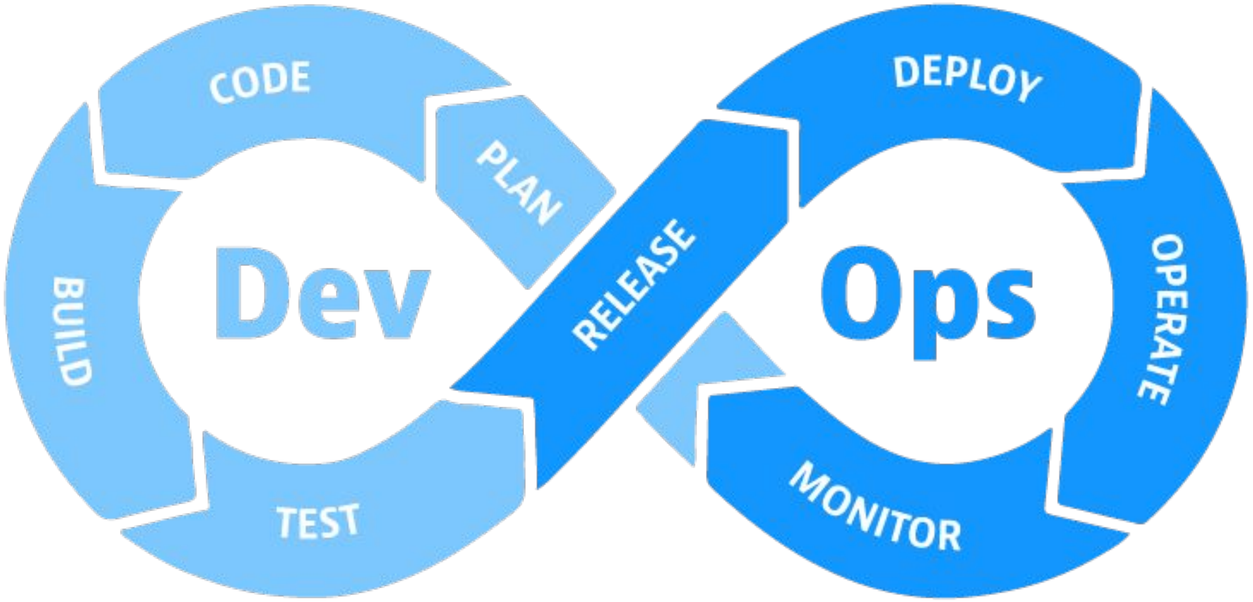


CRISP-DM – Cross-Industry Standard Process for Data Mining

Meanwhile, in software development…

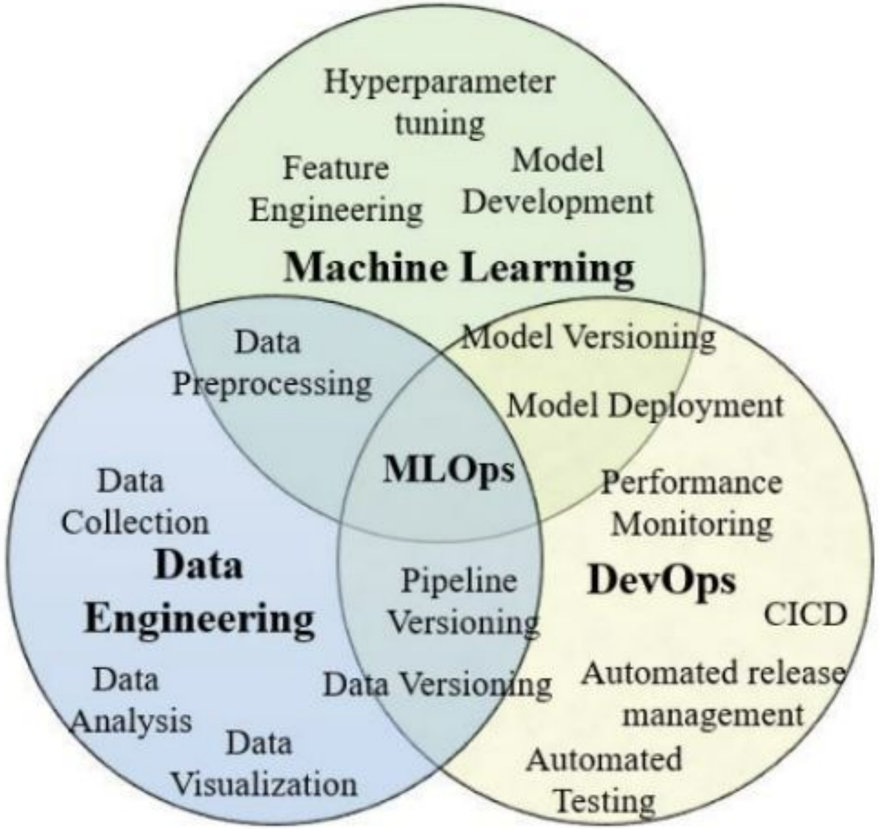


- **DevOps** (Development + Operations)
  - A set of practices, tools and culture focused on building better software faster by gap between software development and operations teams.
- Top 2 practices
  - **Continuous Integration (CI)**
    - Practice of frequently merging code by different devs to central repository, after which automated builds and tests are run.
  - **Continuous Delivery (CD)**
    - Practice according to which, there is constantly a new version of the software under development to be installed for testing, evaluation and then production.
- Also emphasized: **transparency**, **communication**, **collaboration**.



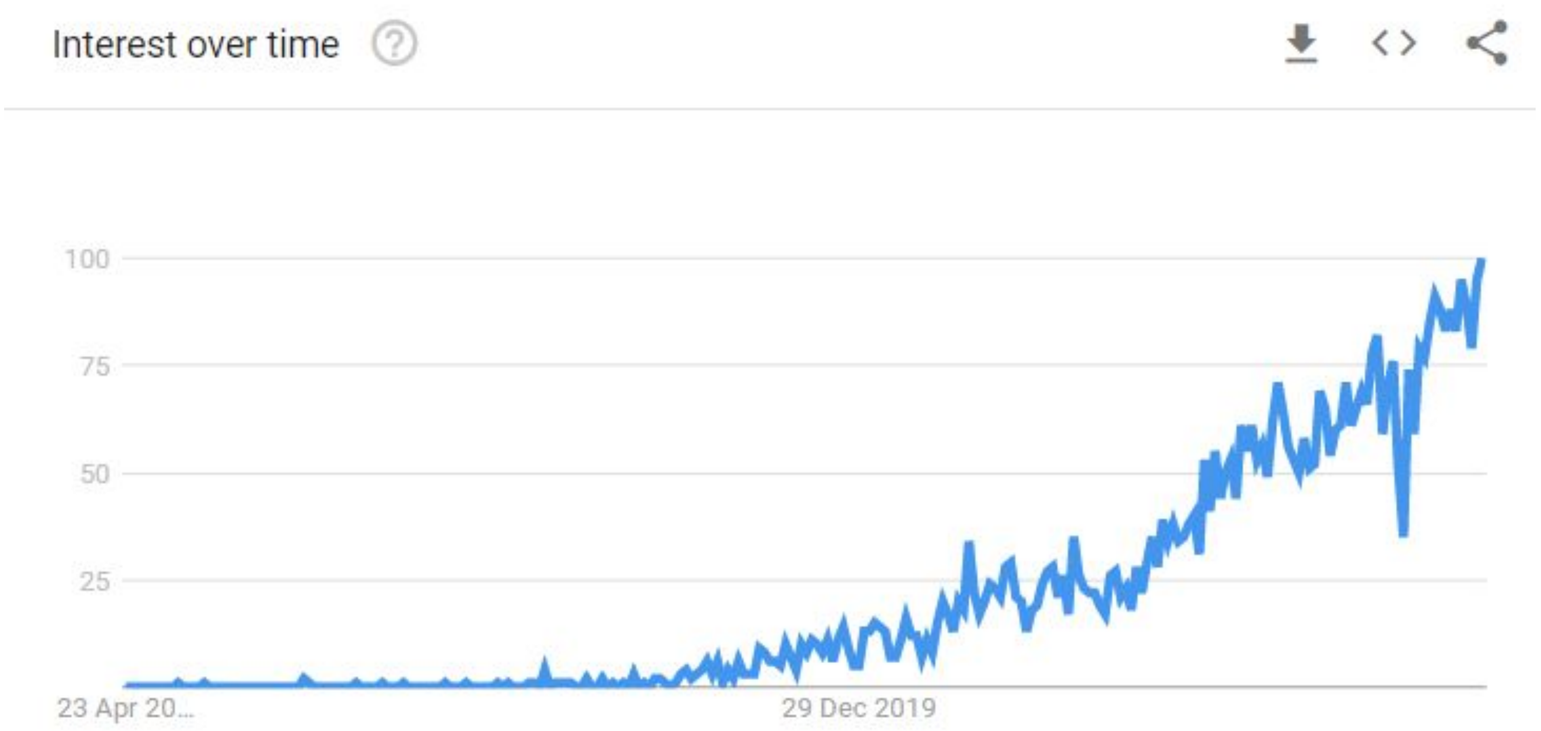
MLOps

- **MLOps** (Machine Learning + Operations)
  - A set of practices focused on rapidly and reliably developing, deploying and maintaining ML models in production.
  - Fosters greater communication and collaboration between data engineers, data scientists and operations professionals.



https://arxiv.org/abs/2202.10169 .

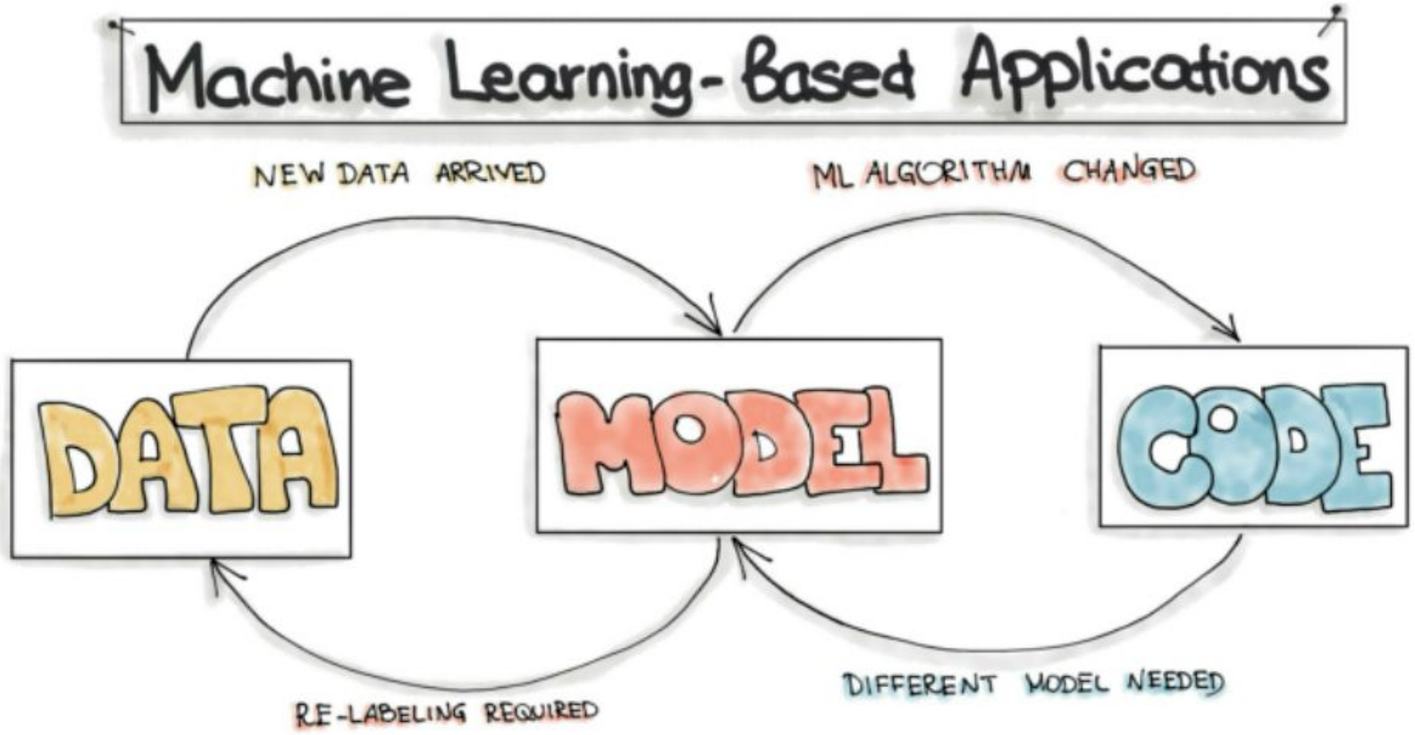
Interest in MLOps (by Google Trends)



1	China	100
2	South Korea	49
3	Singapore	29
4	Israel	18
5	Taiwan	15
6	Hong Kong	13
7	India	9
8	Japan	9
9	Sweden	9
10	Canada	8
11	Portugal	8
12	Finland	7
13	United States	6
14	Netherlands	6
15	Germany	6

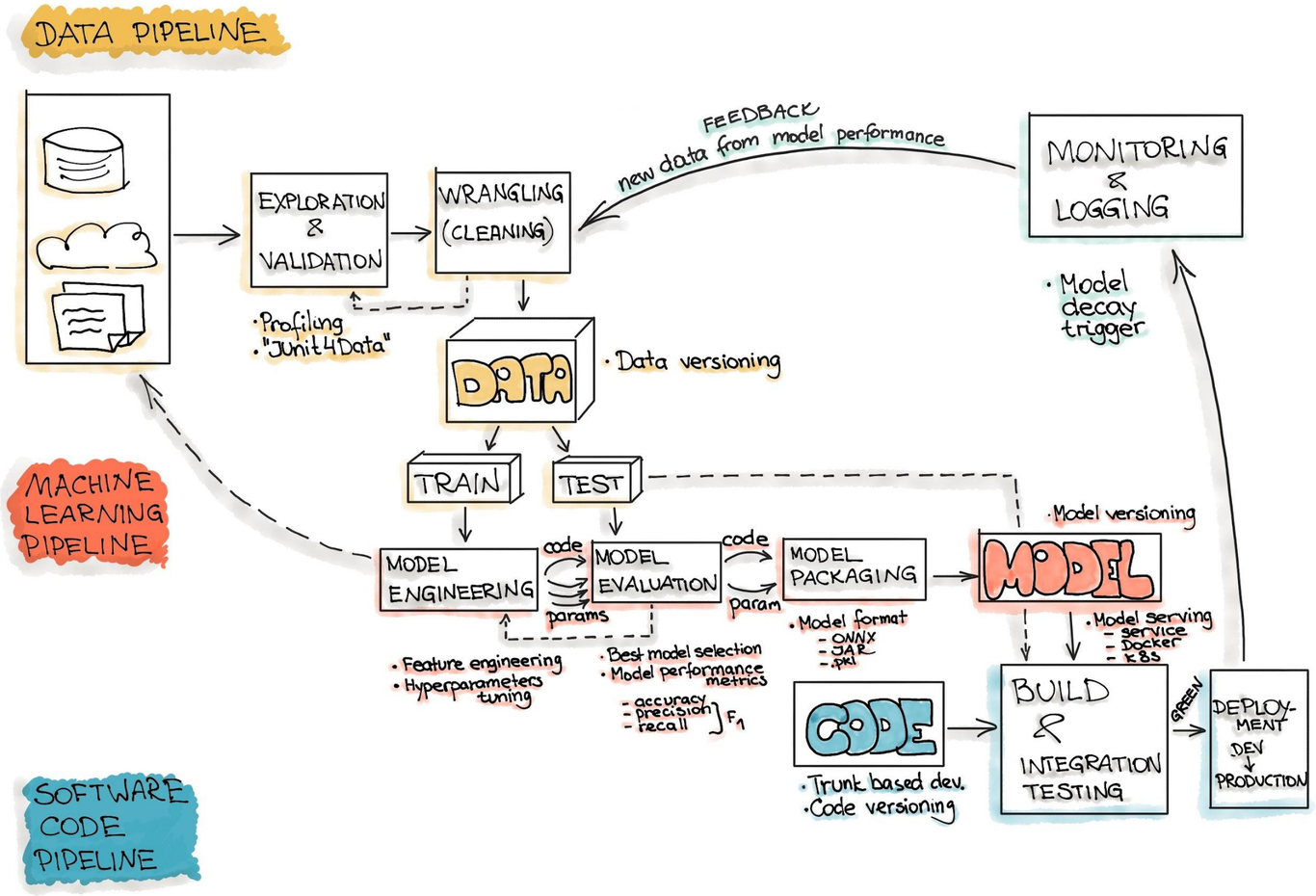
Main Challenge for MLOps

- Machine learning systems: an entanglement of data, models and code.
- Data is always changing, but so do business needs.
- Leading to: Changing Anything Changes Everything Principle
- This is a big challenge, but MLOps accepts it as given and builds on it.



https://ml-ops.org/content/motivation .

Disentangling to different pipelines



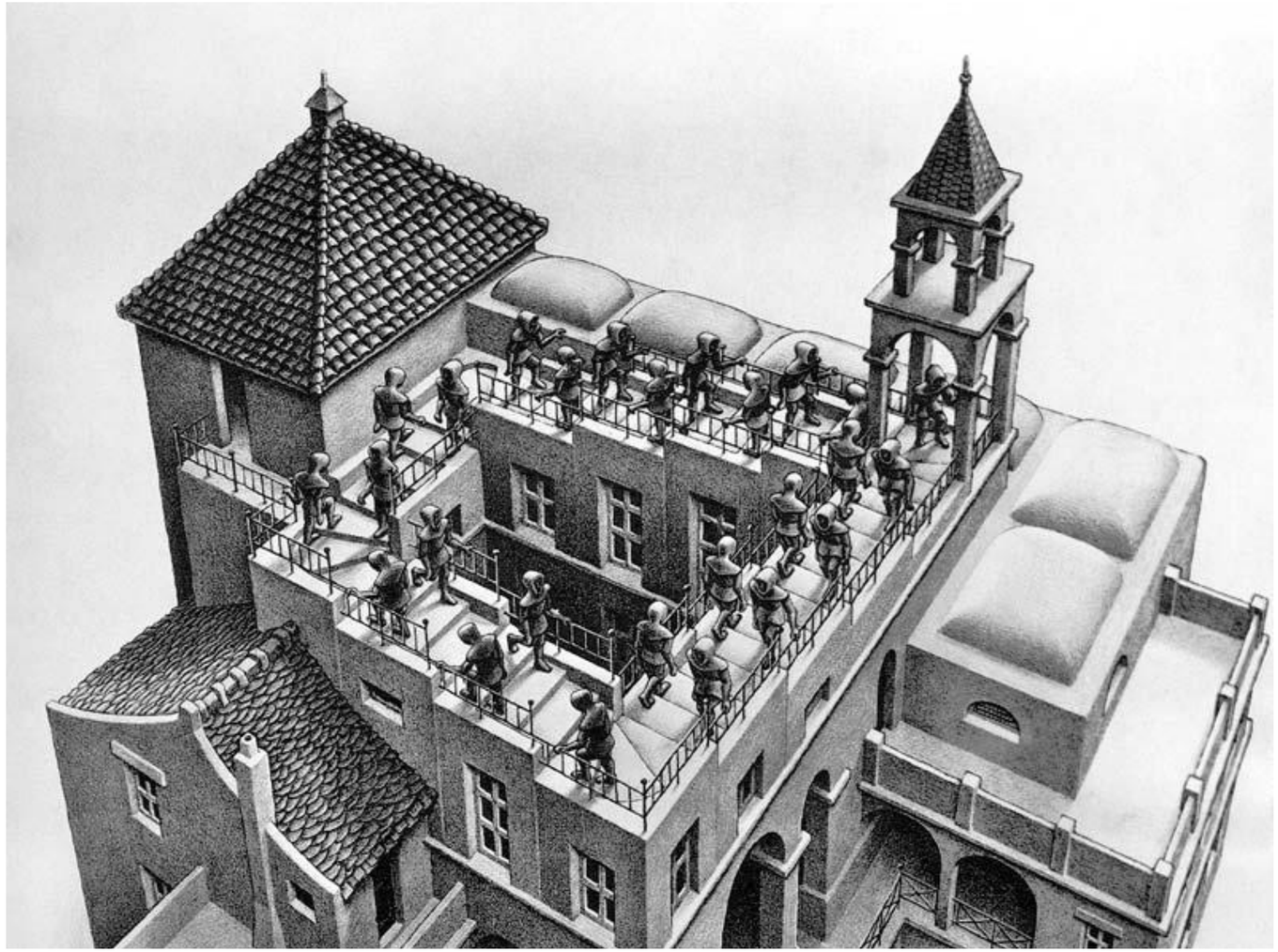
https://ml-ops.org/content/end-to-end-ml-workflow

2

MLOps principles

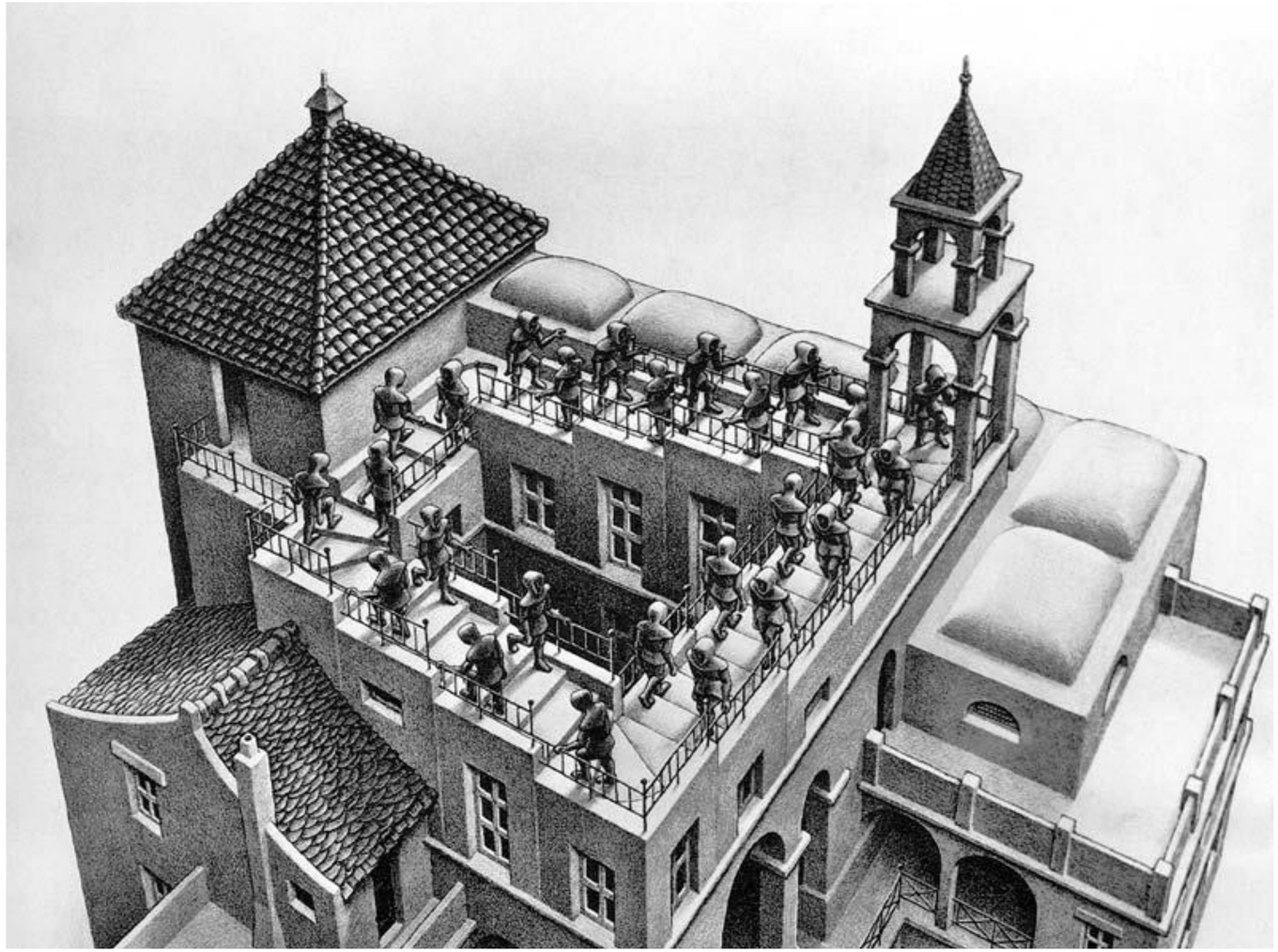
Continuous X

- **Continuous Integration (CI):**
  - frequently merging code by different devs to central repository, after which automated builds and tests are run.
- **Continuous Delivery (CD):**
  - automatically building, testing and deploying your machine learning models



Continuous X

- **Continuous Integration (CI):**
  - frequently merging code by different devs to central repository, after which automated builds and tests are run.
- **Continuous Delivery (CD):**
  - automatically building, testing and deploying your machine learning models
- **Continuous Training (CT):**
  - constantly training different models that can be used to compare and benchmark against each other
- **Continuous Monitoring (CM):**
  - constantly monitoring different aspects of your machine learning systems in produ😊
- **Continuous Documentation**
  - automatically generating and keeping up-to-date documentation of your machine learning models

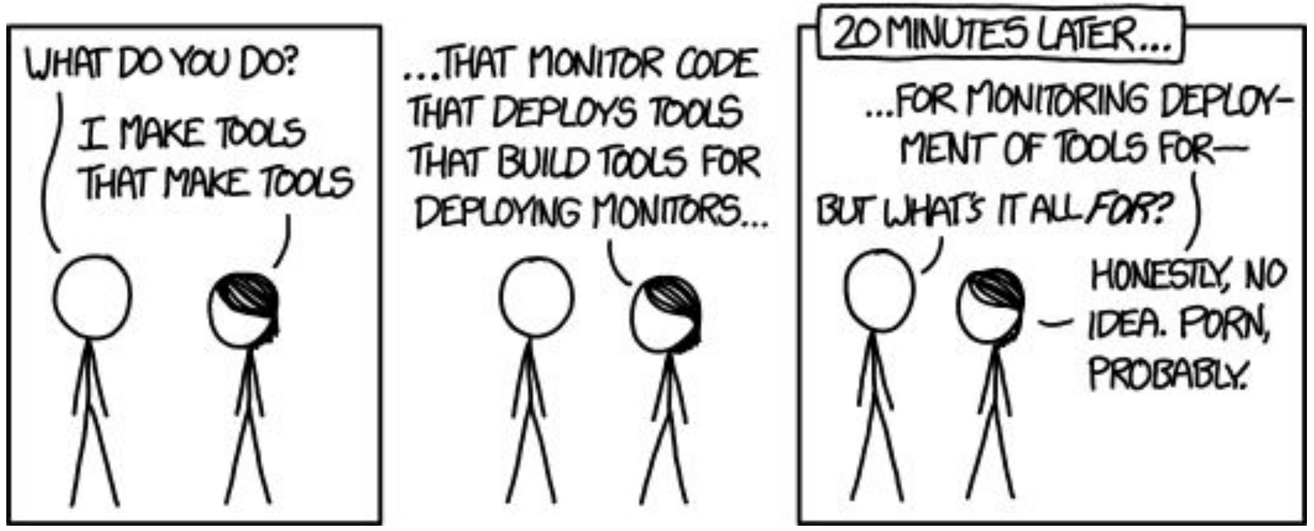


Version everything

- **Artifact:** any tangible by-product produced during development of software.
- Common examples: **Data, code, models, environments, documentation**.
- However, there are many other, and (whenever possible) they should be stored and versioned.
- Storage and versioning of artefacts is the basis for:
  - **Reproducibility:** ability to rerun experiments and get the same results
  - **Traceability:** ability to trace experiments to the code/data that generated them
  - **Collaboration:** enabling multiple users to work with the same artifacts

Automate (whenever allowed)

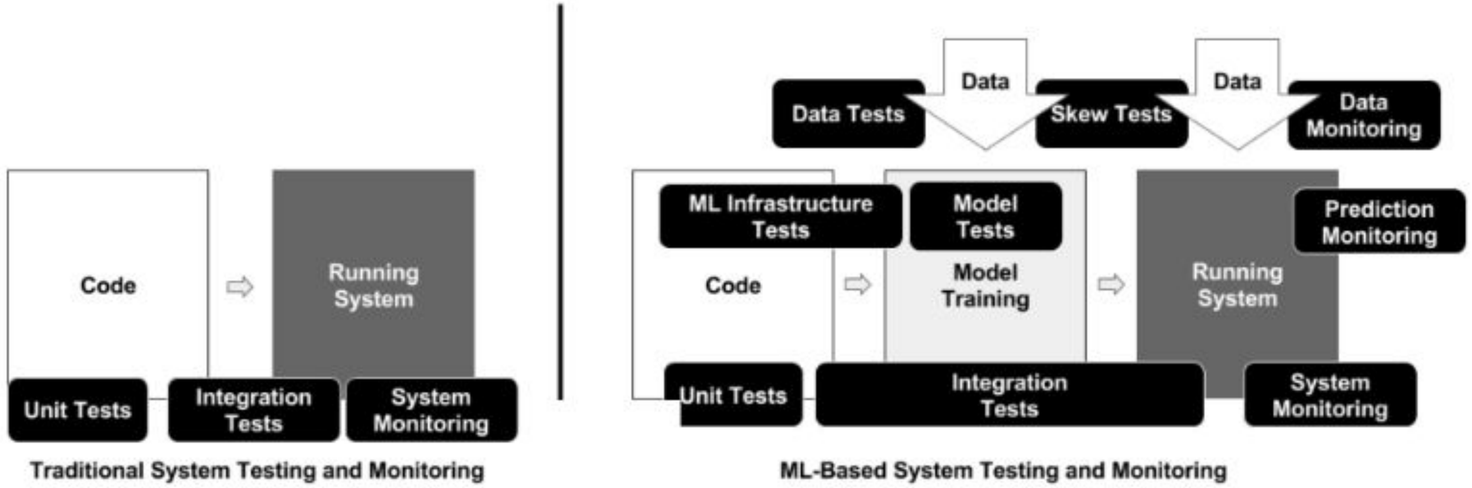
- Automation is key to enable the continuous processes required by MLOps
- It can be **used at every stage** of the MLOps workflow.
- Remove as much manually intensive and error-prone tasks as possible.
- Ideally, **automate everything**, then **reintroduce people** strategically.
- This enables the **teams to focus on adding value**, rather than on maintaining the proces.



https://xkcd.com/

If it isn't tested, it is broken

- Good practices:
  - Enforce **passing of tests** prior to merging code.
  - **Test entire system**, not just the model.
  - **Test different artifacts**, not just code.
  - Use a **variety** of different testing approaches



https://research.google/pubs/pub46555/

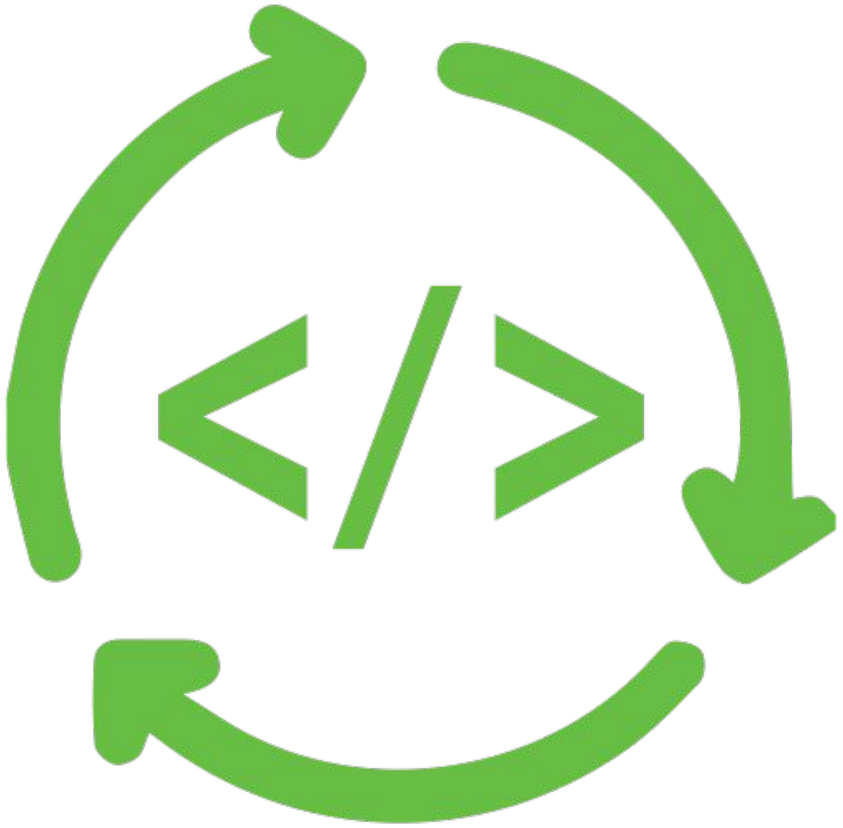
.

Reusability

Ability to use a model or a piece of code for multiple purposes across multiple environments.

Good practices supporting reusability

- **Modularity**: separating a system into independent components.
- **Agnostic and interoperable tools**: avoiding tools that don't play well with others
- **Discoverable code and artifacts:** having clear and consistent way to find what we need



Collaboration and communication

- Good practices:
  - **Frequent communication** between **all roles** involved in MLOps workflow
  - **Transparency**: sharing information about experiments, results, models and deployments with entire team.

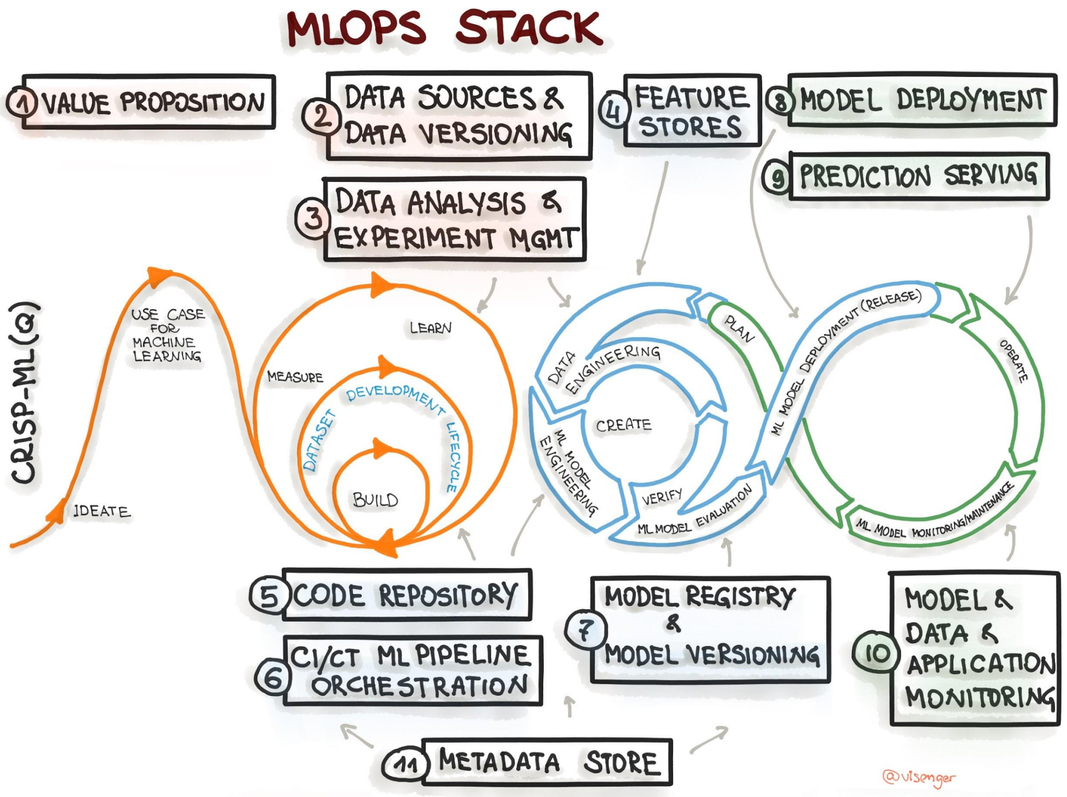
3

MLOps tools

MLOps tools

•DEV TOOLS

- **Code repository:** Central location where code is stored and managed
- **CI/CD pipelines**: Build, test, deploy code changes
- **Model/Data/App monitoring**: Logging, tracing and monitoring of machine learning models, data and applications



https://ml-ops.org/content/motivation

.

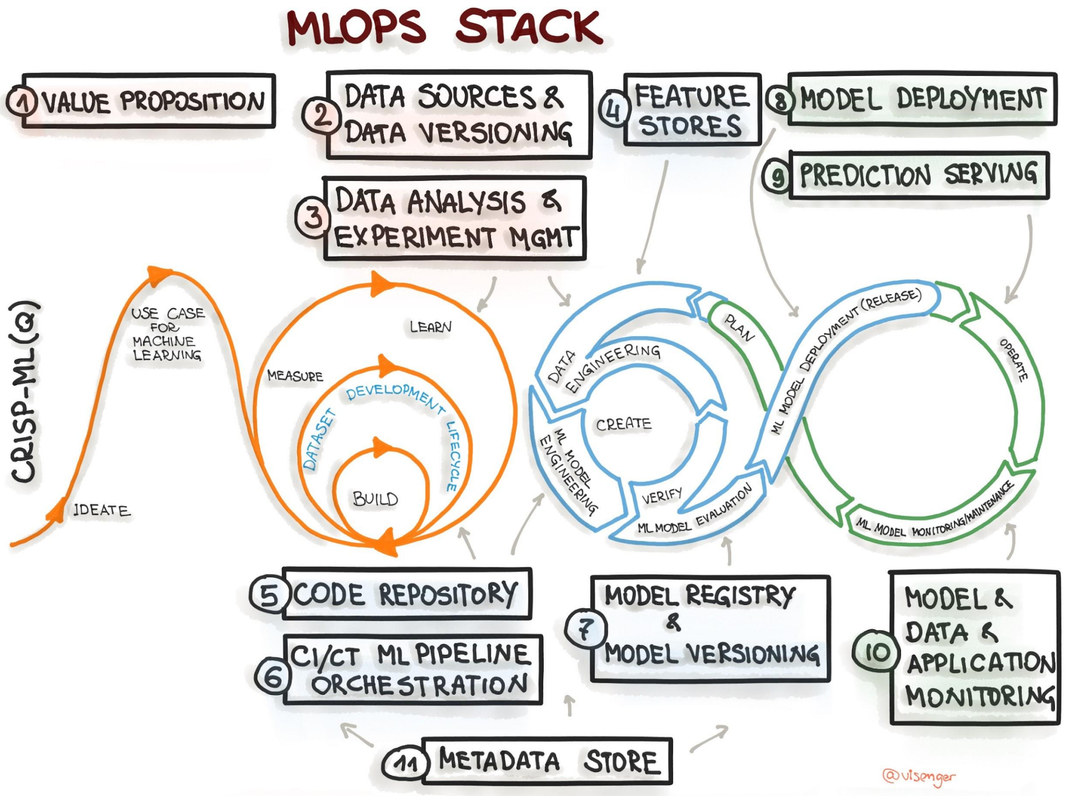
MLOps tools

•DEV TOOLS

- **Code repository:** Central location where code is stored and managed
- **CI/CD pipelines**: Build, test, deploy code changes
- **Model/Data/App monitoring**: Logging, tracing and monitoring of machine learning models, data and applications

•ML TOOLS

- **Data analysis:** Programming languages, IDEs
- **Model training**: Machine learning frameworks
- **Model deployment:** Execution environment to deploy models
- **Prediction serving:** Allows for serving predictions from deployed machine learning models



https://ml-ops.org/content/motivation

.

MLOps tools

•DEV TOOLS

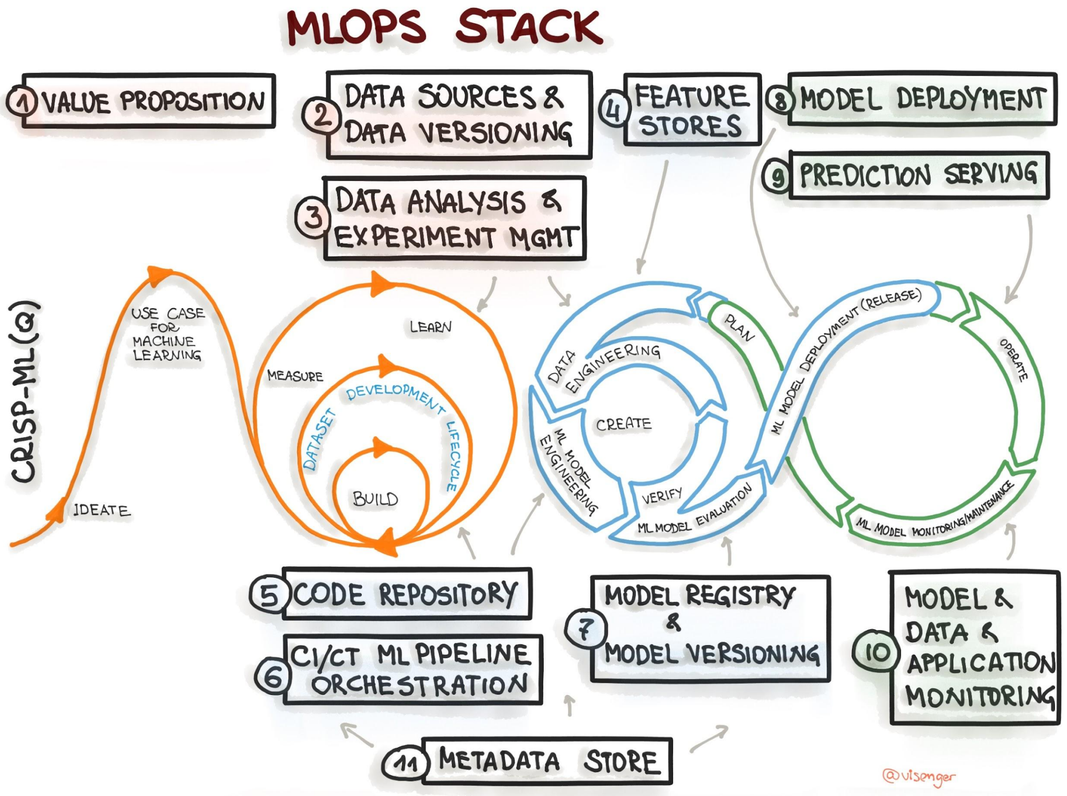
- **Code repository:** Central location where code is stored and managed
- **CI/CD pipelines**: Build, test, deploy code changes
- **Model/Data/App monitoring**: Logging, tracing and monitoring of machine learning models, data and applications

•ML TOOLS

- **Data analysis:** Programming languages, IDEs
- **Model training**: Machine learning frameworks
- **Model deployment:** Execution environment to deploy models
- **Prediction serving:** Allows for serving predictions from deployed machine learning models

•MLOPS SPECIFIC TOOLS

- **Data versioning:** Allows for tracking changes to data.
- **Feature store:** Central repository to store, version and share features
- **Metadata store:** Central repository to store and manage all metadata collected during model lifecycle

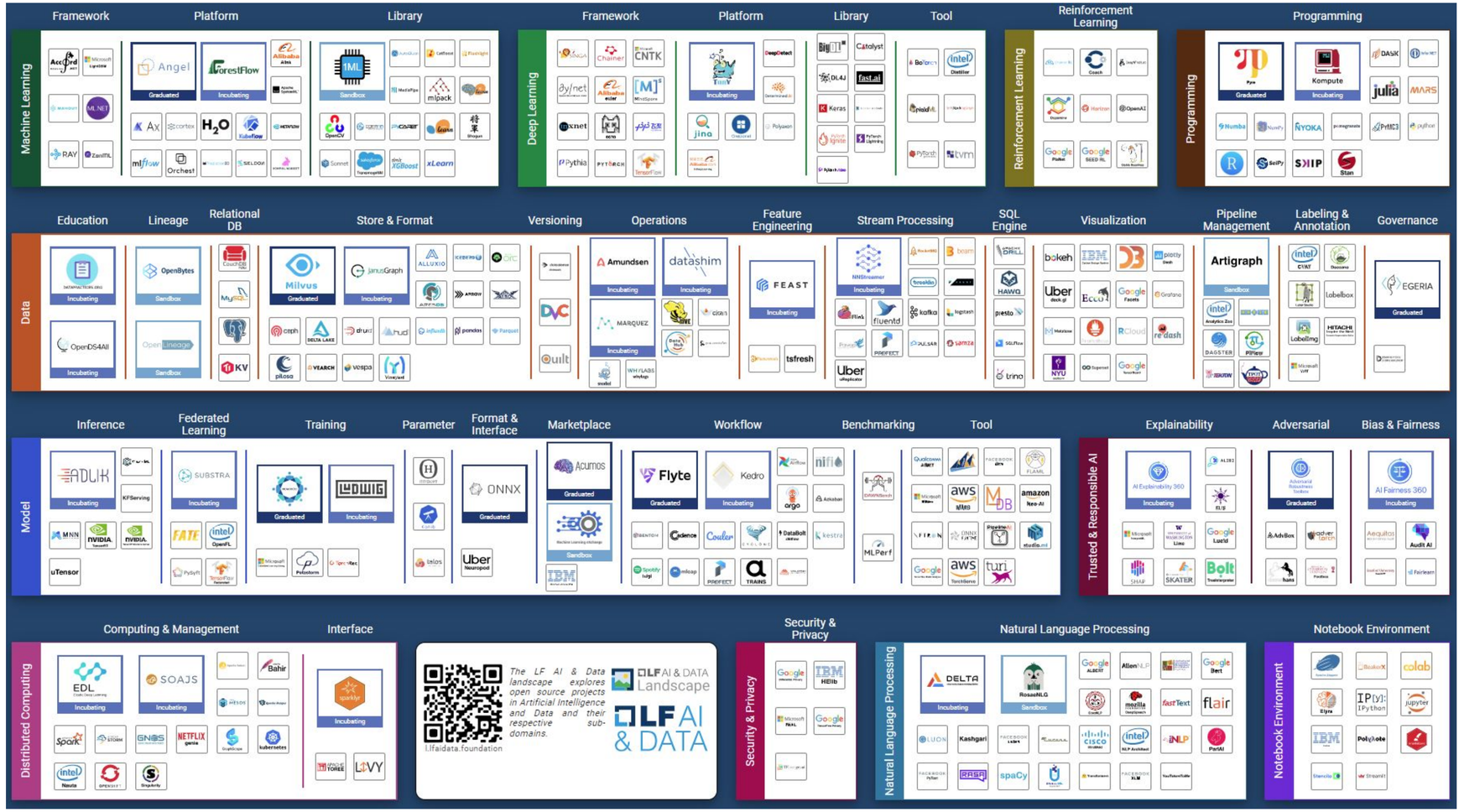


https://ml-ops.org/content/motivation

.

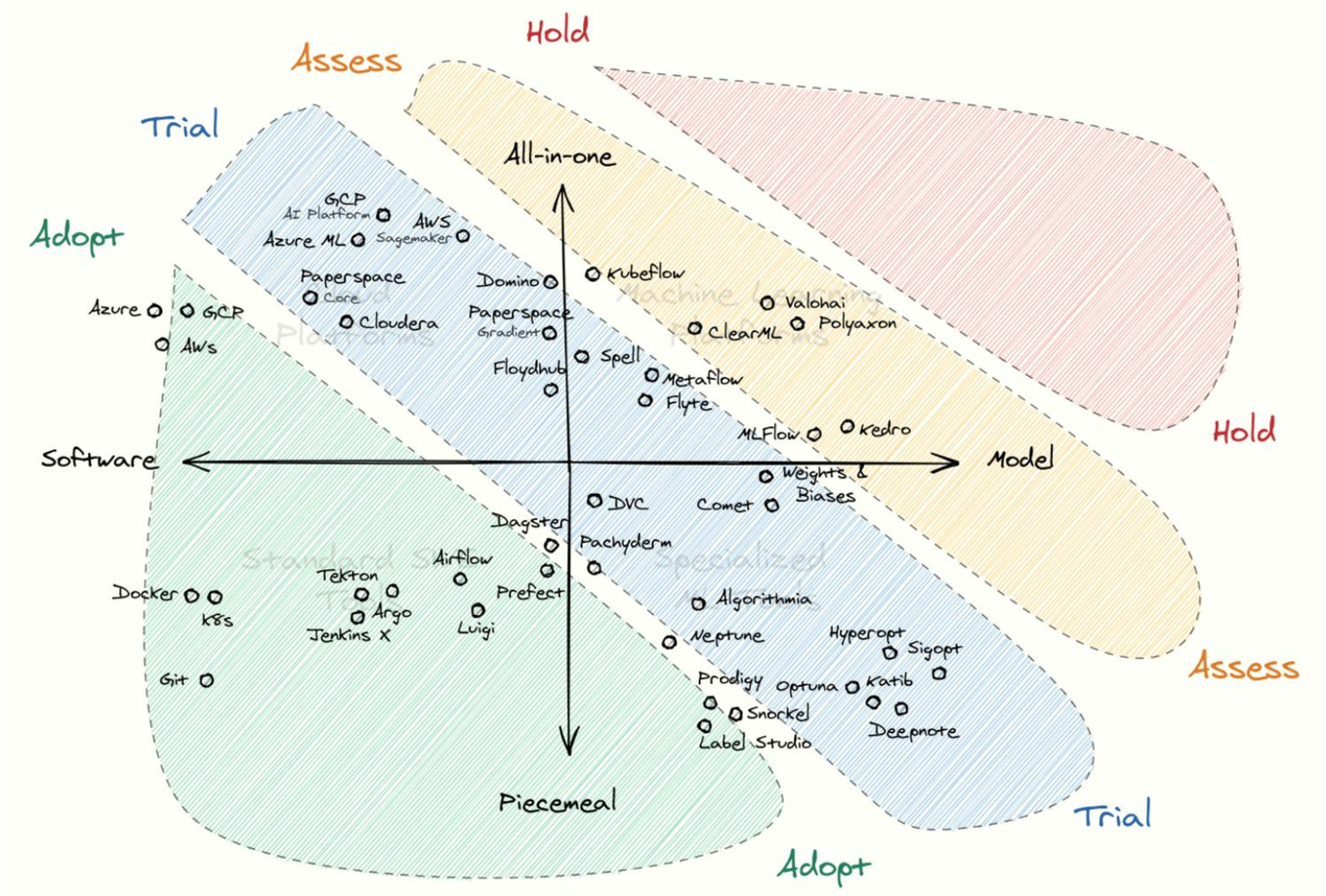
LF AI & Data Landscape

- The landscape of tools that support MLOps is changing constantly.
- A curated interactive landscape available at: https://landscape.lfai.foundation/
- Another good resource to expore tools: https://github.com/EthicalML/awesome-production-machine-learning
- A 1-page MLOps Stack Canvas with questions to help architecting the system: https://miro.com/miroverse/mlops-stack-canvas/



Navigating MLOps tool landscape

- Thoughtworks' Technology Radar
  - **Adopt:** Use or be left behind.
  - **Trial:** Pursue in a low-risk project or environment
  - **Assess:** Explore and understand how it could affect you
  - **Hold:** Don't bother for now, too new to reasonably assess.

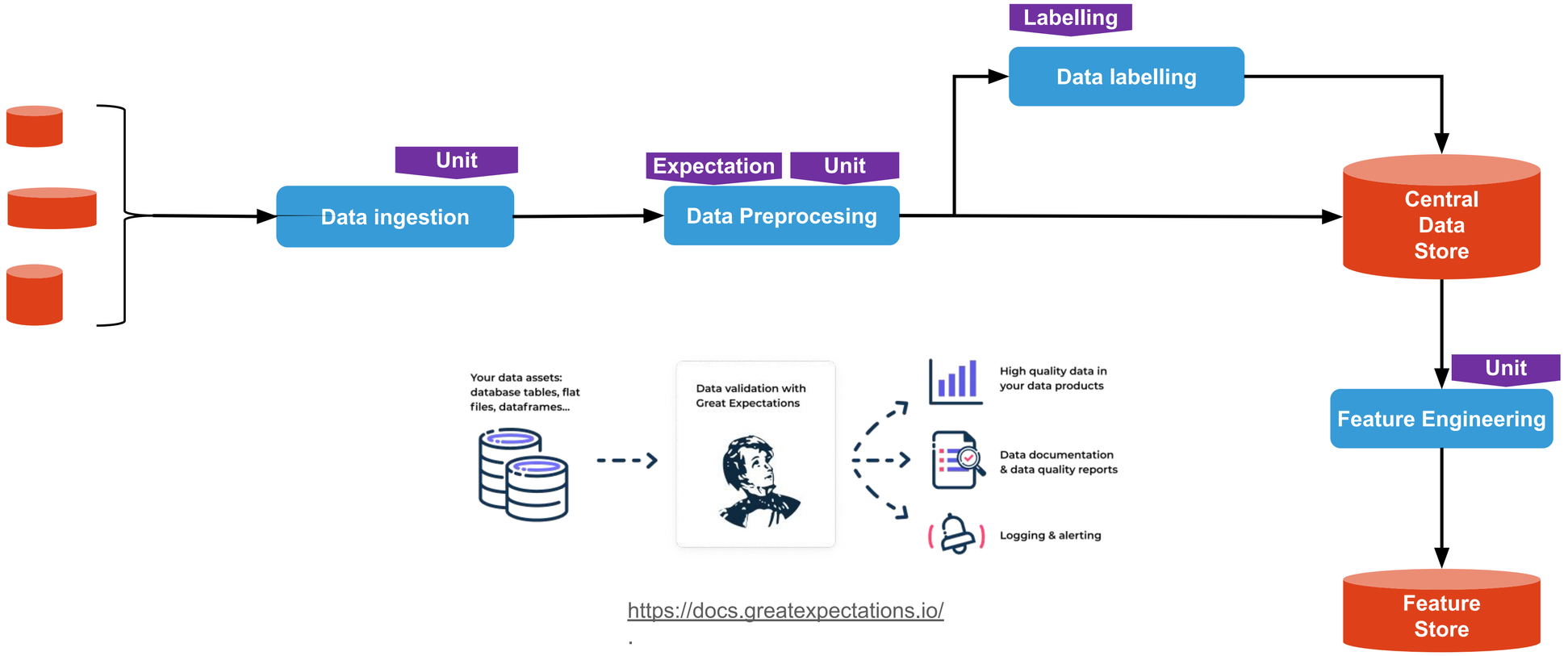


https://ljvmiranda921.github.io/notebook/2021/05/30/navigating-the-mlops-landscape-part-3/ .

4

MLOps workflows

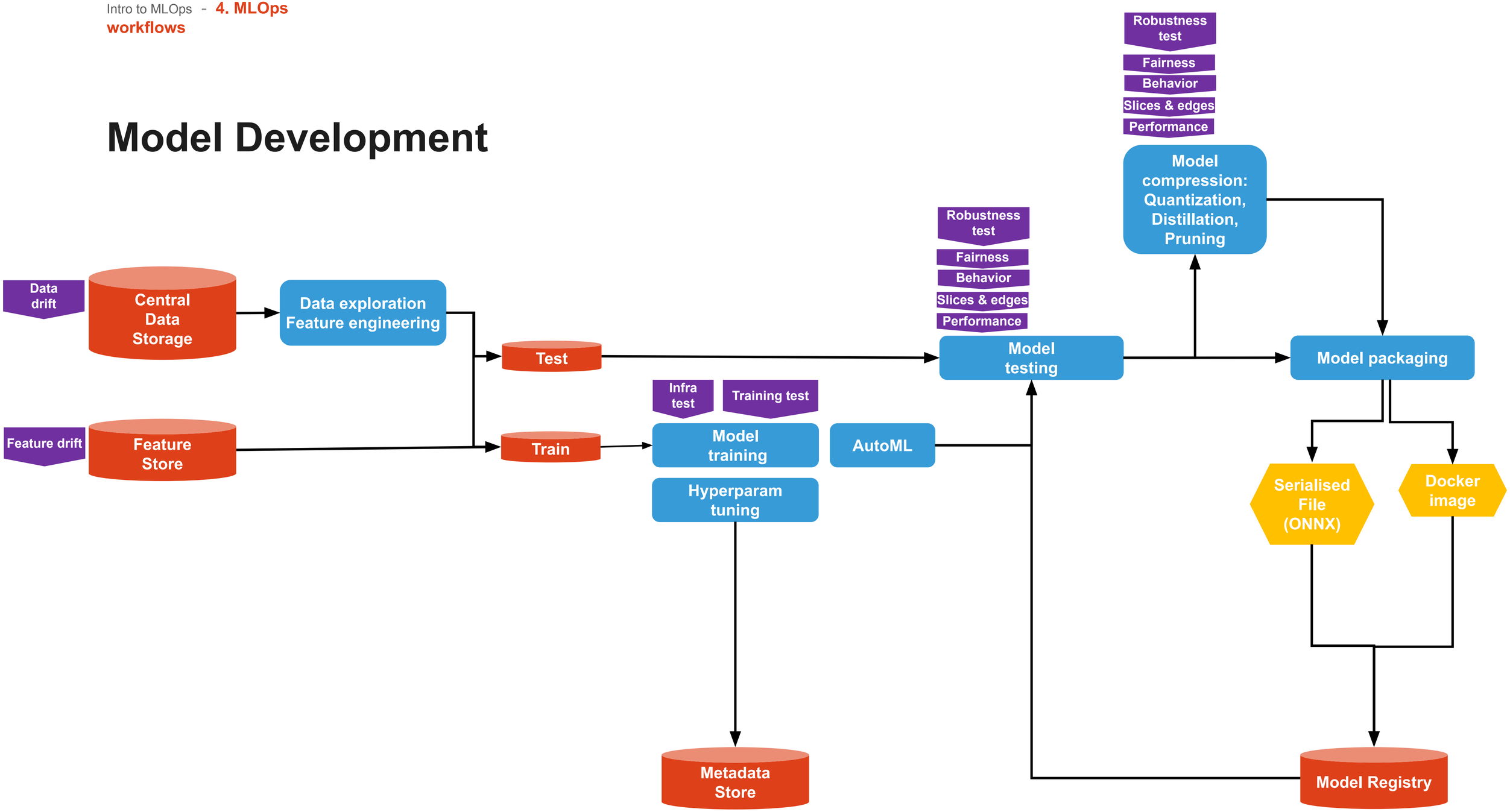
Data management



https://docs.greatexpectations.io/

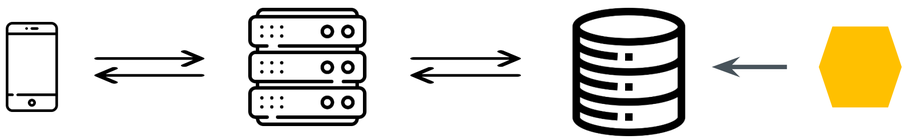
.

Model Development



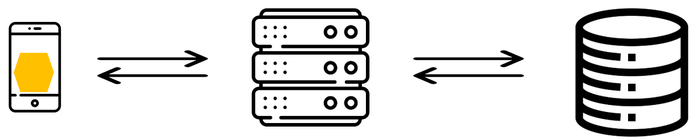
Choosing a model deployment strategy

Batch prediction (periodically run model, cache results in database)



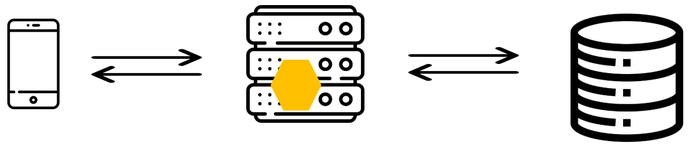
+ simple
+ low latency
- scalability
- staleness

Edge prediction (deploy model on end user device)



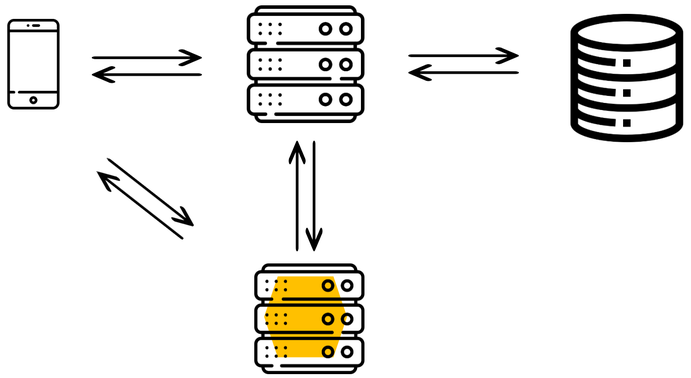
+ low latency
+ offline
+ privacy
- resources
- frameworks
- updates
- monitoring

Model-in-service (deploy model on application server)



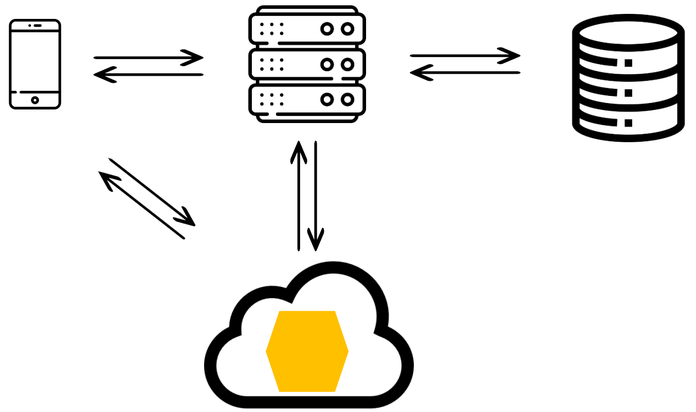
+ infra reuse
- compatibility (code, hardware, scale)
- resources

Model-as-service (deploy model as its own separate service)



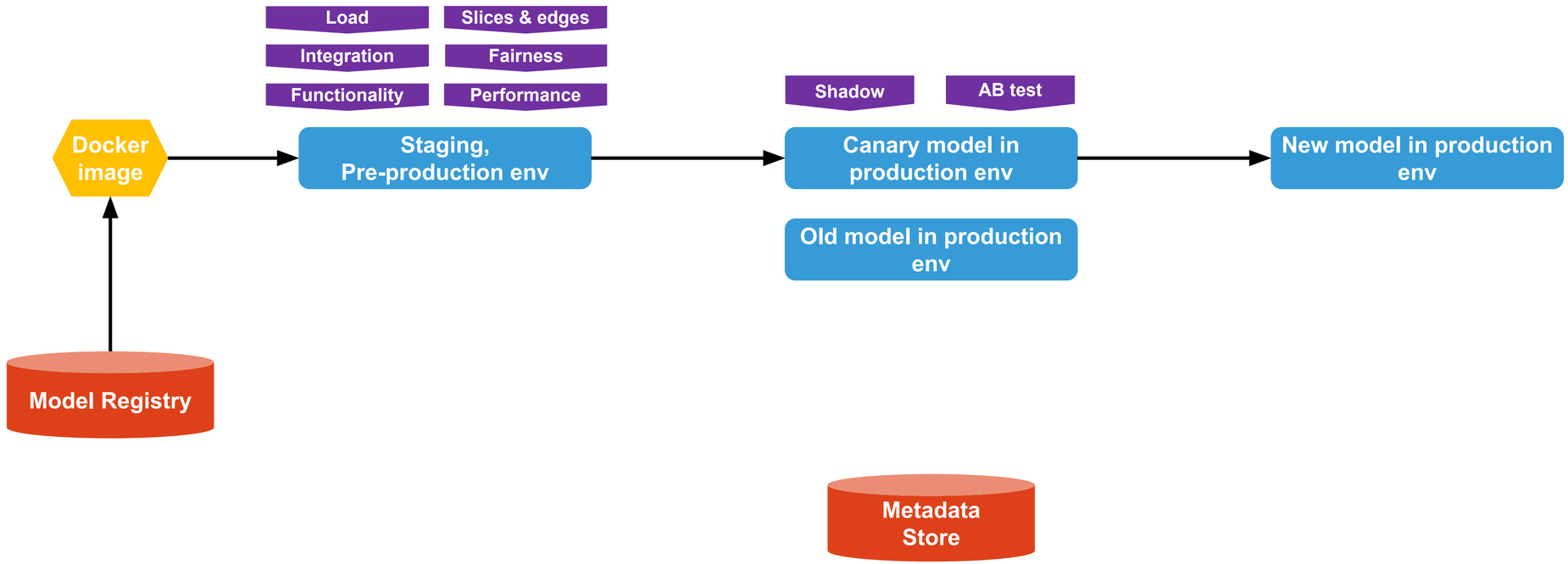
+ dependable
+ scalable
+ flexible
- more latency
- more infra
- More DevOps work

Model as a serverless function

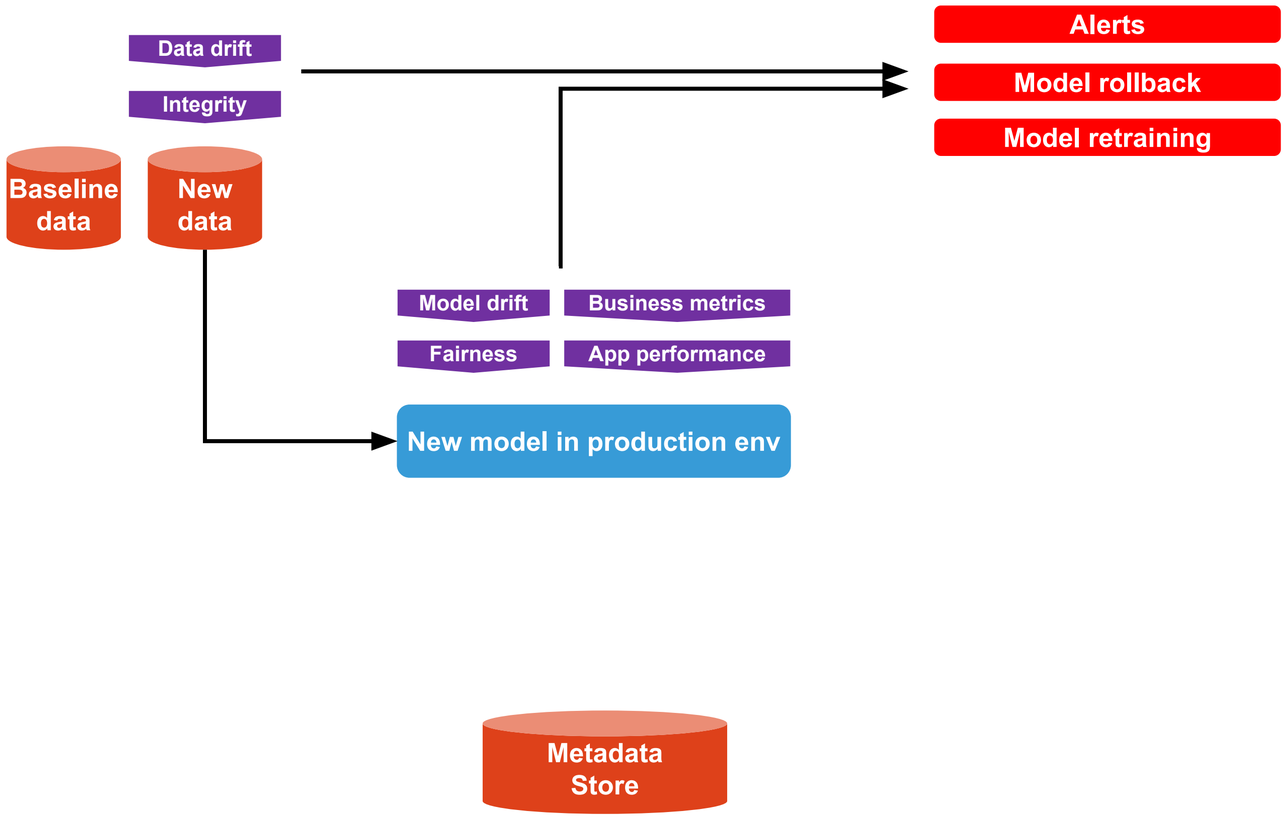


+ only pay for compute
+ no infra
- deployment package size
- generic servers limits

Deploying to Production and Serving

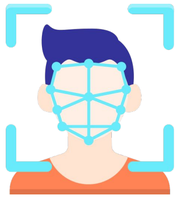


Monitoring and feedback loops



An extra bit of motivation for the end

High-risk use cases of AI



Biometric identification and categorisation
(whenever it is not prohibited)

Management and operation of critical infrastructure

- Examples: safety components in management and operation of utilities, traffic

Education and vocational training

- Examples: AI systems for assessing students, assigning people to trainings, AI systems that impact personal development of children

Employment and worker management

- Examples: AI systems for recruitment, assessment of employees

Access to essential services

- Examples: AI systems for credit scoring, assessment of eligibility for public benefits, emergency services response priority, patient triage

Law enforcement

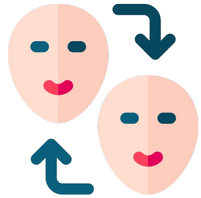
- Examples: AI systems for evidence evaluation, detection of fraudulent content, crime analytics

Border control management

- Examples: AI systems for management of borders, migration, asylum processes

Administration of justice and democratic processes

- Examples: AI systems to assist judges, to influence voters, count voting ballots



Creation of (some) AI generated content

- Examples: AI systems to generate complex text such as news, deep fakes of people.


endava