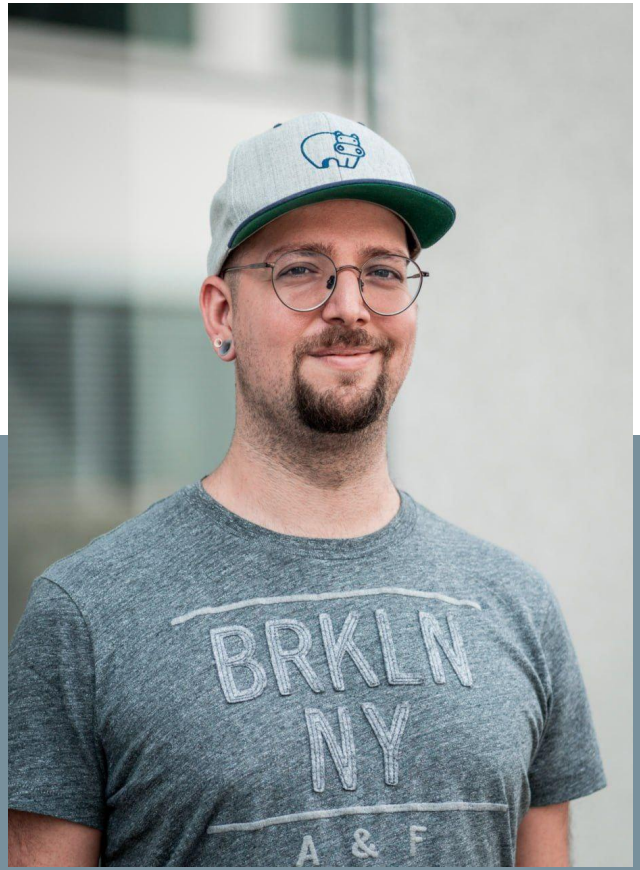
Fabrizio Lazzaretti & Marco Crisafulli

DevOps für KI: Mehr als nur Modelle

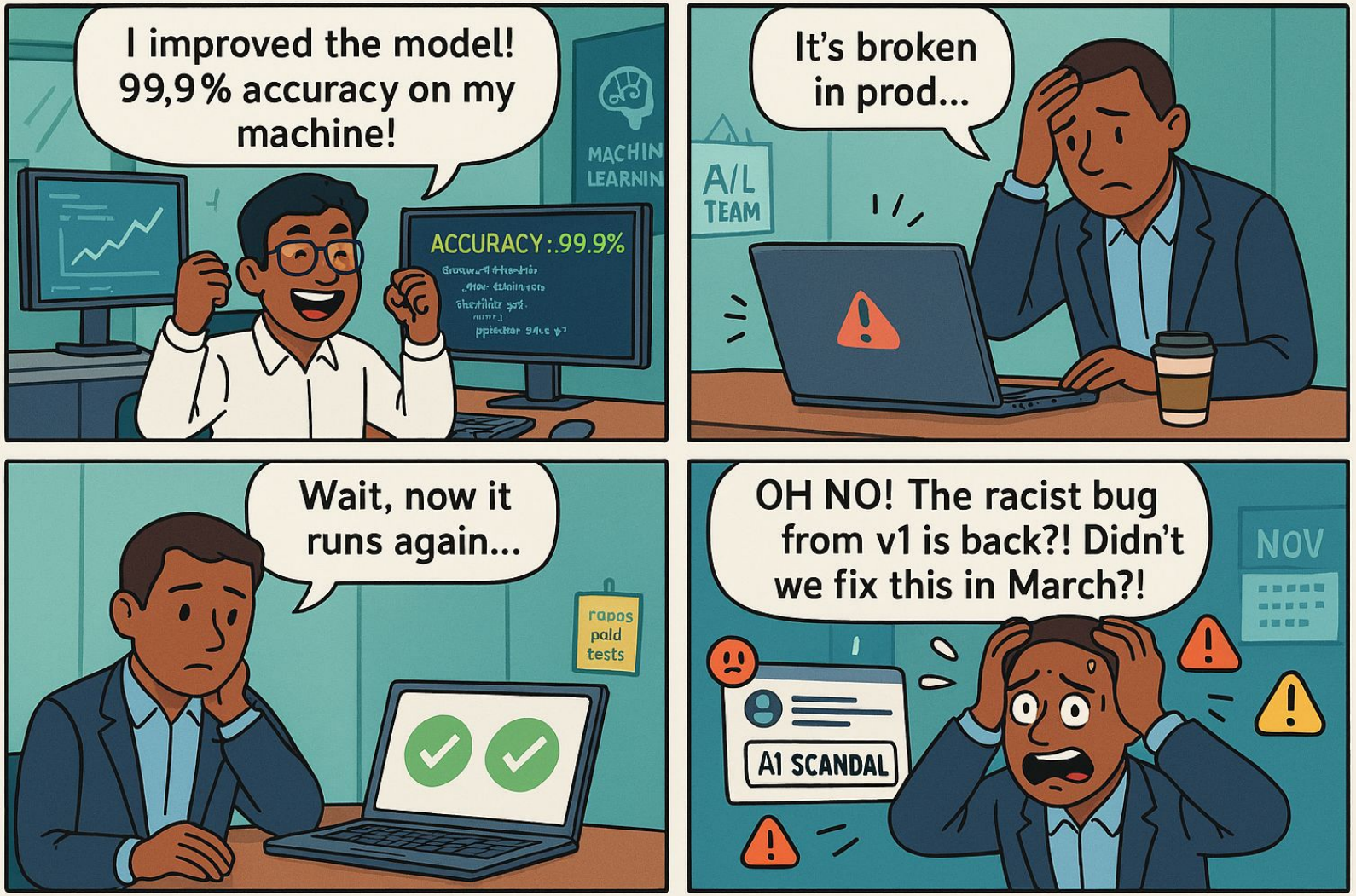
Fabrizio Lazzaretti
Managing Consultant @ Wavestone |
CNCF Ambassador | Speaker | Author

Helping enterprises adopt cloud-native technologies while ensuring future-proof architectures and business-IT alignment



Marco Crisafulli
Co-Founder @ enki |
ML Engineer | Father | Sometimes funny

AI can solve real problems. Helping companies to understand and utilize this technology.


I improved the model! 99,9% accuracy on my machine!
ACCURACY:.99.9%
It's broken in prod...
A/L TEAM
Wait, now it runs again...
rapos pald tests
OH NO! The racist bug from v1 is back?! Didn't we fix this in March?!
NOV
A1 SCANDAL

IT-Tage

Gartner predicts … 85% of AI projects will deliver erroneous outcomes due to bias in data, algorithms or the teams responsible for managing them.

Gartner

Zillow: [CEO] made reference to the company's lack of confidence in its home buying algorithm's ability to accurately predict fluctuations in home prices. … leading to a write-down of over $300 million

GeekWire

Google Flu Trends performing well for two to three years and then failing significantly and requiring substantial revision

Wired

Why is this needed?

“The main challenges people face when developing ML capabilities are scale, version control, model reproducibility, and aligning stakeholders”

2020 State of Enterprise Machine Learning Report

From problem to solution

1. We will not solve all your AI problems in this talk
2. You will not be better than the big ones, but you can learn from it

Our goal for today:

- Understand your business, don’t make wrong expectations and let the business understand the IT
- Reproducibility, Testing, Continuous Deployment with practical implementation in Kubeflow

IT-Tage

Before we start…

WAVESTONE

enki

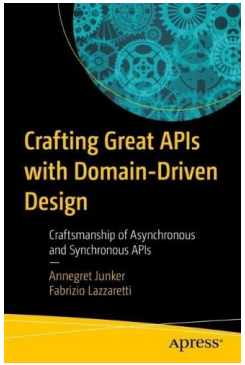
We need AI!

- What business problem do we want to solve?
- Which KPIs can we define to know if the project was successful?
- Do we have the necessary data, know-how and laws to be able to do this?
- Does the problem benefits from applying AI?

How can we find out if we need this AI project?

- Qualify business ideas
  Understand what we want to do, and what to prioritize
  - Business Model Canvas
  - Capability Map
  - Wardley Map
- Gathering business requirements
  Understand how we need to do it and get a common language and understanding
  - Domain Storytelling
  - Visual Glossary
  - Event Storming
  - Context Map

- Tactical design of the components
  Define how we structure the solution
  - Bounded Context Canvas
  - Architecture Communication Canvas
  - API Product Canvas



Getting started with the project

The 4 phases

PoC
Minimal Concept

MVP
Concept
Data Quality
Compliance
Minimal SLAs

Product Strategy
Concept
Data Quality
Compliance
SLAs
Evolution

Scaled Approach	
Concept	Governance
Data Quality	Disaster Recovery
Compliance	Cost Optimization
SLAs	Training Plan
Evolution	

Phase 1: PoC 🥰

We have uncertainty that we want to mitigate before doing an MVP:

- What do we want to prove?
- When did we prove it?
- How do we measure success?

=> Multiple PoCs can be needed, to prove different things

=> If everything is clear, no PoC is needed

Phase 2: MVP 🕵️

PoC can be simple, but what do we need to know to go to prod?

- Budget
  - Initial costs
  - Operational costs
  - Iterative extension/update costs
- When is the project successful?
  - How will we define quality?
  - How do we measure KPIs and see if we achieve them?
- Architecture
  - Loosely coupled
  - Good APIs between components
  - Service cuts
- SLA
  - What happens if it does not work?
  - How fast do we need to recover after a problem?
- Security & Compliance
- Security risk, business risk, risk acceptance, …
  - What happens if we do something wrong?
  - What happens if data get lost or stolen?

IT-Tage

From concept to implementation

WAVESTONE

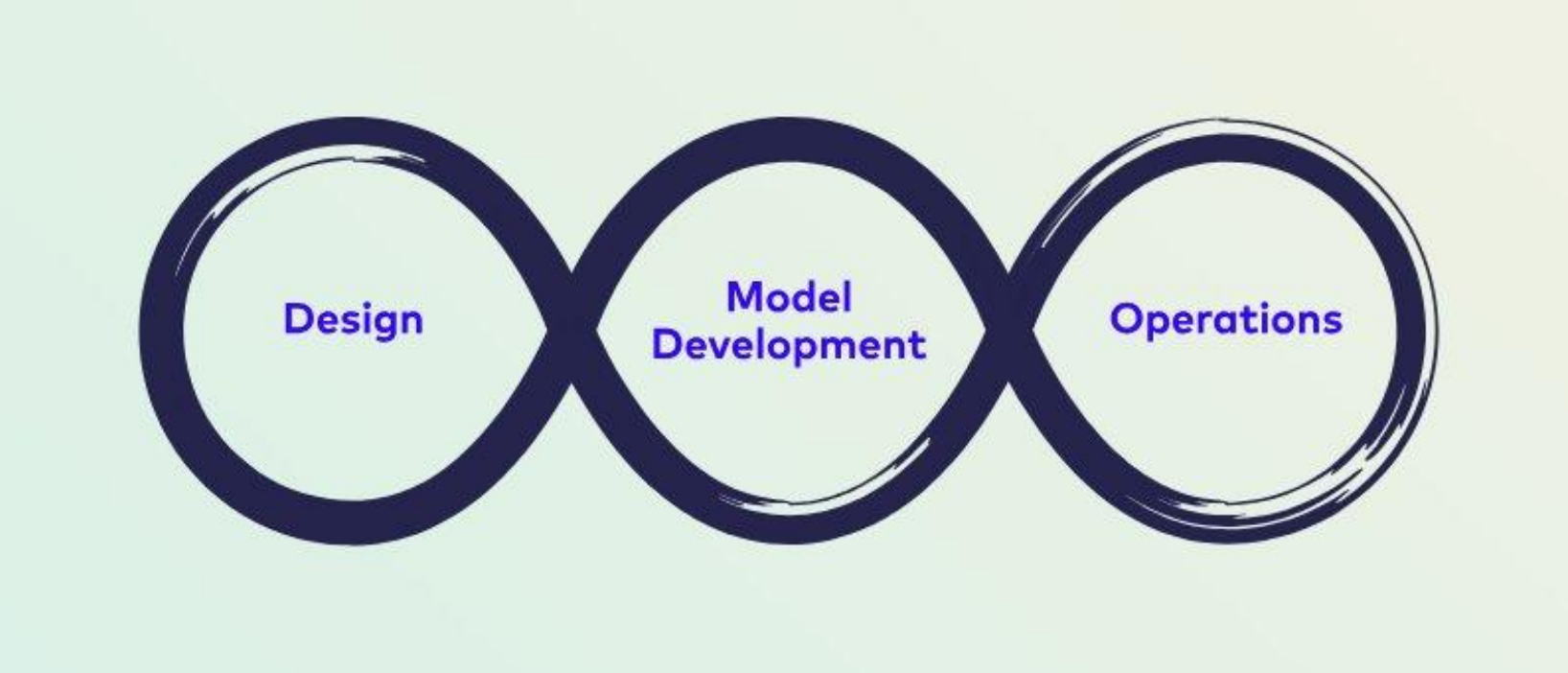
enki

End-to-end ML, right?

Load Data

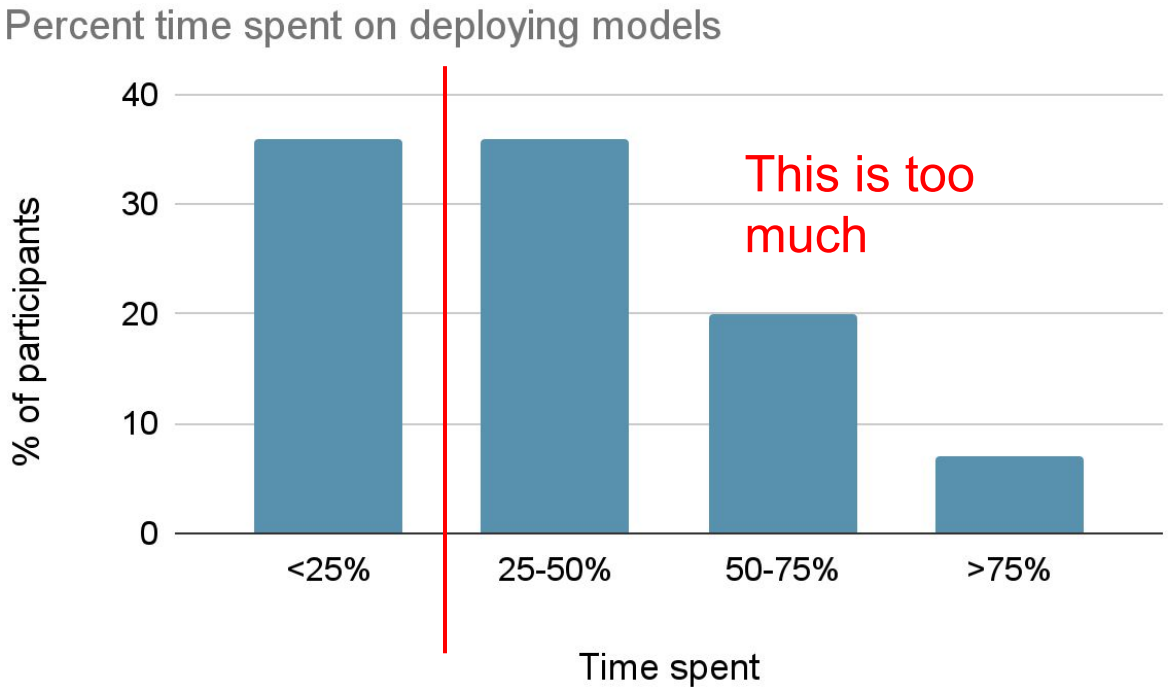
Magic (training)

MLOps - Machine Learning Operations



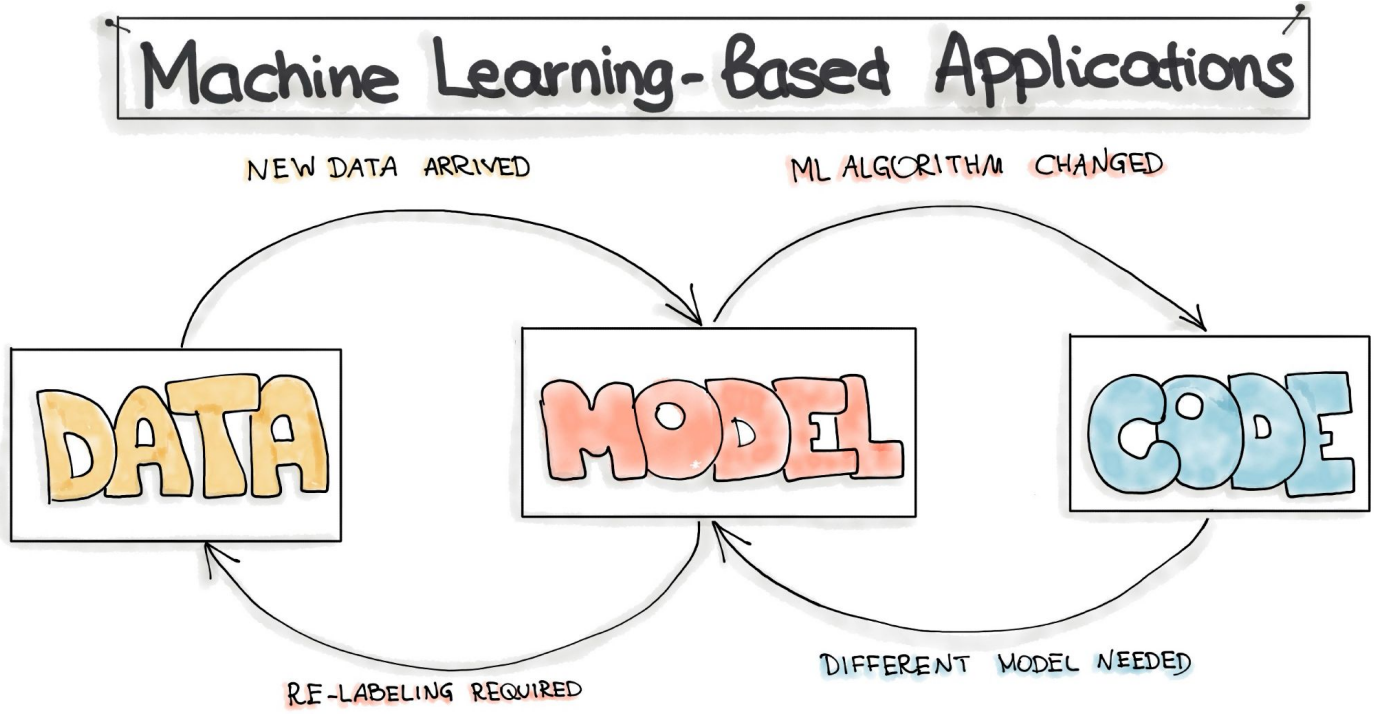
ImgSrc: https://ml-ops.org/

Why is this needed?



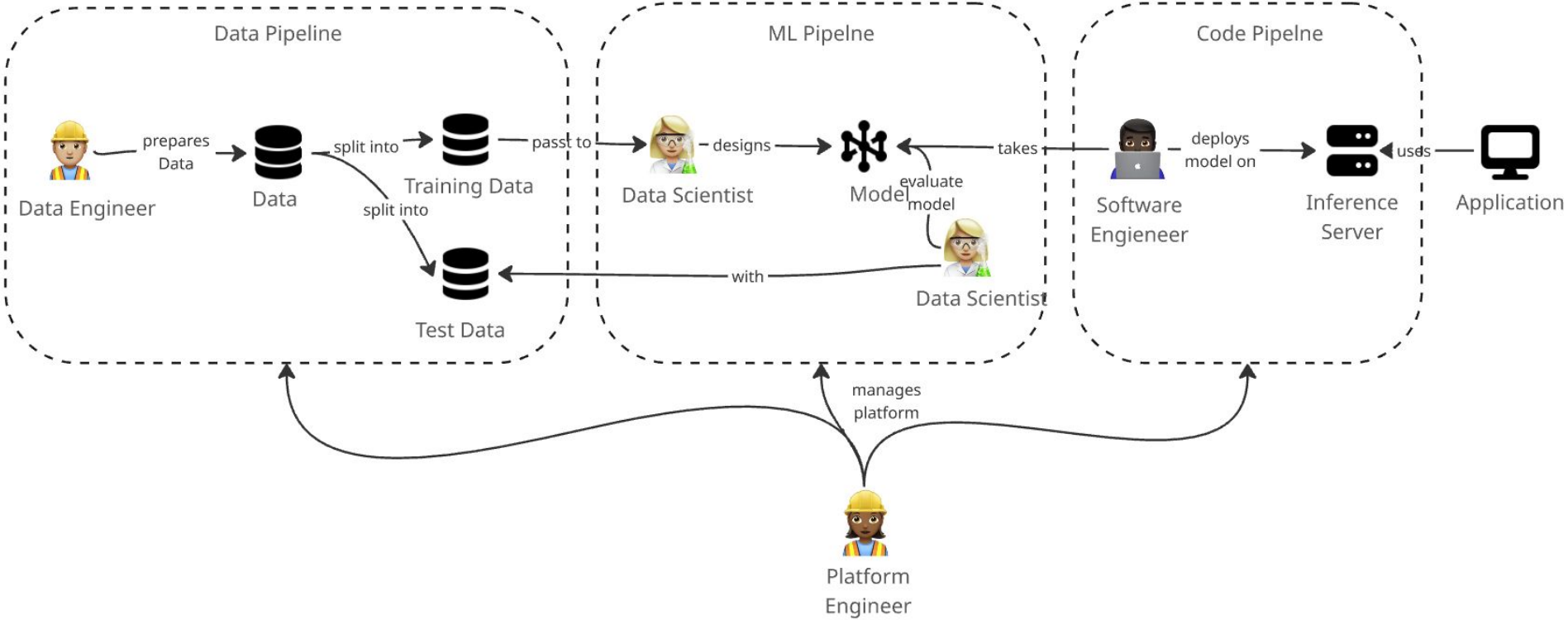
Src: 2020 State of Enterprise Machine Learning Report

MLOps is not DevOps



ImgSrc: https://ml-ops.org

MLOps in detail



IT-Tage

Applying MLOps with Kubeflow

WAVESTONE

enki

Kubeflow is a Cloud Native Computing Foundation project

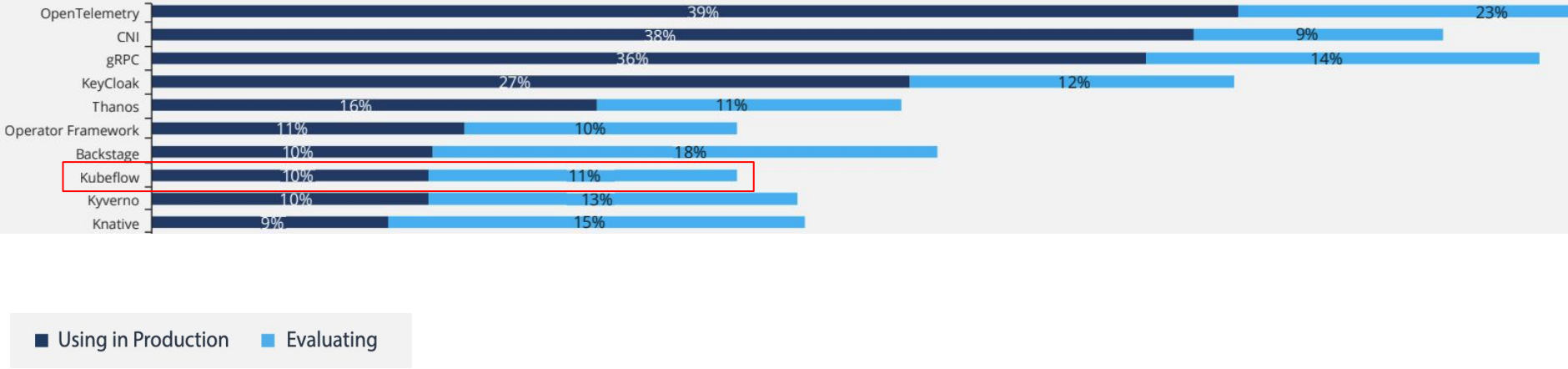
- CNCF is a foundation is a Linux Foundation project
- Enable open-source projects in the cloud & adds governance
- Founded in 2015 with the announcement of Kubernetes 1.0 (donation from Google, before called Borg)
- The big projects: Kubernetes, Prometheus, Envoy, ArgoCD

Kubeflow an incubator project with traction

FIGURE 14

CNCF INCUBATED PROJECTS IN USE OR IN EVALUATION

Which of these incubating CNCF projects is your organization using in production or evaluating?



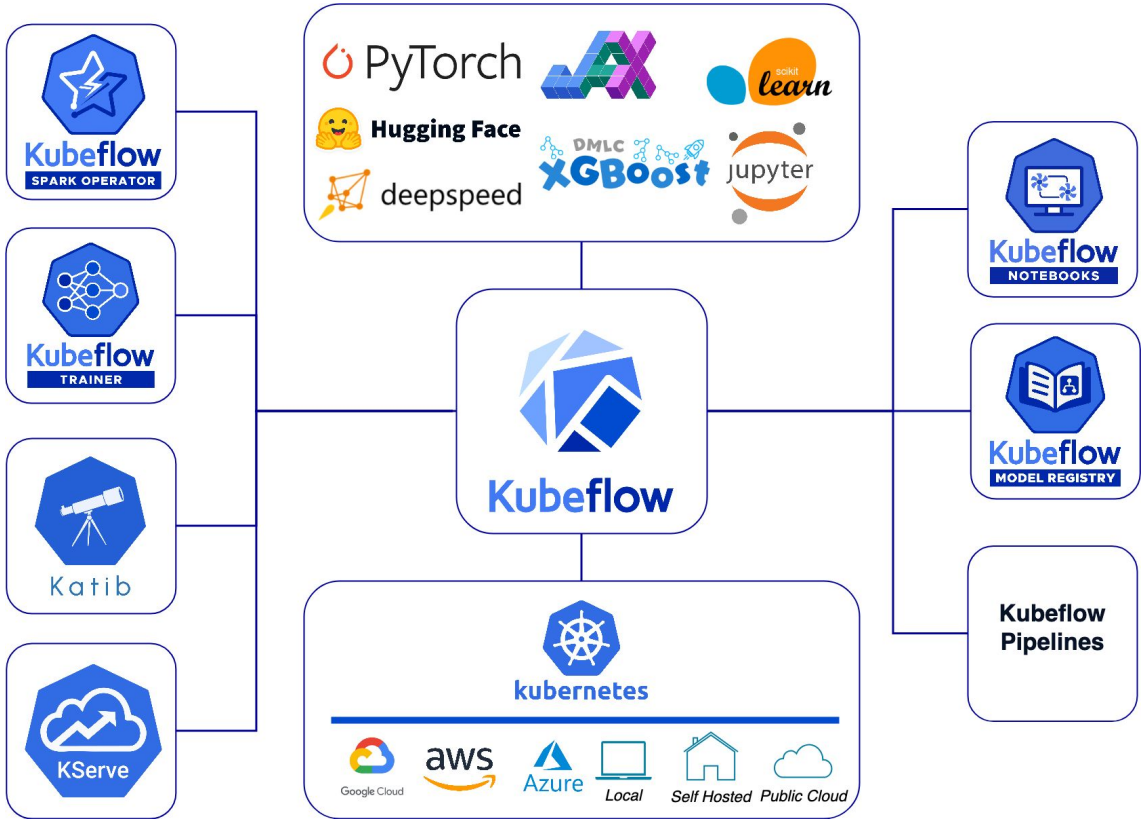
Src: https://www.cncf.io/wp-content/uploads/2025/04/cncf_annual_survey24_031225a.pdf

Who uses Kubeflow?

The Kubeflow user survey [2023] drew responses from 90 members of the community mostly made up of members from the United States (43%), Europe (34%), and Asia-Pacific (10%).

The majority of the respondents were from the Tech industry (49%), followed by Finance (13%) and Consulting (11%).

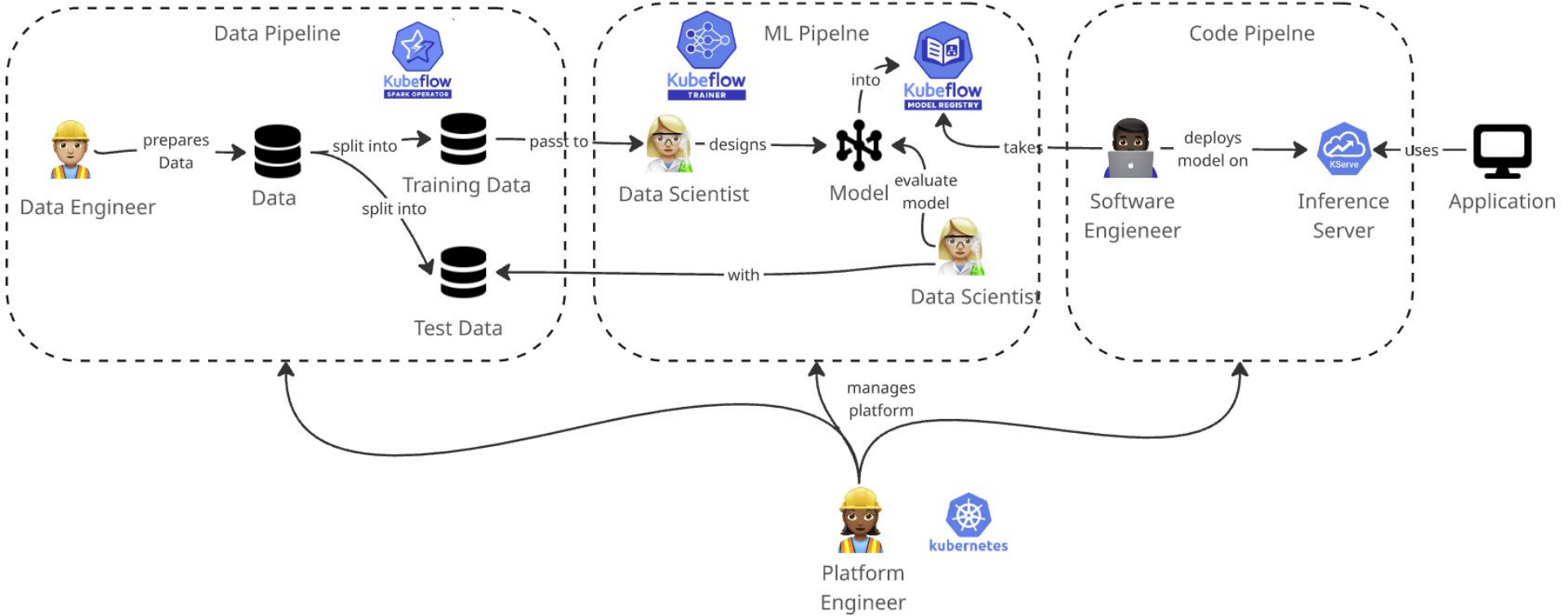
Src:https://blog.kubeflow.org/kubeflow-user-survey-2023/



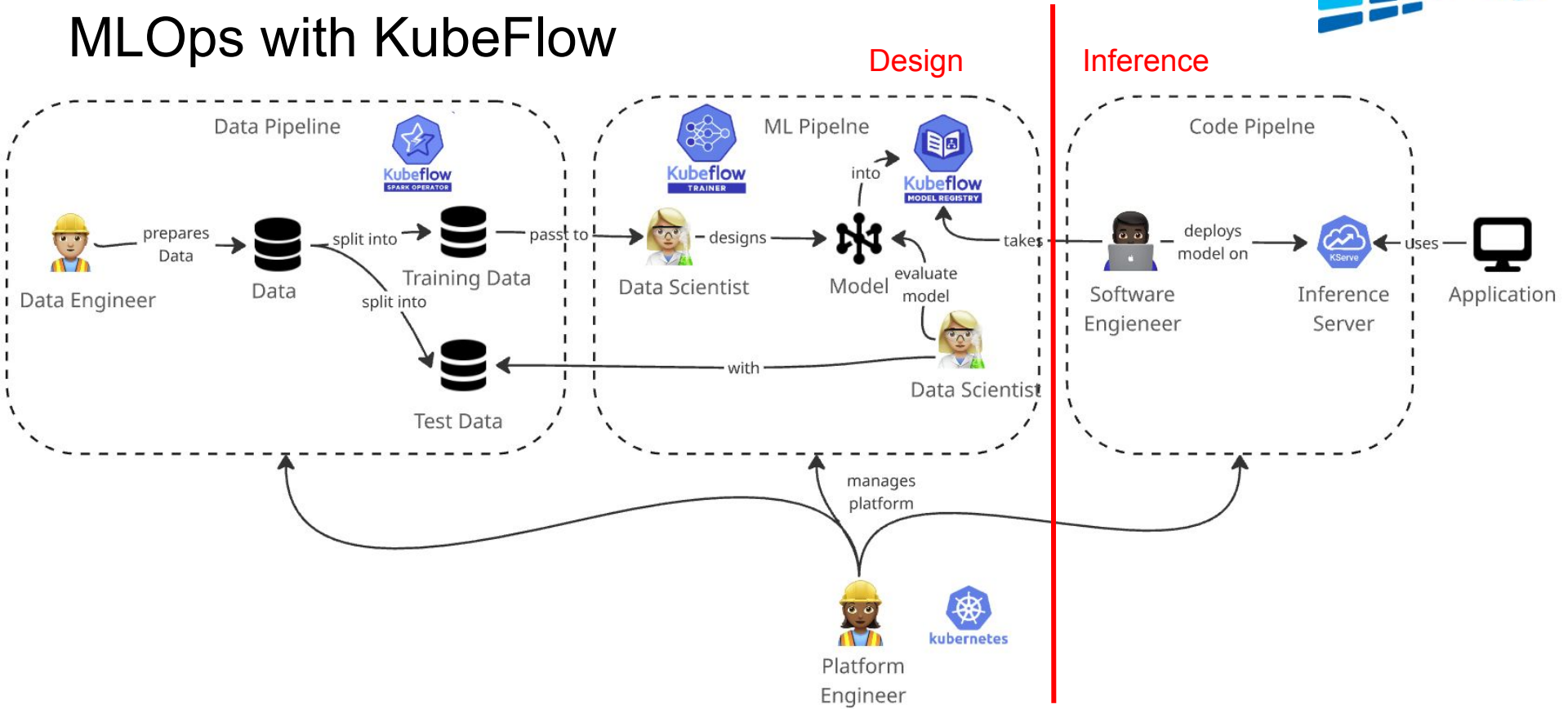
ImgSrc: https://www.kubeflow.org/docs/started/introduction/

Kubeflow and an own model

MLOps with Kubeflow



MLOps with KubeFlow



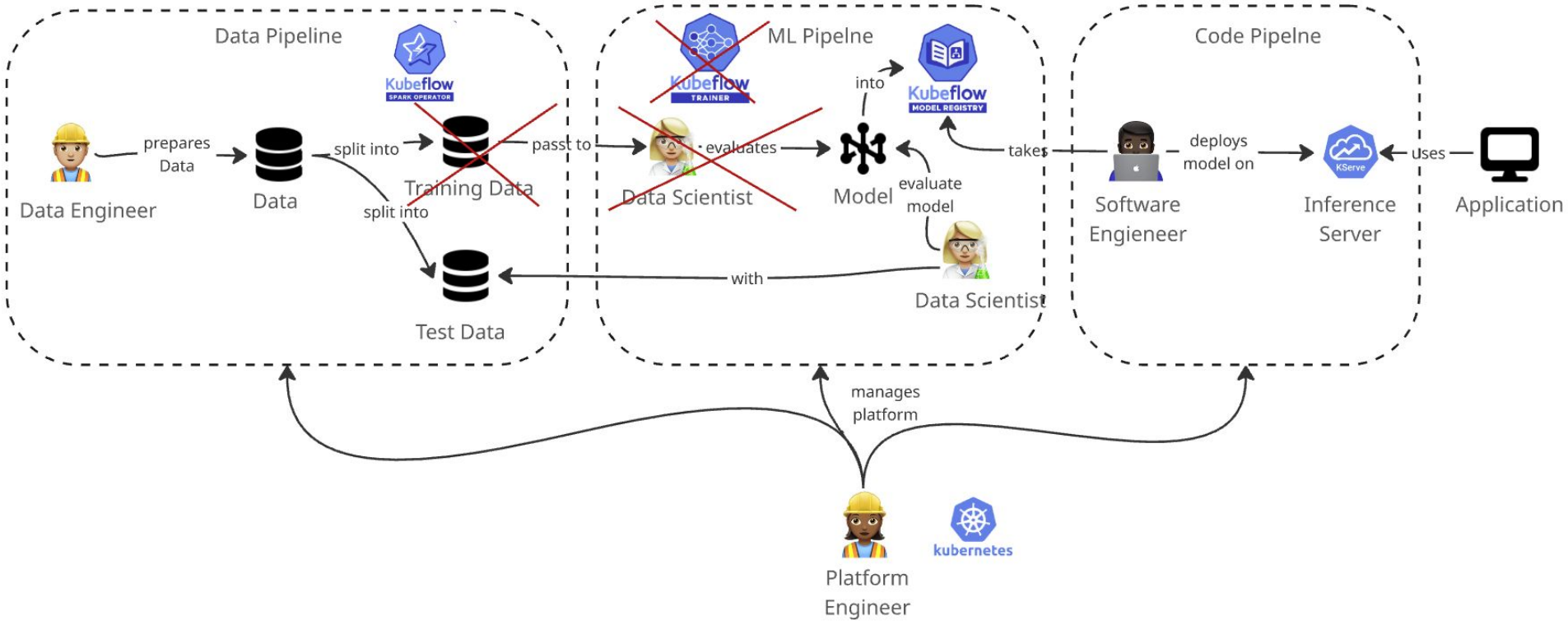
IT-Tage

I want an off the shelf model (ChatGPT, ..)

WAVESTONE

enki

MLOps with off the shelf models



IT-Tage

When to use Kubeflow?

WAVESTONE

enki

Pros and cons

Pros	Cons
Cloud-native, Kubernetes-integrated	High operational complexity
Highly scalable & production-ready	Steep learning curve
Modular, extensible components	Difficult upgrades/maintenance
Vendor-neutral, open source	Heavy infrastructure overhead
Works across clouds & on-prem	Slow for rapid prototyping
Strong community & ecosystem	Documentation needs improvement

Lessons learned

- Start small
  - Kubeflow has a lot of components -> Start with one
  - Attention regarding over-engineering
- Own cluster
  - Managed if possible
- Manage Kubernetes Manifests -> DevOps
  - e.g kustomize
- It is not a magic platform
  - You still need to think about the process

Conclusion

Don’t throw out what you learned from classical IT projects

- Requirement engineering
- Phases
- KPIs

At the end people need to talk!
DevOps + MLOps + Any-Ops

Always evaluate your results, outcome can change dramatically after small changes!

Follow up

Wavestone: Beyond traditional change management whitepaper:
https://www.wavestone.com/en/insight/beyond-traditional-change-management/

Kubeflow Introduction:
https://www.kubeflow.org/docs/started/introduction/

Kubeflow training from CNCF:
https://trainingportal.linuxfoundation.org/courses/introduction-to-aiml-toolkits-with-kubeflow-lfs147

Thank you!

Kontakt Fabrizio
https://github.com/Lazzaretti
https://www.linkedin.com/in/fabrizio-lazzaretti/

Kontakt Marco
https://github.com/mcrisafu
https://www.linkedin.com/in/mcrisafu/

https://lazzaretti.me/ittage25