MLOps Committee in LF AI & Data

Saishruthi Swaminathan

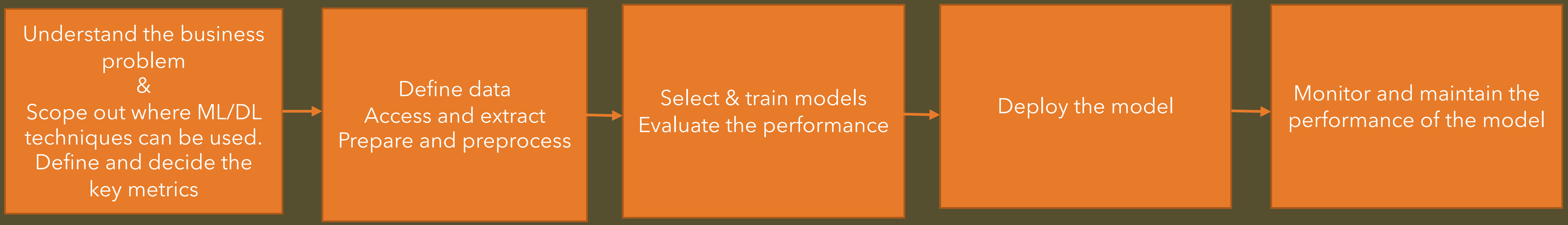
Technical Lead & Data Scientist

IBM CODAIT

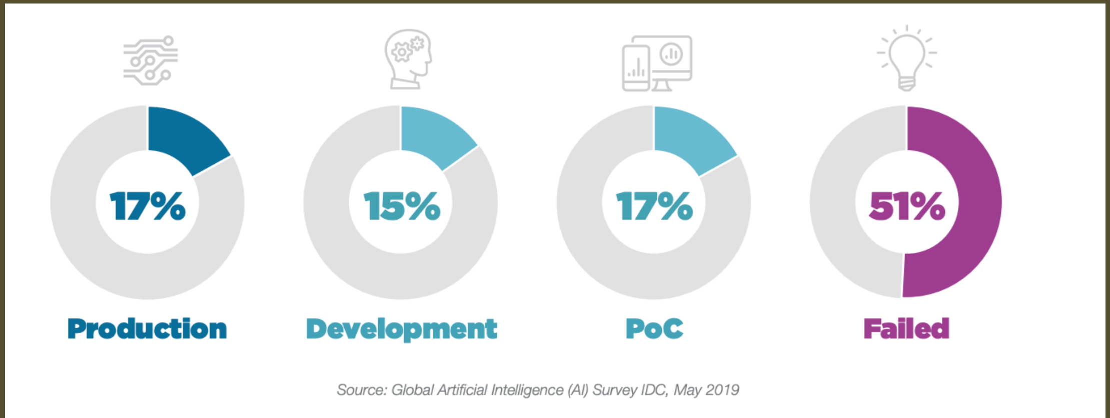
About Myself

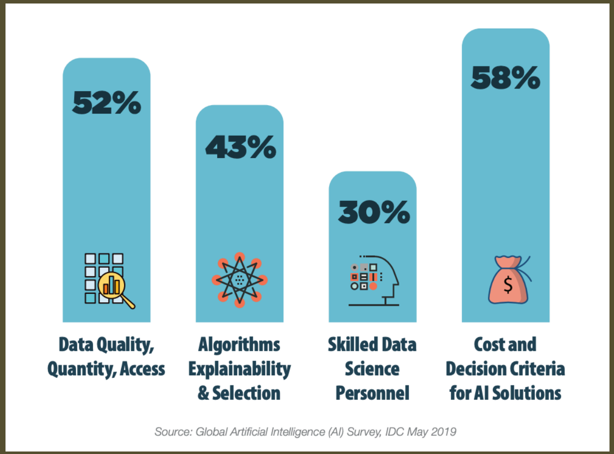
- Open Source Contributor: AI Fairness 360 R, AI Factsheet 360, Model & Data Asset eXchange, TensorFlow
- Member of LF Trusted AI Committee
- Leading Data Science Engagements for clients and partners
- Ethical AI Advocate – Over 100K reached through talks and workshops
- Worked with San Jose City in disaster management using AI research
- Research – Material Engineering and Data Annotation Quality
- Coursera Instructor - Over 110K learners
- Recognized as one of the top startup ideas in Silicon Valley Business Competition

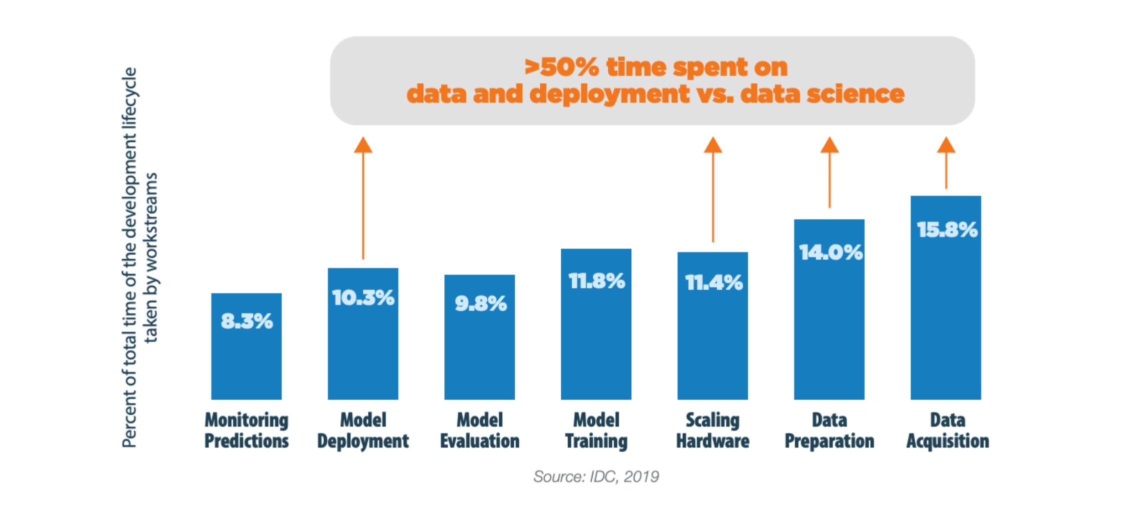
ML Lifecycle



Let’s look at some statistics - 2019



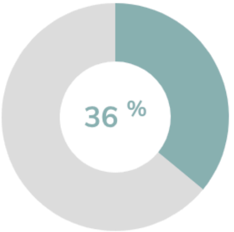




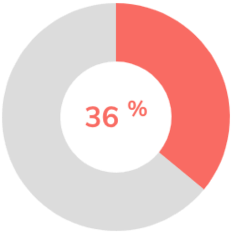
Let’s look at some statistics - 2020

Responses from 582 survey respondents

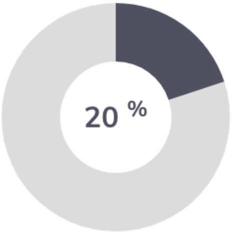
What percentage of your data scientists' time is spent deploying ML models?



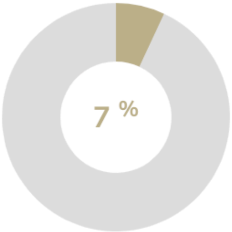
36% of survey participants said their data scientists spend **a quarter** of their time deploying ML models



36% of survey participants said their data scientists spend **a quarter to half** of their time deploying ML models



20% of survey participants said their data scientists spend **half to three-quarters** of their time deploying ML models

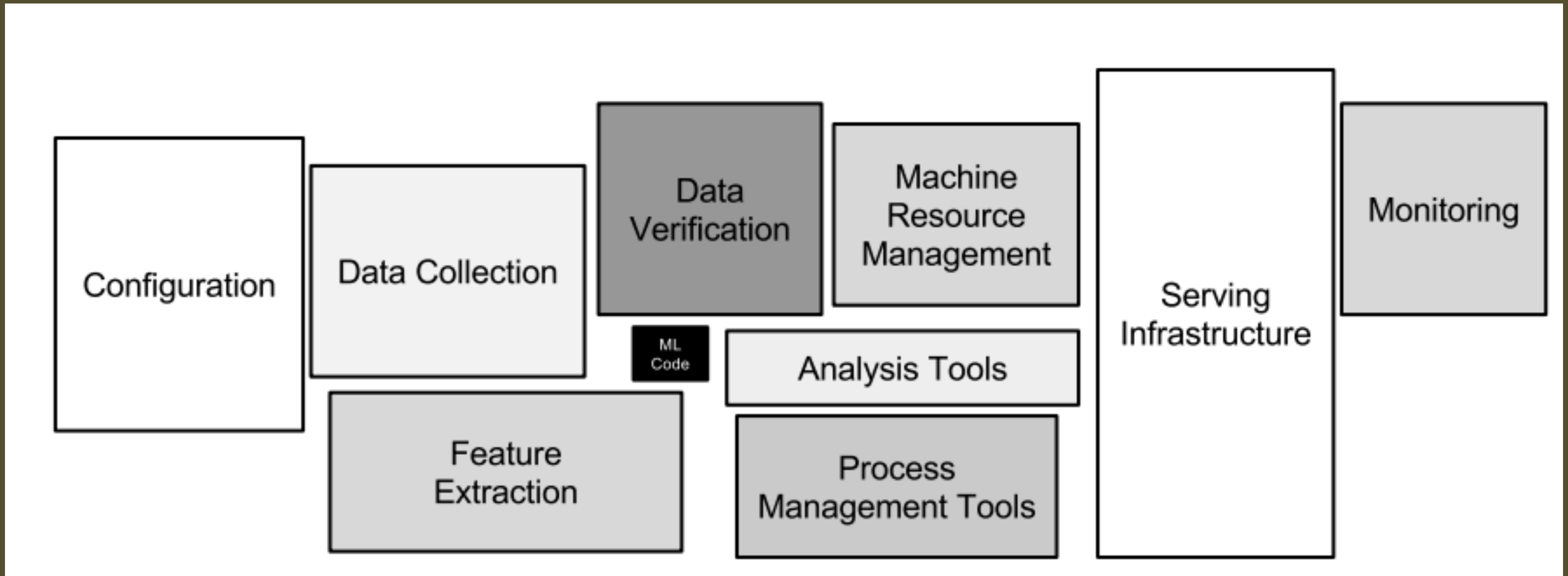


7% of survey participants said their data scientists spend **more than three-quarters** of their time deploying ML models

1% of respondents said they were unsure.

Source: https://algorithmia.com/state-of-ml?utm_medium=website&utm_source=interactive-page&utm_campaign=IC-1912-2020-State-of-ML&_hsenc=p2ANqtz-_WbXKYLnpgf4zi4OZTNYmNgCRPIFFEgmW-Cqi2Px_T1K2wklJvDt7KdCxB5vXAPmGirLi7ukZTykxeUh9vmHdn7dRF9g&_hsmi=81660946

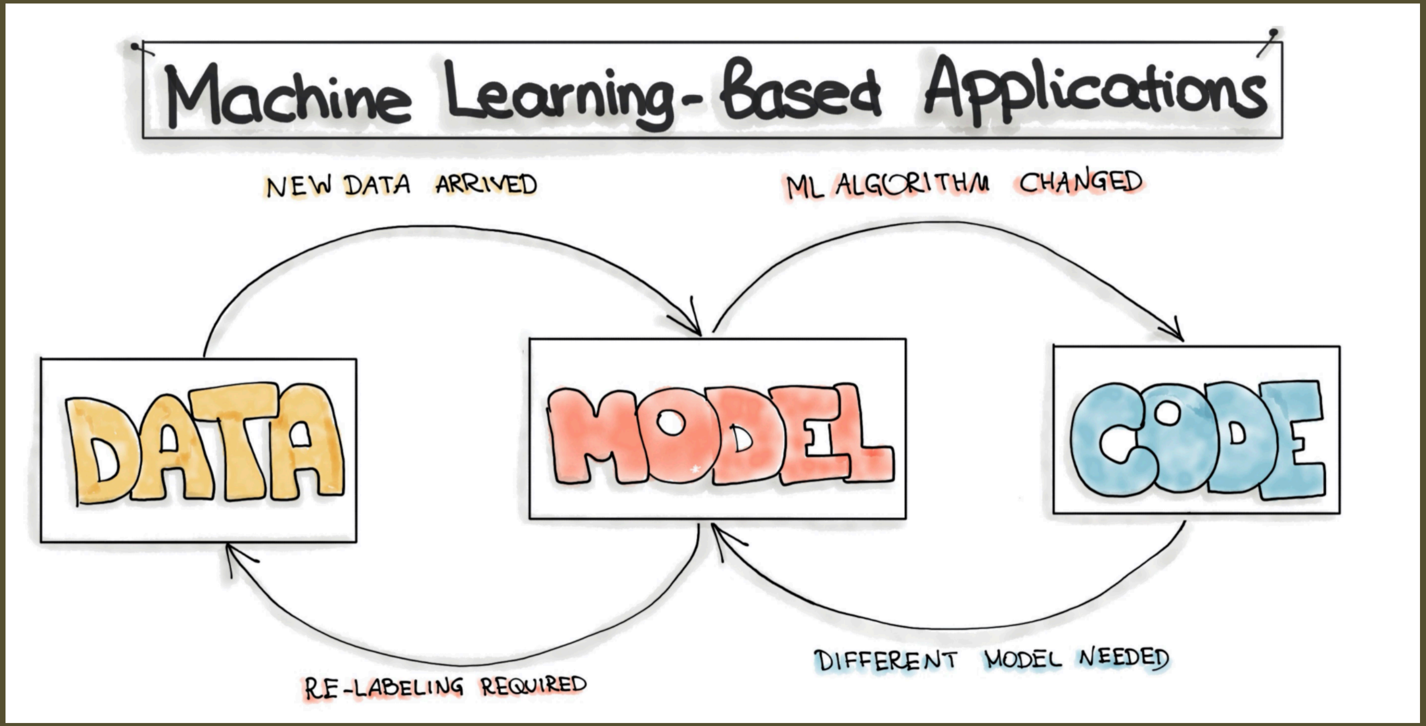
Hidden Technical Debt in ML Systems



https://papers.nips.cc/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf

"Using the software engineering framework of *technical debt*, we find it is common to incur massive ongoing maintenance costs in real-world ML systems. We explore several ML-specific risk factors to account for in system design. These include boundary erosion, entanglement, hidden feedback loops, undeclared consumers, data dependencies, configuration issues, changes in the external world, and a variety of system-level anti-patterns."

Changing Anything Changes Everything

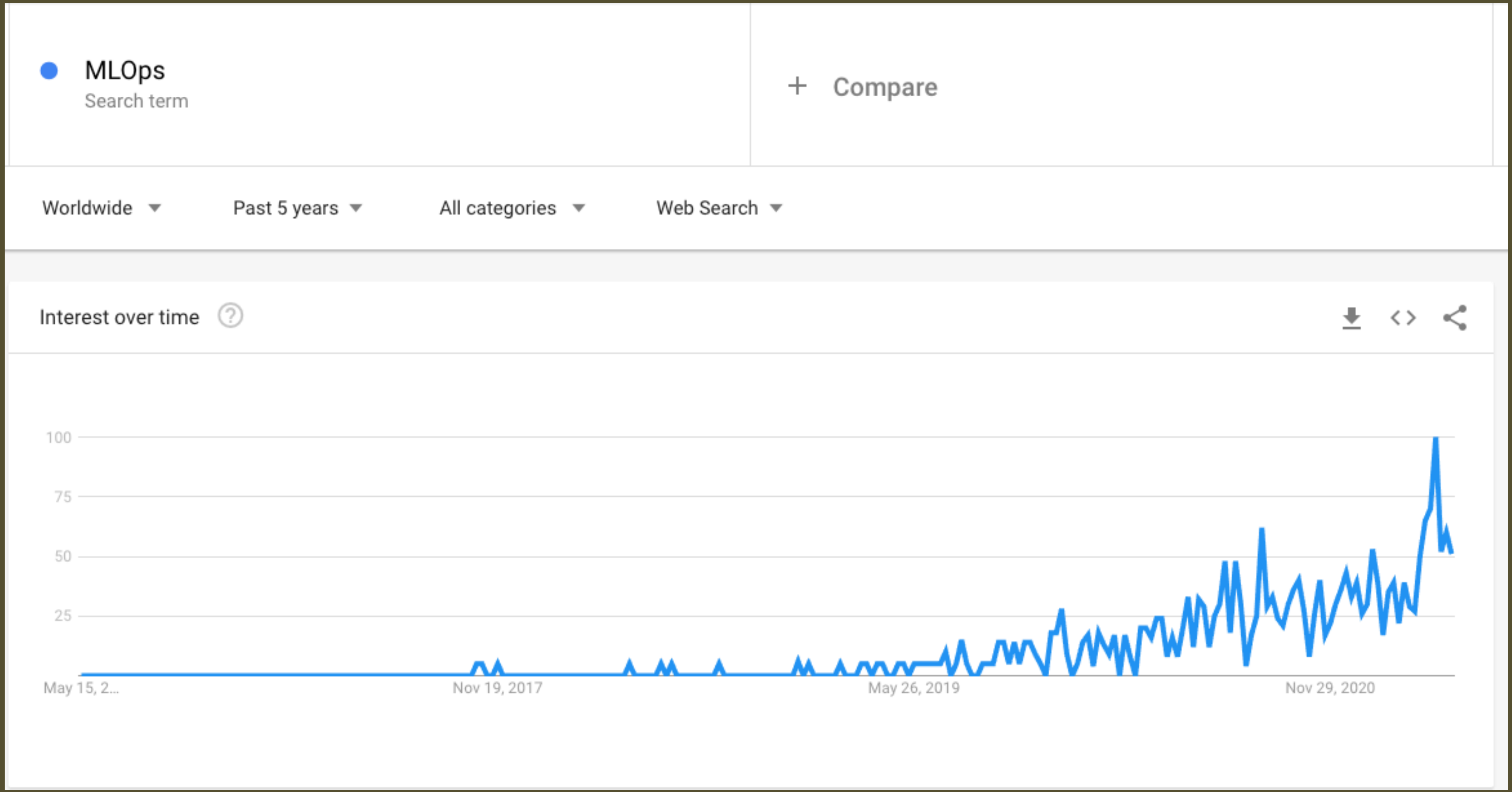


→ **MLOps**

MLOps

Set of tools and principles to support machine learning project lifecycle

Current Trend Worldwide



https://trends.google.com/trends/explore?date=today%205-y&q=MLOps

2021 Point of View

Most artificial intelligence (AI) projects fail. About 80% never reach deployment, according to Gartner, and those that do are only profitable about 60% of the time. When we take a moment to consider the signs of successful AI all around us, these numbers may come as a surprise. We have voice assistants for our phones and homes, optimized online product searches, advanced fraud detection at our banks, and more. Yet as it stands now, we'll never see the fruits of the majority of AI endeavors. In the final part of our five-part series on 2021 predictions, we look at the future of successful AI deployments.

These statistics might seem disheartening for companies that are turning to AI for positive impacts like greater revenue, lower costs, and more personalized, effective customer experiences, but we're seeing signs of promise. In 2021, we predict that companies will start to overcome the 80% failure rate of deployment. Gartner has further predicted that by 2024, 75% of organizations will shift from piloting to operationalizing AI. This change in momentum will be driven by greater accessibility to data and the development of highly flexible models to adapt to specific business needs.

VB The Machine GamesBeat Jobs Special Issue Become a Member | Sign In

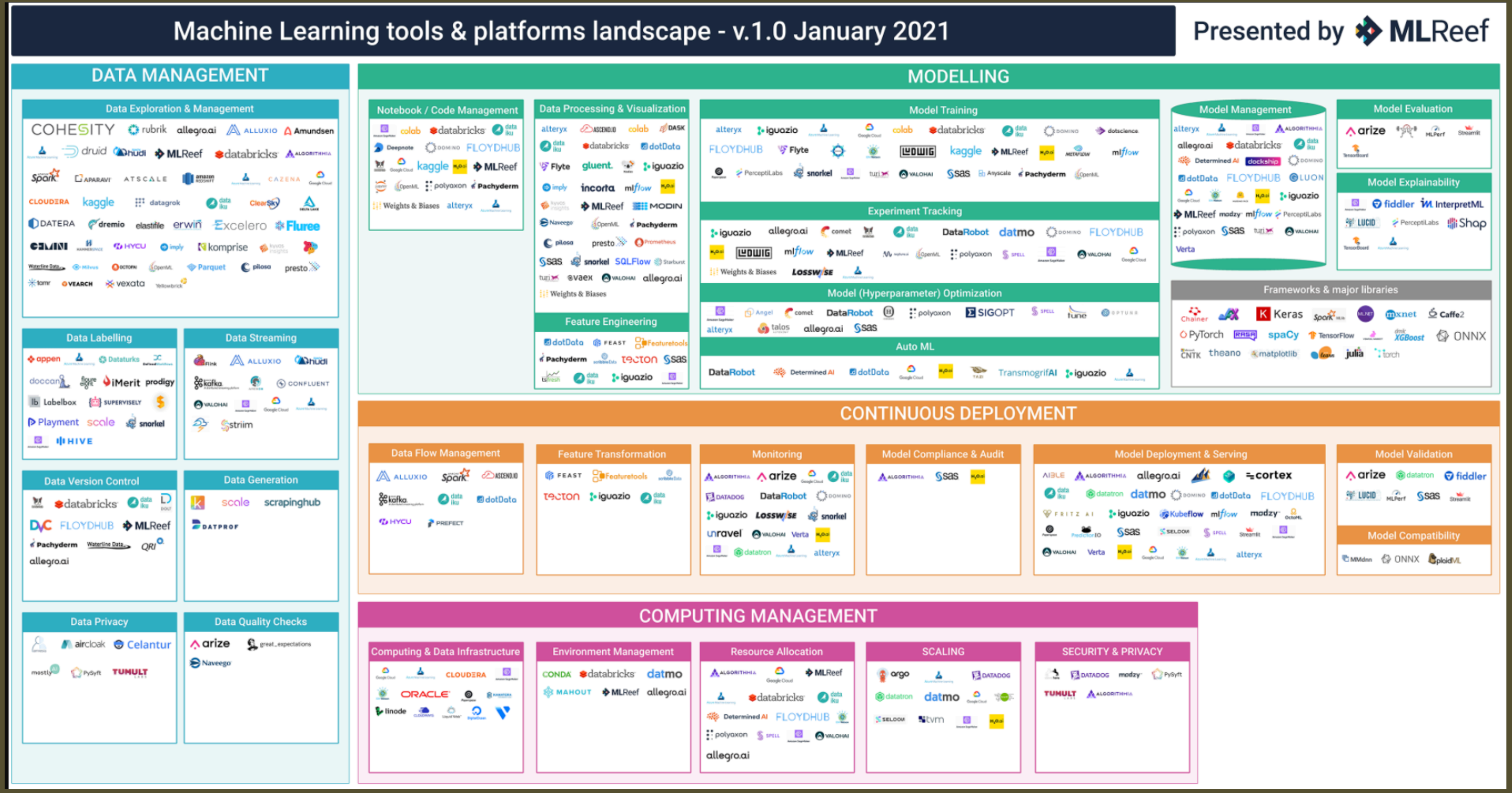
The Machine
Making sense of AI

Sponsored

The future of AI deployments reaching production is bright in 2021

Wilson Pang, Appen December 16, 2020 4:10 AM

Key Projects in MLOps Space



Note:

This list is specific to the MLOps.
Projects might overlap with existing LF AI Landscape

https://about.mlreef.com/blog/global-mlops-and-ml-tools-landscape

LF AI & Data Mission

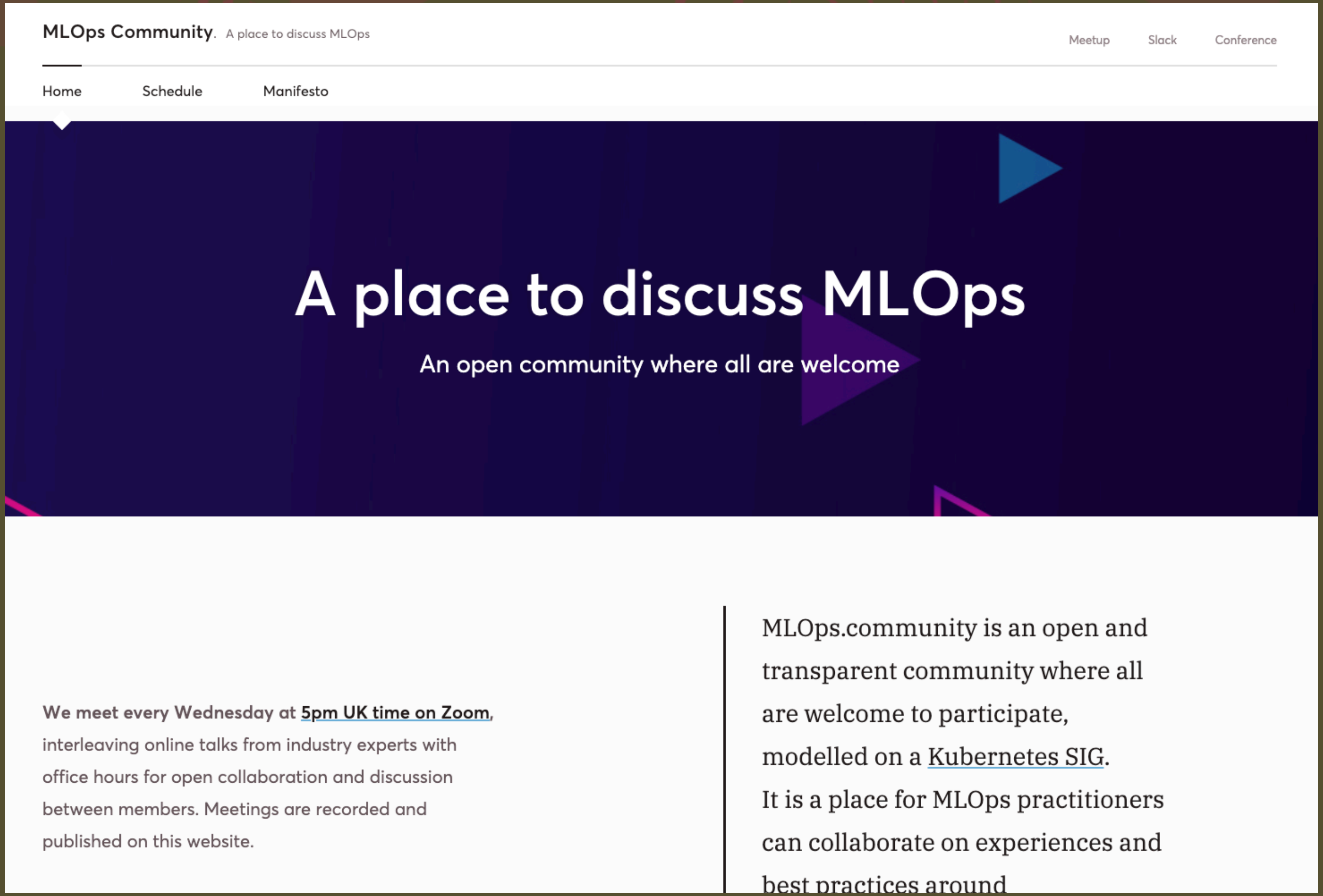
How MLOps fit in LF AI & Data Scope

ABOUT LF AI & Data

LF AI & Data is an umbrella foundation of the Linux Foundation that supports open source innovation in artificial intelligence, machine learning, deep learning, and data. LF AI & Data was created to support open source AI, ML, DL and Data, and to create a sustainable open source AI ecosystem that makes it easy to create AI and Data products and services using open source technologies. We foster collaboration under a neutral environment with an open governance in support of the harmonization and acceleration of open source technical projects.

MLOps tools and methodologies
plays a major role in getting AI models to production

Popular effort in MLOps



https://awesomeopensource.com/project/visenger/awesome-mlops

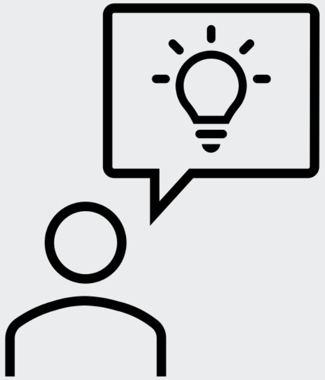
Proposal to Kick off MLOps Committee in LF AI & Data

Different organizations have different MLOps approaches

There is no one correct approach

LF AI & Data being a global group and having global innovators in Data and AI space in one place,
creating a dedicated group to discuss and innovate in MLOps space will be helpful to the community.

Based on Survey Conducted what current committee members expect



- Provide exposure on different industrial approaches
- Gather current practices and create a template architecture that can be a base for organizations trying to adopt MLOps.
- What are the Open Source MLOps project? What are each one's pros and cons?
- Data centric MLOps approach
- Use Case based approach. Learn technology through a use case
- Current industry issues in getting models to production and how to tackle as a community?

MLOps LF AI & Data Focus

Exposure on industrial approaches for managing ML models in production

Create template architecture for managing ML project lifecycle

Identify Projects and tools in MLOps Space

Get community exposed to how these MLOps tools work together and where to use in the pipeline with pros and cons

Understand usage of MLOps tools and practices through industrial use cases (by domain)

Identify gaps in the use case implementation

Discuss solutions that can fill the gap

Take data centric Approach in managing ML model performance in production

Learn tools and best practices on data centric approach

Provide opportunity for committee members to to do research together

Advocate about the work

MLOps relationship to existing Committees

- AI Ethics
- Cross Collaboration
- ML Workflow & InterOp Committee (Identify gaps in portfolio)
- Trusted AI
- Outreach Committee

Thank You

- Michael Tanenbaum
- Ludan Stoecklé
- Vishnu
- Adam Pocock
- DC Martin
- Sebastian Lehrig
- Nancy Rausch
- Meng Wei
- Yuan Liya
- Jim Spohrer
- Ibrahim Haddad