HARMAN
A SAMSUNG COMPANY

HARMAN MLOPS

AGENDA

HARMAN
A SAMSUNG COMPANY

HARMAN MLOps

1 INTRODUCTION

2 HARMAN MLOPS

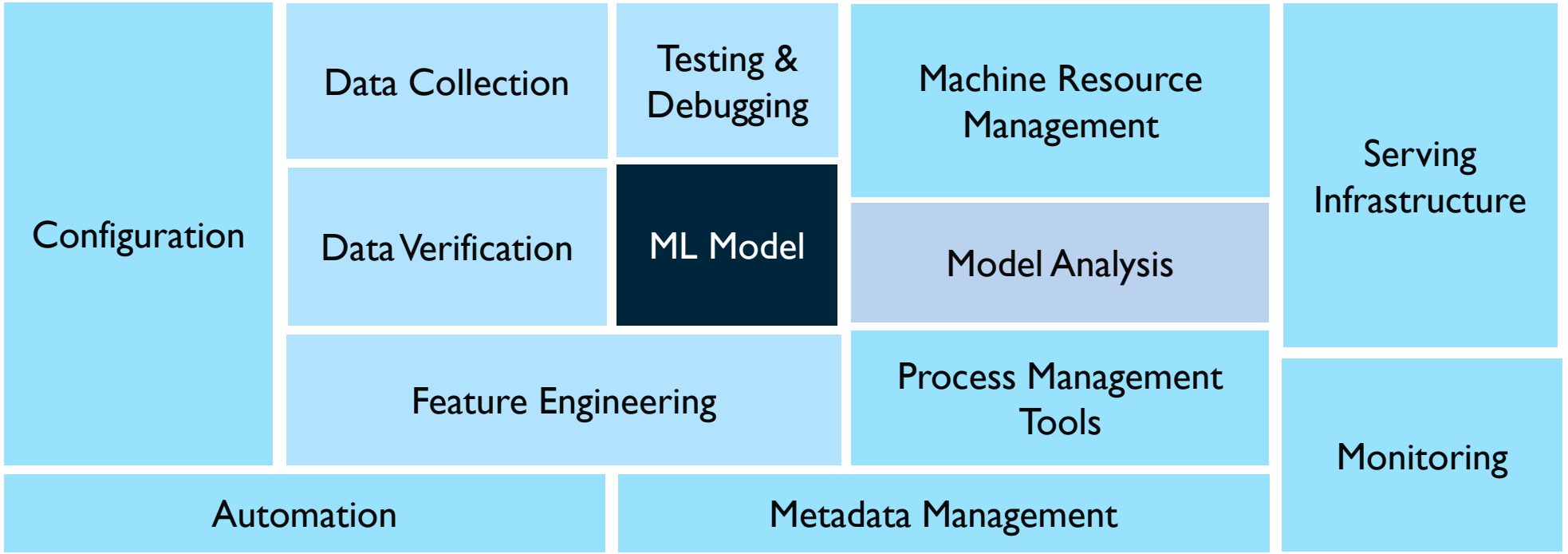
3 HARMAN MLOPS METHODOLOGY & TOOL CHAIN

4 MLOPS CASE STUDIES

HARMAN
A SAMSUNG COMPANY

1 INTRODUCTION

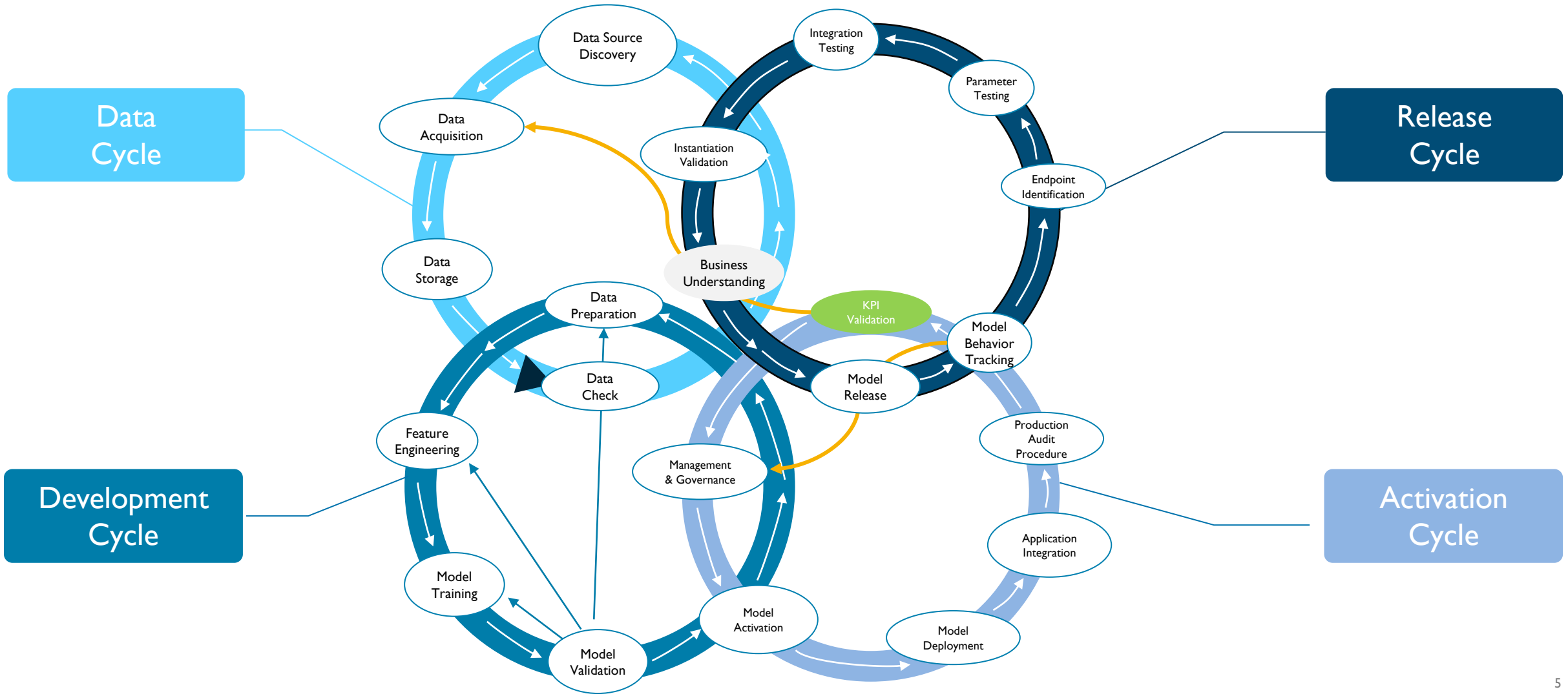
MACHINE LEARNING MODEL LIFECYCLE



Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

[Hidden Technical Debt in Machine Learning Systems (neurips.cc)]

MACHINE LEARNING MODEL LIFECYCLE



MLOPS MATURITY

Level 0

- All manual, script driven
- Prone to Training – Serving skew
- No CI/CD
- Lack of active monitoring
- Source control
- Lack of environment control
- Incompatible with DevOps setup

Level 1

- Rapid experimentation with orchestration and automation
- Automated Retraining
- Experimental-operational symmetry
- Modularized code
- Continuous delivery of models
- Pipeline deployment
- Data Validation
- Model Validation
- Feature Store (optional)
- Metadata management
- ML Pipeline Triggers
- Drift detection
- Model / pipeline performance monitoring

Level 2

- Integrated annotation
- Pipeline continuous integration
- Pipeline continuous delivery
- Automated triggering
- Model continuous delivery
- Monitoring
- Continuous integration
  - Unit test – feature engineering, processing methods
  - Test model training coverage
  - Test prediction of undefined values
  - Test production of expected artifacts
  - Integration testing
- Continuous delivery
  - Verifying the model & infrastructure compatibility
  - Prediction API testing
  - Automated deployment to a test environment
  - Semi-automated deployment to a pre-production environment
  - Manual deployment to a production environment

HARMAN
A SAMSUNG COMPANY

2 HARMAN MLOPS

HARMAN MLOPS

HARMAN MLOps is the key foundational framework for development, deployment and management of machine learning solutions in production. It enables enterprises to automate, scale, and re-use components and best practices across the ML development life cycle.

KEY VALUE PROPOSITIONS OF HARMAN MLOPS

Collaboration

Fairness, Privacy, Security, Transparency

Performance

Automation

Speed

Governance

HARMAN MLOPS

Key Capabilities

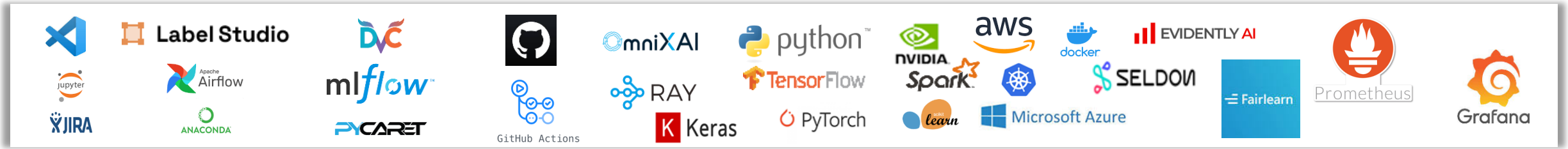
- Integrated annotation – Annotate multiple types of data (Text, Image, Time Series, Tabular)
- Rapid experimentation with orchestration and automation – DAG based scheduling and automation of scripts
- Experimental-operational symmetry – Ensure that the model developed by Data Scientists work with same efficacy in production also with real world data
- Modularized code – All solutions are developed as reusable modules as API or webservices
- Pipeline deployment – CI / CD ensures seamless deployment of developed data and ML pipelines in prod
- Feature Store – Ensure availability and reusability of engineered attributes for rapid experimentation & retraining
- Metadata management – All data, pipeline, experimentation & model metadata available for scrutiny
- Drift detection – Capability to monitor and statistically detect both data & concept drifts
- Model / pipeline performance monitoring – Continuous monitoring of performance & business KPIs

Implementation*

Desired Maturity Level

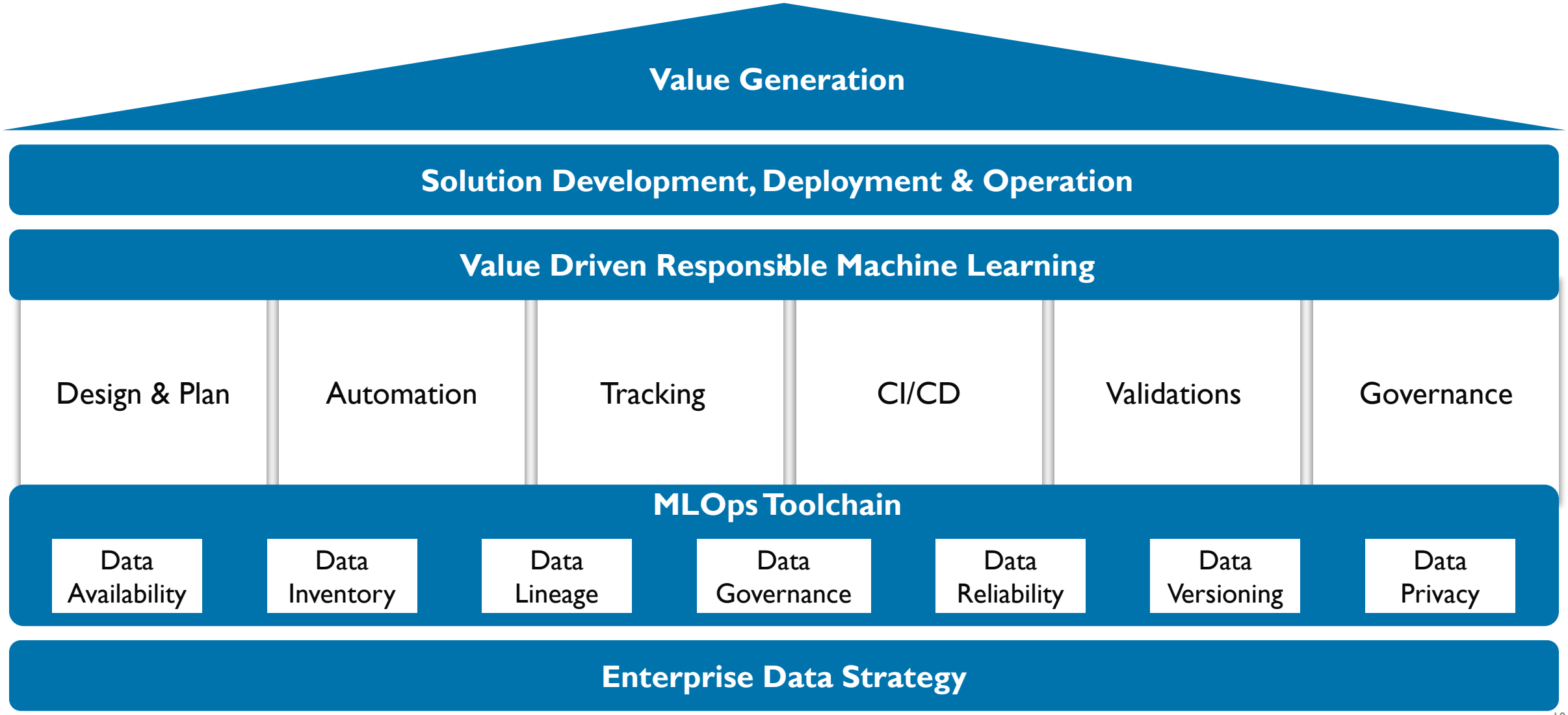
Current Maturity Level		L1: Automated retraining, Manual CI/CD, Responsible AI	L2: Automated CI/CD, Monitoring based triggers
	L0: Manual, No CI/CD, Fragmented	3-5 months	4-7 months
	L1: Automated retraining, Manual CI/CD, Responsible AI		2-4 months

Operations Size



* Indicative, per use-case

DATA TO VALUE ENABLED WITH MLOPS



BUILT ON TENETS OF RESPONSIBLE AI

HARMAN
A SAMSUNG COMPANY

Performance	Transparency	Bias Free	Auditable	Privacy preserving
• High performance machine learning solutions in terms of relevant metrics (precision, recall, F1 score, latency etc.)	• Models which can be interpretable (human can predict the behavior of model), and / or have their mechanism presented in human terms. • Explanations customized for the skill level of user	• Checks and controls built in to avoid any undue advantage or disadvantage to a recipient group	• All the actions and parameters of the machine learning models are recorded immutably and available for audits.	• All solutions designed with privacy preserving frameworks to prevent access of PII and PHI data to any unauthorized actor.
Regulatory compliance	Ensure fairness	Better model governance	User trust & confidence	Cloud / hybrid friendly

KEY BENEFITS OF HARMAN MLOPS

HARMAN
A SAMSUNG COMPANY

Agility & Speed

- Bring ML models to production faster and at scale
- Faster time to market for products & services

Efficiency

- ML model output with reduced effort. Help users build & deploy models faster.

Explainability

- Explain how the ML model works

Effectiveness

- Help users make correct decisions

Trust

- Increase users' confidence in the system

Automation

- Automated model pipeline management reduces manual interventions, decrease time for deployment, enables continuous delivery

Collaboration

- Track model, code and data changes and increase collaboration among teams

Scrutability

- Allow users to tell the system that it is wrong

Debugging

- Allow users to identify biases or defects in the system so that they can be corrected.

Monitoring

- Monitor models in production
- Respond to model performance issues faster
- No broken models in production

Cost of Development

- Reduced cost of development due to automation, CI/CD monitoring & seamless integration

Governance & Compliance

- Reduced risks due to Model explainability, compliance

HARMAN
A SAMSUNG COMPANY

3 HARMAN MLOPS METHODOLOGY & TOOL CHAIN

OFFERINGS

Capability Assessment & Advisory

Activities :

- ML Ops Requirements and Capability Assessment - delivered through workshops, stakeholder interviews, assessment and user journey workshop, identification of frameworks and tooling, gap analysis, understanding and assumptions,
- MLOps Solution Design - understanding key challenges and gaps in the MLOps capability, reviewing technical options, relevant tools, ways of working and defining a target solution architecture and an ML operating model
- Playback and input into an MLOps Solution Roadmap and a Business Case

Deliverables:

- Target MLOps Solution Architecture
- Documented Recommendations for an ML Target Operating Model
- MLOps High level Implementation Roadmap and inputs into a Business Case aligned with the business goals and priorities

Implementation

Activities :

- Project planning including milestones
- Tool requirement assessment
- ML Engineers to setup MLOps pipelines and toolchain

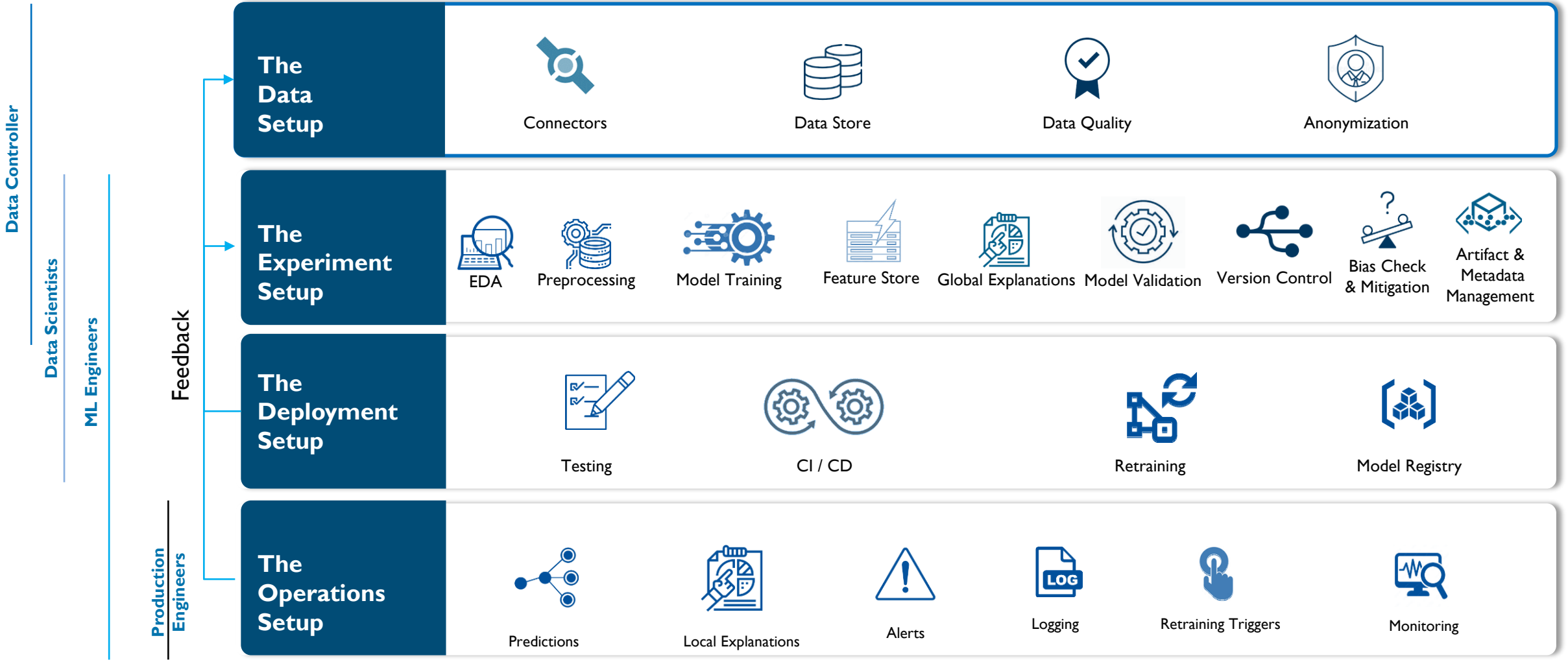
Deliverables:

- Complete toolchain and pipeline setup to run machine learning experiments, test & deploy them in production, and manage them
- Standardized environments
- Role-based Access Controls
- Responsible AI

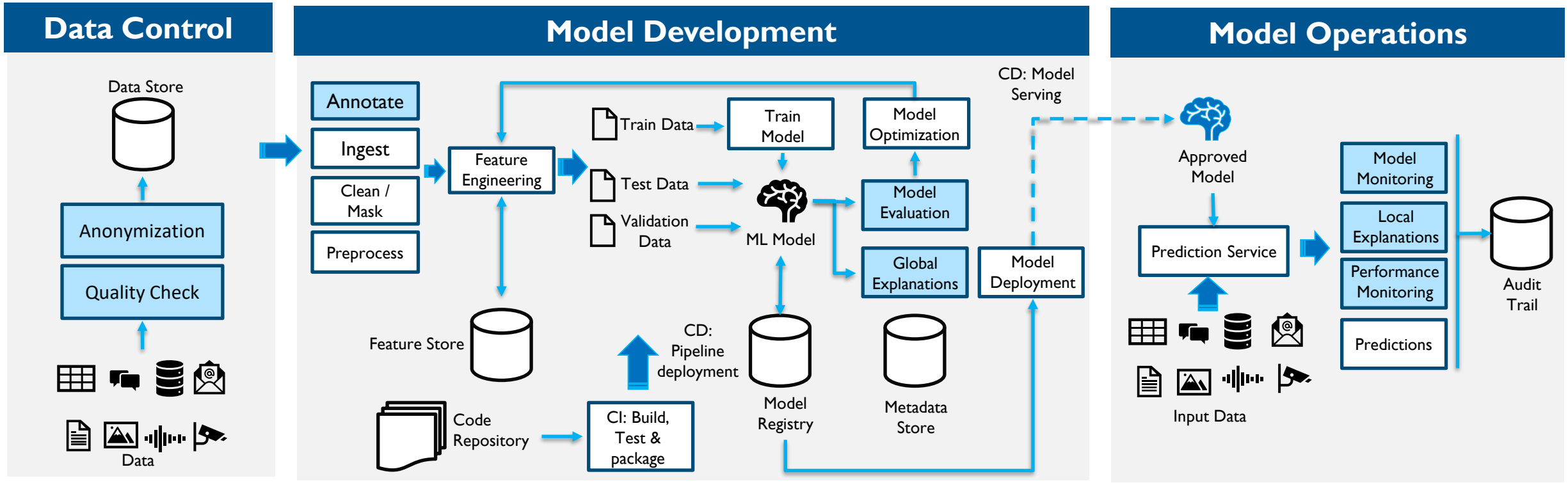
MACHINE LEARNING MODEL LIFECYCLE

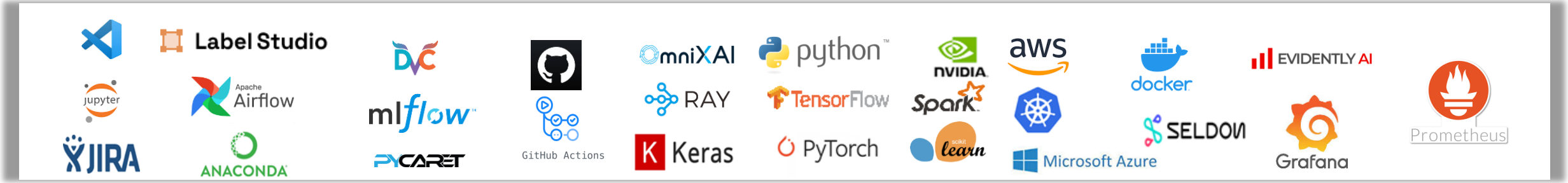
Machine Learning Development				Machine Learning Operations	
Design	**Data**	**Model Development**	**Testing**	**Deployment**	**Production**
Business Understanding	Data pipeline	Model training pipeline	Testing pipeline	Deployment Pipeline	Model Governance
Requirement Gathering	Data Extraction from unstructured sources	Feature Engineering	Model correctness, performance, relevance, explainability	Model portability across different platforms	Model Monitoring, Drift Detection, Retraining triggers
Use-case Prioritization	Data Cleaning & Wrangling	Automated model selection	Model efficiency, robustness, fairness, interpretability	Automated Deployment	Model behavior tracking
Data Acquisition	Version Control of Code and Data	Automated training	Packaging, infra, pipeline, API & integration testing	Standardise Deployment	Model performance
Security & Privacy concerns	Data Tagging & Labeling	Model reproducibility, versioning	Data & model drift testing		Model Explainability
	Version Control of Model, Code and Data	Model Packaging	Automated Testing		Auditing, Compliance
		Build, select and track model versions	Execute experiments in a visual intuitive manner		Model Lifecycle Management

BUILDING A ROBUST MLOPS

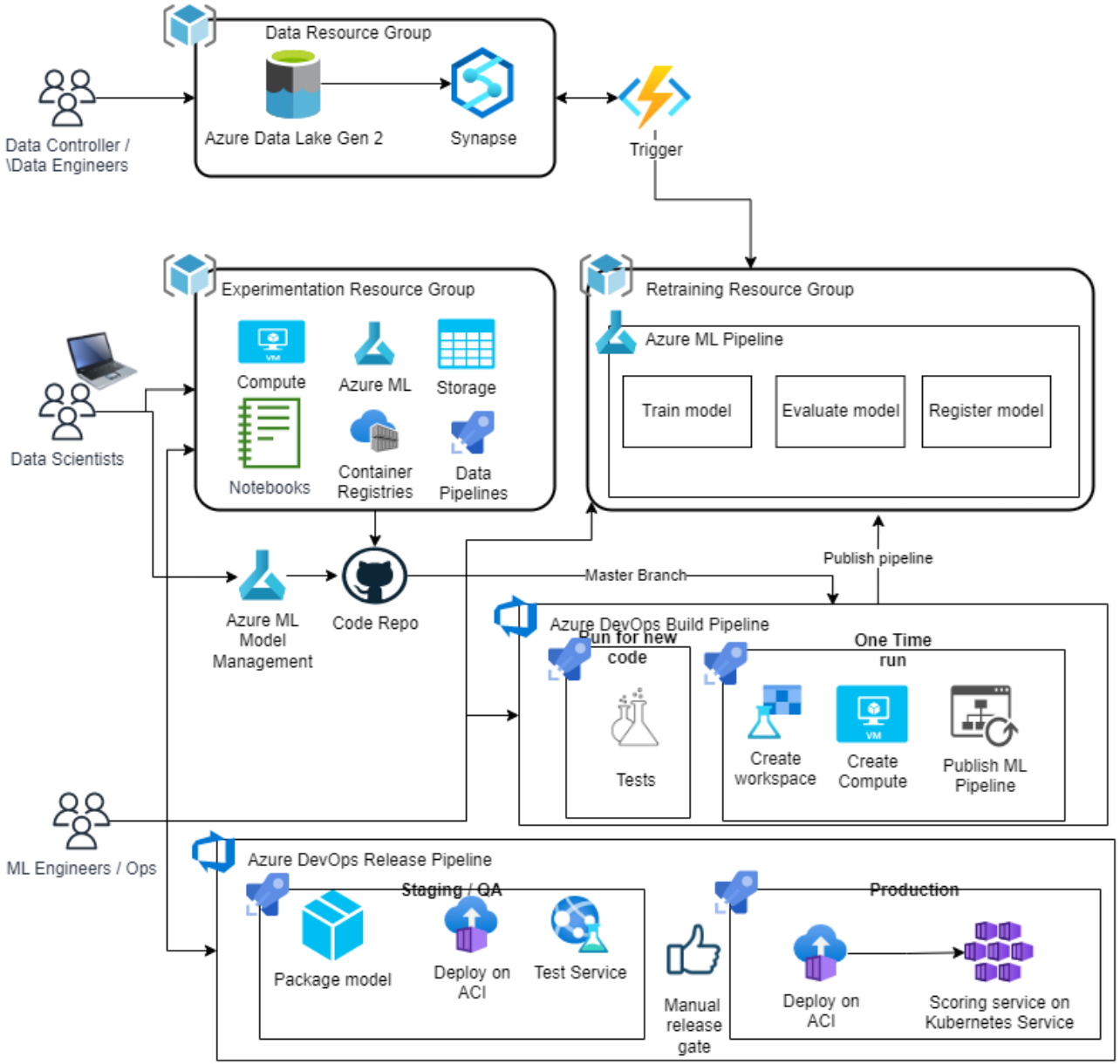


HARMAN MLOPS TOOLCHAIN





AZURE CLOUD NATIVE REFERENCE ARCHITECTURE



DEVELOPMENT / DEPLOYMENT TOOLCHAIN (OPEN – SOURCE STACK)

HARMAN
A SAMSUNG COMPANY

S. No.	Tool	Functionality	Description
1	Label Studio	Annotation	Label every data type. (https://labelstud.io/)
2	Airflow	Scheduler	Airflow is a platform created by the community to programmatically author, schedule and monitor workflows. (https://airflow.apache.org/)
3	Ray / Rapids / Sklearn / Keras / Pytorch	ML Frameworks	
4	MLFlow	Manage ML Lifecycle	MLflow is an open source platform to manage the ML lifecycle, including experimentation, reproducibility, deployment, and a central model registry.
5	DVC	Data & ML pipeline version control	DVC is built to make ML models shareable and reproducible. It is designed to handle large files, data sets, machine learning models, and metrics as well as code.
6	Github	Code versioning	The complete developer platform to build, scale, and deliver secure software.
7	Github Actions	CI/CD	Automate, customize, and execute your software development workflows right in your repository with GitHub Actions. (https://github.com/features/actions)
8	Github Copilot	Code help	Get autocomplete-style suggestions from an AI pair programmer as you code. (https://github.com/features/copilot)
9	Evidently.AI	Monitoring	Tools to evaluate, test and monitor machine learning models (https://www.evidentlyai.com/)
11	Grafana	Monitoring	Multi-platform open source analytics and interactive visualization web application. (https://grafana.com/)

DEVELOPMENT / DEPLOYMENT TOOLCHAIN (OPEN – SOURCE STACK)

S. No.	Tool	Functionality	Description
12	OmniXAI	Explainability	Omni-way explainable AI and interpretable machine learning capabilities to address many pain points in explaining decisions made by machine learning models (https://github.com/salesforce/OmniXAI)
13	Microsoft Presidio	Data Protection	It provides fast identification and anonymization modules for private entities in text and images such as credit card numbers, names, locations, social security numbers, bitcoin wallets, US phone numbers, financial data and more. (https://microsoft.github.io/presidio/)
14	Prometheus	Monitoring	Systems monitoring and alerting toolkit (https://prometheus.io/)
15	Dockers		All tools available as containerized microservices to the extent possible
16	Jupyter Notebook	Experimentation	IDE and environment for ML experimentation
17	VSCode	Coding	IDE with significant integrations available natively

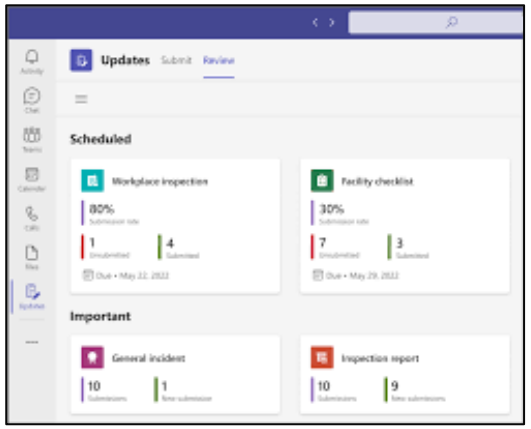
KEY CAPABILITIES

HARMAN
A SAMSUNG COMPANY

Integrated annotation – Annotate multiple types of data (Text, Image, Time Series, Tabular)	Rapid experimentation with orchestration and automation – DAG based scheduling and automation of scripts	Experimental-operational symmetry – Ensure that the model developed by Data Scientists work with same efficacy in production also with real world data	Explainability – Provide both Global (model level) and local (prediction level) human interpretable explanations	Modularized code – All solutions are developed as reusable modules as API or webservices
Pipeline deployment – CI / CD ensures seamless deployment of developed data and ML pipelines in prod	Feature Store – Ensure availability and reusability of engineered attributes for rapid experimentation & retraining	Metadata management – All data, pipeline, experimentation & model metadata available for scrutiny	Drift detection – Capability to monitor and statistically detect both data & concept drifts	Model / pipeline performance monitoring – Continuous monitoring of performance & business KPIs

TOOL CHAIN FOR DESIGN

MS Teams



- Collaboration
- Meetings
- Document Sharing

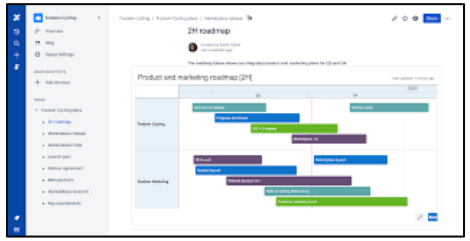
MIRO Board

- Ideation & Brainstorming
- User Journeys
- Workflow Design

JIRA

- Project Management
- Feature Tracking

Confluence



- Team Wiki
- Collaboration
- Knowledge Management

HARMAN
A SAMSUNG COMPANY

THANK YOU