



# Platform Engineering Day EUROPE

**AMSTERDAM, THE NETHERLANDS**

**23 MARCH 2026**

#PLATENGDAY



# Creating an IDP for AI Engineering

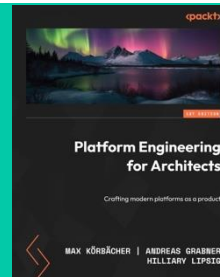
*It's the Platform, Not the Model.*

**Max Körbächer**

CNCF GB Member & Ambassador

LF Europe Advisory Board

CEO & Chief Technology Advisor @Liquid Reply



01

# The Enterprise AI Paradox

# Hype Yay, Value Ney

88%

of orgs use AI

80%

of AI projects fail

72%

of CIOs break even  
or lose money

5%

achieve sustained  
P&L impact



The gap between experimentation and production value has become a stale.

# What's Blocking the Path to Production?





## People

-  94% face AI skill shortages
-  3:1 talent demand vs supply
-  79% of execs fake AI knowledge
-  Only 35% have upskilling programs

## Process

-  Only 18% have AI governance
-  47% had negative consequences
-  81% use unapproved AI tools
-  Shadow AI adds \$\$\$ to breaches

## Technology

-  75%+ GPUs underutilized
-  40-70% resource idle time
-  AI reversed 5yr cloud waste decline
-  1,150+ tools in the landscape

02

# The AI/ML Tooling Complexity Crisis

# The 2025 MAD Landscape: Tool Sprawl is Real

**1,150+** companies mapped  
in the MAD Landscape


























































































**40–50** tools to evaluate before  
selecting 10–15 for production

Stage	Tools	Tension
<b>Experiment Tracking</b>	MLflow, W&B, Neptune, Comet	Heavy overlap; teams often run 2
<b>Orchestration</b>	Kubeflow, Airflow, Prefect, Dagster, Metaflow	7 tools, different philosophies
<b>Model Serving</b>	vLLM, Triton, KServe, BentoML, Seldon	Different layers co-exist
<b>LLM Frameworks</b>	LangChain, LlamaIndex, LiteLLM	Rapidly evolving; agents add more
<b>Monitoring</b>	Evidently, Arize, Langfuse, LangSmith	AI metrics distinct from infra

# CNCF CNAI Landscape is not match for ...

Filters GROUP: [Projects and products](#) [Members](#) [Certified partners and providers](#) [Serverless](#) [Wasm](#) [CNAI](#) VIEW MODE: [Grid](#) [Card](#) ZOOM: [-](#) [+](#)

CNAI

<b>Data Architecture</b> 🔍  CNCI INCUBATING                  	<b>Distributed Training</b> 🔍         	<b>Workload Observability</b> 🔍     	<b>Open Enterprise AI Blueprints</b> 🔍 
<b>General Orchestration</b> 🔍  CNCI GRADUATED  CNCI INCUBATING  CNCI INCUBATING         	<b>Data Science</b> 🔍       	<b>AutoML</b> 🔍   	
<b>Governance, Policy &amp; Security</b> 🔍  CNCI GRADUATED  CNCI INCUBATING     	<b>Model/LLM Observability</b> 🔍       	<b>Vector Databases</b> 🔍   	
<b>ML Serving</b> 🔍  CNCI INCUBATING        	<b>CI/CD - Delivery</b> 🔍     	<b>Agentic AI</b> 🔍  	

APPLICATIONS & AGENTS

ENTERPRISE

Grid of logos for Enterprise AI applications including Zendesk, Salesforce, Microsoft, and others.

Grid of logos for AI Agents and Applications including Anthropic, OpenAI, and various industry-specific solutions.

INDUSTRY

Grid of logos for Industry-specific AI solutions across sectors like Healthcare, Finance, and Robotics.

INFRASTRUCTURE

Grid of logos for Infrastructure providers including Amazon, Google, Microsoft, and various database and storage solutions.

ANALYTICS

Grid of logos for Analytics and BI platforms including Tableau, PowerBI, and various data visualization tools.

MACHINE LEARNING & ARTIFICIAL INTELLIGENCE

Large grid of logos for ML and AI platforms, categorized into Developer, Observability, Safety, and various AI models.

INTEGRATE & OPERATE

Grid of logos for Integration and Operations tools including ETL, Data Governance, and Cloud Observability.

DATA SOURCES & APIS

Grid of logos for Data Sources and APIs including Kaggle, GitHub, and various data marketplaces.

DATA & AI CONSULTING

Grid of logos for Data and AI Consulting firms including Accenture, IBM, and Deloitte.

# The LLM Era Added Entirely New Categories



## LLM Frameworks

LangChain, LlamaIndex,  
Semantic Kernel, Haystack

*Agent frameworks (LangGraph, CrewAI)  
adding yet another layer*



## Vector Databases

Pinecone, Weaviate, Qdrant,  
Milvus, Chroma, pgvector

*Essential for RAG, didn't exist  
as a category 3 years ago*



## AI Gateways

LiteLLM, Portkey, Helicone

*Multi-model routing, caching,  
cost control, observability*



## Fine-Tuning

HuggingFace PEFT, Axolotl,  
Unsloth, TorchTune

*LoRA/QLoRA techniques require  
new tooling and GPU management*



## LLM Observability

Langfuse, LangSmith,  
Arize Phoenix

*Prompt logging, hallucination  
detection, cost tracking*

# The LLM Era Added Entirely New Categories



## LLM Frameworks

LangChain, Llama Index, Semantic Kernel, Haystack

Agent frameworks (LangGraph, CrewAI) adding yet another layer



## Vector Databases

Pinecone, Weaviate, Qdrant, Milvus, Chroma, Redis

Essential for RAG 't exist 10 years ago



## AI Gateways

LiteLLM, Portkey, Helicone

Multi-model routing, caching, cost control, observability



## Fine-Tuning

HuggingFace PEFT, Axolotl, Unsloth, TorchTune

LoRA/QLoRA techniques require new tooling and GPU management



## LLM Observability

Langfuse, LangSmith, Arize Phoenix

Prompt logging, hallucination detection, cost tracking

**Each tool is individually excellent and cloud-native. Together, they create an integration tax.**

03

# Platform Engineering Principles

# Platform Engineering Has Already Solved This

90%

of orgs use an IDP

*DORA 2025*

76%

have dedicated  
platform teams

*DORA 2025*

60%

increase in deployment  
frequency

*DORA 2025*

2x

faster software  
delivery

*Backstage study*

## DORA 2025 Key Finding

When platform quality is high, AI's effect on organizational performance is strong and positive.

**When quality is low, AI's effect is negligible.**

*The platform is the prerequisite for AI value.*



# Three Principles That Transfer to AI



## Golden Paths

Pre-configured, opinionated workflows that reduce cognitive load while remaining optional. Coined by Spotify, called "Paved Roads" at Netflix. A developer fills in a form and the platform creates repos, CI/CD, monitoring, and security automatically.



## Self-Service Infrastructure

Replace ticket queues with form-based provisioning, code-based specs, CLIs, and APIs. Environment provisioning drops from days to minutes. No more waiting for the ops team.



## Platform as a Product

Treat the internal platform as a product with developers as customers. Apply product management discipline: user research, roadmaps, NPS scores, iterative delivery. Start with the "Thinnest Viable Platform."

04

# The AIDP Architecture

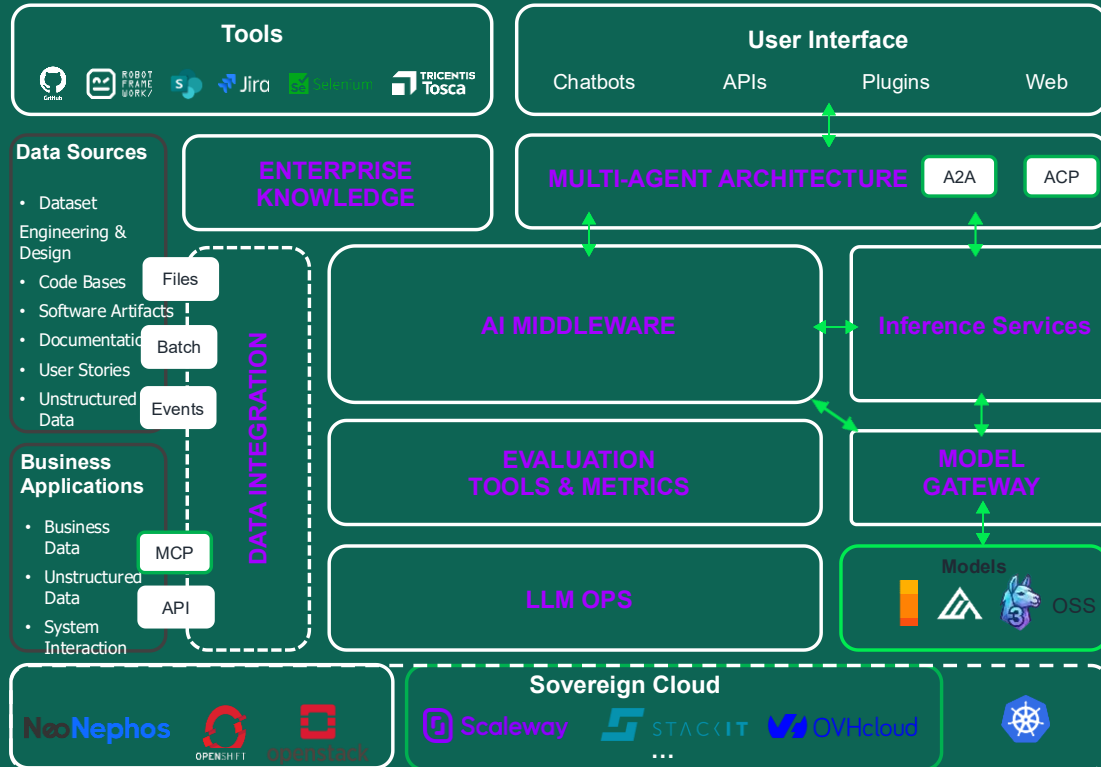
# AIDP Reference Architecture

Six integrated layers from GPU infrastructure to developer portal



# AIDP

Has to cover and integrate for a wide field of domains.



# Layer 5: The AI Developer Portal

## AI-Extended Software Catalog

- ✓ **Services & APIs** (open standards)
- ✓ **ML Models** (versioned, with model cards)
- ✓ **Datasets** (lineage, schema, ownership)
- ✓ **Pipelines** (training, inference, data)
- ✓ **Agents** (LLM-powered, with guardrails)
- ✓ **Prompts** (versioned prompt templates)
- ✓ **Feature Sets** (reusable feature definitions)

## AI Golden Paths (Templates)

- ➔ **Fine-tune an LLM**  
Data prep → training → eval → deploy
- ➔ **Deploy a RAG Pipeline**  
Embeddings → vector DB → retrieval → serve
- ➔ **Create an AI Agent**  
LLM + tools + guardrails → endpoint
- ➔ **Set Up Experiment Tracking**  
MLflow + compute + storage → notebook
- ➔ **Model A/B Test**  
Canary → traffic split → metrics → promote

*Think: developer fills in a form → platform provisions everything end-to-end.*

05

Examples

# Pioneered Internally by Big Tech

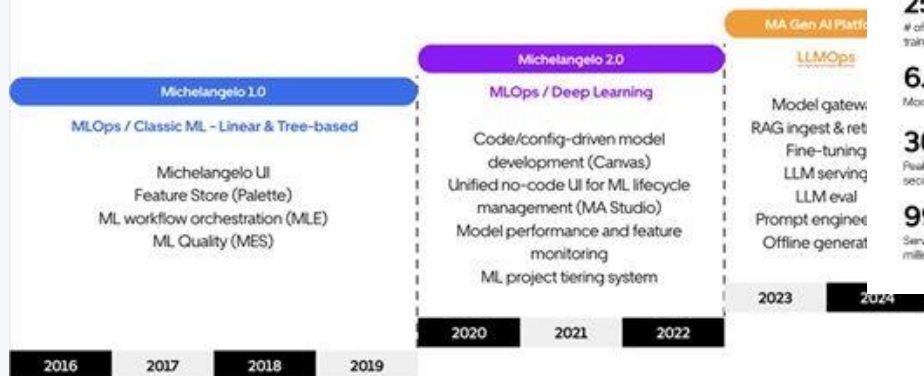
Uber

## Michelangelo 3.0

Unified feature store ("Palette"), multi-framework training via "Lego" plug-and-play architecture, <10ms serving via NVIDIA Triton. Now in GenAI phase.

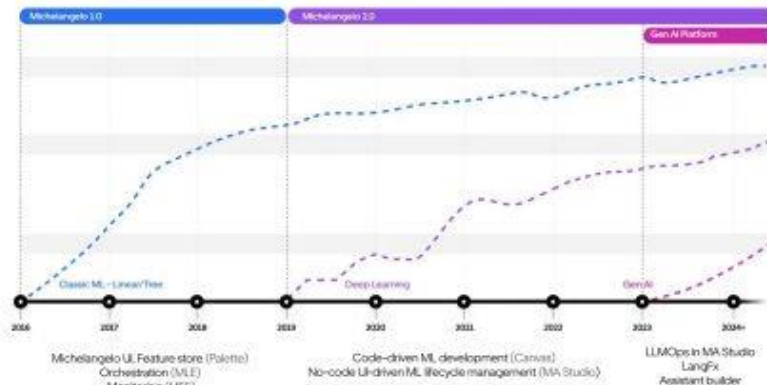
*Michelangelo has been evolving as Uber's business grows*

### Evolution of Michelangelo - Uber's AI Platform



**Michelangelo's Mission:** To provide **best-in-class** ML capabilities and tools so that Uber's ML developers can **rapidly** build, deploy, and iterate **high-quality** ML applications **at scale**

- 100%** Uber's most mission-critical ML user cases on Michelangelo
- 25K** # of model trained/month
- 6.1K** Models in production
- 30M** Peak predictions per second
- 99.99%** Serving reliability w/ millisecond latency

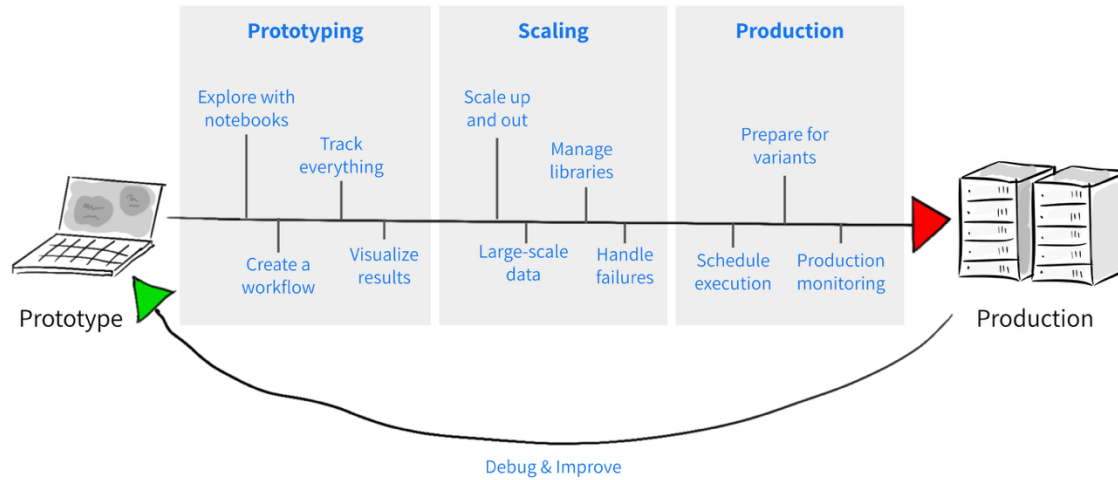


# Pioneered Internally by Big Tech

Netflix

## Metaflow

Handles the "hidden 80% of data science that isn't model building." Now open-source. Scored highest in CNCF Radar for "likelihood to recommend."



06

# What an AIDP Delivers

# Six Outcomes an AIDP Enables



**Reduced Time-to-First-Model**

**60% faster**

Golden paths turn multi-week setups into self-service actions



**Standardized Workflows**

**40-50% less cognitive load**

Curated stacks replace evaluating 40+ tools



**Governance at Scale**

**Embedded, not bolted on**

Model cards, audit trails, bias checks, automatic



**Cost Control & GPU Efficiency**

**85-96% utilization**

Centralized scheduling, LLM caching (87-95% savings)



**Shadow AI Containment**

**7% EBIT breach risk**

Governed self-service beats unapproved alternatives



**Democratization**

**AI as an API**

Non-ML developers deploy AI via golden paths

# The Thinnest Viable AIDP

*Start with the single most painful workflow. Make the sanctioned path easier than the shadow path.*

## Phase 1

### Foundation

Find your root. Catalog your existing AI assets (models, datasets, pipelines). Create your first golden path template.

## Phase 2

### Integration

Connect MLflow for experiment tracking. Add Kubeflow/Airflow for orchestration. Integrate GPU scheduling.

## Phase 3

### Self-Service

Build golden path templates for common workflows (fine-tune, RAG, agents). Add AI gateway (LiteLLM). Enable self-service provisioning.

## Phase 4

### Governance

Add monitoring (Evidently/Arize). Implement FinOps dashboards. Enable audit trails and compliance controls.

The competitive moat in  
enterprise AI isn't the model.

It's the platform.

Thank you.