Lecture HSE · May 23, 2026

DataOps / MLOps for AI Business

How managers govern data, models, release flow and economics for production AI

Alexander Polomodov

Speaker Context

Technical Director & Fellow, T-Technologies

- Architecture and engineering R&D.
- AI adoption in SDLC.
- Focus: governed production AI.

Here DataOps/MLOps is a management loop

Releasing models safely on top of a data platform

Opening | Follow-up

We skip the basics

No data-platform recap.

We look as managers

Decisions, owners, loop control.

We count economics

Data, features, inference, review.



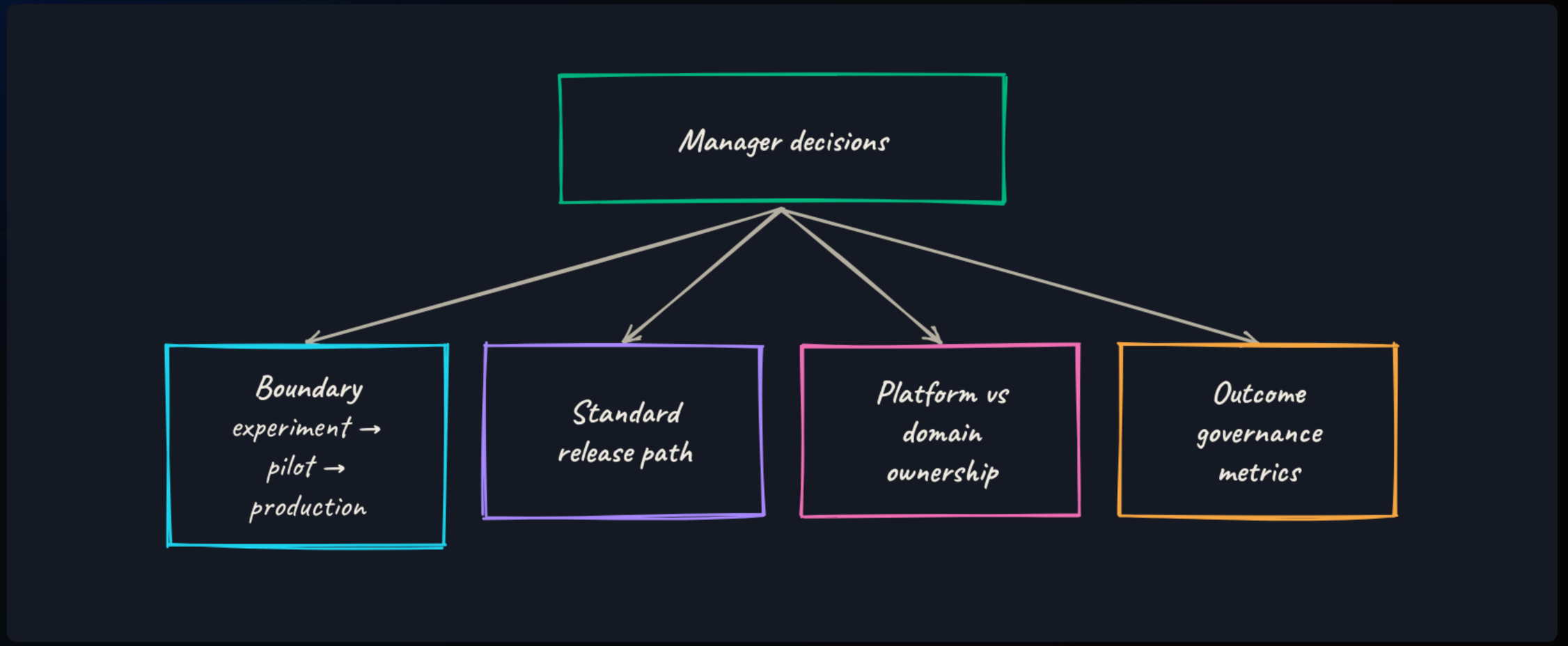
Block 1

01

Management frame

Governed path from data to decision

Manager Decisions



Operating-model options

Managed / self-service	Domain	Hybrid
Common path Platform standards	Hypothesis Quality and risk	Common standard Custom runtime

Evaluate the loop, not the stack

+ What speeds up results

- Hypothesis → live validation.
- Repeatable release path.
- Self-service without losing control.

+ What contains risk

- Reproducible data, features, evals.
- Stop rules, rollback, blast radius.
- Unit cost and path reversibility.



Block 2

02

Data and feature contracts

Which data and features are safe for models

Data platform is not MLOps

Data platform

Catalogs, lineage, access
Reliable path to data products

DataOps / MLOps

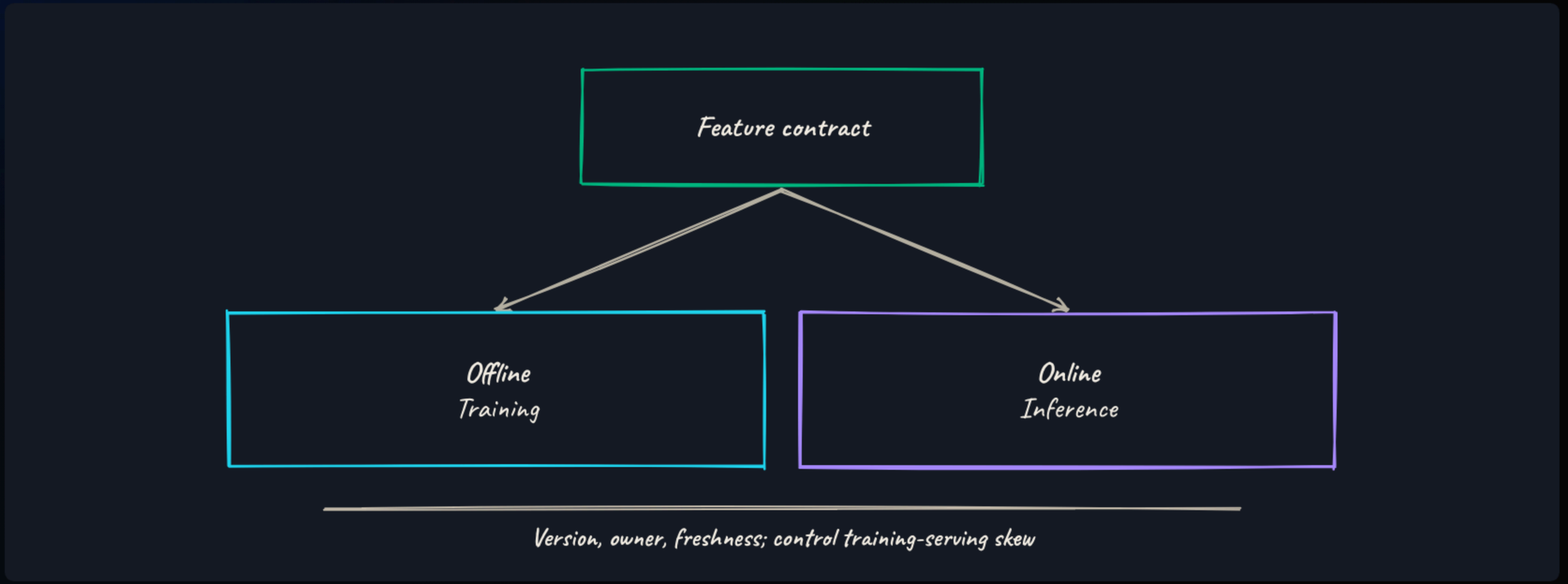
Fit for training,
inference and release
Features, evals, runtime
Release and feedback

The model lifecycle is governed



Feedback returns to the data and the next version

One feature contract for training and inference



Data antipatterns for ML

- Unreproducible training slice.
- Features without owner/version/SLA.
- Offline metrics hide production skew.
- Retrain hides data/policy debt.



Block 3

03

Safe model release

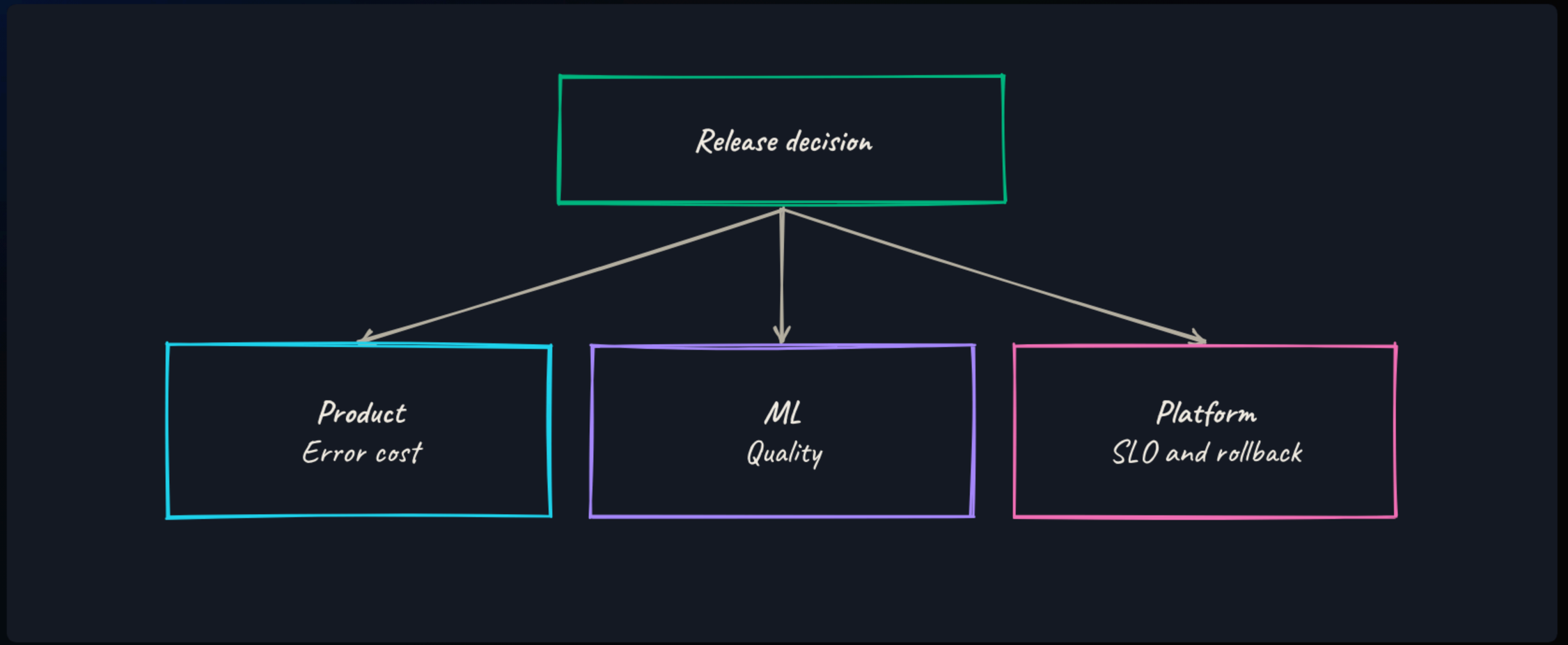
Model release must be governed

Release gates: replay, shadow, canary, A/B, rollback



Quality, runtime and business stop rules; rollback at every step

Who owns a model release



Model Quality Is Not the Decision

- AUC: Area Under Curve.
- Precision and recall measure prediction quality.
- Decision: cost, latency and risk.
- Release card: threshold and fallback.



Block 4

04

Serving and runtime economics

A production model costs money every minute: in features, inference, review and operational support

Online, batch and streaming inference

Online	Batch	Streaming
Reply now Latency SLO	Batch workload Cost and window	Event stream Freshness

Managed or owned runtime

Block 4 · Choice Managed / Own

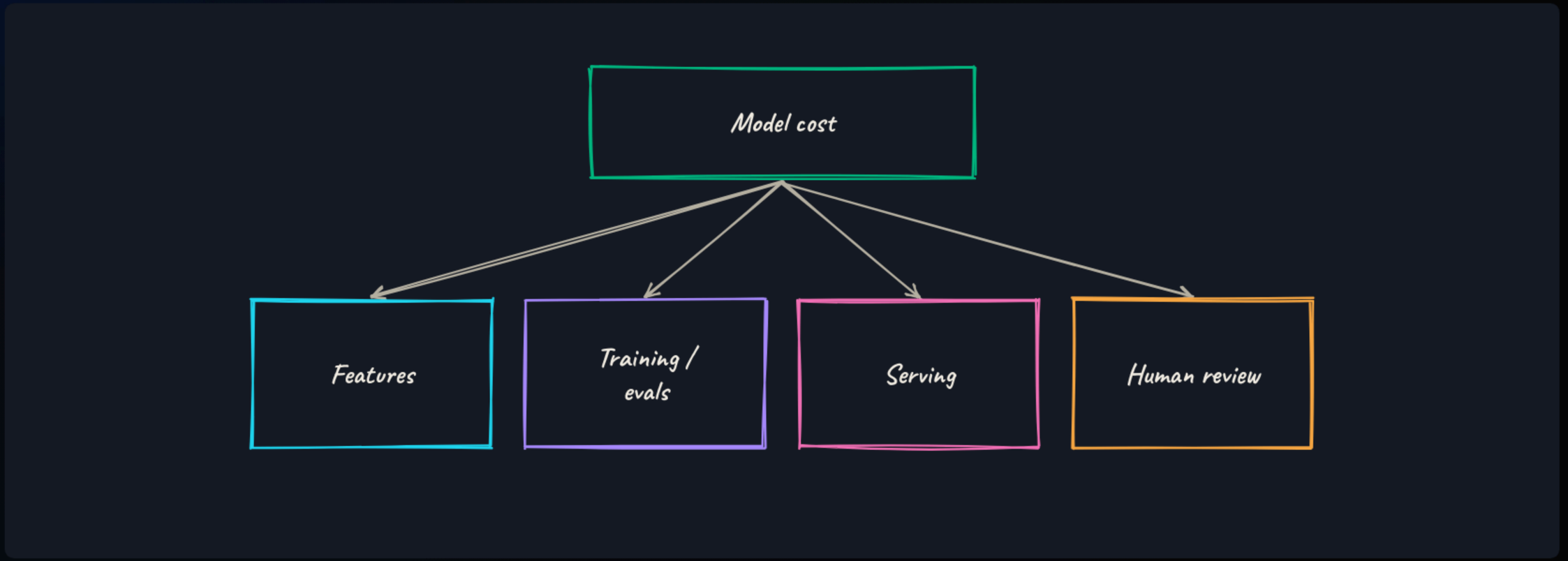
+ When managed is enough

- Fast launch, standard load profile.
- Baseline SLO and monitoring are enough.
- Error cost < platform complexity.

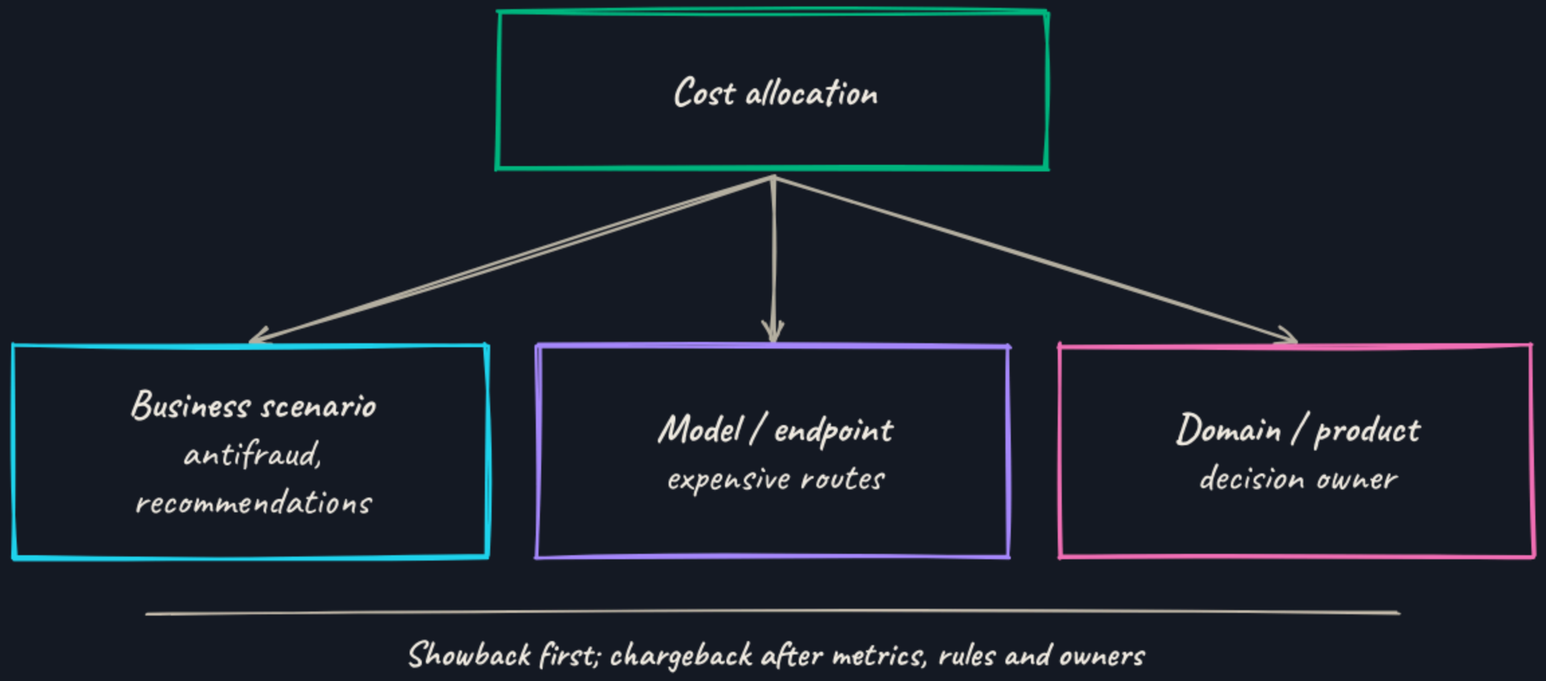
+ When to own

- Custom routing, fallback, strict SLA.
- Predictable volume; unit cost matters.
- Special access/audit/environment policies.

FinOps for models: how to calculate cost



Cost must map to an owner





Block 5

05

Observability, feedback and maturity

After release you need a system that sees degradation, collects feedback and scales without heroics

Human review and quality governance

Block 5 · HITL Quality loop



Correct data, policy or the next model version

Growth requires a more mature operating model

	Pilot	Growth	Regulation	Scale
Data	Fragmented	Contracts	Lineage	Policies
Release	Manual	Canary	Gates	Audit
Runtime	One path	Fallback	Environments	Routes
Feedback	Support	Queues	Reason codes	Closed loop

Final Matrix

Manager’s summary matrix

- Managed service: common path, fast launch.
- Platform: release, observability, SLO, FinOps.
- Domains: hypothesis, error cost, feedback.
- Own runtime: control, scale, economics.

- A production model is a business loop.

Links and materials

Sources References

- Google Cloud — MLOps pipelines.
- Sculley et al. — Hidden Technical Debt.
- Feast Feature Store documentation: ↗docs.feast.dev
- NIST AI Risk Management Framework 1.0: ↗nist.gov