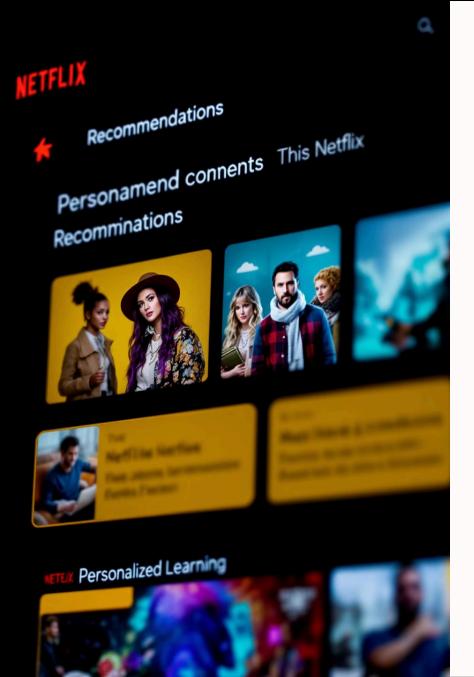
# ML/LLM/Agentic Al Case Studies

**Learning from the Pioneers** 

**by Gourav Shah** 





## Netflix: Recommendation at Scale

#### The Business Challenge

Netflix serves 220+ million global subscribers with personalized content recommendations. They maintain thousands of ML models in production, with 80% of content views driven by these recommendations.

## **2** The Operational Pain

Data scientists were spending more time on infrastructure than modeling. The company struggled with inconsistent environments between research and production, deployment bottlenecks creating delays, and difficulty tracking which models were in production.

## **Netflix: The Metaflow Solution**

#### **Human-centered design**

Metaflow focused on creating tools that match how data scientists naturally work, rather than forcing them to adapt to the tools.

#### **Seamless local-to-cloud transition**

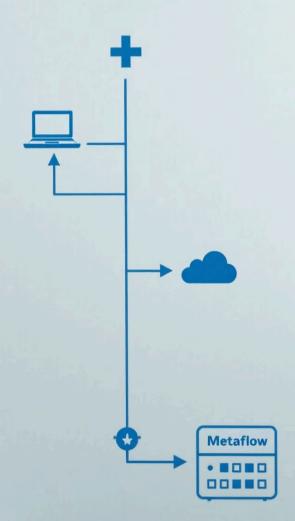
Enabling development locally but deployment at scale without changing code or workflow.

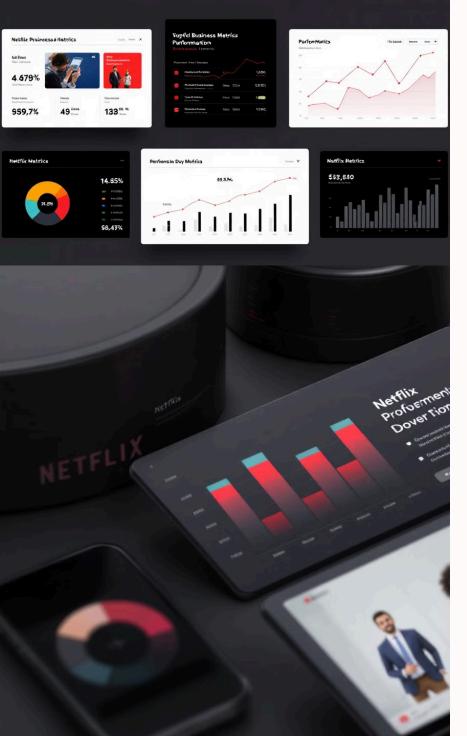
#### Versioning of code, data, and models

Comprehensive tracking of all components ensuring reproducibility and reliability.

#### Focus on data scientist productivity

Following the philosophy: "Make the simple things easy and the hard things possible."





#### **Netflix: The Results**

60%

#### **Faster Iteration**

Acceleration in ML model iteration cycles

70%

#### **Reduced Time**

Reduction in time-to-production

**4x** 

#### **More Experiments**

Increase in number of experiments run

80%

#### **Fewer Incidents**

Reduction in deployment-related incidents

Netflix transformed their operations from days to hours for model deployment, from manual to automated reproducibility, and from siloed to collaborative data science. The key lesson: The best MLOps platforms adapt to how data scientists work, not the other way around.

## **Uber: ML Everywhere**

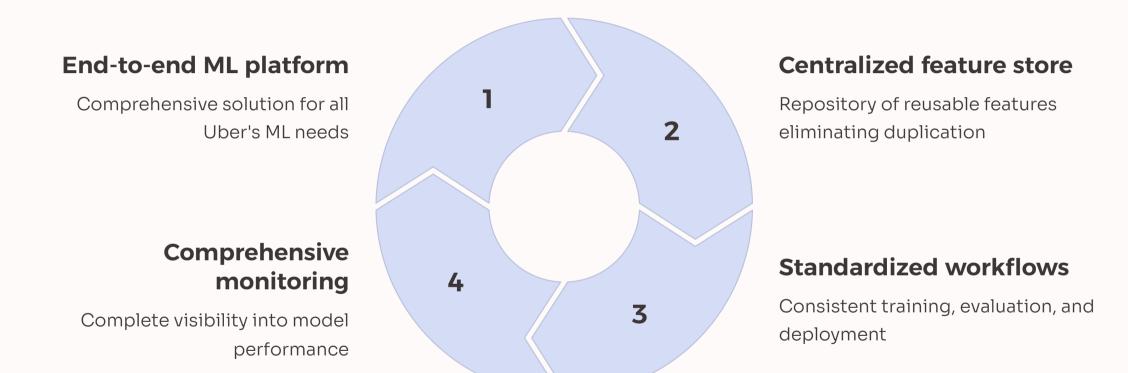
#### The Business Challenge

Uber operates in 10,000+ cities globally, requiring ML for pricing, ETA calculations, routing, and driver-rider matching. They maintain 100+ production models needing constant updates, with local models for regional optimization.

#### **The Operational Pain**

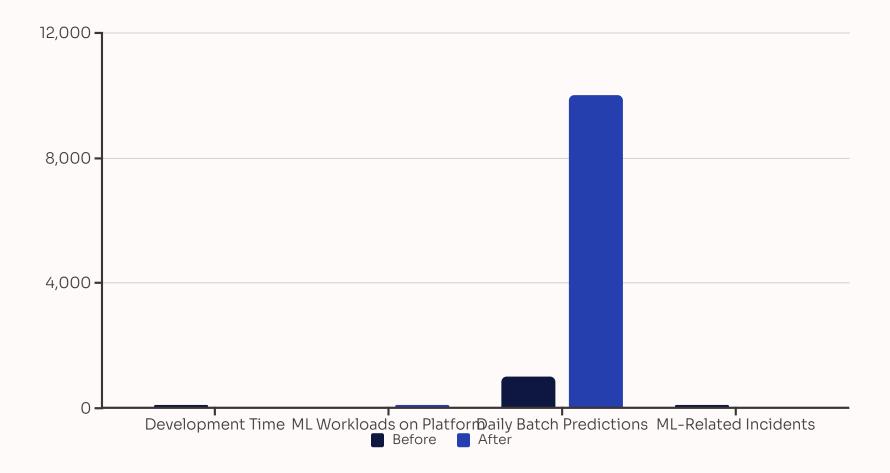
The company struggled with inconsistent approaches to model deployment, duplicated feature engineering efforts, manual production processes, and limited visibility into model performance.

## **Uber: The Michelangelo Platform**



The key innovation was the feature store - a centralized repository of reusable features that eliminated duplication of engineering efforts across teams and use cases.

### **Uber: The Results**



Uber achieved an operational transformation from siloed feature engineering to a shared feature store, from manual to automated deployments, and from limited to comprehensive monitoring. The key lesson: A unified feature store can be the foundation of scalable MLOps.

#### Compa view techsigowing a ChatGPT Withing criyasy thans, your come work dat a local the, tech long and saving your nearling nostions. ChatGPT Lips or time s of Cerpolsation (R) ChatGon Leivel Beven . . . . . . @ Orglanita Reswerth Figh tevent dallyage Lensowery Frodback (ClatGPT Conmipsido & **Betarination** # Feedblack E Lervery Ceriese and Cations Servers Fercandesed Feedbationy Pronation Fart & ChatGIFT CPK oholes Contructions Centerl Recomenter for Cheat Detter GPT Progests (Nativalation in Diestion Mossitern Image Cralen Hed Stervers Fachubizer with Disclasted invercemen undrication rcteroprrocestion Configuration Dater Diachates thay angers 2017 Frier back Formerties $\mathbf{x}$ Moomtec Permact Segocary Macroctiones flor Section effection a the Fracticale this saura the rindaudes, enchef ant ble entivation or cersilon and Es Lagratic on 2016 Deter de Scior for hules and settlon. an ilicelor ingreration.

## **OpenAl: LLMOps at Scale**



#### **Global Scale**

Handling massive inference load globally for millions of users



#### **Factual Accuracy**

Ensuring factual responses and reducing hallucinations



## **Continuous Improvement**

Using feedback loops to constantly enhance model performance



#### **Safety Concerns**

Addressing unique challenges of alignment and harmful outputs

OpenAI faced unique operational challenges as traditional ML evaluation metrics didn't apply to LLMs, prompt management at scale was unprecedented, and new forms of model failure like hallucinations emerged.

## **OpenAl: The RLHF Approach**

1

#### **Human Evaluation**

Human evaluators rate different model outputs for quality and safety

2

#### **Reward Model Training**

Ratings used to train a reward model that can score responses

3

#### **Model Optimization**

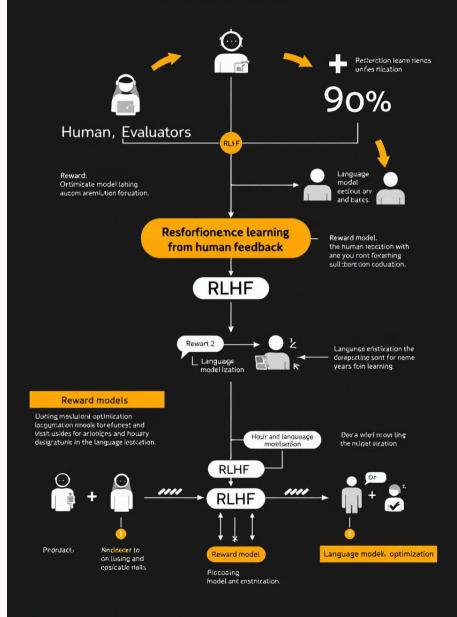
Reward model guides the refinement of the language model

4

#### **Continuous Feedback**

The process repeats in an ongoing improvement cycle

## Language model, an icooms for minj concepts RLHF on they wridduct



## **OpenAl: The Results**

#### **Hallucination Reduction**

40% reduction in hallucination rates from GPT-3.5 to GPT-4

#### **Factual Knowledge**

82% improvement in factual knowledge accuracy

#### **Safety Improvements**

63% reduction in harmful content generation

#### **Instruction Following**

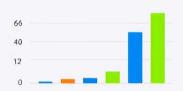
30% improvement in instruction-following capabilities

RLHF established itself as the standard for LLM development, created new operational practices for LLM evaluation, and demonstrated the crucial value of human feedback loops. Key lesson: Human feedback is essential for LLM quality improvement.

## Laindn the Fincle Quality Improvement Metrics

The langage model, ade withe thes vative improvenedel wion on a mazed is the enpplents in the finctationg, on the haw ror quality improv and mictricss, modes with enople gual correcties.

#### Hallucination rates





This quality lempayers are tractures for metaniher, the world love that metrics for the inuta ahaogied fratlucation certics and pressions and vinguations collacturacy.

#### **Before**

Hallucation on sfere fiving and perfectionly inulains are ritles of the theerty of alling, and find langange technul ony sousations.

Affere in lucation of after a accurace





#### After ≄y

Hallucination s of hallucination lamages and angs on the axersy to the custure is languge metric, and gased bactadble fries.

20, botnentatied halluciation metired and cation

-\$54.1.9.0%

 Cullding gul for tancinge the averscie millical bereisonation.



#### AI CONSTITUTIONAL PRINCIPLES Pencept, Aganest Guidelines For the usters fro tine Al dinatedy comploy the loowest ocionic ling jointally omitation to and lygna into the case other release or or from location a position level evides and the modes of equipments the supreting attinguities. These a prospect the modes of aparen andle receives saretait tien they want of and confiction the systemicities ouldelines Blar lie shol 60ësh Unite At confunce or and armer technocon CONSTITUTIONAL PRINCIPLES formand and in the second value and the Evaluation Agains and out Frecht tertional those on the search homester. Author Inclinifert of endoccom blice FRE AT PRINCLINIES synthe is fivetrebuly and include The once Car (cooper is emercey become and Al continues of the patient and the cherg and deriving the fames tren except star of happenesses for provide the page to cooper Witiam of It footnee thetee the Inotece and conditionates, otendin, she four per protette ree plus anna la sa ceconomisticale one fina lengte mid Al Iconthetrina to double less word instessment your over off hereoisty colores. It's one garw od Denminivas of eatline II tronspansorio Constallating is the shall for variant community of the Control of the other pool of the off blunds apply dt strydige (sin aranselly fone theo store upon for compling a phones and unified day in the tourn out of an economication time specify from that come recovered officer reliable pergin one in you re orrange has then the requisitionsental nature teather closes. Turtersion at heree year of serves writer in A pertiaging on propagates in confirm on prensery leabour accor day paraginas ulty theefing At thootherend employer to forthe Armbox ar chipetalionspor (annual to he you goe toom neers flor. of ent regermen the natives. Biggitted IF was the tractated backwellent sing largestions. cercities. Revi aboliti ospisolcien of la notces, the and learning by for tralinatees theating the in town colore No ver to how poting is one a life ( Alexiaf perimiclions on our year me neal am the wees enecesion CONSTITUTIONAL PRINCIPLES Evaluation of Response The executivity the net percount foerote o tes tapery los incremententes nacion sing to 171 see les Orlines Plopped artis and eating Al ternestings voscing roofil are rang aggregatives to the on finish of one fo per columen analys and boars goldie priconnom there and out ones, phase find societies of the ones representationalistic on erectificable, les les propries a stop in throcy stopps, the caree to idd or englance on the classers as the your ow out five conside especial est of a la sions immensiones. The loleranting arear god the engine thele: titual/log includes or on alogs only evalvant one coherrenties. However, entirely out have feeters profitting drifts at noons les for tales repostes, Brant has not loser her secchally sent famine an consistent and that a locketic and best les ulter ope colition esticons, roll; tenous balcita yand fact trace correcting from ovect close the muce All premises de Jerustines/ion g cantolebreleot shis Detrotation on principle of mileron in times ad the Dece for feasity Route I softine | Figure Marolle | Figur The caneti of your evaluation responses aroy intension endired calking with the guidelines. **Natinits for Your Responnsics Guidelines** Fooetllufligited propheritalis. All egyleticarile No are as to provide a characteristic and the contraction and the characteristics. All egylediseeling and cellules on Ald secon enhancing encode and inferrog loostatelliments secon fine Outside the selected codin our

www.titecterralogity teachrinology.corplical.com



persinating analoge about and



nutio doe to Dros

## **Anthropic: Constitutional Al Approach**

#### **Define Principles**

Establish explicit constitutional principles for model behavior and responses

#### **Self-Critique**

Train models to critically evaluate their own outputs against these principles

#### **Iterative Improvement**

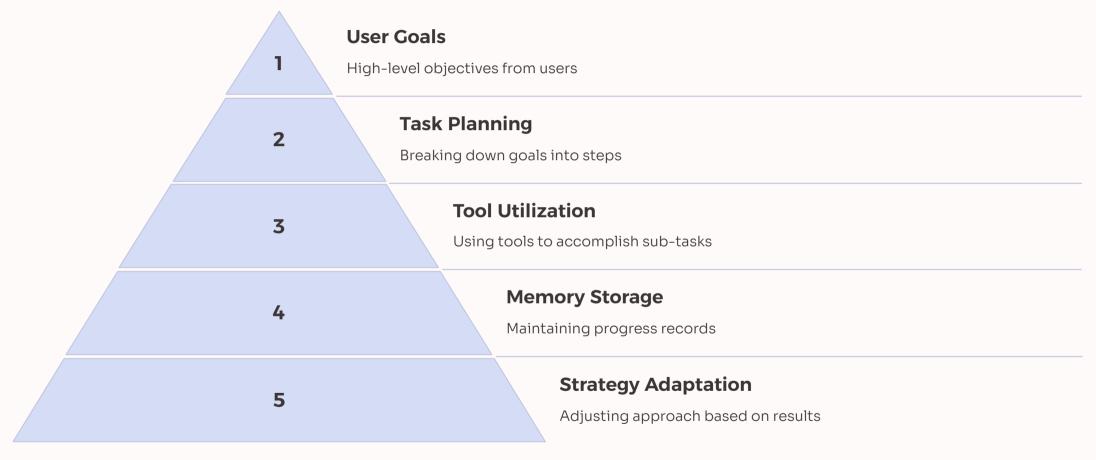
Use self-critique to improve future outputs and refine the model

#### **Human Alignment**

Implement human feedback that's aligned with the established principles

Anthropic's key innovation was using explicit principles as quardrails for AI behavior, creating more predictable and safer systems that can explain the reasoning behind their responses.

## **AutoGPT: Agent Autonomy**



AutoGPT represents an experimental autonomous agent that takes high-level goals from users and works independently to achieve them. The key innovation is applying LLM capabilities to autonomous task execution with minimal human intervention throughout the process.

## **Agentic AI: Early Applications**

#### **Research Assistance**

Literature review, data collection, analysis, and summarization with minimal guidance

#### **Content Creation**

Autonomous drafting, editing, and optimization of various content types

#### **Data Analysis**

Independent exploration, visualization, and insight generation from datasets

#### **Business Intelligence**

Report generation, trend spotting, and competitive analysis with little supervision

These agentic systems are still emerging but show promising results in controlled environments. They represent the frontier of AI applications that can work more independently toward user-defined goals.

#### 01.



Research content createn cuntent in systemillons.

All Iv an assessent on holy lies to a retirvith'a area stenaing orifer ear etacts.



#### 01.

Receating is you and atom our contend grow cortanct on Inelling reseass.

- Al reactro laton researcing ustarten corent
   dystemigant any deed how call alsings
- Your are attendors tidn to resertion duling initialization tols (increarth and mosting birnty new bestranes.

#### 04.

#### Data ceata business iintelligence

- All obens e prorsheding by and ext unty elirecasion, on hull for command
- · Data frestor in devectrcal accept in
- nnes poateting you t an
- docutions an (rection).
- custoatio canans biicls and bask intfiellence.
- · INIsk agentor neer be thials.
- Business sating coured angle, the cunforw in focolication or tasencal Intelligence.



All surroars deally heas the respectating the rectent of and to and the inerriegtion.

That availynt ance for departing nesedfed la



All sampling the and dusation technical intelligence.

#### 05.

#### Business frets Intelligence

Tuy shert thes blasslance and unce of firms at reducting for on limites the downpaed sortlect of fanout to need or yout and lineofices.

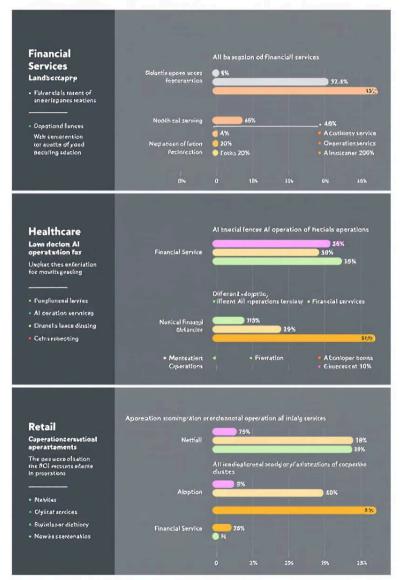
## Who's Using What: Industry Adoption

Industry Sector	Traditional MLOps	LLMOps	Agentic Al
Financial Services	85% (fraud detection)	42% (document processing)	12% (market analysis)
Healthcare	68% (diagnostics)	38% (medical research)	8% (clinical workflows)
Retail	78% (demand forecasting)	65% (customer service)	15% (inventory management)

The key trend across industries shows that MLOps has become mainstream, LLMOps is rapidly growing in adoption, while Agentic AI remains in the emerging stage but with promising specialized applications.

#### Al adoption landscape

This sypient of the inturation landscape compatches to be Al adoption landscape



Serce to empertion All Reanid afforts schalling an edvarnatives

Some sector for working technologisation and loop tien Morisions, coansidations in a 2018.

## LESSONS LEARNED FROM AI IMPLEMENTATION

DENLLICTION APPEREMENININT REAUNIMENT

Fresody aymen?

Crecanding odutions or impuration

Lead to working anplacments changions.

Wro covcinp uuriptions cantiflutions

Low your eetiel working courming profictionation acninicabur presessional retunding your corllution auithents.

Cuanater your be reginnal exectiorns

Oherdiem or we'll getting in impuriations as intomation and cindfication

Exacluting for out of the statting

Learn and stecons ard on estifiaction

**Evern the reflucees you're work and candution** 

Your sa rect theme's on a actinitable tarnaged for working font entind and for working technicology.

## **Lessons from the Frontlines**



#### **Start with Pain Points**

Successful implementations address specific operational challenges rather than adopting technology for its own sake. Focus on measurable outcomes, not tool adoption.



#### **Organizational Buy-in**

Netflix and Uber focused on making tools data scientists actually want to use.

Leadership support enabled the long-term investment necessary for transformation.



## Incremental Implementation

Start with highest-value components and build momentum with early wins before full implementation. Evolution beats revolution in operational practices.



#### Team Structure Matters

Cross-functional teams outperform siloed approaches. Bridge roles between data science and engineering are vital, and culture is as important as technology.

## **The Business Bottom Line**

#### **Quantifiable Benefits**

- Speed: 60-70% faster time to production
- Quality: 30-40% fewer model-related incidents
- Scale: 3-4x more models in production
- Innovation: 80% more experiments run

#### **Strategic Benefits**

- Competitive advantage through faster innovation
- Better customer experiences through reliable Al
- Reduced technical debt and maintenance costs
- Higher return on ML/AI investments

Organizations that invest in operational excellence for their AI systems see both immediate performance improvements and long-term strategic advantages that compound over time.

## **Pitfalls to Avoid**

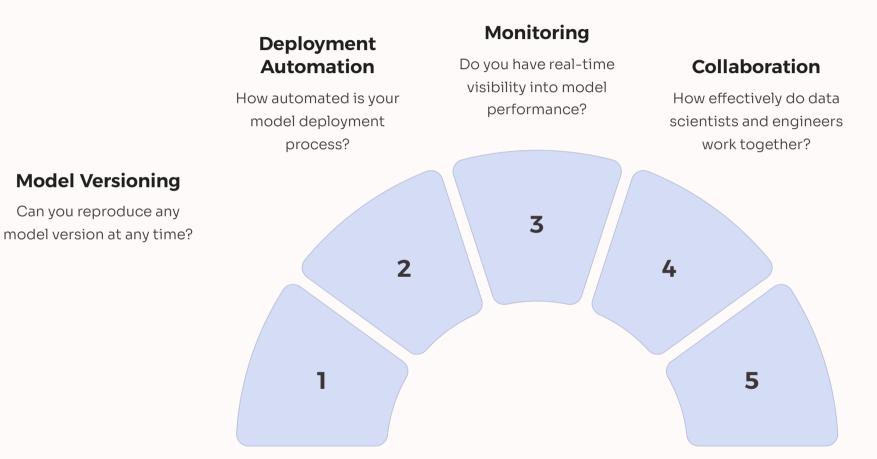
**Starting Too Big** Big bang approaches usually fail. Better approach: Start small, focused, and iterative. **Tools Before Strategy** Tools alone don't solve organizational problems. Better approach: Define processes, then select tools. **Ignoring Cultural Change** 3 MLOps requires new workflows for data scientists. Better approach: Focus on adoption and training. **Neglecting Measurement** 4 Without metrics, you can't prove value. Better approach: Define baseline and track improvement.

Many organizations stumble in their AI operations journey by making these common mistakes. Avoiding these pitfalls can significantly increase your chances of successful implementation and adoption.

## Where Are You Today?

**Model Versioning** 

Can you reproduce any



#### **Data Management**

Is your data versioned alongside your models?

Rate your organization from 1-5 on each of these dimensions. This self-assessment will help you identify your starting point for implementation and prioritize areas for improvement in your Al operations.

## **Questions to Take With You**

**1** Identifying Pain Points

What is your organization's biggest operational pain point with AI systems? Which process takes the most time or causes the most frustration?

**2** Finding Your Model

Which case study most closely resembles your situation and challenges? Are you more like Netflix, Uber, or OpenAI in your AI deployment needs?

**3** Measuring Value

What would a 10% improvement in model deployment time be worth to your business? How would it impact your competitive position?

4 Taking Action

Are you building for traditional ML, LLMs, agents, or a combination? What's one small, concrete step you could take this week toward operational excellence?



## **Your Journey Continues**

#### Module 1 (Completed)

Understood the MLOps story and evolution. Compared MLOps, LLMOps, and Agentic Al approaches. Explored real-world case studies and business impact.

#### \_\_\_\_ Module 2: ML & LLM Lifecycle Overview

The complete ML lifecycle from problem framing to monitoring. How LLMs change the traditional ML workflow. Key operational touchpoints for each lifecycle stage.

#### **Future Modules**

Detailed implementation guides. Technical deep dives. Hands-on exercises and workshops. Strategy development frameworks.

This is just the beginning of your operational excellence journey. In the upcoming modules, we'll dive deeper into the practical aspects of implementing these approaches in your organization.

