



Module 11

Kickoff Workshop



Icebreaker

ML Deployment in Action!

Which of these challenges do you think is most common when deploying ML models at scale?

- ❖ Managing data pipelines
- ❖ Model performance degradation
- ❖ Keeping track of multiple versions
- ❖ Getting stakeholder buy-in



Agenda

- **Welcome and introduction**
- **Module overview**
- **Applying module skills**
- **Project overview**
- **Q&A**
- **Next steps**



Module overview



Your learning journey



Module skills: from learning to applying

Deploy ML model prototypes into live environments



Apply deployment pipelines to move ML models from testing to production, ensuring real-time access and business integration.

Identify and mitigate risks throughout the deployment process



Evaluate model stability and data drift pre-launch to prevent performance degradation and business disruption.

Optimise resource allocation for efficient and cost-effective model deployment



Streamline infrastructure and team resources to ensure model deployments are scalable without unnecessary cost.



Skill 1 in the real-world

Design scalable architectures for ML model deployment



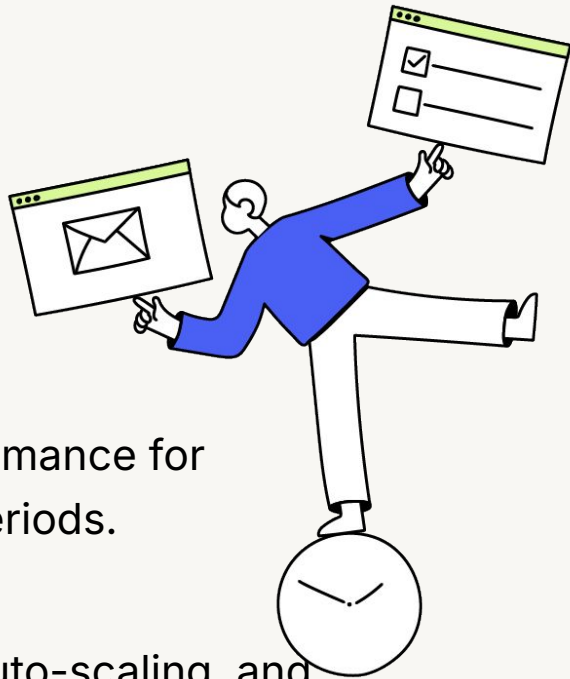
Netflix used a containerised microservices approach with Kubernetes to deploy their recommendation models globally, enabling consistent and scalable user experiences.



This allowed rapid model updates, ensured low-latency performance for millions of users, and maintained uptime during high-traffic periods.



Apply similar scalable practices—like modular deployment, auto-scaling, and monitoring—across different ML pipelines and teams to reduce operational bottlenecks and improve reliability organisation-wide.





Applying module skills to your workplace



Skills application brainstorm

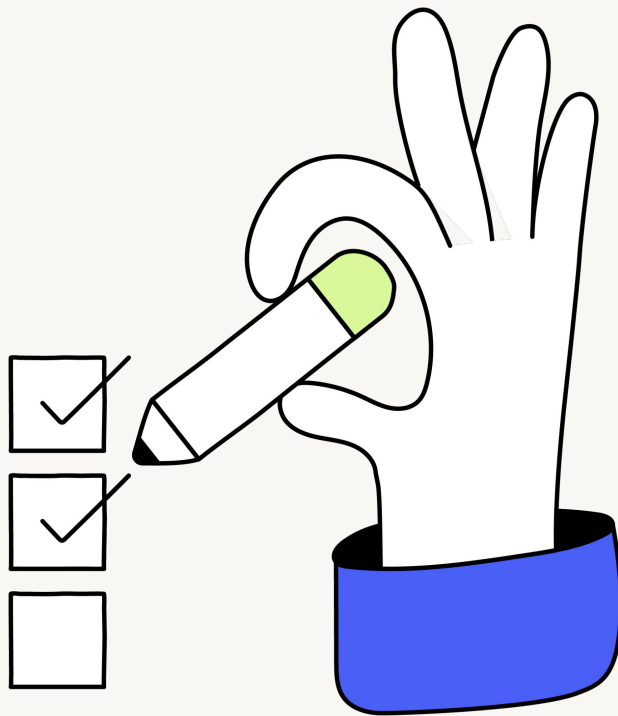
Choose a module skill and think of 3 ways to apply it in your role



Create a post for the group explaining how you'd like to apply this skill this month



Post your explanation in the Skill Application Brainstorm discussion board





Skill application share out

- 1 Review other discussion posts.
- 2 Choose another person's post that you think is a great example of applying the skill.
- 3 Raise your hand to share your thoughts.





Project overview

Phase 3 project: Deploying ML models

Project objective: Deploy an accessible, scalable, secure, and cost-efficient ML model while maintaining and adapting its performance and providing clear communication.

**Module 11
milestone:**
Model
Deployment

**Module 12
milestone:**
Monitoring,
Maintenance, and
Continuous
Learning

**Module 13
milestone:**
Communication,
Documentation,
and Inclusive
Collaboration

We are here!

Module 11 milestone: Model deployment

Phase project objective: Deploy an accessible, scalable, secure, and cost-efficient ML model while maintaining and adapting its performance and providing clear communication.

Milestone task 1



Deploying the ML model: Explore various ML model deployment strategies to determine the most effective approach for ensuring model accessibility, scalability, and security in a production or testing environment. Implement mechanisms for version control, rollback, minimal downtime, and optimised performance.

Milestone task 2



ML/AI platform architecture and resource allocation: Justify the selection of computational resources, including cloud services, GPU acceleration, and storage options, and compare different ML platform architectures to optimise infrastructure for efficiency and cost-effectiveness. Assess security risks and outline mitigation strategies.

Milestone task 3



Risk management in deployment: Identify potential risks associated with ML model deployment and implement comprehensive mitigation strategies to ensure model robustness and reliability. Develop monitoring and detailed rollback procedures to maintain performance and address failures.

Module 11 milestone: KSB requirements

Phase project objective: Deploy an accessible, scalable, secure, and cost-efficient ML model while maintaining and adapting its performance and providing clear communication.

.

Task 1 KSBs



Model deployment: K11

Task 2 KSBs



Monitoring, maintenance, and continuous learning: S22

Task 3 KSBs



Communication, documentation, and inclusive collaboration: None



Questions?



Closing out

- **Action items:** Async Unit 1
- **Have questions?**
 - Chat with Atlas
 - Reach out to your Cohort Coach
 - Sign up for Tutoring
- **Log your OTJ hours**
- **Take the Workshop Survey** on My Multiverse

multiverse