Canonical

A guide to MLOps

Executive summary

With an adoption that doubled compared to 2017 based on the McKinsey State of AI report, AI is going through a shift. Initiatives are going past the experimentation phase. Some noteworthy changes in the space include:

- An increasing number of capabilities being used by organisations as part of their AI initiatives.

- An increasing level of investment allocated to machine learning projects, which goes hand in hand with a higher adoption rate.

- More interest in collection, governance and ethics, aiming to ensure compliance for production deployments.

- Stable, secure, scalable tooling is a priority for enterprises. Having AI that enterprises can benefit from is critical.

- AI is more affordable and performant, with needs that are better addressed, tools that mitigate risk and an ecosystem that is better integrated.

How do you navigate these changes in the fast-paced world of AI? The answer lies in building a well-oiled MLOPs practice. Much like DevOps changed the world of software development, MLOps will enable maturity for AI. This guide offers some guidance for data scientists and ML practitioners who are looking to take their AI models from the experimentation phase to the production stage with MLOPs. We will start with an overview of what MLOps is, and cover its benefits and basic principles. We will then explore the typical MLOps lifecycle, available tooling and how Canonical can help you in your journey.

Contents

An overview of MLOps

MLOps is slowly evolving into an independent approach to the machine learning lifecycle that includes all steps – from data gathering to governance and monitoring. It will become a standard as artificial intelligence is moving towards becoming part of everyday business, rather than just an innovative activity.

MLOps benefits a wide range of functions within an enterprise that are impacted by machine learning initiatives. While data scientists can actually focus on modelling, the data engineering side has a tool to create pipelines and automate processes. At the same time, at a higher level, MLOps is a practice that ensures the stability that solution architects care about and the security that IT auditors demand today to get behind ML initiatives.

From a business angle, MLOps is a practice that brings reliability to machine learning projects. The goal is to have trustworthy projects that perform as expected in production. Depending on the use case, the outcome of an ML project can be different, but MLOps optimises associated functions or tasks, automates processes and plays an important role in time and cost savings. MLOps can be applied in any industry and any enterprise.

What is MLOps?

MLOps is the short term for machine learning operations and it represents a set of practices that aim to simplify workflow processes and automate machine learning and deep learning deployments. It accomplishes the deployment and maintenance of models reliably and efficiently for production, at a large scale.

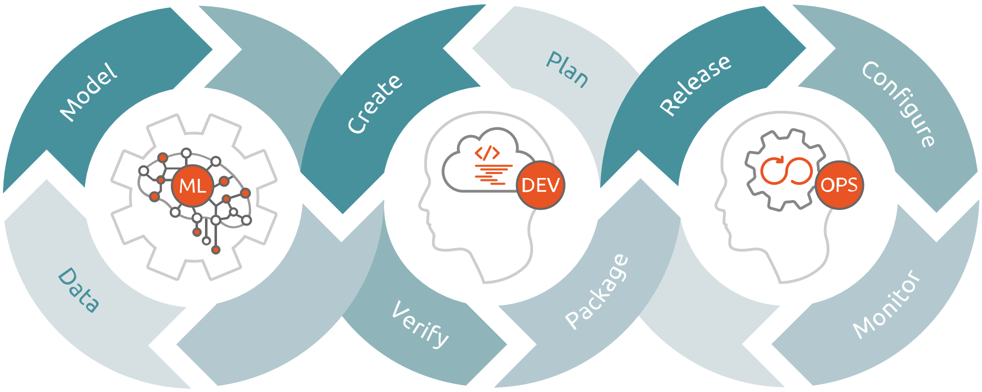

Image source: Nvidia

MLOps includes, besides the best practices, processes and underpinning technologies. They all provide a scalable, centralised and governed means to improve machine learning activities. It is a mix between machine learning development and operations.

From an ML perspective, it includes specific activities, such as model development or data gathering. It then goes further to development, where packaging, model creation and verification play an important role.