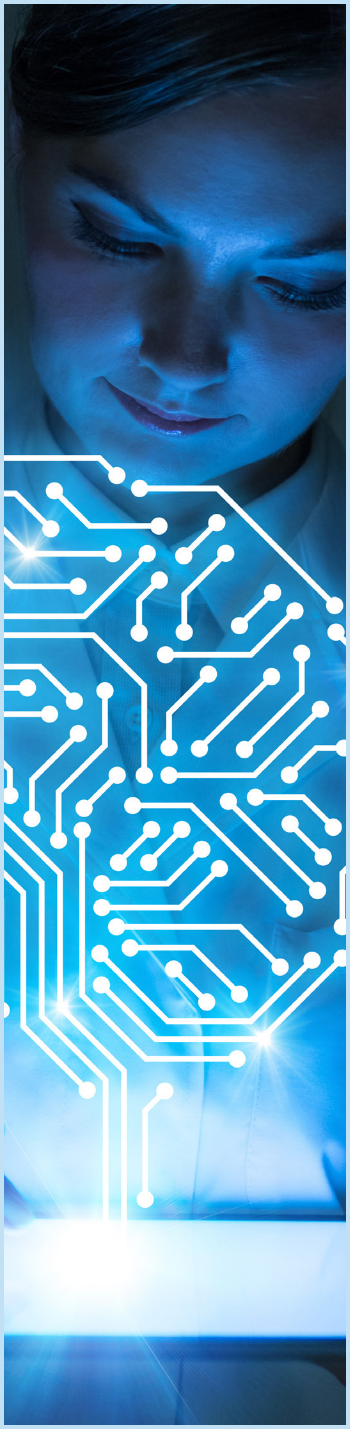
VIEW POINT



ARE YOU PREPARED FOR ANALYTICS AT SCALE THROUGH AI/ML?

Abstract

Organisations across sectors are making significant investments in artificial intelligence (AI) and machine learning (ML) to gain business insights from all the data collected through various channels. Implementing AI at scale is the next step. However, the path is not a smooth one. Replicating AI/ML experiments on a large scale requires technical expertise and infrastructure on a large scale too. Moving from pilot projects to enterprise-wide implementation is difficult. However, there is light at the end of the tunnel. Machine Learning Operations (MLOps) is the new discipline that can help organisations overcome most hurdles. MLOps focuses on the organisation's AI scaling processes, teams and tools in a strategic and integrated manner and guides it to success.

Infosys®
Navigate your next

AI is no longer a domain of the digitally driven alone, it is used by the manufacturing industry, healthcare, financial sector and almost every other sector. In a recent report, Deloitte showed how organisations across industries were using AI to boost their value. Clearly, the rewards of using AI are greatest when the technology can be implemented at scale, beyond pilot projects. In fact, Gartner recently reported that over 75 % of organisations will move from only piloting AI technologies to scaling them by the end of 2024. This is also when the honeymoon ends and life begins.

Running AI at scale is not easy. It requires a very efficient production line where reliable and risk-compliant AI models are designed quickly. However, problems crop up as implementation of AI increases. If an AI or ML model makes an error, and the error is not identified immediately, the magnitude of the error could quickly multiply leading to unfavourable results. This hurdle is forcing many organisations to start development from scratch every time they decide to use AI to gather insights.

To counter such effects, organisations have started adopting a new discipline called Machine Learning Operations (MLOps). It can be defined as a set of best practices across industries that bring together people, processes and technology for implementing analytics at scale. MLOps helps organisations set up best practices and tools to enable AI to be implemented by operations quickly, safely and efficiently.

Standardise processes

The first step in scaling AI processes is to standardise the approved AI models. Standardising helps streamline, implement and refine the AI models. Creating and building AI models and algorithms is a long process. Data scientists, the backbone of any ML project, first prepare the data and create features, then they train and tune the model and its parameters. They also ensure the model works. The organisation's IT and software engineers then deploy it, all the while monitoring its performance and output to ensure the model works during production. Finally, a governance team must ensure that the AI model is compliant with all necessary regulations.

Standardising involves building models that can be repeated and operated by a well defined process. This is the stage that many organisations struggle with. Finalising a bespoke process that can be repeated and used as a production model is fraught with errors unless every possible weakness is investigated and perfected. Points at which the AI project is handed over to another team on the lifecycle must be standardised carefully so that different teams can continue working on it without fear of interruptions.

The MLOps team must collaborate to define a standard process for developing and implementing an AI model and also provide the tools — such as a standard set of libraries — to support adoption of the process.

Data organisation is an important parameter too. Scattered, inaccessible or data with limited access, are all bumps on the road. Data must be centralised and accessible under accepted rules of governance for best results. Data insights and business goals must have a bearing on each other too, or else the process can become insignificant and superfluous to the primary goal of the organisation.

Build the right teams

To scale AI successfully, organisations must build specialised teams to focus on different strategic areas of the project cycle. Each aspect of building and implementing an AI model requires different skills and each team must be able to manage a particular aspect. For example, data scientists, ML engineers, compliance experts and risk managers can all manage distinctly different areas of a project, and organisations must be aware of this fact. As AI is scaled, more and more expertise is required.

Pod model and COE model

As organisations scale AI, two kinds of team structures work best:

- The pod model where a small team comprising a data scientist, data engineer and a software or ML engineer develop the AI model; and
- The Centre of Excellence (COE) model where data science experts are distributed among various product teams. Common standards and best practice guidelines are outlined here.

Both structures have proved themselves although the pod model leads to knowledge being walled off from the outside teams while the COE model leads to excessive knowledge distribution – not always a positive effect. Governance teams work best when outside of both these team structures.

Pick the right tools

This is not easy given the nascent nature of the standardisation process. Data scientists work with about a dozen specialised tools to build a single AI model. IT and governance work with a distinctly different set of tools. The two toolchains rarely connect and that makes building a repeatable workflow rather difficult. It slows down the process of scaling AI and errors can creep in easily.

Clearly collaboration is the only answer. For fast and accurate results, all stakeholders must be able to work on tools and platforms that connect well and also support creativity, safety and speed. The three groups that need to see eye-to-eye, at least in terms of AI development, include:

- the data scientists who build the AI models,
- the IT teams who run the models during production and also maintain the infrastructure, and
- the governance teams who ensure all regulations are complied with.

Factors that must be considered while picking MLOps tools are:

Compatibility

New tools that are picked should be compatible with the existing AI ecosystem, or should be easily extensible to provide the necessary support. Organisations moving to the cloud should find tools that work under hybrid conditions.

Coexist with data science as well as IT

Data scientists and IT teams usually have very different needs when it comes to tools. Data scientists need to be free to choose libraries independently and to work without constant support from the engineering or IT teams. Similarly, IT needs to work on a platform that is fitted with the right restraints and which makes sure that real-time implementation follows IT-approved directions. The perfect MLOps platform would fulfil the needs of both teams but given the few chances of that, both teams often end up picking different platforms to work on. The challenge is to connect the two.

Governance and collaboration

The perfect MLOps tool would be one approved by data scientists, engineers, as well as those in compliance and governance. While the speed of collaboration between scientists and engineers determines the speed to market, collaboration with governance ensures feasibility of the product.

AI governance does not mean ensuring security and control alone, it also means alignment of the application with the organisation's culture, ethics and business goals. Decisions made by the AI model should be unbiased and trustworthy. Consequently, MLOps tools are often fitted with checklists to ensure responsible AI usage, documentation and workflows.

Unlocking the potential of AI

Successfully scaling AI requires smart and strategic decisions and the patience to ensure that every model developed adheres to all the regulatory needs besides being technically strong. Shortcuts of any kind are likely to backfire in the long run. While licensed APIs and pre-trained models are undoubtedly valuable shortcuts, every organisation must focus on how best to operationalise AI under their own operating conditions.

On the other hand, AI tools and technologies have improved dramatically over recent years making AI application lifecycles fast and efficient. It is up to organisations to pick and utilise the right technologies as they prepare to scale AI models.

* For organizations on the digital transformation journey, agility is key in responding to a rapidly changing technology and business landscape. Now more than ever, it is crucial to deliver and exceed on organizational expectations with a robust digital mindset backed by innovation. Enabling businesses to sense, learn, respond, and evolve like a living organism, will be imperative for business excellence going forward. A comprehensive, yet modular suite of services is doing exactly that. Equipping **organizations with intuitive decision**-making automatically at scale, actionable insights based on real-time solutions, anytime/ anywhere experience, and in-depth data visibility across functions leading to hyper-productivity, Live Enterprise is building connected organizations that are innovating collaboratively for the future.

For more information, contact infosysbpm@infosys.com

Infosys®
Navigate your next