UNSW
Institute for
Cyber Security

Securing Machine Learning Operations (MLOps):

An organisational approach

Prof Debi Ashenden, UNSW,

Dr Hung Nguyen, University of Adelaide,

Prof Ganna Pogrebna, University of Sydney

[i] https://www.iiot-world.com/industrial-iot/connected-industry/why-85-of-machine-learning-projects-fail/#:~:text=According%20to%20Gartner%2C%2085%25%20of,way%20it%27s%20applied%20to%20projects

[ii] https://www.forbes.com/sites/cognitiveworld/2022/08/14/the-one-practice-that-is-separating-the-ai-successes-from-the-failures/?sh=44956b2d17cb

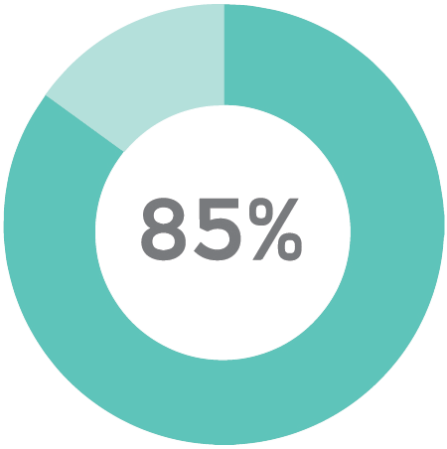
Introduction

AI and ML are being deployed with enormous success across a range of sectors.

Despite their undeniable utility, industry research indicates that a substantial majority of AI/ML projects never get beyond prototype stage, or fail to deliver expected results. A 2019 study by Gartner found that up to 85% failed to deliver expected outcomes[i], while in 2022 Forbes stated a failure rate of between 60% and 80%[ii].

Failure is complex and multifactorial, but security concerns often play a part. This is because many ML models are developed in a research environment or as proof of concept. Security issues are only addressed when the decision is taken to operationalise a model and move it into a real-world environment.

For organisations to get the best possible return on investment from AI/ML, they need to think strategically about operationalisation and security from the outset, and this requires organisational policy that brings together input and expertise from outside technical teams.



85% of AI and ML projects failed to deliver expected outcomes

In many institutions, work remains to be done in terms of establishing end-to-end processes and governance structures to manage security risks effectively.

MLOps governance

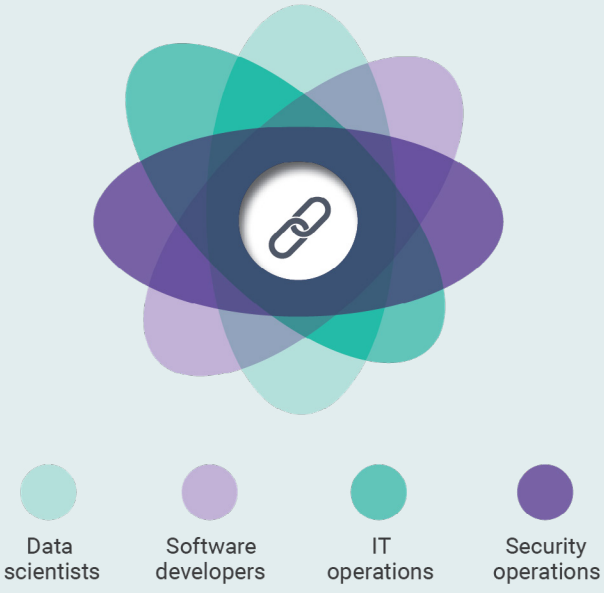
MLOps embrace the lifecycle of machine learning, from model development to deployment. This broad term can mask the vast diversity between different MLOps projects, which range from the application of tried and tested models in a new context or environment, through to the development of new models from scratch. Due to both the technical focus on finding the right toolset to deliver an ML model into production and the uniqueness of each project, organisational decisions about processes, governance and security are often made on a case by case basis. While this might be effective in the early stages of an organisation's use of ML models, or where a model is being developed for a very specific use case, it quickly becomes unsustainable in contexts where there are multiple, highly complex projects.

Due to both the technical focus on finding the right toolset to deliver an ML model into production and the uniqueness of each project, organisational decisions about processes, governance and security are often made on a case-by-case basis.

When an organisation doesn't have established structures of governance and procedures for the management of MLOps projects, it often falls to those working on technical implementation to make decisions without sufficient support. While these decisions may be made with the best of intentions, they can nevertheless lead to exposure to increased risk of security breaches or other forms of project failure, simply because of a lack of strategic security expertise.

Security risks in MLOps

Whether due to a data security breach, legal compliance issues or ethical concerns, failure at any point can expose an organisation to significant reputational damage.



MLOps demands collaboration between data scientists, software developers, IT operations and security, including the CISO. Many ML models are designed in a research environment. While it is vital to identify and address new security vulnerabilities that may be exposed when projects are shifted into a production environment, exposure to other types of risk must also be addressed. First, to avoid potential liability, the way that a model acquires and uses data must be assessed from a regulatory perspective. Secondly, a number of cases have occurred in which models have produced biased, politically embarrassing or discriminatory results.

Whether due to a data security breach, legal compliance issues or ethical concerns, failure at any point can expose an organisation to significant reputational damage.

This risk is aggravated in cases where prototype models are used in production without security risk mitigations, where integration with other organisational systems is weak, or where ML models designed for decision support are in practice used for automated decision making.

In short, MLOps security is not just a question of the technology and tools being used. As in other areas, vulnerabilities can be more effectively managed through a combination of policy, process and personnel across the design, development, implementation, and use of ML models.

In terms of implementation, several challenges need to be addressed.

(i) Developing policy to ensure that all sections of an organisation, including finance, HR, and legal, understand and fulfil their role in supporting the delivery of secure MLOps.

(ii) Generating a holistic understanding of security risks across the MLOps process, and taking steps to mitigate those risks systematically.

(iii) Ensuring that MLOps teams include the right mix of expertise, including data science, machine learning, software engineering and cyber security, with additional support as needed.

Policy development

Security practitioners reading this may be thinking 'we have good security policies already – what's so special about MLOps?' It is true that MLOps security policies share many aims and procedures with existing industry standards, and include procedures for managing access to sensitive data, implementing appropriate security controls, conducting regular audits and assessment, patching vulnerabilities, and responding to incidents. This said, MLOps risk management also needs to take into account additional issues, including adversarial machine learning and the updating and training of models.

Policies must also be tailored to ensure compliance with an organisation's ethical expectations. Model bias, defined as systematic errors or distortions in a machine learning model's ability to make predictions or classifications, can be due to flaws in a model's design or in the data used to train the model. The way in which a model is used in decision making processes is another key ethical issue. Transparency and processesn to continually review datasets and results is thus a key concern in terms of ensuring fairness and protecting an institution's reputation.

Agile as a policy methodology

MLOps policy needs to be agile and flexible in order to respond to rapid technical change, and to support continuous improvement. Agile policy development is rooted in the principles of agile software development. As well as being technology agnostic, this methodology prioritises collaboration, rapid iteration and continuous improvement. Policy makers work collaboratively to continually update requirements, listen to feedback and adjust policy as needed.

Effective collaboration helps manage risk and security, improves internal communication, and builds confidence.

Effective collaboration

Operationalising an ML model is complex and success depends on collaboration between different teams and organisational sectors to ensure consistency in development and deployment. Effective collaboration, in turn, helps manage risk and security, improves internal communication, and builds confidence.

Who needs to be involved in the development of MLOps policy and why?

(i) Finance

Finance departments have an important role to play in ensuring that resources are available to support MLOps projects through their life cycles. This includes funding not only hardware, software and personnel, but also ongoing expenses such as maintaining a digital twin. A digital twin is a virtual replica of a physical asset or system that can be used for purposes including testing, monitoring and optimisation. In MLOps, a digital twin can be an invaluable resource for understanding vulnerabilities and testing and implementing security measures in anticipation of real-world attacks or applications. While the benefits are significant, digital twins can be expensive to implement and maintain. Financial teams need to understand the issues involved in order to make sound decisions about the return on investment at different stages of MLOps timelines, and allocate resources accordingly.

(ii) Legal

Legal departments play a critical role in ensuring that MLOps projects comply with laws and regulations around data privacy, security, and intellectual property. When an organisation uses an ML model that has been designed and developed by third parties (usually either contractors or research teams) there are legal and contractual issues to consider, including intellectual property rights, data privacy, liability and indemnification (particularly with respect to reliance on automated decision making), and service level agreements for performance and updating.

It is also important to address ethical considerations and define an exit strategy that includes procedures for data migration.

(iii) Human Resources

HR has a key responsibility to attract personnel with the right skill sets for MLOps teams. These teams often need a wide base of skills and expertise, and to remain constantly up to date with the latest development in technology and best practices. For this reason, individuals that have both deep expertise in a specific domain and a broad understanding of other relevant areas – a skill set distribution that can be graphically represented as a letter T – are often desirable team members. To give an example, a data scientist with T-shaped skills might have deep expertise in statistical models and machine learning and a broad understanding of software engineering and deployment pipelines. This would enable them to work effectively with software engineers. T-shaped individuals are able to work effectively across different teams and disciplines, and also have a solid foundation of knowledge and expertise that can be built on as new technologies emerge and practices evolve. Tailoring recruiting, role definition and other HR practices to attract and reward T-shaped skills can be key to MLOps success.

(iv) End users

ML models may be used as decision-support tools, which analyse data and provide insight and recommendations, or for automated decision-making, a process in which models make decisions without human intervention based on predefined rules and algorithms. While the use of MLOps for automated decision-making can improve efficiency and reduce costs, it is important to ensure that the potential risks and consequences of automated decision-making are considered and that humans are involved where necessary to ensure accountability, fairness, and safety. Decisions need to be made about whether a model will support or make decisions from the beginning of the development process, rather than improvising use according to circumstances.

In short, as well as helping to ensure return on investment, institution-wide support is key to risk identification and management. Policy developed and implemented collectively is more likely to be comprehensive, effective, and aligned with an organisation's overall strategy and culture. Effective MLOps demand real collaboration between the CISO, data scientists developing ML models, softwaredevelopers deploying and managing ML models, the legal team, HR, finance and end users. By working together with technical teams responsible for MLOps, these actors are critical enablers of MLOps success.

As well as helping to ensure return on investment, institution-wide support is key to risk identification and management.

Recommended actions

- Organisations need to develop a robust MLOps governance structure to manage security risks.
- Finance must understand MLOps and allocate sufficient resources to MLOps projects, including not only hardware, software, and personnel, but also the potential implementation of digital twins and/or targeted experimentation.
- HR should recruit individuals with the required skill sets for MLOps, including 'T-shaped' skills – deep expertise in one area and broad understanding in others. I Legal needs to ensure that MLOps projects comply with laws and regulations related to data privacy, security, and intellectual property.
- Legal needs to ensure that MLOps projects comply with laws and regulations related to data privacy, security, and intellectual property.
- MLOps teams should implement agile policies that are flexible and adaptable, allowing for continuous improvement and rapid response to changing technology landscapes and evolving security threats.
- All involved in MLOps should support a culture of collaboration, especially between data scientists, software engineers, IT operations, and security teams, from the outset of a project, not just when it's time to move the model into production.

Glossary

Adversarial machine learning is a technique that involves intentionally manipulating or adding noise to input data in order to trick or deceive a machine learning model into making incorrect predictions or classifications.

Agile software development is an iterative approach to software development that emphasises flexibility, collaboration, and continuous improvement. Agile development teams work in short cycles or sprints, and prioritize delivering working software quickly, while adapting to changing requirements and feedback.

Decision support/automated decision making refers to the use of technology, including machine learning models, to assist or automate decision-making processes. These technologies can help humans make more informed and accurate decisions by analysing data and providing insights and recommendations, or they can make decisions autonomously based on predefined rules or algorithms.

Digital twins are virtual replicas of physical objects or systems that use real-time data and simulations to model and predict their behaviour and performance. Digital twins are often used in engineering, manufacturing, and other industries to optimise and improve processes, and to identify and address potential problems before they occur.

Ethical expectations refer to the standards and principles that govern the ethical behaviour of individuals and organisations involved in the development and deployment of technology, including machine learning models. These expectations include ensuring that models are transparent, fair, and unbiased, and that they do not harm or discriminate against individuals or groups.

Model bias refers to systematic errors or distortions in a machine learning model's ability to make predictions or classifications due to inherent flaws in the data used to train the model or in the model's design.

Training models refers to the process of using data to teach a machine learning model to make accurate predictions or classifications. The training process involves feeding the model with labelled data, evaluating the model's performance, and adjusting the model's parameters to improve its accuracy.

T-shaped skills refers to a combination of broad, cross-disciplinary knowledge and deep expertise in a specific area. The "T" shape represents the breadth and depth of skills, with the horizontal bar indicating the ability to collaborate and communicate across disciplines, and the vertical bar indicating the depth of expertise in a specific area.

This research was funded by the Australian Government through the Australian Research Council (ARC) under the National Intelligence and Security Discovery Research Grant (NISDRG) NI210100139.

Further Reading

A useful introduction to AI/ML and data science:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/850129/The_Dstl_Biscuit_Book_WEB.pdf

A book chapter on governance, maturity and MLOps:
https://www.oreilly.com/library/view/what-is-mlops/9781492093626/ch04.html

Guidance on how to develop agile policies:
https://www.policyhub.gov.au/sites/default/files/resources/agile-policy-playbook.pdf

A well maintained knowledge base of ML case studies:
https://atlas.mitre.org

Further enquiries

Professor Debi Ashenden
UNSW IFCYBER Director

ifcyber@unsw.edu.au
ifcyber.unsw.edu.au

UNSW