Conceptualizing the Secure Machine Learning Operations (SecMLOps) Paradigm

Xinrui Zhang* and Jason Jaskolka
Systems and Computer Engineering, Carleton University, Ottawa, Canada
xinrui.zhang@carleton.ca, jason.jaskolka@carleton.ca
*corresponding author

***Abstract*—Due to the proliferation of machine learning in various domains and applications, Machine Learning Operations (MLOps) was created to improve efficiency and adaptability by automating and operationalizing ML products. Because many machine learning application domains demand high levels of assurance, security has become a top priority and necessity to be involved at the beginning of ML system design. To provide theoretical guidance, we first introduce the Secure Machine Learning Operations (SecMLOps) paradigm, which extends MLOps with security considerations. We use the People, Processes, Technology, Governance and Compliance (PPTGC) framework to conceptualize SecMLOps, and to discuss challenges in adopting SecMLOps in practice. Since ML systems are often multi-concerned, analysis on how the adoption of SecMLOps impacts other system qualities, such as fairness, explainability, reliability, safety, and sustainability are provided. This paper aims to provide guidance and a research roadmap for ML researchers and organizational-level practitioners towards secure, reliable, and trustworthy MLOps.**

***Keywords*—machine learning; operations; MLOps; SecMLOps; security;**

I. Introduction

Due to the continuous advancement of machine learning (ML), many critical domains, such as healthcare, finance, energy, and transportation, embrace this emerging technology. Security is a paramount concern to be considered in ML-based systems because security failures may severely impact the system quality and our society in negative ways (e.g., unreliable or unsafe operations leading to human harm and/or property damage). However, managing security in ML systems is not straightforward and there exists intrinsic and extrinsic challenges. Machine learning, particularly deep learning, has been highlighted with the issue of lacking explainability, interpretability, and transparency [1]. This concern is often examined from legal and ethical views, and consequently has raised questions in terms of the fairness and accountability [2] of ML-based systems, especially for those related to equality and diversity such as gender, race, or religion [3]. Although many researchers shed light on eXplainable Artificial Intelligence (XAI), aiming at explaining the decision-making process behind ML models as a way of having more trustworthy ML-based systems, considerable efforts are still required to sufficiently understand how the ML model makes decisions [4]. This lack of transparency in ML models substantially contributes to the effectiveness of stealthy adversarial attacks [5]. A well-known example is that a small perturbation in the input that is imperceivable to human eyes can completely fool the ML model. Moreover, due to the complexity of ML systems, adversaries often have larger attack surfaces to exploit, and systems could face different threats throughout the whole development lifecycle [6], [7], including data collection, ML model training, and deployment. Practical attacks include, but are not limited to, data poisoning attacks [8], adversarial example attacks [9], and model theft [10]. Since the explainability and transparency of ML models is still a challenging problem [11] and sophisticated attacks are emerging [12], the integration of security into the ML development lifecycle is more important than ever, but has not yet received widespread attention.

Building ML-based systems is a complex and iterative process, requiring multidisciplinary efforts from experts with different backgrounds. With the growing prevalence of ML in various applications, the generation of ways to effectively develop, deploy, and manage ML models in production is needed. As a result, Machine Learning Operations (MLOps) has been created as the paradigm to introduce automation and monitoring at all stages of building an ML system, including training, integration, testing, releasing, deployment, and infrastructure management [13]. However, MLOps does not clearly provide explicit consideration for security concerns such as those described above. To address this issue, in this paper, we introduce a novel paradigm called Secure ML Operations (SecMLOps), by extending the MLOps with security considerations in the same spirit as DevOps to DevSecOps or SecDevOps [14]. We conceptualize and elaborate SecMLOps under the framework of People, Processes, Technology, Governance and Compliance (PPTGC), which gives full visibility and control of the paradigm in the aspects of who performs it, how to perform it, what to perform with, and under which constraints. Establishing a paradigm to guide and support the secure development and operations of ML systems provides more than a one-time gain. To the best of our knowledge, this is the first paper to conceptualize the SecMLOps paradigm and discuss its associated challenges.

In this paper, our main contributions include:

- Proposing a conceptualization of SecMLOps under the PPTGC framework;
- Discussing the challenges related to adopting SecMLOps in practice and suggesting solutions to overcome them;
- Analyzing how the adoption of SecMLOps can impact other qualities in ML-based systems including fairness, explainability, reliability, safety, and sustainability.

This research was supported by the Ericsson 5G Fellowship Program under the Ericsson-Carleton Partnership.

2693-9177/22/$31.00 ©2022 IEEE
DOI 10.1109/QRS57517.2022.00023

The rest of the paper is organized as follows. Section II provides the necessary background of MLOps and the PPTGC framework as the foundation for conceptualizing SecMLOps. Section III describes a novel paradigm of SecMLOps under PPTGC framework. Section IV discusses the challenges of adopting SecMLOps in practice and suggests several solutions to overcome these challenges. Section V scrutinizes the impacts on other ML-based system qualities including fairness, explainability, reliability, safety, and sustainability when adopting SecMLOps. Related works are discussed in Section VI. Finally, Section VII concludes and highlights directions for future research.

II. Background

This section gives the basic background to support the establishment of the SecMLOps paradigm. Firstly, a review of MLOps, which serves as the foundation for SecMLOps, is presented. Then, an introduction on the People, Processes, Technology, Governance and Compliance (PPTGC) framework, which describes the primary aspects of the SecMLOps conceptualization, is provided.

A. MLOps

Although Machine Learning Operations (MLOps) has received growing interest from both industry and academia, it is still a vague term [15], [16]. Despite this, the essential core idea behind MLOps is to extend DevOps with ML-specific considerations to ensure continuous delivery of high-performance ML models in production [17], [15], [18].

Existing research has shed light on various specific aspects of MLOps. Testi et al. [16] proposed the taxonomy for clustering research papers on MLOps, namely ML-based software systems, ML use case applications, and ML automation frameworks. They also presented a framework for an ML pipeline which includes ten steps: business problem understanding, data acquisition, ML methodology, ML training & testing, continuous integration, continuous delivery, continuous training, continuous monitoring, explainability, and sustainability.

The MLOps workflow or process, maturity level models, tools, and platforms from both grey and scientific literature are presented in [19], [20], and an additional comparison and selection of tools for meeting specific requirements of MLOps is presented in [21]. The ML workflow seems to reach a general agreement, but there are no universal maturity level models for MLOps. For example, Meenu et al. [20] presented a MLOps maturity model from an academic view using four stages: 1) Automated Data Collection, 2) Automated Model Deployment, 3) Semi-automated Model Monitoring, and 4) Fully-automated Model Monitoring. However, two different industrial maturity level models from Google and Microsoft are demonstrated in [19]. Trends and challenges for MLOps are discussed from several diverse perspectives, such as sustainability [22] and data-quality [18]. Kreuzberger et al. [15] conceptualized MLOps by identifying the technical components, principles, roles, and architectures. This work gave a relatively complete picture of MLOps, specifically how different roles follow the relevant principles, supported by technical components, to implement the MLOps architecture. Since SecMLOps is unavoidably based on MLOps, we have chosen to adopt the work by Kreuzberger et al. [15], which is the most detailed and explicit work we found on MLOps, as the foundation upon which to conceptualize SecMLOps.

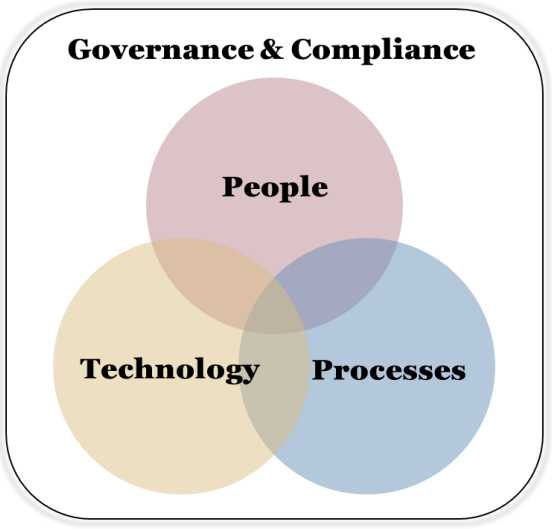

Figure 1. People, Processes, Technology, Governance and Compliance (PPTGC) framework adopted from [23]

B. PPTGC Framework

The People, Processes, Technology, Governance and Compliance (PPTGC) framework, shown in Figure 1, is extended based on the People, Processes, and Technologies (PPT) framework proposed in [23]. The PPT framework is widely used in information technology topics such as product development [24], knowledge management [25] and customer relationship management [26]. Although the Governance and Compliance aspect is often subordinated to processes [23], we agree that it can be a category of its own due to the high importance within the concept of DevOps variants. The *people* aspect refers to the individuals or teams responsible for doing the work, the roles that are involved, and the knowledge that is required to do the work. The *process* aspect refers to ways in which the work is done. The *technology* aspect refers to the tools and platforms to perform specific tasks within the work. Lastly, the *governance and compliance* aspect refers to the standards or guidance that the work should follow and the restrictions and limitations of the work in certain domains. In our context, the PPTGC framework provides the primary aspects to support the conceptualization of the SecMLOps paradigm, and the work means to realize SecMLOps in ML development and operations.

III. Introducing SecMLOps

In this section, we aim to introduce a new paradigm aimed at explicitly considering security concerns throughout ML

operations. The core concept of SecMLOps is proposed and further expanded under each aspect of the PPTGC framework.

A. Our vision of SecMLOps

The core concept of SecMLOps is proposed below.

Concept. *SecMLOps promotes the explicit consideration and integration of security within the whole MLOps life cycle to result in more secure, reliable, and trustworthy ML-based systems.*

This is a multidisciplinary effort composed from the aspects of people, process, technology, and governance and compliance. As mentioned in Section II, our description of SecMLOps in the subsequent sections is largely builds upon the work by Kreuzberger et al. [15].

B. People

It is always people who make systems more secure. According to [15], the roles involved in MLOps include the business stakeholder or product owner, solution architect, data scientist or ML developer, data engineer, software engineer, DevOps engineer, and ML engineer or MLOps engineer. Each role with its purpose and related tasks in MLOps is briefly described in [15]. In the following, we only focus on those roles and responsibilities relevant for SecMLOps:

R1. Business Stakeholder (*similar roles*: Product Owner, Project Manager). The business stakeholder controls the ML systems security by developing high-level security strategies in the business domain. This may require analysis from external security experts in helping the business stakeholders set up security objectives, understand the security requirements [27] in the domains in which ML is applied, especially sensitive domains such as healthcare [28] and cybersecurity [12]. The business stakeholder also supervises the enforcement of those security requirements by making security policies to guide ML development and operations that the rest of roles should follow.

R2. Solution Architect (*similar role*: IT Architect). The solution architect follows the security-by-design paradigm by considering security thoroughly including when designing the ML systems architecture, selecting technologies to be used, and performing the ML systems evaluation. This may require external security experts to identify attack surfaces, perform threat modelling and risk analysis [29], [27], analyze possible attacks, and provide general solutions to mitigate them.

R3. Data Engineer (*similar role*: DataOps Engineer). Data engineers specialized in security or the internal security experts have the knowledge of possible attacks that could happen during data ingestion and feature engineering. Based on the design solution by the solution architect (i.e., role R2), they discover and implement the most suitable defences and mitigations [30] to realize secure data ingestion and management to result in desired datasets.

R4. Data Scientist (*similar roles*: ML Specialist, ML Developer). Data scientists specialized in security or the internal security experts have the knowledge of adversarial ML and various training techniques [31], [32]. They design or optimize the ML algorithm to meet the security requirements demanded by the solution architect (i.e., role R2), such as robustness and privacy-preservation.

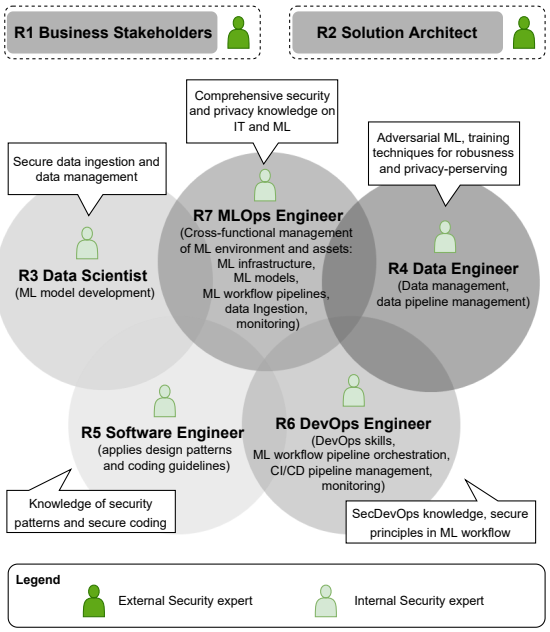

Figure 2. Summary of the roles in SecMLOps and their necessary knowledge, adapted from [15]

R5. Software Engineer. Software engineers specialized in security or the internal security experts are aware of, and follow, secure coding practices and guidelines [33] to lay a foundation for secure and well-engineered ML-based systems. They may need to refactor the code into a more sustainable manner for the purpose of security management.

R6. DevOps Engineer. DevOps engineers specialized in security or the internal security experts have knowledge of SecDevOps [34]. They ensure the safety and security of components to achieve the integration of ML development and operations, such as continuous integration and continuous delivery (CI/CD) automation, ML workflow orchestration, and monitoring.

R7. MLOps Engineer (*similar role*: ML Engineer). MLOps engineers specialized in security or the internal security experts have the comprehensive knowledge of security in ML and IT. They communicate and manage the cooperation between data scientists, data engineers, software engineers, and DevOps engineers (i.e., roles R3, R4, R5 and R6) to maintain the normal operation among interdisciplinary teams, in support of better reaching the goal of secure and robust ML-based systems. They are also responsible for handling incidents whenever human intervention is needed.

Figure 2 summarizes these roles and responsibilities and the necessary knowledge required for each role. Specifically, there are two primary situations. The first is when external security experts needed to assist the roles (e.g., R1 and R2) and the second is when roles are specialized in security with specific technical background (e.g., R3, R4, R5, R6, and R7).

Depending on the organization size, budgets and application domain, other mixed situations can happen. This role description acts as general guidance in achieving SecMLOps and the organization should tailor the role arrangement based on their specific needs, e.g., internal security experts may not be required for every role.

C. Processes

Security engineering activities are encouraged to be performed during the ML lifecycle to achieve secure and trustworthy ML systems. We propose a security process framework for SecMLOps in Figure 3. The proposed framework is built upon existing automated ML pipelines introduced in [15] and [13]. Readers can refer to [15] for detailed description of the ML end-to-end workflow. Here we only focus on the security aspects throughout the process, and they are presented in three stages: *Design*, *Experiment*, *Production*. We also describe security aspects that are applicable to the *Overall* process.

1) Design: In the MLOps project initiation, when performing the business problem analysis to obtain and define the business requirements, the overall security objectives (e.g., confidentiality, integrity, availability, privacy, etc.) and security requirements of the ML system assets should also be defined in terms of the current ML development and long-term operation and maintenance. Since different levels of confidentiality and privacy related to the data and the purpose of the application would result in different levels of required protection, the security requirements for the ML-related data and model should be clearly stated.

During the requirements analysis, potential threats and countermeasures in the entire ML workflow in a particular business domain should also be identified and documented via building ML security models such as ML attack surfaces [35], threat models [7], [35], [36], and attack/adversarial models [7], [35], [37]. This comprehensive understanding plays an important role in achieving security-by-design at the architecture design stage when it is less costly to make changes.

2) Experimentation: Based on the constructed understanding from the previous security activities during the *Design* stage, organization-level security policies should be established, obeyed, and constantly updated during the MLOps lifecycle. The security policies should define the system assets that need protection and how to protect those assets from potential threats, covering ML-related areas such as data security and model security, as well as IT-related areas such as physical security, personnel security, administrative security, and network security [38].

Before exporting the model, security evaluation [6], [39] on the ML model (e.g., using cleverhans [40], secml [41], Adversarial Robustness Toolbox (ART) [42]) should be performed to assess robustness of the ML system in the simulated adversarial setting. For example, a list of attacks that are identified as harmful with high likelihood to the ML systems can be executed to test how systems react and check if the log information can be used to trace back to the attack.

3) Production: In the production stage, the monitoring component is vital to maintain the normal operation of ML systems. Except for preserving the reliability of the ML systems by maintaining the predictive quality especially when handling uncertainty, the monitoring component should function as an anomaly detection mechanism to identify abnormal situations. This is essential, especially for safety or security-critical systems and applications, to ensure fail-safes in the ML production. The monitoring component should also implement incident responses guided by the security policies. When it senses the systems might be under attack, it can react as designed to minimize attack impacts, for example containerizing the uncompromised ML systems to avoid further loss [43]. After the attack, human intervention by the MLOps engineers (see Section III-B) should be involved to ascertain the root cause of failure [43], and update the defences and policies to cover more attacks.

4) Overall: During the whole process, it is important to keep documenting useful knowledge in a shareable way. The MITRE ATT&CK framework [44] is a well-known searchable knowledge base of adversary tactics and techniques for traditional software security. Similar curated repositories of attacks against ML-based systems during lifecycle should also be created [43], along with the defences and mitigation for the incident respond. A great example of this is MITRE ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems) [45].

Some other aspects that could be documented include but not limit to the approach to establish security policies and their effectiveness; rationales of chosen and proven techniques, tools, and platforms involved in the ML workflow; sound metrics to evaluate the processes, and the level of capability maturity model (CMM) of SecMLOps. The eventual goal is to encode the accumulated knowledge and experience into a catalogue of reusable solutions, such as patterns and security patterns [46] so that people can continuously build on and benefit from them.

D. Technology

Kreuzberger et al. [15] listed the technical components that enable MLOps, including Google's Vertex AI [47], Microsoft's Azure Machine Learning [48] and Amazon SageMaker [49], just to name a few. For SecMLOps, we extend those components with possible security controls. When building the CI/CD component, systematic testing such as unit testing and integration testing on both CI and CD components are necessary to guarantee rapid and reliable operations. For example, testing for CI can include ensuring model training convergence and suitable integration between pipeline components. Testing for CD can include the load testing of the service to capture metrics such as queries per seconds (QPS) and model latency [13]. Meanwhile, teams should use automation to replace manual and/or error-prone tasks in the testing process as much as possible.

The source code repository, feature store system, model registry, and ML metadata stores are essential storage systems

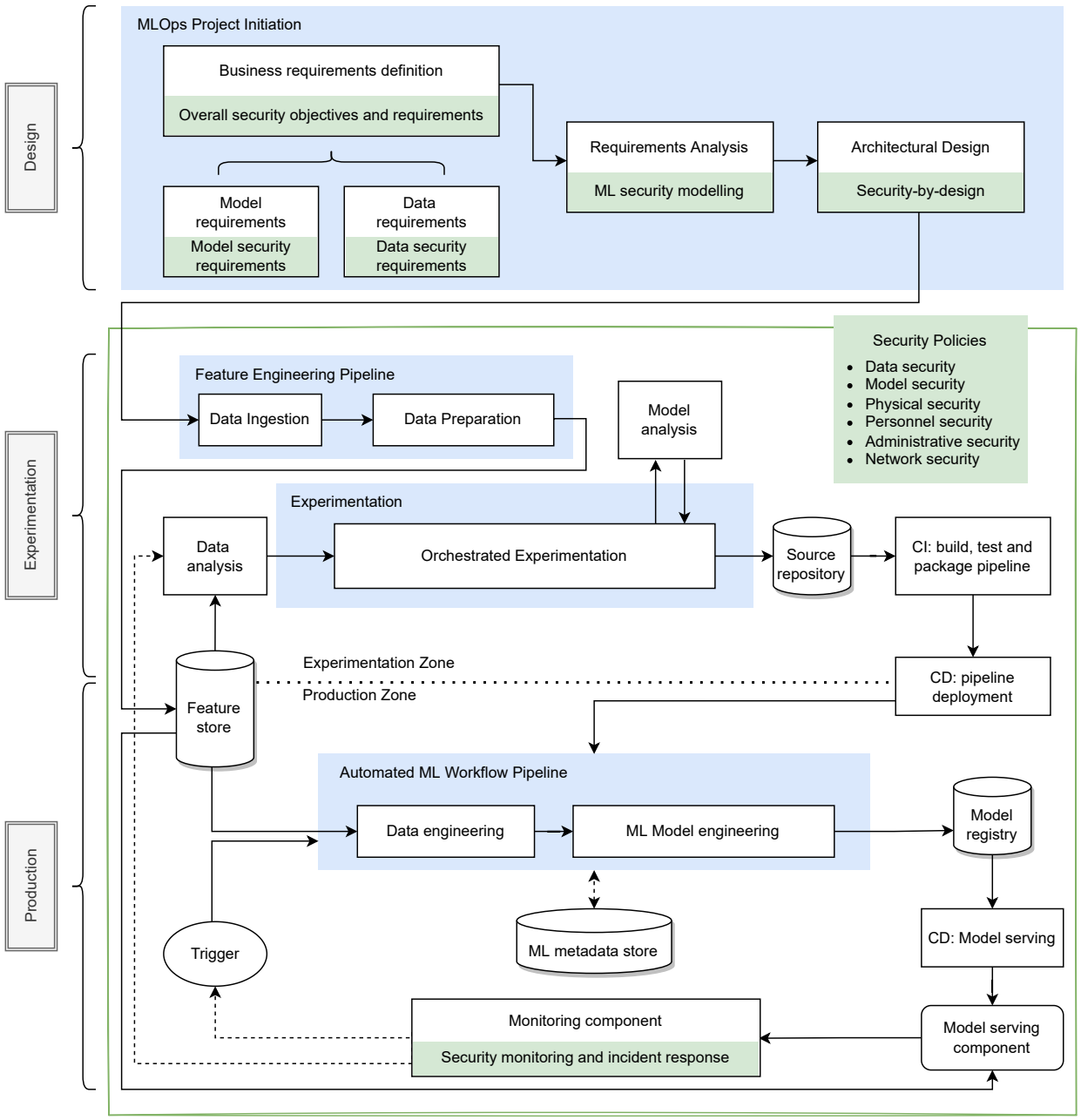

Figure 3. A security process framework for SecMLOps. Blue blocks show a simplified version of the MLOps workflow from [15] with identified security considerations shown in green.

required in MLOps. Therefore, basic security controls guided by security policies such as access control, encryption, and logging functions (i.e., identity information, time, action, etc.) should be equipped to protect from tampering, information disclosure, and repudiation threats. To ensure the security of the model training infrastructure and model serving component, the hardware and digital supply chain should be verified.

There are various tools and platforms for individual technical components in MLOps [15] and some companies also provide ML lifecycle management platforms to realize MLOps. We provide a summary of these tools and platforms in Table I, along with their main security features and compliance if applicable. Take supervisely [52] as an example, it is a commercial data prepossessing platform that embraces role-based access control (RBAC), encryption of data at rest and in transit, single sign-on authentication (SSO), virus scanning, and network isolation as security features to protect from adversaries. This platform is also compliant with GDPR and HIPAA standards, which follows security practices and measures that ensures data confidentiality and privacy.

Different business domains require different levels of security management. Thus, to converge towards SecMLOps,

TABLE I
A SUMMARY OF SECURITY CONTROLS IN TOOLS AND PLATFORMS FOR MLOPS

Use	Name	Status	Main Security Features	Compliance	Obtainable Security Documentation?
Data Preprocessing	Labelbox [50]	Commercial	Encryption at rest, SSO	CCPA, GDPR, SOCII and HIPAA	Yes
	iMerit [51]	Commercial	N/A	ISO 27001, ISO 9001:2015, GDPR, SOCII and HIPAA	No
	supervisely [52]	Commercial	RBAC, encryption at rest and in transit, SSO, virus scan, network isolation	GDPR and HIPAA	Yes
	Segments.ai [53]	Commercial	N/A	ISO 27001 and GDPR	No
Data Versioning	LakeFs [54]	Open-Source	N/A	N/A	No
	Delta Lake [55]	Open-Source	Access control, encryption at rest and in transit	CCPA and GDPR	Yes
Feature Engineering	Tecton [56]	Commercial	SSO, access controls, IAM permissions AWS-standard encryption	SOCII and GDPR	Yes
	Feast [57]	Open-Source	N/A	N/A	No
	HopsWork [58]	Open-Source	N/A	N/A	No
	Rasgo [59]	Commercial	N/A	N/A	No
	dotData [60]	Commercial	N/A	N/A	No
End-to-end MLOps platform	MLflow [61]	Open-Source	N/A	N/A	No
	managed MLflow [62]	Commercial	RBAC	N/A	Yes
	Iguazio [63]	Commercial	RBAC, encryption at rest and in transit, penetration test, vulnerability scan	N/A	Yes
	Weights and Biases [64]	Commercial	RBAC, encryption at rest and in transit, pentest, comprehensive security policies	SOCII	Yes
	Polyaxon [65]	Open-Source	RBAC	N/A	OK
	Vertex AI [47]	Commercial	SSO, encryption at rest and in transit, authentication	N/A	OK
	Amazon SageMaker [49]	Commercial	RBAC	N/A	OK
	Azure Machine Learning [48]	Commercial	RBAC, encryption at rest and in transit, vulnerability scan, network isolation	N/A	Yes
	Snorkel [66]	Open-Source	RBAC, encryption at rest and in transit, penetration test, vulnerability scan	N/A	Yes

the availability and configurability of security features and controls should be considered as a critical factor when selecting an MLOps tool or platform, especially for safety or security-critical systems. For example, if the business domain is data-sensitive (e.g., healthcare or finance), Google's Vertex AI [47] and Microsoft's Azure Machine Learning [48] are both furnished with role-based access control, vulnerability scanning, and encryption mechanisms to protect data in transit and at rest. However, in other domains, we insist that basic security controls such as authentication and access control should not be neglected, no matter whether choosing tools or platforms available on the market or developing self-use tools.

E. Governance and Compliance

Due to the prevalence of ML in many critical domains and applications, security has become a top priority and many countries have taken incremental steps to legislate security when developing ML systems [67]. For example, in the United States, the SPY Car Act was introduced to enhance controls on cybersecurity to all vehicles including automated driving systems (ADSs). According to the law, it requires isolating critical software systems from the rest of a vehicle's internal network and evaluating all vehicles using best practices. It also asks for specifications to ensure the security of collected information in ADSs when the data is on the vehicle, in transit or in any off-board storage [67]. Therefore, securely developing and operating ADSs is obligated by law. Meanwhile, the SecMLOps paradigm should be implemented under the guidance of related laws and regulations to show the system's compliance to specific security requirements.

An organization should follow related standards and guidelines, as well as establish security policies that guide people to protect assets in ML systems and workflows using suitable methods and tools. ISO/IEC JTC 1/SC 42 [68] provides standards that cover a large variety of topics under artificial intelligence, including big data reference architecture and data quality for analytics and machine learning. ETSI also created several standards (e.g., ETSI GR SAR series [69]) to preserve and improve the security of AI/ML technologies. Their standards on hardware for AI/ML, data supply chain and mitigation against threats for ML-based systems are qualified sources as technical guidance. For ML systems which are

launched in specific nations or fields, there are data regulations to comply with such as GDPR standards in Europe and HIPAA standards for medical data. Globally, Europe made the first step towards AI regulation by submitting the EU AI Act, and Canada's Directive on Automated Decision-making (the Canada ADM Directive) was created followed by the Canada AI White Paper [70].

Beyond the commonly used performance metrics, metrics to evaluate how well SecMLOps is implemented should also be built. Metrics are important as quantifiable measures to assess the status of a system and can be used to support compliance efforts. Specifically, they can provide evidence that ML development and operations are maintaining compliance with standards and regulations.

Such metrics should cover the aspect of people; more precisely, the security knowledge of different roles and security culture of the organization. For example, the following metrics can be used:

- the percentage of people who completed security awareness training;
- the percentage of people who attended activities related to promoting security;
- current level of maturity in the Capability Maturity Models (CMM) [71].

The trend of these statistical metrics (increasing or decreasing) can reflect whether a people-centric security culture is forming. The maturity level of CMM can reflect the formality and optimization of the existing process in promoting security and evaluate the capability of these processes.

With respect to process and technology, the following metrics could be used:

- the number of times security policies are violated;
- the number of times that significant performance drop is due to adversarial attacks and (the number of times that applying tools with security controls could have avoided it).
- the time needed to find: (1) the root cause, (2) enforce incident response, and (3) remediate identified issues.

Metrics for ML systems monitoring could include serving latency, throughput, disk utilization, system's up-time and GPU/CPU usage and number of APT calls [21]. Lastly, metrics on the effectiveness of the security policy should also be established. Possible parameters include the coverage and suitability of security policies. In all cases, it is imperative that the selected metrics be sound, meaning that they can be consistently measured in a reproducible, objective, and unbiased fashion while providing contextually relevant, actionable information for decision makers [72].

IV. Challenges in Adopting SecMLOps in Practice

Now that we have conceptualized the SecMLOps paradigm, we must carefully consider the practical challenges in adopting the paradigm in practice. The challenges for adopting MLOps still apply here as SecMLOps is built upon the foundation of MLOps. Open challenges include, but are not limited to, a lack of highly skilled experts for roles involved in MLOps, inefficient communication, managing voluminous and varying data, scalability of the infrastructure [15], model retraining, and monitoring [19]. In what follows, we discuss the challenges in adopting SecMLOps by considering each aspect of the PPTGC framework and provide our suggestions for overcoming them.

A. People

Overall, people underestimate and even downplay the role of security in MLOps. Security activities are usually considered to consume too much time and money, which often conflicts with practical business pressures that are typically centred on generating revenues related to the developed products and solutions. This issue is exacerbated by the fact that a system cannot be guaranteed to be absolutely secure even with security controls applied. Furthermore, there is a perception that the chances that a specific system will be the target of an attack are often underestimated [73]. These issues often lead to complacency and a lack of attention and care towards securing the system. Therefore, it is generally difficult to form a security culture at an organizational level. Another unfortunate fact is that there is a constant shortage on security experts and personnel, and security experts with an ML background are even more rare. As a result, it is practically challenging to fulfill all the roles described in Section III-B with internal or external security experts.

Suggestions to Overcome the Challenges. Since many ML-related attacks are evolving to be more sophisticated, there is a need for an organizational commitment to continuous learning, awareness, and mastery at the most advanced level to understand those attacks and to make and use effective defenses. It is recommended to offer related security training programs for different roles. Activities in different forms are also welcomed to encourage learning, such as security-related topic days (e.g., security knowledge competitions), round-table discussions, seminars, and poster fairs. Assembling AI red teams to take an adversarial approach to expose vulnerabilities with the goal of making the system stronger could also be promoted. Some leading companies, such as Facebook, have already done this [74]. However, the above activities come at a cost and the organization should tailor plans based on their size, budget, and need. Ultimately, the goal is to form a culture of people-centric security and continuous learning and awareness in the organization.

B. Technology

There are a decent number of tools and platforms available (see Table I) for individual technical components involved in MLOps such as data prepossessing, and the end-to-end MLOps lifecycle. Readers can refer to [19] and [21] for detailed lists of supportive tools.

We conducted a small empirical study by randomly selecting 20 tools or platforms and scrutinized how obtainable documentation related to their security features are on their website. The results of this exercise are shown in the last column of Table I. By obtainable, we mean whether it is

straightforward to find information about security controls (and standards compliance) in the tools or platforms webpage (e.g., in the main page or in the documentation). The result shows that 60% (12 out of 20) of them do not provide a clear introduction on security features that can be easily navigated. Among them, 75% (9 out of 12) of tools do not mention security at all, and the remaining ones (3 out of 12) require some non-trivial amount of effort to find documentation related their security controls. As a result, there is a lack of MLOps tools and platforms with well-explained security features. For those platforms equipped well-defined enterprise-level security, the associated cost of adopting such tools or platforms might be unaffordable for some small or medium-sized organizations with limited budgets.

Suggestions to Overcome the Challenges. We suggest platform builders to put forth greater effort in explicitly documenting and presenting security controls as one of the key features in the platform introduction that can be easily found and navigated by users and developers. For those tools that do not have security controls, we encourage contributions to fill the gaps, especially for those open-sourced projects.

C. Governance and Compliance

It generally requires a considerable amount of effort to be compliant with regulations and standards. Take the Canada ADM Directive as an example; it requires to conduct risk assessment during the system development lifecycle and establish appropriate safeguards as proof of security in quality assurance. However, it does not show how to do it and how to demonstrate the effectiveness of doing it a certain way. Similar challenges also apply when it comes to following related standards. Additionally, choosing suitable standards for organizations to follow is also demanding since many of them cover broad topics and are still under development [68]. There are many automated tools for checking GDPR compliance to assist in privacy and data protection [75], but only few tools are specifically designed for ML systems compliance (e.g., compliance.ai [76]).

Suggestions to Overcome the Challenges. We advocate that more work should be put in explaining the AI regulation and standards in terms of how to prove ML systems' compliance. For example, some real-world industrial use cases can be demonstrated to show the whole process including what to do and how to do it. Note that there may be differences in terms of what needs to be done and how it needs to be done system-to-system and the organization should adjust case-by-case. In any case, there is a need on more detailed references (the process depicted in Figure 3 may serve this propose) to guide the compliance process. Further, automated tools for ML systems compliance are encouraged to create and update, to improve the efficiency of compliance checking.

D. Process

The challenges in the process aspect consist of challenges from all other aspect because it requires people, technology, and governance and compliance to be able to finish the process. To be specific, security experts, usable and effective technologies, and the ability to understand and show compliance with applicable regulations and standards are necessary to adopt and implement the SecMLOps paradigm in practice. Besides, ML systems development and operations (as shown in Figure 3) is iterative and experimental. It requires scalability, sufficient compute power, as well as versioning functions that translate into complex infrastructure, which is still challenging to manage and operate. Security as a non-functional requirement is usually sacrificed to resolve the conflicts between performance, time and expenses, with a relatively high probability of leading to problematic times during the production.

Suggestions to Overcome the Challenges. It is encouraged to make efforts in solving all other aspects, and it can largely alleviate the burden in completing the process in an effective and manageable way. Having a detailed process with security consideration and sticking to it is what SecMLOps requires, and it is necessary for the organization to balance the trade-offs in their specific context. Trade-off analysis methods [77] are helpful for designers to understand the advantages and disadvantages of requirements and/or design choices for a system, therefore making reasonable balance between security and other competing goals such as performance and usability.

V. Impact on Other Qualities

The eventual goal of MLOps is to commercialize ML-based systems, where security is a dominating quality to gain customer trust. However, we must not lose sight of the fact that there exist other qualities that need to be considered in the development and operations of ML-based systems. Applying security solutions has been shown to have positive effect on other qualities, for example in blockchains [78] and Internet of Things (IoT) [79]. In this section, we discuss the relationship between security and other important qualities, including fairness, explainability, reliability, safety, and sustainability. We demonstrate how the adoption of SecMLOps impacts other system qualities and vice versa.

A. Security and Fairness

Fairness in ML systems is getting more attention, as increasing number of those systems are used in society-impacting environments and have a direct influence in our lives. Different sources of unfairness in ML have been identified as biases in data, algorithms, and user interaction, and discrimination [80]. There are numerous real-world examples that shows the existence of bias in ML-based systems, such as in AI chatbots, employment matching, and flight routing [81]. The methods for fair ML are categorized as pre-processing, in-processing, and post-processing [80], [82]. The first two categories have the same spirit as mitigating certain adversarial attacks, for example data poisoning, which is to sanitize data and modify learning algorithms, respectively. Most works focus on reducing discrimination while maintaining a decent accuracy [83], [84], but the effect of applying those techniques on the security of the systems is rarely discussed. However, strengthening

the security of ML-based systems seems to have a positive impact on fairness, and adversarial training [85] has been proven to be effective in mitigating unwanted biases in various gradient-based learning models, including both regression and classification tasks.

B. Security and Explainability

Explainability is acknowledged as an essential feature for ML to be practically deployed in any critical system since it determines how much trust and confidence domain experts and users can put into the system. Considerable efforts have been made in the eXplainable AI (XAI) field [11], [86]. XAI [87] proposes creating a suite of ML techniques that result in "more explainable models, and allow humans to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners." It is obvious that XAI has a beneficial impact on ML security. Examples show that it is successfully adopted as predictive maintenance to predict machine errors or tool failures in manufacturing [88], and as fault-detection in real chiller systems [89]. Because the reasons behind how ML model acts are able to be explained by XAI, it also benefits the governance and compliance aspect in SecMLOps. Following the SecMLOps could help release the burden of root cause analysis when unintended explainability issues arise in ML-based systems. Despite that, XAI unavoidably poses a threat to model's confidentiality and causes intellectual property exposure, since model stealing exists even on black-box settings [90]. Works such as [91] have been devoted to solving this problem.

C. Security and Reliability

Reliability in ML systems is defined as not only achieving strong predictive performance, but also consistently performing well over many decision-making tasks involving uncertainty, robust generalization, and adaptation [92]. Making ML-based systems secure is the foundation to make them more reliable. Moreover, security monitoring can be taken advantage of to report uncertainty in the predictions and allow systems to fail gracefully in abnormal situations to improve reliability. Developing ML-based systems with security, safety and reliability can eventually contribute the trustworthiness of ML systems [93], and enable deeper integration of ML in critical application domains such as healthcare, government, and justice.

D. Security and Safety

ML security and ML safety cannot isolate one from another. Hendrycks et al. [94] categorized ML safety problems as robustness, monitoring, alignment, and systemic safety. Similarly, Mohseni et al. [95] grouped safety-related ML research into three strategies: (1) Inherently Safe Models, (2) Enhancing Model Performance and Robustness, and (3) Run-time Error Detection techniques. Their essential ideas largely overlap and align with the goal of SecMLOps, including considering safety early in the design phase, building ML-based systems that can endure unusual, extreme, and adversarial events, and monitoring ML systems to identify abnormal situations and enhance reliability.

E. Security and Sustainability

A famous study by Strubell et al. [96] shows that training a single, deep learning, natural language processing model can cause carbon dioxide emissions as much as in the lifetime of five cars. As reducing carbon emission becomes a global commitment, sustainable AI or sustainability of AI is proposed to "develop AI/ML that is compatible with sustaining environmental resources for current and future generations; economic models for societies; and societal values that are fundamental to a given society" [97]. ML systems designed under SecMLOps would require some extra efforts during the design and experiment stages, but we conjecture that it is beneficial to the ecological and economic cost in the long term, since less re-engineering due to security issues is needed and security mechanisms can be adopted and reused in similar cases.

VI. Related Work

Security considerations have been included in DevOps (the most popular paradigm) as SecDevOps or DevSecOps. SecDevOps and DevSecOps are two terms mostly used interchangeably in academia, although researchers tend to prefer DevSecOps as the primarily used term. The fundamental idea for DevSecOps is to integrate security practices or activities in the DevOps processes through increased collaboration and communication between the development and operations teams with the security team [34], [98]. Adoption challenges are discussed by [99], [98], [100] in terms of tools, practices, infrastructure and people or culture.

In general, the discussion related to securing the ML workflow is far behind the discussions related to utilizing ML as a cybersecurity tool [101], [102], [103]. Among the works of ML security, most are seeking to secure ML systems in a specific application or improve the security or robustness of the ML algorithm and structure by identifying threats and proposing defenses [6], [7], [35], rather than securing general development lifecycle of ML systems. Qayyum et al. [28] formulated the ML pipeline for healthcare applications and described vulnerabilities at each stage that raises security and robustness challenges. Another paper [104] explore various security and privacy threats and consequences of threats to healthcare systems, as well as existing security measures and their limitations. Similar work was also done for autonomous vehicle systems [105]. Papernot et al. [106] provided a generic model of machine learning applications and focused on articulating a threat model for ML, and categorize attacks and defences within an adversarial framework. Alternatively, our work proposes SecMLOps and fills the gaps of the theoretical guidance for considering and integrating security into MLOps from the aspect of people, processes, technology, governance and compliance.

VII. Concluding Remarks

In this paper, we provided a conceptualization of the SecMLOps paradigm, which extends MLOps with explicit security considerations throughout the ML workflow and aims towards secure, reliable, and trustworthy MLOps. Our vision of SecMLOps is defined from the aspects of people, processes, technology, and compliance and governance to give a complete picture of the SecMLOps concept. Challenges in building SecMLOps solutions with possible suggestions to overcome them are discussed under the same aspects. As other qualities are often required in ML systems, we examined how SecMLOps impacts other qualities including fairness, explainability, reliability, safety, and sustainability.

We believe that SecMLOps is a future direction for MLOps due to the increasing number of ML-based systems deployed in the critical application domains requiring high-level of security assurance. We acknowledge that further research and development are required to fully realize our vision for SecMLOps, and we believe that the solutions proposed in this work provide a research roadmap towards achieving this goal.

References

[1] N. Diakopoulos, "Accountability in algorithmic decision making," *Communications of the ACM*, vol. 59, no. 2, pp. 56–62, 2016.

[2] A. Preece, "Asking 'why' in AI: Explainability of intelligent systems–perspectives and challenges," *Intelligent Systems in Accounting, Finance and Management*, vol. 25, no. 2, pp. 63–72, 2018.

[3] S. Leavy, "Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning," in *Proceedings of the 1st International Workshop on Gender Equality in Software Engineering*, pp. 14–16, 2018.

[4] A. Adadi and M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[5] Y. Zhou and M. Kantarcioglu, "On transparency of machine learning models: A position paper," in *AI for Social Good Workshop*, 2020.

[6] M. Xue, C. Yuan, H. Wu, Y. Zhang, and W. Liu, "Machine Learning Security: Threats, Countermeasures, and Evaluations," *IEEE Access*, vol. 8, pp. 74720–74742, 2020.

[7] Q. Liu, P. Li, W. Zhao, W. Cai, S. Yu, and V. C. M. Leung, "A Survey on Security Threats and Defensive Techniques of Machine Learning: A Data Driven View," *IEEE Access*, vol. 6, pp. 12103–12117, 2018.

[8] M. Jagielski, A. Oprea, B. Biggio, C. Liu, C. Nita-Rotaru, and B. Li, "Manipulating machine learning: Poisoning attacks and countermeasures for regression learning," in *2018 IEEE Symposium on Security and Privacy (SP)*, pp. 19–35, IEEE, 2018.

[9] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami, "Practical black-box attacks against machine learning," in *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, pp. 506–519, 2017.

[10] F. Tramèr, F. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart, "Stealing machine learning models via prediction APIs," in *25th USENIX Security Symposium*, pp. 601–618, 2016.

[11] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, 2020.

[12] H. Chen and M. A. Babar, "Security for machine learning-based software systems: a survey of threats, practices and challenges," *arXiv preprint arXiv:2201.04736*, 2022.

[13] Google Cloud, "MLOps: Continuous delivery and automation pipelines in machine learning." https://cloud.google.com/architecture/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning, 2020. Accessed: 2022-05-15.

[14] H. Myrbakken and R. Colomo-Palacios, "Devsecops: a multivocal literature review," in *International Conference on Software Process Improvement and Capability Determination*, pp. 17–29, Springer, 2017.

[15] D. Kreuzberger, N. Kühl, and S. Hirschl, "Machine learning operations (MLOps): Overview, definition, and architecture," *arXiv preprint arXiv:2205.02302*, 2022.

[16] M. Testi, M. Ballabio, E. Frontoni, G. Iannello, S. Moccia, P. Soda, and G. Vessio, "MLOps: A taxonomy and a methodology," *IEEE Access*, vol. 10, pp. 63606–63618, 2022.

[17] S. Alla and S. K. Adari, *Beginning MLOps with MLFlow: Deploy Models in AWS SageMaker, Google Cloud, and Microsoft Azure*. Berkeley, CA: Apress, 2021.

[18] C. Renggli, L. Rimanic, N. M. Gurel, B. Karlas, W. Wu, and C. Zhang, "A data quality-driven view of MLOps," *IEEE Data Engineering Bulletin*, vol. 44, pp. 11–23, March 2021.

[19] G. Symeonidis, E. Nerantzis, A. Kazakis, and G. A. Papakostas, "MLOps - Definitions, Tools and Challenges," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, (Las Vegas, NV, USA), pp. 0453–0460, IEEE, Jan. 2022.

[20] M. M. John, H. H. Olsson, and J. Bosch, "Towards MLOps: A Framework and Maturity Model," in *2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, (Palermo, Italy), pp. 1–8, IEEE, Sept. 2021.

[21] P. Ruf, M. Madan, C. Reich, and D. Ould-Abdeslam, "Demystifying MLOps and Presenting a Recipe for the Selection of Open-Source Tools," *Applied Sciences*, vol. 11, p. 8861, Sept. 2021.

[22] D. A. Tamburri, "Sustainable MLOps: Trends and Challenges," in *2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, (Timisoara, Romania), pp. 17–23, IEEE, Sept. 2020.

[23] M. Vielberth, F. Böhm, I. Fichtinger, and G. Pernul, "Security operations center: A systematic study and open challenges," *IEEE Access*, vol. 8, pp. 227756–227779, 2020.

[24] J. M. Morgan and J. K. Liker, *The Toyota product development system: integrating people, process, and technology*. Productivity Press, 2020.

[25] G. D. Bhatt, "Knowledge management in organizations: examining the interaction between technologies, techniques, and people," *Journal of knowledge management*, 2001.

[26] I. J. Chen and K. Popovich, "Understanding customer relationship management (CRM): People, process and technology," *Business Process Management Journal*, 2003.

[27] C. Wilhjelm and A. A. Younis, "A threat analysis methodology for security requirements elicitation in machine learning based systems," in *2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, pp. 426–433, 2020.

[28] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, "Secure and robust machine learning for healthcare: A survey," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 156–180, 2020.

[29] G. McGraw, H. Figueroa, V. Shepardson, and R. Bonett, "An architectural risk analysis of machine learning systems: Toward more secure machine learning," *Berryville Institute of Machine Learning, Clarke County, VA. Accessed on: Mar*, vol. 23, 2020.

[30] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical secure aggregation for privacy-preserving machine learning," in *proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1175–1191, 2017.

[31] T. Bai, J. Luo, J. Zhao, B. Wen, and Q. Wang, "Recent advances in adversarial training for adversarial robustness," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21* (Z.-H. Zhou, ed.), pp. 4312–4321, International Joint Conferences on Artificial Intelligence Organization, 8 2021. Survey Track.

[32] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *The journal of machine learning research*, vol. 17, no. 1, pp. 2096–2030, 2016.

[33] M. Graff and K. R. Van Wyk, *Secure coding: principles and practices*. " O'Reilly Media, Inc.", 2003.

[34] V. Mohan and L. B. Othmane, "SecDevOps: Is It a Marketing Buzzword? - Mapping Research on Security in DevOps," in *2016 11th International Conference on Availability, Reliability and Security (ARES)*, (Salzburg, Austria), pp. 542–547, IEEE, Aug. 2016.

[35] X. Wang, J. Li, X. Kuang, Y.-a. Tan, and J. Li, "The security of machine

learning in an adversarial setting: A survey," *Journal of Parallel and Distributed Computing*, vol. 130, pp. 12–23, Aug. 2019.
[36] A. Marshall, J. Parikh, E. Kiciman, and R. S. S. Kumar, "Threat Modeling AI/ML Systems and Dependencies." https://docs.microsoft.com/en-us/security/engineering/threat-modeling-aiml, 2022. Accessed: 2022-05-27.
[37] K. Sadeghi, A. Banerjee, and S. K. Gupta, "A system-driven taxonomy of attacks and defenses in adversarial machine learning," *IEEE transactions on emerging topics in computational intelligence*, vol. 4, no. 4, pp. 450–467, 2020.
[38] IBM, "Security policy and objectives." https://www.ibm.com/docs/en/i/7.1?topic=security-policy-objectives, 2021. Accessed: 2022-05-24.
[39] B. Biggio and F. Roli, "Wild patterns: Ten years after the rise of adversarial machine learning," *Pattern Recognition*, vol. 84, pp. 317–331, 2018.
[40] N. Papernot, F. Faghri, N. Carlini, I. Goodfellow, R. Feinman, A. Kurakin, C. Xie, Y. Sharma, T. Brown, A. Roy, A. Matyasko, V. Behzadan, K. Hambardzumyan, Z. Zhang, Y.-L. Juang, Z. Li, R. Sheatsley, A. Garg, J. Uesato, W. Gierke, Y. Dong, D. Berthelot, P. Hendricks, J. Rauber, and R. Long, "Technical report on the CleverHans v2.1.0 adversarial examples library," *arXiv preprint arXiv:1610.00768*, 2018.
[41] M. Melis, A. Demontis, M. Pintor, A. Sotgiu, and B. Biggio, "secml: A python library for secure and explainable machine learning," *arXiv preprint arXiv:1912.10013*, 2019.
[42] M.-I. Nicolae, M. Sinn, M. N. Tran, B. Buesser, A. Rawat, M. Wistuba, V. Zantedeschi, N. Baracaldo, B. Chen, H. Ludwig, I. Molloy, and B. Edwards, "Adversarial robustness toolbox v1.2.0," *CoRR*, vol. 1807.01069, 2018.
[43] R. S. S. Kumar, M. Nyström, J. Lambert, A. Marshall, M. Goertzel, A. Comissoneru, M. Swann, and S. Xia, "Adversarial machine learning-industry perspectives," in *2020 IEEE Security and Privacy Workshops (SPW)*, pp. 69–75, IEEE, 2020.
[44] MITRE, "MITRE ATT&CK." https://attack.mitre.org/, n.d. Accessed: 2022-08-19.
[45] MITRE, "Adversarial Threat Landscape for Artificial-Intelligence Systems (ATLAS)." https://atlas.mitre.org/, n.d. Accessed: 2022-08-19.
[46] X. Zhang and J. Jaskolka, "Security patterns for machine learning: The data-oriented stages," in *The 27th European Conference on Pattern Languages of Programs*, EuroPLoP 2022, (Kloster Irsee, Germany), p. 18, 2022.
[47] Google Cloud Guides, "Monitor and secure." https://cloud.google.com/vertex-ai/docs/general/monitoring-security, n.d. Accessed: 2022-05-16.
[48] M. Baldwin, "Azure security baseline for Azure Machine Learning." https://docs.microsoft.com/en-us/security/benchmark/azure/baselines/machine-learning-security-baseline?context=/azure/machine-learning/context/ml-context, 2022. Accessed: 2022-05-16.
[49] Amazon Developer Guide, "Security in Amazon SageMaker." https://docs.aws.amazon.com/sagemaker/latest/dg/security.html, n.d. Accessed: 2022-05-16.
[50] Labelbox. https://labelbox.com/. Accessed: 2022-08-23.
[51] imerit. https://imerit.net/. Accessed: 2022-08-23.
[52] supervisely. https://supervise.ly/. Accessed: 2022-08-23.
[53] Segments.ai. https://segments.ai/. Accessed: 2022-08-23.
[54] LakeFs. https://lakefs.io/. Accessed: 2022-08-23.
[55] Delta Lake. https://delta.io/. Accessed: 2022-08-23.
[56] Tecton. https://www.tecton.ai/. Accessed: 2022-08-23.
[57] Feast. https://feast.dev/. Accessed: 2022-08-23.
[58] HopsWork. https://www.hopsworks.ai/. Accessed: 2022-08-23.
[59] Rasgo. https://www.rasgoml.com/. Accessed: 2022-08-23.
[60] dotData. https://dotdata.com/. Accessed: 2022-08-23.
[61] mlflow. https://mlflow.org/. Accessed: 2022-08-23.
[62] Databricks. https://www.databricks.com/product/managed-mlflow. Accessed: 2022-08-23.
[63] Iguazio. https://www.iguazio.com/. Accessed: 2022-08-23.
[64] Weights and Biases. https://wandb.ai/site. Accessed: 2022-08-23.
[65] polyaxon. https://polyaxon.com/. Accessed: 2022-08-23.
[66] snorkel. https://www.snorkel.org/. Accessed: 2022-08-23.
[67] A. Taeihagh and H. S. M. Lim, "Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks," *Transport Reviews*, vol. 39, no. 1, pp. 103–128, 2019.
[68] International Organization for Standardization, "ISO/IEC JTC 1/SC 42 Artificial intelligence." https://www.iso.org/committee/6794475.html, n.d. Accessed: 2022-05-09.
[69] ETSI, "Securing Artificial Intelligence (SAI)." https://www.etsi.org/technologies/securing-artificial-intelligence, n.d. Accessed: 2022-08-01.
[70] R. Wright, "Comparing European and Canadian AI regulation," 2021.
[71] M. C. Paulk, B. Curtis, M. B. Chrissis, and C. V. Weber, "Capability maturity model, version 1.1," *IEEE software*, vol. 10, no. 4, pp. 18–27, 1993.
[72] J. Samuel, K. Aalab, and J. Jaskolka, "Evaluating the soundness of security metrics from vulnerability scoring frameworks," in *19th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, (Guangzhou, China), pp. 442–449, 2020.
[73] OWDT, "Cyber security professionals underestimate vulnerabilities until hit." https://owdt.com/cyber-security-professionals-underestimate-vulnerabilities-until-hit/, 2022. Accessed: 2022-08-01.
[74] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. C. Ferrer, "The deepfake detection challenge (DFDC) dataset," 2020.
[75] Y.-S. Martin and A. Kung, "Methods and tools for gdpr compliance through privacy and data protection engineering," in *2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pp. 108–111, 2018.
[76] compliance.ai, "compliance.ai." https://www.compliance.ai/, n.d. Accessed: 2022-08-15.
[77] G. Elahi and E. Yu, "Modeling and analysis of security trade-offs–a goal oriented approach," *Data & Knowledge Engineering*, vol. 68, no. 7, pp. 579–598, 2009.
[78] I. Malakhov, A. Marin, S. Rossi, and D. Smuseva, "On the use of proof-of-work in permissioned blockchains: Security and fairness," *IEEE Access*, vol. 10, pp. 1305–1316, 2022.
[79] L. Chen, S. Thombre, K. Järvinen, E. S. Lohan, A. Alén-Savikko, H. Leppäkoski, M. Z. H. Bhuiyan, S. Bu-Pasha, G. N. Ferrara, S. Honkala, J. Lindqvist, L. Ruotsalainen, P. Korpisaari, and H. Kuusniemi, "Robustness, security and privacy in location-based services for future IoT: A survey," *IEEE Access*, vol. 5, pp. 8956–8977, 2017.
[80] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Comput. Surv.*, vol. 54, jul 2021.
[81] O. A. Osoba and W. Welser IV, *An intelligence in our image: The risks of bias and errors in artificial intelligence*. Rand Corporation, 2017.
[82] M. Choraś, M. Pawlicki, D. Puchalski, and R. Kozik, "Machine learning–the results are not the only thing that matters! what about security, explainability and fairness?," in *International Conference on Computational Science*, pp. 615–628, Springer, 2020.
[83] F. Calmon, D. Wei, B. Vinzamuri, K. Natesan Ramamurthy, and K. R. Varshney, "Optimized pre-processing for discrimination prevention," *Advances in neural information processing systems*, vol. 30, 2017.
[84] F. Kamiran and T. Calders, "Data preprocessing techniques for classification without discrimination," *Knowledge and information systems*, vol. 33, no. 1, pp. 1–33, 2012.
[85] B. H. Zhang, B. Lemoine, and M. Mitchell, "Mitigating unwanted biases with adversarial learning," in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 335–340, 2018.
[86] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, "Explaining explanations: An overview of interpretability of machine learning," in *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 80–89, 2018.
[87] D. Gunning, "Explainable artificial intelligence (XAI)," *Defense Advanced Research Projects Agency (DARPA)*, vol. 2, no. 2, p. 1, 2017.
[88] B. Hrnjica and S. Softic, "Explainable AI in manufacturing: A predictive maintenance case study," in *Advances in Production Management Systems. Towards Smart and Digital Manufacturing* (B. Lalic, V. Majstorovic, U. Marjanovic, G. von Cieminski, and D. Romero, eds.), (Cham), pp. 66–73, Springer International Publishing, 2020.
[89] S. Srinivasan, P. Arjunan, B. Jin, A. L. Sangiovanni-Vincentelli, Z. Sultan, and K. Poolla, "Explainable AI for chiller fault-detection systems: Gaining human trust," *Computer*, vol. 54, no. 10, pp. 60–68, 2021.
[90] T. Orekondy, B. Schiele, and M. Fritz, "Knockoff nets: Stealing functionality of black-box models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
[91] M. Juuti, S. Szyller, S. Marchal, and N. Asokan, "PRADA: Protecting against DNN model stealing attacks," in *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*, pp. 512–527, 2019.

[92] D. Tran, J. Liu, M. W. Dusenberry, D. Phan, M. Collier, J. Ren, K. Han, Z. Wang, Z. Mariet, H. Hu, *et al.*, "Plex: Towards reliability using pretrained large model extensions," *arXiv preprint arXiv:2207.07411*, 2022.

[93] D. Kaur, S. Uslu, K. J. Rittichier, and A. Durresi, "Trustworthy artificial intelligence: A review," *ACM Comput. Surv.*, vol. 55, jan 2022.

[94] D. Hendrycks, N. Carlini, J. Schulman, and J. Steinhardt, "Unsolved problems in ML safety," *arXiv preprint arXiv:2109.13916*, 2021.

[95] S. Mohseni, H. Wang, Z. Yu, C. Xiao, Z. Wang, and J. Yadawa, "Practical machine learning safety: A survey and primer," *arXiv preprint arXiv:2106.04823*, 2021.

[96] E. Strubell, A. Ganesh, and A. McCallum, "Energy and policy considerations for modern deep learning research," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 13693–13696, 2020.

[97] A. van Wynsberghe, "Sustainable AI: AI for sustainability and the sustainability of AI," *AI and Ethics*, vol. 1, pp. 213–218, Aug 2021.

[98] H. Myrbakken and R. Colomo-Palacios, "DevSecOps: A Multivocal Literature Review," in *Software Process Improvement and Capability Determination* (A. Mas, A. Mesquida, R. V. O'Connor, T. Rout, and A. Dorling, eds.), vol. 770, pp. 17–29, Cham: Springer International Publishing, 2017. Series Title: Communications in Computer and Information Science.

[99] R. N. Rajapakse, M. Zahedi, M. A. Babar, and H. Shen, "Challenges and solutions when adopting DevSecOps: A systematic review," *Information and Software Technology*, vol. 141, p. 106700, Jan. 2022.

[100] M. Sánchez-Gordón and R. Colomo-Palacios, "Security as Culture: A Systematic Literature Review of DevSecOps," in *Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops*, (Seoul Republic of Korea), pp. 266–269, ACM, June 2020.

[101] M. Al-Qatf, Y. Lasheng, M. Al-Habib, and K. Al-Sabahi, "Deep learning approach combining sparse autoencoder with svm for network intrusion detection," *IEEE Access*, vol. 6, pp. 52843–52856, 2018. cited By 191.

[102] D. Gibert, C. Mateu, and J. Planes, "The rise of machine learning for detection and classification of malware: Research developments, trends and challenges," *Journal of Network and Computer Applications*, vol. 153, 2020. cited By 112.

[103] Y. Kayode Saheed, A. Idris Abiodun, S. Misra, M. Kristiansen Holone, and R. Colomo-Palacios, "A machine learning-based intrusion detection for detecting internet of things network attacks," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 9395–9409, 2022. cited By 0.

[104] A. I. Newaz, A. K. Sikder, M. A. Rahman, and A. S. Uluagac, "A survey on security and privacy issues in modern healthcare systems: Attacks and defenses," *ACM Trans. Comput. Healthcare*, vol. 2, jul 2021.

[105] J. Cui, L. S. Liew, G. Sabaliauskaite, and F. Zhou, "A review on safety failures, security attacks, and available countermeasures for autonomous vehicles," *Ad Hoc Networks*, vol. 90, p. 101823, 2019. Recent advances on security and privacy in Intelligent Transportation Systems.

[106] N. Papernot, P. McDaniel, A. Sinha, and M. P. Wellman, "Sok: Security and privacy in machine learning," in *2018 IEEE European Symposium on Security and Privacy (EuroS&P)*, pp. 399–414, IEEE, 2018.