

The Impact on the Security of Cloud Computing Platforms When Deploying Artificial Intelligence and Recommendations

Qiuyue Liu

Engineer, Internet Center, Institute of Technology and Standards, China Academy of Information and Communications Technology.

liuqiuyue@caict.ac.cn

Abstract. With the rapid development of artificial intelligence (AI), deployment it on cloud platforms has become a prevalent trend. But it also brings risks to cloud platforms. This study conducts an in-depth analysis of these risks, covering aspects such as the security of cloud platform architectures, data security, and supply chain security. It also puts forward coping strategies, such as improving policies, regulations, and industry standards, and promoting ecological construction. Future research can be carried out from aspects like cross-cloud platform security management and cross-cloud data security.

Keywords: AI; Cloud Platform Risks; MaaS; Data Security.

1. Introduction

1.1 Research Background

Today, Artificial intelligence (AI) occupies a crucial position in scientific and technological fields, and the enthusiasm for its research continues to rise. With the continuous progress of science and technology, AI has become the most transformative technological force in the current era. The text-to-video model Sora officially released by OpenAI (the Open Artificial Intelligence Research Center in the United States), has attracted attention from all sectors. Dreamina has triggered extensive discussions once again. Applications like text-to-video and text-to-music have fueled the incessant expansion of the AI boom. The 2024 Top Ten Frontier Technology Trends in AI makes people full of expectations for the future of AI.

In order to fulfill the high demands for AI applications and realize rapid deployment and extensibility, more and more enterprises and organizations choose to deploy AI on cloud platforms. Cloud platforms are endowed with the merits of flexible resource allocation and convenient maintenance and management. The price war of large language models in China has further spurred the acceleration of AI migrating to the cloud. With the incessant decline in the average price of large language model APIs, the superiority of cloud deployment solutions will be more conspicuous in the future. Nonetheless, this deployment mode also brings along a series of risks.

1.2 Research Purpose

This study aims to conduct an in-depth analysis of the changes introduced by the deployment of AI on cloud platforms and the risks to cloud platforms brought about by these changes, and propose effective coping strategies to ensure the secure deployment and application of AI and cloud platforms.

2. Characteristics of Cloud Platforms for Deploying AI

The deployment of AI on cloud platforms requires specific configurations to meet the needs of its high-performance computing and complex algorithms. To deploy AI on cloud computing platforms and ensure its smooth operation, there are three major requirements for cloud platforms.

2.1 Hardware Environment of Cloud Platforms on which AI Deployment Depends

GPU chips play a crucial role in the deployment of AI. GPUs have powerful parallel computing capabilities and can accelerate the training and inference processes of AI algorithms such as deep learning. According to statistics, using NVIDIA GPU chips for deep learning training can be dozens or even hundreds of times faster than using traditional CPUs.

The data sets and model files of AI are usually relatively large and also need to store training data, model parameters, and inference results. Cloud platforms need to be configured with sufficient storage space. Meanwhile, the storage needs to have a fast read and write speed to reduce data access latency.

In addition, cloud platforms need to have high-bandwidth network connections to ensure the rapid transmission of data and the efficient training of models.

2.2 Software Environment of Cloud Platforms on which AI Deployment Depends

The operating system that supports AI development frameworks is also one of the key configurations. Many AI development frameworks, such as TensorFlow and PyTorch, have better compatibility and performance on Linux systems.

The AI algorithm library is an important part of AI deployment. These algorithms are a series of pre-written functions and algorithms that facilitate developers to quickly build and deploy AI models. There are dozens of common AI algorithm libraries. Among them, Scikit-learn is an open-source Python machine learning library that encapsulates various machine learning tasks. OpenCV consists of a series of C functions and a small number of C++ classes and is an algorithm library based on the Apache2.0 license, implementing numerous general algorithms in image processing and computer vision.

Development tools and platforms are also essential parts for the deployment of AI on cloud platforms. Development tools and platforms on cloud platforms usually provide functions such as code version control, collaborative development, and automated deployment, which can improve development efficiency and team collaboration ability. Jupyter Notebook is a widely used interactive development environment which is widely used for data cleaning and conversion, numerical simulation, statistical modeling, data visualization, machine learning, etc.

2.3 Cloud Services on which AI Deployment Depends

The simplest and most popular way for enterprises to start using AI technology is to use cloud-based AI, Model as a Service (MaaS). MaaS has the advantages of convenience, high efficiency, and strong scalability. Currently, it is an important form of AI applications. It can complete the pre-training of many tasks and provide trained models to customers in the form of services, solving the problem that complex tasks require a large amount of data and powerful computing resources for model training and greatly reducing the threshold for using AI technology.

Fig. 1 shows the functional framework diagram of MaaS.

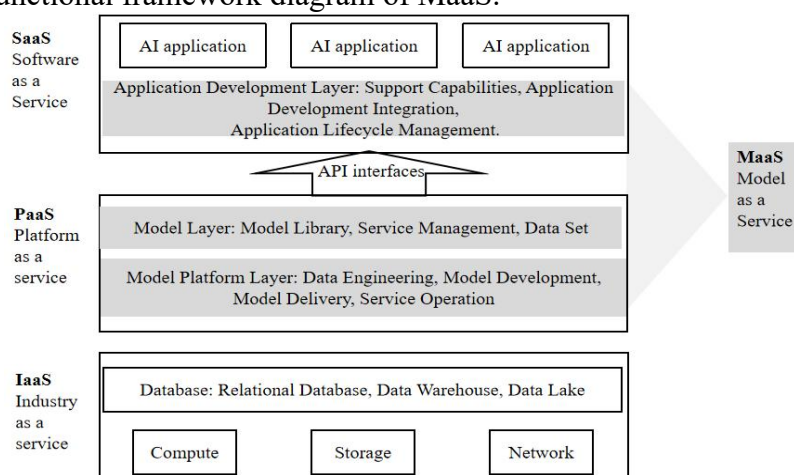


Fig. 1 shows the functional framework diagram of MaaS

In short, the deployment of AI on cloud platforms requires the support of various configurations such as GPU chips, operating systems that support AI development frameworks, AI algorithm libraries, and development tools and platforms. The performance and security of these configurations directly affect the deployment effect and operation stability of AI.

3. Impact of AI on the Security of Cloud Platform Architectures

3.1 Impact on the Security of Cloud Platform Architectures

The introduction of AI has had multiple impacts on cloud platform architectures. In terms of virtualization security, although virtualization technology has brought about an improvement in resource utilization and flexibility for cloud platforms, it has also brought new security challenges. With the increase in AI training tasks, multiple virtual machines may run simultaneously, and the complexity of the associations between virtual machines further increases. Attackers can take advantage of the interactions and resource competition among virtual machines to find vulnerabilities and launch attacks.

In terms of the isolation among cloud platform tenants, the deployment of AI may affect the isolation effect among tenants. AI applications usually need to access a large amount of data, computing resources, network resources, etc., which may lead to more resource competition and sharing. If the isolation measures are not in place, an AI application of one tenant may affect the security and performance of other tenants. For example, an AI training task of one tenant may occupy excessive network bandwidth and affect the network communication quality of other tenants.

In terms of cloud platform boundary protection, the introduction of AI may make the protection strategies of cloud platform boundary firewalls more complex. There may be more network traffic and data exchanges, and traditional firewall rules are difficult to effectively cope with these changes. During the AI training process, there may be frequent communications with external data. If the firewall does not update its rules in a timely manner, it may increase the exposure surface of the cloud platform and increase the risk of data leakage.

3.2 Impact on Data Security

Data leakage is originally one of the major risks faced by cloud platforms. The deployment of AI will increase the channels for data leakage.

Risk of unencrypted sensitive information: The data processed by AI may contain sensitive information such as users' personal privacy data and enterprises' commercial secrets. The storage of cloud platforms must provide powerful data encryption functions. In addition, with the development of AI, cross-cloud deployment may become more common, and data transmission between cloud platforms also needs to be encrypted.

Database security vulnerabilities are also an important channel for data leakage. AI applications usually need to access databases to obtain training data and store model parameters. If the security measures of databases are not in place, data may be stolen. For example, if the user authentication and authorization mechanisms of databases are flawed, unauthorized users may be able to access sensitive data in databases. In addition, vulnerabilities in database software may also be exploited by attackers to steal or destroy data.

Risk of unauthorized access: In terms of database access, multi-user MaaS is slightly different from other PaaS services. All MaaS configuration file templates designed by MaaS developers are stored in the IaaS layer. MaaS directly retrieves data from the IaaS layer database. Model developers on the platform may have full data access and modification rights, while ordinary users may only have read rights. The lack of a strict access control mechanism will lead to unauthorized access to data. Due to the special mechanism of MaaS accessing databases, it can lead to more

serious data leakage. The cloud storage where MaaS is deployed should be able to accurately control who can access, modify, and delete data according to factors such as users' roles, permission levels, and data sensitivity.

3.3 Impact of the AI Dependent Environment on Cloud Platform Security

The deployment of AI on cloud platforms depends on specific configurations. Improper configurations may also trigger a series of security risks, mainly facing denial-of-service attacks, data poisoning, etc.

Denial-of-service attacks: Attackers may launch denial-of-service attacks against the cloud platform environment on which AI depends, especially against high-performance computing resources such as GPU chips. Since the training and inference processes of AI have a huge demand for computing resources, once suffering from a denial-of-service attack, it may lead to the interruption of training and inference tasks and seriously affect the normal operation of business. For example, attackers can control a large number of zombie hosts to send a large number of requests to the cloud platform, exhausting computing resources such as GPU chips and unable to provide services for legitimate AI tasks. According to relevant data statistics, the number of cloud platform service interruption events caused by denial-of-service attacks is on the rise every year, bringing huge losses to enterprises.

Software environment security risks: AI applications deployed on the cloud have a high dependence on underlying development frameworks, operating systems, etc. The security of frameworks and operating systems that support AI development is also very important. If their security vulnerabilities are exploited by attackers, it may also lead to the successful implementation of data poisoning attacks.

3.4 Impact on Supply Chain Security

In terms of hardware, international giants such as NVIDIA, Intel, Xilinx, and Google currently account for the main share in the AI chip market. Domestic enterprises have been accelerating the layout in the field of AI hardware in recent years. At present, domestic enterprises are developing relatively fast in fields such as edge inference chips, but overall, there is still a large gap from the international advanced level.

In terms of software, American technology giants such as Google and Microsoft have invested huge amounts in AI algorithms and software development. Domestic academia and enterprises have also successively launched self-developed large models based on large language models. Currently, the software AI depends on is mainly controlled by American, such as machine learning frameworks TensorFlow and PyTorch, data analysis software Pandas, NumPy and Scikit-learn, natural language processing libraries NLTK (Natural Language Toolkit) and Spacy, and computer vision library OpenCV (Open Source Computer Vision Library).

Compared with the software and hardware of other cloud computing platforms, the domestic autonomous controllability of the software and hardware on which AI depends is lower, and the risks in aspects such as supply interruption, technical barriers and limited development, hardware backdoors, and data sovereignty issues are more serious. The software also faces the following risks:

Operating system and system software risks: MaaS and data preprocessing tools, data annotation tools, monitoring tools, etc. run on operating systems and various system software. These operating systems are basically Windows, Linux, and macOS. If the operating system suppliers stop updating and supporting (such as Microsoft stopping updating certain old versions of Windows Server), MaaS and other tools may face security vulnerabilities and compatibility issues.

Model training and management software risks: MaaS depends on specific model training software (such as TensorFlow, PyTorch, etc.) and model management software. If these software have vulnerabilities, version compatibility issues, or license changes, it will bring risks to MaaS. For example, software vulnerabilities may be exploited by hackers to steal models or user data; version

incompatibility may lead to the inability of models to train or infer normally; license changes may increase software use costs, affect the price competitiveness of MaaS services, or face legal risks.

Supply chain process attack risks: Malicious attackers may launch attacks against the software supply process, such as implanting malicious code during software updates. This situation is extremely dangerous for MaaS and other tools, because once attacked, it may lead to serious consequences such as the modification of model services and the leakage of user data.

4. Coping Strategies and Suggestions

With the continuous development of the deployment of AI on the cloud, it is recommended to further improve policies, regulations, and standards to guide relevant parties to strengthen the security management of cloud platforms and reduce risks.

4.1 Improving Policies and Regulations

Clarify the responsibilities of all parties. Formulate policies and regulations to clarify the responsibilities and obligations of cloud computing service providers, AI developers, and users in terms of cloud platform security. For example, require cloud computing service providers to take effective security measures to protect user data, AI developers to ensure the security of their applications, and users to comply with security regulations and properly keep their account passwords.

Further optimize the security assessment of cloud computing services. Conduct special security assessments for AI services on the cloud, regularly assess the security management capabilities of cloud computing service providers, and regularly track the effectiveness of security protection measures for AI on the cloud.

4.2 Improving the Industry Standard System

In terms of the construction of cloud security technology standards, organize relevant institutions to formulate security standards for AI on the cloud as soon as possible to regulate cloud computing service providers to implement security technology measures for AI services on the cloud. For example, it is stipulated that advanced encryption technologies, access control technologies, and vulnerability management technologies must be adopted to ensure the security of cloud platforms. Meanwhile, require cloud computing service providers to establish security monitoring and emergency response mechanisms, effectively back up data, and timely discover and handle security events.

In terms of data standard construction, improve the data standard system to regulate the storage, transmission, and use of data on cloud platforms. Require that data must be stored in a certain format and specification, encryption technologies must be adopted during the transmission process, cloud platforms should support data storage encryption technologies, and cloud users must comply with relevant laws, regulations, and user authorizations when using data.

In terms of AI security standards, improve the AI security standard system to regulate the deployment and application of AI on cloud platforms. Require AI developers to conduct security tests on applications to ensure that they will not pose a threat to the security of cloud platforms. Meanwhile, stipulate that AI applications must comply with relevant laws, regulations, and ethical norms and must not be used for illegal purposes.

4.3 Promoting Ecological Construction

Strengthen the ecological construction of software and hardware on which AI depends from the following aspects to form a complete AI industry and promote the rapid and healthy development of domestic AI:

Strengthen the independent innovation ability of software and hardware. Increase the research and development (R&D) efforts for AI chips, improve the performance and energy efficiency of

chips, support the development of domestic chip enterprises, strengthen the R&D of chip manufacturing processes, and improve the manufacturing level of domestic chips. Encourage domestic enterprises and scientific research institutions to develop independent AI algorithms and frameworks, improve the efficiency and accuracy of algorithms, and reduce dependence on foreign algorithms and frameworks. Establish an open-source community for AI software to promote exchanges and cooperation among developers, share code and technology, and improve the development efficiency and quality of software.

Strengthen the construction of the infrastructure of the AI industry. Coordinate the development of AI cloud service platforms to provide convenient cloud computing services for enterprises and developers, reduce repeated investments, reduce the development and deployment costs of AI applications, unify security standards, and ensure the implementation of security measures. Establish AI testing and verification platforms to conduct performance tests, security assessments, and reliability verifications on AI products to ensure their quality and security.

Strengthen the promotion of security protection measures for AI on the cloud. Further promote cloud platform maintainers to use intrusion detection systems (IDS) and intrusion prevention systems (IPS) to conduct real-time monitoring of the network traffic of cloud platforms and timely discover abnormal behaviors. Timely monitor the usage of GPU chips, access records of data storage, etc., and discover potential denial-of-service attacks and data poisoning behaviors. Formulate a complete emergency response plan and data backup and recovery plan for the security of AI on the cloud. Once a security event is discovered, immediately start the emergency response process, take measures to isolate the affected systems, and restore data and services. When suffering from a denial-of-service attack, it is possible to quickly restore the normal service of the cloud platform through technical means such as traffic cleaning. For data poisoning attacks, timely clean the contaminated training data and retrain the AI model.

5. Conclusion and Outlook

This study has deeply explored the risks to cloud platforms brought about by the deployment of AI on the cloud and put forward corresponding coping strategies. Through the analysis of the environment on which AI depends, it has analyzed the impact of deploying AI on the security of cloud platforms and put forward protection suggestions, providing support for ensuring the secure deployment of AI on cloud platforms. With the continuous development of AI technology and the continuous evolution of cloud platform architectures, new security risks may continue to emerge. We need to continuously pay attention to these changes and constantly explore new security technologies and solutions.

Future research focuses and directions can be carried out from the following two aspects: Research on security protection systems based on AI that can automatically adapt to the dynamic changes of cloud platforms and continuously expanding application scenarios, and automatically adjust security policies and protection measures; As the deployment of AI on cloud platforms becomes increasingly common, the multi-cloud environment will become a trend. Future research will focus on security strategies in the multi-cloud environment, strategies for unified security management and protection among different cloud platforms, and security issues of data flow and sharing in the multi-cloud environment.

Acknowledgments

This work was supported by the National Key R&D Program of China under Grant 2023YFB2906402.

References

- [1] Daniel Hand. "2024 年生成式 AI 及云领域的五大趋势." 软件和集成电路 12(2023).

- [2] 郑黎明,潘文联,成楠. (2022). 人工智能技术应用及其发展趋势. 科技与创新(17).
- [3] Manju Ramesh,Dheeraj Chahal,Rekha Singhal. (2023). Multicloud Deployment of AI Workflows Using FaaS and Storage Services..
- [4] 志刚. (2018). 人工智能即服务:当人工智能遇到云计算. 大众科学(7).
- [5] 李喆. (2023). 大模型赋能数字经济. 《质量与市场》(17).
- [6] Gan W,Wan S,Yu P S. Model-as-a-service(MaaS): A survey[J]. arXiv preprint arXiv:2311.05804,2023
- [7] Li Z,Yang C,Huang Q,et al. Building Model as a Service to support geosciences[J]. Computers,Environment and Urban Systems,2017,61: 141-152