Risk Analysis of Software Supply Chain in the Communication Industry Under Large Language Models

Chang Xing[1,*], Longbin Zhou[2], Xiuqing Li[1]
[1]China Mobile Communications Group Hebei Co., Ltd., Shijiazhuang, Hebei, China
[2]China Mobile Information Technology Co., Ltd., Beijing, China
**Corresponding Author.*

Abstract: The software supply chain of the communication industry under large language models involves multiple aspects, including product development, testing, and operation and maintenance management, which is a full-lifecycle activity. As an emerging technology, large language models have been widely applied in the software supply chains of various industries in recent years. For traditional software supply chains, they pose relatively significant data security risks to both upstream and downstream sectors. Research has shown that both environmental and data risks need to be analyzed and addressed to effectively reduce the frequency of risks and resolve a series of existing issues. This not only enhances the convenience of people's lives but also mitigates risks caused by data leakage.

Keywords: Large Language Models; Communication Industry Software; Supply Chain Risks

1. Introduction

In recent years, the development of large language models has set off waves time and again in the communication technology industry. Significant changes have occurred in the upstream and downstream of the software supply chain, and the security and stability of software supply chain operations are facing new challenges. As an indispensable part of China's national and social development, the security and quality of software communication directly affect the effectiveness and stability of industry development. Therefore, this study explores how to conduct safety management for communication industry software under large language models and reduce the series of potential security risks.

2. Principles and Functions of Large Language Models

2.1 Principles of Large Language Models

A large language model is a deeply empowered neural network integrating multi-dimensional linguistic intelligence and rule-based units, represented by GPT-4. Based on the pre-trained Transformer architecture, it analyzes and processes subsequent words and phrases according to the contextual content in the text, enabling intelligent content push and analysis to construct a neural network system. In the management of supply chain big data, the communication industry needs to adopt proactive and systematic approaches to strengthen technologies, restructure processes, improve the professional competence of staff, and build a new risk control framework through ecological collaboration and other aspects. This framework will adapt to the specific methods and standards of risk governance for new-type software supply chains in the AI era[1]. While enjoying the dividends of AI, it can improve the network security quality of the communication industry, identify the safest management methods for communication networks, and reduce poor software performance and user privacy violations caused by improper network data security management.

2.2 Functions of Large Language Models

Firstly, large language models integrate pre-training and fine-tuning strategies. In the pre-training phase, they can absorb and quickly process massive amounts of unlabeled text, conduct in-depth learning of linguistic structures and rules, and construct corresponding linguistic representations based on the model. In the fine-tuning phase, for specific tasks, data can be used to further optimize relevant content, with the ultimate goal of meeting specific application needs. Secondly, due to their ability to conduct in-

Copyright @ STEMM Institute Press

depth learning of massive linguistic data in a short time, large language models can accurately capture and understand knowledge and patterns of different languages, quickly comprehend and construct more complete and coherent text content. By selecting a large language model, it can automatically generate customized responses based on the input contextual scenarios, accurately addressing users' various questions and needs[2]. Thirdly, large language models can create scenarios consistent with text content according to contextual environments, covering multiple dimensions such as question answering and vivid description. This endows large language models with application potential in generating suggestions and risk assessment, continuously expanding their practical value and scope of application. Additionally, large language models possess efficient and rapid information extraction capabilities, helping users quickly grasp the core points in relatively long texts through systematic analysis.

3. Risk Analysis of Software Supply Chain in the Communication Industry Under Large Language Models

3.1 Identification of Data Leakage Risks

Large language models rely on extensive datasets in the risk analysis of the communication industry's software supply chain. These datasets include users' personal information (such as names and contact details) and even activity trajectories when using the software. Such content is inherently sensitive privacy information, so staff must attach great importance to analyzing, understanding, and effectively protecting data security and privacy. Data leakage is highly likely under large language models. For example, after introducing large language models into the communication industry, intelligent customer service systems require massive user data for risk assessment, including personal identification information and transaction records[3]. Inadequate security measures during data storage and processing may lead to large-scale data leakage, where illegal actors could use the data for fraud or other malicious activities. This not only causes economic losses to customers but also severely damages the reputation and credibility of communication industry software. Currently, data security and privacy protection are extremely important, and the most rigorous and secure strategies must be adopted-including regular data audits, access control, and data encryption-to effectively prevent the frequent recurrence of similar information leakage incidents. Communication industry software involves a large amount of sensitive personal information, such as transaction records and identification data. The use of large language models for processing these data may result in violations of data-related laws and regulations. Additionally, some data may be protected by copyright, and large language models may use such data without authorization.

3.2 Risks to Code Quality and Maintainability

The application of large language models in communication industry software may lead to reduced code consistency. Code generated by different developers or the same developer at different times may have significant differences in style, architectural patterns, and error-handling mechanisms, lowering the overall quality of the project's codebase and increasing long-term maintenance costs. Furthermore, there are intellectual property and compliance risks: code generated by large language models may inadvertently plagiarize copyrighted source code, triggering intellectual property disputes. In the communication field involving strict export controls (e.g., encryption algorithms) or regional regulations, the use of generated code from unknown sources may result in compliance violations. Another risk is model hallucinations and code defects: large language models may generate "hallucinatory" code that is grammatically correct but logically flawed, contains security vulnerabilities (e.g., buffer overflows, SQL injection variants), or fails to comply with communication protocol standards. If developers over-rely on such code without sufficient review, these defects will directly enter the product, forming "AI-native vulnerabilities." There are also obvious risks of training data contamination and backdoors: if malicious code, open-source projects with hidden backdoors, or biased data are mixed into the training dataset of large language models, the models may "learn" and replicate these patterns. This leads to generated code containing hidden backdoors or logic bombs, threatening the security of network infrastructure.

4. Risk Response Strategies for Software Supply Chain in the Communication Industry Under Large Language Models

4.1 Response to Data Leakage Risks

The communication industry needs to establish a comprehensive data management system, requiring large language models to quickly and scientifically classify and store all data, implement access controls, and adopt advanced encryption technologies and security protection measures. In environments where large language models are used in the communication industry, all data must be effectively transmitted, stored, and processed to enhance data application security. In practical data use, considerations must also include fine-grained data management and encrypted access permissions, ensuring that all user behaviors (such as accessing or retrieving information) are protected and personal information is not leaked. Additionally, it is necessary to strengthen safety education for all employees, conduct regular training on operational standards, and fully integrate all external data—especially through rigorous data cleaning and preprocessing. Only by strictly implementing these processes can potential harmful information be eliminated in a timely manner, ensuring data purity and security[4].

4.2 Response to Technical, Environmental, and Economic Risks

The application of large language models in the software supply chain of China's communication industry poses significant economic and environmental risks to upstream and downstream sectors. Currently, it is necessary to establish a human-led, AI-assisted review process, strengthening the security review of all code configurations and scripts generated by large language models. The dual collaboration of humans and AI can improve the quality of technical inspections. Large language models also need to conduct independent and secure evaluations of key communication software, including fuzz testing, symbolic execution, and formal verification, with particular attention to protocol consistency and boundary conditions. Furthermore, it is necessary to promote secure and controllable localized dedicated models to reduce reliance on uncontrollable large models. To address environmental and economic risks, it is essential to balance risks, adopt energy-efficient algorithms to reduce model parameters, and prioritize green energy and carbon offset methods to power data centers. Continuously promoting and developing AI application strategies can effectively reduce energy consumption through optimized algorithms, improve hardware energy efficiency, and avoid negative environmental impacts[5].

5. Conclusion

In summary, the improvement of software supply chain application effects in the communication industry is closely related to large language models, but these models also bring significant risks. Such risks are multi-dimensional, in-depth, and interrelated, with obvious loopholes in traditional management approaches. Under the new supply chain management framework, it is necessary to identify the optimal and safest management ideas to enhance overall management effectiveness, thereby addressing new challenges such as model data issues, data contamination, and ambiguous accountability.

References

[1] CHEN Gang, SHEN Weijiang, CAI Liming, et al. Innovative Application of Large Language Models in Industrial Software Development, Operation, and Maintenance Services—Practice of ChatOPS Agents[J]. Digital Transformation, 2025, 2(11): 52-58.

[2] ZHANG Tianyi, ZHOU Tong, ZHANG Chenxi, et al. LLM-Extractor: A Constraint Extraction Method Between Software Configurations Based on Large Language Models[J/OL]. Journal of Software, 1-29.

[3] ZHANG Bin, LI Runhao, FENG Chao. An Automatic Heap Memory Layout Method for Software Vulnerabilities Based on Large Language Models[J/OL]. Computer Engineering, 1-16[2025-12-18].

[4] WANG Ying, LI Juntao, LOU Yiling, et al. An Empirical Study on Large Language Model-Assisted Software Development[J/OL]. Computer Applications and Software, 1-11.

[5] HE Ke, JIANG Yazhen, LI Liangchen. Research on Intelligent Question-Answering Workflow of Forging Simulation Software Constructed Based on

the Coze Platform — Collaborative Application of Large Language Models and Structured Knowledge Bases[J]. Forging & Stamping, 2025(07): 20+22+24.