

Contact

18982180486 (Home)  
mixdeers@gmail.com

www.linkedin.com/in/bitliu  
(LinkedIn)  
www.liuxunzhuo.com (Personal)

Top Skills

Large Language Models (LLM)  
AI Gateway  
API Gateway

Languages

英语 (Limited Working)

Certifications

Istio Fundamentals  
CNCF Ambassador  
Envoy Fundamentals  
Program Committee Member:  
KubeCon + CloudNativeCon + Open  
Source Summit + AI\_dev China 2024

Honors-Awards

Gold Asia International Mathematical  
Olympiad Open Contest

Xunzhuo Liu

LLM Inference. Author of vLLM Semantic Router, Chair of K8S AI  
Gateway WG. AI Networks at Tencent  
Chengdu, Sichuan, China

Summary

Bit is exploring the frontier technologies in combination of networking  
and LLM at Tencent. He is currently working on AI Infrastructure.  
Previously, he has been involved in research at UESTC NLP Lab.

Bit is leading the development of vLLM Semantic Router, an  
intelligent auto reasoning router for Efficient LLM Inference on  
Mixture-of-Models, saving tons of cost by advanced routing  
algorithm.

As a CNCF Ambassador and Linux Foundation LFAPAC, Bit serves  
on the Envoy Gateway Steering Committee. He also maintains  
multiple projects including Envoy AI Gateway, vLLM AIBrix, Istio,  
Kiali, Aeraki-Mesh, and Merbridge, as well as the approver of  
Higress and MOSN. Additionally, Bit contributes as a Kubernetes  
Gateway API and Kubernetes Ingress2Gateway reviewer and  
member of Kubernetes.

Experience

vLLM

11 months

Author

January 2025 - Present (11 months)

Co-author of vLLM Semantic Router, an Intelligent Mixture-of-Models Router  
for Efficient LLM Inference

Maintainer

May 2025 - Present (7 months)

Maintainer of vLLM AIBrix, focusing on LLM Inference Networking

Kubernetes

Co-Chair and Founding Member - AI Gateway WG  
September 2025 - Present (3 months)

The AI Gateway Working Group focuses on the intersection of AI and networking, particularly in the context of extending load-balancer, gateway and proxy technologies to manage and route traffic for AI Inference.

腾讯

5 years

AI Networking

July 2022 - Present (3 years 5 months)

Chengdu

At Tencent Cloud, working at Tencent Kubernetes Engine (TKE) team, focusing on LLM Inference and Container Networking, including:

- LLM Inference: AI Gateway and RDMA, vRDMA.
- Linux Kernel: ipvs, iptables, netfilters.
- Container Network: IPAM, CNI, Kubeproxy, eBPF, NetworkPolicy.
- Access Layer: Service, Ingress and CLB.

Service Mesh

November 2021 - Present (4 years 1 month)

Chengdu

At Tencent Cloud, working at Tencent Cloud Mesh (TCM) team, focusing on Service Mesh.

API Gateway

December 2020 - November 2021 (1 year)

Nanshan District

At Tencent Games, working on API Gateway built with EnvoyProxy and extension with WebAssembly.

Envoy

Maintainer

June 2022 - Present (3 years 6 months)

1. Steering Committee member of Envoy Gateway
2. Maintainer of Envoy Gateway and Envoy AI Gateway

The Linux Foundation

LFAPAC

December 2023 - Present (2 years)

Linux Foundation APAC Open Source Evangelist

Cloud Native Computing Foundation (CNCF)

7 years

CNCF Ambassador  
2018 - Present (7 years)  
China

Maintainer  
2018 - Present (7 years)  
China

Maintaining CNCF projects including Envoy Gateway, Envoy AI Gateway, Istio, Kiali, Aeraki-Mesh, and Merbridge.

University of Electronic Science and Technology of China  
NLP Research Assistant  
March 2019 - March 2020 (1 year 1 month)  
中国 成都

A research assistant in UESTC Nature Language Processing Lab, working on NLP/DL/ML Research.

---

## Education

电子科技大学  
Bachelor's degree, Software Engineering · (September 2018 - September 2022)

Chengdu NO.7 High School 成都市第七中学  
Chengdu No.7 High School in Linyin Road · (September 2015 - June 2018)