

# Xiaoyu Chu

Amsterdam, the Netherlands | Machine Learning Engineer | Ph.D. Computer Science

☎ (+31) 06-4528-2610 | ✉ [chuxiaoyu123@gmail.com](mailto:chuxiaoyu123@gmail.com) | 🌐 [github.com/chuxiaoyu](https://github.com/chuxiaoyu) | in [linkedin.com/chuxiaoyu](https://www.linkedin.com/in/chuxiaoyu)

## SUMMARY

---

ML systems engineer specializing in building data and ML/LLM pipelines for analyzing datacenter telemetry and system reliability. Hands-on experience developing monitoring frameworks, LLM-powered data extraction pipelines, and time-series prediction. Skilled in MLOps, AgentOps, and AIOps, passionate about building reliable and observable AI systems.

## SKILLS

---

**Programming:** Python, TypeScript, SQL

**Big Data:** PySpark, MongoDB, MySQL, Jupyter, NumPy, Pandas, Matplotlib, Seaborn

**AI/ML:** PyTorch, TensorFlow, HuggingFace, Scikit-learn, Transformer

**LLM/Agent:** LangChain, LangGraph, Context Engineering, Tool / Function Calling, RAG, Memory, Fine-Tuning

**DevOps:** FastAPI, Git, GitHub, Linux, Docker, CI/CD, AWS, GCP, Vertex AI, SLURM, Prometheus

## EXPERIENCE

---

### Ph.D. Candidate in Computer Systems

Mar 2022 - Present

*Vrije Universiteit Amsterdam*

*Amsterdam, the Netherlands*

- Developed a data pipeline leveraging 6 major **LLM APIs (GPT, Claude, Gemini)** to automate the **extraction and classification of structured information** from over 3,000 cloud incident reports (including AWS, Azure, and GCP), enabling large-scale analysis of service failures. [\[GitHub\]](#)
- Developed an **automated monitoring framework (FAILS)** to track the availability of major LLM services (e.g., ChatGPT, Claude), analyzed over 500 incidents to identify critical failure-recovery patterns. [\[GitHub\]](#) [\[FAILS\]](#)
- Analyzed 26GB of performance, failure, and power usage data from ML workloads in a Dutch **HPC datacenter (SURFLisa)**, collected via **SLURM** and **Prometheus**. [\[GitHub\]](#) [\[PDF\]](#)
- Published **7** papers in high quality venues; presented 4 talks at international conferences and professional events.
- Mentored **10+** BSc. and MSc. theses, course, and honors projects, resulting in 4 publications. [\[LINK\]](#)

### NLP Research Engineer Intern

Mar 2021 - May 2021

*Orange Labs Beijing*

*Beijing, China*

- Prototyped a **deep structured semantic model** to improve the accuracy of personalized news recommendation.
- Optimized **MongoDB** operations and contributed to **API design** for data access, enhancing system usability.

## PROJECTS [\[GITHUB\]](#)

---

### Multimodal blog posts retrieval system | *FastAPI, BM25, CLIP, Docker, GCP, Vertex AI*

- Built a multimodal blog retrieval system supporting both **text and image queries**, combining **BM25** with **MiniLM** and **CLIP embeddings** as **hybrid ranking** to handle diverse query types.
- Implemented model pipelines based on **FastAPI**, achieving a best retrieval score of **MRR@3 at 0.8683**.
- Deployed the retrieval model to **GCP Vertex AI** as a scalable endpoint, with **model preloading**, **batch encoding**, and **matrix multiply** to minimize inference latency.

### ML for time series operational incident prediction | *PyTorch, Scikit-learn, Time Series*

[\[LINK\]](#)

- Built a **time-series anomaly prediction** pipeline on HPC system telemetry (CPU, memory, I/O, network) using **sliding-window** feature engineering to predict incidents 10 minutes ahead.
- Trained and evaluated **Logistic Regression, Random Forest, and XGBoost** models with class-imbalance handling and feature scaling, achieving F1-score up to 0.9.
- Designed an **end-to-end ML workflow** including data preprocessing, model evaluation, and a proposed cloud deployment architecture (AWS Lambda + monitoring pipeline) for real-time alerting.

### Fine-tuning LLMs for query segmentation | *PyTorch, BERT, NLP*

[\[LINK\]](#)

- Fine-tuned **DistilBERT-based transformer** models for query segmentation (NER) on geo-search queries using the HuggingFace transformers framework.
- Implemented **tokenization-label alignment** for subword tokenizers and trained models on BIO-formatted datasets, achieving 0.83 F1-score on multilingual queries.
- Built an **end-to-end NLP pipeline** including dataset preprocessing, model fine-tuning, and evaluation, with analysis of class imbalance, and domain-specific query characteristics.

## EDUCATION

---

### Vrije Universiteit Amsterdam

*Ph.D. in Computer Science*

Feb 2022 – Present  
*Amsterdam, the Netherlands*

### Beijing Normal University

*MSc. in Information Science, GPA: 89.60/100.00 (Top 20%)*

Sep 2018 – Dec 2021  
*Beijing, China*

### Lanzhou University

*B.A. in Accounting, GPA: 4.29/5.00 (Top 15%)*

Sep 2014 – Jun 2018  
*Lanzhou, China*

## SELECTED PUBLICATIONS [[GOOGLE SCOLAR](#)]

---

- **Xiaoyu Chu**, Shashikant Ilager, Yizhen Zang, Sacheendra Talluri, Alexandru Iosup. "Leveraging LLMs for Structured Information Extraction and Analysis of Cloud Incident Reports". ACM ICPEWiP'26. [[Paper](#)]
- Sacheendra Talluri, Dante Niewenhuis, **Xiaoyu Chu**, Jakob Kyselica, Mehmet Cetin, Alexander Balgavy, and Alexandru Iosup. "Cloud Uptime Archive: Open-Access Availability Data of Web, Cloud, and Gaming Services." IEEE TPDS'26. [[Paper](#)]
- Sándor Battaglini-Fischer\*, Nishanthi Srinivasan\*, Bálint László Szarvas\*, **Xiaoyu Chu**, Alexandru Iosup. "FAILS: A Framework for Automated Collection and Analysis of LLM Service Incidents". ACM HotCloudPerf'25. [[Paper](#)]
- **Xiaoyu Chu**, Sacheendra Talluri, Qingxian Lu, Alexandru Iosup. "An Empirical Characterization of Outages and Incidents in Public Services for Large Language Models". ACM ICPE'25. [[Paper](#)]
- **Xiaoyu Chu\***, Daniel Hofstätter\*, Shashikant Ilager, Sacheendra Talluri, Duncan Kampert, Damian Podareanu, Dmitry Duplyakin, Ivona Brandic, Alexandru Iosup (\*Equal contributions). "Generic and ML Workloads in an HPC Datacenter: Node Energy, Job Failures, and Node-Job Analysis". IEEE ICPADS'24. [[Paper](#)]

## AWARDS

---

- ACM/ICPE Travel Grant, ICPE, May 2023, 2025
- CSC Scholarship for Ph.D., VU, Mar 2022 - Feb 2026
- Summer School Scholarship, Technion, Jul 2017 - Aug 2017
- CSC Exchange Scholarship, KNU, Sep 2016 - Jun 2017
- 1st Class Merit Scholarship, BNU, Sep 2020
- 2nd Class Merit Scholarship, LZU, Sep 2015, 2016