

XIAOYU CHU

Amsterdam, the Netherlands | Willing to Relocate | Need KM Sponsorship

 chuxiaoyu123@gmail.com

 (+31) 06-4528-2610

 [linkedin/chuxiaoyu](https://www.linkedin.com/in/chuxiaoyu)

 [github/chuxiaoyu](https://github.com/chuxiaoyu)

SUMMARY

Data scientist with 7+ years experience in statistical analysis, machine learning, and natural language processing. Skilled in system modeling, operational analytics, and LLM applications.

SKILLS

Programming	Python, SQL, SPARQL
Big Data	MongoDB, PySpark
AI/ML	PyTorch, Langchain, Prompt Engineering, RAG, TensorFlow
DevOps	Git, GitHub, Linux, Cursor, Docker
Cloud	AWS, GCP

EDUCATION

Vrije University Amsterdam	Amsterdam, the Netherlands
Ph.D. Candidate in Computer Science	(Expected) 2026
Beijing Normal University (985)	Beijing, China
MSc. in Information Science	Sep 2018 - Dec 2021
GPA: 89.60/100.00 (Top 20%)	
Kyungpook National University	Daegu, South Korea
Exchange	Sep 2016 - Jun 2017
Lanzhou University (985)	Lanzhou, China
B.A. in Accounting	Sep 2014 - Jun 2018
GPA: 4.29/5.00 (Top 15%)	

LANGUAGES

- Chinese (Native)
- English (Advanced)
- Dutch (Duolingo)

PERSONALITY

- ENTP/INTJ
- Resilient
- Collaborative

EXPERIENCE

Ph.D. Researcher in Computer System Analysis.....
Vrije University Amsterdam Mar 2022 - Feb 2026, Amsterdam

- 1 Built pipeline and applied 6 LLM APIs (e.g., GPT, Claude, and Gemini) for structured information extraction, classification, and analysis from 3,000+ cloud incident reports. [\[GitHub\]](#)
- 2 Collected and analyzed failure-recovery patterns of 500+ incidents across 8 widely used LLMs; developed an automated monitoring framework to track the availability of major LLM services (e.g., ChatGPT, Claude, DeepSeek). [\[GitHub\]](#) [\[FAILS\]](#) [\[PDF1\]](#) [\[PDF2\]](#)
- 3 Analyzed 26GB of performance, failure, and power usage data from ML workloads in a Dutch HPC datacenter (SURFLisa), collected via SLURM and Prometheus. [\[GitHub\]](#) [\[PDF\]](#)
- 4 Published 6 papers in high quality venues; presented 4 talks at international conferences and professional events. [\[Pub\]](#) [\[Talk\]](#)
- 5 Mentored 10+ BSc. and MSc. thesis, course, and honor projects, resulting in 4 publications. [\[Link\]](#)

NLP Research Engineer Intern.....
Orange Labs Beijing Mar 2021 - May 2021, Beijing

- Designed and 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 and integration, enhancing system usability.

Data Analyst Intern.....
51Talk Online Education Sep 2020 - Dec 2020, Beijing

- Extracted and organized operational data via SQL on the Hive platform, improving efficiency and accuracy in data-driven decision-making.

PROJECTS

FAILS: Automating Analytics of LLM Service Incidents

- Led a team develop to the first open-sourced framework for automatic collecting and analyzing incident reports on different LLM services and providers. [\[FAILS\]](#) [\[PDF\]](#) [\[Slide\]](#)
- Highlights: (1) Providing 17 types of failure analysis, allowing users to explore temporal trends, and analyze service reliability metrics; (2) Leveraging advanced LLM tools to assist in data analysis and interpretation, enabling users to gain observations and insights efficiently.

Peer Review Text Classification for Literature Retrieval

- Used a pretrained BERT model and softmax classification to classify peer review text.
- Implemented a prototype literature retrieval system using Elasticsearch, integrating peer review information to improve retrieval accuracy and better meet user needs.

XIAOYU CHU

Amsterdam, the Netherlands | Willing to Relocate | Need KM Sponsorship

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

SUMMARY

Data scientist with 7+ years experience in statistical analysis, machine learning, and natural language processing. Skilled in system modeling, operational analytics, and LLM applications.

COURSES

Lifelong learner with completed Coursera courses in:

- Fundamentals of AI Agents Using RAG and LangChain [\[Link\]](#)
- AI Infrastructure and Operations Fundamentals [\[Link\]](#)
- Sequences, Time Series and Prediction [\[Link\]](#)
- Build a Modern Computer from First Principles [\[Link\]](#)

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

INTERESTS

- Bouldering
- History, Art
- CORTIS
- Fortune Telling

SELECTED PUBLICATIONS

The complete list can be accessed through my [Google Scholar](#).

- 1 **Xiaoyu Chu**, Shashikant Ilager, Yizhen Zang, Sacheendra Tallau, Alexandru Iosup. Leveraging LLMs for Structured Information Extraction and Analysis of Cloud Incident Reports. Under Review of ACM/IEEE UCC2025.
- 2 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 HotCloudPerf2025. <https://doi.org/10.1145/3680256.3721320>
- 3 **Xiaoyu Chu**, Sacheendra Talluri, Qingxian Lu, Alexandru Iosup. An Empirical Characterization of Outages and Incidents in Public Services for Large Language Models. ACM ICPE2025. <https://dl.acm.org/doi/abs/10.1145/3676151.3719372>
- 4 **Xiaoyu Chu***, Daniel Hofstätter*, Shashikant Ilager, Sacheendra Talluri, Duncan Kampert, Damian Podareanu, Dmitry Dulyakin, Ivona Brandic, Alexandru Iosup (*Equal contributions). Generic and ML Workloads in an HPC Datacenter: Node Energy, Job Failures, and Node-Job Analysis. IEEE ICPADS2024. <https://ieeexplore.ieee.org/abstract/document/10763608>

TALKS

- 1 An Empirical Characterization of Outages and Incidents in Public Services for Large Language Models. ICPE'25, Virtual, Toronto, Canada, May 7, 2025 [\[Slide\]](#)
- 2 Generic and ML Workloads in an HPC Datacenter: Node Energy, Job Failures, and Node-Job Analysis. ICPADS'24, Belgrade, Serbia, October 12, 2024 [\[Slide\]](#)
- 3 How Do ML Jobs Fail in Datacenters? Analysis of a Long-Term Dataset from an HPC Cluster. HotCloudPerf'23, Coimbra, Portugal, April 16, 2023 [\[Slide\]](#)

SUPERVISION

The complete list can be accessed through this [Link](#).

- 2025, Msc Thesis, Yiren Bai. Understanding Service Reliability of Large Language Models: An Empirical Characterization on Operator and User Reports. [\[PDF\]](#)
- 2024, Bsc Thesis, Yizhen Zang. Enhancing Operational Data Synthesis and Predictive Analysis in HPC Clusters Using Large Language Models. [\[PDF\]](#)
- 2023, Msc Thesis, Shekhar Suman. 'ODABler': Design and Evaluation of an Operational Data Analytics Framework for Energy-efficient management of Workloads in a Data Centre Simulator OpenDC. [\[PDF\]](#)