Xin Cui

chloexincui@gmail.com • Linkedin

Work Experience

Aug 2025 – present	**Research Engineer**, Tzafon AI, San Francisco, CA US
	Multimodal Post-training team
Jul 2024 – Jul 2025	**Research Engineer**, Character AI, Menlo Park, CA US
	LLM Post-training team
May 2018 – Mar 2024	**Senior Software Engineer**, LinkedIn, Mountain View, CA US
	Retention AI Platform, Job Search and Recommendation team
Jun 2015 – Dec 2016	**Full-Stack Software Engineer**, Indeed, Tokyo Japan
	SEO, Page Ranking team
July 2014 – Sep 2014	**Software Engineer Intern**, CITIC Securities, Shanghai China
	Quantitative Trading Service team

Projects Experience

Tzafon AI	### Computer Use Agent Fine Tuning
	• Led the post-training team in building a 0–1 SFT pipeline with the Qwen2.5-VLM series, supporting large-scale data ablation studies, grounding tasks, and OSWorld evaluations.
Character AI	### OSS Fine Tuning
	• Large scale distributed SFT training with ray + deepspeed on the latest Qwen3 dense series
	### Data Curation in Preference Alignment
	• Curated training data for Direct Preference Optimization (DPO), including privacy-preserving chat clustering to support pluralistic alignment. Designed scoring and filtering strategies to select safer and more balanced preference pairs. Fine-tuned language models using iterative DPO and its variants, monitored training via LM loss, KL divergence, and preference margin metrics, and evaluated model performance through online A/B experiments.
	### Safety Modeling
	• Full-stack development and training of transformer-based text models for safety enhancement. Generated synthetic training data through data distillation, LLM and human-assisted labeling, reject sampling, and majority voting. Trained classifier heads on fine-tuned language models and evaluated safety performance via offline prevalence metrics and online A/B testing. Collaborated cross-functionally on safety policy and ontology development, as well as red-teaming.
	Tech-stack – python, spark/k8s, apache beam/Google dataflow, SFT/DPO/pytorch, MySQL/Datalake/BigQuery, WandDB/Hex, Training(Ray/Hydra/Deepseed), Inference(vllm/roller), vscode/git/hex, dbt / apache airflow
LinkedIn	### Contextual Next-Best-Action Recommendation Platform and Product
	• Built contextually triggered cross-pillar (feed, jobs, edge building etc) action recommendations push notification for infrequent LinkedIn members, redesigned Next-Best-Action comms served as a platform and allows partner team to onboard, collaborate with Retention AI team. Impact sitewise +0.1% WAU.
	### Near realtime audience targeting
	• Improved A/B experiment member targeting efficient from offline manner to near realtime, improved data freshness from one day delay to P99 5-10 seconds. Since the A/B experiment platform is serving as one of the largest QPS in LinkedIn, the new pipeline is able to serve high data freshness with low latency.
	### New search index for job seeker and recruiter
	• Led and implemented fair chance employer filter and had multiple experiences on backend index building across LinkedIn's major products – Jobs and Recruiter. Brought Product team's novel search experience to reality.
	### Job's Data Streaming Platform

● Built data streaming pipeline with Apache Samza to standardize implicit job data/features. With combination of rule base and fastText model, able to achieve the data extraction with precision to 97% across all targeted benefits with avg recall 67%

Tech-stack Java, Scala, Python, Spring, Hadoop, Key-Document DB(Espresso), Kafka, Restful APIs, Apache Samza, Apache Lucene

Indeed Tokyo

ENRICH COMPANIES' PAGES WITH THEIR INSTAGRAM PHOTOS

● Link companies' official Instagram accounts to their Indeed accounts and show the latest Instagram photos on their Indeed company pages.

● Brought Hackathon project to Product. Two engineers and one PM hackathon project that aiming to solve less enriched company showcase page by integrating with Instagram account and display most recent instagram photo posts.

IMPROVE SEARCHING ENGINE OPTIMIZATION(SEO) STRATEGY

● Collaborated with XFN (Marking Strategy, PM) teams to implement and experiment various SEO features to drive organic traffic to the site. Worked as a full-stack software engineer.

Tech-stack Java, ReactJS, Spring, MySQL, Google Closure Library, CI/CD

CITIC Securities

QUANTITATIVE TRADING DATA SERIALIZATION SUPPORT

● Implemented a dynamic serialization declaration mechanism and added support for WCF-standard DataContract and DataMember attributes, enhancing compatibility within the existing data exchange framework.

Tech-stack C#, Protocol buffers.

CONTEST

2010   **Bronze Medal** `35th ACM-ICPC Asia Regional Programming Contest`

EDUCATION

2012 – 2015   **Fudan University**, Computer Science, Master Degree
Shanghai Key Laboratory of Intelligent Information Processing | Advisor: Prof. Shuigeng ZHOU
Focused on applying Latent Dirichlet Allocation (LDA) for single-cell RNA-seq (scRNA-seq) analysis.

2008 – 2012   **Northeast Normal University**, Computer Science, Bachelor Degree
National Scholarship, Ministry of Education of China