

Junjie Chen

junjiechen.chris@outlook.com | chrisjic.com | junjiechen-chris | junjie-chen-b6a075240

Summary

LLM/NLP applied scientist with a Computer Science Ph.D. from UTokyo and postdoctoral research experience at the University of Oxford. Built LLM-centric multimodal translation and evaluation systems, including AutoMQM with 0.87 correlation to human MQM judgments. Developed a noise-robust audio-based heartbeat detection algorithm, achieving 0.90 r and 0.87 R^2 on heart rate and RMSSD for high-confidence predictions. Experience in GPU cluster management and supported evaluation and training pipelines for LLMs from 180M to 175B parameters.

Applied & Research Experience

University of Oxford

Oxford, United Kingdom

POSTDOCTORAL RESEARCHER

May, 2026 - Now

- Built audio-based animal heartbeat estimation model under noisy signal conditions.
- Reduced heart rate regression error by 77% and RMSSD error by 90% at 75% prediction availability.
- Achieved 0.90 R^2 on heart rate and 0.87 R^2 on RMSSD for high-confidence predictions.

Amazon

Tokyo, Japan

AMAZON SCIENCE FELLOW

Aug. 2025 - Feb. 2026

- Built LLM-centric multimodal translation system for graphic novels, integrating visual and text-processing modules.
- Designed module fusion method that reduced translation error by approximately 40% versus baseline.
- Designed AutoMQM, an LLM-based translation quality evaluation pipeline with 0.87 correlation to human MQM judgments.

Rakuten

Tokyo, Japan

RESEARCH INTERN

Nov. 2024 - Feb. 2025

- Applied LLM Sparse AutoEncoder (SAE) features to personal information identification for business data workflows.
- Improved PII detection F1 from 72% to 87% through feature analysis and model/pipeline refinement.

National Institute of Informatics

Tokyo, Japan

RESEARCH ASSISTANT

Jun. 2024 - Feb. 2025

- Built evaluation pipeline for Japanese LLMs from 180M to 175B parameters.
- Supported large-scale LLM evaluation and training infrastructure on 800-GPU Slurm cluster.

The University of Tokyo

Tokyo, Japan

TEACHING ASSISTANT

Oct. 2022 - Feb. 2023

- Supervised 3 student teams on their final NLP projects (one published at JNLP-2023 conference).

Research Projects

University of Tokyo

LLM-INFORMED SYNTAX PARSING

2023–2025

- Developed LLM-informed methods for unsupervised constituency and dependency parsing, combining paraphrase-based resampling, reinforcement-learning, and grammatical priors.
- Designed a paraphrase-based resampling method to mitigate spurious textual patterns in unsupervised constituency parsing; improved accuracy by 8 absolute points across four languages.
- Estimated word-word mutual information with grammatical constraints for LLM-based dependency parsing; improved parsing accuracy by over 5%.
- Published related work at ICLR 2025 Spotlight, ACL 2024, and NAACL 2024.

University of Tokyo

SYNTAX-INFORMED SEMANTIC DEPENDENCY PARSING WITH MIXTURE-OF-EXPERTS

2021–2022

- Developed a mixture-of-experts approach to condition semantic dependency labels on automatically discovered syntactic patterns.
- Improved semantic dependency parsing accuracy by ~ 1 absolute F1 point over state-of-the-art methods.
- Published as ACL 2022 Oral work.

Selected Publications

Improving Unsupervised Constituency Parsing via Maximizing Semantic Information

ICLR 2025 (Spotlight)

JUNJIE CHEN, XIANGHENG HE, YUSUKE MIYAO, AND DANUSHKA BOLLEGA

2025

Constituents are Frequent Word Sequences among Sentences with Equivalent Predicate-Argument Structures: Unsupervised Constituency Parsing by Span Matching	ACL 2024 Findings
JUNJIE CHEN, XIANGHENG HE, DANUSHKA BOLLEGALA, AND YUSUKE MIYAO	2024
Language Model Based Unsupervised Dependency Parsing with Conditional Mutual Information and Grammatical Constraints	NAACL 2024
JUNJIE CHEN, XIANGHENG HE, AND YUSUKE MIYAO	2024
Modeling Syntactic-Semantic Dependency Correlations in Semantic Role Labeling Using Mixture Models	ACL 2022 (Oral)
JUNJIE CHEN, XIANGHENG HE, AND YUSUKE MIYAO	2022
Syntactic-Semantic Dependency Correlation in Semantic Role Labeling: A Shift in Semantic Label Distributions	Journal of Natural Language Processing
JUNJIE CHEN	2022
ProsodyFM: Unsupervised Phrasing and Intonation Control for Intelligible Speech Synthesis	AAAI 2025 (Oral)
XIANGHENG HE, JUNJIE CHEN, ZIXING ZHANG, BJÖRN W SCHULLER	2025
Task Selection and Assignment for Multi-Modal Multi-Task Dialogue Act Classification with Non-Stationary Multi-Armed Bandits	ICASSP 2024
XIANGHENG HE, JUNJIE CHEN, AND BJÖRN W SCHULLER	2024
An Improved StarGAN for Emotional Voice Conversion: Enhancing Voice Quality and Data Augmentation	Interspeech 2021
XIANGHENG HE, JUNJIE CHEN, GEORGIOS RIZOS, AND BJÖRN W SCHULLER	2021

Skills

ML/LLM	LLM Application, Multimodal Translation, Automated Evaluation, Unsupervised Learning, Syntax/Semantics Parsing, Audio Biosignals
Programming	Python, TypeScript, Bash
Frameworks	PyTorch, Lightning, vLLM, Hydra, TensorFlow, Hugging Face Accelerate
Infrastructure	AWS, Slurm, Megatron, Docker, Cloudflare
Languages	Chinese (Native), Japanese (N1, Business), English (C1, Business Level)

Education

The University of Tokyo, Advised by Yusuke Miyao	Tokyo, Japan
PH.D. IN COMPUTER SCIENCE	Oct. 2021 - Mar. 2025
• JSPS doctoral fellow; research published at ICLR, ACL, and NAACL.	
The University of Tokyo, Advised by Yusuke Miyao	Tokyo, Japan
M.S. IN COMPUTER SCIENCE	Oct. 2019 - Sep. 2021
University of Liverpool, Advised by Danushka Bollegala	Liverpool, UK
VISITING STUDENT	Oct. 2023 - Mar. 2024
Northeastern University	Shenyang, China
B.S. IN SOFTWARE ENGINEERING	Sep. 2015 - Jun. 2019