Yerin Choi

eriny77c@gmail.com | +82-10-7288-2691 | Yerin Choi | lakahaga

Seoul, South Korea

Research Interests

- Lightweight tuning and adaptation strategies for bias mitigation
- Bias dynamics during model compression and distillation, and their mitigation

Objective

A researcher focused on responsible multimodal AI. During my M.S., I developed dysarthric speech assessment and expressive TTS models, which underscored the importance of equitable and reliable system behavior in sensitive real-world settings. My current work extends this perspective to large multimodal generative models, examining representational fairness in large generative models. My research interests center on two key areas: 1) developing lightweight tuning and adaptation strategies that enable efficient bias mitigation in large language and multimodal models; 2) understanding how biases emerge and amplify during model compression and distillation processes, and designing targeted mitigation approaches for these efficiency-driven transformations.

Education

Sogang University Seoul, South Korea

- **Master of Science in Artificial Intelligence** *Mar 2022 - Feb 2024*
  - CGPA: 3.89/4.30
  - Thesis: **Diffusion-based Dysarthric Speech Augmentation for Enhanced Automatic Severity Classification in Dysarthric Speech**
  - Supervisor: Prof. Myoung-Wan Koo, Intelligent Spoken Dialogue Interface System (ISDS) Lab
- **Bachelor of Science in Computer Science and Engineering**
  Bachelor of Economics *Mar 2018 - Feb 2022*
  - CGPA: 3.25/4.30

Employment

- **Brain Science Institute, Korea Institute of Science and Technology (KIST) [🌐]** *Dec 2024 – Ongoing*
  Research Intern, Laboratory of Computational Neurophysics (LCNP) Seoul, South Korea
  - Contributing to a research project on bias evaluation and mitigation in large multimodal generative models (see Project section).
  - Researching speech-language models to support communication for individuals with severe disabilities, developmental disorders, and neurodegenerative diseases
  - Supervised by Dr. Kyungreem Han.
- **Naver Cloud [🌐]** *Nov 2023 – May 2024*
  Research Intern, Voice AI Team Seongnam, South Korea
  - Developed a non-parallel style transfer Text-to-Speech (TTS) model for expressive speech synthesis.
  - Implemented dual-style embeddings with adversarial learning to improve style similarity and reduce content leakage in TTS outputs.

Publications

[1] Choi, Y. et al. (2025). *Knowledge Graphical Representation and Evaluation of Social Perception of Bias in Text-to-Image Models*. Workshop on LLMs and KRR for Trustworthy AI @ KR 2025. [First Author, Oral].

[2] Choi, Y., Lee, J., & Koo, M.-W. (2024). *Speech Recognition-based Feature Extraction for Enhanced Automatic Severity Classification in Dysarthric Speech*. IEEE Spoken Language Technology Workshop (SLT). [First Co-Author].

[3] Choi, Y. et al. (2024). *Inappropriate Pause Detection in Dysarthric Speech Using Large-Scale Speech Recognition*. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). [First Co-Author].

[4] Choi, Y., & Koo, M.-W. (2023). *DC CoMix TTS: An End-to-End Expressive TTS with Discrete Code Collaborated with Mixer*. Interspeech 2023. [First Author].

Selected Projects

- **Development of self-evolving AI bias detection-correction-explain platform based on international multidisciplinary governance** *Apr 2025 - Ongoing*
  Self-Evolving Bias Evaluation in large multi-modal models KIST
  - Developing knowledge-grounded evaluation frameworks for bias assessment and mitigation in text-to-image and vision-language models.
  - Designing human preference datasets to validate evaluation methods and ensure alignment with diverse perspectives on fairness.
- **Development of AI Technology that Provides Dialog-Based Multi-Modal Explainability** *Mar 2022 - Dec 2023*
  Dialogue-based XAI in Medicine Sogang Univ.
  - Developed a diagnosis and explanation system for dysarthric speech to support clinical therapy applications
  - Implemented automatic evaluation using large-scale ASR models for inappropriate pause detection and severity classification (ICASSP 2024)
  - Designed interpretable feature extraction pipeline to enable user-friendly explainability for clinical practitioners (SLT 2024)
- **Development of Robot and Contents Supporting Children's Reading Activities Based on AI** *Mar 2022 - Dec 2023*
  TTS models for children Sogang Univ.
  - Developed an expressive TTS system for children's reading activity support robot
  - Implemented discrete code-based architecture with mixer models for natural prosody generation (Interspeech 2023)

Conferences & Presentations

- **International Multidisciplinary AI Safety Governance Conference: Development of a Self-Evolving AI Bias Detection, Correction, and Explanation Platform** *Nov 2025*
  Presented "Bias Amplification in Distillation: Risks and Challenges for LLMs and LVLMs" (Talk) Melbourne, Australia
- **KR 2025 Workshop on LLMs and KRR for Trustworthy AI** *Nov 2025*
  Presented "Knowledge Graphical Representation and Evaluation of Social Perception of Bias in Text-to-Image Models" (Oral) Melbourne, Australia
- **Global AI Frontiers Symposium 2025** *Oct 2025*
  Presented "A self-evolving AI bias detection-correction-explanation platform toward Humanistic AI: international governance and human-machine interaction" (Poster) Seoul, South Korea
- **ICASSP 2024** *Apr 2024*
  Presented "Inappropriate Pause Detection in Dysarthric Speech Using Large-Scale Speech Recognition" (Poster) Seoul, South Korea
- **Interspeech 2023** *Aug 2023*
  Presented "DC CoMix TTS: An End-to-End Expressive TTS with Discrete Code Collaborated with Mixer" (Poster) Dublin, Ireland
- **Interspeech 2022** *Sep 2022*
  Student Volunteer Incheon, South Korea

Awards

- **Best Paper Award in Korea Computer Congress 2023** *Jun 2023*
  Korean Institute of Information Scientists and Engineers (KIISE)
  - Automatic Human Transcription Error Detection in ASR Corpora
- **1st Place in the Data-centric AI Hackathon** *Nov 2022*
  National Information Society Agency (NIA) of Korea
  - Automated quality assessment and improvement of Korean speech datasets
- **1st Place in SKT AI Fellowship** *Jun 2022 – Oct 2022*
  SK Telecom, South Korea
  - Personalized lecture recommendation chatbot, deployed in SKT's internal education platform

Scholarships

- **Graduate School Dean's Special Scholarship** *Jul 2022*
  Sogang University
  - Awarded for outstanding research performance
- **Smilegate AI Scholarship** *2022 – 2023*
  Smilegate AI, South Korea
  - Full tuition and research stipend awarded for the M.S. program

Teaching Experience

- Pattern Recognition (CSE5416), Advanced Undergrad/Grad, TA *Spring, 2023*
- Advanced Applied C Programming (STS2008), Undergrad, TA *Spring-Fall, 2022*
- Computational Thinking (COR1009), Undergrad, TA *Fall, 2021*

Skills

- **Technical Skills:** Python, PyTorch, HuggingFace Transformers, Git, AWS
- **Languages:** Korean (Native), English (Fluent - TOEFL 100)