Anna Seo Gyeong Choi

sc2359@cornell.edu · Linkedin · Website

Research interests

AI Ethics, Algorithmic Fairness, Speech/Signal Processing, Clinical Speech Biomarker

Education

2021.08 – 2026.05 (Expected)	**Cornell University** Ph.D. in Information Science M.Sc. in Information Sceience (2021 – 2024) Committee: Allison Koenecke (Chair), Matthew Wilkens, Marten van Schijndel
2019.03 – 2021.08	**Seoul National University** M.A. in English Language and Literature Thesis: *Analysis of Lexical Stress of Korean-accented English and Spanish-accented English for Accent Classification* Committee: Minhwa Chung (Chair), Sunhee Kim, Yongwon Lee
2015.03 – 2019.02	**Seoul National University** B.A. in Hispanic Language and Literature, *cum laude* B.A. in English Language and Literature, *cum laude*

Experience

2025.09 – present	**Kotoba Technologies** *External Collaborator* Full stack Korean speech-to-text model development and iterative improvements of synthetic data generation
2025.05 – present	**Rev AI** *Speech Science Intern* Working with Verbatim/Non-verbatim/Legal mode specialized speech-to-text models for speech dysfluencies
2024.09 – 2025.08	**University of Pennsylvania** *Visiting Scholar* (Remote) Working with Linguistic Data Consortium & UPenn Hospital on clinical projects, including autism-, schizophrenia-, and neurodegeneration-related projects
2024.06 – 2024.08	**NAVER Cloud** *Summer Intern* Worked with Speech-to-Text teams (NSpeech, NEST) on Kaldi & PyTorch based modeling of foreign-accented speech

Working papers

In Preparation — **ClinSpeech: A Holistic Benchmark for evaluating ASR Fairness in Clinical Conversations**
Anna Seo Gyeong Choi, Oresis Papakyriakopoulos, Allison Koenecke, & Alessandro Fabris

Under Review — **PEEC: The Protected Entities Ethics Checklist for Collecting Speech Data from Vulnerable Clinical Populations**
Anna Seo Gyeong Choi, Sunghye Cho, & Iris Nowenstein

Under Review — **Pitfalls of Auditing Practices in Automatic Speech Recognition Technologies: A Case Study of People with Aphasia**
Katelyn X. Mei*, **Anna Seo Gyeong Choi***, Hilke Schellmann, Mona Sloane, & Allison Koenecke (* equal contribution)

In Submission — **SpeechSpectrum: A Framework for User-Controlled Speech-to-Text Representation Along the Linguistic Fidelity Spectrum**
Anna Seo Gyeong Choi*, Maria Teleki*, Miguel del Rio Fernandez, James Caverlee & Allison Koenecke (* equal contribution)

Publications

All conference proceedings are peer-reviewed with full paper

2025 (To appear) — **Analyzing Dialectical Biases in LLMs for Knowledge and Reasoning Benchmarks**
Eileen Pan, **Anna Seo Gyeong Choi**, Maartje Ter Hoeve, Skyler Seto, & Allison Koenecke
Proceedings of Empirical Methods in Natural Language Processing (EMNLP) Findings

2025 (To appear) — **Fairness of Automatic Speech Recognition: Looking Through a Philosophical Lens**
Anna Seo Gyeong Choi & Hoon Choi
Proceedings of AAAI/ACM Conference on AI, Ethics, and Society (AIES)

2025 — **Comparative Evaluation of Acoustic Feature Extraction Tools for Clinical Speech Analysis**
Anna Seo Gyeong Choi, Alex Richardson, Ryan Partlan, Sunny X. Tang†,& Sunghye Cho† († equal contribution)
Proceedings of Interspeech

2025 — **Reasoning-Based Approach with Chain-of-Thought for Alzheimer's Detection Using Speech and Large Language Models**
Chanwoo Park, **Anna Seo Gyeong Choi**, Sunghye Cho, & Chanwoo Kim
Proceedings of Interspeech

| 2025 | **Data-Driven Mispronunciation Pattern Discovery for Robust Speech Recognition**
Anna Seo Gyeong Choi*, Jonghyeon Park*, & Myungwoo Oh (* equal contribution)
Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) |
|---|---|
| 2024 | **Quantification of Automatic Speech Recognition System Performance on d/Deaf and Hard of Hearing Speech**
Robin Zhao*, **Anna Seo Gyeong Choi***, Allison Koenecke†, & Anäis Rameau† (*, † equal contribution)
The Laryngoscope |
| 2024 | **Careless Whisper: Speech-to-Text Hallucination Harms**
Allison Koenecke, **Anna Seo Gyeong Choi***, Katelyn X. Mei*, Hilke Schellmann†, & Mona Sloane† (*, † equal contribution)
Proceedings of ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT) |
| 2024 | **KoFREN: Comprehensive Korean Word Frequency Norms Derived from Large Scale Free Speech Corpora**
Jinseo Kim, **Anna Seo Gyeong Choi**, & Sunghye Cho
Proceedings of Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING) |
| 2024 | **Crosslinguistic Acoustic Feature-based Dementia Classification using Advanced Learning Architectures**
Anna Seo Gyeong Choi, Jinseo Kim, Seo-hee Kim., Minseok Baek & Sunghye Cho
Proceedings of Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING) Workshop on Resources and Processing of Linguistic, Para-linguistic and Extra-linguistic Data from People with Various Forms of Cognitive/Psychiatric/Developmental Impairments (RaPID-5) |
| 2023 | **Augmented Datasheets for Speech Datasets and Ethical Decision-Making**
Orestis Papakyriakopoulos*, **Anna Seo Gyeong Choi***, William Thong, Dora Zhao, Jerome Andrews, Rebecca Bourke, Alice Xiang†, & Allison Koenecke† (*, † equal contribution)
Proceedings of ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT) |
| 2020 | **Building Korean Conversational Speech Data in Emergency Medical Domain**
Kim, Sunhee., Jooyeong Lee, **Seo Gyeong Choi**, Seunghun Ji, Jeemin Kang, Jongin Kim, et al.
Phonetics and Speech Sciences, 12 (4) |

Non-Archival Presentations

| Jul 2025 | **Pitfalls of Auditing Practices in Automatic Speech Recognition Technologies: A Case Study of People with Aphasia** (Oral)
11th International Conference on Computational Social Science (IC2S2 '25) |
|---|---|
| Jul 2023 | **Auditing Korean Speech Datasets for Dialectal Fairness in Speech-to-Text Applications** (Poster)
9th International Conference on Computational Social Science (IC2S2 '23) |

Jul 2023	**Auditing Korean Speech Datasets for Dialectal Fairness in Speech-to-Text Applications** (Poster) *9th International Conference on Computational Social Science (IC2S2 '23)*
Jul 2023	**Augmented Datasheets for Speech Datasets and Ethical Decision-Making** (Poster) *9th International Conference on Computational Social Science (IC2S2 '23)*
Apr 2022	**Phoneme-to-phoneme (P2P) lexicon expansion technology for Brazilian Portuguese** (Poster) *Mid-Atlantic Student Colloquium on Speech, Language and Learning (MASC-SLL 2022)*
Nov 2020	**Building Korean Conversational Speech Data in Emergency Medical Domain** (Oral) *Korean Conference on Speech Sciences*
Oct 2019	**Stereotypical dialects in *Zootopia* and its implications** (Oral) *Coalition of English Departments in Asia*

Awards

2025	Cornell Center for Social Science Qualitative and Interpretive Research Institute Small Research Grant (Cornell University)
2023	Microsoft Azure Cloud Computing Grant (Cornell University)
2022	Cornell Center for Social Science Grant (Cornell University)
2021	Sage Fellowship (Cornell University)
2019 - 2020	Graduate Student Instructor Scholarship (Seoul Nat'l Univ.)
2017 - 2019	CORE Undergraduate Scholarship (Seoul Nat'l Univ.)
2017	CORE Exchange Students Scholarship (Seoul Nat'l Univ.)
2015	Scholarship for Academic Excellence (Seoul Nat'l Univ.)

Teaching experience

Cornell Univ.	**LING 1101: Introduction to Linguistics**, **SPAN 1210: Elementary Spanish I**, **INFO 2950: Introduction to Data Science**
Seoul Nat'l Univ.	**Graduate Seminar on Applied Linguistics**, **Graduate Course on English Phonology**, **English Phonetics**, **English and Society**