JOSHUA MEYER

joshua.richard.meyer@gmail.com

jrmeyer.github.io

Josh is a leader working at the intersection of data, machine learning, generative AI, and language. With over a decade of experience, his career has spanned from large corporations to start-ups, from academia to non-profits. With a PhD in automatic speech recognition, Josh deeply understands the technical underpinnings of AI. He has lead teams to successfully plan, execute, and ship products, and he cares deeply about customers and markets.

EXPERIENCE

rabbit, inc. Head of Applied AI 2024 — Present

Josh leads a team of Applied AIML engineers, working with conversational AI models (LLMs, speech-to-text, text-to-speech) as well as large multimodal models (vision and text). Josh coordinates between various C-level stakeholders from product, research, and engineering teams.

Leadership Team · Recruiting · Strategic Partnerships · Contract Negotiations

Coqui Co-founder 2021 — 2024

As one of four founders, Josh has been at every stage of the business. From pitching the first VCs, to collecting data for AI models, to red-lining with enterprise customers. In the early days Josh spent most time training speech-to-text models and working on the data pipeline. As the business grew, Josh took on product and GTM leadership.

Board Member · Leadership Team · Go-to-Market Strategy · Product Strategy · R&D · B2B Sales · Recruiting · Strategic Partnerships · PR and Marketing · Investor Relations · Customer Relations · Contract Negotiations

Mozilla Foundation Machine Learning Fellow 2020 — 2021

Josh lead the AI strategy for a multi-national mutli-stakeholder project to deploy voice technologies in East Africa, aligned with UN Sustainable Development goals. Target applications were a COVID radio monitoring dashboard for the Ugandan Ministry of Health as well as COVID informational voicebot for the Rwandan Information Society Authority. Partners included the German government, Ugandan academia, and the Rwandan startup sector. Josh was involved at every stage of the project, from ideation to pitching stakeholders, from data collection to model training and deployment. Josh developed the first publicly available speech recognition models for the Kinyarwanda and Luganda languages.

Artie, Inc. Lead Scientist 2019 — 2021

Josh directed and developed all Speech technologies used at Artie. The technologies were integrated into a game engine to power conversational AI avatars. Josh developed the pipeline for data collection and quality control, and he scaled the training of contextually dependent language model decoders for custom speech-to-text engines. Josh also experimented with text-to-speech at the time, but the state-of-the-art wasn't high enough for production use. Josh collected data, trained the core models, and worked with the backend team to deploy.

National Science Foundation Graduate Research Fellow 2015 — 2019

Josh was awarded 1 out of 10 Fellowships nationally in his category. The Fellowship supported Josh during four years of dissertation research on automatic speech recognition. In particular, Josh worked on algorithms for more efficient training of speech recognition models for new languages. During his research, Josh spent one year in Paris as a research fellow at CNRS, and multiple data collection trips to Kyrgyzstan as a research fellow at the American University of Central Asia.

Mozilla Corporation Machine Learning Researcher 2018 — 2019

Josh joined the Machine Learning group at Mozilla in order to bring automatic speech recognition to as many languages as possible. He accomplished this by working closely with the data collection team (i.e. Project Common Voice) as well as training end-to-end speech recognition models via multitask and transfer learning with the core ML team (i.e. Project DeepSpeech).

Various Companies Speech Recognition Consultant 2016 — 2018

Josh worked as an automatic speech recognition consultant, for companies ranging from large telcos to small edtech startups. He designed training and data pipelines, mostly revolving around the Kaldi toolkit. Kaldi was the state-of-the-art in speech recognition before DeepSpeech, and DeepSpeech was state-of-the-art until Whisper.

CNRS Chateaubriand Research Fellow 2016 — 2017

Josh researched transfer-learning for acoustic models as well as text augmentation for language decoders, both in the context of automatic speech recognition. Over this year at CNRS, Josh cemented his knowledge of the underpinnings of speech recognition. Josh's time at CNRS was supported by the Embassy of France to the United States under a Chateaubriand STEM Fellowship.

American University of Central Asia Research Fellow 2013 — 2014

During this research fellowship and multiple visits afterward, Josh conducted acoustic-phonetic research on the Kyrgyz language. This included creation of a spoken dataset of Kyrgyz regional varieties, as well as behavioral studies on human perception of both Russian and Kyrgyz by native speakers. In addition, Josh organized a working group on computational linguistics for students, professors, and researchers.

EDUCATION

University of Arizona 2019
Ph.D. in Computational Linguistics *Tucson, AZ*
Thesis: *Multi-Task and Transfer Learning in Low-Resource Speech Recognition*

University of Arizona 2015
M.A. in Linguistics *Tucson, AZ*

Seton Hall University 2012
B.A. in Liberal Studies *South Orange, NJ*

GRANTS & AWARDS

Graduate Research Fellowship	National Science Foundation	$138,000	2015 — 2019
NSF INTERN Grant	National Science Foundation	$50,000	2018
Chateaubriand STEM Fellowship	Embassy of France to the USA	€5.000	2016
GROW Travel Award	National Science Foundation	$5,000	2016
Visiting Research Fellowship	American University of Central Asia	$5,000	2013 — 2014
Regents Scholarship	Seton Hall University	$111,140	2008 — 2012
Liberal Studies Honors Citation	Seton Hall University		2012

PUBLICATIONS

Google Scholar Profile

XTTS: a Massively Multilingual Zero-Shot Text-to-Speech Model. Casanova, E; Davis, K; Gölge, E; Göknar, G; Gulea, I; Hart, L; Aljafari, A; **Meyer, J**; Morais, R; Olayemi, S; Weber, J. *https://arxiv.org/abs/2406.04904*	INTERSPEECH 2024
AfroDigits: A Community-Driven Spoken Digit Dataset for African Languages. Emezue, C; Gandhi, S; Tunstall, L; Abid, A; **Meyer, J**; Lhoest, Q; Allen, P; Von Platen, P; Kiela, D; Jernite, Y; Chaumond, J; Noyan, M.; Sanseviero, O. *https://arxiv.org/abs/2303.12582*	AfricaNLP 2023 (co-located ICLR 2023)
BibleTTS: A Large, High-fidelity, Multilingual, and Uniquely African Speech Corpus. **Meyer, J**; Adelani, D; Casanova, E; Öktem, A; Whitenack, D; Weber, J; Kabongo, S; Salesky, E; Orife, I; Leong, C; Ogayo, P; Emezue, C; Mukiibi, J; Osei, S; Agbolo, A; Akinode, V; Opoku, B; Olanrewaju, S; Alabi, J; Muhammad, S. *https://arxiv.org/abs/2207.03546*	INTERSPEECH 2022
The Makerere Radio Speech Corpus: A Luganda Radio Corpus for Automatic Speech Recognition. Mukiibi, J; Katumba, A; Nakatumba-Nabende, J; Hussein, A; **Meyer, J**. *https://arxiv.org/abs/2206.09790*	LREC 2022
Multilingual Spoken Words Corpus. Mazumder, M; Chitlangia, S; Banbury, C; Kang, Y; Ciro, JM; Achorn, K; Galvez, D; Sabini, M; Mattson, P; Kanter, D; Diamos, G; Warden, P; **Meyer, J**; Reddi, V. *https://openreview.net/forum?id=c20jiJ5K2H*	NeurIPS 2021 (Datasets and Benchmarks Track)
Few-Shot Keyword Spotting in Any Language. Mazumder, M; Banbury, C; **Meyer, J**; Warden, P; Reddi, V. *https://arxiv.org/abs/2104.01454*	INTERSPEECH 2021
What shall we do with an hour of data? Speech recognition for the un- and under-served languages of Common Voice. Tyers, F; **Meyer, J**. *https://arxiv.org/abs/2105.04674*	2021 arXiv
Common Voice: A Massively-Multilingual Speech Corpus. Ardila, R; Branson, M; Davis, K; Henretty, M; Kohler, M; **Meyer, J**; Morais, R; Saunders, L; Tyers, F; Weber, G. *https://arxiv.org/abs/1912.06670*	LREC 2020
Artie Bias Corpus: An Open Dataset for Detecting Demographic Bias in Speech Applications. **Meyer, J**; Rauchenstein, L; Eisenberg, J; Howell, N. *https://aclanthology.org/2020.lrec-1.796/*	LREC 2020
Multi-Task and Transfer Learning in Low-Resource Speech Recognition. **Meyer, J**. *http://jrmeyer.github.io/misc/MEYER_dissertation_2019.pdf*	2019 PhD Dissertation. University of Arizona.

Unsupervised Task Discovery for Multi-Task Acoustic Modeling. **Meyer, J**. *https://github.com/JRMeyer/mlslp/blob/master/poster/mlslp-poster.pdf*	Machine Learning in Speech and Language Processing Workshop (co-located INTERSPEECH 2018)
The Field is not the Lab, and the Lab is not the Field: Experimental linguistics and endangered language communities. **Meyer, J**; Kloehn, N; Carnie, A. *https://www.degruyter.com/document/doi/10.1515/9783110527018/html*	2018 De Gruyter Mouton: Insights from Practices in Community-Based Research
Navigating the safety implications of doing research and being researched in Kyrgyzstan: cooperation, networks and framing. Bekmurzaev, N; Lottholz, P; **Meyer, J**. *https://www.tandfonline.com/doi/abs/10.1080/02634937.2017.1419165*	2018 Central Asian Survey.
Development of a Kyrgyz Speech Synthesizer: A Demonstration of the Ossian Frontend and Merlin Neural Network Speech Synthesis Toolkit. **Meyer, J**. *http://www.turklang.org/wp-content/uploads/2017/05/%D0%A2.1.pdf*	2017 The 5th International Conference on Computer Processing of Turkic Languages
'Friend' or 'Foreign Agent'? On the Limits of Field Research in Post-Soviet Kyrgyzstan. Lottholz, P; **Meyer, J**. *https://www.academia.edu/33980022*	2016 Exeter Central Asian Studies Network
Conducting Linguistic Fieldwork in Kyrgyzstan. **Meyer, J**. *https://journals.uair.arizona.edu/index.php/arizanthro/article/viewFile/21375/20949*	2016 Arizona Anthropologist, Vol. 27

RESEARCH EXPERIENCE (OTHER)

Arizona Phonological Imaging Lab. Josh developed software to visualize and statistically analyze ultrasonic images of tongue contours for articulatory-phonetics research.	2015
Arizona Scottish Gaelic Project Josh conducted fieldwork with speakers of Scottish Gaelic, and aided in the compilation of an online grammar for the language.	2012 — 2013

Language Acquisition Research Center at Hunter CUNY. Josh designed a text corpus-based research program, transcribed spontaneous child speech, and designed the laboratory's website. 2011 — 2012

KIT/NYU MEG Lab at NYU. Josh aided in the analysis and processing of Magnetoencephalography (MEG) brain imaging data from cognitive science research. 2012

Language Acquisition and Neurolinguistics Lab at Rutgers University. Josh aided in experimental design of Neurolinguistic research studies. 2012

TALKS & PRESENTATIONS

Language Technologies in Central Asia: A Survey. Joint ESCAS-CESS Conference at the American University of Central Asia. Bishkek. 2017

Predicting Language Dominance in Kyrgyz-Russian Bilinguals. w/ Quam, C. Arizona Linguistics Circle 9. Tucson. 2015

Phonological processing in Kyrgyz-Russian bilinguals. w/ Quam, C; Bever, T. The Miniconference on Metrical Structure: Acquisition and processing. Utrecht. 2014

The Kyrgyzstan corpus project: Building a language resource unique to Kyrgyzstan and available to all. CASI Public Seminar at the American University of Central Asia. Bishkek. 2014

Now you hear it, now you don't: Phonological processing in Kyrgyz-Russian bilinguals. CASI Public Seminar at the American University of Central Asia. Bishkek. 2013

Psycholinguistics: Thinking about language differently. Lecture conducted at The Platttform at Werkstatttraum. Berlin. 2013

WRITTEN TUTORIALS

Kaldi Troubleshooting Head-to-Toe. 🔗 Link. 2018

How to Train *practically* any Model from *practically* any Data with TensorFlow. 🔗 Link. 2018

Getting started with the Merlin Speech Synthesis Toolkit. 🔗 Link. 2017

How to Train a Deep Neural Net Acoustic Model with Kaldi. 🔗 Link. 2016

How to add a new language to the eSpeak NG Speech Synthesizer. 🔗 Link. 2016

The Flow of TensorFlow: An Email Classification Tutorial. 🔗 Link. 2016

An Introduction to CMU-Sphinx Speech Recognition Toolkit: First Steps. 🔗 Link. 2016

EVENT ORGANIZATION

Kyrgyz Voice Technology Hackathon. *American University of Central Asia*. Attended by undergraduate students as well as professional developers. 🔗 Link. 2019

DeepSpeech & Common Voice Tutorial. *University of Tartu*. Delivered to attendees of the Fifth International Workshop on Computational Linguistics for Uralic Languages. 🔗 Link. 2019

Speech Synthesis Workshop: Hands-on with Merlin & Ossian. *Higher School of Economics*. Delivered to Computational Linguistics Faculty and Graduate Students. 🔗 Link. 2018

Speech Recognition Workshop: Hands-on with Kaldi & DeepSpeech. *Higher School of Economics*. Delivered to Computational Linguistics Faculty and Graduate Students. 🔗 Link. 2017

OPEN SOURCE WORK

Coqui STT	*Trained Coqui's best STT model on 47k hours of data*. 🔗 Link.	2021
Mozilla DeepSpeech	*Added Transfer Learning*. 🔗 Link.	2020
Multi-Task Kaldi	*Multi-Task Acoustic Modeling*. 🔗 Link.	2018
NVDA	*Added Kyrgyz language to screen reader*. 🔗 Link.	2017
eSpeak NG	*Added Kyrgyz language to speech synthesizer*. 🔗 Link.	2016
Autotrace	*Created visualization tools for ultrasonic imaging*. 🔗 Link.	2015

TECHNICAL STRENGTHS

Computer Languages	Python, Bash, MATLAB, Perl, R, C++ (*some knowledge*)
Other	TensorFlow, PyTorch (*some knowledge*), Kaldi, scikit-learn, pandas, nltk, matplotlib, kenlm, Linux, AWS, git, GitHub