Spoken Conversational Agents with Large Language Models

Huck Yang, Andreas Stolcke, and Larry P. Heck

Website: https://huckiyang.github.io/emnlp-25-tutorial

Spoken conversational agents are converging toward voice-native LLMs. This tutorial distills the path from cascaded ASR/NLU to end-to-end, retrieval- and vision-grounded systems. We frame adaptation of text LLMs to audio, cross-modal alignment, and joint speech–text training; review datasets, metrics, and robustness across accents; and compare design choices (cascaded vs. E2E, post-ASR correction, streaming). We link industrial assistants to current open-domain and task-oriented agents, highlight reproducible baselines, and outline open problems in privacy, safety, and evaluation. Attendees leave with practical recipes and a clear systems-level roadmap.

Huck Yang, Sr. Research Scientist, NVIDIA Research
email: huckiyang@nvidia.com
website: https://huckiyang.github.io/
Huck obtained his Ph.D. and M.Sc. from Georgia Institute of Technology, Atlanta, GA, supported by Wallace H. Coulter Fellowship and B.Sc, from National Taiwan University. Prior to joining NVIDIA, he was a scientist at Amazon and a research intern at Google and Hitachi. His primary research lies in the area of speech-language modeling, robust speech recognition, and multi-modal post-training alignments. He served as area chairs and committee members in IEEE ICASSP 2022 to 2025, EMNLP 2024, SLT 2024, and NAACL 2025. He has served in the IEEE SPS technical committee at Applied Signals Processing Systems (ASPS) and Data Collection Committee (DCC) since 2022. He received the best industry paper honoronable mentioned ACL 25 and best student paper nominee in Interspeech 23.

Andreas Stolcke, Distinguished AI Scientist/VP, Uniphore
email: stolcke@icsi.berkeley.edu
website: https://www.linkedin.com/in/andreas-stolcke
Andreas obtained his PhD from UC Berkeley, followed by senior researcher/scientist positions at SRI International, Microsoft, and Amazon, as well as an external

1

fellow of the Berkeley International Computer Science Institute. He is currently a Distinguished AI Scientist/VP at Uniphore, where he is leading research on conversational AI. His interests include computational linguistics, language modeling, speech recognition, speaker recognition and diarization, and paralinguistics, with over 300 papers and patents in these areas. His open-source SRI Language Modeling Toolkit was widely used in academia. Andreas is a Fellow of the IEEE, the International Speech Communication Association, and the Asia-Pacific Artificial Intelligence Association, as well as an IEEE Distinguished Industry Speaker.

Larry Heck, Rhesa Screven Farmer Jr., Advanced Computing Concepts Chair, Georgia Institute of Technology,
email: `larryheck@gatech.edu`
website: `https://larryheck.github.io/`
Larry Heck is a Professor with joint appointments in ECE and Interactive Computing at the Georgia Institute of Technology, where he holds the Rhesa S. Farmer Advanced Computing Concepts Chair and is a Georgia Research Alliance Eminent Scholar. He is an IEEE Fellow and National Academy of Inventors Fellow, author of numerous papers, and holder of more than 50 U.S. patents. He previously served as CEO of Viv Labs and SVP at Samsung (2017–2021), leading Bixby in North America. From 2014–2017, he was Principal Scientist at Google, leading dialogue research. At Microsoft (2009–2014), he was Chief Scientist of Speech products, Distinguished Engineer in MSR, and co-founder of Cortana. Earlier, he was VP at Yahoo, VP of R&D at Nuance, and began his career at SRI, where his team pioneered deep learning in speech processing. He earned his PhD and MS in Electrical Engineering from Georgia Tech, and his BSEE from Texas Tech University.

2