

Dong Zhang

☎ (+86)13598645410 | ✉ dongzhang22@m.fudan.edu.cn | 🏠 Onutation.github.io | 🌐 Onutation

Research Interests

Multi-modal Foundation Models

Unified multi-modal LLM with perceptive and generative abilities

Multi-modal LLM Agent

Agents built on multi-modal LLM

Generative AI for speech

Spoken language models, Speech tokenizer

Education

Fudan University

M.S. IN COMPUTER SCIENCE

- **Advisor:** Prof. Yaqian Zhou & Prof. Xipeng Qiu, FudanNLP group

Shanghai, China
2022.9-2025.6 (expected)

Fudan University

B.S. IN ELECTRONIC ENGINEERING

Shanghai, China
2018.9-2022.6

Publications & Preprints

SpeechGPT: Empowering Large Language Models with Intrinsic Cross-Modal Conversational Abilities

- **Dong Zhang**, Shimin Li, Xin Zhang, Jun Zhan, Pengyu Wang, Yaqian Zhou, Xipeng Qiu
- This work is the first and only open-source speech-to-speech dialogue LLM, one-year earlier than GPT-4o.
- A Trending Project at Github with **1000+ stars** and is promoted by different media and forums, such as *Heart of Machine*, *Speech Home*, *Twitter* and *Youtube*
- EMNLP 2023 Findings

SpeechTokenizer: Unified Speech Tokenizer for Speech Large Language Models

- **Dong Zhang**^{*}, Xin Zhang^{*} (Equal contribution, order is random), Shimin Li, Yaqian Zhou, Xipeng Qiu
- Github **340+ stars**
- ICLR 2024

SpeechGPT-Gen: Scaling Chain-of-Information Speech Generation

- **Dong Zhang**^{*}, Xin Zhang^{*} (Equal contribution), Jun Zhan, Shimin Li, Yaqian Zhou, Xipeng Qiu
- Preprint

SpeechAgents: Human-Communication Simulation with Multi-Modal Multi-Agent Systems

- **Dong Zhang**, Zhaowei Li, Pengyu Wang, Xin Zhang, Yaqian Zhou, Xipeng Qiu
- Preprint

SpeechAlign: Aligning Speech Generation to Human Preferences

- **Dong Zhang**^{*}, Zhaowei Li^{*}, Shimin Li, Xin Zhang, Pengyu Wang, Yaqian Zhou, Xipeng Qiu
- Preprint

DUB: Discrete Unit Back-translation for Speech Translation

- **Dong Zhang**, Rong Ye, Tom Ko, Mingxuan Wang, Yaqian Zhou
- ACL 2023 Findings

AnyGPT: Unified Multimodal LLM with Discrete Sequence Modeling

- Jun Zhan^{*}, Junqi Dai^{*}, Jiasheng Ye^{*}, Yunhua Zhou, **Dong Zhang**, Zhigeng Liu, Xin Zhang, Ruibin Yuan, Ge Zhang, Linyang Li, Hang Yan, Jie Fu, Tao Gui, Tianxiang Sun, Yugang Jiang, Xipeng Qiu
- This work gets **600+ stars** at Github and is promoted by different media and forums, such as *Heart of Machine*, *Twitter*
- ACL 2024

GroundingGPT:Language Enhanced Multi-modal Grounding Model

- Zhaowei Li, Qi Xu, **Dong Zhang**, Hang Song, Yiqing Cai, Qi Qi, Ran Zhou, Junting Pan, Zefeng Li, Van Tu Vu, Zhida Huang, Tao Wang
- ACL 2024

InferAligner: Inference-Time Alignment for Harmlessness through Cross-Model Guidance

- Pengyu Wang, **Dong Zhang**, Yanglin Li, Xinghao Wang, Chenkun Tan, Ke Ren, Botian Jiang, Xipeng Qiu
- Preprint

SeqXGPT: Sentence-Level AI-Generated Text Detection

- Pengyu Wang, Linyang Li, Ke Ren, Botian Jiang, **Dong Zhang**, Xipeng Qiu
- EMNLP 2023

GenTranslate: Large Language Models are Generative Multilingual Speech and Machine Translators

- Yuchen Hu, Chen Chen, Chao-Han Huck Yang, Ruizhe Li, **Dong Zhang**, Zhehuai Chen, Eng Siong Chng
- ACL 2024

Selected Talks

SpeechGPT Series

- Invited talks at SJTU X-Lance, Bytedance, Agora.ai

SpeechTokenizer: Unified Speech Tokenizer for Speech Large Language Models

- Invited talks at MIT CASIL SLS group

SpeechGPT-Gen: Scaling Chain-of-Information Speech Generation

- Invited talks at AGI Leap Summit 2024 hosted by SuperAGI

Experiences

ByteDance AI-Lab (Shanghai)

RESEARCH INTERN

- Mentor: Rong Ye
- Topic: Speech translation, speech pre-training

Shanghai, China
April 2022 - June 2023

Awards

SCHOLARSHIP FOR OUTSTANDING STUDENT OF FUDAN UNIVERSITY

2023, 2022, 2021, 2020, 2019

Academic Services

REVIEWER: EMNLP (2023), ACL (2024), NEURIPS (2024)