

# Ziyi Yang

✉ yangziyi99@gmail.com   🔗 <https://ziyi-yang.github.io/>   **in** ziyi-yang

## Education

---

<b>Ph.D.</b> Stanford University	2016 – 2021
<b>M.S.</b> Stanford University	2016 – 2018
<b>B.S.</b> Nanjing University	2012 – 2016

## Experience

---

<b>Databricks MosaicML</b> , Staff Research Scientist	Bellevue, WA Jan 2025 – Now
<ul style="list-style-type: none"><li>• <b>Vision LLM.</b> Tech lead for developing document-parsing VLM with hands-on ownership across post-train pipeline: SFT, RL, data curation, evaluation, and serving.</li></ul>	
<b>Microsoft GenAI</b> , Principal Researcher	Redmond, WA Sep 2023 – Dec 2024
<ul style="list-style-type: none"><li>• <b>Lead of Post-Training.</b> Led the end-to-end post-training of Microsoft's Phi model family (<a href="#">Phi-3</a> and <a href="#">Phi-4</a>), driving SFT, RL (RLHF/DPO/PPO), synthetic data generation, and evaluation. Hands-on in model training, data curation, and pipeline optimization to deliver production-ready models.</li></ul>	
<b>Microsoft Research</b> , Senior Researcher	Redmond, WA Jul 2021 – Sep 2023
<ul style="list-style-type: none"><li>• <b>NLP &amp; Multimodal Research.</b> Built language- vision-speech multi-modal model i-Code. Publications on CVPR (spotlight, twice), NeurIPS, EMNLP, AAAI, NAACL, etc.</li></ul>	
<b>Google DeepMind</b> , Research Intern	Mountain View, CA June-Dec 2020
<ul style="list-style-type: none"><li>• <b>Multilingual Sentence Embeddings.</b> Trained a novel transformer-based sentence encoder for over 110 languages. Two first-author papers accepted at <b>EMNLP 21'</b>.</li></ul>	
<b>Microsoft Research</b> , Research Intern	Bellevue, WA June-Sept 2019
<ul style="list-style-type: none"><li>• <b>Unsupervised Text Summarization.</b> Designed a transformer-based unsupervised summarization system with state-of-the-art performance. One first-author paper accepted to <b>EMNLP 20'</b>.</li></ul>	
<b>Microsoft Research</b> , Research Intern	Redmond, WA June-Sept 2018
<ul style="list-style-type: none"><li>• <b>Non-parameterized Sentence Embeddings.</b> Achieved state-of-the-art performances of non-parameterized methods. One first-author paper accepted to <b>EMNLP 19' as oral presentation.</b></li></ul>	

## Selected Publications

---

- **Phi-4-mini technical report: Compact yet powerful multimodal language models via mixture-of-loras**

Lead for post-training.

- **Phi-3 technical report: A highly capable language model locally on your phone.**

Lead for post-training.

- **Self-exploring language models: Active preference elicitation for online alignment**

Shenao Zhang, Donghan Yu, Hiteshi Sharma, Han Zhong, Zhihan Liu, [Ziyi Yang](#), Shuohang Wang, Hany Hassan Awadalla, Zhaoran Wang. *TMLR 2025*

- **Phi-3 safety post-training: Aligning language models with a break-fix cycle**

Lead for post-training.

- **i-Code Studio: A Configurable and Composable Framework for Integrative AI.**

Yuwei Fang, Mahmoud Khademi, Chenguang Zhu, [Ziyi Yang](#), et al. *EMNLP Demo 2024*

- **i-Code V2: An Autoregressive Generation Framework over Vision, Language, and Speech Data.**

[Ziyi Yang](#), Mahmoud Khademi, Yichong Xu, Reid Pryzant, Yuwei Fang, Chenguang Zhu, et al. *NAACL Findings 2024*

- **Soft Convex Quantization: Revisiting Vector Quantization with Convex Optimization**

Tanmay Gautam, Reid Pryzant, [Ziyi Yang](#), Chenguang Zhu, Somayeh Sojoudi. *L4DC 2024*

- **CoDi-2: In-Context Interleaved and Interactive Any-to-Any Generation.**

Zineng Tang, [Ziyi Yang](#)\* (intern mentor), Mahmoud Khademi, Yang Liu, Chenguang Zhu, Mohit Bansal\*. *CVPR 2024, Spotlight*

- **Any-to-any generation via composable diffusion.**

Zineng Tang, [Ziyi Yang](#)\* (intern mentor), Chenguang Zhu, Michael Zeng, Mohit Bansal\*. *NeurIPS 2023*

- **Unifying Vision, Text, and Layout for Universal Document Processing.**

Zineng Tang, [Ziyi Yang](#)\* (intern mentor), Guoxin Wang, Yuwei Fang, Yang Liu, Chenguang Zhu, Michael Zeng, Cha Zhang, Mohit Bansal\*. *CVPR 2023, Spotlight*

- **Global Selection of Contrastive Batches via Optimization on Sample Permutations.**

Vin Sachidananda, [Ziyi Yang](#), Chenguang Zhu. *ICML 2023*

- **i-Code: An integrative and composable multimodal learning framework.**

[Ziyi Yang](#), Yuwei Fang, Chenguang Zhu, Reid Pryzant, Dongdong Chen, Yu Shi, Yichong Xu, Yao Qian, et al. *AAAI 2023*

- **Language models with image descriptors are strong few-shot video-language learners.**

Zhenhailong Wang, Manling Li, Ruochen Xu, Luowei Zhou, Jie Lei, Xudong Lin, Shuohang Wang, [Ziyi Yang](#), Chenguang Zhu, Derek Hoiem, Shih-Fu Chang, Mohit Bansal, Heng Ji. *NeurIPS 2022*

- **Empowering Language Models with Knowledge Graph Reasoning for Open-Domain Question Answering.**

Ziniu Hu, Yichong Xu, Wenhao Yu, Shuohang Wang, [Ziyi Yang](#), Chenguang Zhu, Kai-Wei Chang, Yizhou Sun. *EMNLP 2022*

- **Automatic Rule Induction for Efficient Semi-Supervised Learning.**

Reid Pryzant, Ziyi Yang, Yichong Xu, Chenguang Zhu, Michael Zeng. *EMNLP Findings 2022*

- **UNISUMM and SUMMZOO: Unified Model and Diverse Benchmark for Few-Shot Summarization.**

Yulong Chen, Yang Liu, Ruochen Xu, Ziyi Yang, Chenguang Zhu, Michael Zeng, Yue Zhang. *ACL 2023*

- **Anomaly Detection with Domain Adaptation.**

Ziyi Yang, Iman Soltani Bozchalooi, Eric Darve. *CVPR 2023 workshop*

- **Make lead bias in your favor: A simple and effective method for news summarization.**

Chenguang Zhu, Ziyi Yang, Robert Gmyr, Michael Zeng, Xuedong Huang. *SIGIR 2021*

- **Universal Sentence Representations Learning with Conditional Masked Language Model**

Ziyi Yang, Yinfei Yang, Daniel Cer, Jax Law, Eric Darve. *EMNLP 2021*

- **A Simple and Effective Method To Eliminate the Self Language Bias in Multilingual Representations**

Ziyi Yang, Yinfei Yang, Daniel Cer, Eric Darve. *EMNLP 2021*

- **Memory Augmented Generative Adversarial Networks for Anomaly Detection**

Ziyi Yang, Iman Soltani Bozchalooi, Eric Darve. *IEEE Transactions on Neural Networks and Learning Systems*

- **Filtered Inner Product Projection for Multilingual Embedding Alignment.**

Vin Sachidananda, Ziyi Yang, Chenguang Zhu. *ICLR 2021*

- **Multi-Constitutive Neural Network for Large Deformation Poromechanics Problem.**

Qi Zhang\*, Yilin Chen\*, **Ziyi Yang\***, Eric Darve. *NeurIPS ML4PS Workshop 2020*

- **TED: A pretrained unsupervised summarization model with theme modeling and denoising.**

Ziyi Yang\*, Chenguang Zhu\*, Michael Zeng, Xuedong Huang, Eric Darve. *EMNLP Findings 2020*

- **Regularized Cycle Consistent Generative Adversarial Network for Anomaly Detection.**

Ziyi Yang, Iman Soltani Bozchalooi, Eric Darve. *ECAI 2020*

- **Non-parameterized Sentence Embedding via Orthogonal Basis.**

Ziyi Yang, Chenguang Zhu, Weizhu Chen. *EMNLP 2019, Oral*

- **Embedding Imputation with Grounded Language Information.**

Ziyi Yang, Chenguang Zhu, Vin Sachidananda, Eric Darve. *ACL 2019*