

# Zonghan Yang

✉ yangzh20@mails.tsinghua.edu.cn • 🌐 minicheshire.github.io  
🐦 yang\_zonghan • 🔄 minicheshire

## Education

---

- 2020.9 - **Ph.D. in Computer Science and Technology**  
CS Department, Tsinghua University  
*Advisor: Yang Liu*
- 2016.9 - 2020.6 **B.Eng. in Computer Science and Technology**  
CS Department, Tsinghua University

## Professional Experience

---

- 2018.4 - **CS Department, Tsinghua University**, Research Assistant  
*with Yang Liu*  
Language-based autonomous agents [1, 2];  
LLM alignment from the control theory perspective [5, 9, 11, 12] (Tweets: [here](#));  
Leveraging control theories to design robust neural models [8, 10, 15];  
Natural language processing: LLMs [3, 4, 5, 7], NMT [13, 14, 16].
- 2022.3 - 2022.6 **Microsoft Research Asia**, Research Intern  
*with Xiaoyuan Yi and Xing Xie*  
Responsible natural language generation with feedback control over LLMs [9].
- 2019.9 - 2020.6 **Yau Mathematical Sciences Center, Tsinghua University**, Research Assistant  
*with Chenglong Bao and Zuoqiang Shi*  
Network architecture design from the dynamical system perspective;  
Robustness improvement guaranteed by control theories [15].
- 2019.7 - 2019.9 **Language Technology Institute, Carnegie Mellon University**, Research Intern  
*with Yulia Tsvetkov*  
Error analysis and reduction in continuous-output neural machine translation.
- 2017.3 - 2018.4 **CS Department, Tsinghua University**, Research Assistant  
*with Xiaoyuan Yi and Maosong Sun*  
Ancient Chinese poetry generation in Jiuge system [17];  
Ancient and modernized Chinese poetry generation with multimodal inputs.

## Publications

---

- [1] [Towards Unified Alignment Between Agents, Humans, and Environment](#)  
**Zonghan Yang\***, An Liu\*, Zijun Liu\*, Kaiming Liu, Fangzhou Xiong, Yile Wang, Zeyuan Yang, Qingyuan Hu, Xinrui Chen, Zhenhe Zhang, Fuwen Luo, Zhicheng Guo, Peng Li, Yang Liu. Preprint. Project page [here](#), Tweets [here](#)
- [2] [Scaffolding Coordinates to Promote Vision-Language Coordination in Large Multi-Modal Models](#)  
Xuanyu Lei, **Zonghan Yang**, Xinrui Chen, Peng Li, Yang Liu. Preprint. Project page [here](#), Tweets [here](#)
- [3] [PANDA: Preference Adaptation for Enhancing Domain-Specific Abilities of LLMs](#)  
An Liu, **Zonghan Yang**, Zhenhe Zhang, Qingyuan Hu, Peng Li, Ming Yan, Ji Zhang, Fei Huang, and Yang Liu. Preprint.
- [4] [OneBit: Towards Extremely Low-bit Large Language Models](#)  
Yuzhuang Xu, Xu Han, **Zonghan Yang**, Shuo Wang, Qingfu Zhu, Zhiyuan Liu, Weidong Liu, Wanxiang Che. Preprint.
- [5] [Exploring the Impact of Model Scaling on Parameter-Efficient Tuning](#)  
Yusheng Su, Chi-Min Chan, Jiali Cheng, Yujia Qin, Yankai Lin, Shengding Hu, **Zonghan Yang**, Ning Ding, Xingzhi Sun, Guotong Xie, Zhiyuan Liu, Maosong Sun. EMNLP 2023.
- [6] [Restricted Orthogonal Gradient Projection for Continual Learning](#)  
Zeyuan Yang, **Zonghan Yang**, Peng Li, and Yang Liu. AI Open.
- [7] [Bridging the Gap between Decision and Logits in Decision-based Knowledge Distillation for Pre-trained Language Models](#)  
Qinhong Zhou, **Zonghan Yang**, Peng Li, Yang Liu. ACL 2023 (Oral).
- [8] [Improving Adversarial Robustness of Deep Equilibrium Models with Explicit Regulations Along the Neural](#)

## Dynamics

**Zonghan Yang**, Peng Li, Tianyu Pang, Yang Liu. ICML 2023.

[9] [Unified Detoxifying and Debiasing in Language Generation via Inference-time Adaptive Optimization](#)

**Zonghan Yang**, Xiaoyuan Yi, Peng Li, Yang Liu, Xing Xie. ICLR 2023.

[10] [A Closer Look at the Adversarial Robustness of Deep Equilibrium Models](#)

**Zonghan Yang**, Tianyu Pang, Yang Liu. NeurIPS 2022.

[11] [Parameter-efficient Fine-tuning of Large-scale Pre-trained Language Models](#)

Ning Ding, Yujia Qin, Guang Yang, Fuchao Wei, **Zonghan Yang**, Yusheng Su, Shengding Hu, Yulin Chen, Chi-Min Chan, Weize Chen, Jing Yi, Weilin Zhao, Xiaozhi Wang, Zhiyuan Liu, Hai-Tao Zheng, Jianfei Chen, Yang Liu, Jie Tang, Juanzi Li, Maosong Sun. Nature Machine Intelligence. Cover Article of the March Issue, 2023. Also: Extended version on [ArXiv](#)

[12] [On Robust Prefix-Tuning for Text Classification](#)

**Zonghan Yang**, Yang Liu. ICLR 2022.

[13] [Alternated Training with Synthetic and Authentic Data for Neural Machine Translation](#)

Rui Jiao, **Zonghan Yang**, Maosong Sun, Yang Liu. Findings of ACL 2021.

[14] [Neural Machine Translation: A Review of Methods, Resources, and Tools](#)

Zhixing Tan, Shuo Wang, **Zonghan Yang**, Gang Chen, Xuancheng Huang, Maosong Sun, Yang Liu. AI Open.

[15] [Interpolation between Residual and Non-Residual Networks](#)

**Zonghan Yang**, Yang Liu, Chenglong Bao, Zuoqiang Shi. ICML 2020.

[16] [Reducing Word Omission Errors in Neural Machine Translation: A Contrastive Learning Approach](#)

**Zonghan Yang**, Yong Cheng, Yang Liu and Maosong Sun. ACL 2019 (Short Paper).

[17] [Chinese Poetry Generation with a Working Memory Model](#)

Xiaoyuan Yi, Maosong Sun, Ruoyu Li, **Zonghan Yang**. IJCAI 2018.

## Teaching Experience

---

2021F, 2022F, 2023F **(40240432) Formal Language & Automaton**, Tsinghua University

CS Department, Undergraduate level course, Instructor: Yang Liu

2020F, 2020S

**(14204002) Progressive English Reading & Writing**, Tsinghua University

Language Center, Undergraduate level course, Instructor: Li Yang

## Selected Professional Service

---

2022 **AACL-IJCNLP 2022**, Co-chair of Student Research Workshop (SRW): [[Website](#)] [[Proceedings](#)]

2019 **Machine Translation Reading List**, with 2.4k stars on [[GitHub](#)]

## Selected Honors and Awards

---

2023 **Outstanding Teaching Assistant Award**, Tsinghua University

2023 **Dean's Award**, Institute for AI Industry Research, Tsinghua University

2023 **Sohu Research Scholarship**, CS Department, Tsinghua University

2023 **Travel Award**, ICLR 2023

2022 **Sohu Research Scholarship**, CS Department, Tsinghua University

2022 **"Stars of Tomorrow" Certificate**, Microsoft Research Asia

2020 **Excellent Undergraduate Thesis**, Tsinghua University

2020 **Excellent Graduate (Bachelor)**, CS Department, Tsinghua University

2018 **Enterprise Innovation Award**, 36th Tsinghua "Challenge Cup" Competition.

Awarded to my *Multimodal Poetry Generation* project.

2018 **Sohu Research Scholarship**, CS Department, Tsinghua University