JUNJIE WU

Tel: (+852)-91532290 Email: junjie.wu@connect.ust.hk Website: https://wujunjie1998.github.io

EDUCATION

- **Hong Kong University of Science and Technology** *September 2020 - Present*
  Ph.D. in Artificial Intelligence Advised by **Prof. Dit-Yan Yeung**

- **Sun Yat-sen University, Guangzhou, China** *September 2016 - June 2020*
  Bachelor in Statistics, School of Mathematics, Major GPA: 3.8/4.0
  (Second-class, Third-class Scholarship of Sun Yat-sen University, 2016-2017, 2017-2018)

EXPERIENCES

- **GenAI, Meta** *June 2025 - Sept 2025*
  Research Scientist Intern Advised by **Yide Zhao**
  Overview: *Investigating the application of LLMs in advertisement generation.*
- **Yale NLP Lab, Yale University** *Sept 2024 - Present*
  Visiting Ph.D. Student Advised by **Prof. Arman Cohan**
  Overview: *Investigating the long context understanding capability of large language models.*
- **Pattern Recognition Center, WeChat AI, Tencent** *May 2024 - Present*
  Research Intern Advised by **Mo Yu, Lemao Liu**
  Overview: *Investigating the inductive reasoning capability of large language models.*
- **Tencent AI Lab** *July 2021 - Jan 2024*
  Research Intern Advised by **Lemao Liu, Wei Bi**
  Overview: *Investigating the robustness of machine translation systems.*
- **CoAI Lab, Tsinghua University** *Oct 2019 - Aug 2020*
  Research Intern Advised by **Prof. Minlie Huang**
  Overview: *Tracking and controlling topic transition in document-grounded dialog system.*
- **Blablablab, University of Michigan** *July 2019 - June 2020*
  Research Intern Advised by **Prof. David Jurgens**
  Overview: *Predicting prosocial (defined by many metrics like healthy, supportive, politeness) outcomes in online conversations from a large-scale Reddit dataset.*
- **NGN Lab, Tsinghua University** *July 2018 - Aug 2018, Jan 2019 - May 2019*
  Research Intern Advised by **Prof. Yongfeng Huang**
  Overview: *English and Chinese text emotion analysis and classification.*

PREPRINTS (*: EQUAL CONTRIBUTION.)

1. **Junjie Wu***, Tsz Ting Chung*, Kai Chen* and Dit-Yan Yeung. "Unified Triplet-Level Hallucination Evaluation for Large Vision-Language Models "
   - Introduce a new framework to evaluate LVLMs' hallucination on triplet-level, with a benchmark dataset for evaluation and a mitigation method based on the paper's findings.
2. **Junjie Wu**, Dit-Yan Yeung. "SCAT: Robust Self-supervised Contrastive Learning via Adversarial Training for Text Classication "
   - Propose a novel contrastive learning-based approach to enhance the robustness of NLP classification models against various textual adversarial attacks.

ACCEPTED PAPERS (*: EQUAL CONTRIBUTION.)

1. **Junjie Wu***, Gefei Gu*, Yanan Zheng, Dit-Yan Yeung and Arman Cohan. "Ref-Long: Benchmarking the Long-context Referencing Capability of Long-context Language Models " (**ACL 2025**)
   - Introduce Ref-Long, a long-context benchmark that systematically evaluates the long-context referencing capability of LCLMs, which leads to several findings through experimenting on it.

2. **Junjie Wu**, Mo Yu, Lemao Liu, Dit-Yan Yeung and Jie Zhou "Understanding LLMs Fluid Intelligence Deficiency: An Analysis of the ARC Task " (**NAACL 2025 Oral**)
   - Systematically investigate the challenges LLMs face on inductive reasoning tasks through a series of experiments, and conclude many findings that could facilitate future works.
3. Mo Yu*, Lemao Liu*, **Junjie Wu***, Tsz Ting Chung*, Shunchi Zhang*, Jiangnan Li, Dit-Yan Yeung and Jie Zhou "The Stochastic Parrot on LLM's Shoulder: A Summative Assessment of Physical Concept Understanding " (**NAACL 2025 Oral**)
   - Introduce a novel physical concept understanding task called PhysiCo, revealing that the SOTA LLMs exhibit a significant gap compared to humans, showing evidence of the Stochastic Parrot phenomenon in these LLMs.
4. **Junjie Wu**, Lemao Liu, Wei Bi and Dit-Yan Yeung. "Rethinking Targeted Adversarial Attacks for Neural Machine Translation " (**ICASSP 2024**)
   - Point out a serious issue in current NMT targeted adversarial attacks, then propose a new attack setting to remedy this issue and a novel targeted adversarial attack method that outperforms previous methods.
5. **Junjie Wu**, Lemao Liu and Dit-Yan Yeung. "Towards General Error Diagnosis via Behavioral Testing in Maching Translation " (**EMNLP 2023**)
   (*Presented at the GenBench workshop at EMNLP 2023.*)
   - Design a novel bilingual translation pair generation based behavioral testing approach for machine translation systems, which could provide comprehensive and faithful behavioral testing results for general error diagnosis.
6. Jiajun Bao*, **Junjie Wu***, Yiming Zhang*, Eshwar Chandrasekharan and David Jurgens. "Conversations Gone Alright: Quantifying and Predicting Prosocial Outcomes in Online Conversations " (**The Web Conference (WWW) 2021**)
   - Identify factors that are related to the prosocial outcomes in online conversations, then design a model to predict whether a conversation will lead to prosocial outcomes or not.
7. **Junjie Wu** and Hao Zhou. "Augmenting Topic-Aware Knowledge-Grounded Conversations with Dynamic Built Knowledge Graphs" **(DeeLIO at NAACL 2021)**)
   - Propose a method to dynamically built knowledge graph from the conversation history, which helps to enhance the quality of the generated dialogs.

ACADEMIC SERVICES

Programme Committee: ACL (2023)
Reviewer: NeurIPS (2024), ICLR (2025), ICML (2025), ACL Rolling Review (ARR)

TECHNICAL SKILLS AND OTHERS

Programming: Python, Pytorch, Matlab, R, Latex
TOEFL: 105 **GRE**: V155 Q170 AW4.0
Miscs: I like playing basketball, and I am the team member of the school basketball team from 2016-now.