Wangchunshu Zhou

https://michaelzhouwang.github.io

Education

ETH Zurich   **Zurich, Switzerland**
Ph.D in Computer Science (*quitted*)   *July. 2022 - April. 2023*
Advisor: Prof. Ryan Cotterell and Prof. Mrinmaya Sachan

Beihang University   **Beijing, China**
Master in Computer Science and System Engineering
Diplome d'ingenieur (*French Engineering Degree*)   *Sep. 2018 - June. 2021*
Advisor: Prof. Ke Xu

Beihang University   **Beijing, China**
B.S. in Information and Computation Science (*Sino-French Engineering School*)
Major in Mathematics, Physics, and General Engineering.   *Sep. 2014 - Jul. 2018*

Research Interests

My research interest lies in AGI. My current research goal is to bridge the gap between powerful pre-trained large language models and their impact on the lives of humankind, and develop **Language Technology for All**.

To achieve this goal, I identify two major research topics that I'm currently interested in: **content generation** with LLMs and building LLMs that possess **superintelligence**.

Content generation with LLMs involves methods to better **control** and **customize** LLMs to generate desirable texts of **arbitrary length** that are **personalized**, **creative**, and **useful**. I am also interested in improving human-AI interaction in smart writing and build next-generation human-AI **collaborative** writing platform, which is more efficient and deliver smoother experience compared to previous AI-assissted writing systems.

As for superintelligence research, I am passionate about building algorithms that enable AI Agents or LLMs to **self-evolve** either by exploring situated environments or using large scale web corpus. Meanwhile, I also care about the superalignment problem which aim to align superhuman LLMs with human values.

I am also interested in or working in the following subjects:

o Efficient methods for LLMs [8,9,16,17,21,30] and Low resource NLP [6, 10, 11].
o Generative AI: text generation and its evaluation [1,2-7,12-15,22-25,30,34,36]
o Language model pre-training and transfer learning for NLP [12,15,32,33].
o Commonsense reasoning and knowledge-based reasoning [4,12,13,31].
o Multi-modality [20,22,29,37], text-to-image generation [22].

Publications

* denotes equal contribution

o **X^2-VLM: All-In-One Pre-trained Model For Vision-Language Tasks [37]**
  - Jiao Sun*, Yufei Tian*, Wangchunshu Zhou*, Nan Xu*, Qian Hu, John Frederick Wieting, Xuezhe Ma
  - in Proceedings of **EMNLP 2023**
  - We conduct an throughout evaluation of the abilities of LLMs on controlled text generation.

o **Evaluate Large Language Models on Controlled Generation Tasks [36]**
  - Jiao Sun*, Yufei Tian*, Wangchunshu Zhou*, Nan Xu*, Qian Hu, John Frederick Wieting, Xuezhe Ma
  - in Proceedings of **EMNLP 2023**
  - We conduct an throughout evaluation of the abilities of LLMs on controlled text generation.

o **Towards a Mechanistic Interpretation of Multi-Step Reasoning Capabilities of Language Models [35]**
  - Yifan Hou, Jiaoda Li, Yu Fei, Alessandro Stolfo, Wangchunshu Zhou, Guangtao Zeng, Antoine Bosselut, Mrinmaya Sachan
  - in Proceedings of **EMNLP 2023**
  - We conduct an analysis of the step-by-step reasoning behavior of large language models.

o **Doolittle: Benchmarks and Corpora for Academic Writing Formalization [34]**
  - Shizhe Diao, Yongyu Lei, Liangming Pan, Tianqing Fang, Wangchunshu Zhou, Sedrick Scott Keh, Min-Yen Kan,

Tong Zhang
- in Proceedings of **EMNLP 2023**
- We release a new benchmark for the task of academic writing formalization and set up the baselines.

○ **Let's Synthesize Step by Step: Iterative Dataset Synthesis with Large Language Models by Extrapolating Errors from Small Models** [33]
- Ruida Wang, Wangchunshu Zhou, Mrinmaya Sachan
- in Proceedings of **EMNLP 2023** (**Findings**)
- We propose a novel method to use LLMs to synthesize training data for smaller models by extrapolating errors smaller models make on a validation set.

○ **To Repeat or Not To Repeat: Insights from Scaling LLM under Token-Crisis** [32]
- Fuzhao Xue, Yao Fu, Wangchunshu Zhou, Zangwei Zheng, Yang You
- in Proceedings of **NeurIPS 2023**
- We investigate the problem of insufficient tokens available for pre-training larger LLMs and the impact of repeating pre-training data.

○ **Commonsense Knowledge Transfer for Pre-trained Language Models** [31]
- Wangchunshu Zhou, Ronan Le Bra, Yejin Choi
- in Proceedings of **ACL 2023** (**Findings**)
- We propose commonsense knowledge transfer, a self-supervised method for transferring the commonsense knowledge stored in a commonsense knowledge model to a general large language model.

○ **Modular Transformers: Compressing Transformers into Modularized Layers for Flexible Efficient Inference** [30]
- Wangchunshu Zhou, Ronan Le Bra, Yejin Choi
- in Proceedings of **ACL 2023** (**Findings**)
- We propose to train modularized transformers with multi-grained module replacing. Modularized transformers can be flexibly assembled according to different performance-efficiency trade-offs.

○ **Cross-View Language Modeling: Towards Unified Cross-Lingual Cross-Modal Pre-training** [29]
- Yan Zeng*, Wangchunshu Zhou*, Ao Luo, Ziming Cheng, Xinsong Zhang
- in Proceedings of **ACL 2023**
- We propose CCLM, a unified framework for multi-lingual multi-modal pre-training.

○ **Don't Group, Just Rescore: A Simpler Alternative to Constrained Beam Search** [28]
- Yuchen Eleanor Jiang, Tianyu Liu, Wangchunshu Zhou, Kexun Zhang, Vilém Zouhar, Alex Warstadt, Mrinmaya Sachan, Ryan Cotterell
- We propose dynamic beam rescoring (DBR), a simple yet effective constrained decoding method that only rescore candidates instead of grouping them.

○ **EfficientVLM: Fast and Accurate Vision-Language Models via Knowledge Distillation and Modal-adaptive Pruning** [27]
- Tiannan Wang*, Wangchunshu Zhou*, Yan Zeng, Xinsong Zhang, Lei Yu
- in Proceedings of **ACL 2023** (**Findings**)
- We propose a framework that combines general distillation and task/modal-specific pruning to compress vision-language models.

○ **Learning to Predict Persona Information for Dialogue Personalization without Explicit Persona Description** [26]
- Wangchunshu Zhou, Qifei Li, Chenle Li
- in Proceedings of **ACL 2023** (**Findings**)
- We propose to train a persona predictor to enable dialogue personalization without explicit persona information.

○ **Controlled Text Generation with Natural Language Instructions** [25]
- Wangchunshu Zhou, Yuchen Eleanor Jiang, Ethan Wilcox, Ryan Cotterell, Mrinmaya Sachan
- in Proceedings of **ICML 2023**
- We introduce a InstructCTG, a controlled text generation framework that incorporates different constraints by conditioning on natural language descriptions and demonstrations of the constraints.

○ **Automatic Educational Question Generation with Difficulty Level Controls** [24]
- Ying Jiao, Kumar Shridhar, Peng Cui, Wangchunshu Zhou and Mrinmaya Sachan
- in Proceedings of **AIED 2023**
- We propose a constrained text generation framework for automatic educational question generation with difficulty level controls.

○ **Predicting Reference-Based MT Metrics Without the Reference** [23]
- Vilém Zouhar, Shehzaad Dhuliawala, Wangchunshu Zhou, Nico Daheim, Tom Kocmi, Yuchen Eleanor Jiang, Mrinmaya Sachan.
- in Proceedings of **EACL 2023**
- We show that it is possible to train a model to predict reference-based MT metrics without the reference, which

can benefit quality estimation for machine translation systems.

o **Write and Paint: Generative Vision-Language Models are Unified Modal Learners [22]**
- Shizhe Diao, Wangchunshu Zhou, Xinsong Zhang, Jiawei Wang.
- in Proceedings of **ICLR 2023**
- We show that generative pre-training on image and text modality can be unified and results in a powerful foundation model capable of both vision, text, and vision-language tasks.

o **Efficiently Tuned Parameters Are Task Embeddings [21]**
- Wangchunshu Zhou*, Canwen Xu*, Julian McAuley.
- in Proceedings of **EMNLP 2022**
- We show that inter-task transferability can be efficiently predicted by parameter-efficient fine-tuning on source and target tasks.

o **VLUE: A Multi-Task Benchmark for Evaluating Vision-Language Models [20]**
- Wangchunshu Zhou*, Yan Zeng*, Shizhe Diao*, Xinsong Zhang*.
- in Proceedings of **ICML 2022**
- We release VLUE, a multi-task benchmark for evaluating the "true" transfer/generalize ability of pre-trained vision-language models.

o **Contextual Representation Learning beyond Masked Language Modeling [19]**
- Zhiyi Fu*, Wangchunshu Zhou*, Jingjing Xu, Hao Zhou, Lei Li.
- in Proceedings of **ACL 2022**
- We present a learning to teach method to improve knowledge distillation with meta learning.

o **BERT Learns to Teach: Knowledge Distillation with Meta Learning [18]**
- Wangchunshu Zhou*, Canwen Xu*, Julian McAuley.
- in Proceedings of **ACL 2022**
- We present a learning to teach method to improve knowledge distillation with meta learning.

o **A Survey on Green Deep Learning . [17]**
- Jingjing Xu*, Wangchunshu Zhou*, Zhiyi Fu*, Hao Zhou, Lei Li.
- We present a systematic review of the development of Green deep learning technologies.

o **Beyond Preserved Accuracy: Evaluating Loyalty and Robustness of BERT Compression. [16]**
- Canwen Xu*, Wangchunshu Zhou*, Tao Ge, Ke Xu, Julian McAuley, Furu Wei.
- in Proceedings of **EMNLP 2021** (**Oral**)
- We propose to evaluate BERT/model compression techiniques in terms of loyalty and robustness in addition to accuracy.

o **Improving Sequence-to-Sequence Pre-training via Sequence Span Rewriting. [15]**
- Wangchunshu Zhou, Tao Ge, Canwen Xu, Ke Xu, Furu Wei.
- in Proceedings of **EMNLP 2021**
- We introduce a sequence span rewriting objective which bridge the gap between pre-training and fine-tuning of sequence-to-sequence pre-trained models and provide more informative pre-training signals.

o **Learning from Perturbations: Diverse and Informative Dialogue Generation with Inverse Adversarial Training. [14]**
- Wangchunshu Zhou*, Qifei Li*, Chenle Li.
- in Proceedings of **ACL 2021** (**Oral**)
- We introduce an inverse adversarial learning framework that encourages neural dialogue agents to generate diverse and informative responses that are sensitive to perturbations in the dialogue history.

o **Blow the Dog Whistle: A Chinese Dataset for Cant Understandingwith Common Sense and World Knowledge [13]**
- Canwen Xu*, Wangchunshu Zhou*, Tao Ge, Ke Xu, Julian McAuley, Furu Wei.
- in Proceedings of **NAACL 2021**
- We introduce a large and diverse Chinese dataset for creating and understanding cant from a computational linguistics perspective.

o **Pre-training Text-to-Text Transformers for Concept-centric Commonsense. [12]**
- Wangchunshu Zhou*, Dong-Ho Lee*, Ravi Selvam, Seyeon Lee, Bill Yuchen Lin, Xiang Ren
- in Proceedings of **ICLR 2021**.
- We introduce novel self-supervised pre-training objectives in both generative and contrastive form and a joint training framework to improve the commonsense reasoning ability of pre-trained text-to-text transformers and train CALM, a Concept-Aware Language Model.

o **Connecting the Dots Between Fact Verification and Fake News Detection. [11]**
- Qifei Li*, Wangchunshu Zhou*.
- in Proceedings of **COLING 2020** (**Oral**)

- We propose to connect the dots between Fact Verification, which is better explored, and Fake News Detection, which is less developed, by using a pre-trained summarization model to summarize a news article into a short claim and then use a pre-trained fact verification model to classify the trustworthiness of the news.

o **Towards Interpretable Natural LanguageUnderstanding with Explanations as Latent Variables.** [10]
  - Wangchunshu Zhou*, Jinyi Hu*, Hanlin Zhang*, Xiaodan Liang, Maosong Sun, Chenyan Xiong, Jian Tang.
  - in Proceedings of **NeurIPS 2020**
  - We develop ELV, a general framework for interpretable natural language understanding that requires only a small set of human annotated explanations for training.

o **BERT Loses Patience: Fast and Robust Inference with Early Exit.** [9]
  - Wangchunshu Zhou*, Canwen Xu*, Tao Ge, Julian McAuley, Ke Xu, Furu Wei.
  - in Proceedings of **NeurIPS 2020**
  - We propose PABEE, a straightforward yet effective inference method that can be used as a plug-and-play technique to simultaneously improve the efficiency and robustness of a pretrained language model (PLM).

o **BERT-of-Theseus: Compressing BERT by Progressive Module Replacing.** [8]
  - Canwen Xu*, Wangchunshu Zhou*, Tao Ge, Furu Wei, Ming Zhou.
  - in Proceedings of **EMNLP 2020**
  - We propose a novel model compression approach that progressively replace each component in a large model by smaller modules. We apply our proposed approach to compress BERT and achieved state-of-the-art performance in comparable settings.

o **Pseudo Bidirectional Decoding for Local Sequence Transduction.** [7]
  - Wangchunshu Zhou, Tao Ge, Ke Xu.
  - in Proceedings of **EMNLP 2020** (**Findings**)
  - We propose a pseudo bidirectional decoding approach which utilize the characteristic of local sequence transduction tasks like grammatical error correction and spell correction.

o **Improving Grammatical Error Correction with Machine Translation Pairs.** [6]
  - Wangchunshu Zhou, Tao Ge, Chang Mu, Ke Xu, Furu Wei, Ming Zhou.
  - in Proceedings of **EMNLP 2020** (**Findings**)
  - We propose to use a pair of Machine Translation models with different qualities to synthesize pseudo-parallel data for pretraining Grammatical Error Correction models.

o **Scheduled DropHead: A Regularization Method for Transformer Models.** [5]
  - Wangchunshu Zhou, Tao Ge, Ke Xu, Furu Wei, Ming Zhou.
  - in Proceedings of **EMNLP 2020** (**Findings**)
  - We introduce DropHead, a structured dropout mechanism for the multi-head attention mechanism in the transformer models, and a specifically designed dropout rate schedule for DropHead.

o **CommonGen: A Constrained Text Generation Challenge for Generative Commonsense Reasoning** [4]
  - Bill Yuchen Lin, Wangchunshu Zhou, Ming Shen, , Pei Zhou, Chandra Bhagavatula, Yejin Choi, Xiang Ren
  - in Proceedings of **EMNLP 2020** (**Findings**)
  - We propose a novel constrained text generation task and released a dataset to test the generative commonsense reasoning ability of different NLG models.

o **Self-Adversarial Learning with Comparative Discrimination for Text Generation.** [3]
  - Wangchunshu Zhou, Tao Ge, Ke Xu, Furu Wei, Ming Zhou.
  - in Proceedings of **ICLR 2020**
  - We propose to integrate the self-play mechanism, which is commonly used in the RL community, into training of GANs to reduce the reward sparsity and mode collapse problem and make training more stable.

o **Learning to Compare for Better Training and Evaluation of Open Domain Text Generation Models.** [2]
  - Wangchunshu Zhou, Ke Xu.
  - in Proceedings of **AAAI 2020** (**Oral**).
  - We proposed a novel "Learning to Compare" paradigm and employ the skill rating system, which is commonly used to evaluate human chess players' skill, to evaluate the performance of open domain text generation systems.

o **BERT-based Lexical Substitution.** [1]
  - Wangchunshu Zhou, Tao Ge, Ke Xu, Furu Wei, Ming Zhou.
  - in Proceedings of **ACL 2019**
  - We proposed a novel lexical substitution based on pretrained masked lanuage models (e.g. BERT) to automatically propose substitute candidates and rank them without relying on external lexical resources.

Research/Work Experiences

o **Cofounder & CTO at AIWaves Inc.**   **HangZhou, China**
  Research & Development   *April. 2023 -*
  Leading the R&D team on applications of LLMs on long text generation, HAI, and LLM agents.

Pre-doctoral Young Investigator at Allen Institute for AI **Washington, U.S**
Mosaic team, Mentor: Dr. Ronan Le Bra, Prof. Yejin Choi *July. 2021 - July 2022*
Research Internship, working on commonsense reasoning and efficient text generation.

Research Scientist at Bytedance AI Lab **Beijing, China**
Multi-modal team, Leader: Dr. Hang Li *October. 2021 - June 2022*
Working on multi-modal large language model pre-training.

Research Intern at Bytedance AI Lab **Beijing, China**
NLP group, Mentor: Dr. Jingjing Xu *Mar. 2021 - Jun. 2021*
Research Internship, working on green Deep Learning.

Research Intern at Ink Lab of USC **Los Angeles, U.S.**
Ink Lab, Mentor: Prof. Xiang Ren *Jan. 2020 - Dec. 2020*
Research Internship, working on two projects related to machine commonsense reasoning.

Research Intern at MILA **Quebec, Canada**
Tang's Group, Mentor: Prof. Jian Tang *Mar. 2020 - Jun. 2020*
Research Internship, working on Interpretable Natural Language Understanding.

Research Intern at Microsoft Research Asia **Beijing, China**
Natural Language Computing Group, Mentor: Dr. Tao Ge, Dr. Furu Wei *Dec. 2018 - Present*
Research Internship, working on natural language generation and efficient pre-trained language models.

Services

- Reviewer: NeurIPS 2020-2023, AAAI 2021-2023, ICML 2021-2024, ACL 2021-2023, EMNLP 2021-2023, ICLR 2022-2024, ACL Rolling Review.
- Student Volunteer: ACL 2019, ICLR 2020, EMNLP 2020

Awards

- Baidu PhD Fellowship, class of 2022 (10 recipients worldwide)
- XiaoMi Fellowship, Beihang University, 2020 (25/3000)
- National Scholarship (Graduate Student), 2020 (4/240)
- Student Travel Grant: AAAI 2020, ICLR 2020
- Scholarship of Academic Excellence (Master), Beihang University. 2018, 2019 (top 15%)
- The CASC Award, Beihang University. 2017 (3/115)
- Scholarship of Academic Excellence (Bachelor), Beihang University. 2016, 2017 (top 20%)

Hobbies and Miscellaneous

- Video game: FIFA23: Ranked Top50 (worldwide) among all XBOX players, 20/20 wins in the Weekend League; League of Legends; Red Alert 2; Age of Empires 2.
- Reading sci-fic novels and watching sci-fic movies. Huge fan of Harry Potter series and The Wheel of Time series.
- Football (Real Madrid and CR7), watching F1 (Lewis Hamilton and Mercedes) and playing Sim Racing games.