

# Yizhe Zhang

Redmond, WA, 98052

Email: [yizhe.zhang@hotmail.com](mailto:yizhe.zhang@hotmail.com) | Telephone: +1 (919) 536-8858

Website: <https://dreasysnail.github.io/>

Google scholar: [Link](#)

Github: [Link](#)

## **Research Interest**

My recent research focus on 1) intelligent agent; 2) LLM reasoning and planning; 3) non-autoregressive generations.

## **Experience**

*05/2022 – present*

**Staff Research Scientist | Apple, Cupertino, CA**

*09/2021 – 05/2022*

**Research Scientist | Facebook AI, Menlo Park, CA**

*02/2018 – 09/2021*

**Senior Researcher | Natural Language Processing @ Microsoft Research AI, Redmond, WA**

## **Education**

*08/2013 – 02/2018*

**Ph.D. in Machine Learning | Duke University, Durham, NC**

*08/2015 – 02/2018*

**M.Sc. in Statistics | Duke University, Durham, NC**

*08/2007 – 06/2011*

**B.Sc. in Physics | Nanjing University, Nanjing, China**

## **Publications**

### **Preprint**

- Zhuofeng Wu, He Bai, Aonan Zhang, Jiatao Gu, VG Vydiswaran, Navdeep Jaitly, **Yizhe Zhang**. Divide-or-Conquer? Which Part Should You Distill Your LLM? arXiv (2024)
- **Yizhe Zhang\***, He Bai\*, Ruixiang Zhang\*, Jiatao Gu, Shuangfei Zhai, Josh Susskind, Navdeep Jaitly. How Far Are We from Intelligent Visual Deductive Reasoning? arXiv (2024)

- Xiaogeng Liu, Zhiyuan Yu, **Yizhe Zhang**, Ning Zhang, Chaowei Xiao. Automatic and universal prompt injection attacks against large language models. *arXiv* (2024)

**Peer-reviewed Conferences and Journals** (\* equally contributed)

- **Yizhe Zhang**, Jiarui Lu, Navdeep Jaitly. The Entity-Deduction Arena: A playground for probing the conversational reasoning and planning capabilities of LLMs. *ACL* (2024)
- Pratyush Maini, Skyler Seto, He Bai, David Grangier, **Yizhe Zhang**, Navdeep Jaitly. Rephrasing the web: A recipe for compute and data-efficient language modeling. *ACL* (2024)
- Xingyao Wang, Yangyi Chen, Lifan Yuan, **Yizhe Zhang**, Yunzhu Li, Hao Peng, Heng Ji. Executable code actions elicit better llm agents. *ICML* (2024)
- Jiatao Gu, Shuangfei Zhai, **Yizhe Zhang**, Lingjie Liu, Joshua M Susskind. Boot: Data-free distillation of denoising diffusion models with bootstrapping. *ICML* (2024)
- Jiatao Gu, Shuangfei Zhai, **Yizhe Zhang**, Joshua M Susskind, Navdeep Jaitly. Matryoshka diffusion models. *ICLR* (2023)
- **Yizhe Zhang**, Jiatao Gu, Zhuofeng Wu, Shuangfei Zhai, Joshua Susskind, Navdeep Jaitly. PLANNER: generating diversified paragraph via latent language diffusion model. *NeurIPS* (2023)
- Shuangfei Zhai, Tatiana Likhomanenko, Etai Littwin, Dan Busbridge, Jason Ramapuram, **Yizhe Zhang**, Jiatao Gu, Joshua M Susskind. Stabilizing transformer training by preventing attention entropy collapse. *ICML* (2023)
- Felix Faltings, Michel Galley, Baolin Peng, Kianté Brantley, Weixin Cai, **Yizhe Zhang**, Jianfeng Gao, Bill Dolan. Interactive text generation. *EMNLP* (2023)
- Jiatao Gu, Shuangfei Zhai, **Yizhe Zhang**, Miguel Angel Bautista, Josh Susskind. f-dm: A multi-stage diffusion model via progressive signal transformation. *ICLR* (2023)
- Gyuwan Kim, Jinhyuk Lee, Barlas Oguz, Wenhan Xiong, **Yizhe Zhang**, Yashar Mehdad, William Yang Wang. Bridging the training-inference gap for dense phrase retrieval. *EMNLP Findings* (2022)
- **Yizhe Zhang**\*, Deng Cai\*. Linearizing transformer with key-value memory. *EMNLP* (2022)
- Zhisong Zhang, **Yizhe Zhang**, Bill Dolan. Towards More Efficient Insertion Transformer with Fractional Positional Encoding. *EACL* (2021)
- Jiachang Liu, Dinghan Shen, **Yizhe Zhang**, Bill Dolan, Lawrence Carin, Weizhu Chen. What Makes Good In-Context Examples for GPT-3? *ACL DEELIO workshop* (2021)
- Zeqiu Wu, Michel Galley, Chris Brockett, **Yizhe Zhang**, Bill Dolan. Automatic document sketching: Generating drafts from analogous texts. *ACL Findings* (2021)
- Tianyu Liu, **Yizhe Zhang**, Chris Brockett, Yi Mao, Zhifang Sui, Weizhu Chen, Bill Dolan. A token-level reference-free hallucination detection benchmark for free-form text generation. *ACL* (2022)
- **Yizhe Zhang**, Siqi Sun, Xiang Gao, Yuwei Fang, Chris Brockett, Michel Galley, Jianfeng Gao, Bill Dolan. Joint retrieval and generation training for grounded text generation. *AAAI* (2022)

- Jungo Kasai, Hao Peng, **Yizhe Zhang**, Dani Yogatama, Yi Mao, Weizhu Chen, Noah A Smith. Finetuning Pretrained Transformers into RNNs. *EMNLP (2021, Oral Presentation)*
- Dianqi Li, **Yizhe Zhang**, Hao Peng, Liqun Chen, Chris Brockett, Ming-Ting Sun, Bill Dolan. Contextualized perturbation for textual adversarial attack. *NAACL (2021)*
- Woon Sang Cho, **Yizhe Zhang**, Sudha Rao, Asli Celikyilmaz, Chenyan Xiong, Jianfeng Gao, Mengdi Wang, Bill Dolan. Unsupervised Common Question Generation from Multiple Documents using Reinforced Contrastive Coordinator. *EACL (2021)*
- Zeqiu Wu, Michel Galley, Chris Brockett, **Yizhe Zhang**, Xiang Gao, Chris Quirk, Rik Koncel-Kedziorski, Jianfeng Gao, Hannaneh Hajishirzi, Mari Ostendorf, Bill Dolan. A Controllable Model of Grounded Response Generation. *AAAI (2021)*
- Ramakanth Pasunuru, Asli Celikyilmaz, Michel Galley, Chenyan Xiong, **Yizhe Zhang**, Mohit Bansal, Jianfeng Gao. Data Augmentation for Abstractive Query-Focused Multi-Document Summarization. *AAAI (2021)*
- **Yizhe Zhang**, Xiang Gao, Sungjin Lee, Chris Brockett, Michel Galley, Jianfeng Gao, Bill Dolan. Consistent Dialogue Generation with Self-supervised Feature Learning. *SIGDIAL (2020)*.
- **Yizhe Zhang**, Guoyin Wang, Chunyuan Li, Zhe Gan, Chris Brockett, Bill Dolan. POINTER: Constrained Text Generation via Insertion-based Generative Pre-training. *EMNLP (2020)*
- Xiang Gao, **Yizhe Zhang**, Michel Galley, Chris Brockett and Bill Dolan. Dialogue Response Ranking Training with Large-Scale Human Feedback Data. *EMNLP (2020)*
- Chunyuan Li, Xiang Gao, Yuan Li, Xiujuan Li, Baolin Peng, **Yizhe Zhang**, Jianfeng Gao. Optimus: Organizing Sentences via Pre-trained Modeling of a Latent Space. *EMNLP (2020)*
- Jianqiao Li, Chunyuan Li, Guoyin Wang, Hao Fu, Yuhchen Lin, Liqun Chen, **Yizhe Zhang** and Lawrence Carin. Improving Text Generation with Student-Forcing Optimal Transport. *EMNLP (2020)*
- Yu Cheng, Zhe Gan, **Yizhe Zhang**, Oussama Elachqar, Dianqi Li, Jingjing Liu. Contextual Text Style Transfer. *Findings of EMNLP (2020)*
- Shuyang Dai, Yu Cheng, **Yizhe Zhang**, Zhe Gan, Jingjing Liu and Lawrence Carin. Contrastively Smoothed Class Alignment for Unsupervised Domain Adaptation. *ACCV (2020)*
- Xinnuo Xu, **Yizhe Zhang**, Lars Liden, Sungjin Lee. Datasets and Benchmarks for Task-Oriented Log Dialogue Ranking Task. *Interspeech (2020)*
- Siyang Yuan, Ke Bai, Liqun Chen, **Yizhe Zhang**, Chenyang Tao, Chunyuan Li, Guoyin Wang, Ricardo Henao, Lawrence Carin. Weakly supervised cross-domain alignment with optimal transport Task. *BMVC (2020)*
- **Yizhe Zhang**, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, Bill Dolan. DialoGPT: Large-Scale Generative Pre-training for Conversational Response Generation. *demo track, ACL (2020)*

- Yichen Huang\*, **Yizhe Zhang**\*, Oussama Elachqar, Yu Cheng. INSET: Sentence Infilling with Inter-sentential Generative Pre-training. *ACL (2020)*
- Pengyu Cheng, Renqiang Min, Dinghan Shen, Christopher Malon, **Yizhe Zhang**, Yitong Li and Lawrence Carin. Improving Disentangled Text Representation Learning with Information Theoretical Guidance. *ACL (2020)*
- Xinjie Fan, **Yizhe Zhang**, Zhendong Wang, Mingyuan Zhou. Adaptive Correlated Monte Carlo for Contextual Categorical Sequence Generation. *ICLR (2020)*
- Yuan Li, Chunyuan Li, **Yizhe Zhang**, Xiujuan Li, Guoqing Zheng, Lawrence Carin, Jianfeng Gao. Complementary Auxiliary Classifiers for Label-Conditional Text Generation. *AAAI (2020)*
- Liqun Chen, Ke Bai, Chenyang Tao, **Yizhe Zhang**, Guoyin Wang, Wenlin Wang, Ricardo Henao, Lawrence Carin. Sequence Generation with Optimal-Transport-Enhanced Reinforcement Learning. *AAAI (2020)*
- Xiang Gao, **Yizhe Zhang**, Sungjin Lee, Michel Galley, Chris Brockett, Jianfeng Gao and Bill Dolan. Structuring latent spaces for stylized response generation. *EMNLP (2019)*
- Dianqi Li, **Yizhe Zhang**, Zhe Gan, Yu Cheng, Chris Brockett, Ming-Ting Sun and Bill Dolan. Domain Adaptive Text Style Transfer. *EMNLP (2019)*
- Xinnuo Xu, **Yizhe Zhang**, Lars Liden and Sungjin Lee. Unsupervised Dialogue Spectrum Generation for Log Dialogue Ranking. *SIGDIAL (2019)*, *Best paper nomination*
- Liqun Chen, Guoyin Wang, Chenyang Tao, Dinghan Shen, **Yizhe Zhang** and Lawrence Carin. Improving Textual Network Embedding with Global Attention via Optimal Transport. *ACL (2019)*
- Dinghan Shen, Asli Celikyilmaz, **Yizhe Zhang**, Liqun Chen, Xin Wang, Jianfeng Gao, Lawrence Carin. Towards Generating Long and Coherent Text with Multi-Level Latent Variable Models. *ACL (2019)*
- Xiang Gao, Sungjin Lee, **Yizhe Zhang**, Chris Brockett, Michel Galley, Jianfeng Gao, Bill Dolan. Jointly Optimizing Diversity and Relevance in Neural Response Generation. *NAACL (2019)*
- Liqun Chen, **Yizhe Zhang**, Ruiyi Zhang, Chenyang Tao, Zhe Gan, Haichao Zhang, Bai Li, Dinghan Shen, Changyou Chen, Lawrence Carin. Improving Sequence-to-Sequence Learning via Optimal Transport. *ICLR (2019)*
- **Yizhe Zhang**, Michel Galley, Jianfeng Gao, Zhe Gan, Xiujuan Li, Chris Brockett, Bill Dolan. Generating Informative and Diverse Conversational Responses via Adversarial Information Maximization. *NeurIPS (2018)*
- Liqun Chen, Shuyang Dai, Chenyang Tao, Dinghan Shen, Zhe Gan, Haichao Zhang, **Yizhe Zhang**, Lawrence Carin. Adversarial Text Generation via Feature-Mover's Distance. *NeurIPS (2018)*
- Yunchen Pu, Shuyang Dai, **Yizhe Zhang**, Zhe Gan and Lawrence Carin. Multi-Domain Joint Distribution Learning with Generative Adversarial Nets. *ICML (2018)*

- Dinghan Shen, Guoyin Wang, Wenlin Wang, Martin Renqiang Min, Qinliang Su, **Yizhe Zhang**, Chunyuan Li, Ricardo Henao and Lawrence Carin. On Simple Word-Embedding-Based Models and Associated Pooling Mechanisms. *ACL* (2018)
- Guoyin Wang, Chunyuan Li, Wenlin Wang, **Yizhe Zhang**, Dinghan Shen, Xinyuan Zhang, Ricardo Henao and Lawrence Carin. Joint Embedding of Words and Labels for Text Classification. *ACL* (2018)
- Dinghan Shen, **Yizhe Zhang**, Ricardo Henao, Qinliang Su, Lawrence Carin. Deconvolutional Latent-Variable Model for Text Sequence Matching. *AAAI* (2018).
- Wenlin Wang, Piyush Rai, Yunchen Pu, Kai Fan, **Yizhe Zhang**, Ricardo Henao, Lawrence Carin. A Flexible Probabilistic Framework for Learning to Predict Unseen Classes. *AAAI* (2018).
- Zhe Gan, Liqun Chen, Weiyao Wang, Yunchen Pu, **Yizhe Zhang**, Lawrence Carin. Triangle Generative Adversarial Networks. *NIPS* (2017).
- **Yizhe Zhang**, Changyou Chen, Zhe Gan, Lawrence Carin. Stochastic Gradient Monomial Gamma Sampler. *ICML* (2017).
- **Yizhe Zhang**, Zhe Gan, Zhi Chen, Lawrence Carin. Adversarial Feature Matching for Text Generation. *ICML* (2017).
- **Yizhe Zhang**, Xiangyu Wang, Changyou Chen, Lawrence Carin. Towards Unifying Hamiltonian Monte Carlo and Slice Sampling. *NIPS* (2016).
- Changyou Chen, Nan Ding, Chunyuan Li, **Yizhe Zhang**, Lawrence Carin. Distributed Bayesian Learning with Stochastic Gradient MCMC. *NIPS* (2016).
- **Yizhe Zhang**, Ricardo Henao, Lawrence Carin. Dynamic Poisson Factor Analysis. *ICDM* (2016).
- Kai Fan, **Yizhe Zhang**, Katherine Heller. Triply Stochastic Variational Inference for Non-linear Beta Process Factor Analysis. *ICDM* (2016).
- **Yizhe Zhang**, Ricardo Henao, Jianling Zhong, Lawrence Carin, Alexander Hartemink. Learning a Hybrid Architecture for Sequence Regression and Annotation. *AAAI* (2016).
- **Yizhe Zhang**, Ricardo Henao, Chunyuan Li, Lawrence Carin. Bayesian Dictionary Learning with Gaussian Processes and Sigmoid Belief Networks. *IJCAI* (2016).
- **Yizhe Zhang**, Changyou Chen, Ricardo Henao, Lawrence Carin. Laplacian Hamiltonian Monte Carlo. *ECML* (2016).
- **Yizhe Zhang**, Yupeng He and Chaochun Wei (2015). MOST+: a Motif Finding Approach Combining Genomic Sequence and Heterogeneous Genome-wide Signatures. *BMC Genomics*.
- Yupeng He, **Yizhe Zhang**, Guangyong Zheng and Chaochun Wei (2012). CRF-based Transcription Factor Binding Site Finding System. *BMC Genomics*.
- Jiemeng Liu, Haifeng Wang, Hongxing Yang, **Yizhe Zhang**, Jinfeng Wang, Fangqing Zhao and Ji Qi. (2012). Composition-based Classification of Short Metagenomic Sequences Elucidates the Landscapes of Taxonomic and Functional Enrichment of Microorganisms. *Nucleic Acids Research*.

## **Teaching**

- Advanced Machine Learning @Duke (STA571). Instructor: *Katherine Heller*
- Probabilistic Machine Learning @Duke (CS571). Instructor: *Cynthia Rudin*

## **Rewards**

- NeurIPS top 5% reviewer award. (2018)
- Department Fellowship. (2008-2011)
- National Excellent Graduate Scholarship (top 1%). (2012)
- Travel award: NIPS (2015, 2016), ICML (2017), ICDM (2016), IJCAI (2016), AAAI (2016)

## **Professional Services**

Area Chair: NeurIPS (2010-2024), ICML (2022-2024), ICLR (2023-2024), ACL (2020-2021), EMNLP (2022), NAACL (2023) and AAAI (2018-2021)

Action Editor: TMLR (since 2023), ARR (since 2023)

Organization committee: ACL (2020)

## **Proficiency**

- Pytorch, Tensorflow, C/C++, Python, Java, Lua, MATLAB and R.