Lingyu Gao

Toyota Technological Institute at Chicago (TTIC)
6045 S. Kenwood Ave., Chicago, IL 60637
lygao@ttic.edu | |

Summary

I am a final-year Ph.D. student working in natural language processing (NLP). My primary interest is in text classification and generation, to *identify and enhance the capabilities of the generative components of pretrained language models*.

Education

Toyota Technological Institute at Chicago (TTIC), Chicago, IL, USA
Ph.D. candidate, Computer Science, CGPA: 3.86/4.0, Advisor: *Prof. Kevin Gimpel* — *Sep '17-Present*
M.S. within Ph.D., Computer Science, Advisor: *Prof. Kevin Gimpel* — *Sep '17-Sep '19*

Tsinghua University (THU), Beijing, China
M.E., Electrical Engineering, CGPA: 3.63/4.0, Advisor: *Prof. Xiaohua Jiang* — *Sep '14-Jun '17*
B.E., Electrical Engineering and Automation, CGPA: 3.79/4.0 — *Aug '10-Jul '14*

Awards & Honors

- ETS Pre-Doctoral Fellowship *'21*
- Mitsubishi Heavy Industries Scholarship *'14*
- NARI-RELAYS Scholarship *'13*
- 1^{st} prize in Schneider Electric Programmable Logic Controller Competition *'13*
- 1^{st} grade Academic Excellence Scholarship *'11*
- 2^{nd} grade Freshman Scholarship *'10*

Selected Technical Skills

Python, PyTorch, TensorFlow, LaTeX, NumPy, Pandas

Internship Experience

Research Intern, Google LLC., Mountain View, CA, USA *May-Aug, '23*
Target: Selecting Better In-Context Learning Demonstrations for Text Classification
Key Skills: TensorFlow, Pandas, Python, NumPy, LaTeX
Models: Flan-PaLM 2 (M & L), off-the-shelf retriever (fine-tuned on mT5-base)
- Achieved a +2.6% gain on F1 macro scores over an already high baseline that matches or exceeds current benchmarks.
- Proposed constraints for demonstration selection are potentially adaptable to other applications, including ranking.

Research Intern (Remote), TikTok Inc., Chicago, IL, USA *May-Aug, '22*
Target: Generating Questions of Different Styles Controlled with Keywords
Key Skills: PyTorch, PyTorch Lightning, Python, NumPy
Models: T5, mT5, ByT5 (all base versions)
- Demonstrated that an enhanced T5 model with additional tokens, such as emojis, excels in generating keywords together with topics over other models, surpassing spaCy on keyword extraction by an F1 score of 0.21.
- Generated questions controlled with keywords, topics, and specified length. Determined that using distinct models yields better results for generating questions with different styles.

Intern (Remote), Educational Testing Service, Princeton, NJ, USA *Jun-Aug, '21*
Target: Generating and Ranking Inquisitive Questions Controlled with Question Types
Key Skills: PyTorch, Fairseq, Pandas, Python, NumPy, LaTeX
Models: RoBERTa, BART (all large versions)
- Produced diverse questions tailored to specific question types.
- Leveraged a pairwise ranker to select generated questions that matched the quality of human-crafted queries in terms of syntax, semantics, relevancy, and inquisitiveness, as validated by human assessment.

Ongoing Work and ArXiv Preprints

- **The Benefits of Looking One Step Ahead: A Decoding Strategy Incorporating Lookahead Heuristics**
  Lingyu Gao, Debanjan Ghosh, Kevin Gimpel
  Work in progress.
- **Ambiguity-Aware In-Context Learning with Large Language Models**
  Lingyu Gao, Aditi Chaudhary, Krishna Srinivasan, Kazuma Hashimoto, Karthik Raman, Michael Bendersky

Publications

†: Co-senior authors

- **ToMChallenges: A Principle-Guided Dataset and Diverse Evaluation Tasks for Exploring Theory of Mind**
  Xiaomeng Ma, **Lingyu Gao**, Qihui Xu
  The 27th Conference on Computational Natural Language Learning (CoNLL). 2023
- **The Benefits of Label-Description Training for Zero-Shot Text Classification**
  Lingyu Gao, Debanjan Ghosh†, Kevin Gimpel†
  The 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2023
- **Evaluating Transformer Models and Human Behaviors on Chinese Character Naming**
  Xiaomeng Ma, **Lingyu Gao**
  Transactions of the Association for Computational Linguistics (TACL). 2023
- **How do we get there? Evaluating transformer neural networks on English past tense inflection**
  Xiaomeng Ma, **Lingyu Gao**
  The 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (AACL-IJCNLP). 2022
- **"What makes a question inquisitive?"A Study on Type-Controlled Inquisitive Question Generation**
  Lingyu Gao, Debanjan Ghosh, Kevin Gimpel
  The 11th Joint Conference on Lexical and Computational Semantics (*SEM). 2022
- **Distractor Analysis and Selection for Multiple-Choice Cloze Questions for Second-Language Learners**
  Lingyu Gao, Kevin Gimpel, Arnar Jensson
  The 15th Workshop on Innovative Use of NLP for Building Educational Applications (BEA). 2020
- **A Cross-Task Analysis of Text Span Representations**
  Shubham Toshniwal, Haoyue Shi, Bowen Shi, **Lingyu Gao**, Kevin Gimpel and Karen Livescu
  The 5th Workshop on Representation Learning for NLP (RepL4NLP). 2020
- **Design and Heat Leak Analysis of a HTS DC Cable System**
  Lingyu Gao, Guolin Chai, and Xiaohua Jiang
  Cryogenics and Superconductivity 45.9 (2017): 41-45. (in Chinese)
- **Closed-Loop Distribution Network by a Midvoltage Flexible HTS DC System**
  Xianglong Zhang, Lin Cheng, **Lingyu Gao**, Yiqun Zhang, Zhongxi Li, Yingyu Zeng, Zhenqian Zhang, and Xiaohua Jiang
  IEEE Transactions on Applied Superconductivity 26.4 (2016): 1-4.

Teaching

Teaching Assistant | *Introduction to Machine Learning* | Instructor: *Prof. Kevin Gimpel* *Autumn '19*

Areas of Expertise

♦ Natural Language Processing ♦ Deep Learning & Machine Learning ♦ Data Analysis & Visualization

Services

- **Reviewer** for NAACL-HLT 2021, BEA (2022, 2023), EMNLP (2022, 2023), ACL 2023, TALLIP (2023, 2024), ARR 2024, COLM 2024
- **Secondary Reviewer** for EMNLP 2019 and RepL4NLP 2020
- **Volunteer** in a non-profit organization Circle Cat *May'23-Present*
- **Student Member** of TTIC DEI committee in 2020 - 2021, PhD admissions committee in 2020
- **Teaching Volunteer** in Mabian Yi Autonomous County, Sichuan, China *Jul'11*
- **Member** of Student Association for Science and Technology, EE Department *Feb-Jun, '11*