Jason Zhansheng **Phang**

☐ (+1) 312 420-0746 | ✉ jasonphang@nyu.edu | ⌂ jasonphang.com | ⌨ zphang

Education

New York University   *New York, NY*

PHD IN DATA SCIENCE   *Sep. 2019 - Expected May 2023*

- Advisor: Samuel R. Bowman

MS IN DATA SCIENCE   *Sep. 2017 - May. 2019*

- Specialization: Natural Language Processing
- GPA: 3.95 /4.00 (Cumulative)

The University of Chicago   *Chicago, IL*

BSC IN MATHEMATICS WITH SPECIALIZATION IN ECONOMICS   *Sep. 2011 - Jun. 2015*

- Major: Mathematics (with recognized specialization in Economics), Economics, and Statistics
- GPA: 3.94 / 4.00 (Cumulative), 3.94 / 4.00 (Major)

Research

Google Brain   *Mountain View, CA (Remote)*

STUDENT RESEARCHER / RESEARCH INTERN (HOST: PETER J. LIU)   *May. 2021 - Dec. 2021*

- Applying efficient Transformers to long document summarization

Google Translate   *Mountain View, CA (Remote))*

RESEARCH INTERN (HOST: ADITYA SIDDHANT, MELVIN JOHNSON)   *May. 2020 - Aug. 2020*

- Improving cross-lingual language encoders leveraging both monolingual and parallel data

Machine Learning for Language (ML^2), New York University   *New York, NY*

MEMBER (ADVISOR: SAMUEL R. BOWMAN)   *Sep. 2018 - Present*

- Studied the effectiveness of sentence representation learning methods (GPT, BERT) with respect to multi-task pretraining schemes and downstream evaluation tasks

NYU Langone Health   *New York, NY*

RESEARCH INTERN (ADVISOR: KRZYSZTOF J. GERAS)   *May. 2018 - Present*

- Applied deep learning methods to screening mammography for early breast cancer detection, with a focus on saliency map and model-interpretability methods

Publications

2021   Alicia Parrish, Angelica Chen, Nikita Nangia, Vishakh Padmakumar, **Jason Phang**, Jana Thompson, Phu Mon Htut, Samuel R. Bowman. BBQ: A Hand-Built Bias Benchmark for Question Answering. *Preprint*. [arXiv:2110.08193]

2021   **Jason Phang***, Haokun Liu, Samuel R. Bowman. Fine-Tuned Transformers Show Clusters of Similar Representations Across Layers . *Blackbox NLP 2021*. [arXiv:2109.08406]

2021   Clara Vania*, Phu Mon Htut*, William Huang*, Dhara Mungra, Richard Yuanzhe Pang, **Jason Phang**, Haokun Liu, Kyunghyun Cho, Samuel R. Bowman. Comparing Test Sets with Item Response Theory. *ACL 2021*. [arXiv:2106.00840]

2021   Leo Gao, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, **Jason Phang**, Horace He, Anish Thite, Noa Nabeshima, Shawn Presser, Connor Leahy. The Pile: An 800GB Dataset of Diverse Text for Language Modeling. *Preprint*. [arXiv:2101.00027]

2020   Yiqiu Shen, Nan Wu, **Jason Phang**, Jungkyu Park, Yiqiu Shen, Kangning Liu, Sudarshini Tyagi, Laura Heacock, S. Gene Kim, Linda Moy, Kyunghyun Cho, Krzysztof J. Geras. An interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised localization. *Medical Image Analysis, Vol. 68, 2021.* [Link]

2020   **Jason Phang**, Jungkyu Park, Krzysztof Geras. Investigating and Simplifying Masking-based Saliency Methods for Model Interpretability. *Preprint*. [arXiv:2010.09750]

2020   Nan Wu, Zhe Huang, Yiqiu Shen, Jungkyu Park, **Jason Phang**, Taro Makino, S. Gene Kim, Kyunghyun Cho, Laura Heacock, Linda Moy, Krzysztof J. Geras. Reducing false-positive biopsies with deep neural networks that utilize local and global information in screening mammograms. *Preprint*. [arXiv:2009.09282]

2020 **Jason Phang***, Iacer Calixto*, Phu Mon Htut, Yada Pruksachatkun, Haokun Liu, Clara Vania, Katharina Kann, Samuel R. Bowman. English Intermediate-Task Training Improves Zero-Shot Cross-Lingual Transfer Too. *AACL 2020*. [arXiv:2005.13013]

2020 Yada Pruksachatkun*, **Jason Phang***, Haokun Liu*, Phu Mon Htut*, Xiaoyi Zhang, Richard Yuanzhe Pang, Clara Vania, Katharina Kann, Samuel R. Bowman. Intermediate-Task Transfer Learning with Pretrained Models for Natural Language Understanding: When and Why Does It Work?. *ACL 2020*. [Link]

2020 Yada Pruksachatkun, Phil Yeres, Haokun Liu, **Jason Phang**, Phu Mon Htut, Alex Wang, Ian Tenney, Samuel R. Bowman. jiant: A Software Toolkit for Research on General-Purpose Text Understanding Models. *System Demonstrations, ACL 2020*. [Link]

2019 Nan Wu, **Jason Phang**, Jungkyu Park, Yiqiu Shen, Zhe Huang, Masha Zorin, Stanisław Jastrzębski, Thibault Févry, Joe Katsnelson, Eric Kim, Stacey Wolfson, Ujas Parikh, Sushma Gaddam, Leng Leng Young Lin, Kara Ho, Joshua D. Weinstein, Beatriu Reig, Yiming Gao, Hildegard Toth, Kristine Pysarenko, Alana Lewin, Jiyon Lee, Krystal Airola, Eralda Mema, Stephanie Chung, Esther Hwang, Naziya Samreen, S. Gene Kim, Laura Heacock, Linda Moy, Kyunghyun Cho, Krzysztof J. Geras. Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. *IEEE Transactions on Medical Imaging*. [Link] (Best Paper, AI for Social Good ICML2019 Workshop.)

2019 Alex Warstadt, Yu Cao, Ioana Grosu, Wei Peng, Hagen Blix, Yining Nie, Anna Alsop, Shikha Bordia, Haokun Liu, Alicia Parrish, Sheng-Fu Wang, **Jason Phang**, Anhad Mohananey, Phu Mon Htut, Paloma Jeretič, Samuel R. Bowman. Investigating BERT's Knowledge of Language: Five Analysis Methods with NPIs. *EMNLP 2019*. [Link]

2019 Phu Mon Htut*, **Jason Phang***, Shikha Bordia*, Samuel R. Bowman Do Attention Heads in BERT Track Syntactic Dependencies?. *Extended Abstract, Natural Language, Dialog and Speech (NDS) Symposium.* [arXiv:1911.12246]

2019 Yiqiu Shen, Nan Wu, **Jason Phang**, Jungkyu Park, Gene Kim, Linda Moy, Kyunghyun Cho, Krzysztof J. Geras. Globally-Aware Multiple Instance Classifier for Breast Cancer Screening. *Extended Abstract, MIDL 2019*. [arXiv:1906.02846]

2019 Thibault Févry, **Jason Phang**, Nan Wu, S. Gene Kim, Linda Moy, Kyunghyun Cho, Krzysztof J. Geras. Improving localization-based approaches for breast cancer screening exam classification. *Extended Abstract, MIDL 2019*. [arXiv:1908.00615]

2019 Jungkyu Park, **Jason Phang**, Yiqiu Shen, Nan Wu, S. Gene Kim, Linda Moy, Kyunghyun Cho, Krzysztof J. Geras. Screening Mammogram Classification with Prior Exams. *Extended Abstract, MIDL 2019*. [arXiv:1907.13057]

2018 **Jason Phang***, Thibault Févry* and Samuel R. Bowman. Sentence Encoders on STILTs: Supplementary Training on Intermediate Labeled-data Tasks. *Preprint*. [arXiv:1811.01088]

2018 Thibault Févry* and **Jason Phang***. Unsupervised Sentence Compression using Denoising Auto-Encoders. *Conference on Computational Natural Language Learning (CoNLL) 2018*. [arXiv:1809.02669]

* denotes joint first-authorship.

Work Experience

AQR Capital Management   *Greenwich, CT*
Research Analyst (Managed Futures / Short-term Alpha)   *Jul. 2015 - Aug. 2017*

Singapore Armed Forces (10th C4I Battalion)   *Singapore*
3rd Sergeant   *Sep. 2009 - Oct. 2010*
Private   *Mar. 2009 - Sep. 2009*

Teaching Experience

New York University: Natural Language Processing   *New York, NY*
Teaching Assistant   *Fall 2021*
New York University: Natural Language Understanding   *New York, NY*
Teaching Assistant   *Spring 2020*
New York University: Programming for Data Science   *New York, NY*
Teaching Assistant   *Fall 2019*
New York University: Natural Language Understanding   *New York, NY*
Grader   *Spring 2019*

Honors & Awards

2019-2024	**Data Science Fellowship**, New York University	*New York, NY*
2015	**General Honors**, The University of Chicago	*Chicago, IL*
2012-2015	**UChicago Careers in Business Fellow**, The University of Chicago	*Chicago, IL*
2011-2015	**Dean's List**, The University of Chicago	*Chicago, IL*
2014	**Phi Beta Kappa**, The University of Chicago	*Chicago, IL*

Projects

jiant: A Software Toolkit for Research on General-Purpose Text Understanding Models

https://github.com/nyu-mll/jiant

- Lead developer for 2.0 rewrite of *jiant*, a toolkit for NLP transfer learning research