Shi Feng *Curriculum Vitae*

CONTACT

shi.feng@gwu.edu
ihsgnef.github.io
scholar.google.com/shifeng

ACADEMIC POSITIONS

George Washington University 2024 -
Assistant Professor of Computer Science

New York University 2023 - 2024
Postdoctoral scholar
Advisor: Sam Bowman, He He

University of Chicago 2021 - 2023
Postdoctoral scholar
Advisor: Chenhao Tan

EDUCATION

University of Maryland 2016 - 2021
Ph.D. in Computer Science
Advisor: Jordan Boyd-Graber
Thesis: *Towards Human-AI Cooperation on Sequential Decision Making Problems*

Shanghai Jiao Tong University 2012 - 2016
B.S. in Computer Science
ACM Honor Class

INDUSTRY EXPERIENCE

Salesforce Research June 2020 - Sept 2020
Research Intern Remote
Advisor: Bryan McCann

Microsoft Research June 2018 - Sept 2018
Research Intern Redmond, Washington
Advisor: Jessica Lundin

Microsoft Research Asia August 2015 - February 2016
Research Intern Beijing, China
Advisors: Shujie Liu, Mu Li

PUBLICATIONS

[1] KARL: Knowledge-Aware Retrieval and Representations aid Retention and Learning in Students
Matt Shu, Nishant Balepur, **Shi Feng**, Jordan Boyd-Graber
Empirical Methods in Natural Language Processing (EMNLP), 2024.

[2] A SMART Mnemonic Sounds like "Glue Tonic": Mixing LLMs with Student Feedback to Make Mnemonic Learning Stick
Nishant Balepur, Matthew Shu, Alexander Hoyle, Alison Robey, **Shi Feng**, Seraphina Goldfarb-Tarrant, Jordan Boyd-Graber
Empirical Methods in Natural Language Processing (EMNLP), 2024.

[3] Large Language Models Help Humans Verify Truthfulness—Except When They Are Convincingly Wrong
Chenglei Si, Navita Goyal, Sherry Wu, Chen Zhao, **Shi Feng**, Hal Daumé III, Jordan Boyd-Graber
North American Chapter of the Association for Computational Linguistics (NAACL), 2024.

[4] Measuring Inductive Biases of In-Context Learning with Underspecified Demonstrations
Chenglei Si*, Dan Friedman*, Nitish Joshi, **Shi Feng**, Danqi Chen, He He
Association for Computational Linguistics (ACL), 2023.

[5] Machine Explanations and Human Understanding
Chacha Chen*, **Shi Feng***, Amit Sharma, Chenhao Tan
Transactions on Machine Learning Research (TMLR), 2023
ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT), 2023
Best Paper, *ICML Workshop on Human-Machine Collaboration and Teaming (HMCaT)*, 2022.

[6] Learning Human-Compatible Representations for Case-Based Decision Support
Han Liu, Yizhou Tian, Chacha Chen, **Shi Feng**, Yuxin Chen, Chenhao Tan
International Conference on Learning Representations (ICLR), 2023.

[7] Learning to Explain Selectively
Shi Feng, Jordan Boyd-Graber
Empirical Methods in Natural Language Processing (EMNLP), 2022.

[8] Active Example Selection for In-Context Learning
Yiming Zhang, **Shi Feng**, Chenhao Tan
Empirical Methods in Natural Language Processing (EMNLP), 2022.

[9] Calibrate Before Use: Improving Few-shot Performance of Language Models
Tony Z. Zhao*, Eric Wallace*, **Shi Feng**, Dan Klein, Sameer Singh
International Conference on Machine Learning (ICML), 2021.

[10] Concealed Data Poisoning Attacks on NLP Models
Eric Wallace*, Tony Z. Zhao*, **Shi Feng**, Sameer Singh
North American Chapter of the Association for Computational Linguistics (NAACL), 2021.

[11] Quizbowl: The Case for Incremental Question Answering
Pedro Rodriguez, **Shi Feng**, Mohit Iyyer, He He, Jordan Boyd-Graber
Journal of Machine Learning Research (JMLR), 2021.

[12] What can AI do for me: Evaluating Machine Learning Interpretations in Cooperative Play
Shi Feng, Jordan Boyd-Graber
ACM Conference on Intelligent User Interfaces (ACM IUI), 2019.

[13] Universal Adversarial Triggers for Attacking and Analyzing NLP
Eric Wallace, **Shi Feng**, Nikhil Kandpal, Matt Gardner, Sameer Singh
Empirical Methods in Natural Language Processing (EMNLP), 2019.

[14] Misleading Failures of Partial-input Baselines
Shi Feng, Eric Wallace, Jordan Boyd-Graber
Association for Computational Linguistics (ACL), 2019.

[15] Understanding Impacts of High-Order Loss Approximations and Features in Deep Learning Interpretation
Sahil Singla, Eric Wallace, **Shi Feng**, Soheil Feizi
International Conference on Machine Learning (ICML), 2019.

[16] Trick Me If You Can: Human-in-the-loop Generation of Adversarial Examples for Question Answering
Eric Wallace, Pedro Rodriguez, **Shi Feng**, Jordan Boyd-Graber
Transactions of the Association for Computational Linguistics (TACL), 2019.

[17] Pathologies of Neural Models Make Interpretation Difficult
Shi Feng, Eric Wallace, Alvin Grissom II, Mohit Iyyer, Pedro Rodriguez, Jordan Boyd-Graber
Empirical Methods in Natural Language Processing (EMNLP), 2018.

[18] Improving Attention Modeling with Implicit Distortion and Fertility for Machine Translation
Shi Feng, Shujie Liu, Nan Yang, Mu Li, Ming Zhou, Kenny Q. Zhu
International Conference on Computational Linguistics (COLING), 2016.

Teaching Experience

Courses:

- Teaching assistant: UMD's graduate level NLP course.
- Teaching assistant: UMD's graduate level ML course.
- Teaching assistant: UMD's undergraduate level Objective-oriented Programming course.
- Teaching assistant: SJTU's undergraduate level Computation Theory and Automata course.
- Teaching assistant: SJTU's undergraduate level Algorithms course.

Tutorials:

- NAACL, 2022. *Human-Centered Evaluation of Explanations*

Mentoring

Student Research Mentoring

- Chenglei Si (2022-2023), UMD Undergrad. Published [4]. Now PhD at Stanford.
- Yiming Zhang (2022-2023), UChicago Masters. Published [8] Now PhD at CMU.
- Chacha Chen (2022-2023), UChicago PhD. Published [5].
- Han Liu (2022-2023), UChicago PhD. Published [6]
- Eric Wallace (2018-2019), UMD Undergrad. Published [13, 16] Now PhD at UC Berkeley.
- Matthew Shu (2021-present), Highschool student. Now Undergrad at Yale.

Presentations

Invited Talks & Presentations & External Visits

- GWU, 2023. *Evaluating AI: From Crowdsourcing Truths to Truth-finding Processes*
- UNT, 2023. *Pragmatic Interpretability*
- UAlbany, 2023. *Pragmatic Interpretability*
- USC ISI, 2023. *Pragmatic Interpretability*
- NEC Labs Europe, 2023. *Pragmatic Interpretability*
- UPenn, 2019. *Evaluating Interpretability of NLP Models*

- UCSD, 2019. *Evaluating Interpretability of NLP Models*
- UCI, 2019. *Evaluating Interpretability of NLP Models*
- NLP Highlights Podcast, 2019. *Pathologies of Neural Models Make Interpretation Difficult*

Conference Oral Presentations: ACL 2023 [4], FAccT 2023 [5], NAACL 2021 Virtual [10], EMNLP 2019 Hong Kong [13], IUI 2019 Los Angeles [12], EMNLP 2018 Brussels [17].

Academic Service

Program Committee Member

- Conferences: ACL (2019, 2020), NeurIPS (2020, 2021, 2022), EMNLP (2018, 2019, 2020, 2021), ACL Rolling Review (2021, 2022), ICLR (2021), NAACL (2022), FAcct (2023), CoNLL (2020)
- Workshops: Human-Centered Explainable AI (CHI 2023), Trust and Reliance in AI-Assisted Tasks (CHI 2023), Dynamic Adversarial Data Collection (NAACL 2022)

Awards

- **Best Reviewer**: EMNLP (2018, 2020), NeurIPS (2020)