

# Daniel Fried

---

|                     |  |                        |
|---------------------|--|------------------------|
| CONTACT INFORMATION | dfried@cs.cmu.edu<br>dpfried.github.io   |                        |
| POSITIONS           | <b>Assistant Professor</b> , Carnegie Mellon University<br>Language Technologies Institute, School of Computer Science   | 2022 – present         |
| EDUCATION           | <b>UC Berkeley</b><br>Ph.D. in Computer Science<br><i>Adviser</i> : Dan Klein<br><i>Thesis</i> : Learning Grounded Pragmatic Communication   | 2015 – 2021            |
|                     | <b>University of Cambridge</b> , Churchill College<br>M.Phil. in Computer Science, <i>with distinction</i><br><i>Adviser</i> : Stephen Clark<br><i>Thesis</i> : Low Rank Tensor Approximations for Compositional Distributional Semantics                  | 2014 – 2015            |
|                     | <b>University of Arizona</b><br>B.S. in Computer Science, Mathematics, and Information Science, <i>summa cum laude</i><br><i>Thesis Adviser</i> : Mihai Surdeanu<br><i>Thesis</i> : Predicting Community Traits Using the Language of Food on Social Media | 2010 – 2014            |
| FORMER POSITIONS    | <b>Postdoc</b> , University of Washington<br><i>Host</i> : Luke Zettlemoyer  | 2021 – 2022            |
|                     | <b>Visiting Researcher</b> , Facebook AI Research<br><i>Host</i> : Mike Lewis  | 2021 – 2022            |
|                     | <b>Research Intern</b> , Google DeepMind<br><i>Hosts</i> : Aida Nematzadeh, Stephen Clark, Chris Dyer<br><i>Project</i> : Structured models for language-conditioned video segmentation  | Summer 2019            |
|                     | <b>Research Intern</b> , Microsoft Research<br><i>Hosts</i> : Hoifung Poon, Chris Quirk, Kristina Toutanova, Scott Wen-Tau Yih<br><i>Project</i> : Learning to rank for personalized medicine  | Summer 2016            |
|                     | <b>Undergraduate Research Assistant</b> , University of Arizona<br><i>Advisers</i> : Mihai Surdeanu, Stephen Kobourov, Paul R. Cohen<br><i>Areas</i> : Question answering, language grounding, data visualization  | 2012 – 2014            |
|                     | <b>Research Intern</b> , Nara Institute of Science and Technology<br><i>Host</i> : Kevin Duh<br><i>Project</i> : Incorporating relational semantics in word embeddings   | Fall 2013              |
|                     | <b>Research Intern</b> , RWTH Aachen University<br><i>Hosts</i> : Ali Jannesari, Zhen Li<br><i>Project</i> : Supervised learning for automatic code parallelization  | Summer 2013            |
|                     | <b>Engineering Intern</b> , Microsoft<br><i>Project</i> : Data warehousing, analytics, and visualization for Xbox Live   | Summer 2012            |
| HONORS & AWARDS     | Okawa Research Award   | 2023                   |
|                     | Google Ph.D. Fellowship in Natural Language Processing   | 2019 – 2021            |
|                     | Outstanding Graduate Student Instructor, UC Berkeley   | 2018                   |
|                     | Outstanding Reviewer, ACL  | 2018, 2020, 2021, 2022 |
|                     | Outstanding Reviewer, NeurIPS  | 2019                   |
|                     | Tencent AI Lab Fellowship  | 2018 – 2019            |
|                     | Huawei & Berkeley AI Fellowship  | 2017 – 2018            |
|                     | Best M.Phil. Student Award, Cambridge Computer Laboratory  | 2015                   |
|                     | Churchill Scholarship  | 2014 – 2015            |
|                     | NDSEG Fellowship   | 2014                   |

|   |             |
|---|-------------|
| Finalist, Hertz Graduate Fellowship                                   | 2014        |
| Outstanding Senior Award in Research, U. Arizona College of Science   | 2014        |
| Outstanding Senior Award in Academics, U. Arizona Computer Science    | 2014        |
| Outstanding Senior Award in Academics, U. Arizona Information Science | 2014        |
| Barry M. Goldwater Scholarship  | 2013        |
| National Merit Scholar  | 2010 – 2014 |
| Flinn Scholarship, Flinn Foundation of Arizona                        | 2010 – 2014 |
| Presidential Scholar, U.S. Department of Education                    | 2010        |

## JOURNAL PUBLICATIONS

1. **StarCoder: May the Source Be With You!**  
Raymond Li et al. (68 authors from the BigCode Project)  
*Transactions on Machine Learning Research (TMLR)*, 2023
2. **Human-Level Play in the Game of Diplomacy by Combining Language Models with Strategic Reasoning**  
FAIR Diplomacy Team  
*Science*, 2022
3. **Syntactic Structure Distillation Pretraining for Bidirectional Encoders**  
Adhiguna Kuncoro\*, Lingpeng Kong\*, Daniel Fried\*, Dani Yogatama, Laura Rimell, Chris Dyer, and Phil Blunsom  
*Transactions of the Association for Computational Linguistics (TACL)*, 2020
4. **Higher-Order Lexical Semantic Models for Non-Factoid Answer Reranking**  
Daniel Fried, Peter Jansen, Gustave Hahn-Powell, Mihai Surdeanu, and Peter Clark  
*Transactions of the Association for Computational Linguistics (TACL)*, 2015

## CONFERENCE PUBLICATIONS

5. **TroVE: Inducing Verifiable and Efficient Toolboxes for Solving Programmatic Tasks**  
Zhiruo Wang, Graham Neubig, and Daniel Fried  
ICML, 2024
6. **Amortizing Pragmatic Program Synthesis with Rankings**  
Yewen Pu, Saujas Vaduguru, Priyan Vaithilingam, Elena Glassman, and Daniel Fried  
ICML, 2024
7. **Asking More Informative Questions for Grounded Retrieval**  
Sedrick Keh, Justin T. Chiu, and Daniel Fried  
*Findings of NAACL*, 2024
8. **Generating Pragmatic Examples to Train Neural Program Synthesizers**  
Saujas Vaduguru, Daniel Fried, and Yewen Pu  
*International Conference on Learning Representations (ICLR)*, 2024
9. **Sotopia: Interactive Evaluation for Social Intelligence in Language Agents**  
Xuhui Zhou\*, Hao Zhu\*, Leena Mathur, Ruohong Zhang, Haofei Yu, Zhengyang Qi, Louis-Philippe Morency, Yonatan Bisk, Daniel Fried, Graham Neubig, and Maarten Sap  
*International Conference on Learning Representations (ICLR)*, 2024
10. **WebArena: A Realistic Web Environment for Building Autonomous Agents**  
Shuyan Zhou\*, Frank Xu\*, Hao Zhu\*\*, Xuhui Zhou\*\*, Robert Lo\*\*, Abishek Sridhar\*\*, Xianyi Cheng, Yonatan Bisk, Daniel Fried, Uri Alon, and Graham Neubig  
*International Conference on Learning Representations (ICLR)*, 2024
11. **API-Assisted Code Generation for Question Answering on Varied Table Structures**  
Yihan Cao\*, Shuyi Chen\*, Ryan Liu\*, Zhiruo Wang, and Daniel Fried  
*Empirical Methods in Natural Language Processing (EMNLP)*, 2023
12. **Symbolic Planning and Code Generation for Grounded Dialogue**  
Justin Chiu, Wenting Zhao, Derek Chen, Saujas Vaduguru, Alexander Rush, and Daniel Fried  
*Empirical Methods in Natural Language Processing (EMNLP)*, 2023

13. **Pragmatics in Language Grounding: Phenomena, Tasks, and Modeling Approaches**  
Daniel Fried\*, Nicholas Tomlin\*, Jennifer Hu, Roma Patel, and Aida Nematzadeh  
*Findings of EMNLP*, 2023
14. **Execution-Based Evaluation for Open-Domain Code Generation**  
Zhiruo Wang, Shuyan Zhou, Daniel Fried, and Graham Neubig  
*Findings of EMNLP*, 2023
15. **Data Augmentation for Code Translation with Comparable Corpora and Multiple References**  
Yiqing Xie, Atharva Naik, Daniel Fried, Carolyn Rose  
*Findings of EMNLP*, 2023
16. **AutoReply: Detecting Nonsense in Dialogue Introspectively with Discriminative Replies**  
Weiyang Shi, Emily Dinan, Adi Renduchintala, Daniel Fried, Athul Paul Jacob, Zhou Yu, and Mike Lewis  
*Findings of EMNLP*, 2023
17. **Generating Images with Multimodal Language Models**  
Jing Yu Koh, Daniel Fried, and Ruslan Salakhutdinov  
*Neural Information Processing Systems (NeurIPS)*, 2023
18. **Pragmatic Inference with a CLIP Listener for Contrastive Captioning**  
Jiefu Ou, Benno Kroger, and Daniel Fried  
*Findings of ACL*, 2023
19. **Contrastive Decoding: Open-ended Text Generation as Optimization**  
Xiang Lisa Li, Ari Holtzman, Daniel Fried, Percy Liang, Jason Eisner, Tatsunori Hashimoto, Luke Zettlemoyer, and Mike Lewis  
*Annual Meeting of the Association for Computational Linguistics (ACL)*, 2023
20. **Grounding Language Models to Images for Multimodal Inputs and Outputs**  
Jing Yu Koh, Ruslan Salakhutdinov, and Daniel Fried  
*International Conference on Machine Learning (ICML)*, 2023
21. **Coder Reviewer Reranking for Code Generation**  
Tianyi Zhang, Tao Yu, Tatsunori B. Hashimoto, Mike Lewis, Wen-tau Yih, Daniel Fried, and Sida I. Wang  
*International Conference on Machine Learning (ICML)*, 2023
22. **DS-1000: A Natural and Reliable Benchmark for Data Science Code Generation**  
Yuhang Lai\*, Chengxi Li\*, Yiming Wang\*, Tianyi Zhang\*, Ruiqi Zhong\*, Luke Zettlemoyer, Scott Wen-tau Yih, Daniel Fried, Sida I. Wang, and Tao Yu  
*International Conference on Machine Learning (ICML)*, 2023
23. **InCoder: A Generative Model for Code Infilling and Synthesis**  
Daniel Fried\*, Armen Aghajanyan\*, Jessy Lin, Sida I. Wang, Eric Wallace, Freda Shi, Ruiqi Zhong, Wen-tau Yih, Luke Zettlemoyer, and Mike Lewis  
*International Conference on Learning Representations (ICLR)*, 2023
24. **Natural Language to Code Translation with Execution**  
Freda Shi, Daniel Fried, Marjan Ghazvininejad, Luke Zettlemoyer, and Sida I. Wang  
*Empirical Methods in Natural Language Processing (EMNLP)*, 2022
25. **Neural Theory-of-Mind? On the Limits of Social Intelligence in Large LMs**  
Maarten Sap, Ronan Le Bras, Daniel Fried, and Yejin Choi  
*Empirical Methods in Natural Language Processing (EMNLP)*, 2022
26. **G3: Geolocation via Guidebook Grounding**  
Grace Luo\*, Giscard Biamby\*, Trevor Darrell, Daniel Fried, and Anna Rohrbach  
*Findings of EMNLP*, 2022.
27. **Inferring Rewards from Language in Context**  
Jessy Lin, Daniel Fried, Dan Klein, and Anca Dragan  
*Annual Meeting of the Association for Computational Linguistics (ACL)*, 2022

28. **Reference-Centric Models for Grounded Collaborative Dialogue**  
Daniel Fried, Justin Chiu, and Dan Klein  
*Empirical Methods in Natural Language Processing (EMNLP)*, 2021
29. **Modular Networks for Compositional Instruction Following**  
Rodolfo Corona, Daniel Fried, Coline Devin, Dan Klein, and Trevor Darrell  
*North American Chapter of the Association for Computational Linguistics (NAACL)*, 2021
30. **Learning to Segment Actions from Observation and Narration**  
Daniel Fried, Jean-Baptiste Alayrac, Phil Blunsom, Chris Dyer, Stephen Clark, Aida Nematzadeh  
*Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020
31. **Cross-Domain Generalization of Neural Constituency Parsers**  
Daniel Fried\*, Nikita Kitaev\*, and Dan Klein  
*Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019
32. **Are You Looking? Grounding to Multiple Modalities in Vision-and-Language Navigation**  
Ronghang Hu, Daniel Fried, Anna Rohrbach, Dan Klein, Trevor Darrell, and Kate Saenko  
*Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019
33. **Pragmatically Informative Text Generation**  
Sheng Shen, Daniel Fried, Jacob Andreas, and Dan Klein  
*North American Chapter of the Association for Computational Linguistics (NAACL)*, 2019
34. **Speaker-Follower Models for Vision-and-Language Navigation**  
Daniel Fried\*, Ronghang Hu\*, Volkan Cirik\*, Anna Rohrbach, Jacob Andreas, Louis-Philippe Morency, Taylor Berg-Kirkpatrick, Kate Saenko, Dan Klein\*\*, and Trevor Darrell\*\*  
*Neural Information Processing Systems (NeurIPS)*, 2018
35. **Policy Gradient as a Proxy for Dynamic Oracles in Constituency Parsing**  
Daniel Fried and Dan Klein  
*Annual Meeting of the Association for Computational Linguistics (ACL)*, 2018
36. **Unified Pragmatic Models for Generating and Following Instructions**  
Daniel Fried, Jacob Andreas, and Dan Klein  
*North American Chapter of the Association for Computational Linguistics (NAACL)*, 2018
37. **Effective Inference for Generative Neural Parsing**  
Mitchell Stern, Daniel Fried, and Dan Klein  
*Empirical Methods in Natural Language Processing (EMNLP)*, 2017
38. **Improving Neural Parsing by Disentangling Model Combination and Reranking Effects**  
Daniel Fried\*, Mitchell Stern\*, and Dan Klein  
*Annual Meeting of the Association for Computational Linguistics (ACL)*, 2017
39. **Towards Using Social Media to Identify Individuals at Risk for Preventable Chronic Illness**  
Dane Bell, Daniel Fried, Luwen Huangfu, Mihai Surdeanu, and Stephen Kobourov  
*Language Resources and Evaluation Conference (LREC)*, 2016
40. **Low-Rank Tensors for Verbs in Compositional Distributional Semantics**  
Daniel Fried, Tamara Polajnar, and Stephen Clark  
*Annual Meeting of the Association for Computational Linguistics (ACL)*, 2015
41. **Analyzing the Language of Food on Social Media**  
Daniel Fried, Mihai Surdeanu, Stephen Kobourov, Melanie Hingle, and Dane Bell  
*International Conference on Big Data*, 2014
42. **Maps of Computer Science**  
Daniel Fried and Stephen Kobourov  
*Pacific Visualization Symposium (PacificVis)*, 2014
43. **Predicting Parallelization of Sequential Programs Using Supervised Learning**  
Daniel Fried, Zhen Li, Ali Jannesari, and Felix Wolf  
*International Conference on Machine Learning and Applications*, 2013

44. **A Generative Probabilistic Framework for Learning Spatial Language**  
Colin Dawson, Jeremy Wright, Antons Rebguns, Marco Valenzuela Escárcega, [Daniel Fried](#), and Paul Cohen  
*International Conference on Development and Learning*, 2013. **Best Paper Award**
45. **Bayesian Geometric Modeling of Indoor Scenes**  
Luca Del Pero, Joshua Bowdish, [Daniel Fried](#), Bonnie Kermgard, Emily Hartley, and Kobus Barnard  
*Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012

#### WORKSHOP PUBLICATIONS

46. **SantaCoder: Don't Reach for the Stars**  
Loubna Ben Allal\*, Raymond Li\*, Denis Kocetkov\*, et al. (41 authors from the BigCode Project)  
*Deep Learning for Code Workshop*, 2023. **Best Paper Award**
47. **Modeling Perspective-Dependent Ambiguity in Grounded Collaborative Dialogue**  
Justin Chiu, Wenting Zhao, Alexander M. Rush, and [Daniel Fried](#)  
*Wordplay: When Language Meets Games Workshop*, 2022
48. **Interactive Assignments for Teaching Structured Neural NLP**  
David Gaddy, [Daniel Fried](#), Nikita Kitaev, Mitchell Stern, Rodolfo Corona, John DeNero, and Dan Klein  
*Teaching NLP Workshop at NAACL*, 2021
49. **Challenges for Using Social Media for Early Detection of Type II Diabetes Mellitus**  
Dane Bell, [Daniel Fried](#), Luwen Huangfu, Mihai Surdeanu, and Stephen Kobourov  
*International Workshop on Social Media World Sensors*, 2016
50. **Learning Low-Rank Tensors for Transitive Verbs**  
[Daniel Fried](#), Tamara Polajnar, and Stephen Clark  
*Advances in Distributional Semantics Workshop*, 2015
51. **Incorporating both Distributional and Relational Semantics in Word Representations**  
[Daniel Fried](#) and Kevin Duh  
*International Conference on Learning Representations (ICLR) Workshop*, 2015

#### PREPRINTS

52. **CodeBenchGen: Creating Scalable Execution-based Code Generation Benchmarks**  
Yiqing Xie, Alex Xie, Divyanshu Sheth, Pengfei Liu, [Daniel Fried](#), and Carolyn Rose  
arXiv, 2024.
53. **What Are Tools Anyway? A Survey from the Language Model Perspective**  
Zora Zhiruo Wang, Zhoujun Cheng, Hao Zhu, [Daniel Fried](#), and Graham Neubig  
arXiv, 2024
54. **Repetition Improves Language Model Embeddings**  
Jacob Mitchell Springer, Suhas Kotha, [Daniel Fried](#), Graham Neubig, and Aditi Raghunathan  
arXiv, 2024
55. **VisualWebArena: Evaluating Multimodal Agents on Realistic Visual Web Tasks**  
Jing Yu Koh, Robert Lo\*, Lawrence Jang\*, Vikram Duvvur\*, Ming Chong Lim\*, Po-Yu Huang\*, Graham Neubig, Shuyan Zhou, Ruslan Salakhutdinov, and [Daniel Fried](#)  
arXiv, 2024
56. **Comparative Knowledge Distillation**  
Alex Wilf, Alex Tianyi Xu, Paul Pu Liang, Alexander Obolenskiy, [Daniel Fried](#), and Louis-Philippe Morency  
arXiv, 2023

\*, \*\*: equal contribution

## STUDENTS

PhD thesis advisor:

- Jing Yu (JY) Koh (with Ruslan Salakhutdinov)
- Andy Liu (with Mona Diab)
- Saujas Vaduguru
- Zhiruo (Zora) Wang (with Graham Neubig)

MLT thesis advisor:

- Alex Xie (with Matt Gormley and Vincent Hellendoorn)

PhD committee member:

- Shuyan Zhou
- Aman Madaan
- Ta-Chung Chi
- Nikitha Rao
- Frank Xu
- Ritam Dutt
- Ruohong Zhang

## TEACHING

**Instructor: 11-891, Neural Code Generation**, CMU, Spring 2024

**Instructor: 11-877, Advanced Multimodal Machine Learning**, CMU, Spring 2024

**Instructor: 11-711, Advanced NLP**, CMU, Fall 2023

**Instructor: 11-777, Multimodal Machine Learning**, CMU, Spring 2023

**Instructor: CS188, Artificial Intelligence**, UC Berkeley, Summer 2018

**Teaching Assistant: CS188, Artificial Intelligence**, UC Berkeley, Fall 2017

**Teaching Assistant: CS245, Intro to Discrete Structures**, University of Arizona, Spring 2012

**Teaching Assistant: ISTA100, Great Ideas of the Information Age**, University of Arizona, Fall 2011

**Project Design: CS288, Natural Language Processing**, UC Berkeley, Spring 2020

## INVITED TALKS

*Interacting with LLMs for Grounded Tasks*. NILLI Workshop, EMNLP 2023. Spring 2023

*Interacting with (code) LLMs*. MIT Neurosymbolic Reading Group. Spring 2023

*InCoder, SantaCoder, and StarCoder: Findings from Training Open-Source Code LLMs*. Bloomberg AI; GitHub Next. Spring 2023

*Using Language Strategically in Context*. CoRL Workshop on Strategic Multi-Agent Interactions, 2022; Johns Hopkins University, University of Pennsylvania, UT Austin. Spring 2023

*Contextual Communication in Programming*. LTI Future of Code Generation Seminar, 2022

*Reasoning About Actions and Rewards in Language Interactions*. MIT CPL, 2022. (with Jessie Lin)

*Modularity in Grounded Interaction*. ViGIL Workshop, NAACL 2021. (with Rudy Corona)

*Learning Grounded Pragmatic Communication*. University of Arizona, TTI-Chicago, University of Southern California, Purdue, Carnegie Mellon University, UC Irvine, Université de Montréal/Mila, Allen Institute for Artificial Intelligence, Facebook AI Research, Google Research, Stanford Cognition and Language Workshop. Spring 2021

*Pragmatic Models for Generating and Following Grounded Instructions*. Stanford NLP Seminar, University of Arizona, USC ISI. Fall 2018–Spring 2019

ACADEMIC  
SERVICE

**Workshop co-organizer:** Workshop on Large Language Models for Agents, ICLR 2024

**Workshop co-organizer:** 3rd UnImplicit Workshop, EACL 2024

**Workshop co-organizer:** 2nd UnImplicit Workshop, NAACL 2022

**Workshop co-organizer:** 2nd Advances in Language and Vision Workshop, NAACL 2021

**Workshop advisory board:** Theory of Mind in Communicating Agents, ICML 2023

**Outstanding reviewer awards:** ACL 2018, NeurIPS 2019, ACL 2020, ACL 2021, ACL 2022

**Senior area chair:** ACL 2024

**Area chair:** EMNLP 2022–2023, ACL 2023

**Ethics chair:** NAACL 2024

**Reviewing:** TACL 2022–2024; ACL Rolling Review 2021–2023; ACL 2018–2022; EMNLP 2016–2021; NAACL 2019, 2021; AACL-IJCNLP 2020; EACL 2017, 2021; NeurIPS 2019–2023; ICML 2019, 2023; ICLR 2021–2023; AKBC 2021; COLING 2018; \*SEM 2016–2018; NAACL-SRW 2016, 2018; ACL-SRW 2020; IJCAI 2016; SpLU-RoboNLP 2019, 2021; NeuralGen 2019; ViGIL 2019, 2021; DeepLo 2019; ALVR 2020, 2021; LaReL 2020; Cooperative AI 2021; Meaning in Context 2021