

EDUCATION & PROFESSIONAL EXPERIENCES

Orby AI

Research Scientist

Mountain View, CA

Jun. 2024 – present

Amazon AWS AI

Applied Scientist

Bellevue, WA

Apr. 2022 – Mar. 2024

Senior Applied Scientist

Apr. 2024 – Jun. 2024

- Co-led a team of 10+ applied scientists on launching Amazon Q, a conversational assistant that accesses enterprise knowledge sources and enterprise tools
- Enterprise search, RAG, LLM evaluation, etc.

JD AI Research

Senior Research Scientist

Mountain View, CA

Jul. 2021 – Apr. 2022

Research Scientist

Oct. 2020 – Jun. 2021

- Developed ML algorithms applicable to NLP tasks and beyond with a focus on explainability, data-efficiency, and efficient interaction in human-computer collaboration.

Stanford University

Doctor of Philosophy, Computer Science

Stanford, CA

Sep. 2015 – Sep. 2020

- Research Assistant with Stanford Natural Language Processing Group (Advisor: Chris Manning)

Master of Science, Computer Science

Sep. 2013 – Jun. 2016

- Research Assistant with Stanford Artificial Intelligence Laboratory (Advisor: Andrew Ng)

Master of Science, Department of Statistics

Apr. 2016 – Apr. 2017

Tsinghua University

Research Assistant, State Key Laboratory of Intelligent Technology & Systems (Advisor: Xiaolin Hu)

Beijing, China

Jul. 2011 – Jun. 2013

Bachelor of Engineering (magna cum laude), School of Software
2012

Aug. 2008 – Jul. 2012

SELECTED PUBLICATIONS (* = Equal Contribution)

- [1] Yumo Xu, **Peng Qi***, Jifan Chen*, Kunlun Liu, Rujun Han, Lan Liu, Bonan Min, Vittorio Castelli, Arshit Gupta, Zhiguo Wang. CiteEval: Principle-Driven Citation Evaluation for Source Attribution. In *ACL*, 2025
- [2] Zhengxuan Wu, Yuhao Zhang*, **Peng Qi***, Yumo Xu*, Rujun Han, Yian Zhang, Jifan Chen, Bonan Min, Zhiheng Huang. Dancing in Chains: Reconciling Instruction Following and Faithfulness in Language Models. In *EMNLP*, 2024
- [3] Rujun Han, Yuhao Zhang, **Peng Qi**, Yumo Xu, Jinyuan Wang, Lan Liu, William Yang Wang, Bonan Min, Vittorio Castelli. RAG-QA Arena: Evaluating Domain Robustness for Long-form Retrieval Augmented Question Answering. In *EMNLP*, 2024
- [4] Kaiser Sun, **Peng Qi***, Yuhao Zhang*, Lan Liu, William Yang Wang, Zhiheng

- Huang. Tokenization Consistency Matters for Generative Models on Extractive NLP Tasks. In *Findings of the ACL: EMNLP*, 2023.
- [5] **Peng Qi***, Nina Du*, Christopher D. Manning, Jing Huang. PragmatiCQA: A Dataset for Pragmatic Question Answering in Conversations. In *Findings of the ACL*, 2023.
 - [6] Rujun Han, **Peng Qi**, Yuhao Zhang, Lan Liu, Julia Burger, William Yang Wang, Zhiheng Huang, Bing Xiang, Dan Roth. RobustQA: Benchmarking the Robustness of Domain Adaptation for Open-Domain Question Answering. In *Findings of the ACL*, 2023.
 - [7] Manoj Ghuhan Arivazhagan, Lan Liu*, **Peng Qi***, Xinchu Chen*, William Yang Wang, Zhiheng Huang. Hybrid Hierarchical Retrieval for Open-Domain Question Answering. In *Findings of the ACL*, 2023.
 - [8] Xiyang Hu, Xinchu Chen*, **Peng Qi***, Deguang Kong, Kunlun Liu, William Yang Wang, Zhiheng Huang. Language Agnostic Multilingual Information Retrieval with Contrastive Learning. In *Findings of the ACL*, 2023.
 - [9] Bo Li, **Peng Qi**, Bo Liu, Shuai Di, Jingen Liu, Jiquan Pei, Jinfeng Yi, and Bowen Zhou. Trustworthy AI: From Principles to Practices. *ACM Computing Surveys*, 2023.
 - [10] Chao Shang, Guangtao Wang, **Peng Qi**, and Jing Huang. Improving Time Sensitivity for Question Answering over Temporal Knowledge Graphs. In *Association of Computational Linguistics (ACL)*, 2022.
 - [11] **Peng Qi***, Haejun Lee*, Oghenetegiri “TG” Sido*, and Christopher D. Manning. Answering Open-Domain Questions of Varying Reasoning Steps from Text. In *Empirical Methods for Natural Language Processing (EMNLP)*, 2021.
 - [12] Chao Shang, **Peng Qi**, Guangtao Wang, Jing Huang, Youzheng Wu, Bowen Zhou. Open Temporal Relation Extraction for Question Answering. In *3rd Conference on Automated Knowledge Base Construction (AKBC)*, 2021.
 - [13] Kevin Huang, **Peng Qi**, Guangtao Wang, Tengyu Ma, and Jing Huang. Entity and Evidence Guided Document-Level Relation Extraction. In *6th Workshop on Representation Learning for NLP (RepL4NLP) at ACL 2021*, 2021.
 - [14] Xiaochen Hou, Jing Huang, Guangtao Wang, **Peng Qi**, Xiaodong He, and Bowen Zhou. Selective Attention Based Graph Convolutional Networks for Aspect-Level Sentiment Classification. In *TextGraphs-15 at NAACL 2021*, 2021.
 - [15] Yuhao Zhang, Yuhui Zhang, **Peng Qi**, Christopher D. Manning, and Curtis P. Langlotz. Biomedical and Clinical English Model Packages for the Stanza Python NLP Library. *Journal of the American Medical Informatics Association (JAMIA)*, 2021.
 - [16] Xiaochen Hou, **Peng Qi**, Guangtao Wang, Rex Ying, Jing Huang, Xiaodong He, and Bowen Zhou. Graph Ensemble Learning over Multiple Dependency Trees for Aspect-level Sentiment Classification. In *2021 Annual Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2021.
 - [17] Devendra Singh Sachan, Yuhao Zhang, **Peng Qi**, and William L. Hamilton. Do

- Syntax Trees Help Pretrained Transformers Extract Information? In *The 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 2021.
- [18] Ashwin Paranjape*, Abigail See*, Kathleen Knealy, Haojun Li, Amelia Hardy, **Peng Qi**, Kaushik Ram Sadagopan, Nguyet Minh Phu, Dilara Soylu, and Christopher D. Manning. Neural Generation Meets Real People: Towards Emotionally Engaging Mixed-Initiative Conversations. *The Alexa Prize Proceedings*, 2020.
 - [19] **Peng Qi**. Explainable and Efficient Knowledge Acquisition from Text. (Ph.D. Thesis) *Stanford University*, 2020.
 - [20] **Peng Qi**, Yuhao Zhang, and Christopher D. Manning. Stay Hungry, Stay Focused: Generating Informative and Specific Questions in Information-Seeking Conversations. In *Findings of ACL: EMNLP 2020*, 2020.
 - [21] **Peng Qi***, Yuhao Zhang*, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In *Association of Computational Linguistics (ACL), System Demonstrations*, 2020
 - [22] **Peng Qi**, Xiaowen Lin*, Leo Mehr*, Zijian Wang*, Christopher D. Manning. Answering Complex Open-domain Questions Through Iterative Query Generation. In *2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019.
 - [23] Zhilin Yang*, **Peng Qi***, Saizheng Zhang*, Yoshua Bengio, William W. Cohen, Ruslan Salakutdinov, and Christopher D. Manning. HotpotQA: A Dataset for Diverse, Explainable Multi-hop Question Answering. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018.
 - [24] Yuhao Zhang*, **Peng Qi***, and Christopher D. Manning. Graph Convolution over Pruned Dependency Trees Improves Relation Extraction. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018.
 - [25] **Peng Qi***, Timothy Dozat*, Yuhao Zhang*, and Christopher D. Manning. Universal Dependency Parsing from Scratch. In *CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 2018.
 - [26] Urvashi Khandelwal, He He, **Peng Qi**, and Dan Jurafsky. Sharp Nearby, Fuzzy Far Away: How Neural Language Models Use Context. In *56th Annual Conference of Association of Computational Linguistics (ACL)*, 2018.
 - [27] **Peng Qi** and Christopher D. Manning. Arc-swift: A Novel Transition System for Dependency Parsing. In *55th Annual Conference of Association of Computational Linguistics (ACL)*, 2017.
 - [28] Timothy Dozat, **Peng Qi**, and Christopher D. Manning. Stanford’s Graph-based Neural Dependency Parser at the CoNLL 2017 Shared Task. *CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*. **First place**
 - [29] Arun Chaganty*, Ashwin Paranjape*, Jason Bolton*, Matthew Lamm*, Jinhao Lei*, Abigail See*, Kevin Clark, Yuhao Zhang, **Peng Qi**, and Christopher D. Manning. Stanford at TAC KBP 2017: Building a Trilingual Relational Knowledge Graph. In *Text Analysis Conference (TAC)*, 2017.

- [30] Yuhao Zhang*, Arun Chaganty*, Ashwin Paranjape*, Danqi Chen*, Jason Bolton*, **Peng Qi**, and Christopher D. Manning. Stanford at TAC KBP 2016: Sealing Pipeline Leaks and Understanding Chinese. In *Proceedings of the Text Analysis Conference (TAC) - Knowledge Base Population*, 2016.
- [31] Andrew L. Maas, **Peng Qi**, Ziang Xie, Awni Y. Hannun, Daniel Jurafsky, and Andrew Y. Ng. Building DNN Acoustic Models for Large Vocabulary Speech Recognition. *Computer Speech & Language*, 2016.
- [32] **Peng Qi** and Xiaolin Hu. Learning nonlinear statistical regularities in natural images by modeling the outer product of image intensities. *Neural computation*, 26(4):693–711, 2014.
- [33] Xiaolin Hu, Jianwei Zhang, **Peng Qi**, and Bo Zhang. Modeling response properties of V2 neurons using a hierarchical k-means model. *Neurocomputing*, 134:198–205, 2014.
- [34] **Peng Qi**, Shuochen Su, and Xiaolin Hu. Modeling outer products of features for image classification. In *Advanced Computational Intelligence (ICACI)*, 2013.
- [35] Xiaolin Hu, **Peng Qi**, and Bo Zhang. Hierarchical k-means algorithm for modeling visual area V2 neurons. In *Neural Information Processing (ICONIP)*, pages 373–381, 2012. *Best Paper Award*

PREPRINTS & IN SUBMISSION

- [36] Yu Gu, Kai Zhang, Yuting Ning, Boyuan Zheng, Boyu Gou, Tianci Xue, Cheng Chang, Sanjari Srivastava, Yanan Xie, **Peng Qi**, Huan Sun, Yu Su Is Your LLM Secretly a World Model of the Internet? Model-Based Planning for Web Agents. *arXiv preprint arXiv:2411.06559*, 2024
- [37] Hendrik Schuff, Heike Adel, **Peng Qi**, Ngoc Thang Vu. Challenges in Explanation Quality Evaluation. *arXiv preprint arXiv:2210.07126*, 2022.
- [38] **Peng Qi**, Jing Huang, Youzheng Wu, Xiaodong He, and Bowen Zhou. Conversational AI Systems for Social Good: Opportunities and Challenges. *arXiv preprint arXiv:2105.06457*, 2021.
- [39] **Peng Qi***, Guangtao Wang*, and Jing Huang. SpanDrop: Simple and Effective Counterfactual Learning for Long Sequences. *arXiv preprint arXiv:2208.02169*, 2022.

PATENTS

- Method and system for aspect-level sentiment classification by merging graphs. U.S. Patent US20230267322A1 (pending)

HONORS

Inaugural Yufan Award (Rising Stars Category) at World Artificial Intelligence Conference 2020 (20 top young Chinese AI scientists worldwide), Facebook ParLAI Research Award, China’s National Scholarship (top 3% university-wide at Tsinghua), Freshman Scholarship (provincial top 10 in college entrance exam), and other merit-based awards from undergrad

EXPERIENCE

Facebook AI Research (New York)

Jun. 2017 – Sep. 2017

Learning to Teach through Communication (Mentors: Jason Weston, Douwe Kiela, Kyunghyun Cho)

- Studied emergent teaching behavior of machine learning agents in a constrained communication setting
- Proposed a reinforcement learning-based method for agents to learn to teach
- Implemented the proposed method in ParlAI with positive results on image classification tasks

Contributing Open Source Projects

Stanza [↗](#) (documentation [↗](#))

- PyTorch implementation of Stanford's full system in the 2018 CoNLL Shared Task on Universal Dependency Parsing from Raw Text
- Includes neural network models for tokenization, part-of-speech tagging, lemmatization, dependency parsing, and NER in 60+ languages with pretrained models available and intuitive Python interface
- Pipelines for syntactic analysis and NER for English-language biomedical text
- Main contributor and maintainer

HuggingFace Transformers [↗](#)

- Bugfixes in various standard example scripts for running BERT for question answering

Universal Dependencies [↗](#)

- Maintainer of the Chinese-GSD treebank and its corresponding version in simplified Chinese
- Converted the GSD treebank to simplified Chinese, fixed annotations and segmentation when necessary, engaged in community discussions about annotation standards

ParlAI [↗](#)

- Implemented various features to enable teacher training in ParlAI
- Bugfixes and stability improvements

Caffe [↗](#)

- Implemented a framework for generic solvers
- Implemented Nesterov's Accelerated Gradient solver and AdaGrad solver
- Contributed neuron layers (Leaky ReLU, Mean-variance normalization)
- Compatibility issues / bug fixes for Mac OS

Kaldi [↗](#)

- Contributed the first training recipe for the Fisher/Switchboard mixed speech corpus (the largest speech corpus in use in academia)
- A bug fix for Kaldi's speaker identification for better speaker heldout training

ConvolutionalRBM.m [↗](#) (Owner)

- An implementation of Honglak Lee *et al.*'s convolutional restricted Boltzmann machine model in Matlab, MEX (C++/CUDA)

GraphCut [↗](#)

- A fast CPU/GPU hybrid implementation for min-cut problems in grid graphs

Other Project Experiences (Highlights)

2008 – 2012

- Short Text Classification based on Universal Knowledge Base
- TextTriBute: A Distributed Framework for Semantic Text Retrieval
- Semantic Query Optimization in HyperSQL-DB the Open Source Database
- Semantic-directed LR Grammar Analyzer Generator [↗](#)

INVITED TALKS

- Seminar speaker. Explainable and Efficient Knowledge Acquisition from Text. University of Southern California. (Spring 2020)

- Seminar speaker. Explainable and Efficient Knowledge Acquisition from Text. University of Arizona. (Spring 2020)
- Seminar speaker. Answering Complex Open-domain Questions in the Wild. University of Southern California Information Sciences Institute (USC ISI). (Fall 2019)
- Seminar speaker. Answering Complex Open-domain Questions in the Wild. Stanford OVAL Seminar. (Fall 2019)
- Seminar speaker. Answering Complex Open-domain Questions in the Wild. University of California at Los Angeles. (Fall 2019)
- Invited speaker. Open-domain Text Understanding at Scale with Multi-step Reasoning. Facebook AI Research. (Summer 2019)
- Invited speaker. Open-domain Text Understanding at Scale with Multi-step Reasoning. Cresta, Inc. (Summer 2019)
- Invited speaker. Open-domain Text Understanding at Scale with Multi-step Reasoning. Samsung Research America. (Summer 2019)
- Invited speaker. Open-domain Text Understanding at Scale with Multi-step Reasoning. Salesforce Research. (Spring 2019)
- Seminar speaker. Answering Complex Open-domain Questions Through Iterative Query Generation. University of Washington. (Summer 2019)
- Speaker. A Gentle Introduction to Reinforcement Learning. Stanford NLP Reading Group. (Spring 2019)
- Invited speaker. A tutorial on Deep Learning for Natural Language Processing. LingCon Hackathon training day [🔗](#) (Fall 2017)

TEACHING & MENTORING

- TA, CS 224D Deep Learning for Natural Language Processing (Spring 2015)
- TA, CS 124 From Languages to Information (Winter 2015)
- TA, CS 145 Introduction to Databases (Summer 2014)
- TA, CS 224S Spoken Language Processing (Spring 2014)
- Project Mentor, CS 224n (Winter 2019, Best Custom Project Report Prize; Winter 2018, Best Custom Project Report Prize; Winter 2017)
- Tutor, CS 145 Introduction to Databases, CS 107 Computer Organization and Systems, CS 245 Database Systems Principles

SERVICE

Session Chair: NAACL 2021

Area Chair: EMNLP 2024 (via ARR), ACL 2024 (via ARR), ACL 2023, NAACL 2021

Publicity Chair: NAACL 2021

Program Committee:

2021 ACL Rolling Review, AKBC, MRQA, AAAI

2020 EMNLP (Outstanding Reviewer)

2019 EMNLP-IJCNLP (Outstanding Reviewer), MRQA, ACL, NAACL-HLT

2018 EMNLP (Best Reviewer Award), ACL, CoNLL Shared Task, UDW

2017 ACL

2013 ICACI

Journal: IEEE TNNLS