

783 Dreese Laboratories, 2015 Neil Ave, Columbus, OH 43210, USA

■ su.809@osu.edu | ★http://ysu1989.github.io | 🞧 ysu1989 | Google Scholar

Education _

University of California, Santa Barbara

Santa Barbara, CA, USA

Ph.D. IN COMPUTER SCIENCE

2012 - 2018

- · Advisor: Xifeng Yan
- Dissertation: Towards Democratizing Data Science with Natural Language Interfaces
- Outstanding Dissertation Award

Tsinghua UniversityBeijing, China

B.Eng. IN Computer Science and Technology

2008 - 2012

• Outstanding Freshman & Outstanding Graduate Award

Professional Experience _

Department of Computer Science & Engineering, The Ohio State University

ASSISTANT PROFESSOR

Columbus, OH, USA 2020 - Present

Microsoft Semantic Machines

Berkeley, CA, USA

SENIOR RESEARCHER

2018 - Present

Department of Computer Science & Engineering, The Ohio State University

Columbus, OH, USA

VISITING ASSISTANT PROFESSOR

2018 - 2019

2016, 2017

Microsoft Research

Redmond, WA, USA

U.S. Army Research Laboratory

Adelphi, MD, USA

VISITING RESEARCHER

RESEARCH INTERN

2015, 2016

IBM T.J. Watson Research Center

Yorktown Heights, NY, USA

RESEARCH INTERN

2015

Teaching___

SP22	Instructor, CSE5525: Foundations of Speech and Language Processing (U/G)	OSU
AU21	Instructor (SEI=4.22/5), CSE6521: Artificial Intelligence (G)	OSU
AU20	Instructor (SEI=4.75/5), CSE5539: Cutting-Edge Topics in Natural Language Processing (U/G)	OSU
SP20	Instructor (SEI=4.61/5), CSE5243: Introduction to Data Mining (U/G)	OSU

Grants

Since joining OSU in 2020, I have obtained 9 funded projects, totaling \$36.6M, with approximately \$2.1M for my share. I am the PI or Co-PI for 6 of the projects, totaling \$870K. In addition to core AI research, I also actively participate in large multi-disciplinary endeavors and serve in leadership roles. In the NSF AI Institute - ICICLE, I was the lead writer for both overall vision and AI and currently serve as lead of AI (with Eric Folser-Lussier). In the NSF HDR Institute - Imageomics, I serve as the lead of Machine Learning Foundations. My research is funded through both government agencies including NSF, NIH (pending), and ARL (pending) as well as industry partners (Cisco, Walmart, Amazon, and Fujitsu).

[9] **Cisco:** *Transformer-based Graph Neural Networks for Knowledge Graph Reasoning at Scale* Role: Sole PI. Grant: \$199,876. 2022-2023.

[8] Cisco: Building Transparent Interactive Natural Language Interfaces to Databases

Role: Co-PI. PI: Huan Sun. Grant: \$199,976. My share: \$64,184. 2022-2023.

[7] Walmart: Product Question Answering for Walmart Online Shopping

Role: Sole PI. Grant: \$120,109. 2022-2023.

[6] Amazon: Alexa Prize Simbot Challenge

Role: Sole PI. Grant: \$250,000. 2022-2023.

NSF National AI Research Institute: ICICLE: Intelligent CyberInfrastructure with Computational Learning in the Environment

Role: Senior Personnel (*Lead of AI*). PI: Dhabaleswar K. Panda. Grant: \$19,999,998. My share: approximately \$450,000. 2021-2026.

NSF HDR Institute: Imageomics: A New Frontier of Biological Information Powered by Knowledge-guided Machine Learning

Role: Senior Personnel (*Lead of ML Foundations*). PI: Tanya Berger-Wolf. Grant: \$14,969,077. My share: approximately \$800,000. 2021-2026.

[3] NSF Convergence Accelerator: Actionable Sensemaking Tools for Curating and Authenticating Information in the Presence of Misinformation during Crises

Role: Senior Personnel. PI: Srinivasan Parthasarathy. Grant: \$749,997. 2021-2022.

[2] **OSU TDAI Research Pilot Grant:** *Practical and Comprehensive Social Media Pharmacovigilance* Role: Sole Pl. Grant: \$50,000. 2021-2022.

[1] **Fujitsu Lab of America:** *Large-scale, Ontology-preserving Knowledge Graph Embedding* Role: Sole Pl. Gift grant: \$50,000. 2020-2021.

Students

PHD STUDENTS

- Yu Gu (SP20)
- Vardaan Pahuja (SP20)
- Bernal Jiménez Gutiérrez (SP20)
- Chan Hee Song (AU20)
- Samuel Stevens (AU21)
- Shijie Chen (AU21 , co-advised with Huan Sun)

UNDERGRADUATE STUDENTS

- Clay Washington (OSU CIS; 2021)
- Nikolas McNeal (OSU CIS; 2021)
- Sparsh Kustagishettar (OSU CSE; 2022)
- Alison Zhong (OSU CSE; 2022)
- Samuel Stevens (OSU CSE; 2020 2021). Placement: PhD at OSU.

HIGH SCHOOL STUDENTS

• Evan Huang (2021 summer): Submission to AMIA 2022 Annual Symposium High School Scholars program

GRADUATE DISSERTATION/THESIS COMMITTEE MEMBER

- Xiang Deng (CSE PhD): "Al-powered Data System: Unified Semantic Layer and Natural Language Interface"
- Arpan Jain (CSE PhD): "Novel Parallelization Strategies for High-Performance DNN Training on HPC Systems"
- Zichen Zhang (CSE PhD): "Design of Unmanned Aerial Systems for Precision Agriculture using Machine Learning Techniques"
- Zhen Wang (CSE PhD): "Knowledge-Centric Natural Language Processing: Acquisition, Representation and Reasoning"
- Saumya Sahai (CSE MS): "A Study on Informal Fallacies in Online Discussions"

- Ziyu Yao (CSE PhD): "On Advancing Natural Language Interfaces: Data Collection, Model Development, and User Interaction"
- Md Muntasir Zitu (BMI PhD): "Identification of Adverse Drug Events from clinical free texts of Electronic Health Records (EHR) data using Natural Language Processing"
- Jie Zhao (CSE PhD): "Towards Building Trustworthy Automatic Question Answering Systems"
- Semih Cayci (ECE PhD): "Online Learning for Optimal Control of Communication and Computing Systems"
- Adam Stiff (CSE PhD): "Mitigation of Data Scarcity Issues for Semantic Classification in a Virtual Patient Dialogue Agent"

Publications

I have published <u>36</u> refereed articles, including 2 journal, 2 workshop, and 32 conference articles. <u>26</u> of the conference publications appear in top conferences per csrankings.org including ACL (5), EMNLP (12), NAACL (3), WWW (2), SIGKDD (1), CVPR (1), SIGIR (1), and AAAI (1).

+: My advisee at OSU. #: My intern at Microsoft

REFEREED JOURNAL PAPERS

[2] Random Control Selection for Conducting High-throughput Adverse Drug Events Screening Using Large-scale Longitudinal Health Data

Chien-Wei Chiang, Pengyue Zhang, Macarius Donneyong, You Chen, Yu Su, Lang Li CPT: Pharmacometrics & Systems Pharmacology (**PSP**), 2021

[1] Task-Oriented Dialogue as Dataflow Synthesis

Semantic Machines, Jacob Andreas, John Bufe, David Burkett, Charles Chen, Josh Clausman, Jean Crawford, Kate Crim, Jordan DeLoach, Leah Dorner, Jason Eisner, Hao Fang, Alan Guo, David Hall, Kristin Hayes, Kellie Hill, Diana Ho, Wendy Iwaszuk, Smriti Jha, Dan Klein, Jayant Krishnamurthy, Theo Lanman, Percy Liang, Christopher H. Lin, Ilya Lintsbakh, Andy McGovern, Aleksandr Nisnevich, Adam Pauls, Dmitrij Petters, Brent Read, Dan Roth, Subhro Roy, Jesse Rusak, Beth Short, Div Slomin, Ben Snyder, Stephon Striplin, Yu Su, Zachary Tellman, Sam Thomson, Andrei Vorobev, Izabela Witoszko, Jason Wolfe, Abby Wray, Yuchen Zhang, and Alexander Zotov *Transactions of the Association for Computational Linguistics* (**TACL**), 2020

Deployed as the official conversational interface for Microsoft Outlook

REFEREED CONFERENCE PAPERS

[32] One Step at a Time: Long-Horizon Vision-and-Language Navigation with Milestones

Chan Hee Song⁺, Jihyung Kil, Tai-Yu Pan, Brian M. Sadler, Wei-Lun Chao, Yu Su IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2022

[31] Bridging the Generalization Gap in Text-to-SQL Parsing with Schema Expansion

Chen Zhao[#], Yu Su, Adam Pauls, Emmanouil Antonios Platanios Annual Conference of the Association for Computational Linguistics (**ACL**), 2022

[30] ReasonBERT: Pre-trained to Reason with Distant Supervision

Xiang Deng, Yu Su, Alyssa Lees, You Wu, Cong Yu, and Huan Sun Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2021

[29] A Systematic Investigation of KB-Text Embedding Alignment at Scale

Vardaan Pahuja⁺, Yu Gu⁺, Wenhu Chen, Mehdi Bahrami, Lei Liu, Wei-Peng Chen, and Yu Su *Annual Conference of the Association for Computational Linguistics* (**ACL**), 2021

[28] Compositional Generalization for Neural Semantic Parsing via Span-level Supervised Attention

Pengcheng Yin, Hao Fang, Graham Neubig, Adam Pauls, Emmanouil Antonios Platanios, Yu Su, Sam Thomson, and Jacob Andreas

Annual Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT), short paper, 2021

[27] Beyond I.I.D.: Three Levels of Generalization for Question Answering on Knowledge Bases

Yu Gu⁺, Sue Kase, Michelle Vanni, Brian Sadler, Percy Liang, Xifeng Yan, and Yu Su *The Web Conference* (**TheWebConf, previously WWW**), 2021

[26] An Imitation Game for Learning Semantic Parsers from User Interaction

Ziyu Yao, Yiqi Tang⁺, Scott Wen-tau Yih, Huan Sun, and Yu Su Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2020

[25] KGPT: Knowledge-Grounded Pre-Training for Data-to-Text Generation

Wenhu Chen, Yu Su, Xifeng Yan, and William Yang Wang Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020

[24] Document Classification for COVID-19 Literature

Bernal Jiménez Gutiérrez⁺, Jucheng Zeng, Dongdong Zhang, Ping Zhang, and Yu Su *Findings of EMNLP*, short paper, 2020

[23] Logical Natural Language Generation from Open-Domain Tables

Wenhu Chen, Jianshu Chen, Yu Su, Zhiyu Chen, and William Yang Wang Annual Conference of the Association for Computational Linguistics (ACL), 2020

[22] Model-based Interactive Semantic Parsing: A Unified Formulation and A Text-to-SQL Case Study

Ziyu Yao, Yu Su, Huan Sun, and Scott Wen-tau Yih

Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019

[21] HierCon: Hierarchical Organization of Technical Documents based on Concepts

Keqian Li, Shiyang Li, Semih Yavuz, Hanwen Zha, Yu Su, and Xifeng Yan *IEEE International Conference on Data Mining* (**ICDM**), 2019

Best of ICDM 2019 Selection

[20] Global Textual Relation Embedding for Relational Understanding

Zhiyu Chen, Hanwen Zha, Honglei Liu, Wenhu Chen, Xifeng Yan, and Yu Su *Annual Conference of the Association for Computational Linguistics* (**ACL**), short paper, 2019

[19] How Large A Vocabulary Does Text Classification Need? A Variational Approach on Vocabulary Selection

Wenhu Chen, Yu Su, Yilin Shen, Zhiyu Chen, Xifeng Yan and William Yang Wang Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2019

[18] Learning to Compose Topic-Aware Mixture of Experts for Zero-Shot Video Captioning

Xin Wang, Jiawei Wu, Da Zhang, Yu Su, William Yang Wang *AAAI Conference on Artificial Intelligence* (**AAAI**), 2019

[17] Concept Mining via Embedding

Keqian Li, Hanwen Zha, Yu Su, Xifeng Yan IEEE International Conference on Data Mining (ICDM), 2018

[16] XL-NBT: A Cross-lingual Neural Belief Tracking Framework

Wenhu Chen, Jianshu Chen, Yu Su, Xin Wang, Dong Yu, Xifeng Yan and William Yang Wang 23th Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018

[15] What It Takes to Achieve 100% Condition Accuracy on WikiSQL

Semih Yavuz, Izzeddin Gur, Yu Su, Xifeng Yan 23th Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2018

[14] DialSQL: Dialogue Based Structured Query Generation

Izzeddin Gur, Semih Yavuz, Yu Su, Xifeng Yan 56th Annual Meeting of the Association for Computational Linguistics (ACL), 2018 (Acceptance rate: 258/1018=25.3%)

[13] Natural Language Interfaces with Fine-Grained User Interaction: A Case Study on Web APIs

Yu Su, Ahmed H. Awadallah, Miaosen Wang, Ryen White 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2018 (Acceptance rate: 86/409=21.0%)

[12] Global Relation Embedding for Relation Extraction

Yu Su*, Honglei Liu*, Semih Yavuz, Izzeddin Gur, Huan Sun, Xifeng Yan (*: equal contribution)

16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 2018

(Acceptance rate: 207/665=31.1%)

[11] Unsupervised Neural Categorization for Scientific Publications

Kegian Li, Hanwen Zha, Yu Su, Xifeng Yan

18th SIAM International Conference on Data Mining (SDM), 2018

(Acceptance rate: 87/375=23.2%)

[10] Building Natural Language Interfaces to Web APIs

Yu Su, Ahmed H. Awadallah, Patrick Pantel, Michael Gamon, Madian Khabsa, Mark Encarnacion 26th ACM International Conference on Information and Knowledge Management (CIKM), 2017 (Acceptance rate: 171/820=20.9%)

[9] Cross-domain Semantic Parsing via Paraphrasing

Yu Su, Xifeng Yan

22th Conference on Empirical Methods in Natural Language Processing (EMNLP), 2017

(Acceptance rate: 216/836=25.8%)

[8] An End-to-End Deep Framework for Answer Triggering with a Novel Group-Level Objective

Jie Zhao, Yu Su, Ziyu Guan, Huan Sun

22th Conference on Empirical Methods in Natural Language Processing (EMNLP), short paper, 2017

(Acceptance rate: 107/582=18.4%)

[7] Recovering Question Answering Errors via Query Revision

Semih Yavuz, Izzeddin Gur, Yu Su, Xifeng Yan

22th Conference on Empirical Methods in Natural Language Processing (EMNLP), short paper, 2017

(Acceptance rate: 107/582=18.4%)

[6] On Generating Characteristic-rich Question Sets for QA Evaluation

Yu Su, Huan Sun, Brian Sadler, Mudhakar Srivatsa, Izzeddin Gur, Zenghui Yan, Xifeng Yan 21th Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016

(Acceptance rate: 177/687=25.8%)

[5] Improving Semantic Parsing via Answer Type Inference

Semih Yavuz, Izzeddin Gur, Yu Su, Mudhakar Srivatsa, Xifeng Yan

21th Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016

(Acceptance rate: 177/687=25.8%)

[4] A Fast Kernel for Attributed Graphs

Yu Su, Fangqiu Han, Richard E. Harang, Xifeng Yan

16th SIAM International Conference on Data Mining (SDM), 2016

(Acceptance rate: 96/371=25.9%)

[3] Table Cell Search for Question Answering

Huan Sun, Hao Ma, Xiaodong He, Wen-Tau Yih, Yu Su, Xifeng Yan

25th International Conference on World Wide Web (WWW), 2016

(Acceptance rate: 115/727=15.8%)

[2] Visual Graph Query Formulation and Exploration: A New Perspective on Information Retrieval at the Edge

Sue Kase, Michelle Vanni, Joanne Knight, Yu Su, Xifeng Yan

SPIE 9851, Next-Generation Analyst IV (SPIE Defense+Security), 2016

[1] Exploiting Relevance Feedback in Knowledge Graph Search

Yu Su, Shengqi Yang, Huan Sun, Mudhakar Srivatsa, Sue Kase, Michelle Vanni, Xifeng Yan 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2015

(Acceptance rate: 160/819=19.5%)

REFEREED WORKSHOP PAPERS

[2] An Investigation of Language Model Interpretability via Sentence Editing

Samuel Stevens⁺ and Yu Su BlackboxNLP Workshop at EMNLP, 2021

[1] On the Validity of Geosocial Mobility Traces

Zengbin Zhang, Lin Zhou, Xiaohan Zhao, Gang Wang, Yu Su, Miriam Metzger, Haitao Zheng, and Ben Y. Zhao *12th ACM Workshop on Hot Topics in Networks* (**HotNets**), 2013 (Acceptance rate: 26/110=23.6%)

CONFERENCE TUTORIALS

[2] Scalable Construction and Querying of Massive Knowledge Bases

Xiang Ren, Yu Su, Pedro Szekely, Xifeng Yan *The Web Conference* (**TheWebConf**), 2018

[1] Construction and Querying of Large-Scale Knowledge Bases

Xiang Ren, Yu Su, Xifeng Yan

26th ACM International Conference on Information and Knowledge Management (CIKM), 2017

PATENTS

[2] Natural Language to API Conversion (co-inventor)

US Patent Pending 15/919,209

A framework for building interactive natural language interface to web APIs.

[1] Natural Language Interface to Web API (co-inventor)

US Patent Granted 10496452B2

A framework for building natural language interface to web API via crowdsourcing with a hierarchical probabilistic model and optimization algorithm for crowdsourcing.

Professional Service _____

ORGANIZING COMMITTEE

- Faculty Advisor to Student Research Workshop, NAACL 2021.
- First Workshop on Natural Language Processing for Programming, Bangkok, Thailand, 2021 (co-located with ACL).
- First Workshop on Natural Language Interfaces (NLI), Seattle, United States, 2020 (co-located with ACL).
- First Workshop on Knowledge Base Construction, Reasoning and Mining (KBCOM), Los Angeles, United States, 2018 (co-located with WSDM).

PROGRAM COMMITTEE MEMBER/REVIEWER

- Annual Meeting of the Association for Computational Linguistics (ACL): 2017-2022 (area chair in 2022)
- Conference on Empirical Methods in Natural Language Processing (EMNLP): 2017-2021
- Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL): 2019, 2021, 2022
- ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD): 2018-2022 (session chair in 2019 and 2020)
- AAAI Conference on Artificial Intelligence (AAAI): 2019, 2021
- The International World Wide Web Conference (WWW): 2018
- SIAM International Conference on Data Mining (SDM): 2019
- ACM International Conference on Information and Knowledge Management (CIKM): 2017
- Conference on Automated Knowledge Base Construction (AKBC): 2019, 2020, 2021
- Conference on Computational Natural Language Learning (Conll): 2018
- CCF Conference on Natural Language Processing and Chinese Computing (NLPCC): 2017, 2018, 2020 (area chair)
- ACL Rolling Review (ARR): 2021 present

- Transactions of the Association for Computational Linguistics (TACL): 2021 present
- IEEE Transactions on Neural Networks and Learning Systems (TNNLS): 2020 present
- IEEE Transactions on Knowledge and Data Engineering (**TKDE**): 2017 present

OUTREACH ACTIVITIES

- 2020-present. Faculty mentor in the Ohio State LSAMP (Louis Stokes Alliances for Minority Participation) Program.
 - Mentor female African-American undergraduate students in CSE. Meet twice a semester.
- March 2018. Introduce data science and AI to high school students from Laguna Blanca School, Santa Barbara, CA.
- March 2018. "Afternoon with Academia," UCSB chapter of Society of Asian Scientists and Engineers.

University Service

- Faculty advisor to OSU Competitive Programming Club: 2020-present
- TDAI Fall Forum Organizing Committee: 2021
- Graduate Admission Committee: 2019-2022 (campus visit co-organizer in 2022)
- Space Committee: 2020-2021
- Faculty Search Committee: AI (2021-2022), Human-Centric Computing (2021-2022)

Honors & Awards _____

2019	Best of IEEE ICDM 2019 Selection	IEEE
2019	Outstanding Dissertation Award of Computer Science	UCSB
2018	Best Distinguished Graduate Student Lecture	UCSB
2016	SIAM International Conference on Data Mining Travel Award	NSF
2012	Outstanding Graduate Award	Tsinghua University
2009 - 2011	Merit-based Fellowships	Tsinghua University
2008	Outstanding Freshman Fellowship	Tsinghua University

Invited Talks

- [22] A New Paradigm of Conversational AI OSU TDAI Deep Learning Summer School. June 2, 2022. Columbus, Ohio, United States.
- [21] Emerging Frontiers of Conversational Al **JD.com, Inc**. April 28, 2022. Virtual.
- [20] Emerging Frontiers of Conversational Al **Nanjing University**. April 20, 2022. Virtual.
- [19] Emerging Frontiers of Conversational Al **Institute for Interdisciplinary Information Sciences, Tsinghua University**. November 2, 2021. Virtual.
- [18] Emerging Frontiers of Conversational AI **USC**. November 2, 2021. Virtual.
- [17] Emerging Frontiers of Conversational AI **Beijing Academy of Artificial Intelligence**. May 6, 2021. Beijing, China.
- [16] Task-Oriented Dialogue as Dataflow Synthesis Faculty-in-Focus Seminar, CSE@OSU. November 09, 2020. Columbus, OH, United States.
- [15] Towards Democratizing Data Science with Al-powered Knowledge Engines **Stanfard NLP Seminar**. February 07, 2019. Stanford, CA, United States.
- [14] Bridging the Gap between Human and Data with AI

 Microsoft Semantic Machines. August 13, 2018. Berkeley, CA, United States.
- [13] Bridging the Gap between Human and Data with Al **Fudan University**. July 24, 2018. Shanghai, China.
- [12] Bridging the Gap between Human and Data with Al **Tsinghua University**. June 21, 2018. Beijing, China.

- [11] Bridging the Gap between Human and Data with Al **Fujitsu Laboratories of America**. March 9, 2018. Sunnyvale, California, United States.
- [10] Bridging the Gap between Human and Data with AI

 University of Massachusetts. February 15, 2018. Amherst, Massachusetts, United States.
- [9] Bridging the Gap between Human and Data with AI BMI@OSU. January 26, 2018. Columbus, Ohio, United States.
- [8] Bridging the Gap between Human and Data with AI **CSE@OSU**. January 25, 2018. Columbus, Ohio, United States.
- [7] On Building Natural Language Interface to Data and Services **Toutiao Al Lab**. October 17, 2017. Beijing, China.
- [6] On Building Natural Language Interface to Data and Services **Tsinghua University**. October 16, 2017. Beijing, China.
- [5] On Building Natural Language Interface to Data and Services **Fudan University**. October 10, 2017. Shanghai, China.
- [4] Building Natural Language Interfaces to Web APIs Microsoft Research, MSR AI Seminar. July 25, 2017. Redmond, Washington, United States.
- [3] Learning from Zero: Recent Advances in Bootstrapping Semantic Parsers using Crowdsourcing **USC/ISI**. October 28, 2016. Los Angeles, California, United States.
- [2] Exploiting Relevance Feedback in Knowledge Graph Search IBM Research. August 15, 2015. Yorktown Heights, New York, United States.
- [1] Search over Large-scale Knowledge Graphs
 UCSB, IGERT Network Science Seminar. August 14, 2014. Santa Barbara, California, United States.