Sadra Sabouri

213-675-4649 sabourih@usc.edu linkedin.com/sadrasabouri github.com/sadrasabouri
Los Angeles, California

Education

University of Southern California — Los Angeles, California
PhD in Computer Science — *Aug. 2023 – May 2028 (expected)*

University of Southern California — Los Angeles, California
Master's of Computer Science — *Dec. 2025*

Sharif University of Technology — Tehran, Iran
BSc in Electrical Engineering and Computer Science — *May 2023*

Publications

Auditing and Controlling AI Agent Actions in Spreadsheets — Under submission, 2026
S Sabouri, *Z Saghi, R Huang, S Maladi, E Eufracio, S Gulwani, S Chattopadhyay*

How LLMs' Anthropomorphism and Sycophancy Compromise Expert Evaluation of Therapeutic AI — Under submission, 2026
S Sabouri, *Y Daryani, S Haghighi, J May, M Dehghani, S Chattopadhyay*

Prompting Reflection: Encouraging Users to Critically Evaluate AI Suggestions — Under submission, 2026
S Sabouri, *Y Wang, F Gu, R Huang, AK Zhao, D Peskov, JB Graber, J May, JK Kummerfeld, S Chattopadhyay*

Cognitive Biases in LLM-Assisted Software Development — ICSE, 2026
Xinyi Zhou, Zeinabsadat Saghi, ***Sadra Sabouri****, Rahul Pandita, Mollie McGuire, Souti Chattopadhyay*

PyMilo: A Python Library for ML I/O — JOSS, 2026
AmirHosein Rostami, Sepand Haghighi, ***Sadra Sabouri****, Alireza Zolanvari*

Did You Think This Through?: Effect of Varying Advice Styles on Users' Performance and Contemplation — VL/HCC, 2026
S Sabouri, *JB Graber, J May, JK Kummerfeld, S Chattopadhyay*

Synthesizing Program Analyzers to Help Programmers Answer Questions About Code — VL/HCC, 2026
A Nazari, ***S Sabouri****, W B Zhu, R Jia, M Raghothaman, S Chattopadhyay*

ELI-Why: Evaluating the Pedagogical Utility of Language Model Explanations — ACL, 2025
Brihi Joshi, Keyu He, Sahana Ramnath, ***Sadra Sabouri****, Kaitlyn Zhou, Souti Chattopadhyay, Swabha Swayamdipta, Xiang Ren*

ParsiPy: NLP Toolkit for Historical Persian Texts in Python — ALP @ NAACL, 2025
Farhan Farsi, Parnian Fazel, Sepand Haghighi, ***Sadra Sabouri****, Farzaneh Goshtasb, Nadia Hajipour, Ehsaneddin Asgari, Hossein Sameti*

Persian in a Court: Benchmarking VLMs in Persian Multi-Modal Tasks — EvalMMG, 2025
F Farsi, SS Motlagh, S Bali, ***S Sabouri****, S Momtazi*

Trust Dynamics in AI-Assisted Development: Definitions, Factors, and Implications — ICSE, 2025
Sadra Sabouri*, Philipp Eibl, Xinyi Zhou, Morteza Ziyadi, Nenad Medvidovic, Lars Lindemann, Souti Chattopadhyay*

ReaxFF Parameter Set for Boron Clusters and Icosahedral Boron Crystals — J. Phys. Chem. C, 2025
Amin Ahmadisharaf, Adri C. T. van Duin, Bin Liu, Dylan Evans, ***Sadra Sabouri****, Jeffrey Comer*

PahGen: Generating Ancient Pahlavi Text via Grammar-Guided Zero-Shot Translation — LoResMT @ NAACL, 2025
Farhan Farsi, Parnian Fazel, Farzaneh Goshtasb, Nadia Hajipour, ***Sadra Sabouri****, Ehsaneddin Asgari, Hossein Sameti*

Samila: A Generative Art Generator — Preprint, arXiv, 2025
Sadra Sabouri*, Sepand Haghighi, Elena Masrour*

Beyond the Page: Enriching Academic Paper Reading with Social Media Discussions — UIST, 2025
Run Huang, Anna Katherine Zhao, Zeinabsadat Saghi, ***Sadra Sabouri****, Souti Chattopadhyay*

Exploring the Challenges and Opportunities of AI-assisted Codebase Generation — VL/HCC, 2025
Philipp Eibl, ***Sadra Sabouri****, Souti Chattopadhyay*

A Semidefinite Relaxation Approach for Fair Graph Clustering — Ethical AI Workshop @ KDD, 2024
S Baharlouei, ***S Sabouri***

naab: A Ready-to-Use Plug-and-Play Corpus for Farsi — JAIAI, 2024
Sadra Sabouri*, Elnaz Rahmati, Soroush Gooran, Hossein Sameti*

Nafas: Breathing Gymnastics Application Preprint, arXiv, 2024
***S Sabouri**, S Haghighi*

SynTran-fa: Generating Comprehensive Answers for Farsi QA Pairs via Syntactic Transformation Preprint, Preprints.org, 2024
*F Farsi, **S Sabouri**, K Kashfipour, S Gooran, H Sameti, E Asgari*

Representative Sample Size for Estimating Saturated Hydraulic Conductivity via Machine Learning Water Resources Research, 2024
*Amin Ahmadisharaf, Reza Nematirad, **Sadra Sabouri**, Yakov Pachepsky, Behzad Ghanbarian*

A Review of the Recent Speech Recognition Methods J. Vibration and Sound, 2023
*H Hadian, S Gooran, **S Sabouri**, S Sadeghi, Y Amini, H Sameti*

Docalog: Multi-Document Dialogue System Using Transformer-Based Span Retrieval DialDoc @ ACL, 2022
*Sayed Hesam Alavian, Ali Satvaty, **Sadra Sabouri**, Ehsaneddin Asgari, Hossein Sameti*

Experimental Dataset of Electrochemical Efficiency of a Direct Borohydride Fuel Cell (DBFC) Preprint, ChemRxiv, 2022
*Sarmin Hamidi, **Sadra Sabouri**, Sepand Haghighi, Kasra Askari*

A Literature Review on Rater Agreement Metrics Technical report, pycm.io, 2022
*A Zolanvari, S Haghighi, **S Sabouri***

Research Experience

Affiliated Researcher 2025 – Present
USC Information Sciences Institute, [CUTE LAB NAME] (PI: Jonathan May) *Los Angeles, California*

Research Assistant Aug. 2023 – Present
Adaptive Computing Experience (ACE) Lab, USC (PI: Souti Chattopadhyay) *Los Angeles, California*

Research Assistant Aug. 2021 – Nov. 2022
Speech and Language Processing Lab, Sharif University of Technology (PI: Hossein Sameti) *Tehran, Iran*

Research Assistant Intern
University of Sydney

Industry Experience

Applied Scientist Intern Summer 2026
Microsoft PROSE (PROgram Synthesis using Examples)

Machine Learning Engineer Intern 2020 – 2023
Asr Gooyesh Pardaz *Tehran, Iran*

Co-founder 2019
Open Science Laboratory (OpenSciLab)

Open-Source Projects

GitHub and PyPI figures are a snapshot taken August 2026; downloads are for the preceding 30 days.

Memor: conversational memory across LLMs 2025
Lead developer, OpenSciLab. Manages the memory of a user's interactions with LLMs: past conversations can be tapped when prompting, and parts of an exchange with one model carried over to another, bridging otherwise isolated LLM instances.
67 stars, 4 forks, 960 downloads/month

ToCount: lightweight token estimator 2025
Developer, OpenSciLab. Estimates token counts for LLM input using rule-based and ML methods, for prompt analysis and token budgeting. The design point is the cost of the answer: an exact count means loading a whole tokenizer.
23 stars, 2 forks, 32 downloads/month

XNum: universal numeral system converter 2025
Developer, OpenSciLab. Converts digits across English, Persian, Hindi, Arabic-Indic, Bengali, and other numeral systems, auto-detecting mixed formats and converting only the numbers while leaving surrounding text untouched.
35 stars, 1 fork, 50 downloads/month

IPSpot: system IP address fetcher 2025
Developer, OpenSciLab. Reports both the private address a machine holds on its own network and the public address the internet sees it from, over IPv4 and IPv6.
104 stars, 6 forks, 97 downloads/month

Drux: drug release analysis framework 2025
Developer, OpenSciLab. Simulates and plots drug release profiles from mathematical models, an extensible and reproducible platform for quantitative analysis in pharmaceutical research.
25 stars, 1 fork, 41 downloads/month

PyMilo: Python library for ML I/O 2024
Co-developer, OpenSciLab. A flexible I/O interface for ML pipelines that serializes models into a transparent, non-executable format, so a model can be read before it is loaded rather than run on trust.
169 stars, 7 forks, 73 downloads/month

OPR: optimized primer design tool 2024
Developer, OpenSciLab. Designs, validates, and optimizes primers for PCR, qPCR, and sequencing, consolidating primer design tooling into abstracted Python objects.
27 stars, 0 forks, 70 downloads/month

Nava: OS-native sound engine in Python 2023
Lead developer, OpenSciLab. Plays sound in Python with no dependencies and no platform restrictions on Windows, macOS, and Linux. Recognized on the **Real Python podcast**. Funded by a **PSF Software Development Grant**.
161 stars, 6 forks, 4,329 downloads/month

naab: Farsi text corpus 2022
Lead, Sharif SLP Lab. A large, cleaned, ready-to-use open-source Farsi corpus: ∼130GB, 250 million paragraphs, 15 billion words. Used to train the **first T5 models for Farsi**. Released plug-and-play, with the cleaning already done.

Sharif-Wav2Vec2.0: Farsi speech recognition model 2022
Developer, Sharif SLP Lab. Wav2Vec2.0 fine-tuned on 108 hours of Farsi audio from Common Voice, with the token set and language models reworked for Farsi-specific features rather than retrained alone.

Samila: generative art generator 2021
Creator and co-developer, OpenSciLab. A Python prototyping tool for artists that makes an image by scattering many thousands of points, each placed by a formula with random parameters. Gained significant traction in Web3 and NFT communities.
1,176 stars, 79 forks, 281 downloads/month

Nafas: breathing gymnastics application 2021
Co-developer, OpenSciLab. A collection of breathing exercises for people who have been at a computer too long, aimed at improved concentration through raised oxygen levels.
196 stars, 10 forks, 33 downloads/month

MyTimer: a timer for command-line enthusiasts 2021
Developer, OpenSciLab. A simple but comprehensive timer set directly from the command line, so people who spend the day in a shell do not have to leave it to find one.
147 stars, 11 forks, 66 downloads/month

Art: ASCII art library for Python 2019
Creator and co-developer, OpenSciLab. Places typed characters so that together they form a visual shape across several lines: ASCII art delivered as a library a program can call. Added to **conda-forge**; cited in scientific papers on LLM security research. Funded by a **PSF Software Development Grant**.
2,492 stars, 156 forks, 1,578,035 downloads/month

PyCM: multi-class confusion matrix library 2019
Co-developer, OpenSciLab. Evaluates classifiers from predictions already made, supporting most per-class and overall statistics, for the case where accuracy alone tells you nothing useful. Added to **conda-forge**, adopted as a dependency by **Keras** and **TensorBoard**, and used in **200+ research papers**, mostly medical.
1,506 stars, 126 forks, 16,075 downloads/month

PyRGG: Python random graph generator 2019
Developer, OpenSciLab. Synthesizes graph data at any requested size for network simulation, writing several graph file formats including DIMACS-Graph. Adopted by **10+ research teams worldwide**.
222 stars, 31 forks, 507 downloads/month

OPEM: open-source PEM fuel cell simulation tool 2019
Developer, ECSIM. Evaluates proton exchange membrane fuel cell performance and predicts optimum operating parameters, bundling static and dynamic models of the cell in one package.
231 stars, 65 forks, 560 downloads/month

Mycoffee: coffee brewing recipes for terminal users
Developer, OpenSciLab. A community-based coffee recipe CLI application, well received by the programming community on social networks, notably Reddit.

Honors and Awards

Tinker Research Grant, Thinking Machines Lab Jan. 2026
One-year compute grant for fine-tuning and post-training large language models. The funded project is on **Persian prosody-informed LLMs**.

NLnet Foundation Grant (PyCM) Feb. 2025
NLnet Foundation, via NGI0 Commons / ZERO COMMONS and NGI ASSURE open-source calls. Funding to develop distance/similarity metrics and benchmarking in PyCM.

PSF Software Development Grant (Nava) Mar. 2025
Python Software Foundation. Funding to develop OS-based sound engines for Nava, part of the PSF's support for Python packages that foster open science and accessible knowledge.

Trelis AI Micro-Grant (PyCM) Sep. 2024
Development micro-grant to build an accessible RESTful API for PyCM, making the library usable by people not fluent in Python during the ML evaluation phase.

PSF Software Development Grant (Art) May 2024
Python Software Foundation. Funding to add multi-line and custom font features to the Art library.

Vector Scholarship in AI
Vector Institute. Stipend awarded to top students, recognizing exceptional candidates in Vector Institute-recognized or AI-focused study paths.

USC Presenter for ViterbiTrek
USC Viterbi School of Engineering. Selected as a presenter at ViterbiTrek, a program connecting students with industry professionals; participated in a day tour of **SupplyFrame** and **TradeDesk**.

Talks and Posters

PyCon US 2026, Memor Poster May 2026
Poster, Long Beach, California. Presented Memor: managing and transferring conversational memory across LLMs.

Teaching and Outreach

Section Leader, Stanford Python Program
Stanford University. Section leader for a global program teaching the Python programming language.

Workshop Instructor, NLP and Speech Processing
Sharif University of Technology, Tehran. Taught a session on natural language processing and speech processing.

Service

Peer review

- ACM Conference on Human Factors in Computing Systems (CHI) 2026: two research papers
- ACM Conference on Intelligent User Interfaces (IUI) 2026: two research papers
- ACM Symposium on User Interface Software and Technology (UIST) 2026: two research papers
- Creativity & Cognition Conference (C&C) 2026: two poster papers
- International Conference on Human-Agent Interaction (HAI) 2026: one research paper
- International Journal of Human-Computer Interaction (IJHCI): one research paper
- Transactions on Software Engineering and Methodology (TOSEM): one research paper
- The Journal of Systems & Software: one research paper
- Journal of Open Research Software (JORS): one research paper
- Nature Scientific Data: one dataset paper

Skills

Programming languages: Python, C, C++, Java, JavaScript, TypeScript, HTML, CSS, MATLAB, Assembly, Verilog, VHDL, LaTeX

Machine learning and AI: Machine learning, deep learning, natural language processing, large language models, convolutional neural networks, agentic AI development, agent-based modeling

ML frameworks: Transformers, PyTorch, TensorFlow, TRL, LangChain, DSPy

Speech and signal: Speech processing, speech recognition (ASR), signal processing, digital image processing

Immersive and visual computing: Augmented reality, virtual reality, holography, Unity3D

Web and backend: React, Node, FastAPI, Flask, Django, PostgreSQL, web applications, chatbot development

Data: Data science, data visualization, statistics, mathematics, Hadoop, Spark, Google Firebase, Google Analytics

Software and infrastructure: Software development, software design, software infrastructure, Docker, Git, CI/CD, DevOps, GNU/Linux, open-source development

Hardware and embedded: FPGA, Arduino, Raspberry Pi, circuit analysis, Android, AutoCAD

Study design: Research design, user interviews, surveys and questionnaire design, A/B testing, heuristic evaluation, mixed-methods research, hypothesis formulation and testing

Analysis: Qualitative and quantitative data analysis, thematic analysis, statistical analysis, behavioral data analysis, parametric and nonparametric tests

User experience: UX research, user experience testing, user behavior, human-centered AI design

Research and design tools: Qualtrics, Prolific, Figma, Notion

Professional: Teamwork, leadership, project management, public speaking, presentation skills, teaching

References

Souti Chattopadhyay, Assistant Professor of Computer Science, University of Southern California. PhD advisor; runs the Adaptive Computing Experience (ACE) Lab.

Jonathan May, Associate Professor of Computer Science, USC (ISI). Secondary lab affiliation at ISI.

Sumit Gulwani, Microsoft PROSE team. Hosted the summer 2026 applied scientist internship; co-author on the spreadsheet-agents submission.

Jonathan K. Kummerfeld, Assistant Professor of Computer Science, University of Sydney. Co-author on the AI-advice line of work.

Nenad Medvidovic, Professor of Computer Science, University of Southern California. Co-author on the trust dynamics study.

Sepand Haghighi, Independent Researcher, Open Science Laboratory. Long-running open-source collaborator since 2019.