

Stephen Macke[†]

Department of Computer Science
University of Illinois, Urbana-Champaign

Email: smacke@illinois.edu
Website: <https://smacke.net>

- Education*
- ◇ **University of Illinois at Urbana-Champaign**, Urbana, IL
Ph.D. in Computer Science · 2015—
Advisor: Aditya Parameswaran
 - ◇ **Stanford University**, Stanford, CA
M.S. in Computer Science · 2013—2015
 - ◇ **University of Tulsa**, Tulsa, OK
B.S. in Computer Science, Applied Math Major · 2009—2013
Summa Cum Laude, University Honors
Departmental Honors in Computer Science, Mathematics
- Research Interests*
- ◇ Scalable analytics and machine learning, with a focus on bringing *safety* and *structure* to common analytics tasks, while preserving tight feedback and interactivity.
- Publications*
- ◇ **S. Macke**, H. Gong, D. Lee, A. Head, D. Xin, A. Parameswaran
Fine-Grained Lineage for Safer Notebook Interactions
Under revision for VLDB 2021.
 - ◇ **S. Macke**, M. Aliakbarpour, I. Diakonikolas, A. Parameswaran, R. Rubinfeld
Rapid Approximate Aggregation with Distribution-Sensitive Interval Guarantees
To appear at ICDE 2021.
 - ◇ D. Petersohn, **S. Macke**, D. Xin, W. Ma, D. Lee, X. Mo, J. Gonzalez, A. Joseph, J. Hellerstein, A. Parameswaran
Towards Scalable Dataframe Systems
46th Int. Conf. on Very Large Data Bases (VLDB), Tokyo, Japan, Aug. 2020
 - ◇ D. Lee^{*}, **S. Macke**^{*}, D. Xin^{*}, A. Lee, S. Huang, A. Parameswaran
A Human-in-the-loop Perspective on AutoML: Milestones and the Road Ahead
IEEE Data Engineering Bulletin, June 2019
 - ◇ D. Xin, **S. Macke**, L. Ma, J. Liu, R. Ma, S. Song, A. Parameswaran
Helix: Holistic Optimization for Accelerating Iterative Machine Learning
45th Int. Conf. on Very Large Data Bases (VLDB), Los Angeles, California, Aug. 2019
 - ◇ **S. Macke**, A. Beutel, T. Kraska, M. Sathiamoorthy, D. Cheng, E. Chi
Lifting the Curse of Multidimensional Data with Learned Existence Indexes
ML for Systems Workshop at NeurIPS Int'l Conf. on Neural Information Processing Systems, Dec. 2018 *and*
Bay Area Machine Learning Symposium, Oct. 2018
 - ◇ **S. Macke**, Y. Zhang, S. Huang, A. Parameswaran
Adaptive Sampling for Rapidly Matching Histograms
44th Int'l Conf. on Very Large Data Bases (VLDB), Rio de Janeiro, Brazil, Aug. 2018
 - ◇ D. Xin, L. Ma, J. Liu, **S. Macke**, A. Parameswaran
Accelerating Human-in-the-loop Machine Learning With Helix (Demo)
44th Int. Conf. on Very Large Data Bases (VLDB), Rio de Janeiro, Brazil, Aug. 2018
 - ◇ A. El-Kishky, F. Xu, A. Zhang, **S. Macke**, J. Han
Entropy-Based Subword Mining for Word Embeddings
SCLeM Workshop at NAACL Conf. on Computational Linguistics, New Orleans, USA. June 2018
 - ◇ D. Xin, L. Ma, J. Liu, **S. Macke**, S. Song, A. Parameswaran
Accelerating Human-in-the-loop Machine Learning: Challenges and Opportunities
DEEM Workshop at SIGMOD Int'l Conf. on Management of Data, Houston, USA. June 2018
 - ◇ T. Wattanawaroon, **S. Macke**, A. Parameswaran
Towards a Theory of Data-Diff: Optimal Synthesis of Succinct Data Modification Scripts
ArXiv preprint.

[†]he/him/his

^{*}Equal contribution

- Experience*
- ◇ **Graduate Researcher**, University of Illinois at Urbana-Champaign · 2015—
 - ◇ **Visiting Student Researcher**, University of California, Berkeley · 2019—2020
 - ◇ **Software Engineering Intern**, Facebook · Summer 2019
 - Mentored by Peng Wang on the Dangerous Content ML team.
 - Developed a Bayesian label propagation algorithm that resulted in more than **50,000** deletes of terrorism-related posts (as of October 3, 2020). When this semisupervised method was used to compute derived features for Dangerous Content's post classifier, such features ranked in the **top 3** in terms of importance out of more than 100 features, with an estimated increase in confident deletes by **20%**.
 - Expanded the counter-terrorism label propagation pipeline to other problem areas such as hate organizations.
 - Provided extensive documentation for tuning the label propagation pipeline and porting to other problem areas.
 - Demonstrated other miscellaneous classifier improvements besides those resulting from my primary project.
 - ◇ **Software Engineering Intern**, Google Brain · Summer 2018
 - ◇ **Software Engineering Intern**, Alation · Summer 2016
 - ◇ **Software Engineering Intern**, Palantir Technologies · Summer 2014
 - ◇ **Software Engineering Intern**, Palantir Technologies · Summer 2013
 - ◇ **NSF REU Researcher**, University of Illinois at Urbana-Champaign · Summer 2012
- Teaching*
- ◇ **Teaching or Course Assistant for the following courses:**
 - UIUC CS589-CCC (Cloud Computing Capstone) · Spring 2019
 - Stanford CS149 (Parallel Computing) · Winter 2013 and Winter 2014
 - Stanford CS103 (Mathematical Foundations of Computing) · Fall 2013 and Fall 2014
 - Stanford CS101 (Introduction to Computing Principles) · Spring 2014
- Software*
- ◇ **Safety-First Jupyter Kernel (nbsafety)**
 - Drop-in replacement for Jupyter's Python 3 kernel that leverages implicit lineage metadata present in notebook state to highlight cells that are unsafe to execute due to *staleness*. Also highlights cells that resolve such staleness issues, helping reduce cognitive overhead in managing notebook state.
 - ◇ **Automatic Subtitle Synchronizer (ffsubsync)**
 - Project synchronizes subtitles to video using voice audio detection to detect and fix constant drift.
 - **Honorable Mention** at HackIllinois 2019 (among the top 5 submissions that did not win company awards).
 - ◇ **Java Direct I/O (jaydio)**
 - Project provides direct I/O functionality for Java (bypassing FS cache)
 - Development funded by a Sourcegraph Open Source Fellowship (\$1000 stipend)
- Service*
- Organizer for UC Berkeley's Spring 2020 Database Seminar
 - External reviewer for TKDD, SDM, SIGMOD Record
 - Organized reading group for my lab to discuss recent DB research papers
 - CS grad ambassador (host for students admitted to UIUC's CS PhD program) in 2018
 - Served as mentor to undergraduate and master's students
- Honors & Awards*
- 2019 **Awardee**, State Farm Companies Foundation Doctoral Scholarship
 - 2019 **Honorable Mention**, HackIllinois open source hackathon
 - 2015 **Awardee**, Andrew and Shana Laursen Fellowship
 - 2015 **Awardee**, Diffenbaugh Graduate Fellowship
 - 2014 **Awardee**, NSF Graduate Research Fellowship*
 - 2014 **Awardee**, Sourcegraph Open Source Fellowship
 - 2013 **1st Place**, DWR Governor's Cup Business Plan Competition, Oklahoma division (\$22,000 prize)
 - 2013 **Top-150 Score**, Putnam Competition (exact rank: 142.5, highest in Oklahoma/Arkansas)*
 - 2012 **Finalist**, SignalFire University Hacker Olympics in San Francisco
 - 2012 **Awardee**, Goldwater Scholarship*
 - 2012 **World Finalist**, ACM ICPC World Finals in Warsaw, Poland
 - 2009 **Awardee**, University of Tulsa Presidential Scholarship (covering all tuition and living expenses)

*Nationally competitive