

Ankur Taly

Email: ankur.taly@gmail.com

Phone: (732) 513-5690

Homepage: <http://theory.stanford.edu/~ataly>

Summary

I am a computer science researcher with a broad set of interests spanning **machine learning**, **computer security**, **programming languages**, and **formal methods**. My current work at Fiddler labs (2019 — present) is focussed on making explainable AI practical, while previous research within the Google Brain team (2016 — present) focussed on analysis of deep neural networks. Prior research at Google (2012 — 2016) was on authorization mechanisms for distributed systems. Dissertation research at Stanford University (2007 — 2012) was on language security and program verification. I have published papers in top-tier venues of each of these fields.

Employment

- **Fiddler Labs, Mountain View, CA, USA**

May 2019 — present

Head of Data Science

My responsibilities include:

- Overseeing all data science efforts at Fiddler labs, and managing the data science team (3 FTEs + interns).
- Driving research on pushing the state of the art on explainable AI (**refer Publications [5, 27]**).
- Serving as the technical expert in all sales conversations, and helping with setting up product strategy and roadmap.
- Evangelizing explainable AI technology across industry and academia
 - * I co-organized tutorials on explainable AI at several top-tier academic conferences, presented Fiddler’s explainable AI technology at 10+ companies (across healthcare, finance and technology), and delivered guest lectures in several graduate courses on machine learning.

- **Google Inc., Mountain View, CA, USA**

Aug. 2012 — May 2019

Staff Research Scientist

Google Brain (May 2016 — May 2019)

- *Interpreting deep neural networks (DNNs)*. My research spanned the following objectives — (1) Increasing transparency of DNNs (2) Evaluating robustness of DNNs, and (3) Extracting human-intelligible rules from DNNs. (**refer Publications [1, 2, 6, 7, 8, 9, 10, 11, 28]**)
 - * I co-developed a method called “**Integrated Gradients**” for attributing a DNN’s predictions to its input features. For instance, in an object recognition network, it could tell us which pixels of the image were responsible for a certain label being picked.
 - * The method is used by **20+ product teams** inside Google. Prominent application include (1) generating explanations for Verily’s Diabetic retinopathy screening network to assist doctors, (2) detecting bias in text sentiment networks used by product teams (e.g., Youtube), (3) debugging feature importances in predictive targeting models for content ads.
- *Robust question-answering*. My research in this area includes — (1) adversarial analysis of state-of-the-art question-answering networks (**refer Publications [9, 29]**), and (2) techniques for blending machine learning inside a traditional semantic parser to improve its recall. (**refer Publication [30]**)
- *ML fairness*. Developed techniques for assessing and addressing fairness concerns in text classification models. (**refer Publication [7]**)

Senior Research Scientist**Security Research** (Aug. 2012 — May 2016)

- Designed and implemented the security model for the *Vanadium* application framework; see <https://v.io>. The model supports fully decentralized identities, mutual authentication and authorization, fine-grained delegation, and end-to-end encryption. (refer **Publications [3, 12, 13]**, **Patents [2, 3, 4]**)
- Developed the core technology for *Macaroons* — a flexible authorization credential for decentralized and controlled delegation of authority among principals in the Web, Cloud and other distributed systems. Macaroons have seen widespread adoption inside and outside Google with open-source implementations in 9 different languages; see <http://macaroons.io> (refer **Publication [14]**, **Patent [1]**)

Education• **Stanford University****Ph.D. in Computer Science**

Jun. 2012

Thesis title: “Sandboxing Untrusted JavaScript”Advisor: Prof. John C. Mitchell**Google Ph.D. Fellow****M.S. in Computer Science**

Jun. 2010

• **Indian Institute of Technology (IIT), Bombay****B. Tech in Computer Science and Engineering**

May 2007

Selected Awards and Honors

- **Outstanding paper award** for Publication [4], 21st European Symposium on Research in Computer Security (ESORICS), Sep. 2016
- 3rd **prize** for Publication [2], **AT&T Best Applied Security Paper Award** competition, Nov. 2011.
- **Google PhD Fellowship in Language Security (2010-2012)**, Jun. 2010.
- Selected for **best papers from VMCAI 2009**, 10th Int’l conference on Verification, Model checking and Abstract Interpretation, for Publication [10], Dec. 2009.
- **Stanford Computer Forum Fellowship**, Sep. 2007.
- **All India Rank 69** out of 180,000 students, IIT joint entrance examination, Jun. 2003.
- **Gold Medal**, Indian National Physics Olympiad, May 2003.
- **Gold Medal**, Indian National Chemistry Olympiad, May 2003.

Professional Activities

- Co-organized tutorials on “Explainable AI in industry” at **KDD 2019**, **FAT* 2020**, and **WWW 2020**.
- **Teaching and Mentoring**:
 - Taught a **short course** on “Distributed Authorization” at the International **summer school** on Foundations of Security, Analysis, and Design (FOSAD), held at Bertinoro, Italy in Sep. 2016.
 - **Guest lectures in graduate courses**:

- * ECE 739: “Security and Fairness of Deep Learning”, CMU Silicon Valley, USA Feb. 2020
 - * ECE 737: “Engineering Safe Software Systems”, CMU Silicon Valley, USA, Oct. 2019
 - * CS 223: “Advanced Computer Security”, UC Santa Cruz, USA Oct. 2012
 - * CS 258: “Programming Language Theory”, Stanford University, USA Mar. 2009
 - * CS 242: “Programming Languages”, Stanford University, USA Oct. 2008
- **Students mentored:** Sahaj Garg (Stanford University), Pramod Kaushik Mudrakarta (The University of Chicago), Siddhartha Jayanti (Princeton University), Andres Erbsen (MIT).
- **Google Research outreach:**
 - Reviewer for Faculty Research Award (FRA) and Google PhD fellowship proposals. Official FRA liaison for Prof. Stephen Chong and Prof. David Evans.
 - Panelist, 2014 Google Bay Area PhD Summit.
 - **Program Committees:** ACM PLDI (ERC) 2019, ACM PLDI (ERC) 2014, ETAPS POST 2014, ACM PLAS 2013, HOTSPOT 2013.

Research Themes

- **Explaining Machine Learning models [EML]:** Explaining the prediction logic of ML models.
- **Distributed Authorization [DA]:** Identity and Authorization mechanisms for highly distributed services and applications.
- **Language Security [LS]:** Mechanisms for securing untrusted code by restricting the language in which it is written.
- **Verification and Synthesis [VS]:** Techniques for verifying and synthesizing hardware, software, and hybrid systems.

Publications

Journals and Book Chapters:

1. [EML] Kevin McCloskey, Ankur Taly, Federico Monti, Michael P. Brenner, Lucy Colwell — “Using Attribution to Decode Dataset Bias in Neural Network Models for Chemistry”, In: *Proceedings of National Academy of Science (PNAS)* Preprint, 2019.
2. [EML] Rory Sayres, Ankur Taly, Ehsan Rahimy, Katy Blumer, David Coz, Naama Hammel, Jonathan Krause, Arunachalam Narayanaswamy, Zahra Rastegar, Derek Wu, Shawn Xu, Scott Barb, Anthony Joseph, Michael Shumski, Jessie Smith, Arjun B. Sood, Greg S. Carrado, Lily Peng, Dale R. Webster — “Using a Deep Learning Algorithm and Integrated Gradients Explanation to Assist Grading for Diabetic Retinopathy”, In: *American Academy of Ophthalmology*, 2018.
3. [DA] Ankur Taly, Asim Shankar—“Distributed Authorization in Vanadium”, Book Chapter in: *Lecture Notes on Foundations of Security, Analysis and Design (FOSAD)*, Springer, 2016.
4. [VS] Ankur Taly, Sumit Gulwani, Ashish Tiwari — “Synthesizing Switching Logic using Constraint Solving”, In: *International Journal on Software Tools for Technology Transfer (STTT)*, Springer, 2011.

Conferences and Workshops:

5. [EML] Umang Bhatt, Alice Xiang, Shubham Sharma, Adrian Weller, Ankur Taly, Yunhan Jia, Joydeep Ghosh, Ruchir Puri, Josãl M. F. Moura, Peter Eckersley — “Explainable Machine Learning in Deployment”, In: *ACM Fairness, Accountability and Transparency (FAT*)*, 2020.
6. [EML, VS] Divya Gopinath, Hayes Converse, Corina Pasareanu, Ankur Taly — “Finding Contracts in Deep Neural Networks”, In: *Automated Software Engineering (ASE)*, 2019.

7. [EML] Sahaj Garg, Vincent Perot, Nicole Limtiaco, Ankur Taly, Ed H. Chi, Alex Beutel — “Counterfactual Fairness in Text Classification through Robustness”, In: *ACM Artificial Intelligence, Ethics, and Society (AIES)*, 2019.
8. [EML] Mukund Sundararajan, Jinhua Xu, Ankur Taly, Rory Sayres, Amir Najmi — “Exploring Principled Visualization of Deep Network Attributions”, In: *IUI Workshop on Explainable Smart Systems (ExSS)*, 2019.
9. [EML] Pramod Kaushik Mudrakarta, Ankur Taly, Mukund Sundararajan, Kedar Dhamdhere — “Did the model understand the question?”, In: *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2018.
10. [EML] Rory Sayres, Ankur Taly, Ehsan Rahimy, Katy Blumer, David Coz, Naama Hammel, Jonathan Krause, Arunachalam Narayanaswamy, Zahra Rastegar, Derek Wu, Shawn Xu, Lily Peng, Dale R. Webster — “Assisted reads for diabetic retinopathy using a deep learning algorithm and integrated gradient explanation” (extended abstract), *Annual meeting of the Association for Research in Vision and Ophthalmology (ARVO)*, 2018.
11. [EML] Mukund Sundararajan, Ankur Taly, Qiqi Yan — “Axiomatic Attribution for Deep Networks”, In: *International Conference on Machine Learning (ICML)*, 2017.
12. [DA] David Wu, Ankur Taly, Asim Shankar, Dan Boneh — “Privacy, Discovery and Authentication for Internet of Things”, In: *European Symposium on Research in Computer Security (ESORICS)*, 2016 (**award paper**).
13. [DA] Martin Abadi, Mike Burrows, Himabindu Pucha, Adam Sadovsky, Asim Shankar, Ankur Taly — “Distributed Authorization With Distributed Grammars”, In: *Programming Languages with Applications to Biology and Security (PLABS)*, 2015.
14. [DA] Arnar Birgisson, Joe Politz, Ulfar Erlingsson, Ankur Taly, Michael Vrable, Mark Lentzner — “Macarons: Cookies with Contextual Caveats for Decentralized Authorization in the Cloud”, In: *Network and Distributed System Security Symposium (NDSS)*, 2014.
15. [VS] Patrice Godefroid, Ankur Taly — “Automated Synthesis of Symbolic Instruction Encodings from I/O Samples”, In: *ACM Programming Language Design and Implementation (PLDI)*, 2012.
16. [LS, VS] Ankur Taly, Ulfar Erlingsson, John C. Mitchell, Mark S. Miller, Jasvir Nagra — “Automated Analysis of Security-Critical JavaScript APIs”, In: *IEEE Symposium on Security and Privacy (S&P)*, 2011 (**award paper**).
17. [VS] Ankur Taly, Ashish Tiwari — “Switching Logic Synthesis for Reachability”, In: *ACM International Conference on Embedded Software (EMSOFT)*, 2010.
18. [LS] Sergio Maffei, John C. Mitchell, Ankur Taly — “Object Capabilities and Isolation of Untrusted Web Applications”, In: *IEEE Symposium on Security and Privacy (S&P)*, 2010.
19. [VS] Ankur Taly, Ashish Tiwari - “Deductive Verification of Continuous Dynamical Systems”, In: *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FST&TCS)*, 2009.
20. [LS] Sergio Maffei, John C. Mitchell, Ankur Taly — “Isolating JavaScript with Filters, Rewriting, and Wrappers”, In: *European Symposium on Research in Computer Security (ESORICS)*, 2009.
21. [LS] Sergio Maffei, Ankur Taly - “Language-Based Isolation of Untrusted JavaScript”, In: *IEEE Symposium on Computer Security Foundations (CSF)*, 2009.
22. [LS] Sergio Maffei, John C. Mitchell, Ankur Taly — “Run-Time Enforcement of Secure JavaScript Subsets”, *Web 2.0 Security and Privacy (W2SP)* workshop, 2009.
23. [VS] Ankur Taly, Sumit Gulwani, Ashish Tiwari — “Synthesizing Switching Logic using Constraint Solving”, In: *International Conference on Verification, Model Checking and Abstract Interpretation (VMCAI)*, 2009 (**selected as one of the best papers**).

24. [LS] Sergio Maffei, John C. Mitchell, Ankur Taly — “An Operational Semantics for JavaScript”, In: *Asian Programming Languages Symposium (APLAS)*, 2008.
25. [VS] Stephane Gaubert, Eric Goubault, Ankur Taly, Sarah Zennou — “Static Analysis by Policy Iteration on Relational domains”, In: *European Symposium on Programming (ESOP)*, 2007.
26. [VS] Sudeep Juvekar, Ankur Taly, Varun Kanade, Supratik Chakraborty — “Efficient Symbolic Reachability of Networks of Transition Systems”, In: *General Motors Workshop on Next Generation Design and Verification Methodologies for Distributed Embedded Control Systems*, 2007.

Preprints and Technical Reports:

27. [EML] Luke Merrick, Ankur Taly — “The Explanation Game: Explaining Machine Learning Models using Shapley Values”, Technical Report, on arxiv, 2020
28. [EML] Mukund Sundararajan, Ankur Taly — “A Note about: Local Explanation Methods for Deep Neural Networks lack Sensitivity to Parameter Values”, Technical Report, on arxiv, 2018.
29. [EML] Pramod Kaushik Mudrakarta, Ankur Taly, Mukund Sundararajan, Kedar Dhamdhere — “It was the training data pruning tool!”, Technical report, on arxiv, 2018.
30. [EML] Kedar Dhamdhere, Kevin Mccurley, Mukund Sundararajan, Ankur Taly — “Abductive Matching in Question Answering”, Technical report, on arxiv, 2017.
31. [LS] Ankur Taly - “Separation Logic and Mashup Isolation”, Technical report, Stanford University, 2010.

Theses:

32. Ankur Taly - “Sandboxing Untrusted JavaScript”, Doctoral thesis, Stanford University, 2013.
33. Ankur Taly - “Efficient Guided Symbolic Reachability Analysis”, Bachelor’s thesis, IIT Bombay, 2007.
34. Ankur Taly - “Automata on Infinite Inputs”, Junior thesis, IIT Bombay, 2006.

Patents

1. Michael Burrows, Himabindu Pucha, Raja Daoud, Jatin Lodhia, Ankur Taly — “Signatures Of Updates Exchanged In A Binary Data Synchronization Protocol”, 2017.
2. Martin Abadi, Mike Burrows, Himabindu Pucha, Adam Sadovsky, Asim Shankar, Ankur Taly — “Authorization in a Distributed System Using Access Control Lists and Groups”, 2017.
3. Ankur Taly, Asim Shankar, Gautham Thambidorai, David Presotto — “Security model for identification and authentication in encrypted communications using delegate certificate chain bound to third party key”, 2016.
4. Ulfar Erlingsson, Ankur Taly, Michael Vrable, Mark Lentzner - “Macaroons: Methods and Systems of Generating and Using Authentication Credentials for Decentralized Authorization in the Cloud”, 2016.

Selected Invited Talks

- Intuit Tech Con, Mountain View, USA Sep. 2019
- O’Reilly Artificial Intelligence Conference, San Jose, USA Sep. 2019
- Fiddler Explainable AI Summit (keynote), San Francisco, USA Aug. 2019
- GFMI Model Risk Conference, New York, USA Jun. 2019
- DREAM seminar, UC Berkeley, Berkeley, USA Feb. 2019

- CSL seminar, SRI International, Menlo Park, USA Dec. 2018
- Samsung AI Research, Mountain View, USA Sep. 2018
- Dagstuhl Workshop on Machine Learning and Formal Methods, Schloss Dagstuhl, Germany Aug. 2017
- CSL seminar, SRI International, Menlo Park, USA Jun. 2017
- Bhabha Atomic Research Center, Mumbai, India Dec. 2015
- Keybase.io, San Francisco, USA Aug. 2015
- Vail Computer Elements Workshop, Vail, USA Jun. 2015
- CS Colloquium, Microsoft Research Silicon Valley, Mountain View, USA Mar. 2012
- Programming Languages group, Adobe Systems, San Jose, USA Feb. 2012
- Dagstuhl Workshop on Foundations of Scripting Languages, Schloss Dagstuhl, Germany Jan. 2012
- Computer Security Awareness Week (CSAW), NYU-Polytechnique, Brooklyn, USA Nov. 2011
- Logic and Semantics group, Queen Mary College, London, UK Feb. 2011
- Software Validation group, Fujitsu Labs America, Sunnyvale, USA Feb. 2010

Research Internships

- **Research Intern, Microsoft Research, Redmond, USA** Jun.—Sep. 2011
Mentor: Dr. Patrice Godefroid
- **Research Intern, Google Inc., Mountain View, USA** Jun.—Sep. 2010
Mentors: Dr. Jasvir Nagra, Dr. Ulfar Erlingsson
- **Research Intern, SRI International, Menlo Park, USA** Jun.—Sep. 2008
Mentor: Dr. Ashish Tiwari
- **Research Intern, Ecole Polytechnique, Paris, France** May—Jul. 2005 and May—Jul. 2006
Mentors: Prof. Eric Goubault, Prof. Stephane Gaubert

Programming Skills

- **Programming Languages:** Python, Go, JavaScript, C/C++.
- **Tools:** TensorFlow (Deep learning framework), ProVerif (Theorem prover), Z3 (SMT solver).

Activities

- Attended Dagstuhl workshop on Foundation of Scripting Languages, at Schloss Dagstuhl, Germany, 2012.
- Finalist at the 2011 CSAW Applied Security Quiz competition.
- Teaching Assistant at Stanford University for the courses: CS155: “Computer and Network Security” (Mar.— May 2011), CS242: “Programming Languages” (Sep. — Nov. 2008)
- Coordinated the Stanford security seminar from Sep. 2008 to Jun. 2012.