

Curtis G. Northcutt

Mobile: 859-559-5716 • Email: cg@mit.edu

Address: Massachusetts Institute of Technology, 32 Vassar St., Bldg 26-209, Cambridge, MA 02139

EDUCATION:

Massachusetts Institute of Technology (MIT)

M.S. in Electrical Engineering and Computer Science (EECS) (June, 2017)

Ph.D. in Electrical Engineering and Computer Science (EECS) (2013-ongoing)

[National Science Foundation GRFP Fellow](#) & [MITx Research Fellow](#)

Winner of [2016 MIT Master's Thesis Morris Joseph Levin Award](#)

Vanderbilt University - GPA: 4.00/4.00 (Valedictorian)

B.S. in Honors Computer Science and Mathematics (May 2013)

[Founder's Medalist for School of Engineering](#)

Areas of Specialization:

- Machine Learning
 - Semi-supervised learning
 - Robust learning (noisy labels)
- Data Analytics
- MOOCs / Cheating Detection
- Algorithms
- Cybersecurity (Detection)
- Social Networks

National Scholarships:

- [2012 Barry M. Goldwater Scholarship](#) (\$7,500)

National Fellowships:

- [2013 NSF GRFP Fellowship](#) (\$126,000)
- [USAID Research Innovation Fellowship](#) (\$9,000)

Grants:

- [Centre for Big Data on EdTech at Beijing Normal University Collaborative Research](#) (\$15,000)

Academic Programs:

- [LearnLab Educational Data-Mining Track at Carnegie Mellon University](#) (2015)
- [MIT-Imperial Global Fellows Program](#) (2015)
- [MIT EECS Start6 Entrepreneurship](#) (2014)
- [NSF USAID Research Innovation Fellow](#) (2014)
- [James Geddes Memorial Scholarship](#) (\$28,000)
- Jesse Taylor, Jr. Scholarship (\$90,000)
- Judi Hale Memorial Scholarship (\$500)
- [National Science Foundation \(NSF\) Research Experience for Undergraduates \(REU\)](#) (2011)
- [NASA INSPIRE & OLC](#) (2009)
- [Governor's Scholars](#) (2008)
- UK BEST Engineering Program (2007)

PROFESSIONAL EXPERIENCE (RESEARCH & INDUSTRY):

Massachusetts Institute of Technology (MIT), Office of Digital Learning (ODL), edX, MITx

Cambridge, MA | Sept. 2013-(ongoing)

Machine Learning and Data-Mining, Ph.D. Candidate in Computer Science, Advisor: Dr. Isaac Chuang

Two Point Summary: (1) Algorithm/Design and (2) High Impact Implementation

- Develop new machine learning algorithms to infer user-intent from noisy, large interaction datasets.
- Implement inference & detection algorithms across 300+ MITx and HarvardX MOOC courses.

Areas of Machine Learning Research (Focus: semi-supervised learning):

- Learning with noisy training labels (1. Flipped labels or 2. added noise drawn from other distribution).
- Deep Learning, embeddings, and NLP for ASR, keyword-spotting, and comment re-ranking algorithms.
- **News!** Our algorithm Rank Pruning ([Paper](#), [Github](#)) is **state-of-the-art** for classification with noisy labels.

Areas of MOOC (edX) Research:

- Automated Community Detection, Student Social Network Inference
- Practical Cheating Detection in MOOCs and Sybil account / bot detection.
- Quantifying MOOC certification value globally to address the goal of "Democratization of Education."

Previous and Other Ongoing MOOC Research:

- Developed a generalized, distributed, automated gaming and cheating detection algorithm for MOOCs.
- Developed a fingerprint generation algorithm to uniquely identify students in MOOCs.
- Invented the [CAMEO Detection Algorithm](#) to detect “Multiple Account” Cheating in online courses.

Amazon Alexa Machine Learning Team

Cambridge, MA | May 2017-August 2017

Research Scientist Intern (semi-supervised learning | ASR | deep learning), Manager: Dr. Ming Sun

- Pitched and developing a new wake-up (“Alexa”) model with potential impact for millions of customers and millions of dollars saved using [Rank Pruning](#). Still in progress. Details omitted for proprietary reasons.

Facebook Artificial Intelligence Research (FAIR)

Manhattan, New York City, NY | May 2016-September 2016

Research Intern (Machine Learning /Natural Language Processing), Mentor: Dr. Y-Lan Boureau

- Developed and experimented with comment ranking diversification algorithms.
- Added ~800 lines of PHP for online linear algebra / NLP in the backend of [facebook.com](#).
- Technical skills learned: PHP, HACK, Mercurial source control, HIVE SQL, Presto SQL, FB Learner.
- NLP / ML techniques used: Starspace/WSABIE, Deepwalk, Word2Vec, TF-IFD, TSNE, etc.

Microsoft Research (MSR) India, Technologies for Emerging Markets Group

Bangalore, Karnataka, India | June 2014-July 2014

Visiting Scientist on Massively Employed Classroom (MEC) project, Mentor: Dr. Bill Thies

- Built an extensible keylogger to store all input events in MSR India’s MOOC, MEC.
- Invented Codelogger – tool to replay online programming assessments.
- Studied JavaScript key-logging: event handling, capturing and delegation, ACE Editor.

Founder of Reverse Definition Start-up, RevDef™: Connecting People to the Words They Need

Cambridge, MA | January 2014-(ongoing)

Co-founder with Jonas Mueller and Albert Kim

- Used to find colloquial/scientific terms you forget, choosing titles via descriptions, and semantic compression.
- Outperforms Google Search by 250% on queries in search of a single term via NLP semantic analysis.
- Currently not publically accessible, awaiting legal consultation and an updated development phase.

Founder of MIT4Harvard™: MIT Math & Computer Science Expert Tutors for Harvard Students

Cambridge, MA | October 2014-March 2016

CEO, founder, and tutor for a computer science and mathematics tutoring company

- Connects MIT experts to Harvard students requiring tutoring services in Math & Computer Science.
- Facilitated 15+ MIT students as expert Computer Science tutors with premium price ranges and services.
- Personally tutored 2 Harvard students over 2+ years (over 50+ non-Harvard students in lifetime).

Massachusetts Institute of Technology (MIT) Lincoln Laboratory, U.S. Department of Defense

Lexington, MA | June 2013-August 2013

Network Intrusion Detection, Cybersecurity Research, P.I. Joshua W. Haines

- Developed a data-generation tool to characterize performance of a DoD network intrusion detection system.
- Identified errors in the deployed system; Built highly modularized tool for extensibility of new system versions.
- Trained in security networking, Python, Networkx, JSON modules, Apache Tomcat, online algorithms, SQL.

Coursera Researcher with Vanderbilt Mathematic Department

Nashville, TN | January 2013-May 2013

Educational Data-mining and Clustering Analysis Research, Dr. Derek Bruff

- Used Hierarchical Agglomerative Clustering (HAC) to classify student types within the Coursera platform.
- Performed big-data feature extraction of Vanderbilt’s Coursera data, resulting in a research paper.

- Trained in SQL, MATLAB, Weka, HAC, SBAC, COBWEB, EM, clustering analysis.

Vanderbilt University EECS Undergraduate Research 2013

Nashville, TN | January 2013-May 2013

Educational Data-mining and Intelligent Learning Environments Research, Dr. Gautam Biswas

- Designed the Indirectly Collaborative LE (ICLE): crowd-sourcing & narrative to teach mathematical induction.
- Developed a model for CBLEs to teach inductive proofs without natural language processing (NLP).
- Studied intelligent tutoring systems (ITS), open-ended learning, cognitive tools and tutors, educational games.

Vanderbilt University EECS Undergraduate Research 2012

Nashville, TN | August 2012-December 2012

Cyber-Physical Security (CPS) Systems AI Research, Dr. Xenofon Koutsoukos

- Described the relationships of CPS against general cyber security to suggest the value of CPS protocols.
- Combined Bayesian Modelling, Hybrid Automata, and Digital Logic for a generalized security model for CPS.
- Published “*Generalized Algorithm for CPS Intrusion Detection and Security Robustness via Logical Truth Tables.*”

Microsoft Corporation, Windows Phone Division,

Redmond, WA | May 2012-August 2012

Technical Program Manager Intern for Windows Phone, Camera and Photos Team, Dr. Eric Bennett

- Created, designed, and specified full-stack implementations of all (3) Microsoft’s WP8 third-party camera apps.
- Invented FX Suite, PhotoStrip, Photofuse Jr. apps receiving worldwide acclaim in [CNet](#), [Wired](#), and [other media](#)
- Managed 3 PhD and 2 undergrad interns, successfully demonstrating all WP8 APIs for ISVs by our deadlines.
- Selected as a Microsoft Ambassador (50 of 1800 interns selected) and received a full-time job offer.

National Science Foundation REU at University of Notre Dame

Notre Dame, IN | May 2011-August 2011

Experimental Research on Wireless Networking (ERWiN) NSF REU in Computer Science, Dr. A. Striegel

- Developed a C++ application to detect a user’s heart rate via webcam, exposing fallacies in previous research.
- Received Second Place in the NSF REU Poster Competition of the 2011 ERWiN REU for my research.
- Learned C++, OpenCV, C#, SSH clients, Digital Signal Processing, Lomb-Scargle Fourier Analysis.

General Electric (GE) Appliances

Louisville, KY | May 2010-August 2010

Engineering Intern, Database Construction and Website Design and Development

- Designed and implemented a website training database for new employees, used by GE for years afterward.
- Recognized by Gen Manager R. Simpson of GE Sourcing & won world 2nd place in GE’s REEL Global Contest.
- Learned website design/coding/programming, databases, manufacturing, management, and corporate business.

National Aeronautics and Space Administration (NASA) and ViGYAN Inc.

Langley Base, Newport News, VA | June 2009-August 2009

NASA Inspire Internship, Aerospace Research, Mentor: Norm Crabill (VA Aviation Hall of Fame)

- Conducted aerospace engineering and oscillatory dynamics research.
- Analyzed aircraft implementation ARIS (wind gust mitigation of 90%) and reduced exterior oscillations.
- Trained in JAVAfoil, MATLAB, JAVA, wind-tunnel research, harmonic motion, and vector spaces.

University of Kentucky, Department of Physiology and Neurobiology

Lexington, KY | June 2008-June 2009

Neurobiology Research on Glutamatergic Synaptic Transmission, Dr. Robin Cooper

- Studied the effects of residual intracellular calcium alterations on glutamatergic synaptic transmission.
- Presented at Intel-International Science and Engineering Fair (world); awarded Overall First-Place in Kentucky.
- Trained in dissections, pH solutions, staining cells, cell stimulation, focal electrodes, synaptic transmission, etc.

TEACHING:

TEACHING:

- **Teaching Assistant (TA) for 6.867 Graduate Machine Learning at MIT** – Responsibilities include teaching weekly recitation, grading, planning, developing examination questions, and office hours (Fall, 2017)
- **Taught MIT IAP course on Cool Math Proofs** – Topics included Fermat's Last Theorem, Derangements, Handshaking, 4 color theorem, Banach Fixed Point Theorem, Infinite, Euler's Formula, GCD (Euclidean Algorithm), heuristics (2014, 2015)
- **Taught (2) MIT IAP one-day courses in Discrete Math** – Topics included Induction, Deduction Proofs, Contradiction, Graph Theory, Graph Search Algorithms (BFS, DFS, Dijkstra's, A*), Google maps, Minimum Spanning Tree (Kruskall's, Prims), 6 Color Theorem / 4 Color Theorem, Inclusion-Exclusion Principle, Hats (Derangements) Problem, Alpha Beta Pruning on Trees (2014, 2015)
- **Vanderbilt VUceptor** – Co-Instructor with a faculty member of seminar course for first-year students (2010)
- **Vanderbilt Students Volunteering for Science** – Taught hands-on science experiments in poor schools(2010)

TUTORING:

- **Founded MIT4Harvard™ Tutoring** – MIT math & CS expert tutors for Harvard students (2013-ongoing)
- **Mathematics Tutoring with Prof. Rittle-Johnson** – Studied cognitive theory of “*How Children Learn Math*,” and tutored struggling students in Nashville Metro Public Schools (2013)
- **Mathematics Tutor** – Created for-profit STEM, CS, ACT/SAT tutoring business (2008-ongoing)
- **Employee of Vanderbilt Child and Family Center** – Taught simple tasks to toddlers (2009)

GUEST LECTURES:

- **Mathematics Guest Lecturer for Lafayette High School (annually)** – “*Discrete Mathematics: How Mathematicians Formally Prove Things*” - Taught induction, contradiction, and logical deduction. (2012, 2013)
- **CS Guest Lecturer for The Vanderbilt School for Science and Math** – “*How Google Maps works: Theory and Application*” – Taught algorithms to advanced high school students in Metro Nashville Public Schools (02/28/13)

RESEARCH ADVISOR:

- MIT EECS [SuperUROP](#) Undergraduate Advisor – [Ruth Park](#) (2015-2016)
- MIT EECS [UROP](#) Undergraduate Advisor – [Abhinav Venigalla](#) (2017)

PUBLICATIONS:

Available [here](#).

Upcoming (accepted as a conference publication at UAI 2017):

Northcutt, Curtis G., Wu, Tailin, Chuang, Isaac L. (2017). [Learning with Confident Examples: Rank Pruning for Robust Classification with Noisy Labels](#). *Uncertainty in Artificial Intelligence (UAI)*, August 2017. [[arXiv](#)]

RESEARCH, TALKS, & CONFERENCE PRESENTATIONS:

1. **August 12, 2017. [invited] Learning from Confident Examples: Rank Pruning for Robust Classification with Noisy Labels.** Poster (brief 1-minute talk). Uncertainty in Artificial Intelligence Conference (UAI), Sydney, Australia.
2. **June 7, 2017. Rank Pruning for Classification with Noisy Labels.** Talk. Amazon Alexa Machine Learning Wake-word Brown Bag, Cambridge, MA.
3. **May 8, 2017. Classification with Significantly Mislabeled Training Data.** Talk. ML Tea – MIT’s Machine Learning Seminar, Massachusetts Institute of Technology, Cambridge, MA.
4. **April 30, 2017. [invited] Measuring Assessment Authenticity in Open Online Learning.** Talk. 2017 National Conference on Measurement in Education (NCME 2017), San Antonio, TX.
5. **April 22, 2017. [invited] The Challenges of Openness versus Value in Massive Open Online Courses.** Talk. Emerging Challenges in Digital Higher Education, Harvard University, Cambridge, MA.
6. **April 20, 2017. [invited] Comment Ranking Diversification in Forum Discussions.** Poster. 2017 ACM Learning@Scale (L@S) Conference, Cambridge, MA.
7. **April 18, 2017. Practical Cheating Detection in MOOCs using Classification with Significantly Mislabeled Training Data.** MIT EECS Masterworks 2017, Poster. Cambridge, MA.
8. **February 1, 2017. Multiple-Account Cheating Detection and Forum Diversification in Online Courses.** Poster. MIT Festival of Learning, Cambridge, MA.
9. **December 8, 2016. Comment Ranking Diversification in MOOC Forum Discussions.** Poster. Graduate Natural Language Processing (NLP) Final Presentations, Cambridge, MA
10. **September 25, 2016. [invited] Detection and prevention of cheating in MOOCs.** Talk. Speaker at AI With the Best Conference, Virtual Online Confernece.
11. **April 20, 2016. Practical Cheating Detection in Massive Open Online Courses.** MIT EECS Masterworks 2016, Poster, Cambridge, MA.
12. **March 23, 2016. [invited] Detection and prevention of “multiple account” cheating in MOOCs.** Lecture at Centre for Big Data on Technology-Mediated Education, Beijing Normal Univ, Beijing, China.
13. **January 20, 2016. Inference of prerequisite structure in online courses.** Office of Digital Learning MITx Fellows Research Group, NE35 Grand Central Station, Cambridge, MA.
14. **September 23, 2015. [invited] Detection and prevention of “multiple account” cheating in MOOCs.** Office of Digital Learning Brown Bag Lunch Talk, NE35 Grand Central Station, Cambridge, MA.
15. **July 7, 2015. Detection and prevention of “multiple account” cheating in MOOCs.** Private presentation to Dave Pritchard’s group at MIT. Massachusetts Institute of Technology, Cambridge, MA.
16. **May 18, 2015. [invited] Anomalous response patterns among MITx & HarvardX learners on edX.** Privately presented to Harvard/MIT faculty, including Deans & Vice Provosts. HarvardX, Cambridge, MA.
17. **January 15, 2015. [invited] How to start up a start-up: creating RevDef.** MIT Start6 Welcome Speaker, Blade, Boston, MA.
18. **March 6, 2015. IFACAs (Inhumanly Fast Adjacent Correct Answers): automated bot detection in MOOCs.** HarvardX, Cambridge, MA.
19. **January 15, 2015. [invited] How to start up a start-up: Creating RevDef.** MIT Start6 Welcome Speaker, Blade, Boston, MA.
20. **December 9, 2014. SemanticTextDB – When natural language processing (NLP) meets databases,** MIT Stata Center, Computer Science and Artificial Intelligence Lab (CSAIL), Cambridge, MA.
21. **September 19, 2014. [invited] Breaking the glass ceiling: transcending socioeconomic expectations through MOOCs.** Frontiers in Development Conference, Ronald Reagan Building, Washington D.C.

22. **August 4, 2013. Data generation for characterizing the performance of a DoD network intrusion detection monitoring system. 2 presentations.**
 - a. Group 51 Cybersecurity Summer Presentations, MIT Lincoln Laboratory, B Building, (Aug. 4, 2013)
 - b. Summer Student Presentation, MIT Lincoln Laboratory, S Building, Lexington, MA (Aug. 2, 2013)
23. **July 14, 2012. Lens construction to demonstrate Windows Phone 8 camera API.** Windows Phone 8 Intern Poster Competition, Microsoft Corporation, Building Studio, Redmond, WA.
24. **Aug. 5, 2011. Webcam-based heart rate detection via facial photoplethysmography. 2 presentations.**
 - a. Summer Undergraduate Research Symposium, University of Notre Dame (Aug. 5, 2011)
 - b. Notre Dame Collegiate Competition of the 2011 ERWiN NSF REU (Aug. 5, 2011) – 2nd place
25. **August 3, 2010. [invited] Tree-structured database for General Electric (GE) employee online training.** Presented to Richard Simpson, General Manager of Sourcing for GE, Louisville, KY.
26. **August 8, 2009. Wind gust mitigation with Active Ride Improvement Project (ARIS).** NASA INSPIRE Final Presentation at Langley Research Center (LRC), NASA LRC Auditorium, Hampton, VA.
27. **May 12, 2009. Kinetics of the vesicle fusion pore regarding the physiological function at the neuromuscular junction of crayfish.** Poster/Oral Presentation. 4 presentations.
 - a. Intel International Science and Engineering Fair, Reno, Nevada (May 12, 2009)
 - b. KY Science and Engineering Fair, Eastern KY University (April 04, 2009) – 1st place overall
 - c. Bluegrass Society for Neuroscience, Univ. of KY (March 18, 2009)
 - d. Central KY Regional Science and Eng. Fair, Univ. of KY (March 7, 2009) – 1st place category

AWARDS & SOCIETIES:

SCHOLARSHIPS:

- **2013 NSF GRFP Fellowship** (\$146,000) – premier graduate fellowship by the National Science Foundation (2013)
- **USAID Research and Innovation Fellowship** (\$9000) – awarded to aid MSR in developing India (2014)
- **2012 National Goldwater Scholarship** (\$7500) - premier undergraduate award in STEM (2012)
- **James Geddes Memorial Scholarship** (\$28,000) – selective academic scholarship; only 10 awarded (2011-2012)
- **Jesse Taylor, Jr. Scholarship** (\$90,000) – selective award for academic excellence and need (2010-2012)
- **Judi Hale Memorial Scholarship** (\$500) – awarded to one student for academic excellence (2009)

COMPETITION AWARDS:

- **Founder's Medalist, Vanderbilt University School of Engineering** – awarded for first honors, valedictorian of graduating class (2013)
- **Vanderbilt Arts & Science College Scholars Program** – highest academic program, 3% acceptance (2010)
- **First Place in 2011 ICPC ACM** – Local Programming Contest at Vanderbilt (2011)
- **Kentucky Colonel** – commissioned by governor of KY (highest civilian honor in KY) (2008)
- **Rensselaer Medalist** – Awarded to the top math student at Lafayette H.S. (2009)
- **Second Place** in Vanderbilt Commons Quiz Bowl (2010)
- **Second Place American Chemical Society Exam** – Kentucky (2009)

RESEARCH & INDUSTRY AWARDS:

- **MIT Master's Thesis Morris Joseph Levin Award** – MIT faculty-judged thesis award (2016)
- **First Place Overall** in Intel KY State Science and Eng. Fair (also first in local and regional) (2009)
- **REEL Intern Global Competition Finalist** – top 3 team in the world for GE recruitment video (2010)
- **Second Place** in the Collegiate Poster Competition of the ERWiN National Science Foundation REU (2011)
- **American Physiological Society Best Presentation Award** –glutamate synaptic transmission research (2009)

HONORARY SOCIETIES:

- Tau Beta Pi
- Pi Mu Epsilon
- Order of Engineer
- ACM Student Member

ACTIVITIES & LEADERSHIP:

- **MIT Graduate Residential Advisor** – Full-time GRA for MIT fraternity Nu Delta (2015-ongoing)
- **MIT EECS Start6 Entrepreneurial Program** – Founded RevDef, pitched to Andreessen Horowitz (2014)
- **2014-15 MIT EECS Visiting Committee Student Advisory Board** – Decided 2015 (EECS) changes needed
- **MIT EECS Basketball Team** – Integral part of MIT graduate EECS intramural team (2013-ongoing)
- **MIT EECS Graduate Student Association Athletic Chair** – Managed EECS grad intramurals (2014-2015)
- **MIT Outing Club Hiking Trip Leader** – Experienced outdoorsman for White Mountains Natl. Forest (2014-)
- **MIT Outing Club Winter Trip Leader** – Lead official 10-person subzero day expeditions for MIT (2015)
- **Certified in Wilderness First Aid** – Trained in CPR, anaphylaxis, assisted breathing, hypothermia, etc. (2014)
- **President of Tau Beta Pi, Vanderbilt Chapter** – Most selective Eng. Honor Society in U.S. (2012-2013)
- **Selection Committee Member for Ingalls Award for Excellence in Undergraduate Teaching** – Served Vanderbilt faculty award selection committee. Only 1 Engineering student selected (2013)
- **Microsoft Ambassador** – Mentored Vanderbilt students in CS and promoted values through events (2012-2013)
- **Vanderbilt Residential Adviser (RA)** – Advised 40 Vanderbilt students in Blakemore House (2012)
- **V-Squared Mentor for VU Engineering** – Mentor and unofficial advisor to first-year students (2011-2013)

LANGUAGE SKILLS:

- **Spanish Basic Proficiency**
- **Programming Languages** – SQL, Python, JavaScript, PHP, C++, LISP, Java, MATLAB, HTML, CSS, etc.
- **Experience** with C#, XNA, VHDL, MIPS, Ruby, Perl, R, LabVIEW, Prolog, Stata, etc.

OTHER PROJECTS:

- **[SemanticTextDB](#)** — PostgreSQL database with integrated NLP models trained online (2014)
 - database for document-storage/retrieval
 - automated curation and structure discovery
 - automatically-inferred latent features including semantics, topics, sentiment, eloquence, etc.
 - <https://github.com/jwmueller/SemanticTextDB>