

CONTACT  
INFORMATION

e-mail: [critch@eecs.acritch.com](mailto:critch@eecs.acritch.com)  
website: <http://acritch.com/>  
phone and mailing address available upon email request

---

## Summary

**Quantitative talent:**

- (2006) Ranked 5th in Canada on the Putnam mathematics competition during sophomore year.
- (2006) Completed B.Sc. (Honours) in Pure Mathematics in 20 months.
- (2013) UC Berkeley PhD in mathematics completed.
- (2013) Awarded a 3-year research faculty position at the NSF-funded Mathematical Biosciences Institute (deferred).
- (2014-2015) Employed as an algorithmic stock trader at Jane Street Group in NYC.
- (2015-2016) Employed as a research fellow at the Machine Intelligence Research Institute.
- (2017-present) Employed as a theoretical AI researcher at UC Berkeley.

**Technical skill:**

- Extensive coding experience in OCaml, VBA, Bash, and Macaulay2, and some experience with R and Python, for business, trading, and mathematical computing applications.
- Published thesis on applications of algebraic geometry to Hidden Markov models.
- Experience analyzing EEG data using Bayesian network models.
- Developed a tensor manipulation package for Macaulay2.

**Management and team project experience:**

- Founder and Fund Advisor to the Survival and Flourishing Fund ([survivalandflourishing.fund](http://survivalandflourishing.fund))
- Founder and Executive Director of the Berkeley Existential Risk Initiative ([existence.org](http://existence.org))
- Co-founded the Center for Applied Rationality (CFAR, <http://rationality.org/>), a non-profit which teaches workshops on improving decision-making by using recent advances in cognitive science, earning over \$500K in admissions revenue to date, now with 9 full-time employees.
- Co-developed Werewolves social computer game for International House, Berkeley.
- Co-developed Credence Calibration game for CFAR.

**Well-regarded as a speaker and consultant:**

- Invited to present at diverse venues, including the 2020 and 2019 Foresight AGI Strategy Meetings, 2019 Singapore Prime Minister's Office Foresight Conference, 2019 Partnership on AI All Partners' Meeting, 2019 Deepmind AI Safety Workshop in Iceland, 2017 UC Berkeley EECS Faculty retreat, 2017 Beneficial AI conference at Asilomar, 2016 Society for Risk Analysis annual meeting, 2016 Princeton Envision conference, 2016 and 2017 Effective Altruism Global summits, 2015 Oxford University Future of Humanity graduate lecture series, Harvard Effective Altruism 2014 speaker series, Thiel 20 Under 20 Summit in New York City, TEDxYouth@Tallinn in Tallinn, Estonia, the 2013 Effective Altruism Summit, the Carnegie Mellon statistics seminar, the Duke University algebraic geometry seminar, the NCSU symbolic computation seminar, McGill University causal inference seminar, and the Founders Fund 2012 annual business meeting in Maui, Hawaii.
- Consulted on decision processes and training for numerous companies, including Founders Fund venture capital, MetaMed research, the US Army's University of Foreign Military and Cultural Studies, Facebook, and Twitter.
- Advisor for startups NumerAI (homomorphically encrypted algorithmic stock prediction competition), AlphaSheets (multi-language spreadsheet IDE), Guesstimate (probabilistic programming spreadsheet interface), and Arbital (social media platform for navigating and scoring argument structures).

---

## Details

### EDUCATION

PhD in mathematics, UC Berkeley (May, 2013)  
Visiting scholar in mathematics, Uni Roma Tre (January - July 2010)  
M.Sc. in Mathematics, University of Toronto (August 2008)  
B.Sc. (Hons) in Pure Mathematics, Memorial University (May 2006)

### RESEARCH INTERESTS

Shared ownership protocols for AI systems; game theory for artificial agents; optimal bounded reasoning; safety for highly advanced AI systems; algebraic statistics and other applications of algebraic geometry to machine learning and statistical modeling; human cognition and de-biasing; causal inference; hidden Markov models; graphical models; singular learning theory; tensor network models; parameter identifiability; financial market research.

### SELECTED PUBLICATIONS

*A Benchmark for Robust Imitation*, with Sam Toyer et al (2020). Submitted for review.

*Aligning AI With General Human Values*, with Dan Hendrycks et al (2020). Submitted for review.

*Emergent Complexity and Zero-shot Transfer via Unsupervised Environment Design*, with Michael Dennis et al (2020). Submitted for review.

*Symmetry, Equilibria, and Robustness in Common-Payoff Games*, with Scott Emmons et al (2020). Submitted for review.

*Distinguishing Adversarial Robustness from (Un)exploitability*, with Michael Dennis et al (2020). Submitted for review.

*AI Research Considerations for Human Existential Safety*, with David Kruger (2020). Submitted for review.

*Neural Networks are Surprisingly Modular*, with Daniel Filan, Shlomi Hod, Cody Wild, Stuart Russell (2020). arXiv:2003.04881.

*A parametric, resource-bounded generalization of Löb's theorem, and a robust cooperation criterion for open-source game theory* (2019). The Journal of Symbolic Logic.

*Negotiable reinforcement learning for pareto optimal sequential decision-making*, with Nishant Desai, Stuart J Russell (2018). Advances in Neural Information Processing Systems.

*Servant of Many Masters: Shifting priorities in Pareto-optimal sequential decision-making*, with Stuart Russell (2017). arXiv:1711.00363.

*A formal approach to the problem of logical non-omniscience*, with Scott Garrabrant, Tsvi Benson-Tilsen, Nate Soares, Jessica Taylor (2017). arXiv:1707.08747.

*Toward negotiable reinforcement learning: shifting priorities in Pareto optimal sequential decision-making*, (2016). arxiv.org:1701.01302

*Logical Induction*, with Scott Garrabrant, Tsvi Benson-Tilsen, Nate Soares, and Jessica Taylor (2016). arxiv.org:1609.03543.

*Parametric Bounded Löb's Theorem and Robust Cooperation of Bounded Agents*, (2016). arXiv:1602.04184.

*Alignment for advanced machine learning systems*, with Jessica Taylor, Eliezer Yudkowsky, Patrick LaVictoire (2016). Machine Intelligence Research Institute report.

*An Algorithm that Inductively Learns to Assign Well-Behaved Probabilities to Logical Sentences*, with Scott Garrabrant, Tsvi Benson-Tilsen, Nate Soares, and Jessica Taylor (2016). Submitted to Innovations in Theoretical Computer Science (ITCS).

*Algebraic geometry of matrix product states*, with Jason Morton (2014). arxiv.org:1210.2812. Symmetry, Integrability and Geometry: Methods and Applications.

*Inconsistency Evaluation in Preference Relations: A Characterization Based on Metrics Induced by a Norm*, with Michele Fedrizzi, Nino Civolani (2013). ODAM 2013.

*A note on the proportionality between some consistency indices in the AHP*, with M. Brunelli, and M. Fedrizzi (2010). arXiv:1203.6431v1. Applied Mathematics and Computation 219 (14) (2013), 7901-7906.

*Algebraic geometry of hidden Markov and related models*, (2013). PhD thesis in mathematics at UC Berkeley.

*Binary hidden Markov models and varieties* (2012). arXiv:1206.0500. Journal of algebraic Statistics, Vol 4, No 1 (2013): AS2012 Special Volume, part 2

*Polynomial constraints on representing entangled qubits as matrix product states*, with Jason Morton (2012). SIGMA 10 (2014), 095.

*Resolving the Banach-Tarski paradox: inseparability of rigid bodies* (2006). B.Sc. Honors thesis.

*Pushing the limit (generalized limits and limit extrema)* (2006), AEJM 1 no. 1, pp. 47-55.

#### SOFTWARE

*Robust Rental Harmony*, with Jacob Tsimmerman and Chelsea Voss (2015), a novel algorithm and on-line app for assigning envy-free prices to rooms in a shared house, in a way that is maximally robust to changes in preferences.

*Tensors.m2*, with Claudiu Raicu (2012), a Macaulay2 package for studying varieties of tensors and tensor networks.

*Credence Game*, with Alexei Andreev and Zachary Alethia (2012), a game for Windows, Mac, Android and iPhones for calibrating reported subjective credence levels to actual success rates.

#### PROGRAMMING LANGUAGES

Proficient in OCaml, VBA, Bash, and Macaulay2.  
Some experience in Python and R.

#### INVITED TALKS (SELECTED)

Foresight 2020 AGI Strategy Meeting	March 2020
Partnership on AI All Partners' Meeting	September 2019
Singapore Prime Minister's Office Foresight Conference	July 2019
Deepmind AI Safety Workshop in Iceland	October 2019
Foresight 2019 AGI Strategy Meeting	June 2019
NASA/LMI Workshop on Developing Trustworthy Autonomous Systems	October 2017
American Political Science Association Panel Annual Meeting,	September 2017
UC Berkeley EECS Faculty Retreat	April 2017
University of Toronto CS department lecture on open source game theory	March 2017
MIT CS department lecture on logical induction	December 2016
OpenAI research seminar	September 2016
Human Compatible AI research seminar	August 2016
Oxford University Future of Humanity Graduate Lecture Series	November 2015
Canadian IMO team Summer Camp	June, 2015
Thiel Foundation Summit in San Francisco	June 2014

Harvard Effective Altruism speaker series	April 2014
TEDxYouth@Tallinn in Tallinn, Estonia	November 2013
Thiel 20 Under 20 Summit in New York City	November 2013
Effective Altruism Summit	July 2013
Thiel Foundation Appathon	February 2013
Carnegie Mellon statistics seminar	February 2013
Duke University algebraic geometry seminar	January 2013
NCSU symbolic computation seminar	January 2013
McGill University causal inference seminar	December 2012
UC Berkeley algebraic geometry seminar	November 2012
Griffiths cognitive science laboratory	October 2012
RTG Workshop on Tensors and their Geometry at UC Berkeley	September 2012
York University statistics seminar	September 2012
Queen's University algebraic geometry seminar	September 2012
Autonomous electrical engineering and computer science seminar at UC Berkeley	August 2012
Summer Program on Applied Rationality and Cognition in Berkeley	August 2012
CFAR Westminster Retreat in Alamo, California	July 2012
Center for Applied Rationality workshop in Berkeley, California	June 2012
Algebraic Statistics in the Alleghenies at Pennsylvania State University	June 2012
Center for Applied Rationality workshop in Pescadero, California	May 2012
Founders Fund annual business meeting in Maui	May 2012
Stanford student algebraic geometry seminar	February 2012
COGS Causal Inference Symposium at UC Berkeley	February 2012
UC Berkeley algebraic geometry seminar	January 2012
Mathematicians Against Police Violence at UC Berkeley	November 2011
UC Berkeley political psychology working group	September 2011

PROFESSIONAL  
QUALIFICATIONS

FINRA Series 7 and Series 55 certified General Securities Representative and Equity Trader  
 Start-up and management experience (cofounded CFAR, rationality.org)  
 Management experience (as Executive Director of existence.org)  
 Freelance consulting experience (solicited by Founders Fund venture capital and MetaMed research)  
 Trilingual fluency in English, French, and Italian.  
 Bertini experience (numerical algebraic geometry package)

HONORS AND  
AWARDS

NSERC PDF Postdoctoral Fellowship – \$80,000	March 2013
Outstanding Graduate Student Instructor Award	March 2012
DARPA Graduate Student Research grant – \$24,000	January 2012
NSERC PGS-Doctoral Scholarship – \$42,000	September 2008
NSERC CGS-Doctoral Scholarship – \$105,000 (declined after 1 year)	September 2007
NSERC CGS-Masters Scholarship – \$17,500	September 2006
NSERC Undergraduate Summer Research Assistantship – \$6,000	May 2006
Governor General's Medal for Memorial University, awarded to the overall top undergraduate student across all faculties and departments.	May 2006
Memorial University Medal for Excellence in Mathematics	May 2006
Putnam mathematics competition – Honorable Mention (~ 5th in Canada)	December 2005
Centenary of Responsible Government Scholarship – \$1,000	November 2005

Memorial University undergraduate mathematics competition – 1st Prize	October 2005
Memorial University Faculty of Science Dean’s departmental prize, awarded annually to the top non-graduating mathematics student.	October 2005
APICS (Atlantic Provinces) mathematics team competition – 1st Prize	October 2005
Memorial University undergraduate mathematics competition – 1st Prize	January 2005
APICS (Atlantic Provinces) mathematics team competition – 1st Prize	October 2004
Memorial University Alumni Scholarship – \$25,000	2004/09

PROFESSIONAL  
SERVICE

Cofounder and Fund Advisor to the Survival and Flourishing Fund (2019 - present)	
Organizer of the weekly graduate seminar of the Center for Human-Compatible AI (2017 - present)	
Co-founder and investigator for BERI’s Grants Program (2017 - present)	
Thiel Under 20 Fellowship mentor (2013 - 2016)	
Lead curriculum developer for UC Berkeley’s <i>Sense, Sensibility, and Science</i> course, under physics Nobel laureate Saul Perlmutter, philosophy professor John Campbell, and law professor Robert MacCoun (fall 2012 - spring 2013)	
Cofounder and instructor for the Summer Program on Applied Rationality and Cognition (August 2012 - present)	
Signatory for THINK at UC Berkeley, facilitating discussion and implementation of high-impact altruism (fall 2012 - present)	
Curriculum Developer and Cofounder of the Center for Applied Rationality (spring 2012 - present)	
Co-organizer for the UC Berkeley student algebraic geometry seminar, first with Charley Crissman, and later with Andrew Dudzik (spring 2011 - spring 2012)	
Co-organizer for the COGS Causal Inference Symposium with psychology PhD student Michael Pacer (February, 2012)	
Co-organizer of the UC Berkeley algebraic statistics seminar with Shaowei Lin (fall 2011)	
Co-organizer of the Math, Productivity, Happiness and Decision-making seminar at UC Berkeley with Stanford mathematics PhD student Nisan Stiennon (fall 2011)	
Co-organizer of the Many-Algebro-Geometrically Important Concepts seminar at UC Berkeley (spring 2009)	

EMPLOYMENT  
HISTORY

<i>Research Scientist</i>	2017/02 – present
Employer: UC Berkeley Center for Human Compatible AI	40 hours per week
Work: researching methods to ensure beneficial human/AI interaction	
<i>Research Fellow</i>	2015/09 – 2017/02
Employer: Machine Intelligence Research Institute	40 hours per week
Work: researching the theory of artificial agents	
<i>Algorithmic Trader</i>	2014/04 – 2015-08
Employer: Jane Street Capital	50 hours per week
Work: researching, developing and supervising stock trading algorithms	
<i>Postdoctoral Research Fellow</i>	2013/09 (deferred)
Employer: Mathematical Biosciences Institute (NSF)	40 hours per week
Work: researching applications of algebraic geometry to neuroscience and machine learning	

<i>Cofounder and Curriculum Developer</i>	2013/05 – 2014/04
Employer: Center for Applied Rationality	60 hours per week
Work: Developing curriculum for statistical improvements in decision-making and cognitive debiasing.	
<i>Graduate Student Instructor</i>	2010/08 – 2011/12 and 2012/08 – 2013/05
Employer: UC Berkeley	20 hours per week
Work: classroom instruction, office hours, and grading.	(except summer)
<i>Graduate Student Researcher</i>	2012/01 - 2012/07
Employer: DARPA / UC Berkeley	20 hours per week
Research on applications of algebraic geometry to machine learning models under Professor Bernd Sturmfels	
<i>Graduate Student Instructor</i>	2008/08 – 2009/12
Employer: UC Berkeley	10 hours per week
Work: classroom instruction, office hours, and grading.	(except summer)
<i>Lecturer (vector calculus)</i>	2009/06 – 2009/08
Employer: UC Berkeley	15-20 hours per week
Work: classroom instruction, office hours, and grading.	
<i>Teaching Assistantships</i>	2006/10 – 2008/05
Employer: University of Toronto	5-10 hours per week
Work: tutorial classes, tutoring, and grading.	
<i>Research Assistantship (NSERC-USRA)</i>	2006/05 – 2006/08
Employer: NSERC / University of Toronto	full time
Work: researching division algorithms for analytic and quasi-analytic function classes under Professor Edward Bierstone.	
<i>Mathematics Tutor</i>	2005/05 – 2006/05
Employer: Memorial's Undergraduate Career Experience Program (MUCEP)	variable hours
Work: Tutoring mathematics regularly to residents of Paton College under Student Affairs and Services.	
<i>Problem Solver/Designer</i>	2003/06 - 2003/08
Employer: Student Work and Service Program (SWASP)	full time
Work: Creating, solving and typesetting problems for a contest training database under Professor M. Parmenter.	