



Curriculum Vitae of James A. Faville

☎ (914) 361-9016 • ✉ jamesfav7@gmail.com • 🌐 jamesfaville.com

Born 19 February 1995

Education

Haverford College, Haverford, PA

May 2018

Graduated with a Bachelor of Arts (B.A.) after advanced coursework in linguistics (major), classical languages (minor), and mathematics. Awarded the Howard Comfort Prize in Latin.

Bronxville School (K-12), Bronxville, NY

September 2000–May 2013

Graduated first in class with an award in Overall Academic Excellence.

Work Experience

Stanford Existential Risks Initiative, ML Alignment Theory Scholar

June 2023 - Present

Cooperative AI Stream; Mentor: Jesse Clifton.

Various - Independent & Contract Research

June 2021 - June 2023 (intermittent)

Various - Freelance Web Design and Technology Consulting

September 2011 - June 2023 (intermittent)

Center on Long-Term Risk, London, UK - Summer Research Fellow

June 2022–August 2022

Researched agential risks of astronomical future suffering, focusing on conflict between AI systems.

Center for Reducing Suffering, Remote - Research Intern

January 2021–May 2021

Wild Animal Initiative, Remote - Strategic Planning Intern

June 2020–December 2020

Animal Charity Evaluators, Remote - Research Intern

December 2019 - May 2020

Epic Systems, Verona, WI - Technical Problem Solver

August 2018 - January 2019

Haverford Classics Department, Haverford College, Haverford, PA

Teaching Assistant for Introductory Latin

January 2016 - January 2018

Programmer for <https://bridge.haverford.edu/>

June 2016 - August 2016

Tri-College Linguistics Department, Haverford College, Haverford, PA

Grader, Peer Tutor, Research Assistant for Professor Shizhe Huang

February 2016 - March 2018

Teaching Assistant for Syntax I

September 2014 - December 2014

John Hopkins Center for Talented Youth, Lancaster, PA - Linguistics TA

July 2015 - August 2015

Campus Activities

Problem Solving Group, *President* (2017-2018)

Joined Fall 2013

Problem Solving Group met for several hours each week to solve difficult problems in a variety of mathematical subfields, preparing students for the annual Putnam Exam.

Haverford Effective Altruism, *Co-Head* (2017-2018)

Joined Spring 2015

The effective altruism club organised campus-wide events and met weekly to discuss issues in applied ethics.

Machine Learning / AI Safety Reading Group, *Organiser*

Spring 2018

This reading group hosted informal discussions of AI safety and capabilities research.

High-Impact Career Fellowship, *Instructor*

Fall 2017, Spring 2018

The High-Impact Career Fellowship, sponsored by Haverford Center for Career and Professional Advising, helped undergraduate students plan for careers that have large positive impact on the world.

Beyond the Reals: An Exploration of Mathematics in Fiction, *Student Seminar Leader*

Fall 2014

Selected Papers

Eventual strategies for helping artificial sentience

Report for Center for Reducing Suffering

Discusses potential non-research interventions for benefiting artificial sentience, in order to better inform near-term research and capacity building efforts, and to begin an object-level discussion of which interventions seem most robust. This report has not yet been published; a draft is available upon request.

Animal Advocacy & Longtermism

Report for Animal Charity Evaluators

This report has not yet been published; a draft is available upon request.

Coindexation under Attitude Complements

Undergraduate Thesis for Linguistics

Presents a novel finding pertaining to interpretation of bound anaphors in embedded clauses, and argues in favor of a new derivation of 'de re' readings which also accounts for several puzzling aspects of Control Theory.

Selected Coursework

Swarthmore College, Seminar in Tense and Modality

Instructor: Peet Klecha, Term Paper: *Interpretation of the Ancient Greek optative mood*

Haverford College, (Abstract) Algebra I & II

Instructor: Elizabeth Milićević, Term Paper: *Multiplicative groups of p-adic fields*

University of Pennsylvania, Topics in Syntax and Semantics: Causatives

Instructor: Alison Biggs, Term Paper: *Successive Causation, a modification of Copley & Harley's Force-theoretic semantics*

Udacity (online), Deep Learning by Google and Introduction to AI

Coursera (online), Game Theory I, II and Mind & Machine Specialization

Talks

Paideia Institute, Living Latin and Greek Online 2021

February 2021

Topic: *De nullitate aliorum* (on the non-existence of others); discusses solipsism, open individualism, & Ibn Rushd.

AI Safety Support, AI Safety Unconference

March 2021

Topic: *AI alignment and artificial sentience*; describes three types of AS and discusses implications for alignment work.

Effective Altruism NYC, Various Lightning Talks

July 2019, April 2020, October 2020

Topics: *What is a moral circle?*, *Forms of value lock-in*, *Global mental health as Cause X*

NYC Rationality

February 2020

Topic: *Comparison of decision theories*

Professional Development

Alignable Structures, Effective Altruism Philadelphia

November 2022

Other selected events: Ctr. for Effective Altruism (EA Global Washington, D.C., Oct. 2022; EA Global San Francisco, June 2020 & Aug. 2022); EA Student Summit, Oct. 2020; EA Global X Virtual; École 42 (Intensive Basic Training in C, Aug. 2019); AI Safety Unconference, Mar. 2021 - Speaker; Ctr. on Long-Term Risk (Intro Seminar to Risks of Astronomical Suffering, Mar. 2021); Paideia Inst. (Living Latin Online, Feb. 2021 - Speaker; Living Latin in NYC, Feb. 2016, 2017, 2020; Living Greek in Greece, July 2017; Living Latin in Rome, June-July 2015); Penn Effective Altruism (EA Global X Philadelphia, April 2017)

Additional Information

Programming Languages & Computer Skills:

- Python, Javascript (including ES6+, jQuery, & CoffeeScript), HTML, CSS, XML, \LaTeX (experienced)
- Java, C, PHP, MySQL, TensorFlow, Numpy, BASH, Excel, Git, Adobe Premiere, Audacity (familiar)

Modern Languages: English (native speaker); French (experienced); Italian, Spanish (familiar)

Ancient Languages: Latin, Ancient Greek (advanced study); Biblical Hebrew, Ancient Egyptian (familiar)

Travel Experience: Italy, Greece, United Kingdom

Other Skills: Amateur Violist (out of practice), Adult & Pediatric CPR/AED/First Aid (certification lapsed)

Shortlisted candidate for the Research Scholars Programme of the Future of Humanity Institute at Oxford