

Pavilion Seminar: How will Artificial Intelligence change Humanity?

Seminar Description

Artificial intelligence has made remarkable advances in the past decade, leading to machines that can outperform humans on many of the tasks that once defined what it means to be human: understanding language, recognizing images, playing games, and even creating art. According to many prognosticators, within just a few decades we may reach a world where the traditional purposes of human existence, and the work the preponderance of humans do today, will no longer exist. This seminar will explore the validity of such predictions, and consider what the future of humanity is in a world that may not need us. We will explore these issues from a variety of perspectives, spanning economics, politics, philosophy, computer science, and anthropology. We will include both historical and fictional readings to understand how humanity has adapted to past dramatic shifts, technical readings to understand the present and future of artificial intelligence, philosophical and political readings to understand how society might adapt to increasingly intelligence machines, and various other media including computer simulations, music, and movies.

Assignments

Reactions to Readings: Short weekly essays reacting to readings and questions posted about the readings. These essays will be posted in a class discussion website, and available to all students in the class for further discussion.

Course blog: Small groups of students will be responsible for producing a public web page summarizing each reading and discussion, with the goal of producing a visible and useful externally-facing site by the end of the seminar. (This will follow a model the instructor has used successfully in past seminars, e.g., <https://tlseminar.github.io/>.)

Papers: Two major papers one due mid-way through the semester, and the other due near the end. For both of the papers, students would submit a preliminary draft to the instructor for feedback, and would be expected to revise the final paper in response to comments and discussion. For the first paper, students will focus on one aspect of how artificial intelligence has already impacted society, describing the impact of technological advances on a social, political, economic, or psychological aspect of human existence. For the second paper, students will speculate on the future, grounding their arguments in technical understanding of the expected capabilities of artificial intelligence, and considering how humanity may adapt to a future with intelligent machines.

Readings

Historical Perspective:

- Edgar Allan Poe, *Maelzel's Chess-Player*, 1836.
- Yuval Noah Harari. *Sapiens: A Brief History of Humankind*, Harper, 2015. (Selected chapters)
- A. M. Turing. *Computing machinery and intelligence*. Mind, 1950.

Technical perspective:

- Ray Kurzweil. *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*. Penguin Books, 1999.

- Aylin Caliskan, Joanna J. Bryson, Arvind Narayanan. *Semantics derived automatically from language corpora contain human-like biases*. Science, 14 Apr 2017.
- Rodney Brooks. *The Seven Deadly Sins of AI Predictions*. Technology Review, 6 October 2017.

Societal perspective:

- Bertrand Russell, *In Praise of Idleness*. Harper's Magazine, 1932.
- Cathy O'Neil. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown, 2016.
- Erik Brynjolfsson and Andrew McAfee. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company, 2014.
- Martin Ford. *Rise of the Robots: Technology and the Threat of a Jobless Future*, Basic Books, 2015.

Humanist perspective:

- Maureen Dowd. *Elon Musk's Billion-Dollar Crusade to Stop the A.I. Apocalypse*. Vanity Fair Magazine, April 2017.
- Max Tegmark. *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf, 2017.
- Yuval Noah Harari. *Homo Deus: A Brief History of Tomorrow*. Harper, 2017.

Justification for Pavilion Seminar

Focuses on big topics with enduring relevance: The emergence of artificial intelligence presents humanity with the largest transformation we have faced since the agricultural revolution. As machines become better than humans at tasks we once considered uniquely human, humanity will need to rethink what is important about being human and remake our societies in fundamental ways.

Engages multiple modes of inquiry: The seminar will explore the potential and impact for artificial intelligence from many perspectives, including both humanistic and scientific ones.

Interdisciplinary and pedagogically innovative: The seminar will include readings from philosophers, computer scientists, economists, novelists, and anthropologists. For both of the major writing assignments, students will be expected to integrate technical understanding with depth in at least one other domain.

Fulfills second writing requirement: The course will include two major writing assignments. Although the instructor is not in the College of Arts & Sciences, he was the Founding Director of the Interdisciplinary Major in Computer Science (and current director of the Computer Science Distinguished Major Program). Since its founding in 2006, the major has grown to the ninth largest major in the College with 133 students graduating with BA degrees in 2017.

Novelty: This is a new course, and no similar course has been offered at UVA (or any other institution so far as I am aware). The most similar UVA course is probably PHIL 2340: The Computational Age (taught by Paul Humphreys), which has some overlap in topics in considering how computing changes how humans think of ourselves and our role in the world, but does not focus on artificial intelligence or societal change.

David Evans - Condensed CV

Professor of Computer Science
University of Virginia

evans@cs.virginia.edu
<https://www.cs.virginia.edu/evans>

Appointments

University of Virginia, *Professor of Computer Science* (since 2013).

Associate Professor of Computer Science, 2006–2013.

Assistant Professor of Computer Science, 1999–2006.

Founding Director of the *Interdisciplinary Major in Computer Science*; Director of *Distinguished Major in Computer Science*. Program approved in 2006, has grown to the College of Arts & Sciences ninth most popular major. It is the only undergraduate degree program that spans the University's two largest undergraduate schools.

First Udacity Professor, 2012 (on leave from UVA). Developed an open, on-line introductory Computer Science course that has enrolled over 500,000 students, as well as an Applied Cryptography course (over 70,000 students), and assisted other instructors in developing courses that have been taken by over a million students from 190 countries.

Selected Awards and Honors

University Teaching Fellow, 2001.

National Science Foundation CAREER Award, 2001. National Science Foundation's "most prestigious awards in support of junior faculty who exemplify the role of teacher-scholars through outstanding research, excellent education and the integration of education and research..."

ACM Jefferson Undergraduate Teaching Award, 2002.

Harold Morton Jr. Award for Teaching, 2004. School of Engineering's highest teaching award.

All-University Teaching Award, 2008.

State Council of Higher Education for Virginia Outstanding Faculty Award, 2009. This is the Commonwealth of Virginia's highest honor for faculty, awarded by the Governor for "superior accomplishments in teaching, research, and public service."

Distinguished Research Award, 2014. Awarded by the University's Vice President of Research to two recipients each year to recognize "our most promising and creative new full professors".

Selected Publications

See <http://www.cs.virginia.edu/evans/pubs> for links to all publications. My publications have been cited over 9300 times (h-index = 38) (citation counts are from <http://scholar.google.com>, 23 October 2017). Student co-authors for whom I am the primary supervisor are indicated with *.

David Evans. *Static Detection of Dynamic Memory Errors*. In *SIGPLAN Conference on Programming Language Design and Implementation*. May 1996. (295 citations)

Lingxuan Hu* and David Evans. *Localization for Mobile Sensor Networks*. In *10th Annual International Conference on Mobile Computing and Networking*. September 2004. (1311 citations)

Jinlin Yang*, David Evans, Deepali Bhardwaj, Thirumalesh Bhat, and Manuvir Das. *Perracotta: Mining Temporal API Rules From Imperfect Traces*. In *28th International Conference in Software Engineering*. May 2006. (351 citations)

Karsten Nohl*, David Evans, Starbug, and Henryk Plötz. *Reverse-Engineering a Cryptographic RFID Tag*. In *17th USENIX Security Symposium*. August 2008. (260 citations)

Yan Huang*, David Evans, Jonathan Katz, and Lior Malka. *Faster Secure Two-Party Computation Using Garbled Circuits*. In *20th USENIX Security Symposium*. August 2011. (423 citations)

David Evans. *Dori-Mic and the Universal Machine! (A Tragicomic Tale of Combinatorics and Computability for Curious Children of All Ages)*. 2014.

Weilin Xu* and Yanjun Qi and David Evans. *Automatically Evading Classifiers: A Case Study on PDF Malware Classifiers*. In *Network and Distributed System Security Symposium*. February 2016. (48 citations)

Selected Invited Talks

USENIX Security Symposium. *What Biology Can Teach Us About Security*. August 2004.

University of Oregon, Distinguished Speaker. *Scaling Secure Computation*. Oregon Security Day. Eugene, Oregon. 5 April 2013.

Microsoft Research Workshop on Applied Multi-Party Computation, Keynote Talk. *Secure Computation in 2029: Boom, Bust, or Bonanza*. Redmond, WA. 21 February 2014.

USENIX Security Symposium Doctoral Colloquium, Invited Talk. *How To Live In Paradise: A Guide for Prospective, New, and Disgruntled Professors*. Washington, DC. 13 August 2015.

USENIX Enigma. *Classifiers Under Attack*. Oakland, CA. 1 February 2017.

Saarland University, CISPA Distinguished Lecture. *Adversarial Machine Learning: Are We Playing the Wrong Game?* Saarbrücken, Germany. 7 July 2017.

Grants

Principal or Co-Principal Investigator on over \$30M (\$14M to UVa) of research funding from the National Science Foundation, NASA, and Department of Defense. Principal Investigator on ten National Science Foundation grants (including an NSF CAREER award, \$1M award on RFID security, and a \$3M award on secure computation) and two Google Research Awards.

Selected Media Coverage

UVa Today. *Influencing Eternity: University of Virginia to Honor Outstanding Teachers*. April 2008.

UVa Today. *U.Va. Computer Scientist David Evans Wins Statewide Outstanding Faculty Award*. 29 January 2009.

Prospect Magazine. *Professors without Borders*. 28 June 2012.

New York Times. *A Surge in Growth for a New Kind of Online Course*, 25 September 2013.

UVA Today. *Alumna-Turned-Internet Security Expert Listed Among Nations Top Young Innovators*. 21 September 2017.