Syllabus

PAVS 4500 (004): How will Artificial Intelligence change Humanity?

University of Virginia, Fall 2018

Meetings: Mondays, **3:30-6:00PM** in **Rice 536**.

Coordinator: David Evans (evans@virginia.edu). My office is Rice 507.

Office Hours: I will have office hours on **Thursdays**, **9:00-10:30am**. To schedule another time, please use https://davidevans.youcanbook.me/. When my door is open, you are welcome to stop by anytime.

Course Website: https://aipavilion.github.io

Forum: We will use the subreddit, /r/aipavilion, for discussions and sharing.

Description

Artificial intelligence has made remarkable advances in the past decade, leading to machines that can out-perform 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 intelligent and powerful machines, and various other media including computer simulations, music, and movies.

Expected Background

This is a Pavilion Seminar, open to students in all majors and targeted to third and fourth-year students in the College. Enrollment is by instructor permission, and strictly limited to 15 students.

Students are not required to have any particular background in computing or artificial intelligence, but it is hoped that all student swill bring some interesting experience and background to the seminar.

Assignments

Students in the seminar will be expected to complete a variety of different types of assignments during the semester, including:

Readings and Reactions: Most weeks we will have reading (and sometimes viewing) assignments that will typically be some chapters from a book and a few short articles. Short weekly responses to readings and questions posted about the readings. These essays will be posted on the class forum, and available to all students in the class for further discussion and comments.

Papers: There will be two major papers in the seminar. For both of the papers, students will develop an idea for the paper and discuss it with the class, 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. Students are encouraged to develop creative ideas for alternative topics for the papers, as well as alternate communications tools (such as a scripted video or podcast instead of a paper), and to discuss and gain approval for proposed alternatives with the course instructor.

The deliverables for the paper assignments are:

– Sunday, 30 September: Idea for First Paper Due
– Wednesday, 17 October: First Paper Draft Due
– Tuesday, 30 October: First Paper Final Due
– Sunday, 11 November: Proposal for Final Paper Due
– Tuesday, 20 November: Draft of Final Paper Due
– Monday, 10 December: Final Paper Due

(Deadlines are by 11:59pm on the date given.)

Readings

The main books we will read for the first part of the class will be:

– Yuval Noah Harari, *Sapiens: A Brief History of Humankind*. February 2015 (first published in Hebrew in 2011) [Amazon] [Barnes & Noble]

– Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*. 2014. [Amazon] [Barnes & Noble]

The focus books for the second half of the class will be selected based on students' interests and the discussions we have in class, but the most likely books include:

– Cathy O'Neil. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. 2016.

– Martin Ford. *Rise of the Robots: Technology and the Threat of a Jobless Future*. 2015.

– Yuval Noah Harari. *Homo Deus: A Brief History of Tomorrow*. 2017.

– Jerry Kaplan, *Humans Need Not Apply: A Guide to Wealth and Work in the Age of Artificial Intelligence*. 2015.

– Garry Kasparov, *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*. 2017.

– Max Tegmark. *Life 3.0: Being Human in the Age of Artificial Intelligence*. 2017.

In addition to the books, there will be readings from articles and other materials. The weekly reading assignments will be posted on the course site.

Honor

We believe strongly in the value of a *community of trust*, and expect all of the students in this class to contribute to strenghtening and enhancing that community.

As a student at the University of Virginia, you are trusted to be honorable and expected to behave in ways that merit that trust. We take advantage of this trust to provide a better learning environment for everyone. The course will be better for everyone if everyone can assume everyone else is trustworthy, and we start from the assumption that all students at the university deserve to be trusted.

For most assignments in this course, you will be encouraged to discuss ideas and work with others to develop your ideas. You will always be expected to credit any collaborators and document any resources you use. The honor expectations for each assignment should be clearly stated and make it unambiguous what is and is not permitted. If it is ever unclear what is considered acceptable on an assignment, please check with me.

Expectations and Accommodations

It is my goal to create a learning experience that is as engaging, worthwhile, and accessible as possible. We hope the topics we discuss in this class will be ones where many students feel passionately and have strong views they want to share. Some of the readings will contradict your deeply held beliefs and challenge your assumptions. It is important that everyone is respectful and that discussions focus on ideas and evidence, and that everyone has an opportunity to share their thoughts without interruption.

Since the class only meets once a week, and is in a seminar style, it is essential that everyone attend every class, prepares well and contributes outside of class, and is fully engaged during the seminar time. That said, I understand that personal circumstances arise that may make this difficult or impossible for some students some weeks. If you are encountering issues that make it difficult to participate fully in the seminar, please let me know as soon as possible, and we will work something out.

If you anticipate any issues related to the format, materials, or requirements of this course, please meet with me outside of class so we can explore potential options. Students with disabilities may also wish to work with the Student Disability Access Center to discuss a range of options to removing barriers in this course, including official accommodations. Please visit their website for information on this process and to apply for services online: sdac.studenthealth.virginia.edu. If you have already been approved for accommodations through SDAC, please send me your accommodation letter and meet with me so we can develop an implementation plan together.

Evaluation

My hope is students will focus on learning and producing something of value.

Students will be evaluated primarily based on:

– Overall contribution to the seminar. This includes contributions in class and one the course forum, as well as any extraordinary contributions to making the seminar worthwhile.

– Quality of the major writing assignments. Papers will be evaluated for their creativity, novelty, persuasiveness, and clarity. The grades will be based on the final papers, after revisions based on feedback from the proposal and first draft.

I don't provide a generic percentage breakdown for each of the components, since excellent performance on any one of them (along with acceptable effort on the others) will be enough to justify an A in the course. For a typical students, I would expect the grade is determined by 50% for the final paper, 25% for the first paper, and 25% for class contributions. But a student who does an outstanding job on the first paper, or who consistently makes outstanding contributions to the course, will have heavier weighting for those components.

