INST1550 – MILLENNIALS: PIONEERS OF THE ZETTABYTE AGE

COURSE SYLLABUS

Instructor

Philip H. Schroeder
phs5eg@virginia.edu

Faculty Sponsor

Dr. Michael P. Timko
mpt9g@virginia.edu

Course Description & Objectives

"...Tyger Tyger burning bright,
In the forests of the night:
What immortal hand or eye,
Dare frame thy fearful symmetry?"

– William Blake, "The Tyger"

Without you knowing when, why, or even how, machines make decisions every day about you and for you that shape society's perceptions of you and your perceptions of society and self. Whether it be by filtering your social media page, rating you against other job applicants or insurance payers, diagnosing your medical complaints, or determining how you should be policed, machine-learning systems are becoming increasingly empowered at all levels of society to take your personal data and make autonomous decisions about who you are, what you ought to be given, or how you ought to be treated. That is, for the first time in human history, there exist nonhuman agents making invasive, unregulated decisions that carry dangerous cultural and ethical implications.

From self-driving cars to lethal autonomous weapons, machines will progressively replace more humans in executing a myriad of tasks as the pursuit of artificial intelligence reaches full swing. Such progression of machine-learning systems will undoubtedly provide for revolutionary advancements in the fight against humanity's greatest challenges, from poverty and morbidity to violence and war. However, as machine behavior and decision-making replaces that of humans, what is lost? What is the difference between you and an intelligent machine? Perhaps nothing at all. What aspects of you really can't be effectively replaced by sophisticated algorithms? Is it your intelligence? ...your faith? ...your mindfulness? ...your emotional insight? ...your moral nucleus? Where exactly do these aspects of you come from and why can or can't they be programmed? These questions challenge us to reflect upon the very core of our humanity, and how our existence is, or is not, distinguishable from that of highly sophisticated systems. Such reflection can then help to guide a determination of which activities or tasks, and to what extent, machines can and should be prohibited from replacing humans. The issues of this course will call upon insight from a diverse array of perspectives, ranging from policy, philosophy, and religion to the engineering, physical, and life sciences. This course seeks to foster thoughtful discourse across and within each of these perspectives, with the ultimate goal of developing knowledgeable positions regarding:

1) The neurophysical, psychophysiological, philosophical, and theological arguments for how and why the human is or is not distinguishable from intelligent machine systems and, based on this, the extent to which there is an inevitable loss in moral guidance, mindful thought, or conscious intentions when a machine replaces a human
2) The types of activities or occupations, if any, for which a machine should be restricted from replacing a human
   Ex: Should a machine be permitted to care for an infant? ...execute a medical therapy? ...engage in sex? ...drive a car? ...take a human life?
3) The policy and regulatory mechanisms needed to enforce such restrictions and monitor machine activity
   Ex: Should humans be free to object to decisions made about them or for them by machines? Who should be held accountable for unintended consequences of machine behavior? How can and should black-box algorithms be audited?

Please note, the objective of this course is <u>not</u> to excite provocative theorizing based on caricatured illusions of robot-driven worlds, but instead to equip and to train disciplined dialogue founded upon a sober depiction of the current status of machine-learning systems, the potential for artificial intelligence, and the real-world implications of such technology within different sociocultural and professional contexts. Further, in covering the broad landscape of perspectives relevant to this course, the intention is not to be exhaustive, but instead to raise more questions than answers and to preserve a constructive malleability that, in turn, can foster growth of intellectual curiosity as we, as Millennials, prepare to take on the challenges of the coming Zettabyte Age.

Assessments of Learning

The principle objective of this course is to provide a setting, in which a variety of perspectives regarding the nature of the human relative to sophisticated machine systems can be explored and reflected upon to develop knowledgeable positions regarding potential sociocultural issues, and the pragmatic challenges of addressing such issues, with intelligent machines replacing humans in executing a variety of different tasks and occupations. Thus, assessments of learning will be focused on reflections of course content and development of sound positions regarding such content. These focused assessments, respectively, will take the form of Weekly Reading Reflections and 2 Position Papers.

| Weekly Reading Reflections

- To be submitted by midnight on the Saturday prior to the Monday class meeting of each week of the semester via UVACOLLAB
    - Reflections will be made available for classmates to read via the course UVACOLLAB page once the submission deadline has passed, allowing for an additional means of sharing ideas and providing a platform for in-class discussion
- Expected length of 200-400 words
- Should thoughtfully reflect upon concepts discussed in assigned readings
    - Submissions are not expected to be exhaustive or formal, but should demonstrate knowledge of the readings and issues at hand

| 2 Position Papers

- To be submitted for any two, nonconsecutive, weeks of the semester via UVACOLLAB
- Expected length of 2-3 pages - standard margins, font, and formatting
- Should take a clear position on a question raised for that week's topic, defend the position with appropriate references to the readings, and address countering or alternative positions for the chosen question

Grading

INST courses are recognized as 'outside the College' and taught on a CR/NC basis. For this course, feedback regarding assignments and overall contributions to class will be provided to each student on an individual basis throughout the semester. Performance will be assessed according to the following criteria:

Assignment Description	Percentage of Final Grade
Weekly reflections to writing prompts	30%
2 Response Papers	20%
In-class participation	50%

Attendance Policy

Students are expected to attend all course meetings, barring exceptional circumstances (i.e., illness, family emergencies, religious holidays, required academic or athletic activities). Please contact me via email (phs5eg@virginia.edu) if you anticipate an absence.

Course Materials

All assigned readings and course materials will be provided via UVACOLLAB or email.

Course Schedule & Readings

INTRODUCTION

Wk 1. What makes us human? | **January 22**

Joubert, C. 2011. "What makes us human, and why it is not the brain: A creationist defense of the soul"

Wk 2. Machine Learning: Power, Limitations, & Dangers | **January 29**

"Who can make the muddy water clear?" – poem by Lao Tzu

Perisic, I. 2016. "Making hard choices: The quest for ethics in machine learning"

Collins, H. 1995. "Humans, machines, and the structure of knowledge" from *Constructions of the Mind: Artificial Intelligence and the Humanities*

Caliskan, A. 2017. "Semantics derived automatically from language corpora contain human-like biases."

In the News:

- Liedtke, M. 2015. "Google's new app blunders by calling black people 'gorillas'"
- Johnston, I. 2017. "AI robots learning racism, sexism, and other prejudices from humans"
- Olon, S. 2017. "The future of fake news: Don't believe everything you read, see or hear"

Wk 3. Human Interactions | **February 05**

Weizenbaum, J. 1966. "ELIZA - A computer program for the study of natural language communication between man and machine"

Suchman, L. 1987 "Chapter 7: Human-machine communication" from *Plans and Situated Actions: The problem of human-machine communication*

Suzuki, Y. et al. 2015. "Measuring empathy for human and robot hand pain using electroencephalography"

WHAT CAN A MACHINE LEARN?

Wk 4. The Mind-Brain Interaction | **February 12**

Qur'an 17:85

Fieser, J. 2017. "Chapter 3: Mind" from *Great Issues of Philosophy*

Schwartz, J. et al. 2005. "Quantum physics in neuroscience and psychology: A neurophysical model of mind-brain interaction"

Frank, A. 2017 "Mind, matter and materialism"

Wk 5. Morality | **February 19**

Ayala, F. 2010. "The difference of being human: Morality"

Sherwin, Y. 2012 "Machine morality: Computing right and wrong"

Saptawijaya, A., Pereira, L., 2015."Counterfactuals in logic programming with applications to agent morality"

Wk 6. Free Will | **February 26**

Fieser, J. 2017. "Chapter 4: Free will" from *Great Issues of Philosophy*

Libet, B. 1999. "Do we have free will?"

Wegner, D. 2004. "Precis of the illusion of conscious will"

Spring Recess | **March 05**

Wk 7. Idea of Self | **March 12**

Hotz, R. 2007. "How your brain allows you to walk in another's shoes"

Koch, C. 2009. "A theory of consciousness"

Kanai, R. 2017. "We need conscious robots"

Wk 8. God | **March 19**

Herzfield, N. 2002. "Creating in our own image: Artificial intelligence and the image of god"

Dennett, D. 2012. "'A perfect and beautiful machine': What Darwin's theory of evolution reveals about artificial intelligence"

Pickover, C. 2008. "Physics and religion" from *Archimedes to Hawking: Laws of Science & the Great Minds Behind Them*

"The Tyger" – poem by William Blake

POLICY & REGULATION

Wk 9. Accountability for Machine Decisions | March 26

Diakopoulos, N. 2016. "Accountability in algorithmic decision-making"
Baard, E. 2003. "The guilt-free soldier: New science raises the specter of a world without regret"
Goodman, B., Flaxman, S. 2017 "European Union regulations on algorithmic decision-making and a 'right to explanation'"

Wk 10. Human Scoring: Controlling Access to Information & Opportunity | April 02

Citron, D., Pasquale, F. 2014 "The scored society: Due process for automated predictions"
Newman, M. 2016. "AI is helping job candidates bypass resume bias and black holes"
Denyer, S. 2016. "China's plan to organize its society relies on 'big data' to rate everyone"

Wk 11. Human Scoring: Law Enforcement & Criminal Justice | April 09

Angwin, J. et al. 2016. "Machine bias"
"Workshop discussion notes: Predictive policing" & "Workshop discussion notes: Courts and predictive algorithms" from *Data & Civil Rights: A New Era Of Policing And Justice*
Madrigal, A. 2016. "Predictive policing: The future of crime-fighting, or the future of racial profiling?"

Wk 12. Black-Box Medicine | April 16

"What the doctor said" – poem by Raymond Carver
"Moral Residue" – poem by Elizabeth Epstein & Sarah Delgado
Drought, T. 2002. "The privilege of bearing witness"
Anderson, M. et al. 2006. "MedEthEx: A prototype medical ethics advisor"
Price, N. 2015. "Black-box medicine" from *Harvard Journal of Law & Technology*

Wk 13. Lovotics: Artificial Friendship, Love, Sex | April 23

Samani, H. 2012. "A multidisciplinary artificial intelligence model of an affective robot"
Richardson, K. 2016. "The asymmetrical 'relationship': Parallels between prostitution and the development of sex robots"
Coeckelbergh, M. 2010. "Artificial companions: Empathy and vulnerability mirroring in human-robot relations"

Wk 14. Mechanical Tasks: Driverless Cars, Weapons of War | April 30

Villasenor, J. 2014. "Products liability and driverless cars: Issues and guiding principles for legislation"
Belay, N. 2015. "Robot ethics and self-driving cars: How ethical determinations in software will require a new legal framework"
Asaro, P. 2012. "On banning autonomous weapon systems: Human rights, automation, and the dehumanization of lethal decision-making"