From Frankenstein to Alexa: A Humanistic Inquiry into the Ethics of Artificial Intelligence
Hurford Center for the Arts and Humanities, Student Seminar Proposal
Nicky Rhodes and Katya Olson Shipyatsky
Faculty Advisor: Dr. Craig Borowiak
March 26th, 2018

"Surpassing human limits is so human a quest, maybe the most ancient one of all, from an age when dreams were omens dipped in moonlight, and godlike voices raged inside one's head."
- Diane Ackerman, poet[1]

"Computers are useless. They can only give you answers."
- Pablo Picasso

For many, the term "Artificial Intelligence" conjures images of complex robots and incomprehensible advances in modern technology. However, humans have been attempting to create sentience for as long as they have walked the earth. Looking at the history of humankind from ancient civilizations dreaming about gods, through Mary Shelley's 1818 *Frankenstein*, we see that humankind has been contemplating the possibility of creating intelligence from nothing since long before the invention of the first computer. In recent decades, however, mathematicians and computer scientists have attempted to define human consciousness in terms of algorithms and equations, in hopes of eventually being able to replicate them and, by doing so, replicate the elusive essence of the human mind. But can human consciousness be manufactured? Can it be described with logic and symbols?

As scientists approach simulations of thought that more closely resemble our own minds, we are faced not only with the question of whether or not such replicas are "genuine," but also with the ethical and political implications of these new technologies. Who controls these new forms of technology? What obligations does a constructed mind have? Artificial Intelligence is evolving at a breakneck speed, faster than moral and political boundaries can be devised to regulate it and faster than citizens are able to comprehend its impact. As such, the ethical questions concerning Artificial Intelligence are becoming increasingly urgent: scholars, artists, and thinkers from across disciplines are inquiring into the potential for Artificial Intelligence to either heighten or destroy what it means to be human.

Mankind's historical relationship with technology - one of a master with a tool - is growing more complex. Our notions of "tool" and "user" are changing, as humans become data points that our machines collect and process. As our technologies that we interact with become enshrouded in obscurity and unknowability, it is becoming clear that these complications cannot be addressed by science alone. Through the varied lenses of literature, art, film, philosophy, politics, and ethics, we aim to wholistically examine the implications that Artificial Intelligence will have on humanity. We aim to go beyond technical scientific papers and delve into fictional novels, contemporary and historical movies and

[1] https://www.brainpickings.org/2014/10/06/diane-ackerman-human-age/

television shows, poetry, and more. These humanistic meditations on different interpretations of the "mind," combined with political and philosophical texts from many eras, will allow us to gain a deeper understanding of humanity's relationship with technology in this critical point in history.

This course will thrive with a group of students from diverse backgrounds, with different areas of knowledge and insight. Students within fields like Anthropology, Sociology, and Religion may be drawn to the course for its inquiry into human societies and the challenges they face. Students of Philosophy and Political Science will likely find the course compelling for the opportunity it presents to examine new challenges to the human experience and its governance. Those in disciplines like English, History, or Visual Studies might gain valuable insight into the various historical and contemporary representations of the human, the mind, and the machine. Students in the sciences could bring a technical understanding of the human mind and physical body, and may gain diverse perspectives on the ethical and political implications of their work. Computer Science students may have the opportunity to develop an empathetic and humanistic perspective of the technology that they may have a hand in shaping. This list is in no way exhaustive; students from every background will be able to add a valuable and unique voice to the discussion, ensuring a creative, cross-disciplinary examination of humanity and technology.

Two centuries ago, the ubiquity of industrial machinery revolutionized the way humans work and engaged with the environment and each other. Decades ago, the Internet revolutionized the way we think. Today, Artificial Intelligence has the potential to revolutionize who we fundamentally are as humans. As the forefront of this generation, it is our duty to think about this critically, before our machines do so for us.

Mechanics:

Every class session will be framed around central case studies and texts. At the end of each meeting, we will look over the material for the next session and collectively write a number of questions and prompts for participants to keep in mind while reading the material. During the two weeks in between sessions, participants will write down thoughts, questions, and answers as they read, which they will bring in to the following session to further stimulate conversation.

We will ask students to read Mary Shelley's historical novel *Frankenstein* (currently celebrating its 200th year anniversary), and watch the contemporary TV show *Westworld* over the summer of 2018. We will examine these stories from evolving perspectives over the course of the semester, as they raise critical questions about what it means to be a human, what it means to be a machine, and where the line is drawn.

We also may host optional weekly or bi-weekly movie nights, as there is an endless cornucopia of brilliant films on this topic.

Proposed Weekly Syllabus:

Week 1 — What is AI?: Defining the Evolution of Technology, Ethics, and the Mind

Creating thinking simulations of humanity is not a novel pursuit: Ancient Greek mythology included automatons and artificial beings in the courts of the god Hephaestus and Pygmalion; pre-science Jewish folklore depicted "Golems," intelligent beings composed of mud and clay; and in the past century, scientific minds such as Thomas Hobbes and Bertrand Russell strove to recreate the mind through formal symbols and logic. Through the lens of this historical narrative, we will discuss the evolution of Artificial Intelligence as a concept. Beginning with defining what "Artificial Intelligence" really is, we will then explore the emotions and inner-thoughts of manufactured beings in Mary Shelley's *Frankenstein*, compared to the contemporary equivalents in *Westworld* episodes. We will critically analyze various visual and literary representations of thinking technologies and examine how they have changed throughout history. We will ask what our relationships with technology have been in the past, and how that differs from the way they are now. Similarly, we will explore the ethical framework historical mankind has constructed around their technologies. This session will serve to establish a language for discussing the new paradigm of our evolving relationship with increasingly human-like technology.

Readings & Media:

Novel:

Shelley, Mary. *Frankenstein* **280 pages (read over summer)**

TV Series:

Westworld **Season 1, 10 episodes, 40 minutes each (watch over summer)**

Article:

Cesran International: Center for Strategic Research and Analysis, "The Current State of AI as We Begin 2018" **2 pages**

Bossman, Julia. "Top 9 Ethical Issues in Artificial Intelligence" **2 pages**

Video:

CGP Grey. "Humans Need Not Apply" **15 minutes**

Website:

Future of Life Institute, "Benefits and Risks of Artificial Intelligence."
https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/

Week 2 — AI and Humanity

The development of Artificial Intelligence necessarily implies defining what we as humans believe intelligence to be, hence, defining humanity as an essence. This process forces us to confront our conceptions of humanness and non-humanness, leaving us to grapple with difficult questions concerning the nature of consciousness, creativity, and emotion. We will question the difference between intelligence, knowledge, and wisdom. We will survey different definitions of what it means to be human, ranging from philosophical definitions based in ontological analysis of humans as thinking agents, to biological definitions of the human as a living organism. Should we even equate Artificial Intelligence to human intelligence, or are they incomparable phenomena? In hopes of answering these questions, we will probe contemporary impressions of human/computer relationships through Spike Jonze's movie, *Her*, while also looking at the construction of computer and human emotions. We will attempt to define creativity in the human mind, and examine the capabilities of creative machines. We will discuss how to value human life in an age where our decision-making capabilities may be obsolete, exploring various systems of worth. We may explore different ways of interfacing with Artificial Intelligence. Ultimately, this session aims to answer the question: What does Artificial Intelligence show us about Humanity?

Pre-Session Engagement:

Participants will find a piece of art online generated by AI to discuss in this session. For example, Aiva (http://www.aiva.ai/) uses AI to "compose" emotional pieces of music.

Readings & Media:

Movie:

Her **2 hours, group screening**

Journal Article:

Havlík, Vladimír. "The naturalness of artificial intelligence from the evolutionary perspective" **10 pages**

Fazi, M. Beatrice. "Can a machine think (anything new)? Automation beyond simulation" **12 pages**

Shah, Huma, et. al. "Machine humour: examples from Turing test experiments" **8 pages**

Video Lecture:

Dennet, Daniel. "From Bacteria to Bach and Back: The Evolution of Minds" **1 hour 15 minutes**

Encyclopedia Entry:

Internet Encyclopedia of Philosophy. "Consciousness" **Sections 1,2,5,6: 8 pages**

Website:

Space10. "Do you speak human?" https://space10.io/do-you-speak-human/

Further Reading (Optional):

Benjamin, Walter. *The Work of Art in the Age of Mechanical Reproduction*

Week 3 — AI and Religion

Humankind's relationship with the divine has historically been one born from a drive to understand the world. Religious thought helps us to understand life, death, and morality. If (or when) humans are no longer the only ones attempting to understand these concepts, how will our relationship with the divine change? Will artificially intelligent beings also probe these questions through religion? Or will their superior intelligence eliminate our need for religion entirely? Further, are we at the forefronts of constructing a new type of God, one that may truly watch over us from a heavenly perch within "the cloud" — a God that can see, interpret, and understand everything that happens around us, and make decisions accordingly? How should we design this god-like being? Would one worship this being, and if so, how? Would the data it embodies be closer to "truth" or "myth"?

Artificial Intelligence may soon allow humans to craft and control their own virtual realities, allowing us, in a sense, to "play god." Will our ability to act as creative forces ourselves threaten our faith in the divine ones? Humankind's relationships with the divine have also historically been shaped by uncertainty regarding death. Soon faced with the possibility of "uploading" our consciousness and in doing so becoming "immortal," will we still need religion? This week, we will look to history to examine representations of God as a projection of the human to consider possible relationships between humanity, Artificial Intelligence, and the divine in hopes of having a discussion about the future of religion in an Artificially Intelligent world.

Readings & Media:

Novel:

Mitchell, David. *Cloud Atlas*. **Chapters 5 and 6**

Book:

Aslan, Reza. *God: A Human History* **Intro-Chapter 3, 50 pages**

Journal Article:

Herzfeld, Noreen. "Creating in Our Own Image: Artificial Intelligence and the Image of God." **15 pages**

Video Lecture:

Harari, Yuval Noah. "The Rise of Data Religion" **55 mins**

Magazine Articles:

Merritt, Johnathan. "Is AI a threat to Christianity?" **3 pages**

Vlahos, James. "A Son's Race to Give His Dying Father Artificial Immortality" **3 pages**

Week 4 — AI and Politics

One of the most promising aspects of Artificial Intelligence is its potential to analyze infinitely massive amounts of social and environmental data, and "understand" it on a scale and with objectivity that surpasses the capabilities of humans. No more, some argue, will we have to rely upon democracy, with its inherent human imperfections and impercisions to make our political and economic decisions. Nor will we have to rely on the human mind, with its flaws and biases, to grapple with complex ethical dilemmas. However, while machines may be good at crunching numbers, what they do with these data stems into an entirely new field of ethics and politics. Scientists are even developing so-called ethical decision-making technology which will allow us to pass on our moral dilemmas to our machines. Using the 1970 film, *Colossus: The Forbin Project*, as a historical counterpoint, we will examine the implications of offloading our moral and political questions and processes to our machines, as well as the potential ways AI will interact with existing political, economic, and moral systems. We will discuss questions like: Do the fields of ethics and politics need humans, or can our biases be corrected by our artificial counterparts? What roles have machines played in the process of governance in the past, and how can this change with Artificial Intelligence? Who should control Artificial Intelligence? Who is to blame when inhuman, data-based decisions go wrong? Using various contemporary case studies in which we have entrusted machines to administer justice and save lives, we will discuss the capabilities and limitations of building "unbiased" artificial intelligence in an inherently biased world. Who should machines make decisions for: individual people, humanity, the environment, the economy — or perhaps themselves? We will explore these questions using Isaac Asimov's canonical "Three Laws of Robotics," and discuss the realistic implications of this intentionally fictional structure.

In Class Activity:
After discussing Asimov's "Three Laws of Robotics" as well their limitations, we will together come up with our own laws for Artificial Intelligence that draw on our discussions from previous weeks.

Readings & Media:
Movie:
Colossus: The Forbin Project **1 hour 40 minutes, group screening**

Journal Articles:
Harari, Yuval Noah. "Reboot for the AI Revolution" **6 pages**
Verhulst, Stefaan. "Where and when AI and CI meet: exploring the intersection of artificial and collective intelligence towards the goal of innovating how we govern" **5 pages**

Encyclopedia Entry:
Stanford Encyclopedia of Philosophy. "Civic Humanism" **Sections 1,2,6,7: 8 pages**

Magazine Article:
Solon, Olivia. "Artificial Intelligence is Ripe for Abuse, Tech Researcher Warns: 'A Fascist's Dream'" **3 pages**

Short Story Collection:

Asimov, Isaac. *I, Robot* **Select short stories**

Video:
Computerphile. "Why Asimov's Laws of Robotics Don't Work" **8 mins**

Government Report:
Executive Office of the President. "Artificial Intelligence, Automation, and the Economy." **Skim, <35 pages**

Website:
MIT Media Lab. "Moral Machine." http://moralmachine.mit.edu/
Automato.farm. "Ethical Things." http://automato.farm/portfolio/ethical_things/

Week 5 — AI and the Future

Despite all of our inquiry, it remains impossible to truly know what Artificial Intelligence will become. No amount of interdisciplinary, humanistic examination can bring us to a definite answer to the questions we have raised so far. In the face of this uncertainty, we turn to stories for potential futures. In this final session, we will reflect back on the entirety of the past sessions and ask: What inherent value do humans have that cannot be captured in code? Is there knowledge that machines cannot possess? We ask: What will societies look like if organized around an invisible, omniscient mind? Is Artificial Intelligence simply a natural evolution of biological life? More concisely, is AI humanity? Using discussions and questions from previous weeks, we will form our conversations this week around normative predictions for the future of the relationship between humans and their artificial counterparts. By examining passages from the movie *Blade Runner*, we will explore representations of the future of human/computer life in contemporary media. Will an infusion of AI into the genepool create a super-intelligent human race? Will political mishandling create an inescapable oppressive force? Will future humans worship AI as an omniscient being? Though this session (as well as this seminar as a whole) will likely leave us with more questions than answers, our hope is to leave the seminar able to engage more thoughtfully and wholistically in determining the future of Artificial Intelligence.

Readings & Media:
Movie:
Blade Runner **2 hours, group screening**

Book:
Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies* **Sections 8, 13, 15: 27 pages**

Journal Articles:
Lorenc, T. "Artificial Intelligence and the Ethics of Human Extinction" **21 pages**
Sun, Ron. "Potential of full human–machine symbiosis through truly intelligent cognitive systems" **12 pages**

Website:
Future of Life Institute. “Research Priorities for a Robust and Beneficial Artificial Intelligence: An Open Letter” **1 page**

Potential Guest Speakers:

Ronald Sandler, Professor of Philosophy and Religion at Northeastern University

Sandler is an associate professor of philosophy at Northwestern University and the director of the University’s Interdisciplinary Ethics Institute. He authored a textbook on the ethics of Artificial Intelligence entitled *Ethics and Emergent Technologies* and teaches classes in the philosophy of technology, ethical theory, and contemporary moral issues.

Kate Crawford, Director of AI Now Research Institute at New York University

Crawford is a Distinguished Research professor at New York University, a Principal Researcher at Microsoft Research New York, and a Visiting Professor at the MIT Media Lab. Her recent publications address data bias and fairness, social impacts of Artificial Intelligence, predictive analytics and due process, and algorithmic accountability and transparency. She is also the co-founder and co-director of the AI Now Research Institute, an interdisciplinary research center that studies the social impact of Artificial Intelligence.

James Hughes, Director of the Institute for Ethics and Emerging Technologies

Hughes, a bioethicist and sociologist, is the author of *Citizen Cyborg: Why Democratic Societies Must Respond to the Redesigned Human of the Future*, amongst other books. His work explores Artificial Intelligence in the future of human work and interaction. As the Director of the Institute for Ethics and Emerging Technologies, Hughes plays a primary role in determining real, impactful policies regarding Artificial Intelligence development.

Yuval Noah Harari, Professor of History, the Hebrew University of Jerusalem

Harari is an Israeli historian who specializes in the relationships between history and biology. His recent work looks at emergent technologies such as Artificial Intelligence, and critically analyzes the future of humanity grounded in a study of the past.

Proposed Bibliography, by Source Media:
We are more than happy to work with the Hurford Center to edit or reduce our proposed reading and media list

Journal Article:

Fazi, M. Beatrice. "Can a machine think (anything new)? Automation beyond simulation." 12 February 2018. https://doi.org/10.1007/s00146-018-0821-0

Harari, Yuval Noah. "Reboot for the AI Revolution." *Nature,* 550:7676. 17 October 2017. https://www.nature.com/news/reboot-for-the-ai-revolution-1.22826

Havlík, Vladimír. "The naturalness of artificial intelligence from the evolutionary perspective." *AI & Society,* 19 February 2018. https://doi.org/10.1007/s00146-018-0829-

Herzfeld, Noreen. "Creating in our own image: artificial intelligence and the image of God." *Zygon*, 37:2, 303-316. June 2002. http://search.ebscohost.com/login.aspx?direct=true&db=rfh&AN=ATLA0001287875&site=ehost-live

Lorenc, T. "Artificial Intelligence and the Ethics of Human Extinction." *Journal of Consciousness Studies*, 22:9, 194-214. 2015. http://www.ingentaconnect.com/content/imp/jcs/2015/00000022/F0020009/art00012

Shah, Huma, et. al., "Machine humour: examples from Turing test experiments." *AI & Society,* 07 June 2016. https://link.springer.com/article/10.1007/s00146-016-0669-0

Sun, Ron. "Potential of full human–machine symbiosis through truly intelligent cognitive systems." *AI & Society,* 28 November 2017. https://link.springer.com/article/10.1007/s00146-017-0775-7

Verhulst, Stefaan. "Where and when AI and CI meet: exploring the intersection of artificial and collective intelligence towards the goal of innovating how we govern." 21 February 2018. https://link.springer.com/article/10.1007/s00146-018-0830-z

Magazine Articles:

Bossman, Julia. "Top 9 Ethical Issues in Artificial Intelligence." *World Economic Forum,* 21 Oct 2016. https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/

Merritt, Jonathan. "Is AI a Threat to Christianity?" *The Atlantic,* 3 February 2017. https://www.theatlantic.com/technology/archive/2017/02/artificial-intelligence-christianity/515463

Solon, Olivia. "Artificial Intelligence is Ripe for Abuse, Tech Researcher Warns: 'A Fascist's Dream'." *The Guardian,* 13 March 2017.

https://www.theguardian.com/technology/2017/mar/13/artificial-intelligence-ai-abuses-fascism-donald-trump

Vlahos, James. "A Son's Race to Give His Dying Father Artificial Immortality." *Wired,* 18 July 2017. https://www.wired.com/story/a-sons-race-to-give-his-dying-father-artificial-immortality/

Book, Novel & Short Story:

Asimov, Isaac. *I, Robot.* New York City: Gnome Press, 1950. Print.

Aslan, Reza. *God: A Human History*. London: Transworld, 2017. Print.

Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2014. Print.

Mitchell, David. *Cloud Atlas*. London: Sceptre, 2004. Print.

Shelley, Mary. *Frankenstein*. London: Lackington, Hughes, Harding, Marvor, and Jones, 1818. Print.

Movie:

Blade Runner. Dir. Ridley Scott. Perf. Harrison Ford, Rutger Hauer, Sean Young, and Edward James Olmos. The Ladd Company, 1982.

Colossus: The Forbin Project. Dir. Joseph Sargent. Perf. Eric Braeden, Susan Clark, Gordon Pinset, William Schallert. Universal Pictures, 1970.

Her. Dir. Spike Jonze. Perf. Joaquin Phoenix, Amy Adams, Rooney Mara, Olivia Wilde, Scarlett Johansson. Annapurna Pictures, 2013.

TV Series:

Nolan, Jonathan and Joy, Lisa. (2016). *Westworld (TV Series 2016 -).* HBO Entertainment.

Video:

CGP Grey, "Humans Need Not Apply." 13 August 2014. https://www.youtube.com/watch?v=7Pq-S557XQU

Computerphile. “Why Asimov's Laws of Robotics Don't Work.” 6 November 2015. https://www.youtube.com/watch?v=7PKx3kS7f4A

Dennet, Daniel. “From Bacteria to Bach and Back: The Evolution of Minds.” *Google Talk,* 14 February 2017. https://www.youtube.com/watch?v=IZefk4gzQt4

Harari, Yuval Noah. “The Rise of Data Religion.” 12 September 2017. https://www.youtube.com/watch?v=Qry-yp33HoI

Government Report:

Executive Office of the President. “Artificial Intelligence, Automation, and the Economy.” 20 December 2016. https://permanent.access.gpo.gov/gpo75989/Artificial-Intelligence-Automation-Economy.PDF

Website:

Automato.farm. “Ethical Things.” http://automato.farm/portfolio/ethical_things/

Cesran International: Center for Strategic Research and Analysis. “The Current State Of AI As We Begin 2018.” https://cesran.org/the-current-state-of-ai-as-we-begin-2018.html

Future of Life Institute, “Benefits and Risks of Artificial Intelligence.” https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/

Future of Life Institute, “An Open Letter: Research Priorities For Robust And Beneficial Artificial Intelligence.” https://futureoflife.org/ai-open-letter/

MIT Media Lab, “Moral Machine.” http://moralmachine.mit.edu/

Space10, “Do you speak human?” https://space10.io/do-you-speak-human/

Potential Movies (for extra movie nights):

2001 Space Odyssey. Dir. Stanley Kubrick. Perf. Keir Dullea, Gary Lockwood. Metro-Goldwyn-Mayer, 1968

Ex Machina. Dir. Alex Garland. Perf. Domhnall Gleeson, Alicia Vikander, and Oscar Isaac. DNA Films, 2014.

Terminator. Dir. James Cameron. Perf. Arnold Schwarznegger, Michael Biehn, Linda Hamilton, and Paul Winfield. Pacific Western Productions, 1984.

Transcendence. Dir. Wally Pfister. Perf. Johnny Depp, Morgan Freeman, Rebecca Hall, Kate Mara, Cillian Murphy, Cole Hauser, Paul Bettany. Warner Bros. Pictures, 2014.

Suggested Further Reading:

Harari, Yuval Noah. *Homo Deus: A Brief History of Tomorrow*. London: Harvill Secker, 2015. Print.

Benjamin, Walter. *The Work of Art in the Age of Mechanical Reproduction*. Berlin: Independent Publishing Platform, 1936. Print.

Hall, Louisa. *Speak*. New York: HarperCollins, 2015. Print.

Brooker, Charlie. (2011). *Black Mirror (TV Series 2011 -)*. Netflix Series.

Turkle, Sherry. *The Second Self: Computers and the Human Spirit*. Cambridge: MIT Press, 2005. Print.

Asimov, Isaac. "The Last Question." *The Last Question.* New York: Columbia Publications, 1956. Print.

O'Gorman, Marcel. *Necromedia*. Minneapolis: University of Minnesota Press, 2015. Print.