AI and Existential Threats to Civilization

This course will discuss some fundamental problems in Artificial Intelligence (AI) and Machine Learning (ML), in the context of existential threats to civilization. We will consider the use of AI and ML to contribute to the fight against climate change and global pandemics. We will also consider the role of AI and ML, both positive and negative, in the threat of nuclear war. And we will consider whether AI and ML themselves pose a threat to civilization, including whether AI with general intelligence is a risk, and issues of fake news and deep fakes.

This class will be a seminar and is not a qualifying course for MS or PHD students. We will study some fundamental techniques of modern machine learning, including reinforcement learning, the use of deep learning in image analysis, and generative adversarial networks. We will read technical CS papers, but also read work from other fields to provide context. Students will develop a research proposal to address some of the issues raised in class.

Assignments

Students will work throughout the semester to develop a research proposal addressing one of the issues discussed in class. After class 12, a five page white paper describing their proposal will be due. At the end of the semester, a complete 15 page NSF-style proposal will be due, in NSF format https://www.nsf.gov/pubs/policydocs/pappguide/nsf16001/gpg_index.jsp. We will distribute sample NSF proposals. The proposal should have a compelling overall vision, take account of all relevant past work to explain what is innovative, and provide a detailed description of proposed work. The best proposals will often contain persuasive initial results. Students may work on these proposals in groups of up to three students. The more students who are working on a common proposal, the better we expect the proposal to be.

After class 16, students will be required to prepare a half hour talk on their proposal. The talk should emphasize a description of the prior work, background and motivation for the proposal. But it should also contain a persuasive pitch for the proposed work. Presentations will be posted on youtube. The TAs and professor will select several presentations to be given live to the class. Students selected for these presentations will receive extra credit.

In addition, students will be required to turn in one page summaries of some of the papers assigned. Each summary should contain two paragraphs. The first paragraph should summarize the paper. The second paragraph should provide some critical insight into the paper. These summaries are due prior to the class in which the paper will be discussed; late summaries will not be accepted.

Recording and Online Privacy
Our class sessions may be recorded for use by enrolled students, including those who attend online. Students who participate with their camera engaged or utilize a profile image are consenting to have their video or image recorded. If you are unwilling to consent to have your profile or video image recorded, be sure to keep your camera off and do not use a profile image. If you are not willing to consent to have your voice recorded during class, you will need to keep your mute button activated and communicate exclusively using the "chat" feature.

Guides that may be useful for online instruction

- **Campus Resources for Students**
- **Resources for Students Learning Online**
- **Best Practice for Learning Online**
- **Navigating Zoom as a Student**
- Managing Technical Difficulties

Basic Needs & Security
If you have difficulty affording groceries or accessing sufficient food to eat every day, or lack a safe and stable place to live, please visit **UMD's Division of Student Affairs website** for information about resources the campus offers you and let me know if I can help in any way.

Disability Support
Any student eligible for and requesting reasonable academic accommodations due to a disability is requested to provide, to the instructor in office hours, a letter of accommodation from the Office of Disability Support Services (DSS) *within the first TWO weeks of the semester*.

Academic Integrity
In this course you are responsible for both the University's Code of Academic Integrity and the University of Maryland Guidelines for Acceptable Use of Computing Resources. Any evidence of unacceptable use of computer accounts or unauthorized cooperation on tests, quizzes and assignments will be submitted to the Student Honor Council, which could result in an XF for the course, suspension, or expulsion from the University.

Any work that you hand in must be your own work. Any sources that you draw from, including other students, should be appropriately acknowledged. Plagiarism is a serious offense, and will not be tolerated.

Anti-Harassment
The open exchange of ideas, the freedom of thought and expression, and respectful scientific debate are central to the aims and goals of this course. These require a community and an environment that recognizes the inherent worth of every person and group, that fosters dignity, understanding, and mutual respect, and that embraces diversity. Harassment and hostile behavior are unwelcome in any part of this course. This includes: speech or behavior that intimidates, creates discomfort, or interferes with a person's participation or opportunity for participation in the course. We aim for this course to be an environment where harassment in any form does not happen, including but not limited to: harassment based on race, gender, religion, age, color, national origin, ancestry, disability, sexual orientation, or gender identity. Harassment includes degrading verbal comments, deliberate intimidation, stalking, harassing photography or recording, inappropriate physical contact, and unwelcome sexual attention. Please contact an instructor or CS staff member if you have questions or if you feel you are the victim of harassment (or otherwise witness harassment of others).

Course evaluations
We welcome your suggestions for improving this class, please don't hesitate to share it with the instructor or the TAs during the semester! You will also be asked to give feedback using the CourseEvalUM system at the end of the semester. Your feedback will help us make the course better.

Office Hours

Prof. Jacobs will have office hours on Tuesday, 2-3. A zoom link can be found in the zoom section of ELMS. In addition, students should feel free to schedule meetings with Prof. Jacobs at other times.

Class Schedule

This schedule is tentative. We welcome suggestions from the class for additional topics, papers to discuss, or guest speakers. We expect that things may change quite a bit.

There is no text; readings are linked to below. Some films that are relevant to the course include:
1984 (I've only seen the version released in 1984, which I recommend)
WALL-E
Blade Runner
The Terminator
Ex Machina
(More suggestions welcome)

	Topic	Assigned Reading
1	Introduction	
2	Introduction to Climate Change	IPCC Fifth Assessment https://archive.ipcc.ch/report/ar5/wg1/ Read summary for policymakers Short summary from Royal Society, 2020: https://royalsociety.org/~/media/royal_society_content/policy/projects/climate-evidence-causes/climate-change-evidence-causes.pdf Tackling climate change with machine learning https://arxiv.org/pdf/1906.05433.pdf Ezra Klein also has an excellent series of podcast interviews on climate change. Some links are at: https://www.vox.com/podcasts/2019/12/16/210243 saul-griffith-solve-climate-change
	Intro to ML and Neural Nets	If you are unfamiliar with machine learning, Hal's book (http://ciml.info/) provides a good undergraduate introduction. You should probably read it. The Deep Learning book (https://www.deeplearningbook.org/) provides a comprehensive discussion of deep learning. Chapter 5 provides an introduct serve as a reference for this lecture. Chapters 6-9 will introduce concepts in Deep Learning that we'll use in class. http://neuralnetworksanddeeplearning.com/ provides a short, very clear introduction to neural networks.
	CNNs	Lecture
	Satellite image analysis for forest assessment	https://www.nature.com/articles/s41586-020-2824-5?fbclid=IwAR3YO4sZA_RfEFBa-GA_VPZWK7WWi2kRsdvJv44dA3U6FlbqtwngU9TK1Q4 https://openaccess.thecvf.com/content_ICCV_2017/papers/He_Mask_R-CNN_ICCV_2017_paper.pdf https://www.nature.com/articles/s41586-020-2824-5?fbclid=IwAR3YO4sZA_RfEFBa-GA_VPZWK7WWi2kRsdvJv44dA3U6FlbqtwngU9TK1Q4
	Open Catalyst	Guest Speaker, Larry Zitnick https://arxiv.org/pdf/2010.09435.pdf
	AI and nuclear war	https://www.rand.org/pubs/perspectives/PE296.html?source=post_page---------- https://www.tandfonline.com/doi/abs/10.1080/0163660X.2020.1770968?journalCode=rwaq20
	Fake news overview	**https://science.sciencemag.org/content/sci/359/6380/1094.full.pdf?casa_token=LiNuhDd_9IQAAAAA:bL5BheLfx4evJduE5-DgY3RbaZLWQ04XD8IWziY6VfTtndBDIKZ36_WWCK8eNHAu-7I** **https://dl.acm.org/doi/pdf/10.1145/3137597.3137600?casa_token=lYmS5fzwFZIAAAAA:c6Nfbk_4lpc1gD_XcVVfFGKthRHtf9Yh4f8m-L_IuSpp6cjbDvzDHbxK__ej8TigZFY_YxFZHuSr1Q** https://www.technologyreview.com/2020/01/08/130983/were-fighting-fake-news-ai-bots-by-using-more-ai-thats-a-mistake/ Yellow journalism https://daily.jstor.org/to-fix-fake-news-look-to-yellow-journalism/ Politics and the English Language, Orwell https://www.orwellfoundation.com/the-orwell-foundation/orwell/essays-and-other-works/politics-and-the-engli I also highly recommend reading the book 1984. There is also a good movie of it: https://www.imdb.com/title/tt0087803/?ref_=fn_al_tt_1
	AI and Policy	Guest, Chris Meserole, Brookings Institution https://www.brookings.edu/wp-content/uploads/2019/08/FP_20190826_digital_authoritarianism_polyakova_meserole.pdf https://www.brookings.edu/blog/order-from-chaos/2018/05/25/the-west-is-ill-prepared-for-the-wave-of-deep-fakes-that-artificial-intelligence-could-unlea
	Reinforcement Learning	Lecture -- http://incompleteideas.net/book/the-book.html is an excellent text on RL. Reading the first six chapters will give you a good intro.
	RL and Data Centers	Transforming Cooling Optimization for Green Data Center via Deep Reinforcement Learning https://ieeexplore.ieee.org/stamp/stamp.jsp? arnumber=8772127&casa_token=loYZjw7FEWAAAAAA:gP1GUt89HAZ_zL5xLr_MqglAGttvqKVA9PjnZEMAeaDyFX5ui4BRlGttpqnLkFOGHSFRuSOk http://papers.nips.cc/paper/7638-data-center-cooling-using-model-predictive-control.pdf
	GANs	Lecture
	GANs and Deep Fakes	Papers – TBD CNN-generated images are surprisingly easy to spot... for now
	Text embeddings	Lecture Word2vec, Transformers An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale
	Story generation	Bert, GPT-3 http://faculty.washington.edu/ebender/papers/Stochastic_Parrots.pdf On the dangers of Stochastic Parrots

	Fake News Detection	https://www.aclweb.org/anthology/D17-1317.pdf https://papers.nips.cc/paper/2019/file/3e9f0fc9b2f89e043bc6233994dfcf76-Paper.pdf Defending against neural fake news – GROVER Defending Against Neural Fake News, "Liar, Liar Pants on Fire"A New Benchmark Dataset for Fake News Detection
	Protein folding	Overview, ML and covid https://www.nature.com/articles/s42256-020-0181-6 Survey on computational protein folding https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6790072/ Alphafold https://dasher.wustl.edu/bio5357/discussion/nature-577-706-20.pdf Blog post on alphafold and covid https://deepmind.com/research/open-source/computational-predictions-of-protein-structures-associated-with-COVID-19
	Covid	https://www.pnas.org/content/117/41/25904 Modeling between-population variation in COVID-19 dynamics in Hubei, Lombardy, and New York City https://journals.plos.org/plosone/article/authors?id=10.1371/journal.pone.0239474
	Is AI a threat?	**https://edisciplinas.usp.br/pluginfile.php/4906801/mod_resource/content/1/Yuval%20Noah%20Harari%20on%20Why%20Technology%20Favo%20The%20Atlantic.pdf** TBD
	AI and the Singularity	http://lib.21h.io/library/CNKLVMSA/download/LQAYX97Q/2020_Guide_To_Deep_Learning_Basics_-_Logical%2C_Historical_And_Philosophical_Sprin https://www.researchgate.net/profile/Rodney_Brooks/publication/3001510_I_Rodney_Brooks_am_a_robot/links/5559daec08aeaaff3bfa3bb3/I-Rodney-robot.pdf https://reader.elsevier.com/reader/sd/pii/S0004370207001464?token=6308AD9B5898B4500632F9C9B66DC95D521819F5BCBCD0B39A9C0C60BA663F1B1A6A688812D7E0861CE05CC0D463CABB https://www.youtube.com/watch?v=qc4v7AvqigU Section 2, beginning at 21.50. However, the whole video is worthwhile.
	Wind farms	Real-time optimization of wind farms using modifier adaptation and machine learning **https://wes.copernicus.org/articles/5/885/2020/**