A 7-Dimensional Model of AI Risk

Colin Garvey

PhD Candidate, Science & Technology Studies Dept.
Rensselaer Polytechnic Institute

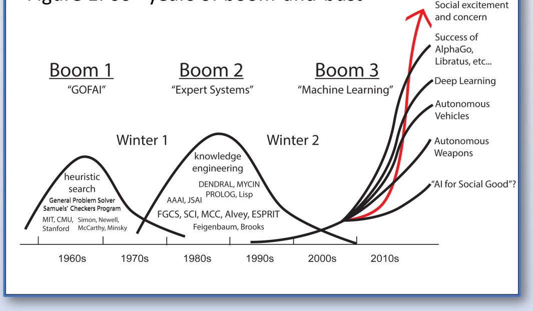


Figure 1. 60+ years of boom-and-bust

ABSTRACT

Concerns about the negative social impacts of AI have been growing in recent years (Fig. 1) as the rapid technological developments Boom 3 produce benefits for some and risks for others. High-profile figures in the tech industry disagree about the risks, with Elon Musk stoking fears and Mark Zuckerberg denying their reality. In the frenzied media coverage of this debate more heat than light has been generated, leaving most people still wondering if AI is dangerous, or not. The disagreement and uncertainty about the risks of AI raise the following two questions, which my dissertation research seeks to answer: What are the risks of AI? And what can be done to mitigate them? In this poster, I present the 7-dimensional model of AI risk I developed to answer the first question. My other poster ("A Framework for the Evaluation of Barriers to the Democratization," AAAI Student Track, Tues Feb 6th, 6:30pm) explains the framework for the democratic governance of technological R&D that I use to answer the second.

Methods

In addition to news coverage and the scholarly literature on AI and robotics, data sources analyzed for this project include: primary documents from AI-focused institutions and tech companies; AI policy documents from governments and private organizations; online ethnographic observation in public fora like Twitter as well as private communities; interviews with technical experts, social scientists, and laypeople; as well as participant observation at AI conferences and laboratories in the USA and Japan. In fact, being here right now at the AAAI/AIES conference is part of my anthropological "fieldwork!"

The State of Risk in AI

What types of AI risk are being created, and by whom? Historically the field of AI has paid little attention to risk [2]. This changed in 2014 when Stephen Hawking began sounding the alarm about the threat AI posed to humankind. More prominent figures quickly followed suit, while others rushed to defend AI. Almost all framed AI impacts in terms of ambivalent extremes—either utopia or dystopia, heaven or hell. These initial conditions locked the emerging conversation into a trajectory that stifled more nuanced views even as the issue gained media attention. Utopians play down or ignore risks, focusing exclusively on potential benefits, while dystopians see only "existential risk," the danger that AI will somehow make humanity extinct [3]. This extremism impairs the publics' ability to understand the risks of this emerging technology and leaves little room to steer AI toward robustly beneficial futures for a majority of humanity. Many now say either nothing needs to be done, or nothing can be done. My dissertation project attempts to disrupt this dichotomous framing by articulating seven dimensions of AI risk.

Economic

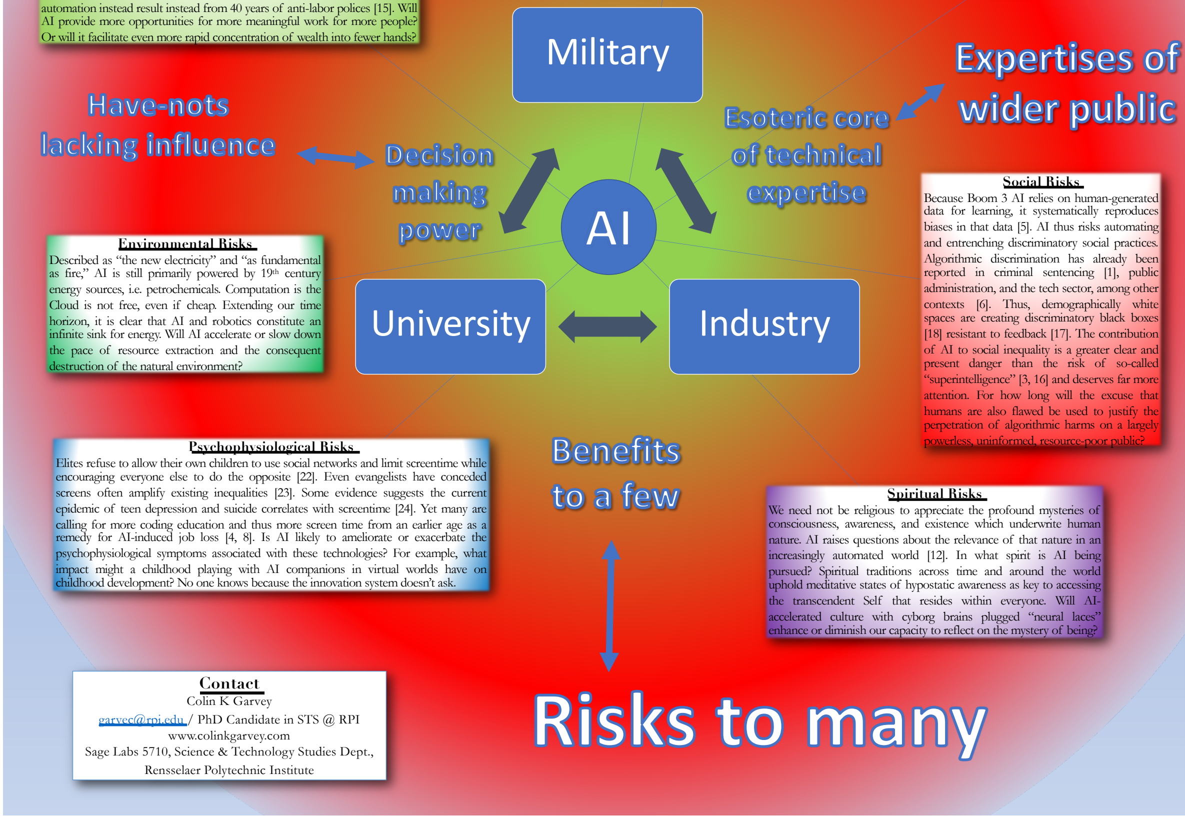
Many have argued AI threatens jobs [3, 4, 8]. The most-cited figure is that "47% of the US workforce is at risk of automation" [9]. Though other studies offer different numbers, the wide variation in quantitative estimates highlights experts' uncertainty about the economic risks of AI. A survey of over 1900 hundred AI scientists found them evenly split over whether AI will improve or destroy the economy [19]. With ample evidence that economic inequality is increasing globally [14], at the very least, AI only needs to support the status quo in order to amplify the trend. Some argue this is the inevitable result of technological evolution [3, 4]; something we should adapt to. Others argue that technology is malleable, and that the negative effects of automation instead result instead from 40 years of anti-labor polices [15]. Will AI provide more opportunities for more meaningful work for more people? Or will it facilitate even more rapid concentration of wealth into fewer hands?

Military

Bracketing Terminator-like scenarios altogether, the military applications of AI still pose serious risks to humanity. Led by the USA, China, and Russia, national militaries are producing a new generation of Autonomous Weapons Systems (AWSs). Proponents argue they will save lives, but experience with semi-autonomous weapons in the US "drone war" suggests AWSs are likely to introduce as many new problems as they solve. Not only are risks of malfunction and hacking nontrivial, "arms race" dynamics have already taken hold [13]. The US military's superiority in AI and robotics may soon be challenged, as China's multi-billion dollar investments in AI reflect the nation's intention to gain dominance in AI through "military-civil fusion" by 2030 [7]. Open letters calling for a ban on AWSs were garnered many signatures when they were introduced in 2015 and 2017, and the Campaign to Stop Killer Robots continues to lobby the UN. But prospects for a ban look dim with no major powers in support. Some experts continue to deny the salience of the risk altogether [25].

Political

AI technologies provide unprecedented tools for elites to manipulate opinion and exploit have-nots [17, 18]. The 2016 US presidential election provided a powerful recent example. The AI technologies powering Facebook's newsfeeds and Google's search results led to partisan isolation, keeping voters in private "echo chambers"; right-wing groups used AI to rapidly disseminate "fake news" and divisive messages designed to stoke suspicion of certain ethnic and religious groups; new modeling techniques allowed for "micro-targeting" of specific demographics most susceptible to manipulation [11]. Called before congressional hearing, the tech-titans have begun to admit some responsibility for the problem [20]. The geo-political stakes are high. US Senator Lindsey Graham recently said, "these technologies also can be used to undermine our democracy and put our nation at risk" [10]. Russian leader Vladmir Putin has asserted, "Whoever leads in AI will rule the world" [21]. Whether Elon Musk is correct that AI will be the cause of WW3 or not, at the least, AI technologies increasingly underwrite the "post-truth era" throwing American democracy into crisis.



Military
University
Industry
AI
Have-nots lacking influence
Decision making power
Esoteric core of technical expertise
Expertises of wider public
Benefits to a few
Risks to many

Social Risks

Because Boom 3 AI relies on human-generated data for learning, it systematically reproduces biases in that data [5]. AI thus risks automating and entrenching discriminatory social practices. Algorithmic discrimination has already been reported in criminal sentencing [1], public administration, and the tech sector, among other contexts [6]. Thus, demographically white spaces are creating discriminatory black boxes [18] resistant to feedback [17]. The contribution of AI to social inequality is a greater clear and present danger than the risk of so-called "superintelligence" [3, 16] and deserves far more attention. For how long will the excuse that humans are also flawed be used to justify the perpetration of algorithmic harms on a largely powerless, uninformed, resource-poor public?

Environmental Risks

Described as "the new electricity" and "as fundamental as fire," AI is still primarily powered by 19th century energy sources, i.e. petrochemicals. Computation is the Cloud is not free, even if cheap. Extending our time horizon, it is clear that AI and robotics constitute an infinite sink for energy. Will AI accelerate or slow down the pace of resource extraction and the consequent destruction of the natural environment?

Psychophysiological Risks

Elites refuse to allow their own children to use social networks and limit screentime while encouraging everyone else to do the opposite [22]. Even evangelists have conceded screens often amplify existing inequalities [23]. Some evidence suggests the current epidemic of teen depression and suicide correlates with screentime [24]. Yet many are calling for more coding education and thus more screen time from an earlier age as a remedy for AI-induced job loss [4, 8]. Is AI likely to ameliorate or exacerbate the psychophysiological symptoms associated with these technologies? For example, what impact might a childhood playing with AI companions in virtual worlds have on childhood development? No one knows because the innovation system doesn't ask.

Spiritual Risks

We need not be religious to appreciate the profound mysteries of consciousness, awareness, and existence which underwrite human nature. AI raises questions about the relevance of that nature in an increasingly automated world [12]. In what spirit is AI being pursued? Spiritual traditions across time and around the world uphold meditative states of hypostatic awareness as key to accessing the transcendent Self that resides within everyone. Will AI-accelerated culture with cyborg brains plugged "neural laces" enhance or diminish our capacity to reflect on the mystery of being?

References

1. Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. "Machine Bias." *ProPublica*.
2. Barrat, James. 2013. *Our Final Invention*. New York: Thomas Dunne Books.
3. Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
4. Brynjolfsson, Erik, and Andrew McAfee. 2012. *Race against the Machine*. Lexington, Mass.: Digital Frontier Press.
5. Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. 2017. "Semantics Derived Automatically from Language Corpora Contain Human-like Biases." *Science* 356 (6334):183–186.
6. Crawford, Kate. June 25, 2016. "Artificial Intelligence's White Guy Problem." *The New York Times.*
7. The Economist. "War at Hyperspeed: Getting to Grips with Military Robotics." *The Economist*, January 25, 2018.
8. Ford, Martin. 2015. *Rise of the Robots: Technology and the Threat of a Jobless Future*. New York: Basic Books, a member of the Pereus Books Group.
9. Frey, Carl Benedikt, and Michael A. Osborne. 2013. "The Future of Employment: How Susceptible Are Jobs to Computerisation." Oxford Martin School of Business.
10. Graham, Lindsey. "Opening Statement At Hearing On Russian Disinformation, Extremist Content Online." October 31, 2017.
11. Halpern, Sue. "How He Used Facebook to Win." *The New York Review of Books*, June 8, 2017
12. Harari, Yuval Noah. *Homo Deus: A Brief History of Tomorrow*. Vintage, 2017.
13. Kania, Elsa. March 9, 2017. "The Next U.S.-China Arms Race: Artificial Intelligence?" Text. The National Interest.
14. Kottasová, Ivana. "World's Richest 1% Grabbed 82% of All Wealth Created in 2017, Oxfam Study Finds." CNNMoney, January 21, 2018.
15. Mishel, Lawrence, and Josh Bivens. 2017. "The Zombie Robot Argument Lurches on: There Is No Evidence That Automation Leads to Joblessness or Inequality." Washington, DC: Economic Policy Institute.
16. Müller, Vincent C., ed. 2016. *Risks of Artificial Intelligence*. Boca Raton, FL: CRC Press.
17. O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Broadway Books.
18. Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press.
19. Pew Research Center. "AI, Robotics, and the Future of Jobs." Digital Life in 2025. Pew Research Center, August 2014.
20. Phillips, Kristine, and Brian Fung. "Facebook Admits Social Media Sometimes Harms Democracy." *Washington Post*, January 22, 2018.
21. RT Int'l. "'Whoever Leads in AI Will Rule the World': Putin to Russian Children on Knowledge Day." *RT International*, September 1, 2017.
22. Selk, Avi. "Apple CEO Tim Cook Says He Wouldn't Let a Child Use Social Media." *Washington Post*, January 21, 2018.
23. Toyama, Kentaro. 2015. *Geek Heresy: Rescuing Social Change from the Cult of Technology*. New York: PublicAffairs.
24. Twenge, Jean M. "Have Smartphones Destroyed a Generation?" *The Atlantic*, September 2017
25. Williams, Chris. "AI Guru Ng: Fearing a Rise of Killer Robots Is like Worrying about Overpopulation on Mars." News. *The Register*, March 19, 2015.

Contact

Colin K Garvey
garvec@rpi.edu / PhD Candidate in STS @ RPI
www.colinkgarvey.com
Sage Labs 5710, Science & Technology Studies Dept.,
Rensselaer Polytechnic Institute