SciencesPo
PARIS SCHOOL OF INTERNATIONAL AFFAIRS

COURSE DESCRIPTION

GOVERNING THE RISE OF ARTIFICIAL INTELLIGENCE: A GLOBAL PERSPECTIVE

Short biography: Nicolas co-founded The Future Society in 2014 and incubated it at the Harvard Kennedy School of Government. Incorporated as a 501 (c) (3) in Massachusetts, the think-and-do tank specializes in questions of impact and governance of emerging technologies (starting with Artificial Intelligence through its 'AI Initiative' launched in 2015), innovation (especially urban innovation through its 'CitiX' initiative launched in 2016) and new media. A recognized thought-leader, strategist and implementer, Nicolas advises cities, governments, international organizations, NGOs and industrial players. He has over fifteen years of professional experience building partnerships across Europe, America and Asia. Nicolas is the co-Convener of the AI Civic Forum (AICF), the Global Governance of AI Roundtable (GGAR) organized yearly during the World Government Summit in Dubai. He is also a member of the AI Group of experts at OECD, of the World Bank's Digital Economy for All Initiative (DE4ALL), and of the Global Council on Extended Intelligence (MIT Media Lab). Nicolas teaches at the Paris School of International Affairs (Sciences Po), at the IE School of Global and Public Affairs in Madrid, and at the Mohamed bin Rashid School of Government in Dubai. He is also a Senior Visiting Research Associate with the Program on Science, Technology and Society at Harvard Kennedy School, and a Fellow with the Center for the Governance of Change at IE Business School in Madrid. Nicolas is a member of three Committees (Policy, Economics, and General Principles) of the IEEE Global Initiative on Ethically Aligned Design of Autonomous & Intelligent Systems. An Arthur Sachs Scholar, Nicolas holds a Master in Public Administration from Harvard Kennedy School, a Master in Defense, Geostrategy & Industrial Dynamics from Panthéon-Assas University, and a Bachelor of Arts in European Affairs and International Relations from Sciences Po Strasbourg.

Course description: Based on a mix of lectures and group discussions, this course will equip students with a foundational understanding of the dynamics of the rise of Artificial Intelligence and its consequences, as well as how they need to be governed to maximize benefits, minimize risks and ensure that the benefits reach everyone.

If the definitional boundaries of Artificial Intelligence (AI) remain contested, experts agree that we are witnessing a global revolution. "Is this time different?" is the question that they worryingly argue over when analyzing the socio-economic impact of the AI revolution as compared with the three previous industrial revolutions of the 19th and 20th centuries. Like before, this Schumpeterian wave may prove to be a *creative destruction* raising incomes, enhancing quality of life for all and generating previously unimagined jobs to replace those that get automatized. Or for the first time it may turn out to be a *destructive creation* leading to mass unemployment, abuses, or loss of control over decision-making processes. This depends on the velocity and magnitude of the development and diffusion of AI technologies, a complex question over which experts diverge widely. Moreover, societies' abilities to shape the AI revolution into a "creative destruction" and diffuse its benefits to all will mostly depend on how societies react, both individually and collectively.

Throughout the duration of the course, we'll see that technology is certainly not destiny, and that policy as well as institutional choices will matter greatly. We'll explore the benefits expected from the AI revolution: a wave of productivity gains with the potential to sustain growth and development over the next decades, counterbalancing the decreasing working-age population; enhanced quality of life for all, through revolutions in healthcare, transportation, education, security, justice, agriculture, retail, commerce, finance, insurance and banking, as well as other domains.

We'll then explore how making the AI revolution work for everyone will require the reform and the potential reinvention of social security, redistribution mechanisms, as well as education, training and

skill development systems, to allow for repeated and viable professional transitions. We'll also discuss the re-balancing of policy and regulatory frameworks needed to protect the most vulnerable from socio-economic exclusion, to prevent algorithmic discrimination and privacy abuses, to ensure cybersecurity, safety, explainability, control and accountability, as well as to avoid an exacerbation of wealth and opportunity inequalities.

Finally, the course will present the challenges and opportunities associated with the need for a more active international coordination to harmonize regulation, value-systems and to ethically align design principles of AI systems. Analyzing AI as a matter of power and sovereignty, the course will also discuss the tension between the need for scale and excessive power concentration tendencies, and explore possible solutions to rein-in adverse competition dynamics.

Course requirements & Grading policy:

Grades will be calculated as follows:

- *Class Participation: Every student is expected to be prepared for and attend every class, and to participate in the discussions.* ***(30%)***
- *Final Policy Paper: Students get in group of three to write one short (1000–1600 word) policy memo, recommending an AI transformation strategy for a country, region or city from the list provided* ***(50%)***
- *Students will have to write individually 1 essay (500–800 words) during the semester reflecting on the readings for one session* ***(20%).*** *The essays will be discussed in class.*

Students are required to participate in class discussions, and to hone their analytical, research, and writing skills through the written assignment. Students are expected to: 1) attend all classes; 2) be on time; 3) respect the no laptops no cell phones in class policy (except when needed for class exercise), 4) submit assignments on time; 5) be respectful of each other and of the instructor; 6) be prepared to be cold-called; and 7) do their best to prepare professional products for their assignments.

Session 1: **Defining Artificial Intelligence (04/02)**

Required readings:

Nicolas Miailhe & Cyrus Hodes, Making the AI revolution work for everyone, The Future Society, AI Initiative, Report to OECD, February 2017 – PART 1, CHAPTER A

Russell, Stuart (2016) "Q&A: The Future of Artificial Intelligence". *University of Berkeley*.

Future of Life institute (2016) "The Top Myths about Advanced AI", http://futureoflife.org/background/aimyths/

Urban, Tim (2015), 'The Artificial Intelligence Revolution: The Road to Superintelligence' *Wait But Why*: http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html

Future of Life Institute, "AI Policy Challenges & Recommendations," https://futureoflife.org/ai-policy-challenges-and-recommendations.

Recommended readings:
Andrew Ng, "The State of Artificial Intelligence," (Video).

Google Cloud Platform, "What is machine learning?" (Video).

JASON, The MITRE Corporation, *Report on Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to DoD*, January 2017. https://fas.org/irp/agency/dod/jason/ai-dod.pdf - Executive Summary & Introduction, skim rest of report.

Google's Deep Mind Explained! - Self Learning A.I. (https://www.youtube.com/watch?v=TnUYcTuZJpM)

What is a Deep Neural Network (https://www.youtube.com/watch?v=aircAruvnKk) and how it learns (https://www.youtube.com/watch?v=IHZwWFHWa-w)

Session 2: The Geopolitics of AI (07/02)

Required readings:

Nicolas Miailhe, R. Buse Çetin, Caroline Jeanmaire, "The Geopolitics of Artificial Intelligence: The Return of Empires?" Politique étrangère, Vol. 83, No. 3, Autumn 2018, (English version in class folder)

Nicolas Miailhe & Cyrus Hodes, Making the AI revolution work for everyone, The Future Society, AI Initiative, Report to OECD, February 2017 – PART 1, CHAPTER B

Manuel Muniz, Marietje Schaake, (2019), "Making the Future Work for Us". Project Syndicate," p. 3.

Grace, Katja, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans. (2017) "When Will AI Exceed Human Performance? Evidence from AI Experts." *arXiv:1705.08807 [Cs]* , May 24, 2017. http://arxiv.org/abs/1705.08807

Recommended readings:

Jessi Hempel, (2017), How Baidu Will Win China's AI Race - and, Maybe, the World's," Wired.

Cremer, J., de Montjoye, Y-A. and H. Schweizer, 2019, "Competition Policy for the Digital Era", European Commission - executive summary page 2-11.

Nicolas Wright, (2018), "How Artificial Intelligence will Reshape the Global Order", Foreign Affairs.

"The Race For AI: Google, Baidu, Intel, Apple In A Rush To Grab Artificial Intelligence Startups," CBInsights, 2017

McKinsey Global Institute, Artificial intelligence, the next digital frontier, June 2017 – Appendix

Good, Irving J. (1965). "Speculations Concerning the First Ultraintelligent Machine," Advances in Computers, (6) 99, 31-83. http://www.kushima.org/is/wp-content/uploads/2015/07/Good65ultraintelligent.pdf | *classic work which first coined the concept of an 'intelligence explosion'*

Ian Hogarth, (2018), "AI Nationalism", Ian Hogarth.

Session 3: **Efficiency of public and private management and new waves of productivity gains, economic growth and revolution in sectors (25/02)**

Required readings:

Nicolas Miailhe & Cyrus Hodes, Making the AI revolution work for everyone, The Future Society, AI Initiative, Report to OECD, February 2017 – PART 2, CHAPTER A, B, C, D, E, F

DeepMind Blog, 'AI reduces Google data centre cooling bill 40 percent."

The Economist (2018), "How AI is spreading throughout the supply chain"

Ajay Agrawal, Joshua Gans, and Avi Goldfarb, "The Simple Economics of Machine Intelligence", Harvard Business Review, November 2016.

James Manyika, Michael Chui, Mehdi Miremadi, Jacques Bughin, Katy George, Paul Willmott, and Martin Dewhurst, *Harnessing Automation for a Future that Works*, McKinsey Global Institute, January 2017 – Five Case Studies (skim)

Recommended readings:

Mark Purdy & Paul Daugherty, How AI boosts industry profits and innovation, Accenture, 2017 (skim)

Ajay Agrawal, Joshua Gans, and Avi Goldfarb. (2018). Prediction Machines: The Simple Economics of Artificial Intelligence. Harvard Business Review Press.

Mark Purdy & Paul Daugherty, (2016) Why Artificial Intelligence is the future of growth, Accenture - (skim)

Ajay Agrawal, Joshua Gans, Avi Goldfarb, (2018), "Economic Policy for AI," Vox CEPR Policy Portal.

Jason Furman and Robert Seamans, (2018), AI and the Economy, NBER Working Paper 24689- (Introduction and Conclusion)

McKinsey Global Institute (2017), A future that works: automation, employment and productivity– Executive Summary and Chapter 5

McKinsey Global Institute, (2017), Artificial intelligence, the next digital frontier, – All except Appendix (Skim)

Wadhwa, Vivek. 2017. The Driver in the Driverless Car: How Our Technology Choices will Create the Future, (Chapters 6, 7, 9-12)

PwC, (2017), *What Doctor? Why AI and Robotics Will Define New Health*.

Session 4: Jobs lost and jobs created 1/2 (3/03)

Required readings:
'Managing the Economic and Social Impacts of the AI Revolution - Part I' (PDF in class folder)

McKinsey Global Institute, Jobs lost jobs gains: workforce transitions in a time of automation, December 2017 – In Brief and Summary of Findings, skim Chapters 1-3

Calum Chace (2018), Our Jobless Future: Essays on Artificial Intelligence and the Economic Singularity. (skim)

Winick, E. 2018. "Every study we could find on what automation will do to jobs, in one chart." MIT Technology Review.

Recommended readings:

Erik Brynjolfsson, Andrew McAfee, The Second Machine Age – Work, Progress and Prosperity in a Time of Brilliant Technology, MIT Press, 2014

Session 5: Jobs lost and jobs created 2/2 (10/03) Guest Lectures: Luis Aranda, Artificial Intelligence and Social Inclusion Economist at the OECD & Andrew Green, Labour market Economist at the OECD.

Required readings:
McKinsey Global Institute, Jobs lost jobs gains: workforce transitions in a time of automation, December 2017 – Skim Chapters 4-6

Jae-Hee Chang and Phu Huynh, (2016), *ASEAN in Transformation: The future of jobs at risk of automation,* ILO. (skim)

James Manyika, Michael Chui, Mehdi Miremadi, Jacques Bughin, Katy George, Paul Willmott, and Martin Dewhurst, Harnessing automation for a future that work, McKinsey Global Institute, January 2017 – Executive Summary

Recommended readings:

Muro, M., Maxim, R, and Whiton, J. 2019, *Automation and Artificial Intelligence: How machines are affecting people and places.* Metropolitan Policy Program at Brookings.

Melanie Arntz, Terry Gregory, and Ulrich Zierahn, (2016). *The risk of automation for jobs in OECD countries: A comparative analysis*, OECD Social, Employment and Migration working paper number 189, OECD. (Skim)

Session 6: Workshop: Adapting Social Security & Income redistributive mechanisms (19/03)

Break-out class discussion and policy analysis of trade-offs, objectives, feasibility and costs

Required readings:

Nicolas Miailhe & Cyrus Hodes, Making the AI revolution work for everyone, The Future Society, AI Initiative, Report to OECD, February 2017 – PART 3, CHAPTER C

'Managing the Economic and Social Impacts of the AI Revolution - Part II' (PDF in class folder)

James Manyika, Michael Chui, Mehdi Miremadi, Jacques Bughin, Katy George, Paul Willmott, and Martin Dewhurst, *Harnessing Automation for a Future that Works*, McKinsey Global Institute, January 2017 – Chapter 6

Exponential View: #6 UBI, Automation & Society in the US: Andrew Yang & Azeem Azhar in Conversation

Abhijit Banerjee, Paul Niehaus, Tavneet Suri (2019), Universal basic income in the developing world, MIT and UC San Diego.

Esther Duflo and Abhijit Banerjee (2019), Economic Incentives Don't Always Do What We Want Them To, On their own, markets can't deliver outcomes that are just, acceptable — or even efficient., New York Times.

Recommended readings:

Kai-Fu Lee. 2018. *AI Superpowers : China, Silicon Valley, and the New World Order.* Houghton Mifflin Harcourt. (Chapter 8.)

James Manyika, Michael Chui, Mehdi Miremadi, Jacques Bughin, Katy George, Paul Willmott, and Martin Dewhurst, *Harnessing Automation for a Future that Works*, McKinsey Global Institute, January 2017 – Rest of the Report

Session 7: Ethical challenges of an increasing delegation to autonomous agents (24/03)
Guest Lecture: Raja Chatila (Fellow of the Institute of Electrical and Electronics Engineers)

Required readings:

Nicolas Miailhe & Cyrus Hodes, Making the AI revolution work for everyone, The Future Society, AI Initiative, Report to OECD, February 2017 – PART 3, CHAPTER B

Yolanda Lannquist. 2018. "Policy & Ethics in AI". Presentation at ReWork AI for Government Summit, Toronto, Canada. (20-min video): http://videos.re-work.co/videos/1197-ai-policy-ethical-dilemmas

Crawford, et al. (2019), AI Now 2019 Report, New York: AI Now – Executive Summary & Intro

Kate Crawford, 'The Trouble with Bias', NIPS Conference 2017. (Video)

Miles Brundage et al., (2018), Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation. (Executive Summary)

High-Level Expert Group on AI with the European Commission, 2019, Ethics guidelines for trustworthy AI: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

International Committee of the Red Cross (ICRC), 2018, Ethics and autonomous weapon systems: An ethical basis for human control? https://www.icrc.org/en/document/ethics-and-autonomous-weapon-systems-ethical-basis-human-control

Recommended readings:

Lannquist, Y., Loke, J., Miailhe, N., Hodes, C. and R. Yampolskiy, 2020, *The Intersection and Governance of Artificial Intelligence and Cybersecurity*. (PDF in class folder)

Exponential View: #3 AI, Warfare & Global Security: Mariarosaria Taddeo & Azeem Azhar in Conversation https://www.youtube.com/watch?v=v6kat-LjkFw (Discusses risks of AI & Cybersecurity)

Kristian Lum, 'Predictive Policing,' Data & Society Research Institute, 2016. (video)

'Interpretability is Necessary in Machine Learning?' Debate at NIPS Conference 2017 among Yann Lecun, Rich Caruana, Patrice Simard, Killian Weinberger.

Session 8: **The Case for 21st Century Education (31/03) Guest Lecture: Michaela Horvathova (International Education Policy Expert)**

Required readings:

'Managing the Economic and Social Impacts of the AI Revolution - Part III' (PDF in class folder)

Nicolas Miailhe & Cyrus Hodes, Making the AI revolution work for everyone, The Future Society, AI Initiative, Report to OECD, February 2017 – PART 3, CHAPTER D

World Economic Forum, (2016), *The Future of Jobs: Employment, skills and workforce strategy for the fourth Industrial Revolution.* (Skim)

Alex Gray, (2016), "The 10 skills you need to thrive in the 4th Industrial Revolution," World Economic Forum.

Artificial intelligence, automation, and the economy, Executive Office of the President of the United States, December 2016. (Pages 30-34)

Niki Iliadis, (2018),'learning to learn, the future-proof skill", Big Innovation Centre, the Secretariat for the All-Party, Parliamentary Group on AI. http://www.appg-ai.org/wp-content/uploads/2018/10/learning-to-learn_final_report_bic_kpmg.pdf Document also in the class Google drive folder. (Skim)

Maya Bialik & Charles Fadel (2018), Knowledge for the Age of Artificial Intelligence: What Should Students Learn?, Center for Curriculum Redesign https://curriculumredesign.org/wp-content/uploads/CCR_Knowledge_FINAL_January_2018.pdf Document also in the class Google drive folder. (Executive Summary).

Recommended readings:

Elliott, S. (2017), *Computers and the Future of Skill Demand*, Educational Research and Innovation, OECD Publishing, Paris.

Fadel, Bialik and Trilling, (2015), *Four-dimensional education: The competencies learners need to succeed.* Center for Curriculum Redesign.

Wayne Holmes, Maya Bialik, Charles Fadel (2019), Artificial Intelligence in Education, Center for Curriculum Redesign.

Session 9: Pathways to AI Governance and Policy (07/04)

Required readings:

Nicolas Miailhe & Cyrus Hodes, Making the AI revolution work for everyone, The Future Society, AI Initiative, Report to OECD, February 2017 – PART 3, CHAPTER A

Tim Dutton et. al, 2018, *Building an AI World: Report on National and Regional AI Strategies*. CIFAR.

Future of Life Institute, "Global Governance, Race Conditions, and International Cooperation," ('Global Governance, Race Conditions, and International Cooperation' section).

Paul De Hert, Vagelis Papakonstantinou, Gianclaudio Malgieria, Laurent Beslayc, Ignacio Sanchez, (2017), "The right to data portability in the GDPR: Towards user-centric interoperability of digital services", Computer Law & Security Review. (Introduction).

Stix, Charlotte, 2019, A Survey of the European Union's Artificial Intelligence Ecosystem. https://www.charlottestix.com/european-union-ai-ecosystem (Executive Summary)

Ding, J. (2018). *Deciphering China's AI Dream: the context, components, capabilities, and consequences of China's strategy to lead the world in AI.* Future of Humanity Institute, Oxford University. (Executive Summary)

EU Commission, (2019), White Paper On Artificial Intelligence: A European Approach to Excellence and Trust. (Skim)

The Economist, (2020), The EU wants to set the rules for the world of technology

Recommended readings:

Exponential View, Diplomacy in the Age of GAFA: Casper Klynge and Azeem Azhar in Conversation.

Furman, J., 2019, Unlocking digital competition: Report of the digital competition expert panel - Summary p. 8.

"Big Data: Bringing Competition Policy to the Digital Era - Executive Summary," OECD, 2017.

Article 29 Working Party, Guidelines on the right to data portability, December 2016,

Peter Galdis, A summary of the EU General Data Protection Regulation, DataIQ, October 2017.

Michaela Ross, (2017), "Artificial Intelligence Pushes the Anti-Trust Envelope," Bloomberg Law.

Session 10: **AI use cases simulation (The Ethics of Facial Recognition) (14/04)**

In-class workshop moderated by Sacha Alanoca and Buse Çetin, The Future Society

Required readings:

European Union, (2019), Ethics guidelines for trustworthy AI, https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (skim)

Harvard Law School, Cyber Law Clinic, Berkman Klein Centre for Internet and Society, (2019), Introducing the Principled Artificial Intelligence Project

Jobin, A., Ienca, M., & Vayena, E. (2019). Artificial Intelligence: the global landscape of ethics guidelines (skim)

Nicolas Miailhe, Niki Iliadis and members of the World Leadership Alliance, Club de Madrid, Policy Dialogue, October, (2019), "Fundamental Rights for the Digital Age". http://www.clubmadrid.org/wp-content/uploads/2019/06/Booklet-Policy-Dialogue-2019.pdf (Skim page 49-54). Document also in the class Google drive folder.

Kate Crawford, (2019), Halt the use of facial-recognition technology until it is regulated**,** World View.

Recommended readings:

CNIL, (2019), Reconnaissance Facial pour un debat a la hauteur des enjeux.

BAAI, (2019), Beijing AI Principles

Mittelstadt, B. (2019). AI Ethics--Too Principled to Fail?

Session 11: Pathways of Governance: Global Governance of AI (21/04)
Tentative Case discussions: The GPAI, AI Commons, AI for SDGs

Required readings:
IEEE, 2019, Ethically Aligned Design, First Edition (register to download)

Nicolas Miailhe & Yolanda Lannquist, (2018), A Challenge to Global Governance, IDB INTAL, Planet Algorithm: Artificial Intelligence for a Predictive and Inclusive form of Integration in Latin America, Page 207-217

Nicolas Miailhe, (2018), "AI & Global Governance: Why we need an intergovernmental panel for artificial intelligence." UN University Center for Policy Research.

Nicolas Miailhe, Yolanda Lannquist, R. Buse Çetin and Nicolas Moës at The Future Society, (2019), "Governing AI Adoption in Developing Countries" Report presented at the Global Governance of AI Roundtable (GGAR) at the World Government Summit in Dubai.

Vinuesa, R. et al, 2020, "The role of artificial intelligence in achieving the Sustainable Development Goals," https://arxiv.org/pdf/1905.00501.pdf.

Recommended readings:
Interview with Urs Gasser, (2018), "GSR-18: Dr URS GASSER, Professor, Harvard; ED, Berkman Klein Center," ITU Global Symposium for Regulators, Geneva, Switzerland. (9-minute video).

Interview with Stuart Russell (2018), "AI FOR GOOD 2018 INTERVIEWS: STUART RUSSELL, Professor of Computer Science, UC-Berkeley," AI for Good Summit 2018, Geneva, Switzerland. (9-min video)

McKinsey Global Institute (2018), *Applying Artificial Intelligence for Social Good.* Discussion Paper.

Session 12: Conclusions & wrap-up: Governance of AI & Long Term Existential Risks & Laws (29/04)

Required readings
Stuart Russell, (2017), "3 Principles for Creating Safer AI" (TED Talk)

Nick Bostrom, (2015), "What happens when our computers get smarter than we are?" TED Talk.

Stuart Russell, 'Take a stand on AI weapons,' Nature, May 2015. (First article by Stuart Russell)

Paul Scharre, Why you shouldn't fear 'Slaughterbots', IEEE Spectrum, December 2017

Stuart Russell, Anthony Aguirre, Ariel Conn and Max Tegmark, "Why You Should Fear 'Slaughterbots'—A Response," IEEE Spectrum, January 2018

World Government Summit, (2018), Summary Report: Global Governance of AI Roundtable Report. (Executive Summary)

Allan Dafoe, (2018), AI Governance: A Research Agenda. Future of Humanity Institute, University of Oxford. http://www.fhi.ox.ac.uk/govaiagenda (Read intro pages 6-13; skim other sections)

Recommended readings:
The Future Society, (2018), *A Global Civic Debate on Governing the Rise of AI,* (Executive Summary)

Floridi, L. et al. 2018. *Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*. AI4People. (Skim)

Russell, Stuart. 2019. Human Compatible: Artificial Intelligence and the Problem of Control. Penguin Publishing Group.

Urs Gasser, (2018), "AI and the Law: Setting the Stage," Berkman Klein Center of Internet and Society at Harvard University, Medium.

Additional Recommended Readings

Tom Upchurch, (2018), "How China Could Beat the West in the Deadly Race for AI Weapons," Wired.

Sotala, Kaj, and Roman V. Yampolskiy. (2013) "Responses to Catastrophic AGI Risk: A Survey." Technical Report. Berkeley, CA: Machine Intelligence Research Institute. http://intelligence.org/files/ResponsesAGIRisk.pdf (Skim Section 4. External AGI Constraints & Section 5. Internal Constraints)

Robert Wiblin, (2017), Positively Shaping the Development of Artificial Intelligence, 80,000 Hours

Bostrom, Nick (2014). "Superintelligence: Paths, Dangers, Strategies". Oxford University Press.

Future of Life Institute Podcast, (2020), On Lethal Autonomous Weapons with Paul Scharre.

Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Cristiano, John Schulman, Dan Mané (2016) "Concrete Problems in AI Safety," *Google Brain, OpenAI, UC Berkeley & Stanford University.*

Video from the campaign to ban Lethal Autonomous Weapons.