Beneficial AI 2017

Participants & Attendees

future of life INSTITUTE

Anthony Aguirre is a Professor of Physics at the University of California, Santa Cruz. He has worked on a wide variety of topics in theoretical cosmology and fundamental physics, including inflation, black holes, quantum theory, and information theory. He also has strong interest in science outreach, and has appeared in numerous science documentaries. He is a co-founder of the Future of Life Institute, the Foundational Questions Institute, and Metaculus (http://www.metaculus.com/).

Sam Altman is president of Y Combinator and was the cofounder of Loopt, a location-based social networking app. He also co-founded OpenAI with Elon Musk. Sam has invested in over 1,000 companies.

Dario Amodei is the co-author of the recent paper Concrete Problems in AI Safety, which outlines a pragmatic and empirical approach to making AI systems safe. Dario is currently a research scientist at OpenAI, and prior to that worked at Google and Baidu. Dario also helped to lead the project that developed Deep Speech 2, which was named one of 10 "Breakthrough Technologies of 2016" by MIT Technology Review. Dario holds a PhD in physics from Princeton University, where he was awarded the Hertz Foundation doctoral thesis prize.

Amara Angelica is Research Director for Ray Kurzweil, responsible for books, charts, and special projects. Amara's background is in aerospace engineering, in electronic warfare, electronic intelligence, human factors, and computer systems analysis areas. A co-founder and initial Academic Model/Curriculum Lead for Singularity University, she was formerly on the board of directors of the National Space Society, is a member of the Space Development Steering Committee, and is a professional member of the Institute of Electrical and Electronics Engineers (IEEE).

Stuart Armstrong is Alexander Tamas Fellow in Artificial Intelligence and Machine Learning at the Future of Humanity Institute (FHI), Oxford. His research at FHI centers on formal decision theory, general existential risk, the risks and possibilities of artificial intelligence, assessing expertise and predictions, and anthropic (self-locating) probability. He has been analysing and improving artificial agents' ability to learn human values, presenting these results at the Conference on Neural Information Processing Systems (NIPS) 2016. His collaboration with DeepMind on Interruptibility has been mentioned in over 100 media articles.

Peter Asaro is an Assistant Professor in the School of Media Studies at The New School in New York City. He is also the co-founder and vice-chair of the International Committee for Robot Arms Control, and has written on lethal robotics from the perspective of just war theory and human rights. Peter's research examines agency and autonomy, liability and punishment, and privacy and surveillance as it applies to consumer robots, industrial automation, smart buildings, aerial drones and autonomous vehicles. His current project is "Regulating Autonomous Artificial Agents: A Systematic Approach to Developing AI & Robot Policy."

Kareem Ayoub is the assistant to Mustafa Suleyman, who is co-founder and Head of Applied AI at DeepMind Technologies. Previously, Kareem co-founded World STEM Works, a non-profit organization dedicated to closing the gap between science and the public. In 2012, he worked collaboratively on his DPhil with the National Institutes of Health (NIH) as a Marshall Scholar at the University of Oxford, and the Medical Scientist Training Program at Washington University in St. Louis.

Guru Banavar is vice president and chief science officer for cognitive computing at IBM. He is responsible for advancing the next generation of cognitive technologies and solutions with IBM's global scientific ecosystem, including academia, government agencies and other partners. Guru leads the Cognitive Horizons Network, a set of research collaborations with leading institutions. Most recently, Guru led the team responsible for creating new AI technologies and systems in the family of IBM Watson, designed to augment human expertise in all industries, from healthcare to financial services to education.

Yoshua Bengio is a professor in the Department of Computer Science and Operations Research at the University of Montréal, and a pioneer of deep learning. He is also head of the Montreal Institute for Learning Algorithms (MILA), program co-director of the Canadian Institute for Advanced Research (CIFAR) Neural Computation and Adaptive Perception program, and Canada Research Chair in Statistical Learning Algorithms. Bengio is currently the action editor for the Journal of Machine Learning Research and an associate editor for Neural Computation.

Nicolas Berggruen is the Chairman of the Berggruen Institute which addresses fundamental political and cultural questions in our rapidly changing humanity with its Great Transformations work. Committed to the arts, he sits on the boards of the Museum Berggruen, Berlin, and the Los Angeles County Museum of Art and has built with leading architects.

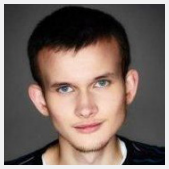
Irakli Beridze is the Senior Strategy and Policy Advisor for the United Nations – Interregional Crime and Justice Research Institute (UNICRI). Prior to joining UNICRI he served as a special projects officer at the Organisation for the Prohibition of Chemical Weapons (OPCW) undertaking extensive missions in politically sensitive areas around the globe. He is a recipient of recognition on the awarding of the Nobel Peace Prize to the OPCW in 2013. Since 2015, he initiated and headed the UN Centre on AI and Robotics with the objective to enhance understanding of the risk-benefit duality of AI through improved coordination, knowledge collection and dissemination, awareness-raising and global outreach activities. He is a member of various international task forces and working groups advising governments and international organisations on numerous issues related to international security, emerging technologies and global political trends.

Margaret Boden is a Research Professor of Cognitive Science at the University of Sussex, where she helped develop the world's first academic programme in cognitive science. She holds degrees in medical sciences, philosophy, and psychology, and integrates these disciplines with AI in her research. She is a Fellow of the British Academy, and of the Association for the Advancement of Artificial Intelligence. Her books include The Creative Mind: Myths and Mechanisms (1990/2004) and Mind as Machine: A History of Cognitive Science (2006). Her latest book, AI, Its Nature and Future, was published last June.

Grégory Bonnet is an Assistant Professor at the University of Caen Normandy, where he works in the GREYC Laboratory on the Models, Agents and Decisions (MAD) team. He received a Ph.D degree in Artificial Intelligence and Embedded Systems at ONERA, Toulouse, France in 2008, and a Master's Degree in Artificial Intelligence at the Toulouse Institute of Computer Science in 2005. His research topics are multi-agent systems, reputation systems, coalition formation, moral reasoning and computational ethics.

Nick Bostrom is Professor at Oxford University, where he is the founding Director of the Future of Humanity Institute. He also directs the Strategic Artificial Intelligence Research Center. He is the author of some 200 publications, including Anthropic Bias (Routledge, 2002), Global Catastrophic Risks (ed., OUP, 2008), Human Enhancement (ed., OUP, 2009), and Superintelligence: Paths, Dangers, Strategies (OUP, 2014), a New York Times bestseller. He is best known for his pioneering work in five areas: (i) existential risk; (ii) the simulation argument; (iii) anthropics; (iv) impact of future technology, especially AI; and (v) macrostrategy (links between long-term outcomes and present actions).

Erik Brynjolfsson is Director of the MIT Initiative on the Digital Economy, Schussel Family Professor at the MIT Sloan School, and Research Associate at The National Bureau of Economic Research (NBER). His research examines the effects of information technologies on business strategy, productivity and performance, digital commerce, and intangible assets. At MIT, his courses include the Economics of Information, and the Analytics Lab. He is one of the most cited scholars in information systems and economics, and co-author of the New York Times bestseller The Second Machine Age: Work, Progress and Prosperity in a Time of Brilliant Technologies.



Vitalik Buterin is a programmer & writer. He is primarily known as a co-founder of Ethereum, a decentralized mining network and software development platform that facilitates the creation of new cryptocurrencies and programs that share a single block chain. In 2014, Buterin won the World Technology Award for the co-creation and invention of Ethereum. Additionally, he is a co-founder of Bitcoin Magazine and has contributed as a developer to various open source bitcoin-related projects. He also contributed to DarkWallet, and the cryptocurrency marketplace site Egora.

Craig Calhoun is president of the Berggruen Institute. From 2012-2016, he was director and president of the London School of Economics and Political Science, where he remains as Centennial Professor. His research has ranged broadly through social science addressing culture, social movements, education, religion, nationalism, the impact of technology, capitalism and globalization, and combining critical theory with both contemporary and historical empirical research. Calhoun is the author of several books including The Roots of Radicalism (2012), Neither Gods nor Emperors (1994), and Nations Matter (2007).

Ryan Calo is the Lane Powell and D. Wayne Gittinger Assistant Professor at the University of Washington School of Law and an Assistant Professor (by courtesy) in the Information School. He is a faculty co-director of the University of Washington Tech Policy Lab, an interdisciplinary research unit that spans the School of Law, Information School, and Department of Computer Science and Engineering. Calo researches the intersection of law and emerging technology, with an emphasis on robotics and the Internet. His work on drones, driverless cars, privacy, and other topics has appeared in law reviews, technical publications, and major news outlets, including Science, Nature, the New York Times, the Wall Street Journal, and NPR.

Stephen Cave is Executive Director of the Leverhulme Centre for the Future of Intelligence and Senior Research Fellow at the University of Cambridge. Previously, he worked for the British Foreign Office as a policy advisor and diplomat. He has written on a wide range of philosophical and scientific subjects, including for the New York Times, The Atlantic and many others. His book 'Immortality' was a New Scientist book of the year. He has a PhD in philosophy from the University of Cambridge.

David Chalmers is University Professor of Philosophy and co-director of the Center for Mind, Brain, and Consciousness at NYU, and also Professor of Philosophy at the Australian National University. He has a Ph.D. in philosophy and cognitive science from Indiana University, where he worked with neural networks and genetic algorithms in Douglas Hofstadter's AI research group. He is the author of The Conscious Mind (1996) and Constructing the World (2012). His ideas include the "hard problem" of consciousness and the "extended mind", whereby technology extends the mind into the world. He co-founded the Association for the Scientific Study of Consciousness and is co-founder and co-director of the PhilPapers Foundation.

Nancy Chang is a research scientist at Google building more meaningful models of language structure, use and learning. For her doctoral research at UC Berkeley and post-doctoral research at Sony Computer Science Laboratory and the Université Sorbonne Nouvelle, she developed a computational framework motivated by embodied theories of meaning, construction-based approaches to grammar and usage-based accounts of language acquisition. At Google, she focuses on leveraging such cognitively inspired approaches to address the fundamental challenges involved in building intelligent assistants.

Meia Chita-Tegmark is a Ph.D. candidate in Developmental Sciences at Boston University and an alumna of the Harvard Graduate School of Education. She conducts research in the Social Development and Learning Lab at Boston University, and she is interested in a variety of topics in developmental psychology, such as atypical social development, attention mechanisms and learning strategies. Meia has strong interests in the future of humanity and big picture questions, and she is a co-founder of the Future of Life Institute.

Paul Christiano is a PhD student in theoretical computer science at UC Berkeley and a research associate at the Future of Humanity Institute. His work in algorithms and learning theory has won best paper and best student paper awards at the Symposium on Theory of Computing. He recently coauthored a paper on the practical problems in AI safety, titled Concrete Problems in AI Safety, and he blogs about AI and philanthropy at medium.com/ai-control. Paul will be a researcher at OpenAI starting in January 2017.

Jack Clark is the Strategy and Communications Director at OpenAI, where he helps with community outreach, policy, communications, and strategy. Jack has spent the past few years writing about artificial intelligence and distributed systems, most recently at Bloomberg and BusinessWeek. His articles have covered technologies like memory networks, image generation, and reinforcement learning for robots, and issues like diversity within AI.

Vincent Conitzer is the Kimberly J. Jenkins University Professor of New Technologies and Professor of Computer Science, Professor of Economics, and Professor of Philosophy at Duke University. Most of his research is on artificial intelligence (especially multiagent systems) and economic theory (especially game theory, social choice, and mechanism design). Conitzer has received several awards for papers and service at the Association for the Advancement of Artificial Intelligence (AAAI) and Autonomous Agents and Multiagent Systems (AAMAS) conferences. Conitzer and Preston McAfee are the founding Editors-in-Chief of the Association for Computing Machinery (ACM) Transactions on Economics and Computation (TEAC).

Ariel Conn oversees digital media and communications for the Future of Life Institute (FLI) at futureoflife.org. She specializes in all forms of online science communication, including writing, social media and web design. She has bachelors degrees in English and physics and a masters in geophysics. She created a got milk? commercial, interned with NASA, researched induced seismology at both Virginia Tech and the National Energy Technology Laboratory, and worked as a science writer for the Idaho National Laboratory.

Owen Cotton-Barratt is a Research Fellow at the Future of Humanity Institute at the University of Oxford, and a Research Advisor at the Centre for Effective Altruism. He has a PhD in pure mathematics. His research interests are centred on how to prioritise actions in situations of great uncertainty, with a focus on the value of research and understanding the long-term effects of actions today. Owen is the Principal Investigator on a grant from the Future of Life Institute into identifying decision-relevant uncertainty for AI safety.

Kate Crawford is a Principal Researcher at Microsoft Research in New York City, a Visiting Professor at MIT's Center for Civic Media, and a Senior Fellow at NYU's Information Law Institute. Her research addresses the social implications of large scale data, machine learning, data discrimination, social impacts of AI, predictive analytics and due process, ethical review for data science, and algorithmic accountability. She is a member of the World Economic Forum's Global Agenda Council on AI and Robotics, as well as the UN Thematic Network on Data for Development. In July 2016, she was the co-chair of the White House symposium AI Now: The Social and Economic Implication of AI in the Near-Term.

Andrew Critch is a Research Fellow at the Machine Intelligence Research Institute (MIRI), and a visiting postdoc at the UC Berkeley Center for Human Compatible AI. He earned his PhD in mathematics at UC Berkeley, and during that time, he cofounded the Center for Applied Rationality and the Summer Program on Applied Rationality and Cognition (SPARC). Andrew has been with MIRI since 2015, and his current research interests include logical uncertainty, open source game theory, and avoiding arms race dynamics between nations and companies in AI development.

Allan Dafoe is an Assistant Professor of Political Science at Yale University and a Research Associate at the Future of Humanity Institute, Oxford University. His research broadly examines the causes of great power war, and statistical methods for causal inference and scientific transparency. Specifically, his current research focusses on the global politics of artificial intelligence, including the risks of militarization of AI, possibilities for global governance of AI, and international institutions promoting beneficial AI.

Tucker Davey is a writer for the Future of Life Institute, where he has recently been interviewing and writing about FLI's AI safety researchers. Tucker is interested in existential risks and effective altruism, and how we can best communicate these ideas to a general audience. After spending this past summer working at a children's home in Honduras, he has become a determined advocate for the EA movement and hopes to make a difference through his research and writing.

Abram Demski is a computer science Ph.D student at the University of Southern California (USC) and a research assistant at the USC Institute for Creative Technologies. Abram is interested in the relationship between artificial intelligence and the foundations of cognition, including decision theory, universal theories of intelligence, and the complex relationship between logic and probability. On the latter theme, he published Logical Prior Probability at the International Conference on Artificial General Intelligence (AGI 2012).

Daniel Dewey is program officer for the Open Philanthropy Project, where he leads the OpenPhil's work on supporting technical research to mitigate potential risks from advanced artificial intelligence. He was previously a research fellow at the Future of Humanity Institute and Oxford Martin School, a software engineer at Google Seattle, and a student researcher at Intel Labs Pittsburgh.

Thomas Dietterich is a professor and Director of Intelligent Systems in the School of Electrical Engineering and Computer Science at Oregon State University. He is the past President of the Association for the Advancement of Artificial Intelligence (AAAI). From 1992-1998 he held the position of Executive Editor of the journal Machine Learning. In 2000, he co-founded a free electronic journal: The Journal of Machine Learning Research, and he is currently a member of the Editorial Board. He was Technical Program Chair of the Neural Information Processing Systems (NIPS) conference in 2000 and General Chair in 2001. He is Past-President of the International Machine Learning Society (IMLS).

Anca Dragan is an Assistant Professor in the EECS Department at UC Berkeley. Her goal is to enable robots to work with, around, and in support of people. She runs the InterACT lab, which focuses on algorithms that move beyond the robot's function in isolation, and generate robot behavior that also accounts for interaction and coordination with end-users. She works across different applications, from assistive robots, to manufacturing, to autonomous cars, and draws from optimal control, planning, estimation, learning, and cognitive science. She serves on the steering committee for the Berkeley AI Research Lab and is a co-PI for the Center for Human-Compatible AI.

Eric Drexler is a pioneering nanotechnology researcher and author. His 1981 paper in the Proceedings of the National Academy of Sciences established fundamental principles of molecular engineering and identified development paths leading to advanced nanotechnologies. In his 1986 book, Engines of Creation, he introduced a broad audience to the promise of high-throughput atomically precise manufacturing. He has worked with the World Wildlife Fund to explore nanotechnology-based solutions to global problems such as energy and climate change. He is currently an Academic Visitor in residence at Oxford University, where he is a Senior Research Fellow, Alexander Tamas Initiative on Artificial Intelligence and Machine Learning at the Future of Humanity Institute.

Stefano Ermon is an Assistant Professor in the Department of Computer Science at Stanford University, where he is affiliated with the Artificial Intelligence Laboratory. His research interests include techniques for scalable and accurate inference in graphical models, statistical modeling of data, large-scale combinatorial optimization, and robust decision making under uncertainty. His research is motivated by a range of applications, in particular ones in the emerging field of computational sustainability.

Oren Etzioni is is Chief Executive Officer of the Allen Institute for Artificial Intelligence. He has been a Professor in the University of Washington Computer Science department since 1991, where he received several awards including GeekWire's Hire of the Year (2014). He was also the founder or co-founder of several companies including Farecast (sold to Microsoft in 2008) and Decide (sold to eBay in 2013), and the author of over 100 technical papers. The goal of Oren's research is to solve fundamental problems in AI, particularly the automatic learning of knowledge from text.

Owain Evans is a postdoctoral scientist at the University of Oxford, where he works with the Future of Humanity Institute as a Research Fellow in the Alexander Tamas Initiative on Artificial Intelligence and Machine Learning. He also leads a collaborative project on Inferring Human Preferences with Andreas Stuhlmueller, John Salvatier and Noah Goodman. He works on incorporating human preferences and constraints into reinforcement learning systems. Owain received his Ph.D at MIT, where he worked on MIT's Probabilistic Computing Project and on Bayesian cognitive science.

Sebastian Farquhar is the Director of the Global Priorities Project - an Oxford-based think tank developing policy in effective government decision-making, unprecedented technological risk, and global public health. Before joining the Global Priorities Project, Sebastian worked as a management consultant at McKinsey & Co. with experience in healthcare, public policy and innovation strategy. He was on the core team launching 80,000 Hours - a career impact social enterprise. He has a Masters in Physics and Philosophy from the University of Oxford.

Chelsea Finn is a graduate student at UC Berkeley, where she works on machine learning for robotic perception and control. She is interested in how learning algorithms can enable robots to autonomously acquire complex sensorimotor skills. She recently spent time at Google Brain, where she worked on scaling self-supervised robot learning with an array of 10 robots. Before joining Berkeley AI Research, she received a Bachelors at MIT, where she worked on computer vision for assistive technologies in the Computer Science and AI Lab (CSAIL).

Kay Firth-Butterfield is the Executive Director of AI-Austin.org, and an adjunct Professor of Law at the University of Texas at Austin. Kay has advanced degrees in Law and International Relations, and advises governments, think tanks and nonprofits about artificial intelligence, law and policy. She co-founded the Consortium for Law and Policy of Artificial Intelligence and Robotics at the University of Texas and teaches its first course: Artificial Intelligence and emerging technologies: Law and Policy. Kay is Vice Chair of an Institute of Electrical and Electronics Engineers (IEEE) Industry Connections Committee considering Artificial Intelligence and ethical design.

Dileep George is a Co-Founder at Vicarious Systems, Inc. Before cofounding Vicarious, Dileep was CTO of Numenta, an AI company he cofounded with Jeff Hawkins and Donna Dubinsky. Before Numenta, he was a Research Fellow at the Redwood Neuroscience Institute. Dileep has authored 22 patents and several influential papers on the mathematics of brain circuits. His research on hierarchical models of the brain earned him a PhD in Electrical Engineering from Stanford University. He earned his MS in EE from Stanford and his BS from IIT in Bombay.

Ian Goodfellow is a Research Scientist at OpenAI and has formerly been affiliated with Google Brain, Université de Montréal, and Stanford AI Lab. He is best known as the inventor of generative adversarial networks and as the lead author of the Deep Learning textbook. As a machine learning security researcher, Ian believes that studying security is key to guaranteeing safety, and that developing techniques to resist human abuse and misuse of AI systems is of paramount importance.

Stephen Goose is director of Human Rights Watch's Arms Division, and was instrumental in bringing about the 2008 convention banning cluster munitions, the 1997 treaty banning antipersonnel mines, the 1995 protocol banning blinding lasers, and the 2003 protocol requiring clean-up of explosive remnants of war. He and Human Rights Watch co-founded the International Campaign to Ban Landmines (ICBL), which received the 1997 Nobel Peace Prize. Goose created the ICBL's Landmine Monitor initiative, the first time that non-governmental organizations around the world have worked together in a sustained and coordinated way to monitor compliance with an international disarmament or humanitarian law treaty. In 2013, he and Human Rights Watch co-founded the Campaign to Stop Killer Robots.

Joseph Gordon-Levitt is is an American actor and filmmaker. He has starred in many films, including 500 Days of Summer, Inception, 50/50, The Dark Knight Rises, and Snowden (2016). He also founded the online production company hitRECord in 2004 and has hosted his own TV series, HitRecord on TV, since January 2014. As one of his recent projects with hitRECord, he is developing a short animated series called USAI, based on the question: Could a machine beat a human at the game of electoral politics?

Katja Grace is a researcher at the Machine Intelligence Research Institute (MIRI) in Berkeley. She contributes to AI Impacts, an independent research project focused on social and historical questions related to artificial intelligence outcomes. Her analyses include Algorithmic Progress in Six Domains (2013). She writes the blog Meteuphoric, and is a part-time PhD student in Logic, Computation, and Methodology at Carnegie Mellon University. Katja previously studied game theory—especially in signaling and anthropic reasoning.

Joshua D. Greene is Professor of Psychology, a member of the Center for Brain Science faculty, and the director of the Moral Cognition Lab at Harvard University. His research has focused on the psychology and neuroscience of moral judgment and decision-making. His broader interests cluster around the intersection of philosophy, psychology, and neuroscience. He is the author of Moral Tribes: Emotion, Reason, and the Gap Between Us and Them.

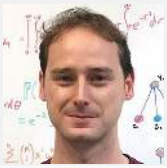
Tom Gruber is a product designer and entrepreneur who uses technology to augment human intelligence. He was cofounder, CTO and VP Design for Siri, and he leads Siri Advanced Development at Apple. His research at Stanford in artificial intelligence, particularly ontology engineering, helped lay the groundwork for semantic information sharing and the Semantic Web. He invented HyperMail, the open-source application that turns email conversations into collective memories on the Web. He has also created companies whose products foster the creation and sharing of collective knowledge: Intraspect for the virtual enterprise and RealTravel for the consumer Web. He is passionate about ocean conservation.

Marta Halina is a University Lecturer in the Philosophy of Cognitive Science at the University of Cambridge. She was a McDonnell Postdoctoral Fellow in the Philosophy-Neuroscience-Psychology Program at Washington University in St. Louis. Her current research focuses on issues related to nonhuman animal mindreading, ape gestural communication, and mechanistic explanation in biology. Marta is a fellow of Selwyn College where she directs studies in History and Philosophy of Science and the Psychological and Behavioural Sciences. She also coordinates the subproject Kinds of Intelligence at the Leverhulme Centre for the Future of Intelligence.

Verity Harding is the Public Policy Manager for DeepMind. She joined DeepMind from Google, where she was Head of Security Policy in Europe, having joined the company in 2013. Prior to this she was Special Adviser to the Deputy Prime Minister of the UK, the Rt Hon Nick Clegg MP, with responsibility for the Home and Justice Departments. She is a graduate of Pembroke College, Oxford University, and was a Michael Von Clemm Fellow at Harvard University Graduate School of Arts and Sciences. In her spare time, Verity sits on the Advisory Board of the social enterprise Women On Boards UK.

Sam Harris is a neuroscientist and the author of five New York Times best sellers. His books include The End of Faith, Letter to a Christian Nation, The Moral Landscape, Free Will, Lying, Waking Up, and Islam and the Future of Tolerance (with Maajid Nawaz). The End of Faith won the 2005 PEN Award for Nonfiction. His writing and public lectures cover a wide range of topics—neuroscience, moral philosophy, religion, meditation practice, human violence, rationality—but generally focus on how a growing understanding of ourselves and the world is changing our sense of how we should live. He also regularly hosts a popular podcast, and in September 2016 he released a TED talk titled: Can we build AI without losing control over it?

Demis Hassabis is the Co-Founder and CEO of DeepMind, which develops neuroscience-inspired general-purpose learning algorithms. Before founding DeepMind, Demis was a Research Fellow at Gatsby Computational Neuroscience Unit at UCL, and a visiting scientist at MIT and Harvard. Demis obtained his PhD in Cognitive Neuroscience from UCL where he investigated the neural mechanisms behind memory and imagination. In his early career he worked on several pioneering AI simulation games such as Syndicate, Theme Park, Black & White, and Republic, after obtaining a BA in Computer Science from Cambridge.



Nick Hay is a researcher at Vicarious, and earned his PhD at UC Berkeley under Stuart Russell. His research applied reinforcement learning and Bayesian analysis to the metalevel control problem: how an agent can learn to control its own computations. More broadly, he is interested in how AI systems can be safely developed for the benefit of humanity, and how this pursuit might be informed by the cognitive sciences.

John Hering is Executive Chairman and Founder of Lookout. He has grown Lookout's footprint to tens of millions of users globally across consumer, enterprise, and government sectors. John is a frequent presenter at mobile and technology industry events including: RSA, Mobile World Congress, Black Hat Technical Security Conference, DEFCON, and Fortune Brainstorm. Additionally, John is an investor in dozens of technology startups focused on the areas of cybersecurity, artificial intelligence, enterprise software, and transportation technologies.

José Hernández-Orallo is Professor of Information Systems and Computation at the Technical University of Valencia, Spain. His academic and research activities have spanned several areas of artificial intelligence, machine learning, data science, cognitive science and information systems. He has published five books and more than a hundred journal articles and conference papers on these topics. His most recent book addresses an integrated view of the evaluation of natural and artificial intelligence (Cambridge University Press, 2017).

Reid Hoffman is the co-founder and executive chairman of LinkedIn and a partner at venture capital firm Greylock Partners. Prior to LinkedIn, Reid served as executive vice president at PayPal, where he was also a founding board member. He currently serves on the boards of Airbnb, Convoy, Edmodo, Xapo, Linkedin, and a number of not-for-profit boards, including Kiva, Mozilla Corporation, Endeavor, and Do Something. He is also the co-author of two New York Times best-selling books: The Start-up of You and The Alliance. Reid earned a master's degree in philosophy from Oxford University, where he was a Marshall Scholar, and a bachelor's degree with distinction in symbolic systems from Stanford University.

ShaoLan Hsueh is an entrepreneur, geek, writer, board game designer, traveler and dreamer. She is the founder and creator of award-winning Chineasy, which is one of the most popular methods of learning Chinese. This project is the culmination of her life's journey through the East and West. Her aim is to help people to understand China, Chinese culture, its language and to bridge the gap between East and West, using technologies and design.Her podcast Talk Chineasy, is to be launched on 1st January 2017 on Amazon Echo and subsequently iTunes and GooglePlay.

Tim Hwang is at Google, where he coordinates the company's public policy portfolio on issues surrounding artificial intelligence and machine learning. Previously, he was the principal investigator leading Intelligence & Autonomy, a two-year interdisciplinary research project supported by the MacArthur Foundation exploring cross-sector policy challenges raised by machine intelligence. In the past, he has served in research roles at the Berkman-Klein Center at Harvard, Oxford Internet Institute, the Stanford Center for Legal Informatics, and the Electronic Frontier Foundation. Dubbed "The Busiest Man on the Internet" by Forbes Magazine, he was named as one of their "30 Under 30" for Law & Policy in 2014.

Daniel Kahneman is a Senior Scholar at the Woodrow Wilson School of Public and International Affairs. He is also Professor of Psychology and Public Affairs Emeritus at the Woodrow Wilson School, the Eugene Higgins Professor of Psychology Emeritus at Princeton University, and a fellow of the Center for Rationality at the Hebrew University in Jerusalem. He was awarded the Nobel Prize in Economic Sciences in 2002 for his pioneering work integrating insights from psychological research into economic science, especially concerning human judgment and decision-making under uncertainty. During the past several years, the primary focus of Professor Kahneman's research has been the study of various aspects of experienced utility (that is, the utility of outcomes as people actually live them).

Rao Kambhampati is Professor of Computer Science & Engineering at Arizona State University (ASU). Rao's research interests are in artificial intelligence with particular emphasis on planning, machine learning, analogical and case-based reasoning, and their applications to automated manufacturing. Rao is the recipient of a three-year National Science Foundation (NSF) research initiation award in 1992 and a five-year NSF Young Investigator award in 1994. Rao is currently serving as the president of the Association for the Advancement of Artificial Intelligence (AAAI).

Angela Kane is a senior fellow at the Vienna Centre for Disarmament and Non-Proliferation. She also teaches at Sciences Po in Paris. She has had a long and distinguished career at the United Nations; her functions included High Representative for Disarmament Affairs (2012–2015), Under-Secretary-General for Management (2008–2012), Assistant Secretary-General for Political Affairs (2005–2008) and Assistant Secretary-General for General Assembly and Conference Management. She served as Deputy Special Representative of the Secretary-General for the United Nations Mission in Ethiopia and Eritrea (UNMEE), and had postings in the Democratic Republic in the Congo, Indonesia, and Thailand.

Holden Karnofsky is a Co-Founder and Co-Executive Director of charity evaluator GiveWell, which he co-founded in mid-2007 after spending several years in the hedge fund industry. He is also the Executive Director of the Open Philanthropy Project, which is a collaboration between GiveWell and Good Ventures.

Viktoriya Krakovna is a research scientist in AI safety at DeepMind and a co-founder of the Future of Life Institute. Her PhD thesis in statistics and machine learning at Harvard University focused on building interpretable models. Viktoriya gained numerous distinctions for her accomplishments in math competitions, including a silver medal at the International Mathematical Olympiad and the Elizabeth Lowell Putnam prize.

János Kramár is a deep learning and AI safety researcher. In 2016 he interned at the Montreal Institute for Learning Algorithms, co-developing zoneout, a state-of-the-art regularization method for recurrent neural nets. He also coauthored a paper surveying and calling for prospective research on secure environments for testing advanced AI systems. He holds a Masters in Statistics from Harvard University, and has worked in algorithmic trading. He was a top competitor in math and programming competitions at the national level in Canada, earning a bronze medal at the International Math Olympiad.

Lawrence M. Krauss is a theoretical physicist with wide research interests, including the interface between elementary particle physics and cosmology, where his studies include the early universe, the nature of dark matter, general relativity and neutrino astrophysics. He is currently Foundation Professor in the School of Earth and Space Exploration and Physics Department at Arizona State University, and Inaugural Director of the Origins Project, a national center for research and outreach on origins issues, from the origins of the universe, to human origins, to the origins of consciousness and culture. Krauss has authored over 300 scientific publications, as well as numerous bestselling popular books. He serves as the chair of the Board of Sponsors of The Bulletin of the Atomic Scientists, and is on the Board of Directors of the Federation of American Scientists.

Ramana Kumar is a researcher in the Trustworthy Systems group of Data61 at The Commonwealth Scientific and Industrial Research Organisation (CSIRO), and conjoint lecturer at The University of New South Wales (UNSW). He earned a degree at The Australian National University in 2010 before obtaining an MPhil in Advanced Computer Science at the University of Cambridge in 2011. Kumar was awarded a Gates Scholarship and completed his Ph.D. at Cambridge in 2016 with the dissertation Self-compilation and Self-verification. Kumar is a lead developer of CakeML, a verified functional programming language. His research interests include scaling up formal methods, and ensuring that, when computer systems become substantially more powerful, their impact is beneficial.

Martina Kunz is a PhD candidate at the University of Cambridge doing research on the design features of international legal systems addressing environmental problems and aims to increase the efficiency of global governance through automated data collection, analytics, visualization and reporting applications. She also collaborates with researchers at the Centre for the Study of Existential Risk, the Centre for the Future of Intelligence and the Future of Humanity Institute on AI policy issues among others. Prior to her PhD at Cambridge, Martina studied and worked at the University of Geneva, Tsinghua University and the Graduate Institute of International and Development Studies.

Ray Kurzweil is an author, computer scientist, inventor and futurist. He was the principal inventor of the first CCD flat-bed scanner, the first omni-font optical character recognition program, and the first print-to-speech reading machine for the blind - just to name a few. He has written many books, including New York Times bestsellers The Singularity Is Near (2005) and How To Create A Mind (2012). He is Co-Founder and Chancellor of Singularity University and in 2012 he was appointed a Director of Engineering at Google, heading up a team developing machine intelligence and natural language understanding.

Neil Lawrence is a leader for Amazon Research Cambridge. He is on leave of absence from the University of Sheffield where he was a Professor in Computational Biology and Machine Learning jointly appointed across the Departments of Neuroscience and Computer Science. He moved to Sheffield in August 2010 from the School of Computer Science in Manchester, where he was a Senior Research Fellow. His research interests are in probabilistic models with applications in computational biology and personalized health. At Sheffield he worked with a team to develop the Open Data Science Initiative, an approach to data science designed to address societal needs.

David Leake is Executive Associate Dean and Professor of Computer Science in the School of Informatics and Computing at Indiana University. He is also an associate of the Indiana University Data to Insight Center. His research interests are in artificial intelligence and cognitive science, including case-based reasoning, intelligent information systems, intelligent user interfaces, knowledge management, knowledge modeling, multimodal reasoning, multistrategy learning, and introspective reasoning. He has authored/edited over 150 publications in these areas. He is Editor in Chief Emeritus of AI Magazine, the official magazine of the Association for the Advancement of Artificial Intelligence (AAAI), after 17 years as Editor in Chief.

Yann LeCun is is Director of AI Research at Facebook, and Silver Professor of Data Science, Computer Science, Neural Science, and Electrical Engineering at New York University (NYU). His current interests include AI, machine learning, computer perception, mobile robotics, and computational neuroscience, and he has published over 180 technical papers and book chapters on these topics. He has held research positions at AT&T, Bell Labs and NEC Labs, He was the founding director of the NYU Center for Data Science. He is the recipient of the 2014 Institute of Electrical and Electronics Engineers (IEEE) Neural Network Pioneer Award, the 2015 IEEE-PAMI Distinguished Researcher Award, and the 2016 Lifetime Achievement Award from the International Academy of Digital Arts and Sciences.

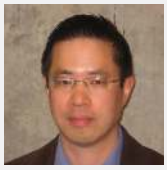
Sean Legassick is a policy advisor at DeepMind. He was a lead developer at Demon Internet and led the design, development and deployment of a multi-billion-pound web trading platform for Winterflood Securities. He is a recognised expert on web and network software development, having received accolades for his work with the Apache Foundation and as technical editor for The Definitive Guide to Django. Sean also created the core of the Chisimba web development framework.

Shane Legg is a machine learning researcher and founder of DeepMind. He is interested in measures of intelligence for machines, neural networks, artificial evolution, reinforcement learning and the theory of learning. He obtained his PhD from IDSIA in Switzerland, and his thesis proposed a formal definition of machine intelligence, for which he was awarded the $10,000 Canadian Singularity Institute research prize. He spent a postdoctoral year at the Swiss Finance Institute building models of human decision making, followed by two years at the Gatsby Computational Neuroscience Unit at UCL, where he is now an honorary fellow.

Jan Leike is a Research Scientist at DeepMind, and a Research Associate at the Future of Humanity Institute, University of Oxford. He is working on long-term technical problems of robust and beneficial artificial intelligence. Previously he was a PhD student with Marcus Hutter and wrote his dissertation on general reinforcement learning. He works on problems in reinforcement learning orthogonal to capability: How do we design or learn a good objective function? How can we design agents such that they are incentivized to act in our best interests? How can we avoid degenerate solutions to the objective function?

Sergey Levine is an Assistant Professor in the Department of Electrical Engineering and Computer Sciences at UC Berkeley. He received a Ph.D. in Computer Science from Stanford University in 2014. His work focuses on machine learning for decision making and control, with an emphasis on deep learning and reinforcement learning algorithms. Applications of his work include autonomous robots and vehicles, as well as computer vision and graphics. His research includes developing algorithms for end-to-end training of deep neural network policies that combine perception and control, scalable algorithms for inverse reinforcement learning, deep reinforcement learning algorithms, and more.

Fuxin Li is an assistant professor in the School of Electrical Engineering and Computer Science at Oregon State University. His research direction is machine learning and computer vision, with a major interest in using and designing new machine learning algorithms to attack the structural data in images and videos, especially big data originating from videos. He obtained a Ph.D. in 2009 from the Institute of Automation, Chinese Academy of Sciences and has since held postdoctoral appointments in the University of Bonn and Georgia Institute of Technology. Li has won the PASCAL VOC Segmentation challenge from 2009-2012, a Microsoft research award and is currently co-leading 2 National Science Foundation (NSF) projects.



Patrick Lin is the director of the Ethics + Emerging Sciences Group, based at California Polytechnic State University, San Luis Obispo, where he is an associate philosophy professor. Other current and past affiliations include: Stanford Engineering, Stanford Law, US Naval Academy, Dartmouth College, Notre Dame, World Economic Forum, and UNIDIR. He is well published in technology ethics, especially on robotics and AI—including the books Robot Ethics (MIT Press, 2012) and Robot Ethics 2.0 (Oxford University Press, forthcoming in 2017)—as well as cyberwar/security, nanotechnology, human enhancement, space exploration, and other areas. He regularly gives invited briefings to industry, media, and government; and he teaches courses in ethics, political philosophy, philosophy of technology, and philosophy of law.

Moshe Looks is Software Designer and Researcher at Google where he conducts research in program induction and artificial general intelligence. Moshe is a practitioner of artificial intelligence as a sister discipline of cognitive science, concerned with the mechanization of human thought. He has published over twenty peer-reviewed papers in venues such as IJCAI (International Joint Conference on AI) and ICCS (International Conference of the Cognitive Science), on topics such as probabilistic approaches to learning programs, AGI architecture, and the intersection of AI and cognitive science.

William MacAskill is the CEO and cofounder of the Centre for Effective Altruism and an Associate Professor in Philosophy at Oxford University. He helped to create the effective altruism movement: the use of evidence and reason to help others by as much as possible with our time and money. MacAskill also cofounded 80,000 Hours, a YC-backed non-profit that provides research and advice on how you can best make a difference through your career. His academic research is on the fundamentals of effective altruism, with a particular focus on how to act given moral uncertainty. He is also the author of Doing Good Better.

Richard Mallah is Director of AI Projects at the Future of Life Institute (FLI), where he works to support the robust, safe, beneficent development of advanced artificial intelligence via meta-research, analysis, research organization, and advocacy. Mallah serves on the Executive Committee of IEEE's (Institute of Electrical and Electronics Engineers) initiative on autonomous systems ethics, and chairs associated working groups. Richard serves as a senior advisor to both Cambridge Semantics, Inc., where he led creation of the industry's highest-rated enterprise text analytics system, and to The AI Initiative of The Future Society at the Harvard Kennedy School.

Jason Matheny became director of the Intelligence Advanced Research Projects Activity (IARPA) in 2015. Before IARPA, he worked at Oxford University, the World Bank, the Applied Physics Laboratory, the Center for Biosecurity and Princeton University, and is the co-founder of two biotechnology companies. Matheny holds a Ph.D. in applied economics from Johns Hopkins University, an M.P.H. from Johns Hopkins University, an M.B.A. from Duke University and a B.A. from the University of Chicago. He received the Intelligence Community's Award for Individual Achievement in Science and Technology.

Yutaka Matsuo is a project associate professor at the Institute of Engineering Innovation, the University of Tokyo. His current research interests are in web mining (especially on social network mining), text processing, and semantic web in the context of artificial intelligence research. Matsuo received the Japanese Society for Artificial Intelligence (JSAI) Best Paper Award in 2002, JSAI Anniversary Project Award in 2006, and Information Processing Society of Japan (IPSJ) Nagao Special Researcher Award in 2008. He served as editor-in-chief of JSAI from 2012-2014.

Andrew Maynard is a Professor in the School for the Future of Innovation in Society at Arizona State University, and Director of the Risk Innovation Lab. He was previously Chair of the Environmental Health Sciences department in the University of Michigan School of Public Health. His research and professional activities focus on risk innovation, and the responsible development and use of emerging technologies. He is a member of the World Economic Forum Global Future Council on Technology, Values and Policy, and writes a regular column for the journal Nature Nanotechnology, and the news website The Conversation.

Andrew McAfee is a principal research scientist at MIT, where he studies how digital technologies are changing business, the economy, and society. He is the coauthor, with Erik Brynjolfsson, of the 2014 bestseller The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies and of Machine, Platform, Crowd: Harnessing the Digital Revolution, which will be published in mid-2017. McAfee blogs for the Financial Times, and writes for publications including Foreign Affairs, Harvard Business Review, The Economist, The Wall Street Journal, and The New York Times. He was educated at Harvard and MIT, where he is the co-founder of the Institute's Initiative on the Digital Economy.

Tasha McCauley is a technology entrepreneur. Her current work with GeoSim Systems centers around a new technology that produces high-resolution, fully interactive virtual models of cities. Prior to her involvement with GeoSim, she co-founded Fellow Robots, a robotics company based at NASA Research Park in Silicon Valley. She was formerly on the faculty of Singularity University, where she taught students about robotics and was Director of the Autodesk Innovation Lab. She sits on the Board of Directors of the Ten to the Ninth Plus Foundation, an organization focused on empowering exponential technological change worldwide.

Tom Mitchell is the E. Fredkin University Professor at Carnegie Mellon University, where he founded the world's first Machine Learning Department. He has served on AI technical advisory committees for the U.S. Department of Justice, and the Department of Defense, has testified to the U.S. House Committee on Veteran's Affairs on possible roles for AI, and recently co-chaired a U.S. National Academies study on "Information Technology, Automation, and the U.S. Workforce." Mitchell is a member of the U.S. National Academy of Engineering, a member of the American Academy of Arts and Sciences, and Past President of the Association for the Advancement of Artificial Intelligence (AAAI). In his current research, Mitchell is developing a never-ending machine learning system which is learning to read the web (http://rtw.ml.cmu.edu), and is using brain imaging to study how the human brain comprehends language.

Elon Musk is the founder, CEO and CTO of SpaceX and co-founder and CEO of Tesla Motors. In recent years, Musk has focused on developing competitive renewable energy and technologies (Tesla, Solar City), and on taking steps towards making affordable space flight and colonization a future reality (SpaceX). He has spoken about the responsibility of technology leaders to solve global problems and tackle global risks, and has also highlighted the potential risks from advanced AI.

Andrew Ng is VP & Chief Scientist of Baidu, Co-Chairman and Co-Founder of Coursera, and an Adjunct Professor at Stanford University. Today, Coursera partners with some of the top universities in the world to offer high quality online courses, and is the largest MOOC (Massive Open Online Courses) platform in the world. Andrew also founded and led the "Google Brain" project which developed massive-scale deep learning algorithms. He is currently working on deep learning and its applications to computer vision and speech, including such applications as autonomous driving.

Peter Norvig is Director of Research at Google Inc. Previously he was head of Google's core search algorithms group, and of NASA Ames's Computational Sciences Division, making him NASA's senior computer scientist. His publications include the books Artificial Intelligence: A Modern Approach (with Stuart Russell), Paradigms of AI Programming: Case Studies in Common Lisp, Verbmobil: A Translation System for Face-to-Face Dialog, and Intelligent Help Systems for UNIX. He is a fellow of the Association for the Advancement of Artificial Intelligence (AAAI), the Association for Computing Machinery (ACM), and American Academy of Arts & Sciences.

Seán Ó hÉigeartaigh is the Executive Director of the Centre for the Study of Existential Risk (CSER) at the University of Cambridge. He is also a Senior Research Associate at the Leverhulme Centre for the Future of Intelligence, where he leads CFI's Policy and Responsible Innovation project, and is a co-investigator at the Strategic AI Research Centre. Seán's research spans technology policy and strategy, catastrophic risk, and horizon-scanning and foresight.

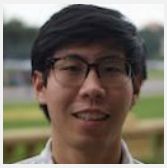
Catherine Olsson is a Software Engineer at OpenAI working on the Universe project. Catherine graduated in Computer Science and Brain & Cognitive Science from MIT, and completed a Master's degree in Neuroscience at NYU.

Steve Omohundro is president of both Possibility Research and Self-Aware Systems, a think tank working to ensure that intelligent technologies have a positive impact. He was a computer science professor at the University of Illinois at Champaign-Urbana and cofounded the Center for Complex Systems Research. He published the book Geometric Perturbation Theory in Physics, designed the programming languages StarLisp and Sather, wrote the 3D graphics system for Mathematica, and built systems which learn to read lips, control robots, and induce grammars.

Toby Ord is a moral philosopher at Oxford University. He created an international society called Giving What We Can, whose members have pledged over $600 million to the most effective charities helping to improve the world. He also co-founded the wider effective altruism movement, encouraging thousands of people to use reason and evidence to help others as much as possible. His current research is on avoiding the threat of human extinction and thus safeguarding a positive future for humanity, which he considers to be among the most pressing and neglected issues we face. He is a leading expert on the potential threats and opportunities posed by advanced artificial intelligence over the coming decades.

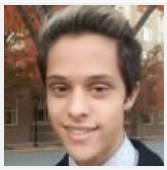
Laurent Orseau is a research scientist in AI safety at DeepMind, and previously an associate professor at AgroParisTech, Paris, France. In 2003, he graduated from a professional master in computer science at the National Institute of Applied Sciences in Rennes and from a research master in artificial intelligence at University of Rennes 1. His goal is to build a practical theory of artificial general intelligence. With his co-author Mark Ring, Orseau was awarded the Solomonoff AGI Theory Prize at AGI'2011 and the Kurzweil Award for Best Idea at AGI'2012.

Pedro Ortega is a Research Scientist in AI safety at DeepMind. His work includes the application of information-theoretic and statistical mechanical ideas to sequential decision-making, which has led to contributions in novel bounded rationality models and recasting adaptive control as a causal inference problem. He obtained his PhD in Engineering from the University of Cambridge, and he has been a postdoctoral fellow at the Department of Engineering in Cambridge, at the Max Planck Institute for Biological Cybernetics/Intelligent Systems, at the Hebrew University in Jerusalem, and at the University of Pennsylvania.



Long Ouyang is an independent research scientist. His research focuses on AI, and he is currently studying topics related to program synthesis, probabilistic programming, and concept learning. Previously, he studied cognitive psychology as a PhD student and then as a postdoc at Stanford.

Claudia Passos-Ferreira is a philosopher and a clinical psychologist. She studied psychology at Rio de Janeiro State University and obtained a Ph.D in Public Health in 2005. She has been a postdoctoral research fellow at the Social Medicine Institute (UERJ) where she taught bioethics and issues in neuroscience, psychopathology and psychoanalysis. She has also been a postdoctoral fellow in philosophy at Rio de Janeiro Federal University working on ethics and biotechnologies. She is currently visiting at Columbia University and New York University, working on infant consciousness, self-consciousness, and moral agency.



Lucas Perry is Project Coordinator for the Future of Life Institute (FLI). Lucas is passionate about the role that science and technology will play in the evolution of all sentient life. He has a background in philosophy, and has studied at a Buddhist monastery in Nepal and engaged in a range of meditative retreats and practices, which work to inform his study of transhumanism, effective altruism, and existential risks.

Tomaso Poggio is the Eugene McDermott Professor in the Dept. of Brain & Cognitive Sciences at MIT and the director of the NSF Center for Brains, Minds and Machines at MIT. He is a member of the Computer Science and Artificial Intelligence Laboratory and of the McGovern Brain Institute. He received the Laurea Honoris Causa from the University of Pavia for the Volta Bicentennial, the 2003 Gabor Award, the Okawa Prize 2009, the American Association for the Advancement of Science (AAAS) Fellowship, and the 2014 Swartz Prize for Theoretical and Computational Neuroscience. A former Corporate Fellow of Thinking Machines Corporation and a former director of PHZ Capital Partners, Inc., Tomaso is a director of Mobileye and was involved in starting, or investing in, several other high tech companies including Arris Pharmaceutical, nFX, Imagen, Digital Persona and DeepMind.

Gill Pratt is the Chief Executive Officer of Toyota Research Institute (TRI). Launched in 2016, TRI's mission is to enhance the safety of automobiles, with the ultimate goal of creating a car that is incapable of causing a crash. Pratt also serves as the Executive Technical Advisor to Toyota Motor Corporation. His primary interest is in the field of robotics and intelligent systems. Pratt holds a Doctor of Philosophy in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), where he later was an Associate Professor and Director of the Leg Lab. Subsequently, he became a Professor at Franklin W. Olin College, and before joining DARPA and then Toyota, was Associate Dean of Faculty Affairs and Research.

Huw Price is Bertrand Russell Professor of Philosophy and a Fellow of Trinity College at the University of Cambridge. He was previously Challis Professor of Philosophy at the University of Sydney, where he was Founding Director of the Centre for Time. In 2012 he co-founded the Centre for the Study of Existential Risk (CSER) in Cambridge with Martin Rees and Jaan Tallinn. He is presently Academic Director both of CSER and of the new Leverhulme Centre for the Future of Intelligence (CFI).

Martin Rees is based at Cambridge University and is the UK's 'Astronomer Royal'. He has written more than 500 research papers on the big bang, galaxy formation, black holes, cosmic explosions and the multiverse, as well as eight books, and hundreds of magazine and newspaper articles. He has received many awards (and academy memberships) for his research. He has been Master of Trinity College, Cambridge, President of the Royal Society (the UK's academy of sciences) and is a member of the UK's House of Lords. Ever since his book 'Our Final Century?' was published, he and been increasingly engaged with long-term environmental and technological threats, and is a co-founder of the Centre for Study of Existential Risks (CSER).

Heather Roff is a Senior Research Fellow at the University of Oxford, a Research Scientist in the Global Security Initiative at Arizona State University, a national Cybersecurity Fellow at the New America Foundation, and a Research Associate at the Eisenhower Center for Space and Defense Studies at the United States Air Force Academy. Her research interests pertain to issues of international security and global justice, principally in relation to emerging military technologies. She is the author of Global Justice, Kant and the Responsibility to Protect. Her new book project, Lethal Autonomous Weapons and the Future of War, examines the moral, legal and policy implications of autonomous systems.

Anthony Romero is the executive director of the American Civil Liberties Union. He has led the ACLU in its legal challenge to the patents held by a private company on the human genes associated with breast and ovarian cancer; in its landmark lawsuit challenging Arizona's anti-immigrant law that invites law enforcement to engage in racial profiling; in its high-profile litigation and lobbying efforts to win the freedom to marry for same-sex couples; and in its nationwide Campaign for Smart Justice, which aims to achieve a 50 percent reduction in the number of Americans behind bars. In 2007, Romero and co-author and NPR correspondent Dina Temple-Raston published In Defense of Our America: The Fight for Civil Liberties in the Age of Terror, a book that takes a critical look at civil liberties in this country at a time when constitutional freedoms are in peril.

Francesca Rossi is a research scientist at the IBM T.J. Watson Research Centre, and an professor of computer science at the University of Padova, Italy, currently on leave. Her research interests focus on artificial intelligence, specifically they include constraint reasoning, preferences, multi-agent systems, computational social choice, and collective decision making. She is also interested in ethical issues in the development and behaviour of AI systems. She has published over 170 scientific articles in journals and conference proceedings, and as book chapters. She is a AAAI and a EurAI fellow, and a Radcliffe fellow 2015. She has been president of IJCAI, an executive councillor of AAAI, and she is Editor in Chief of JAIR. She co-chairs the AAAI committee on AI and ethics and she is a member of the scientific advisory board of the Future of Life Institute. She is in the executive committee of the IEEE global initiative on ethical considerations on the development of autonomous and intelligent systems and she belongs to the World Economic Forum Global Council on AI and robotics.

Jonathan Rothberg is best known for inventing high-speed, "Next-Gen" DNA sequencing and was awarded the National Medal of Technology by president Obama for this innovation. Jonathan brought to market the first new method for sequencing genomes since Sanger and Gilbert won the Nobel Prize in 1980. He sequenced the first individual human genome (Watson Genome), initiated the Neanderthal Genome Project with Svante Paabo, and with the sequencing of Gordon Moore paved the way to the sub $1,000 genome. Under his leadership his team helped understand the mystery behind the disappearance of the honey bee, uncovered a new virus killing transplant patients, and elucidated the extent of human variation—work recognized by Science magazine as the breakthroughs of the year for 2006 & 2007. He currently runs 4catalyzer, a medical device incubator, and is an adjunct professor at Yale.

Daniela Rus is the Andrew (1956) and Erna Viterbi Professor of Electrical Engineering and Computer Science and Director of the Computer Science and Artificial Intelligence Laboratory (CSAIL) at MIT. She serves as the Director of the Toyota-CSAIL Joint Research Center and is a member of the science advisory board of the Toyota Research Institute. Rus' research interests are in robotics, mobile computing, and data science. Rus is a fellow of the Association for Computing Machinery (ACM), the Association for the Advancement of Artificial Intelligence (AAAI), and the Institute of Electrical and Electronics Engineers (IEEE). She earned her PhD in Computer Science from Cornell University. Prior to joining MIT, Rus was a professor in the Computer Science Department at Dartmouth College.

Stuart Russell is Professor of Computer Science at Berkeley and director of the Center for Human-Compatible Artificial Intelligence. His research covers many areas of artificial intelligence, with a particular focus on machine learning, probabilistic modeling and inference, theoretical foundations of rationality, and provably beneficial AI. He is a co-author (with Peter Norvig) of the standard textbook, Artificial Intelligence: a Modern Approach. He is a recipient of the Presidential Young Investigator Award of the National Science Foundation, the IJCAI Computers and Thought Award, and Outstanding Educator Awards from both ACM and AAAI. From 2012 to 2014 he held the Chaire Blaise Pascal in Paris. He is a Fellow of the Association for the Advancement of Artificial Intelligence, the Association for Computing Machinery, and the American Association for the Advancement of Science.

Jeffrey Sachs is University Professor at Columbia University, in the fields of economics, sustainable development, international affairs, and public health. He serves as Special Advisor to the UN Secretary General, a position he has held under Kofi Annan (2002-6), Ban Ki-moon (2007-2016), and Antonio Guterres (2017-). He is a best-selling author and a globally syndicated columnist. He served as the Director of the Earth Institute from 2002 to 2016. Sachs is Director of the Columbia University Center for Sustainable Development, and the UN Sustainable Development Solutions Network under the auspices of UN Secretary-General Antonio Guterres. He co-leads the Ethics in Action Program with the Chancellor of the Pontifical Academy of Sciences and the Secretary General of Religions for Peace. He is a Distinguished Fellow of the International Institute of Applied Systems Analysis in Laxenburg, Austria. He is co-founder and Chief Strategist of Millennium Promise Alliance, and was director of the Millennium Villages Project. Sachs is also one of the UN Secretary-General's SDG Advocates, and is a Commissioner of the ITU/UNESCO Broadband Commission for Development.

Anna Salamon is the co-founder and president of the Center for Applied Rationality (CFAR). She has previously done machine learning research for NASA and applied mathematics research on the statistics of phage metagenomics. She holds a degree in mathematics from UC Santa Barbara.

David Sanford is Chief of Staff, Office of Reid Hoffman at Linkedin Inc, and hee is also an Advisory Council Member at New America California. David received his BA in Entrepreneurial Management from Stanford University.

Matt Scherer is an attorney and legal scholar based in Portland, Oregon who writes and speaks on the intersection of law and artificial intelligence. He is the author of Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. When he's not writing about the law of robots and algorithms, Matt practices in the thoroughly human field of employment law at Buchanan Angeli Altschul & Sullivan LLP. He has a blog titled Law and AI, which is devoted to studying the emerging legal and policy issues surrounding artificial intelligence and autonomous machines.

Jürgen Schmidhuber is Scientific Director of the Swiss AI Lab IDSIA (USI & SUPSI), Professor of AI at USI, Switzerland, former head of the CogBotLab at TU Munich, and President of NNAISENSE, which aims at building the first practical general purpose AI. His team's Deep Learning Neural Networks have revolutionized machine learning and AI, and are now available to billions of users through Google, Apple, Microsoft, IBM, Baidu, and many other companies. His research group also established the field of mathematically rigorous universal AI and optimal universal problem solvers. His formal theory of creativity & curiosity & fun explains art, science, music, and humor. He is recipient of numerous awards including the 2016 Institute of Electrical and Electronics Engineers (IEEE) Neural Networks Pioneer Award.

Eric Schmidt received his PhD from UC Berkeley and BS from Princeton and is an American software engineer, businessperson, and the Executive Chairman of Alphabet, Inc. He joined Google's board of directors and served as a chairman in 2001, and then became the company's CEO that same year. Schmidt stepped down as CEO of Google in 2011, and since then he has served as the Executive Chairman of Google's board of directors and an advisor to Google's co-founders Larry Page and Sergey Brin. Additionally, he is a founding partner at venture capital firm Innovation Endeavors, which invested in tech startups including Uber and helicopter booking service Blade. Prior to his time at Google, Schmidt had stints as CEO of Novell and chief technology officer at Sun Microsystems, where he led the development of Java. In 2006, he was elected to the National Academy of Engineering, which recognized his work on "the development of strategies for the world's most successful Internet search engine company."

Bart Selman is a Professor of Computer Science at Cornell University. He previously was at AT&T Bell Laboratories. His research interests include computational sustainability, efficient reasoning procedures, planning, knowledge representation, and connections between computer science and statistical physics. He has (co-)authored over 100 publications, including six best paper awards. His papers have appeared in venues spanning Nature, Science, and a variety of conferences and journals in AI and Computer Science. He has received an NSF Career Award and an Alfred P. Sloan Research Fellowship. He is a Fellow of the American Association for Artificial Intelligence (AAAI) and a Fellow of the American Association for the Advancement of Science (AAAS).

Andrew Serazin is President of Templeton World Charity foundation, where he is building interdisciplinary teams of scientists, technologists, and humanities scholars to pursue big questions of human purpose, the natural world, and ultimate reality. As a malaria researcher at Oxford and Notre Dame, as well as an executive at the Bill & Melinda Gates foundation in Seattle, he worked to harness insights from unconventional thinkers to advance solutions for nutrition, maternal and child health, and infectious diseases. He also founded Matatu, a venture-backed biotechnology company, to demonstrate the commercial and societal value of the vast community of beneficial bacteria.

Carl Shulman is a Research Associate at the Future of Humanity Institute, Oxford Martin School, Oxford University, where his work focuses on the long-run impacts of artificial intelligence and biotechnology. He is also an Advisor to the Open Philanthropy Project. Previously, he was a Research Fellow at the Machine Intelligence Research Institute and held positions at Clarium Capital Management and Reed Smith LLP. He attended New York University School of Law and holds a degree in philosophy from Harvard University.

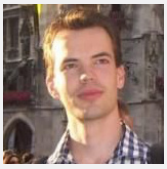
Scott Siskind practices psychiatry in Michigan, is interested in rationality and existential risk, and blogs at slatestarcodex.com.

Andrew Snyder-Beattie is Director of Research at the Future of Humanity Institute (FHI) at Oxford University, where he coordinates the institute's research activities, recruitment, and academic fundraising. While at FHI, Andrew obtained over $2.5m in research funding, and led the FHI-Amlin industry research collaboration. His research interests include ecosystem and pandemic modelling, anthropic shadow considerations, and existential risk. He holds a M.S. in biomathematics and has done research in a wide variety of areas such as astrobiology, ecology, finance, risk assessment, and institutional economics.

Nate Soares is the executive director of the Machine Intelligence Research Institute (MIRI). He first joined MIRI in 2014 as a Research Fellow, quickly earning a strong reputation for his strategic insight and high productivity. Soares is the primary author of most of MIRI's technical agenda, including the overview document Agent Foundations for Aligning Superintelligence with Human Interests (2014) and the Association for the Advancement of Artificial Intelligence (AAAI) paper Corrigibility (2015). Prior to MIRI, Soares worked as a software engineer at Google.

Marin Soljačić is a professor of Physics at MIT. His main research interests are in electromagnetic phenomena, focusing on nanophotonics, non-linear optics, and wireless power transfer. Among many other awards, he received the Adolph Lomb medal from the Optical Society of America in 2005, the TR35 award of Technology Review magazine in 2006, MacArthur "Genius" Grant in 2008, as well as Blavatnik National Award in 2014. Soljačić has been a correspondent member of the Croatian Academy of Engineering since 2009, and in 2011, he became a Young Global Leader (YGL) of the World Economic Forum.

Jacob Steinhardt is a fifth-year graduate student in artificial intelligence at Stanford University working with Percy Liang. His main research interest is in designing machine learning algorithms that are reliable and easy for humans to reason about. He is interested in computationally-bounded reasoning, and he has done work on the foundations of resource-constrained learnability as well as online learning. Outside of research, he is a coach for the USA Computing Olympiad and an instructor at the Summer Program in Applied Rationality and Cognition. Steinhardt also consults part-time for the Open Philanthropy Project.



Bas Steunebrink is a postdoctoral researcher at the Swiss AI lab IDSIA and co-founder of NNAISENSE. His research interests include artificial general intelligence, machine learning, bounded rationality, and AI safety & ethics. Steunebrink earned his Ph.D. in Artificial Intelligence at Utrecht University in 2010. In 2015 he was awarded a grant from the Future of Life Institute to study how the growth of the understanding and ethics of a nascent AI should be shaped and measured by teaching.

Mustafa Suleyman is co-founder and Head of Applied AI at DeepMind, where he is responsible for integrating the company's technology across a wide range of Google products. In February 2016 he launched DeepMind Health, which builds clinician-led, patient-centred technology in the NHS. He is also co-founder of Reos Partners, a global conflict resolution firm specialising in addressing complex social challenges. As a skilled negotiator and facilitator, Mustafa has worked all over the world for a wide range of clients, such as the UN, the Dutch Government and WWF.

Ilya Sutskever received his Ph.D. in computer science from the University of Toronto, under the supervision of Geoffrey Hinton. He was a postdoctoral fellow with Andrew Ng at Stanford University for a brief period, after which he dropped out to co-found DNNResearch which Google acquired the following year. Sutskever joined the Google Brain team as a research scientist, where he developed the Sequence to Sequence model, contributed to the design of TensorFlow, and helped establish the Brain Residency Program. He is a co-founder of OpenAI, where he currently serves as research director.

Richard Sutton is a professor of computer science and iCORE chair at the University of Alberta. He is known for his contributions in the field of reinforcement learning, and he is the author of the original paper on temporal difference learning. Sutton is also co-author of the textbook Reinforcement Learning: An Introduction from MIT Press. He is a fellow of the Association for the Advancement of Artificial Intelligence (AAAI), and has been elected fellow of the Royal Society of Canada. His research papers have been cited nearly 50,000 times and his Youtube videos have been viewed about 18,000 times.

Jaan Tallinn is a founding engineer of Skype and Kazaa. He is a co-founder of Future of Life Institute, a co-founder of the Cambridge Centre for the Study of Existential Risk, and he philanthropically supports other existential risk research organizations such as the Future of Humanity Institute, the Global Catastrophic Risk Institute and the Machine Intelligence Research Institute. He is also a partner at Ambient Sound Investments, an active angel investor, and has served on the Estonian President's Academic Advisory Board. He has given a number of talks highlighting the potential risk from advances in artificial intelligence.

Alexander Tamas is a Founding Partner of Vy Capital, a global technology investment company. Previously, he was a founding member of DST Global where he spearheaded many of DST's investments including into companies such as Facebook, Twitter, Alibaba, Zalando, Airbnb and Spotify. Alexander was also a Managing Director and Board Member of Mail.ru Group in Russia which he helped to IPO in 2010. Beforehand he worked in the technology M&A group of Goldman Sachs in London. Together with the Future of Humanity Institute in Oxford, Alexander established an AI Safety Fellowship in 2011 with the aim of pushing forward the state of the art in control problem solutions.

Jessica Taylor is a research fellow at the Machine Intelligence Research Institute (MIRI). She is interested in questions related to probabilistic modeling: how to design software agents whose world-models integrate both higher and lower levels of analysis, agents that can reason with logical uncertainty, and ones that can acquire human concepts. She has an MSc in computer science from Stanford, where she studied machine learning and probabilistic programming.

Max Tegmark is a Professor of Physics at MIT, President of the Future of Life Institute, and Scientific Director of the Foundational Questions Institute. His research has ranged from cosmology to the physics of cognitive systems, and is currently focused at the interface between physics, AI and neuroscience. He is the author of over 200 publications and the book Our Mathematical Universe: My Quest for the Ultimate Nature of Reality. His work with the Sloan Digital Sky Survey on galaxy clustering shared the first prize in Science magazine's "Breakthrough of the Year: 2003."

Sam Teller is Director at the Office of the CEO at SpaceX and Tesla. He also serves as an adviser for OpenAI. Previously, Sam was Managing Director and Founding Partner of Launchpad LA, the leading startup accelerator in Southern California. He also co-founded Charlie in 2010, where he led new media strategy for top entertainment, corporate, and political clients, launched new media brands, and made angel investments.

Josh Tenenbaum is a professor of Computational Cognitive Science at MIT. His research focuses on one of the most basic and distinctively human aspects of cognition: the ability to learn so much about the world, rapidly and flexibly. His current passion is to understand the nature and origins of common sense: What makes any human toddler more intelligent than any machine ever built? Tenenbaum also works actively in artificial intelligence, believing that if we can build machines that learn, see, think and act in more human-like ways, this will lead to more useful and beneficial AI systems as well as more powerful tools for understanding the human mind.

Marty Tenenbaum is an AI and ecommerce pioneer and is currently Founder and Chair of Cancer Commons, a non-profit network of physicians, scientists, and patients that Newsweek dubbed the "LinkedIn of Cancer." He began his career in artificial intelligence and machine learning, leading elite research groups at SRI International and Schlumberger Ltd. He is a fellow and former board member of the AAAI, and a Director of CommerceNet and Efficient Finance. Previously, he served as a Director at Patients Like Me and the Public Library of Science (PLoS), and as a Consulting Professor of Computer Science at Stanford.

Kristinn Thórisson is an associate professor at the School of Computer Science at Reykjavik University, founder of the Icelandic Institute for Intelligent Machines (IIIM), and co-founder and former co-director of CADIA (Center for Analysis and Design of Intelligent Agents). He has worked extensively on systems integration for artificial intelligence systems in the past, contributing architectural principles for infusing dialogue and human-interaction capabilities into the Honda ASIMO robot. He was co-founder of semantic web startup company Radar Networks, and served as its Chief Technology Officer 2002-2003.

Helen Toner is a senior research analyst at the Open Philanthropy Project, where she works on AI strategy and policy (among other things). Before joining Open Philanthropy, Helen graduated from the University of Melbourne with a BSc in Chemical Engineering and a DipLang in Arabic.

Moshe Vardi is a professor in Computational Engineering and Director of the Ken Kennedy Institute for Information Technology at Rice University. He is the author and co-author of many papers and two books: Reasoning about Knowledge and Finite Model Theory and Its Applications. He is a Fellow of the Association for Computing Machinery (ACM), the American Association for Artificial Intelligence (AAAI), and the American Association for the Advancement of Science (AAAS). His interests focus on automated reasoning, a branch of artificial intelligence with broad applications to computer science, including database theory, computational-complexity theory, knowledge in multi-agent systems, computer-aided verification, and teaching logic across the curriculum.

Manuela Veloso is Herbert A. Simon University Professor in the School of Computer Science, and the Head of the Machine Learning Department at Carnegie Mellon University. She researches in the area of artificial intelligence, with focus on planning, execution, and learning in robotics. Veloso and her students have researched and developed a variety of autonomous robots, including teams of soccer robots, and mobile service robots. Her CoBot mobile service robots interact with humans in a symbiotic merge of strengths and limitations. Veloso is a Fellow of the ACM, IEEE, AAAS, and AAAI. She is the Past President of AAAI, and the co-founder, Trustee, and Past President of RoboCup. As of 2016, Veloso has graduated 34 PhD students and co-authored more than 300 journal and conference publications.

Kent Walker is Senior Vice President and General Counsel of Google Inc. He oversees Google's legal and policy teams, its trust & safety group working on product policies and enforcement, and Google.org. He previously worked for eBay, AOL, Netscape, Liberate Technologies, AirTouch Communications, and the U.S. Department of Justice. He graduated from Harvard College and Stanford Law School.

Wendell Wallach is a scholar, consultant, and author at Yale University's Interdisciplinary Center for Bioethics, where he chairs Technology and Ethics studies. He is also a senior adviser to The Hastings Center. His books include, A Dangerous Master: How to keep technology from slipping beyond our control, a primer on emerging technologies, and Moral Machines: Teaching Robots Right From Wrong (co-authored with Colin Allen), which mapped the then new field of enquiry called machine ethics or machine morality. He received the World Technology Award for Ethics in 2014 and for Journalism and Media in 2015, as well as a Fulbright Research Chair at the University of Ottawa for 2015-2016. The World Economic Forum appointed Mr. Wallach co-chair of its Global Future Council on Technology, Values, and Policy for the 2016-2018 term.

Toby Walsh is Guest Professor at Technical University of Berlin, Professor of Artificial Intelligence at the University of New South Wales, and leads the Algorithmic Decision Theory group at Data61, Australia's Centre of Excellence for ICT Research. He has also been Editor-in-Chief of the Journal of Artificial Intelligence Research, and of AI Communications. Walsh has been elected a fellow of the Australian Academy of Science, and has won the prestigious Humboldt research award. He has played a key role at the UN and elsewhere on the campaign to ban lethal autonomous weapons (aka "killer robots").

Daniel Weld is Professor of Computer Science & Engineering and Entrepreneurial Faculty Fellow at the University of Washington. He landed a Ph.D. from the MIT Artificial Intelligence Lab in 1988, was named Association for the Advancement of Artificial Intelligence (AAAI) Fellow in 1999, and was deemed Association for Computing Machinery (ACM) Fellow in 2005. Additionally, Weld is an active entrepreneur with several patents and technology licenses. He co-founded Netbot Incorporated, creator of Jango Shopping Search, AdRelevance, a monitoring service for internet advertising, and data integration company Nimble Technology.

Adrian Weller is a senior researcher in the Machine Learning Group at the University of Cambridge, in the Computational and Biological Learning Lab. He received a PhD in computer science (machine learning) at Columbia University. Most of his academic research relates to graphical models but he is also very interested in other areas including: finance, anything on intelligence (natural or artificial), networks, robust learning, interpretability, deep learning, reinforcement learning, ethics, social policy, music and methods for big data. He is a faculty fellow at the Alan Turing Institute (ATI) and an executive fellow at the Leverhulme Centre for the Future of Intelligence (CFI). He previously held senior positions in finance.

Michael Wellman is Lynn A. Conway Collegiate Professor of Computer Science & Engineering at the University of Michigan. He received a PhD from the Massachusetts Institute of Technology in 1988 for his work in qualitative probabilistic reasoning and decision-theoretic planning. For the past 20+ years, his research has focused on computational market mechanisms and game-theoretic reasoning methods, with applications in electronic commerce, finance, and cybersecurity. Wellman previously served as Chair of the ACM (Association for Computing Machinery) Special Interest Group on Electronic Commerce (SIGecom), and as Executive Editor of the Journal of Artificial Intelligence Research.

Meredith Whittaker is the founder and lead of Google's Open Research Group, where she architected and created novel models for cross-industry and cross-sector research. In this capacity, she founded Measurement Lab, a globally distributed open measurement system that provides the largest open, verifiable source of data on internet performance. She co-founded Simply Secure, helped build and currently advises the Open Technology Fund, and is a leader in the development of core infrastructure security work within Google and beyond. She has advised the White House, the FCC, the City of New York, the European Parliament, and many civil society organizations on artificial intelligence, internet policy, measurement, data ethics, privacy, and security.

Roman Yampolskiy is an Associate Professor in the department of Computer Engineering and Computer Science at the Speed School of Engineering, University of Louisville. He is the founding and current director of the Cyber Security Lab and an author of many books including Artificial Superintelligence: a Futuristic Approach. Yampolskiy is a Senior member of the Institute of Electrical and Electronics Engineers (IEEE) and a Research Advisor for the Machine Intelligence Research Institute (MIRI) and Associate of the Global Catastrophic Risk Institute (GCRI). His research focuses on AI safety, behavioral biometrics, cybersecurity, digital forensics, games, genetic algorithms, and pattern recognition.

Eliezer Yudkowsky is a decision theorist who is widely cited for his writings on the long-term future of artificial intelligence. His views on the social and philosophical implications of AI have had a major impact on ongoing debates in the field, and his work in mathematical logic has heavily shaped MIRI's research agenda. He is Senior Research Fellow at MIRI and has written a number of popular introductions to the science of human rationality.

Brian Ziebart is an Assistant Professor in the Department of Computer Science at the University of Illinois-Chicago (UIC). He received his PhD from Carnegie Mellon University where he was also a postdoctoral fellow. His research interests include machine learning, decision theory, game theory, robotics, and assistive technologies. Applications of his research include Pedestrian Trajectory Forecasting, which improves robots' movements around people, and Assistive Navigation Systems that can reason about observed driving to learn the driver's preferences.

Shivon Zilis is a partner and founding member of Bloomberg Beta, an early stage venture capital fund, where she invests in entrepreneurs using machine intelligence to solve real-world problems. Beta has invested in 40 machine intelligence companies in areas such as agriculture, education, healthcare, and enterprise intelligence. She's a fellow at University of Toronto's machine learning accelerator, board member at Alberta Machine Intelligence Institute, and advisor at OpenAI. Hockey, photography, and piano make her smile during her free time.