The report editor can be reached at

globalaigovernance@gmail.com

We welcome any comments on this report

and any communication related to AI

governance.

SHANGHAI INSTITUTE FOR SCIENCE OF SCIENCE

weisite：www.siss.sh.cn

mailbox：siss@siss.sh.cn

AI GOVERNANCE IN 2019
A YEAR IN REVIEW

OBSERVATIONS OF 50 GLOBAL EXPERTS

April, 2020

Shanghai Institute for Science of Science

ALL LIVING THINGS ARE NOURISHED
WITHOUT INJURING ONE ANOTHER,
AND ALL ROADS RUN PARALLEL
WITHOUT INTERFERING WITH
ONE ANOTHER.

—— *CHUNG YUNG*, SECTION OF THE *LI CHI*

萬物并育而不相害，
道并行而不相悖。

——《禮記·中庸》

TABLE OF CONTENTS

FOREWORD

Artificial intelligence (AI) is an important driving force for a new round of scientific and technological revolution and industrial transformation, which will bring significant changes to people's lives.

In recent years, countries around the world have continued to issue AI strategies and policies. The technological R&D and the industrial application of AI is thriving. In 2017, the State Council of China issued "Development Planning for a New Generation of Artificial Intelligence" as China's national strategic plan on AI development, which outlined the basic framework for China's AI development before 2030. In February 2019, the National New Generation AI Governance Expert Committee consisting of AI experts from academia and industry was established by China's Ministry of Science and Technology. In June 2019, the Committee released the "Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence", addressing eight governance principles: harmony and human-friendliness, fairness and justice, inclusiveness and sharing, respect for privacy, security and controllability, shared responsibility, open collaboration, and agile governance. With these strategies and principles, China hopes to better coordinate the development and governance of the emerging technology and to ensure secure, controllable and reliable AI. In Shanghai, AI has been designated as a priority development area and an efficient tool for future urban governance. However, the effective governance of AI is the key to ensuring its success. Meanwhile, China, at the national level, also pins high expectations on Shanghai's AI development and governance. In 2019, Shanghai was designated as the National New-Generation AI Innovation and Development Pilot Zone, which emphasized its role of exploring issues related to the AI governance and ethics. Shanghai is also expected to become a national exemplar of AI development.

Established in January 1980, the Shanghai Institute for Science of Science (SISS) is one of China's earliest soft science research institutes. It conducts research to inform decision-making on innovation policy. It focuses on fields such as science, technology and innovation strategies, public policies and industrial technology innovation. It is dedicated to building a professional and platform-type science, technology and innovation think tank.

This year marks the 40th anniversary of SISS. 40 years ago, China started its process of Reform and Opening Up. Two major questions were considered at the time, with aims to bring order and to restore normality for the country's governance system: What is the development pattern for science and technology? How do they influence the economy and society? The founders of SISS called for study on the subject "science of science",in order to bring answers to those questions. They conducted in-depth discussions on the emerging science and technology on the topic of "new science and technology revolution", which influenced China's national and Shanghai's local science and technology strategies.

40 years later, the understanding of science and technology in China has changed deeply and its capacity in science and technology development is strengthened. However, we are still facing complex issues from the subject area "science of science". In recent years, various technologies including big data, internet and AI have emerged, exerting profound and transformative influences on the economy, society, culture and international relations.

We are very fortunate that there is a general global consensus on building cooperative relations in science and technology. This is particularly the case for AI governance, which shapes the common fate of humanity. Therefore,through this report, we hope to work with our global colleagues, track progress made by various parties in this field and lay the foundation for exchanges and cooperation. Together, we can achieve more.

EDITOR-IN-CHIEF : SHI QIAN

SHI Qian is the director of the Shanghai Institute for Science of Science (SISS). Before joining SISS, Professor SHI was the vice president of the Shanghai Academy of Sciences & Technology and concurrently the vice president of the Shanghai Institute of Industrial Technology. He has been long engaged in the general planning for science and technology development, research project management, innovation platform building, and services for innovation and entrepreneurship. Professor SHI participated in the formulation of a number of national industrial development plans and the implementation of major national science and technology projects, where he presided over several soft science research projects, such as "Research on Shanghai's Medium and Long-Term (2021-2035) Developmental Strategy of Science and Technology" from the government of Shanghai. Professor SHI obtained the Shanghai Special Award for Scientific and Technological Progress in 2016. Professor SHI is also the director of Technology Foresight Committee of the Chinese Association for Science of Science and S&T Policy, and the deputy director of the Expert Advisory Committee of the National New-Generation AI Innovation and Development Pilot Zone in Shanghai.

INTRODUCTION

The impact of emerging technologies might be a seminal inflection point in human history that will continually impact all aspects of society over the coming decades. In that, AI is the linchpin accelerating and amplifying the development of all the fields of research. With the rapid development of machine learning in recent years, the governance of the technology has gradually come under the spotlight. It was once possible to keep track of all the research institutes, conferences and policy developments. In 2019, this became an arduous task for researchers and policymakers. The number of initiatives continued to grow. There is a much greater variety of regional perspectives. The diversity of stakeholders participating in this dialogue has increased. The idea that the world urgently needs to find a path towards developing ethical and beneficial AI for all of humanity has become front-and-center in our media and public conversations. Despite the scientific and policy difficulties, it seems that the world is willing to rise up to this challenge.

One way to think of the governance of AI is that it is a 'wisdom race'. The late Stephen Hawking once said that "our future is a race between the growing power of our technology and the wisdom with which we use it. Let's make sure that wisdom wins." To take stock of and share the wisdom, we decided to invite 50 world-class experts (44 institutions) to share their views on the key progress in AI governance in 2019. We hope that this can help separate the signal from the noise for interested readers.

These experts include scientists who have made major contributions to the field of AI.

They approach the question of social impact scientifically and offer technical solutions to the challenge of AI governance. For example, John Hopcroft, a professor at Cornell University and a winner of the Turing Award, points out that the development of current AI systems has the possibility of bias caused by bias in the training data. Stuart Russell, a professor at the University of California, Berkeley, wrote an AI textbook used by more than 1,300 universities in 116 countries. He and his colleague, Caroline Jeanmaire, high-light the importance of conducting technical research on provably beneficial AI as argued in his recent book *Human Compatible*. Yang Qiang, a professor at the Hong Kong University of Science and Technology and General Chair of AAAI 2021, advocates the development of federated learning for addressing privacy issues, which is among the top concerns in AI governance today. Pascale Fung, professor at the Hong Kong University of Science and Technology, makes a general case for developing formal processes for ethical AI systems and specifically proposes the establishment of a standardized algorithm review system. Roman Yampolskiy, an expert in AI security at University of Louisville in the United States, argues that we should not only discuss ethical issues, but also pay attention to the safety and security issues of AI systems. These views from the scientists suggest a technically grounded direction for AI governance in 2019 and beyond.

The emergence of AI governance issues has attracted the attention of experts in the field of traditional humanities and social sciences, which helped open up new research directions.

Allan Dafoe, an expert in international relations studies and Director of the Centre for the Governance of AI, University of Oxford, and his colleague Markus Anderljung, survey the sudden proliferation of professional research institutions, company initiatives and government agencies dedicated to addressing the social impact of AI. It indicates that the field of AI governance research is becoming rapidly institutionalized. Legal scholar Gillian K. Hadfield recently established a new research institute at the University of Toronto, with the mission of focusing on the methodological question of effective value alignment in AI. SU Jun, a professor at the School of Public Policy & Management at Tsinghua University, shares his experience of using social experiments to conduct policy research during the transformation of the social, political or technological environment. Thilo Hagendorff, an AI ethicist at the University of Tübingen, stresses that a transition from 'soft law' to 'hard law' is the next step in AI governance. These discussions are signs that AI governance is becoming a serious intellectual discipline.

At the frontiers of AI applications, industry leaders and investors are paying closer attention to the influence of AI governance on the future of innovation.

As a member of the National New Generation Artificial Intelligence Governance Expert Committee, and the founder of the Chinese AI unicorn company Megvii, Yin Qi suggests that companies need to take more responsibilities in advancing AI governance. Don Wright, former President of the IEEE Standards Association, introduces IEEE's code of AI

ethics first released in 2017 within the framework of corporate governance. Being at the center of the controversy with the language learning model GPT-2, members of OpenAI's policy team offer their reflections on publication norms. This is followed by the perspectives on the malicious use of AI by two observers, namely Seán Ó hÉigeartaigh, Director of the "AI: Futures and Responsibility" Programme at the Leverhulme Centre for the Future of Intelligence (LCFI) of University of Cambridge, and Helen Toner, Director of Strategy at the Center for Security and Emerging Technologies (CSET) of Georgetown University. Millie Liu, Managing Partner at First Star, provides a practical point of view from the frontline by listing some of the key challenges for industry implementation of AI ethics. Steve Hoffman, a Silicon Valley investor, suggests that policymakers should harness the market forces for AI governance as companies would play an inevitable role in making progress in the field.

The issue of AI governance is a concern to scientists, scholars of humanities and the social sciences, as well as policy makers.

2019 might turn out to be the year when AI governance became a truly global issue with significant implications for global governance. We began this section with the discussion from Irakli Beridze, the Head of the Centre for AI and Robotics, at the United Nations, who was one of the recipients of the Nobel Peace Prize awarded to the Organisation for the Prohibition of Chemical Weapons. He argues that we should appreciate both the ethical issues and the positive effect of AI on solving global challenges in the context of law enforcement. Wendell Wallach, a professor and a science and technology ethicist at Yale University, proposes agile, cooperative and comprehensive governance. Three experts including Cyrus Hodes, Nicolas Miailhe, and Jessica Cussins Newman all share the reflection that the OECD made substantial progress in the governance of AI in 2019. From their discussions, we observe that there is a converging consensus from around the world. CHEN Dingding, an expert in international issues and professor at Jinan University in China, discusses the issues of AI governance from the perspective of international relations.

While being increasingly globalized, there is a parallel trend of localizing AI principles in different regions of the world.

The European Union is an active leader in the field of AI governance. Eva Kaili, a member of the European Parliament, presents the European Parliament's main work on AI governance and plans for the future. In 2019, the European Union released the "Ethics Guidelines for Trustworthy AI", which attracted global attention. Francesca Rossi, the AI Ethics Global Leader and a Distinguished Research Staff Member at IBM Research and a member of the EU High-Level Expert Group on Artificial Intelligence, believes that such multi-disciplinary and multi-stakeholder composition of the expert group should serve as a leading example for AI governance. Charlotte Stix, a wellrespected analyst of European AI policy, analyzes the European Union's approach towards "trustworthy AI". Shortly after Brexit, Angela Daly from Strathclyde University discusses the British government's understanding of AI governance, especially the role of the Centre for Data Ethics and Innovation as a specialized institution.

There were also significant developments in other parts of Asia. Danit Gal, technology advisor to the UN Secretary General High-level Panel on Digital Cooperation, observes that the region has a significant traditional cultural imprint on AI ethics and governance. Arisa Ema from the University of Tokyo, who participated in the formulation of the Japanese Cabinet's Social Principles of Human-centric AI, discusses the shift from the government to the industry as the key driver for AI governance development in Japan. Singapore made great achievements in AI governance in 2019 and won the highest award at the World Summit on the Information Society Forum, an UN-level platform. Having contributed to such an achievement, Director of the Singapore Management University Centre for AI & Data Governance (CAIDG) Goh Yihan and his colleague Nydia Remolina, research associate at CAIDG, introduce the Singaporean approach of translating ethical principles into pragmatic measures that businesses can adopt. Based in India, Urvashi Aneja from Tandem Research suggests that the key challenge for Indian policy is striking a balance between equity and growth in the AI era.

Although China has made remarkable achievements in AI R&D and industrial applications, there is a relative lack of international discussions about its approach and progress in AI governance.

Therefore, we invited some of the key policy advisors and experts on China's AI governance to introduce the current status in the country.

FU Ying, former Vice Minister of Foreign Affairs of China and Director of the Center for International Strategy and Security at Tsinghua University, makes a powerful case that the world should cooperate on the issue of AI governance, which requires first and foremost the partnership between China and the United States as major countries. ZHAO Zhiyun , Director of New-Generation Artificial Intelligence Development Research Center of Ministry of Science and Technology, shares the Chinese government's views and recent progress on AI governance. LI Xiuquan, Research Fellow of Chinese Academy of Science and Technology for Development, emphasizes the approach of inclusive development in China's AI governance, with a focus on protecting the vulnerable groups in the society. DUAN Weiwen, a professor and philosopher of science at the Chinese Academy of Social Sciences, discusses the need to construct trust mechanisms for AI for building an agile governance framework. LUAN Qun from the China Center for Information Industry Development under the Ministry of Industry and Information Technology of China surveys the progress in ethical governance in China's AI industry. GUO Rui from Renmin University of China, who participated in related work of the

China Artificial Intelligence Standards Committee, discusses the foundational philosophy in the formulation of standards. It is worth mentioning that in the promotion of AI governance by the Chinese government, one of the key policy tools is setting some provinces and cities as AI "pilot zones". As the largest city in China, Shanghai was approved as such a pilot zone in 2019. Dr. WANG Yingchun from the Shanghai Institute of Science introduces the current situation. The experts we invited this time are representatives from the government and academia. We hope to have the opportunity to extend the conversations with the industry, given that many Chinese companies are actively exploring the issue of AI governance.

From the comments of all experts – from the standpoint of science and technology, of humanities and social sciences, of international relations and of countries and regions, progress in general consensus can be observed in 2019. For example, there is an increasing number of professional institutions being established, a growing degree of global consensus, and a convergence of attention from industry and policymaking communities.

We welcome the readers to share their view on commonalities by reading these contributions from experts. Ultimately, we hope that this report can serve as a launchpad for this consequential conversation of our generation. As the late Alan Turing would say, "we can only see a short distance ahead, but we can see plenty there that needs to be done."

EXECUTIVE EDITORS: LI HUI; BRIAN TSE (INVITED)

LI Hui is an associate professor at the Shanghai Institute for Science of Science. He regularly participates in the formulation of AI strategies for Shanghai as well as on a national level. He also frequently publishes his views on AI governance in major Chinese media such as *People's Daily, Guangming Daily and Wenhui Daily*. He has played a prominent role in organizing the Governance Forum of the World Artificial Intelligence Conference 2019. He earned his PhD in history of science from Shanghai Jiao Tong University in 2011. His background led to his research interests on issues related to AI governance with a long-term perspective and global thinking.

Brian Tse is an independent researcher and consultant working on the governance, safety and international relations of AI. Brian is a Senior Advisor at the Partnership on AI and a Policy Affiliate at the University of Oxford's Centre for the Governance of AI. He has advised organizations including Google DeepMind, OpenAI, Baidu, Tsinghua University Institute of AI, Beijing Academy of AI and Carnegie Endowment for International Peace.

ACKNOWLEDGEMENT

The motivation of this report is to promote exchanges and communication between academic researchers, policy makers, and industry practitioners in this rapidly changing field. It is fortunate that our initiative has received extensive attention and support from our global peers. First and foremost, we would like to express our appreciation to all the 50 experts for their contributions.

Our sincere appreciation goes to John Hopcroft, who has extended his very generous offer in providing guidance to our work. In addition, we would like to express our gratitude to Stuart Russell, Wendell Wallach and Irakli Beridze for their valuable suggestions on the overall framework of the report after reading the first draft.

From the initial idea of the report to its final release, YU Xindong, WANG Yingchun and SONG Jia from the Shanghai Institute for Science of Science gave valuable support to the development and promotion of the project.

LI Xiuquan (China Academy of Science and Technology Development Strategy), Cyrus Hodes (Future Society), Dev Lewis (Digital Asia Hub), Herbert Chia (Sequoia Capital China), DUAN Weiwen (Chinese Academy of Social Sciences) and HE Jia, has provided valuable supports in bringing all the contributors together.

In the process of editing the report, young scholars such as Caroline Jeanmaire (University of California at Berkeley), Thilo Hagendorff (University of Tuebingen), Jessica Cussins Newman (University of California at Berkeley), Charlotte Stix (Eindhoven University of Technology), Angela Daly (Strathclyde University), Kwan Yee Ng (University of Oxford), Jeff Cao (Tencent) , XU Nuo (Shanghai Institute for Science of Science), QU Jingjing (Shanghai Institute for Science of Science) and ZHANG Chaoyun (Shanghai Institute for Science of Science) provided valuable support in editing and proofreading the report. ZHANG Dazhi (Central China Normal University) helped us design the illustration in the report.

Interns ZHANG Jie, SONG Zhixian, SUN Hui, NI Jiawei, and LIANG Xinyi has undertaken a large volume of operational work.

To all colleagues and friends that have provided help, we would like to express our sincere gratitude.

PART 1 TECHNICAL PERSPECTIVES FROM WORLD-CLASS SCIENTISTS

The Importance of Talent in the Information Age

By John Hopcroft

Deep learning has had a major impact on AI even though it is only one technique in the AI tool box. It has been applying in many experimental areas such as image recognition, machine translation, finance, etc. Now that AI is having significant applications, it has raised many issues. If an AI program is making decision say for loans, people want to know why the program made a decision. At the current state of knowledge, we do not know how to answer question like these. Another issue concerns the possibility of bias caused by bias in the training data.

It is clear that a revolution is occurring with AI as a major driver. In the future talent will be the main contribution to a nation's economy and standard of living. The most important issue for China is to improve the quality of undergraduate education to provide the talent for China to become the leading economy in the information age.

ABOUT THE AUTHOR **John E. Hopcroft**

John E. Hopcroft is the IBM Professor of Engineering and Applied Mathematics in Computer Science at Cornell University. From January 1994 until June 2001, he was the Joseph Silbert Dean of Engineering. After receiving both his M.S. (1962) and Ph.D. (1964) in electrical engineering from Stanford University, he spent three years on the faculty of Princeton University. He joined the Cornell faculty in 1967, was named professor in 1972 and the Joseph C. Ford Professor of Computer Science in 1985. He served as chairman of the Department of Computer Science from 1987 to 1992 and was the associate dean for college affairs in 1993. An undergraduate alumnus of Seattle University, Hopcroft was honored with a Doctor of Humanities Degree, Honoris Causa, in 1990.

Hopcroft's research centers on theoretical aspects of computing, especially analysis of algorithms, automata theory, and graph algorithms. He has coauthored four books on formal languages and algorithms with Jeffrey D. Ullman and Alfred V. Aho. His most recent work is on the study of information capture and access.

He was honored with the A. M. Turing Award in 1986. He is a member of the National Academy of Sciences (NAS), the National Academy of Engineering (NAE), a foreign member of the Chinese Academy of Sciences, and a fellow of the American Academy of Arts and Sciences (AAAS), the American Association for the Advancement of Science, the Institute of Electrical and Electronics Engineers (IEEE), and the Association of Computing Machinery (ACM). In 1992, he was appointed by President Bush to the National Science Board (NSB), which oversees the National Science Foundation (NSF), and served through May 1998. From 1995-98, Hopcroft served on the National Research Council's Commission on Physical Sciences, Mathematics, and Applications.

In addition to these appointments, Hopcroft serves as a member of the SIAM financial management committee, IIIT New Delhi advisory board, Microsoft's technical advisory board for research Asia, and the Engineering Advisory Board, Seattle University.

From the Standard Model of AI to Provably Beneficial Systems

By Stuart Russell and Caroline Jeanmaire

AI governance made notable progress on 2019. First, important sets of principles were published, notably the Beijing AI principles and the OECD Principles on AI. Both focus particular attention on ensuring the security of AI systems in the short and long terms, an essential aspect of AI development.

Principles are a good foundation for action, and indeed we also saw instances of concrete action. California became the first state to require all automated online accounts attempting to influence residents' voting or purchasing behaviors to openly identify as robots. This law represents an important first step towards curbing deceptive new technology and making AI systems trustworthy; it is a step towards establishing a basic human right to know whether one is interacting with another human or with a machine. The law will also hinder the spread of misinformation. We hope that the law will develop beyond commercial and voting issues to become a general right, and also serve as a precedent for other states and countries.

In some areas, however, governance dangerously lags behind. Our global community made very little progress in regulating Lethal Autonomous Weapons (LAWs) such as drones, tanks, and other computer-controlled machinery. These technologies run on AI systems and are programmed to locate, select and attack targets without human control. At the November 2019 meeting of member states of the Convention on Certain Conventional Weapons (CCW) at the United Nations in Geneva, diplomats could not agree on a binding common approach towards this issue. As a result, the next two years will be spent on non-binding talks instead of concrete legal work in order for us to move towards a global ban on lethal autonomous weapons to safeguard our common future.

As we develop increasingly capable AI systems that become highly competent and self-sustaining, humans must ensure that these AI systems remain beneficial and safe. Russell, one of the co-authors of this article, just published a book on this topic: *Human Compatible: Artificial Intelligence and the Problem of Control* (Viking/Penguin, 2019). The problem of control over AI systems is not the science fiction plot that preoccupies Hollywood and the media with a humanoid robot that spontaneously becomes conscious and decides to hate humans. It is rather the creation of machines that can draw on more information and look further into the future than humans can, exceeding our capacity for decision making in the real world. With our present conception of AI and our technical approach, there is no plausible prospect of retaining control over machines more powerful than ourselves. To solve this problem, the research community needs to undertake a vast effort to change the standard model in AI towards provably beneficial systems. The AI community is becoming aware of this issue, which makes us hopeful that we will be able to achieve this transformation, but there is much work to do.

ABOUT THE AUTHOR

Stuart Russell

Stuart Russell received his B.A. with first-class honors in physics from Oxford University in 1982 and his Ph.D. in computer science from Stanford in 1986. He then joined the faculty of the University of California at Berkeley, where he is Professor (and formerly Chair) of Electrical Engineering and Computer Sciences, holder of the Smith-Zadeh Chair in Engineering, and Director of the Center for Human-Compatible AI. He has served as an Adjunct Professor of Neurological Surgery at UC San Francisco and as Vice-Chair of the World Economic Forum's Council on AI and Robotics. He is a recipient of the Presidential Young Investigator Award of the National Science Foundation, the IJCAI Computers and Thought Award, the World Technology Award (Policy category), the Mitchell Prize of the American Statistical Association, the Feigenbaum Prize of the Association for the Advancement of Artificial Intelligence, and Outstanding Educator Awards from both ACM and AAAI. From 2012 to 2014 he held the Chaire Blaise Pascal in Paris, and he has been awarded the Andrew Carnegie Fellowship for 2019 to 2021. He is an Honorary Fellow of Wadham College, Oxford; Distinguished Fellow of the Stanford Institute for Human-Centered AI; Associate Fellow of the Royal Institute for International Affairs (Chatham House); and Fellow of the Association for the Advancement of Artificial Intelligence, the Association for Computing Machinery, and the American Association for the Advancement of Science. His book *Artificial Intelligence: A Modern Approach* (with Peter Norvig) is the standard text in AI; it has been translated into 14 languages and is used in over 1400 universities in 128 countries. His research covers a wide range of topics in artificial intelligence including machine learning, probabilistic reasoning, knowledge representation, planning, real-time decision making, multitarget tracking, computer vision, computational physiology, and philosophical foundations. He also works for the United Nations, developing a new global seismic monitoring system for the nuclear-test-ban treaty. His current concerns include the threat of autonomous weapons and the long-term future of artificial intelligence and its relation to humanity. The latter topic is the subject of his new book, *Human Compatible: Artificial Intelligence and the Problem of Control* (Viking/Penguin, 2019).

Caroline Jeanmaire

Caroline has a Master's degree in International Relations from Peking University and a Master's degree in International Public Management from Sciences Po Paris. She received her Bachelor's degree in political sciences from Sciences Po Paris. She also studied at the Graduate Fletcher School of Law and Diplomacy and at Tufts University. Caroline researches international coordination models to ensure the safety and reliability of Artificial Intelligence systems at the Center for Human-Compatible AI (CHAI) at UC Berkeley. She also leads CHAI's partnership and external relations strategy, focusing on building a research community around AI safety and relationships with key stakeholders. Before working at CHAI, she was an AI Policy Researcher and Project Manager at The Future Society, a thinktank incubated at Harvard's Kennedy School of Government. She notably supported the organization of the first and second Global Governance of AI Forums at the World Government Summit in Dubai. In the 2019 edition, she managed two committees: Geopolitics of AI and International Panel on AI research. She published articles and reports on the Geopolitics of AI, US-China industry levers of cooperation on AI and the results of a global civic debate on AI governance. Before this, she participated in numerous climate negotiations and technical intersessions since 2015, including with the French Delegation for COP23 and COP24. Caroline speaks English, French, Spanish and Mandarin Chinese.

The Importance of Federated Learning

By YANG Qiang

As AI moves out of the laboratory and into large-scale application, its potential ethical problems and impacts gradually arouse public concern. Looking back on 2019, the public discussions related to AI ethics focused on the protection and governance of user data privacy. Internationally, Facebook has been fined $5 billion by the US Federal Trade Commission (FTC) for illegally leaking user data. Also, Google was fined tens of millions of euros by French regulators for breaching the GDPR by making its privacy terms too complex for users to understand and too difficult for users to manage the way their personal data was used. In China, data companies have been intensively investigated by regulators for abusing and selling unauthorized users' privacy data. And a large number of data companies have been penalized by business suspension, app removal and even criminal liability for serious cases. This series of events shows that, on the one hand, the public's awareness of data rights related to personal privacy is gradually rising, so these events have attracted wide attention in the media and the public; and on the other hand, the shocking truths of the incidents also indicate that the protection and governance of private data is seriously lagging behind and missing.

Tracing back to the source, these problems are caused by the objective incentives that AI technology relies heavily on massive data collection, but more by the neglect of social responsibility and subjective reckless manners of relevant stakeholders. How to dig out the knowledge and value behind the data on the premise of fully respecting and protecting user data privacy is an imminent challenge facing AI researchers.

Fortunately, 2019 also witnessed AI researchers who have realized the seriousness of the problem and come up with a set of solutions. Among them, Federated Learning, as a promising user data privacy protection scheme, has demonstrated its unique advantages in promoting the implementation of industrial applications. Federated Learning refers to a technical scheme to realize joint modeling of multiple participants by exchanging encryption parameters on the premise that the data is not out of the locality and data is not shared, and its modeling effect is the same as or not much different from that of the aggregation modeling of the entire data set. A variety of encryption techniques are used in the Federated Learning technology framework, such as secure multiparty computing, homomorphic encryption (HE), Yao's garbled circuit and differential privacy (DP). From the perspective of technology application, current Federated Learning has been applied in such fields as small and micro enterprise credit, anti-money laundering, anti-fraud, insurance, and computer vision. In addition, it has been explored for application in such fields as smart medical treatment, autonomous driving, smart city, and government governance. To sum up, Federated Learning can be regarded as an integrator of machine learning technology and privacy protection technology, and also a universal privacy protection machine learning technology with wide application prospect.

ABOUT THE AUTHOR **YANG Qiang**

Prof. YANG is the the Chief AI Officer at WeBank and a Chair Professor and former Head of the Department of Computer Science and Engineering of the Hong Kong University of Science and Technology.

He is a leading researcher of "transfer learning" technology in the international AI community, and he is spearheading a new research direction of "Federated Learning". He was elected a fellow of AAAI (Association for the Advancement of Artificial Intelligence) in July 2013, and the Conference Chair of AAAI 2021 conference. Between 2017 and 2019, he was elected the President of the Board of Trustees of IJCAI, the world's oldest and most popular AI society.

Towards A Formal Process of Ethical AI

By Pascale Fung

Much has been discussed about the governance of AI in different government and societal contexts. New AI strategies and governance documents were proposed in 2019 by the UN, UNESCO, the EU, European Parliament, the governments of China, the US, Japan, the UAE, etc. Top AI companies in the world are working actively in research and development of ethical and beneficial AI, as well as good governance. The latest pronouncement by the CEO of Google that AI applications cannot be determined by market forces alone but needs good governance illustrates the general consensus in the AI community.

All machines make mistakes, but AI errors provoke more fear among people because, just like AI decisions, AI errors are so human-like. Consumers tend to associate such errors with nefarious human-like intentions. If a speaker recorded my conversations or a camera sent me images of someone else's homes, then the AI is "spying". If a search result is biased, it is "sexist" or "racist". If a chatbot gives the wrong answer, it can sound "scary" or "offensive". Suddenly, engineers who are used to dealing with system performance as numbers in a metric are confronted with a society of users who are constantly seeking for philosophical and even legalistic answers. Therefore, our research community is caught off guard. At the level of AI algorithm and system development, researchers and engineers strive for a fair, accountable and transparent process by virtue of both best practice guidelines and formal processes while mitigating and minimizing machine bias and machine error. Nowadays, it is common practice for researchers and developers to release databases, trained models and software codes to the public domain for others to use. Therefore, inherent biases in these databases and models can be propagated to all systems developed based on them.

Professional organizations like the IEEE have provided best practice guidelines in the form of Ethically Aligned Design process. We can apply these principles to all areas of AI algorithm and system development. NGOs such as the Partnership on AI has dedicated working groups aimed at providing best practice guidelines, with expert input from its members of engineers, philosophers, and civil society representatives. The International Organization for Standardization (ISO) with 164 member nations, including the US and China, is working on standardizations in the area of AI. There have been increasing calls for a formal process of AI and ML development that parallels that of the software engineering process as an integral part of AI software product development. A formal process recognized by AI professionals will ensure common standard, a more explainable and verifiable development process, and fewer system errors. A formal process can include standards for

1) Database collection: Data bias should be mitigated before it is released to the larger AI community;

2) Software and algorithm design: Conversational AI should be non-discriminatory; instead of just relying on voice print or facial recognition, biometric recognition should be multimodal to reduce errors;

3) Model training: Specific model architecture and parameter settings are recorded so that the process can be reproduced and interpreted down the pipeline without the need for human trial and error;

4) Testing and verification: Machine fairness and bias can also be evaluated and tested on standard test sets. Many AI conferences already run shared tasks where different groups compare their systems using common training and testing sets. This can abstract and formalize the development of AI algorithms and systems without stifling creativity and safety of research and safe guarding academic independence.

The European Parliament has called for a central regulatory body, much like the Food and Drug Administration, to assess the impact of algorithms before they are deployed. This proposal faces two challenges – 1) algorithms evolve at a breakneck speed and are modified and updated every few months; 2) there might not be enough experts available with the technical knowledge required for algorithm evaluation. Instead, I suggest that such a regulatory body be tasked to assess AI products and applications, rather than the underlying algorithms. Algorithm evaluation should be incorporated into the normal peer-review process of research publications. Editors and technical program chairs tasked to curate these publications should ask reviewers to provide explicit opinions on the ethical issues of the work they are reviewing. With AI professionals' increasing awareness of the ethics of their work, it is my hope that our collective wisdom will improve on this.

More international cooperation is required in AI governance as AI technologies developed today have become open resources and are shared quickly around the world. AI research and education are global today. Companies are working together on standards for autonomous driving. Countries are working together on regulating autonomous weapons. Applications of AI in the areas of security, healthcare, and finance are subject to existing regulations of each region, even though additional regulations are needed to account for algorithm and methodology evolution. Social media and information integrity remains a challenging area where social media companies are currently regulating themselves without consensus. More international cooperation is required and regulatory bodies need to be set up with AI experts and other stakeholders. In 2019 we have seen a more detailed AI governance plan and even more public awareness of its need. In 2020 and beyond, we need to work actively in implementing the proposed good practice guidelines and a formal software process to ensure fairness, accountability and transparency of AI systems.

ABOUT THE AUTHOR **Pascale Fung**

Pascale Fung is a Professor at the Department of Electrical and Electronic Engineering at The Hong Kong University of Science & Technology (HKUST). She is an elected Fellow of the Institute of Electrical and Electronic Engineers (IEEE) for her "contributions to human-machine interactions", and an elected Fellow of the International Speech Communication Association for "fundamental contributions to the interdisciplinary area of spoken language human-machine interactions". She is the Director of HKUST Center for AI Research (CAiRE), the leading interdisciplinary research center among all four schools at HKUST. She is an expert at the Global Future Council, a think tank for the World Economic Forum. She represents HKUST on Partnership on AI to Benefit People and Society. She is a board member of Governors of the IEEE Signal Processing Society. Prof. Fung was born in Shanghai to professional artist parents but found her calling in AI when she became interested in science fiction as a child. Today, her research interest lies in building intelligent systems that can understand and empathize with humans. To achieve this goal, her specific areas of research are the use of statistical modeling and deep learning for natural language processing, spoken language systems, emotion and sentiment recognition, and other areas of AI. As a fluent speaker of seven European and Asian languages, Prof. Fung is particularly interested in multilingual speech and natural language issues.

From AI Governance to AI Safety

By Roman Yampolskiy

AI Governance in 2019 saw an explosion of interest with over 30 countries having established strategies and initiatives to date, to influence development of AI in a direction beneficial to the fulfilment of their domestic and international plans. The hope is to create standards and norms for research, deployment and international cooperation, with multi-national strategies already proposed by European Union, Nordic-Baltic region, and UN. At the same time a number of research centers are now active at the world's top universities and are explicitly devoted to questions related to the governance of AI. See Future of Life's report on Global AI Policy for the review of many national and multinational initiatives: https://futureoflife.org/ai-policy/.

AI Ethics in 2019 likewise experienced near exponential growth, at least in the number of sets of ethical "principles" proposed by over 30 organizations. Careful comparison of proposed ethical guidelines shows convergence on importance of privileging human rights, human values, professional responsibility, privacy, human control, fairness and non-discrimination, transparence, explainability and accountability. At the same time proposals differ in degree to which they place importance on each category and do not converge on common language for expressing areas of agreement. It is likely that in the future many additional organizations will propose their own ethical principles, further complicating landscape and standardization efforts. See Harvard's Berkman Klein Center report which attempts to analyze and map ethical and rights-based approaches to development of Principled AI: https://ai-hr.cyber.harvard.edu/primp-viz.html.

AI Safety also saw a lot of progress in 2019 with multiple companies and universities establishing AI Safety groups. However, it is very important to differentiate between AI Governance/Ethics and technical AI Safety and Security research. While the first two is needed to provide direction, resources, coordination and framework for performing AI research, neither one directly improves safety of intelligent systems. Only direct AI Safety research can do so and a significant danger exists in misinterpreting progress in governance and ethics as progress in safety, giving us a false sense of security. It is my hope that 2020 brings us wisdom to differentiate between governance, ethics and safety and to realize importance and limitations of each in isolation.

ABOUT THE AUTHOR — Roman V. Yampolskiy

Dr. Roman V. Yampolskiy is a Tenured Associate Professor in the department of Computer Science and Engineering 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*. During his tenure at UofL, Dr. Yampolskiy has been recognized as: Distinguished Teaching Professor, Professor of the Year, Faculty Favorite, Top 4 Faculty, Leader in Engineering Education, Top 10 of Online College Professor of the Year, and Outstanding Early Career in Education award winner among many other honors and distinctions. Yampolskiy is a Senior member of IEEE and AGI; Member of Kentucky Academy of Science, and Research Associate of GCRI. Dr. Yampolskiy's main areas of interest are AI Safety and Cybersecurity. Dr. Yampolskiy is an author of over 100 publications including multiple journal articles and books. His research has been cited by 1000+ scientists and profiled in popular magazines both American and foreign, hundreds of websites, on radio and TV. Dr. Yampolskiy has been an invited speaker at 100+ events including Swedish National Academy of Science, Supreme Court of Korea, Princeton University and many others.

PART 2 INTERDISCIPLINARY ANALYSES FROM PROFESSIONAL RESEARCHERS

The Rapid Growth in the Field of AI Governance

By Allan Dafoe & Markus Anderljung

2019 has been an eventful year in AI governance. AI companies and the AI research community have started responding to the challenges of AI governance, new AI governance research institutes have been set up, and there have been promising developments in the AI policy sphere. While there is much work left to be done, it is heartening to see how rapidly this field is growing, and exciting to be part of that growth.

Many large tech companies have started setting up and amending their processes and structures to explicitly address AI ethics and governance concerns. Some of these attempts have backfired such as Google's proposed Ethics Board shutting down after little more than a week following controversy regarding the selection of board members. Other attempts, such as Facebook's independent oversight board for content moderation have caused less controversy. Open AI's decision to conduct a staged release of their natural language model GPT-2 caused significant controversy, but also much needed discussion of publication norms. Navigating these issues forces us to answer some very difficult questions, which will only become more so as the capabilities of AI systems improve.

We have seen some encouraging developments in the AI policy sphere. The EU has shown great interest in AI policy. Its High Level Expert Group on AI delivered a set of ethics guidelines and a set of policy and investment recommendations, and the new Commission President Ursula von der Leyen pledged to initiate comprehensive legislation on AI. Policy actors who have previously been largely silent on AI governance issues have made themselves heard, for example in the release of the Beijing AI Principles and the US Department of Defense's AI principles. Though such principles are a far cry from action on AI governance issues, they provide much-needed foundation for deliberation of some of the most crucial questions of our generation.

A number of new AI governance and ethics institutes and organizations have been announced including the Schwartz Reisman Institute for Technology and Society at the University of Toronto, the Center for Security and Emerging Technology in Washington, D.C., not to mention the activity here in Oxford, such as the announcement of the Institute for AI Ethics and the establishment of the Governance of Emerging Technologies Programme at the Oxford Internet Institute. We look forward to collaborating with these new colleagues.

At the Centre for the Governance of AI, we have been busy growing our team and producing research. We now have a core team of seven researchers and a network of sixteen research affiliates and collaborators. Most importantly, we have had a productive year. We have published reports (such as our *US Public Opinion on Artificial Intelligence and Standards for AI Goverance*), op-eds (e.g. *Thinking About Risks from AI: Accidents, Misuse and Structure and Export Controls in the Age of AI*) and academic papers (*How does the offense-defense balance scale?* and five papers accepted to the AAAI/ACM conference on Artificial Intelligence, Ethics and Society).

ABOUT THE AUTHOR

Allan Dafoe

Allan Dafoe is Associate Professor in the International Politics of AI and Director of the Centre for the Governance of AI at the Future of Humanity Institute, University of Oxford. His research examines the causes of great power war and the global politics surrounding transformative technologies, in particular concerning the risks from artificial intelligence. To help scientists better study these and other topics he also works on methods for causal inference and for promoting transparency.

Markus Anderljung

Markus Anderljung is the AI Strategy Project Manager at the Centre for the Governance of AI at the Future of Humanity Institute, University of Oxford. Markus focuses on growing the Centre, making its research relevant to important stakeholders, acting as an enabler for research, and contributing to some of its research. He has a background in History and Philosophy of Science with a focus on the Philosophy of Economics and Evidence-Based Policy, several years' experience in Management Consulting and as the Executive Director of Effective Altruism: Sweden.

Towards Effective Value Alignment in AI: From "Should" to "How"

By Gillian K. Hadfield

How should we regulate AI? This is the question that has dominated the discussion of AI governance for the last several years. The question has taken the form of moral philosophical puzzles such as the trolley problem. It has been raised by activists and critics drawing attention to the dangers of discrimination and bias in algorithms and facial recognition technology. Concern about the impact of highly targeted political advertising on the stability of politics and social relationships has raised questions about whether we should regulate speech on social media platforms or constrain the collection of personal information.

At the broadest level there is widespread agreement that AI should, as the European High-Level Expert Group on AI put it in 2019, "respect all applicable laws and regulations, ethical principles and values."

But how will that alignment of AI with our human values happen? In practice, what will ensure that AI is lawful and ethical?

It will not be enough to pass laws that say AI must follow the laws. Nor is it feasible to catalogue human values and ethics and embed them into our AI systems. Our world is far too complex, dynamic, and evolving for that.

As I have explored in my work and discuss in my book, *Rules for a Flat World: Why Humans Invented Law and How to Reinvent It for a Complex Global Economy*, long before the challenge of AI, our legal and regulatory systems have faced substantial limits in putting our policy choices-our 'shoulds'-into practice. The legal and regulatory technology that we perfected over the twentieth century-legislation, regulation, regulatory agencies, courts, legal reasoning-is increasingly unable to keep up with the complexity, speed, and global nature of twenty-first century economies and societies. AI accelerates the rate at which the chasm between what we aim to do through law and regulation and what is achieved in practice widens.

While most AI governance projects in 2019 continued to focus on the 'how should we regulate AI' questions, in 2019, a major new initiative began at the University of Toronto to shift the focus to 'how can we regulate AI?'. The mission of the Schwartz Reisman Institute for Technology and Society, under my leadership, is to do the fundamental cross-disciplinary research we need to build the technical, legal, and regulatory systems that can implement our politically-determined goals for AI. We will not ask, should facial recognition be regulated, for example. We will ask, if we put rules into place, such as non-discrimination or legitimate limits to surveillance, how can we ensure that facial recognition systems follow the rules? What technical challenges do we need to solve? What innovations can we develop in regulatory technologies? How can we build AI that helps to ensure AI stays within the bounds of what we, collectively, have decided is right or acceptable? How can we make sure that our efforts at value alignment are effective?

In 2020 and beyond, the Schwartz Reisman Institute will be aiming to broaden the global conversation about AI governance beyond "should" to "how". We will be aiming to contribute to the pool of knowledge and tools available to ensure that AI is deployed where we decide it should be and not where we decide it shouldn't be and that it follows the rules humans have set when it is.

ABOUT THE AUTHOR **Gillian K. Hadfield**

Gillian Hadfield, B.A. (Hons.) Queens, J.D., M.A., Ph.D. (Economics) Stanford, is Professor of Law and Professor of Strategic Management at the University of Toronto and holds the Schwartz Reisman Chair in Technology and Society. She is the inaugural Director of the Schwartz Reisman Institute for Technology and Society. Her research is focused on innovative design for legal and dispute resolution systems in advanced and developing market economies; governance for artificial intelligence; the markets for law, lawyers, and dispute resolution; and contract law and theory. Professor Hadfield is a Faculty Affiliate at the Vector Institute for Artificial Intelligence in Toronto and at the Center for Human-Compatible AI at the University of California Berkeley and Senior Policy Advisor at OpenAI in San Francisco. Her book *Rules for a Flat World: Why Humans Invented Law and How to Reinvent It for a Complex Global Economy* was published by Oxford University Press in 2017.

China Initiative: Applying Long-Cycle, Multi-Disciplinary Social Experimental on Exploring the Social Impact of Artificial Intelligence

By SU Jun

"People-oriented" principle is the consistent aim of China to develop AI and other emerging technologies. Chinese government and academia are highly concerned about the impact of AI on human society and are striving to explore the AI social governance scheme, so as to advance the AI technologies to better serve the well-being of human beings. Encouragingly, China has taken a leading step in AI governance by conducting the social experiment to explore the social impact of AI.

As the irreplaceable driving force of S&T revolution, the opportunities and challenges brought by AI have been profoundly recognized. The consensus to keep vigilant to the threats and risks of incontrollable technology development and severe social inequity has also been well established.

In response to the challenges, we are supposed to not only advocate a responsible R&D and innovation value system, but also strengthen the focus on ethical issues in the process of scientific and technological innovation. We should especially return to "humanism" and reinforce the research on social impact mechanisms, law and trend and improve the social policy system for the development of AI from the perspective of humanities and social sciences. Achieving effective governance of AI requires systematic knowledge and accurate understanding on the social formation and characteristics of the AI era. The establishment of this recognition depends on the application of empirical research, especially the development of social experimental research.

Social experiment is a classic social science research method. It aims at observing people and organizations during the transformation of the social, political or technological environment, which simulates the ideal experimental environment to propose and testify social science theories. Facing the new problems of social governance in the era of intelligence, Chinese government, academia and varied sectors of the society have committed to formulate, promote and apply AI social experimental solutions in multiple areas including academic research, policy practice, and social impact.

In 2019, experts and scholars from Tsinghua University, Zhejiang University and other institutes brought together intellectual resources and took the lead in proposing the policy suggestions to conduct long-cycle, wide-field, multi-disciplinary AI social experiments based on abundant preliminary work.

Based on the achievements from academic research, China's policy practices are rapidly taking shape and continuously developing. In 2019, the Ministry of Science and Technology of China issued the Guidelines for the Construction of the National New-generation Artificial Intelligence Innovation Development Pilot Area, which marked that AI social experiments were being conducted nationwide. The guidelines propose different application scenarios such as education, transportation, government administration, medical care, environmental protection, manufacturing, finance, agriculture, etc., and put forward the comprehensive objectives of social experiment such as social risk prevention, organizational reinvention, data security, and technological adaptation.

Chinese society's consensus on the social governance of AI is taking shape, and the public's support for social experimental schemes is also growing. In October 2019, the First National Conference on Artificial Intelligence Social Experiments was held in Tsinghua University in China. More than 100 experts and scholars exchanged and shared the latest research results of AI social experiments, and discussed the further research plan. *Guangming Daily* and other mainstream media have published articles such as Exploring the Chinese Solution to the Social Governance of Artificial Intelligence, which has earned wide acclaim from all walks of life. The public foundation and social influence of AI social experiment are steadily on the increase.

Evaluating China's initiatives and achievements in the social governance of AI, we have become clearer that conducting AI social experiments could help us accurately identify the challenges and impacts of AI on human society, deeply understand the social characteristics and trends of AI and provide a scientific reference for the establishment of a humanistic intellectualized society.

ABOUT THE AUTHOR **SU Jun**

SU Jun is the Cheung Kong Scholar Chair Professor in School of Public Policy and Management at Tsinghua University. He serves as the Dean of Institute of Intelligence Society Governance (ISG), Tsinghua University, the Director of the Center for Science, Technology and Education Policy (CSTEP) at Tsinghua University and the Director of Think Tank Center of Tsinghua University, and the Deputy Director of the Advisory Committee of the Public Administration under the Ministry of Education. Jun Su has been awarded the special allowance from the State Council.

In addition, SU Jun is an associate at Harvard Kennedy School and senior research fellow at the Fletcher School of Law and Diplomacy, Tufts University. He is also the Chair of the First National Conference on Artificial Intelligence Social Experiment and the co-chair of Harvard-Tsinghua Workshop on Low Carbon Development and Public Policy (2014-2018).

Going Beyond AI Ethics Guidelines

By Thilo Hagendorff

In 2019, discussions on AI ethics were omnipresent. Various academic, governance as well as industry initiatives have come up with their own AI ethics guidelines. News media were swamped with articles demanding for AI ethics. Additionally, countless commissions congregated to set up norms and standards. Besides the virulent discourse on AI ethics, 2019 was also the year in which researchers and practitioners commenced to stress that abstract ethical principles are not worth much without putting them into practice. However, this is easier said than done. All over the world, ethics initiatives agree that privacy, fairness, transparency, safety, and accountability are the minimal requirements for building and using "ethical sound" AI applications. Nevertheless, what those tenets mean in day-to-day decision-making of organizations that develop and deploy such applications is rather unclear. At least empirical studies show that merely reading documents on ethical principles does not have any significant effect on practice.

The existence of ethics codes is only a tiny piece of the bigger puzzle of AI governance. If the aim is to strengthen the likelihood of ethical behavior in AI research and development, governance efforts first and foremost have to address measures for code enforcement, but also things like working climates or ethical cultures in organizations, virtue education, or the shift from competition to cooperation. Regarding the latter, the fierce competition and the related race rhetoric on "global leadership" in AI bears the risk of a reckless race for being first in accomplishing certain technical systems, especially in the context of military applications. This race is to the detriment of values like safety, privacy, or fairness. An important step towards achieving "trustworthy AI" is to attenuate competition in favor of cooperation between nations, companies, but also research institutes.

AI governance in 2020 should focus on strengthening the ties between industry stakeholders but also governance initiatives themselves. This would have the effect of saving a lot of redundancy in deliberating governance tenets and principles. Moreover, 2020 should be the year in which soft laws are increasingly translated into hard law, that gives clear rules for algorithmic non-discrimination, prohibitions for AI in high-stake areas, safety and privacy standards, as well as rules for dealing with labor displacement induced by AI applications.

ABOUT THE AUTHOR **Thilo Hagendorff**

Dr. Thilo Hagendorff is working for the "Ethics and Philosophy Lab" at the "Machine Learning: New Perspectives for Science" Excellence Cluster at the University of Tuebingen, Germany. Moreover, he works for the "AI Ethics Impact Group" of the technical-scientific association VDE (Association for Electrical, Electronic & Information Technologies). His research focusses on ethics in machine learning as well as broader questions in the field of media and technology ethics. Furthermore, he works as a research associate at the University of Tuebingen's International Center for Ethics in the Sciences and Humanities (IZEW). He is also a lecturer at the University of Potsdam's Hasso Plattner Institute.

Interdisciplinary Approach to AI Governance Research

By Petra Ahrweiler

Artificial Intelligence (AI), and especially the ethics of AI in areas of automated decision making, enjoys high priority in national policy strategies of many countries including China and Germany. International cooperation targets a joint research and governance network of a common AI-in-society ecosystem with shared ethical framing.

To improve AI algorithms for automated decision making depends to a large degree on the availability and quality of relevant training data. However, especially for high-risk decision contexts, empirical data is hardly available. Imagine automated decision making in case of an accident in a nuclear power station, a tsunami, or a terror attack in a megacity: Such events are, fortunately, too rare to produce sufficient training data. Furthermore, decision contexts involve people, who behave and interact in largely unpredictable ways according to their respective historical, cultural and social upbringing. Societal frameworks display much variety across the globe thus further restricting the utility of available training data in terms of generalizability and applicability.

Where then to get the models and the training data from to improve algorithms for better AI with a close fit to context-specific norms and values of world societies? This is where expertise of interdisciplinary research institutions such as TISSS Lab or the larger scientific community of the European Social Simulation Association ESSA comes in: for substituting missing empirical data, the innovative suggestion is to generate and exploit artificial data produced by simulations, which computationally represent the social environments AI algorithms have to operate in. In TISSS Lab, technical sciences cooperate with disciplines that are empirically researching, explaining, and anticipating human behaviour and societal developments, such as sociology, psychology, philosophy, law, and other social sciences. Realistically simulating social systems will provide sufficient high-quality training data to improve and validate AI algorithms in automated decision making. The starting international cooperation between Chinese SISS and German TISSS Lab to connect AI and social simulation can significantly further this line of cutting-edge research.

As recently emphasized by the World Artificial Intelligence Conference in Shanghai, cooperation – also transdisciplinary cooperation between science and other areas of society - is key to future progress. Perceptions, attitudes, discussions and acceptance of AI use vary between countries, as do the types and degrees of AI implementation, with reference to norms and values in-use, but also related to technology status, economic models, civil society sentiments, and legislative, executive and judicial characteristics. Building better, i.e. context-sensitive, ethically-acceptable, and socially-informed AI for future societies and realizing the international aspirations of global AI governance require the involvement of non-scientists, i.e. many relevant stakeholders and practitioners from all over the world and from all parts of society, in research. Here, the young partnership between SISS and TISSS Lab has already started to connect to participatory approaches within international funding schemes (e.g. cooperative research project AI FORA funded in the programme "Artificial Intelligence and the Society of the Future" of the German Volkswagen Foundation). Further funding schemes in this direction should be set on the policy agendas to promote progress in AI research and governance.

ABOUT THE AUTHOR **Petra Ahrweiler**

Prof. Dr. Petra Ahrweiler is Full Professor of Sociology of Technology and Innovation, Social Simulation at Johannes Gutenberg University Mainz, Germany.

Her appointment at JGU started in 2013 with getting leave for obtaining the position of Director and CEO at the EA European Academy of Technology and Innovation Assessment in Bad Neuenahr-Ahrweiler, Germany, until 2017. Before 2013, she had been Full Professor of Technology and Innovation Management at Michael Smurfit School of Business, University College Dublin, Ireland, and Director of its Innovation Research Unit IRU. Furthermore, she was Research Fellow of the Engineering Systems Division at Massachusetts Institute of Technology (MIT), Cambridge/USA.

She started her professional career with studying Social Sciences at the University of Hamburg, Germany. At Free University Berlin, Germany, she received her PhD for a study on Artificial Intelligence, and got her habilitation at the University of Bielefeld, Germany, for a study on simulation in Science and Technology Studies.

Her main interests in research and teaching are the mutual relationship of new technologies and society, inter-organisational innovation networks, and agent-based models as methodological innovation in the Social Sciences.

Petra won various research prizes, has long experience in coordinating and completing international, mostly European research projects, publishes inter-disciplinarily in international journals, and has been awarded with fellowships of various scientific societies such as the German Academy of Technical Sciences acatech or AcademiaNet, the network of excellent female scientists in Germany.

European Perspectives on the Anticipatory Governance of AI

By Robin Williams

In his 1980 book, *The Social Control of Technology*, David Collingridge reflected upon the unanticipated risks that accompanied many emerging technologies. He highlighted a dilemma confronting attempts to control the undesired impacts of technology.

'[...] attempting to control a technology is difficult, and not rarely impossible, because during its early stages, when it can be controlled, not enough can be known about its harmful social consequences to warrant controlling its development; but by the time these consequences are apparent, control has become costly and slow' (Collingridge, 1980: 19).

This insight has inspired the proposals for anticipatory governance of new and emerging science and technology, that reflect upon pathways for the development and use of technology and their potential impacts on health, the environment and social life. The UK Engineering and Physical Sciences Research Council today invites the researchers it funds to "anticipate, reflect, engage and act" to achieve Responsible Innovation.

Responsible Innovation is a process that seeks to promote creativity and opportunities for science and innovation that are socially desirable and undertaken in the public interest.

https://epsrc.ukri.org/research/framework/

These ideas are closely related to European Union proposals for Responsible Research and Innovation. How then might these apply to Artificial Intelligence (AI)?

The success of private initiatives by firms like Google and Amazon has driven enormous public and policy interest in AI and has stimulated major public research and training investments worldwide to develop AI capabilities. These have been accompanied by compelling visions of the beneficial applications of AI: autonomous vehicles; care robots; advances in medical science and diagnosis etc. These expectations – sometimes unhelpfully informed by science fiction accounts - often run far ahead of currently demonstrated capabilities. Alongside this growing concern are being articulated about potential risks – to privacy, to autonomy. Complaints have been made about the lack of transparency of algorithmic decision-making systems e.g. in finance or in public administration – and about algorithmic bias where these systems have been shown to disadvantage groups – and which may conflict with equal opportunity legislation applying women and ethnic minorities. This has inspired calls for Fair, Ethical, Transparent Machine Learning systems. Philosophers and ethicists have been enlisted into public and private AI ethics panels (with today over 40 such initiatives in Europe and North America).

However ethical principles per se will not deliver ethical outcomes. AI is not a 'thing' with determinate properties. It refers to a general purpose set of capabilities, applicable to a range of settings, and rapidly advancing through the rapid cycles of developing using and refining new tools and techniques. And the outcomes of AI are rooted not just in the design of these models but in the overall configuration of the algorithmic system. This includes the variables selected as proxies for intended outcomes, metrics and visualisations and above all in the data sets – and especially the training data for machine learning systems. And attempts to develop 'unbiased' AI systems need to confront the fact that social inequalities in society are deeply embedded in the data available – there is no 'view from nowhere'.

However, though there has been much discussion of the opacity of proprietary algorithmic systems, their operation is amenable to probing by those with moderate technical capabilities – for example submitting to recruitment algorithms job applications with different gender, age, racial identifiers. In this respect their operation and biases may be more readily made visible than traditional systems based solely on human judgement. Though it may be hard to 'open the black-box' of algorithmic system, the performance of the black box under different circumstances can be made visible.

The pathway towards Responsible Innovation of Artificial Intelligence is thus through critically scrutinising AI components, configurations, and OUTCOMES – to open up the choices made by those developing/applying them in particular contexts and make them accountable.

Responsible Innovation is thus not a one-off task but a complex bundle of activities. It will best be achieved through interdisciplinary dialogue between AI practitioner communities, stakeholders and citizen groups - what Stilgoe (2018) has characterised as "constructively engaging with the contingencies" of AI practice.

ABOUT THE AUTHOR **Robin Williams**

Robin Williams is Professor of Social Research on Technology at The University of Edinburgh, where he is Director of the Institute for the Study of Science, Technology and Innovation (ISSTI).

Since his recruitment to Edinburgh in 1986 to lead its Centre under the ESRC Programme on Information and Communications Technologies, he has developed an interdisciplinary research programme into 'the social shaping of technology' through over 50 externally funded projects. His personal research has focused upon the design and use of Enterprise Systems, eCommerce and eHealth, and more recently mobile and web 2.0 technologies. He is developing with co-authors, the Biography of Artefacts perspective to address the design and implementation of information infrastructures.

Recent books include *Social Learning in Technological Innovation: Experimenting with Information and Communication Technologies*, (Edward Elgar: 2005) with James Stewart and Roger Slack and *Software and Organisations: The Biography of the Enterprise-Wide System - Or how SAP Conquered the World* (Routledge: 2009) with Neil Pollock and *How Industry Analysts Shape the Digital Future* (Oxford University Press: 2016) with Neil Pollock.

The Impact of Journalism

By Colin Allen

The most important progress related to AI governance during the year 2019 has been the result of increased attention by journalists to the issues surrounding AI. They have brought attention to problems ranging from "algorithmic bias" to the risks to human freedom and democratic ideals that arise from AI-assisted large-scale surveillance by governments and corporations. However, effective governance of AI requires accurate understanding of the technology and its applications. Journalists, business leaders, politicians, and the general public all struggle to understand the technical aspects of AI. The lack of understanding contributes both to excessive optimism and to excessive pessimism about AI, as well as to leading to poorly calibrated levels of trust and mistrust of AI among the people who use it. Miscalibrated trust includes having too much trust in AI when the technology doesn't warrant it (for example, people trusting their self-driving capacities of their cars too much) as well as having too little trust in AI in situations where it perhaps could do a better job than a human.

The promotion of good technical understanding is an important missing component in most journalistic coverage. For example, the widely-reported idea of "algorithmic bias" is potentially misleading because it fails to distinguish biases in the data on which algorithms operate from biases in programmers leading them to design algorithms which ignore relevant information or put too much weight on some factors. Sensible policies for AI governance depend not just on balancing the risks and opportunities provided by AI, but on the understanding the very significant role that humans continue to have in the design and implementation of AI applications, and in their use. Journalistic coverage is important because it has shifted the debate about AI to the important issues of governance, but the process of attaining wisdom in human use of AI has only just begun. Academics, journalists, and software engineers all need to address the question of how to develop wise use policies in a safe way, free from the risks entailed by the nearly unlimited public experimentation that is currently practiced by governments and industry.

ABOUT THE AUTHOR **Colin Allen**

Colin Allen is Distinguished Professor in the department of History & Philosophy of Science at the University of Pittsburgh. From 2015-2019, he held the title of "Chair Professor" at Xi'an Jiaotong University, Xi'an, China, and in 2017 he was appointed Changjiang Scholar by the Ministry of Education in the People's Republic of China.

Allen's research concerns the philosophical foundations of cognitive science. He is particularly interested in the scientific study of cognition in nonhuman animals and computers, and he has published widely on topics in the philosophy of mind, philosophy of biology, and artificial intelligence. He has over 100 research articles and several edited and co-authored books, including *Moral Machines: Teaching Robots Right from Wrong* (Oxford University Press 2009) which has been translated into Korean, Chinese, and Japanese.

Since 1998 Allen has been consulting and programming for The Stanford Encyclopedia of Philosophy and is its Associate Editor. He is director of the Internet Philosophy Ontology project (InPhO) which has received multiple grants for its work in computational humanities. From 2020-2022 he is the recipient of an award from the Templeton World Charity Foundation for a project titled "Wisdom in the Machine Age".

Future of Work in Singapore: Staying on Task

By Poon King Wang

In 2019, the Lee Kuan Yew Centre for Innovative Cities (LKYCIC) at the Singapore University of Technology and Design (SUTD) made two research contributions to show how society can use tasks as building blocks to design human-centric jobs and to uplift lives in the future of work.

The first contribution was a collaboration that was recognized by Singapore's National AI Strategy as contributing to building a Trusted and Progressive Environment for AI in Singapore's Smart Nation journey. Working with France-Singapore think tank Live with AI, AI consultancy Data Robot, and several companies, we used tasks to first track the speed and scale of disruption of AI on jobs. We then incorporated the ethical, social and human considerations, and created one-page step-by-step task-by-task transformation road maps to future jobs that people would find valuable.

Our second contribution was a partnership with the labor unions. We worked with them to identify several jobs that are at high risk of AI displacement. We then used AI to chart clear and concrete task-by-task transition pathways to new jobs for the workers who might be displaced, including pathways to jobs within and outside of the workers' professions and sectors. This combination of clear pathways and expanded choices means workers can be empowered with greater confidence and certainty, and the partnership was cited by the Deputy Prime Minister in an International Labour Organization conference.

These two contributions build on the LKYCIC's future of work research where we have made tasks central for three reasons. First, as long as AI remains narrow, its impact on jobs will be task-by-task, and not job-by-job. Second, there is growing consensus amongst experts that tasks provide the right level of resolution to study the future of work. Third, tasks are increasingly used to explain trends at different scales -- from the impact of specific AI innovations on specific skills, to the macro-economic changes in the labor market in the last few decades.

Our research advances the use of tasks by developing task databases and strategies to help governments, companies, and individuals (such as the abovementioned two contributions). They all take advantage of the fact that any job can be broken down into its constituent tasks, and by assessing which and when tasks will be disrupted, we can track AI disruption risk and transformation potential. At the same time, each job will have tasks that are similar to tasks in other jobs – these can be used to identify new tasks, jobs, and pathways.

In every past Industrial Revolution, even when more jobs were created than destroyed, there were always segments of society who struggled or suffered. In our current Revolution, we are already seeing such signs worldwide.

We have to help more people thrive. Tasks provide the building blocks, databases, and strategies for the public, private, and people sectors to do so clearly, concretely, and confidently.

Together, we can uplift lives if we stay on task.

ABOUT THE AUTHOR — Poon King Wang

Poon King Wang is the Director of the Lee Kuan Yew Centre for Innovative Cities at the Singapore University of Technology and Design (SUTD), where he also heads the Smart Cities Lab and the Future Digital Economies and Digital Societies initiative. He is concurrently Senior Director of Strategic Planning at SUTD.

King Wang is on the World Economic Forum's Expert Network on Cities and Urbanization, and the Board of Live with AI (an independent France-Singapore think tank on Artificial Intelligence). His and his teams' multi-disciplinary research focus on the human dimensions of smart cities and digital economies, and the impact of digital transformation on the future of work, education, and healthcare, and on society at large. He pays particular attention to how leaders of cities and companies can design strategies and policies to lift the lives of their citizens and workers, with the same technologies that are disrupting work, economy and society.

King Wang holds a MSc (Industrial Engineering and Engineering Management) from Stanford University, a BSc (Electrical Engineering) from the University of Illinois at Urbana-Champaign, and a Rocket Engineering Certificate from Moscow State Technical University.

Developing AI at the Service of Humanity

By Ferran Jarabo Carbonell

The short space of this article only allows to enunciate some of the topics. Ethics is making a great contribution to the reflection on Artificial Intelligence. This contribution supposs an aid to the development of this science. In the first place, it offers a walker for the harmonic growth at the service of humanity, and, in the second place, it forces it to keep in mind that the aim is to offer some help to human beings and their safeguard.

Ethical reflection on artificial intelligence must start from a profound conception of what to be a person means. It is not simply a question of referring to the 'Charter of Human Rights'. AI is at the service of men and the human being is an ethical subject by nature. That is, every man needs to know he is doing good things for his personal development. Good is neither a mere feeling, nor a coercion of freedom. We must understand that "good" is everything that is good for oneself and for all human beings. This is not relative, there is consensus (one is the Universal Declaration of Human Rights) and more must be sought so that the science of we speak of is at our service. The human being must not do everything that can be done; insurmountable limits must be established for the good of all.

Below, I list only three fundamental points on which researchers and thinkers should converge. The list could be much longer, but hopefully these three points will serve to initiate reflection:

1.The inherent value of every human being. I am not only talking about the non-discrimination on the basis of race and sex; the human being, with independence of anything else, must be safeguarded and loved. It has already happened many times before: supposedly intelligent algorithms have discriminated people because of their race or sex. This is totally inadmissible in a plural and equal society such as ours. From here we draw a limit: artificial intelligence must always be at the service of the person and not the other way around.

2.Artificial intelligence can never be autonomous. The human being is the ultimate responsible for all his actions. No action coming from artificial intelligence can be detached from its maker. There is an inescapable responsibility of the one who creates the algorithm which the machine works with. Therefore, Artificial Intelligence must always have human control. To be more specific: a) everything that refers to autonomous lethal weapons (LAWS) must be banned for the sake of subsistence. The control of such weapons must never escape human control. b) other systems that can become autonomous (driving, BOTS...) must always depend on human decision. They cannot be left to their own free will.

3.It must be at the service of humanity as a whole without excluding the poor. This point is of utmost importance. It is inconceivable that countries and people with no economic power are excluded from any advance that is made for the good of all. We must find ways to make technological advances for all. There can be no discrimination on any grounds, let alone economic ones.

And to finish: the control of Artificial Intelligence must always be human, as well as its responsibility. Another obvious thing is that the moral decision cannot be made a posteriori, it must always be made a priori. That is, moral laws must be respected and used before making an algorithm and ethics must be observed before any digitization. This is for the sake of the dignity of human nature and in defense of its privacy. Algorithms must be analyzed before being executed.

ABOUT THE AUTHOR **Ferran Jarabo Carbonell**

Prof. Dr. Ferran Jarabo Carbonell, born in Alicante on February 17, 1967. He lives all his life in Girona where he begins his studies.

Degree in Philosophy, Philosophy and Letters and Dogmatic Theology from the Pontifical University of Salamanca. The year 1997 is ordained diocesan priest in Girona. In 2006 she received a PhD in Philosophy from the same pontifical university.

Professor of Philosophical Anthropology and Phenomenology of Religions at the Institute of Religious Sciences of Girona in different periods for almost 16 years. Professor at the Redemptoris Mater seminar in Berlin for four years in various philosophical subjects: Ethics, Philosophical Anthropology, Cosmology, Ontology.

He has participated with different communications in international SITAE Days. Collaborate in various publications with popular articles. He currently collaborates at the University of Mainz with the AI FORA project as a representative of the University of Girona and works pastorally for the diocese of Limburg.

Enhance Global Cooperation in AI Governance on the Basis of Further Cultural Consensus

By WANG Xiaohong

In 2019, substantial progress has been made in AI governance from principle to practice; transdisciplinary cooperation between engineers and humanities scholars has converged on the "human-oriented" approach; all sectors of society including major international organizations, more and more national governments, ICT leading enterprises, academia, media, education circles have made concerted efforts to build a wideranging network of AI governance. But from the perspective of cultural comparison, there is a potential worry about the AI governance environment in 2019 and beyond. The increasingly intensified competition among countries and interregional conflicts make the cooperation and sharing of the frontier technology of AI governance full of uncertainty. The root is the increasingly prominent differences in cultural values among countries and nations, and the danger of being torn from cultural unity faced by the human community. Confronting severe challenges in global governance, AI governance needs to conduct more practical cultural accommodation and further promote value consensus.

The cultural value plays an implicit role for the technical and explicit measures. In recent years, engineers and ethicists have been cooperating to explore and solve specific problems, clarifying ethics as the practical value of AI design framework, and making the process of AI governance increasingly clear. Taking deep neural networks as an example, from the definition of tasks, data collection until designing, training, testing, evaluation and application debugging of models, governance principles (security, transparency, privacy, fairness, etc.) can be added in every link, and the improvement of technical means will approach ethical expectations. However, the abstract principle of "human-centric" may lead to differences in practical value due to cultural differences in the actual situation of AI governance, or even the countermeasures of AI governance. An ethical consensus of AI governance needs to take root in the major issues of the common destiny of mankind and the eternal values accumulated through cultural heritage.

The wisdom of "harmony but difference" (Analects) in Chinese culture means cultural diversity. Future AMAs (artificial moral agents with high autonomy and high sensitivity to values) will choose to cooperate with human beings rather than exterminate human beings. Any intelligent agent needs more freedom, and the greater the diversity, the greater the informational entropy, and the greater the freedom of choice for each individual. The study of information ethics and machine morality has repeatedly revealed that the integration of Chinese and Western cultures is the source of moral insight. "Do as you would be done by" and " I want to stand firm, but also want to let others stand firm, I want to develop, but also want to let others develop" in Analects are consistent with Kant's categorical imperative: only when you are willing to act on this criterion can you make this criterion a norm. In addition, "self-restraining in privacy" (Doctrine of Mean), and self-cultivation practice inherited and developed by the Neo-Confucians, together with the virtue ethics advocated by Aristotle, reflect the common wisdom of the ancient Eastern and Western cultures.

Human beings need the wisdom of cultural integration to realize the moral principles of AI. Human beings must act in concert and in a coordinated way, or any barrel effect will bring all efforts to naught. In 2020, AI governance can focus on the core of AI ethics and strengthen substantive measures to enhance the value consensus among different countries and regions.

ABOUT THE AUTHOR **WANG Xiaohong**

Wang Xiaohong received her Ph.D. in Philosophy of Science and Technology from Peking University in 2004. She is a Fulbright Visiting Research Scholar (IU, 2006-2007). Presently, she works in department of philosophy at XJTU as the co-director and Professor of Research Center for Computational Philosophy. She also serves as a member of the Big Data & AI Working Group of World Federation of Engineering Organizations (WFEO) (since 2019), and an executive committee of China Scientific Methodology Commission (since 2011).

Professor Wang's research concerns the philosophy of cognitive science. She is particularly interested in philosophy of AI machine discovery, computational analysis of Chinese philosophy, and interested in information ethics, and integration of science and humanities.

Three Modes of AI Governance

By YANG Qingfeng

An article on AI governance has caught my attention. This article pointed out that AI governance is 'an unorganized area' (James Butcher et al. 2019). James Butcher (2019) has provided an overview of the practice of different stakeholders in the AI governance activities. According to this article, the key point is to maximize the benefits and minimize the risks. Public sectors and non-public sectors have different responsibilities in AI governance.

AI governance is certainly a new field waiting for exploration. The reason for this is on the controversy over the understandings of what AI is and what AI governance is. Therefore, the primary issue is to clarify the definitions of AI and AI governance. I distinguish three modes of governance based on the AI definition., namely, governance based on governmental bodies, governance based on technologies, and governance based on humanistic values.

The first AI governance is based on governmental bodies. In this view AI is considered as a tool related to different bodies. AI is used by different bodies such as governments, companies, individual, etc. The safety and reliability is the key to good use or rational use. However, problems from rational use will be ignored in this view.

The second AI governance is based on human values. AI is seen as embodiment of human values. AI needs to follow human values such as responsibility, safety, fairness and trust. AI governance is focused on the designing process and how to guard or embed human values into agents. The ethical framework and ethical decision-makers have been emphasized. By Glass-Box, we can 'implement transparent moral bounds for AI behavior' (Andrea Aler Tubella et al. 2019).

The third AI governance is based on the technologies. AI in the view is regarded as technologies or technological system. The view is useful to cover philosophical problems, technological problems and some problems entangled between AI and society. In this view, AI governance focuses on how to tackle such problems as the societal and humanistic impact of AI. The partnership on AI (PAI) 2019 has discussed the influence of AI on people and society, especially algorithmic biases and errors in AI.

Logically, AI governance has experienced a transition from 'use context' to 'prediction context'. Most researches have focused on entities that use and design AI. Rational use or responsible use is the inevitable path. However, AI has strong autonomy and ability to learn. Algorithm has been used to predict human behavior in the future. The basic problem is to tackle with relationship between AI and human being. Coexistence is a good relation model (Beena Ammanath, 2019). Some technological problems such as AI algorithmic bias are more important. Many media have concerned AI bias from algorithms. Many governments and organizations are increasingly concerned about AI bias. Explainable and unbiased algorithms are possible direction. How do we use AI tools to give us a predictive representation of the status of major social practice and predict its development is a question needing to consider? Maybe BlueDot is a good case. It has sent us many real-time infectious disease alerts.

ABOUT THE AUTHOR **YANG Qingfeng**

Yang Qingfeng (1974) received his Ph. D. from Fudan University in 2003. Currently, he is a professor at Center for Applied Ethics and Fudan Development Institute of Fudan University. He also serves as the Executive Director of the Technology Philosophy Committee of the Chinese Society for Philosophy of Nature and the Secretary General of Shanghai Nature of Dialectics Association in China. He is visiting Scholar of Dartmouth College, USA and Swinburne University of Technology, Australia. His current research includes the philosophy of technology, data ethics, philosophy of memory and AI ethics.

PART 3 RESPONSIBLE LEADERSHIP FROM THE INDUSTRY

Companies Need to Take More Responsibilities in Advancing AI Governance

By YIN Qi

There is a consensus that AI governance should be a global priority. In terms of policy making, many countries have successively announced AI strategies and singled out the importance of AI governance. In 2019, China's Ministry of Science and Technology high-lighted the critical nature of this work by announcing the establishment of its National New Generation AI Governance Expert Committee. In terms of media scrutiny, more and more attention has been paid to issues such as the ethical boundaries and technical interpretability of AI and data privacy protection, which are all essentially AI governance issues.

AI governance is not only the responsibility of the government and relevant institutions. Enterprises, as the main force in the R&D and application of AI and the front-line practitioners of AI technologies, should fulfill their responsibilities and take the initiative to achieve enterprise autonomy. Today, many international and Chinese companies, including MEGVII, have launched their own AI Ethics Principles and criteria, elaborating on their initiatives to ensure responsible governance of AI technology.

For companies, effective implementation of AI governance measures is a major area of focus. Let me summarize my thinking based on MEGVII's own firsthand experience:

1. First, we need to maintain a rational focus on and continue to engage in constructive discussions on AI governance. In January of this year, we invited experts across the fields of law, ethics and AI technology, as well as the general public, to join candid and constructive online discussions on the 10 mostly heavily-debated AI ethics issues. We received thousands of comments across social media platforms, and top concerns include privacy, information security and sufficient protection of user rights.

2. Second, we recognize the importance of conducting in-depth research on key issues. Data security and privacy protection are top priorities, for both the public and the enterprises. Megvii has a research partnership with the Beijing Academy of Artificial Intelligence that will focus on these issues. We are working to implement an AI platform to best manage the collection, transmission, storage and usage of data for the full life-cycle protection of data and establish a set of relevant AI data security and privacy protection mechanisms. Megvii was also tasked by the Ministry of Science and Technology to build a National Open Innovation Platform for Next Generation Artificial Intelligence on Image Sensing, where industry-wide research results and practical experience of enterprises will be shared to promote the healthy and rapid development of the AI industry.

3. Third, we need sustained action. A robust and effective organizational framework is required to oversee, implement, and foster collaboration on our AI ethics principles. This is why Megvii has set up an AI Ethics Committee under its Board of Directors, consisting of founders, core executives and external experts, to oversee the implementation of Megvii's AI Ethics Principles. The Committee is supported in its work of coordination and in-depth research by a secretariat and an AI Governance Research Institute.

Although in 2019, we saw some difficult questions arise in AI governance around the world, we hope and expect that 2020 will become the "Year of AI Governance." AI governance is effective solution for maintaining controls in the new era of AI. AI governance must become part of everything we do as an industry, and these types of preventative and protective measures need to be more widely recognized and practiced through a combination of learning and practice. I want to take this opportunity to call on everyone to take a long-term view and face the challenges of AI governance head on. I hope that together we can power humanity with AI.

ABOUT THE AUTHOR **YIN Qi**

Yin Qi (who also goes by "Inch"), is co-founder and CEO of Megvii Technology Limited, a world-class AI company with core competencies in deep learning. He chairs the company's board-level AI Ethics Committee, which is committed to positively contributing to the society with Megvii's AI technology. Yin is a member of the National New Generation Artificial Intelligence Governance Expert Committee, an expert committee established by China's Ministry of Science and Technology engaged in research on AI-related laws, ethics, standards and social issues and international exchanges and cooperation on AI-related governance.

Yin was a member of the 2019 Young Global Leaders of the World Economic Forum. He was named to Fortune's "40 under 40" list of Chinese elites for three consecutive years, and was ranked No. 1 on Forbes Asia's "30 under 30" Enterprise Technology entrepreneurs. MIT Technology Review has also included him in their global "Innovators under 35" list.

Trustworthy AI and Corporate Governance

By Don Wright

Organizations must create ethical systems and practices for the use of AI if they are to gain people's trust. This is not just a compliance issue, but one that can create a significant benefit in terms of loyalty, endorsement, and engagement.

- Capgemini

The proliferation of A/IS (autonomous and intelligent systems) presents a profoundly human moment. Collectively, we are standing in the nexus of history.

While it's always been essential to know your customer and their needs, the specific nuances of AI, where interacting with people demands a higher level of awareness around things like bias, identity, emotion, and cultural relevance, make obtaining and using this knowledge of the customer even more difficult. It also means recognizing that, outside of anyone's positive intentions for what they create, an end-user's experience is not fully up to the designer — it is up to each end-user. This is why IEEE created Ethically Aligned Design, 1st Edition and why it focused on end-users and how they and their values can be a part of AI design.

According to McKinsey Global Institute, "AI has the potential to deliver…global economic activity of around $13 trillion by the year 2030." While the monetary benefits of AI have increased in recent years, so have the concerns around its ethical implementation for people and society as a whole. Beyond the need to combat negative unintended consequences in the design of AI, the analysis, utilization, and honoring of end-user values in design is providing a growing trend of driving innovation in corporate governance.

As a way to highlight this trend, IEEE recently created the Ethically Aligned Design for Business Committee as part of its Global Initiative on Ethics of Autonomous and Intelligent Systems. Comprised of participants from Google, IBM, Intel, Salesforce, Microsoft, and others, the committee launched its first paper in Q1 of 2020 called A Call to Action for Businesses Using AI featuring:

• The Value and Necessity of AI Ethics;

• Creating a Sustainable Culture of AI Ethics; and,

• AI Ethics Skills and Hiring.

While created with corporations in mind, much of its contents will also provide useful guidance for certain governments and NGOs. The paper also features an "AI Ethics Readiness Framework" allowing readers to assess where their organization, public or private, lies on a four-tiered scale highlighting issues such as training, leadership buy-in, organizational impact, and key performance indicators (KPIs) beyond financial metrics alone.

Corporate governance for AI cannot rely on simply adhering to basic compliance criteria regarding mandated legislation like the GDPR. Organizations need to proactively create and prioritize transparent and accountable practices that honor end-user values to establish genuine trust with their employees, customers, and all stakeholders throughout their value chain.

"We want to design healthy relationships with our users. The potential of AI is wrapped up in its longevity as a solution-meaning everything we design must address current and future needs for users. To truly understand those needs, we need an inclusive and ethical approach to the entire process. Globally, we are starting to see the repercussions that come when companies do not prioritize AI ethics in their solutions. We want to make sure that ethical practices are ingrained on our teams so they can then be embedded into the products themselves."

– EAD for Business Committee Member Milena Pribec of IBM

ABOUT THE AUTHOR **Don Wright**

Mr. Don Wright is the President of Standards Strategies, LLC, an ICT Standardization consulting firm. He is the retired Director of Worldwide Standards for Lexmark International and previously IBM and has over 40 years of experience in standards, engineering, software development and marketing. Mr. Wright is a Senior Member of the IEEE and served as President of the IEEE Standards Association (2017-2018), and a member of the IEEE Board of Directors (2017-2018). He previously served as Computer Society VP of Standards, IEEE-SA Standards Board Chair, IEEE-SA Treasurer, IEEE-SA Awards and Recognition Chair, IEEE Admission and Advancement Chair, and on the IEEE Awards Board. He is a member of the Computer Society, Communications Society, Consumer Electronics Society, Society on the Social Implications of Technology, and Technology and Engineering Management Society. He is a member of the Board of Directors of the IEEE-ISTO and previously served as Chairman. He previously served as Chair of the INCITS Executive Board, US HoD to ISO/IEC JTC 1, and two terms as a member of the Board of Directors of ANSI. He graduated from the University of Louisville with BSEE and MEng EE degrees. He is a member of Tau Beta Pi and Eta Kappa Nu.

A Year of Action on Responsible Publication

By Miles Brundage, Jack Clark, Irene Solaiman and Gretchen Krueger

Deepfakes. GPT-2 and issues of synthetic text. Gender-guessing systems. These were some of the things that the AI community reckoned with in 2019, as ethical considerations relating to the publication of AI research came to the fore.

This growing attention to publication norms in the AI community was the result of two factors.

First, a subset of AI systems known as generative models--which can be used to generate samples that look similar to real data--improved in performance and flexibility, sparking concerns about such systems being used to deceive people online with synthetically generated content such as images, audio, and text. (In 2019 it was revealed that realistic-looking but AI-generated images were used as part of an online influence campaign by Epoch Media Group, and researchers explored the potential misuse of language models for generating deceptive or abusive text.)

Second, evidence continued to mount that existing publication practices in the AI community are insufficient to address such risks, and that experimentation with new technical and policy approaches is needed. Continued publishing of deepfakes research, for example, is making it easier and easier to produce misleading videos of people saying or doing things that never occurred, while detection efforts are in their early stages. These trends have raised deep concerns not only about the direct deception of people with AI-generated media, but also the risk of people not believing authentic media because it could have been generated by AI.

One high-profile case of evolving publication norms involved our organization, OpenAI. In February 2019, OpenAI announced its GPT-2 language model, which displayed state of the art performance in various language modeling tasks (predicting what comes next in a text sequence) and surprising performance on other tasks like text summarization, question-answering, and translation. At the same time, we shared our concern that GPT-2 could be used to generate abusive or misleading text. We then took the unusual step of releasing increasingly powerful versions of the model in stages, rather than all at once (a process we call Staged Release), and explored new ways to get expert input on the ease of misusing the system throughout the process. As a result, we were able to work with experts at other research organizations to incrementally improve and share our understanding of GPT-2's characteristics at each stage in the release process.

While our decisions on GPT-2 sparked significant debate, OpenAI was not alone in calling attention to these misuse concerns. Blog posts and papers by other organizations such as Salesforce, Google, Hugging Face, the Allen Institute for AI, and the University of Washington highlighted different societal implications and challenges of large-scale language models. In our view, there is still much to learn about how to responsibly publish language models, as well as AI systems more generally.

Beyond improving documentation of AI systems and the release process associated with them, there was also significant attention paid in 2019 to preparing for instances of misuse through detection and policy changes. Google released a dataset to aid in detecting synthetic voices, while Facebook, the Partnership on AI, and other organizations launched competitions for "deep fake" video detection. Legislators in various countries, and online platforms such as Twitter, also began to formulate policies aimed at addressing related risks.

As technical progress continues and the impacts of AI in the real world become clearer, we expect the AI community to continue grappling with these issues in 2020. We are excited to see how norms evolve in the year ahead as researchers' experiment with new ways of maximizing the benefits of publishing powerful AI systems while minimizing the risks. Because progress in AI can move unusually quickly, we need to be prepared for surprising challenges to arise.

ABOUT THE AUTHOR

Miles Brundage

Miles Brundage is a Research Scientist on OpenAI's Policy team, where he researches issues related to coordination among AI developers and responsible publication of misusable models. He is also a Research Affiliate at the University of Oxford's Future of Humanity Institute, where he previously worked for two years as a Research Fellow. He earned his PhD in Human and Social Dimensions of Science and Technology in 2019 from Arizona State University.

Jack Clark

Jack Clark is the Policy Director for OpenAI, where he leads OpenAI's policy outreach efforts. Jack researches the measurement and analysis of AI systems. He sits on the steering committee of the AI Index, part of the Stanford 100 Year Study on AI project. He is also an external research fellow at the Center of Security and Emerging Technology in Washington DC. Jack has testified in Congress three times and was a technical expert for the OECD's AI Principles initiative in 2019.

Irene Solaiman

Irene Solaiman is a policy researcher at OpenAI. She conducts social impact and fairness analysis and policymaker engagement as part of the Policy Team. She was a fellow at Harvard's Berkman Klein Center as part of the Assembly Student Fellowship (formerly known as Techtopia) researching the ethics and governance of AI. Irene holds a Master in Public Policy from the Harvard Kennedy School and a self-designed B.A. in International Relations from the University of Maryland.

Gretchen Krueger

Gretchen is the project manager for the Policy Team at OpenAI, and works on projects related to responsible publication, coordination, and scenario planning. Prior to joining OpenAI, Gretchen worked at the AI Now Institute at New York University, and at the New York City Economic Development Corporation. Gretchen holds an MS from Columbia University and an AB from Harvard University.

AI Research with the Potential for Malicious Use: Publication Norms and Governance Considerations

By Seán Ó hÉigeartaigh

My heart, why come you here alone?

The wild thing of my heart is grown

To be a thing,

Fairy, and wild, and fair, and whole

GPT-2, 2019[1]

On Valentine's Day 2019, technology company OpenAI announced a language generation model of unprecedented performance.[2] However, as an "experiment in responsible disclosure" it only released a limited version of the language model. In doing so OpenAI brought attention to a governance debate that has since gained a great deal of momentum. OpenAI's decision was due to its researchers' concerns that their technology could have potentially malicious applications. While the technology would have many positive uses, such as in language translation and digital assistants, they reasoned that effective and freely available language generation could also have more harmful impacts. These might include automating fake news generation, helping fraudsters impersonate others online, or automating phishing for cyberattacks.

These concerns related to broader issues around the potential malicious use of synthetic media generation, where machine learning advances are playing a key role. But they also highlighted pressing questions about the responsibilities of AI research groups and companies with regard to malicious uses of their technologies. This discussion is not unique to AI; it has been debated extensively in other technology and security contexts, often under the heading of 'dual use' research. One high-profile instance was a debate in 2011-12 over whether it was appropriate to publish risky influenza research.[3] Due to recent advances in machine learning technologies, the increasingly varied contexts in which they are being deployed, and the more widespread availability of powerful techniques, a growing number of researchers, civil society groups, and governments are now giving attention to concerns over malicious uses of AI.[4, 5]

OpenAI's move to restrict their technology resulted in vigorous debate. Critics argued that the decision not to release was sensationalist and raised undue fears,[6] and that the decision not to release to academics endangered norms of open publication and research-sharing.[7] Others argued that caution was justified,[8] and that delaying publication allowed time to prepare against malicious uses.[9]

A growing interdisciplinary research community is exploring these issues, including at forums such as the Partnership on AI.[10] OpenAI's researchers have written an analysis of what they themselves had learned from their experiment in responsible publication norms,[11] and finally released the full, most high-performance version of their model in November 2019. Many open questions remain about what should constitute research of concern in AI, and what the ideal process should be when advances with the potential for misuse are made.[12] However, one thing is certain: now is an excellent time for this debate. AI technologies will continue to become more powerful, and more widespread in their uses in society. Developments made with the best of intentions will be put to malicious purposes. Now is the time for the AI research and governance communities to explore these questions with a broad set of stakeholders, and to develop appropriate norms, safeguards and best practices for the dual-use AI technologies of tomorrow.

[1]Gwern.net (2019). GPT-2 Neural Network Poetry

[2]OpenAI Blog (2019). Better Language Models and Their Implications

[3]Butler & Ledford (2012). US biosecurity board revises stance on mutant-flu studies

[4]Brundage & Avin (2018). The Malicious Use of Artificial Intelligence

[5]House of Lords (2019). AI in the UK: ready, willing and able?

[6]Lipton, Z. Approximately Correct (2019). OpenAI Trains Language Model, Mass Hysteria Ensues

[7]Li & O'Brien. Electronic Frontiers Foundation (2019). OpenAI's Recent Announcement: What Went Wrong, and How It Could Be Better

[8]Metz & Blumenthal. New York Times (2019). How A.I. Could Be Weaponized to Spread Disinformation

[9]Howard, J. Fast.AI (2019). Some thoughts on zero-day threats in AI, and OpenAI's GPT-2

[10]Leibowitz, Adler & Eckersley. Partnership on AI (2019). When Is It Appropriate to Publish High-Stakes AI Research?

[11]OpenAI blog (2019). GPT-2: 6-Month Follow-Up

[12]Crootof, R. Lawfare (2019). Artificial Intelligence Research Needs Responsible Publication Norms

ABOUT THE AUTHOR **Seán Ó hÉigeartaigh**

Seán Ó hÉigeartaigh is the Director of the AI: Futures and Responsibility programme (AI: FAR) at the Leverhulme Centre for the Future of Intelligence (CFI), an interdisciplinary centre that explores the opportunities and challenges of artificial intelligence. The AI: FAR programme focuses on foresight, security and governance related to artificial intelligence.

He is also the Co-Director of Cambridge's Centre for the Study of Existential Risk (CSER), a research centre focused on emerging global risks and long-term challenges.

Seán's research spans the impacts of artificial intelligence and other emerging technologies, horizon-scanning and foresight, and global risk. He led research programmes on these topics at the Future of Humanity Institute (Oxford) from 2011-2015, was founding Executive Director of the Centre for the Study of Existential Risk from 2014-2019, and co-developed both the Strategic AI Research Centre, and the Leverhulme Centre for the Future of Intelligence. His paper An AI Race: Rhetoric and Risks (with Stephen Cave) recently won joint best paper at the inaugural AI Ethics and Society Conference. He has a PhD in genome evolution from Trinity College Dublin.

GPT-2 Kickstarted the Conversation about Publication Norms in the AI Research Community

By Helen Toner

For me, the most attention-grabbing AI governance discussion of 2019 concerned responsible publication norms, and it was sparked by OpenAI's decision to delay the release of GPT-2, a language model trained to predict the next word in a text.

First announced in a blog post and paper in February, GPT-2 (a successor to GPT, or "Generative Pre-Training") showed a remarkable ability to generate multiple paragraphs of fairly coherent writing in a wide range of styles. But what drew even more attention than GPT-2's performance on language generation was OpenAI's announcement that it would not be publishing the full model. The reasoning: it might be used "to generate deceptive, biased, or abusive language at scale," and OpenAI wanted to take this occasion to prompt discussion in the machine learning (ML) community about responsible publication norms.

The post certainly succeeded at prompting discussion. Initial reactions were mixed, with many ML researchers criticizing what was perceived as a deliberate effort to create hype and attract media attention. Many also felt that OpenAI's strategy was damaging to academic norms of openness, making it harder to replicate and verify their work. By contrast, reactions in AI policy and governance circles were largely positive, expressing appreciation for the effort to begin developing norms around publication of research that could be used in harmful ways, even if this particular work was not especially risky.

Over the course of 2019, OpenAI continued to post about GPT-2, providing updates on their conversations with other groups and their plans going forward. In a May update, OpenAI announced that it would be releasing the model in stages—publishing a "medium" version (following the "small" version with the original post), which was succeeded by a "large" version in August and an "extra-large" version in November.

During this period, multiple researchers attempted to replicate OpenAI's work, and several succeeded in whole or in part. In one particularly interesting case, an independent researcher named Conor Leahy announced on Twitter that he had replicated the model and intended to release it publicly, in deliberate defiance of OpenAI's release strategy. After discussions with OpenAI and other researchers, however, he changed his mind, and decided to keep his work private.

Of course, 2019 was not the year in which the ML community agreed on firm norms around responsible publishing—these questions are complex, and will require further experimentation and debate. But against a backdrop of increasingly convincing deepfake videos, ML research being turned to authoritarian purposes, and other concerning trends, the discussion kickstarted by OpenAI stands out to me as a step in the right direction.

ABOUT THE AUTHOR **Helen Toner**

Helen Toner is Director of Strategy at Georgetown University's Center for Security and Emerging Technology (CSET). She previously worked as a Senior Research Analyst at the Open Philanthropy Project, where she advised policymakers and grantmakers on AI policy and strategy. Between working at Open Philanthropy and joining CSET, Helen lived in Beijing for nine months, studying the Chinese AI ecosystem as a Research Affiliate of Oxford University's Center for the Governance of AI.

The Challenges for Industry Adoption of AI Ethics

By Millie Liu

Artificial Intelligence technology continues its fast development in 2019. Yet despite the promising adoption, there are real-world challenges with the implementations and ethical concerns from the industry. While academia tends to see things from a theoretical perspective, the below observations are made from a more practical point of view from the frontline. These challenges and concerns, in particular, deserve policymakers' attention. The industry can benefit or be hindered by policymaking, which is an undertaking that requires an appreciation of practical nuances.

Challenges with implementation:

-Infrastructure & data automation: modern applications are better built on modern infrastructures. While many companies are moving to microservices in the cloud, a large number still remains on-premise. Existing legacy architecture and the inertia of pulling data across many, many ERPs still lead to bottlenecks.

-Explainable AI & model deployment ownership: Who is responsible for the models deployed in the real world that are also constantly learning and evolving? How do companies protect their customers and their own reputation from the AI model bias and the black box when it's making real-world decisions every day? A common platform for collaboration, deployment and continuous monitoring becomes a pain for companies investing in AI/ML.

Challenges with AI ethics:

-Discrimination: the AI explainability issue not only brought challenges to accuracy and efficiency of decision making, but it also poses major ethical concerns. AI models are trained on real-world historical datasets. If bias exists in a real-world system, then an AI algorithm can exacerbate it. For example, while facial recognition technology has achieving 90%+ accuracy, in racially diverse countries this accuracy may be as low as 65% on women, children, and ethnic minorities. Apple Card was in the recent controversy that it approved much lower credit spending limit on a wife's application than her husband's, with the same family household income. Even if gender or race was not specifically considered in the ML model, related features in the dataset can still embed these biases and lead to unfair decisions. Immediate investment is needed in algorithm interpretability and testing, in addition to executive education around the subtle ways that bias can creep into AI and machine learning projects.

-Security: biometric identity fraud deserves just as much caution as physical identity fraud. Applications like easy purchases with biometric identity verification such as facial recognition are tempting for its convenience, but also leaves vulnerability for exploitation.

-Privacy: personal identifiable information is already collected for purposes such as advertising. Clear guidance on consent giving process not by default, but by affirmative action, and data handling compliance requirement coupled with an enforceable penalty is a high priority for policymakers around the world.

In addition to the AI-specific ethical challenges, there are lots of ethical dilemmas that human being already faced but should be careful handing the decision-making power to algorithms. For example, a classic moral dilemma is the "trolley problem" – if you see a trolley speeding down the track and kill 5 people, there's a lever you can pull to switch the trolley to another track where there stands 1 person, will you pull the lever? How should we design the algorithms for autonomous cars when they face a similar dilemma? Instead of blaming the algorithm for making any decision, it's on us to understand what should be handed to machines to make the decisions for us.

ABOUT THE AUTHOR **Millie Liu**

Millie Liu has focused her career on helping entrepreneurs with deep technology turn their ideas into great businesses with global reach.

She was previously at APT, an enterprise data analytics startup acquired by Mastercard for $600m where she helped Fortune 50 clients such as Walmart and P&G make better strategic decisions leveraging data. She was also the co-founder of an MIT startup working on unsupervised event detection, which later pivoted and became Infervision, an AI precision healthcare platform backed by Sequoia China. Millie is on the advisory board of MIT CSAIL (Computer Science and Artificial Intelligence Lab). She holds a Master of Finance degree from MIT and B.S. in Mathematics from the University of Toronto.

A Call for Policymakers to Harness Market Forces

By Steve Hoffman

Governments around the world, for the most part, have taken a hands-off approach on regulating the use of artificial intelligence for fear of stifling innovation and holding back domestic industries. While this is a wise strategy, AI is becoming integrated into so many aspects of our society and is having such a profound impact that the necessity for careful oversight and governance is becoming increasingly necessary. From the perspective of industry development, it is urgent to solve the problems of algorithm bias, data privacy, content filtering and network security.

Governments cannot just sit back and see what happens. Things are progressing too fast and the stakes are too high. If the wrong software gets into the wrong hands, the consequences can be devastating and irreversible. We've already seen how Facebook's lax oversight of Cambridge Analytica led to the mass dissemination of misinformation that had a direct impact on US elections. With the prevalence of deep fakes and AI bots that can churn out misleading news, there's potential for far greater abuse in the future.

Is banning certain AI applications that manipulate human images and autogenerate news stories the answer? Where do we draw the line between the legitimate and criminal uses of these technologies? The software that can create a deep fake may also be the future of the entertainment industry, as more movies and videos turn to digitally manipulating actors' faces and superimposing them on scenes. The same is true for news generating algorithms, which are being used widely to disseminate legitimate financial updates, weather reports, and other information.

A lot comes down to intent, not the technology itself. Once the algorithms and software are out there, it's too late. Banning them will only keep the software out of the hands of those who want to use them for legitimate purposes. The bad actors will be able to get ahold of them. What we need to do is quickly punish those who use the technologies in ways that harm society, while at the same time encouraging our institutions, researchers, and corporations to come up with countermeasures.

It's wishful thinking that technology, like AI, can be controlled. It can't, and there will always be abuses. The question for policymakers is how can we respond to those abuses quickly? What policies will stimulate and reward those who want to prevent these technologies from causing irreparable harm?

Let's take social networks as an example. Can we put in place legislation that makes it in a social network's best interest to more responsibly manage its data, thoroughly vet and monitor all third-party access, and develop countermeasures to fake news and other emerging threats before they become a major debacle? Increasing the punishments for both intentional abuse of new technologies and gross negligence when it comes to their management, would incentivize entrepreneurs and companies to proactively come up with solutions.

In the future, we'll undoubtedly see a steady stream of new social problems with AI, big data, and other technologies. Trying to legislate all the details surrounding each new technology is too unwieldly and can backfire in terms of developing lasting solutions. Instead, governments should enact policies that promote a rapid market response to existing problems, while encouraging the participants to invest in preventative measures to ward off anticipated threats. Only by harnessing market forces and directing their attention to the most serious dangers can policymakers best reign in the destructive power of emerging technologies.

ABOUT THE AUTHOR **Steve Hoffman**

Steve Hoffman, or Captain Hoff as he's called in Silicon Valley, is the CEO of Founders Space, one of the world's leading incubators and accelerators, with over 50 partners in 22 countries. He's also an angel investor, limited partner at August Capital, serial entrepreneur, and author of *Make Elephants Fly*, the award-winning book on radical innovation.

Always innovating on his life, Captain Hoff has tried more professions than cats have lives, including serial entrepreneur, venture capitalist, angel investor, studio head, computer engineer, filmmaker, Hollywood TV exec, published author, coder, game designer, manga rewriter, animator and voice actor.

Hoffman has a BS from the University of California in Computer Engineering and an MFA from the University of Southern California in Cinema Television. He currently resides in San Francisco but spends most of his time in the air, visiting startups, investors and innovators all over the world.

PART 4 GLOBAL EFFORTS FROM THE INTERNATIONAL COMMUNITY

Mastering the Double-Edged-Sword in Governance of AI

By Irakli Beridze

Scientific progress is yielding new technological tools that can deliver great benefits for society. Artificial Intelligence (AI) in particular, is having a worldwide impact on many sectors – from healthcare to finance. AI could even help us to achieve the 17 ambitious global goals world leaders have set in the 2030 Agenda for Sustainable Development. We should, however, exercise a great care and effort in multilateral policy-making and cross-disciplinary cooperation to discuss the legal and ethical implications of the large-scale use of AI.

To date, self-regulatory approaches by various entities have tried to curb possible harmful effects of AI use in specific disciplines. For instance, American Medical Association proposed a regulatory framework for the responsible evolution of AI in health care. The Netherlands Central Bank released a guidance document containing principles for the responsible use of AI in the financial sector to prevent any harmful effects for banks, their clients, or even the credibility or reputation of the financial sector as a whole.

However, this does not mean that there is no need for action by governments. Regulation in some shape or form may be necessary to reduce the public risks that AI may pose. Although there are some early deliberations on national or international regulations, we are still far from creating real international governance mechanisms. Technological advances are happening faster than our ability to respond and, if governments cannot keep pace, they may fall into a practice of prohibiting or banning in an event to minimise the risk that come with the use of AI. However, these approaches may restrict technology development and stifle innovation.

At the United Nations Interregional Crime and Justice Research Institute (UNICRI), we have established a specialized Centre for AI and Robotics and are one of the few international actors dedicated to looking at AI vis-à-vis crime prevention and control, criminal justice, rule of law and security. We seek to support and assist national authorities, such as law enforcement agencies, in understanding the risks and benefits of these technologies and exploring their use for contributing to a future free of violence and crime. In line with that aim, we are developing pilot projects involving the use of AI to combat corruption, human trafficking, child pornography, the financing of terrorism and to develop solutions for deepfake videos.

In terms of AI governance within this specific domain, we have created a global platform together with INTERPOL to discuss advancements in and the impact of AI for law enforcement. Starting in 2018, we organize an annual Global Meeting on Artificial Intelligence for Law Enforcement. The products of these meetings, which include a joint report in 2019, represents a contribution to advancing the AI governance panorama in the law enforcement community. In connection with the third edition of the global meeting later this year, we will be elaborating a toolkit for responsible AI innovation by law enforcement that will contain valuable guidance and support for law enforcement in developing, deploying and using AI in a trustworthy and lawful manner.

With the emergence of the novel SARS-CoV-2 coronavirus, (COVID-19) and the resulting imposition of lockdowns, limitations of movement of people and closure of borders, the operating environment of law enforcement agencies and security services has suddenly become ever more complex. In response to this growing crisis, many are again turning to AI and related technologies for support in unique and innovative ways, particularly to enhance surveillance. Although governments must do their utmost to stop the spread of the virus, it is still important to not let consideration of fundamental principles and rights and respect for the rule of law be set aside. It is essential that, even in times of great crisis, we remain conscience of the duality of AI and strive to advance AI governance.

Therefore, more than ever, it is essential to guarantee that we do not derail progress toward responsible AI. The positive power and potential of AI is real. However, to truly access it, we must work towards ensuring its use is responsible.

Soft law approaches such as this toolkit can make a valuable contribution to AI governance, particularly in the law enforcement domain where the use of AI is truly an edge case. The positive power and potential of AI is real, however, to access it, we must first work towards ensuring its use is responsible, taking into consideration principles and respect for international law.

ABOUT THE AUTHOR **Irakli Beridze**

Head, Centre for Artificial Intelligence and Robotics

He has more than 20 years of experience in leading multilateral negotiations, developing stakeholder engagement programmes with governments, UN agencies, international organisations, think tanks, civil society, foundations, academia, private industry and other partners on an international level.

Since 2014, he initiated and managed one of the first United Nations Programmes on Artificial Intelligence and Robotics. Initiating and organizing number of high-level events at the United Nations General Assembly, and other international organizations. Finding synergies with traditional threats and risks as well as identifying solutions that AI can contribute to the achievement of the United Nations Sustainable Development Goals.

Mr. Beridze is advising governments and international organizations on numerous issues related to international security, scientific and technological developments, emerging technologies, innovation and disruptive potential of new technologies, particularly on the issue on crime prevention, criminal justice and security.

He is a member of various of international task forces, including the World Economic Forum's Global Artificial Intelligence Council, and the High-Level Expert Group on Artificial Intelligence of the European Commission. He is frequently lecturing and speaking on the subjects related to technological development, exponential technologies, artificial intelligence and robotics and international security. He has numerous publications in international journals and magazines and frequently quoted in media on the issues related to artificial intelligence.

Irakli Beridze is an International Gender Champion supporting the IGC Panel Parity Pledge. He is also recipient of recognition on the awarding of the Nobel Peace Prize to the OPCW in 2013.

Agile, Cooperative and Comprehensive International Mechanisms

By Wendell Wallach

Over the past decade, continual calls have been made in international circles for agile and adaptive governance mechanisms that provide a degree of coordination between the many concerned stakeholders. This becomes particularly critical for the governance of emerging technologies, whose speedy development and deployment pose a serious mismatch for traditional approaches to ethical/legal oversight. As readers of this collection of essays will know, AI has received much attention this past year with more than fifty-five lists of broad principles and an array of specific policy proposals being considered by governmental bodies.

AI offers a perfect pilot project for the creation of new, more agile international governance of emerging technologies. A few different mechanisms have already been proposed. These include recommendations by the UN Secretary General's Higher-Level Panel on Digital Cooperation to the IEEE Ethically Aligned Design Initiative. The OECD has begun work on an AI Policy Observatory. Scholars have proposed other vehicles for monitoring the development of AI, flagging gaps, and developing tools to address those gaps.

Plans are underway for the 1st International Congress for the Governance of AI, which will be hosted by the City of Prague. It was originally scheduled from April 2020 but was postponed until October due to the Covid-19 pandemic. The Congress will go beyond lists of broad principles and specific policy proposals to forge first concrete steps towards implementing the agile governance of AI. In preparation for the Congress a series of experts workshops are being convened to discuss:

- Agile, Cooperative and Comprehensive International Governance Mechanisms
- Hard Law and Soft Law in the Governance of AI
- AI and International Security
- Minimizing and Managing System Failures
- Corporate Self-Governance and Accountability
- Inclusion, just transformation of work and society, and addressing the needs of small nations and underserved communities

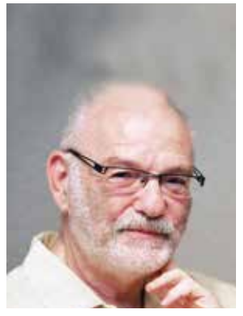
Each of these workshops will develop proposals to put before the ICGAI participants. Should the ICGAI participants overwhelming support any of these proposal, then first steps will be taken for their implementation. The first of these expert workshops was hosted by the Stanford University Digital Policy Incubator on January 6-7, 2020. It proposed the creation of a global governance network as an additional needed institution in the distributed governance of AI.

It is hoped that the Congress will usher in a true multi-stakeholder approach to the governance of emerging technology, including voices from marginalized communities. Of particular importance will participation by representatives from China. While China is the leading implementer of AI solutions in the world, it has to date either not participated in or always been included in many of the other international forums considering the governance of new applications.

For those who feel they can contribute to this conversation, and who wish to participate in ICGAI, registration is available at:

https://www.eventbrite.com/e/the-1st-international-congress-for-the-governance-of-ai-icgaiprague-2020-tickets-86234414455

ABOUT THE AUTHOR **Wendell Wallach**



Wendell Wallach chaired Technology and Ethics studies for the past eleven years at Yale University's Interdisciplinary Center for Bioethics, is senior advisor to The Hastings Center, a fellow at the Carnegie Council for Ethics in International Affairs, and a fellow at the Center for Law and Innovation (ASU). His latest book, a primer on emerging technologies, is entitled, *A Dangerous Master: How to Keep Technology from Slipping Beyond Our Control*. In addition, he co-authored (with Colin Allen) *Moral Machines: Teaching Robots Right from Wrong. The eight volume Library of Essays on the Ethics of Emerging Technologies* (edited by Wallach) was published by Routledge in Winter 2017. 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 in 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, and he is a member of their AI Council for the next two years. Wendell is the lead organizer for the 1st International Congress for the Governance of AI (ICGAI), which will convene in Prague, October 2020.

A Significant Realization by the International Community

By Cyrus Hodes

It seems to me that 2019 will be remembered as a point in time when the international community (governments, private sector, civil society and supranational bodies) had a realization that global governance of an emerging set of intelligent systems maybe a good thing for Humanity.

These are the events I took part in that were, and are, shaping this realization:

- The Beneficial AGI conference in Puerto Rico, led by the Future of Life Institute was an important event realizing the upmost need for a dialog with China on AI Safety, transcending economic tensions.

- The 2nd Global Governance of AI Roundtable: a multi-stake holder / collective intelligence approach set in Dubai as part of the World Government Summit. Besides bringing together 250 international experts in the fields of AI, this year was marked by:

* UNESCO and IEEE meeting to discuss ethics of AI. The IEEE has been presenting its seminal work on AI Ethics while UNESCO has prepared to embark on the leadership on AI Ethics issues within the UN apparatus;

* Gathering of the Council on Extended Intelligence (MIT Media Lab-IEEE);

* First workshop on the Global Data Commons was held with the help of Oxford and McKinsey, over 40 position papers. The GDC is now part of the AI Commons global effort and was taken to AI for Good in Geneva, the UN General Assembly in NY and is about to close the cycle with a presentation at the World Bank Spring Meetings in April with 3 use cases that could be replicated and scaled up globally on sharing data to get to specific Sustainable Development Goals solutions;

* The gathering of AIGO, the OECD expert group on AI in charge of laying out the AI Principles.

- The OECD Principles adopted by the G20 and some partner countries, is an important exercise in summarizing the main recommendations for societies to progress with the use of Beneficial AI.

As a reminder, these principles center on:

• Transparency and explainability

• Robustness, security and safety

• Accountability

• Investing in AI research and development

• Fostering a digital ecosystem for AI

• Shaping an enabling policy environment for AI

• Building human capacity and preparing for labor market transformation

• International cooperation for trustworthy AI

- The resulting OECD AI Policy Observatory to be launched in February with the aim "to help countries encourage, nurture and monitor the responsible development of trustworthy artificial intelligence (AI) systems for the benefit of society".

- The G20 adopting the OECD AI Principles in June 2019 is a consequential step forward keeping in mind that both world leaders in AI (US and China) are part of it.

- UNESCO global AI ethics series: started in North Africa, France, China and Brazil and brought to the table multidisciplinary points of view on a humanistic approach towards the use of AI advancing the discussion with human values for sustainable development.

- In the same vein, The Future Society's AI Initiative has been working with the World Bank to prepare frameworks for developing countries for their national AI Strategies announces the importance of governance of AI and how policy makers could approach it.

- Finally, the Global Forum on AI for Humanity, chaired by French President Emmanuel Macron as part of France's G7 presidency and served as a precursor to the International Panel on AI. The goal of this panel (a bit like the Intergovernmental Panel on Climate Change, IPCC, did), is to become a global point of reference for understanding and sharing research results on AI issues and best practices, as well as convening international AI initiatives.

ABOUT THE AUTHOR Cyrus Hodes

Cyrus Hodes is a Partner at FoundersX Ventures, a silicon-valley based VC firm focusing on early and growth stage AI and robotics startups.

Cyrus co-founded and chairs the AI Initiative, within The Future Society—a 501(c)3 incubated at Harvard Kennedy School—where he engages a wide range of global stakeholders to study, discuss and shape the governance of AI.

He co-leads the Global Data Commons project, together with the UN Secretary General Executive Office and McKinsey, with over 100 global institutions (international organizations, governments, municipalities, private sector and academia).

Cyrus served as the Advisor to the UAE Minister of Artificial Intelligence at Prime Minister's Office. Leading for the past 2 years the Global Governance of AI Roundtable at the World Government Summit in Dubai.

Member of the OECD Expert Group on AI (AIGO), now part of OECD Network of AI Experts (ONE AI)

Member of the Council on Extended Intelligence (MIT-IEEE).

Member of 3 committees of the IEEE Ethically Aligned Design since 2016.

Advisor on AI Ethics at Smart Dubai.

Member of the Steering Committee of AI Commons.

Cyrus was educated at Sciences Po Paris, where he later was a Lecturer.

M.A. (Hons) from Paris II University and M.P.A. from Harvard.

Shifting from Principles to Practice

By Nicolas Miailhe

The global governance of AI has made significant progress in 2019, shifting from principles to practice during what we could call a pivotal year.

By publishing its "Principles on AI" on May 22nd, the OECD established a global reference point. These ethics and governance principles aim to promote artificial intelligence (AI) that is innovative and trustworthy and that respects human rights and democratic values. They were the first set of global principles on AI coming out of a leading multilateral organization and were based on rigorous development process led by a group of independent experts. Their resonance was confirmed by the endorsement, in June 2019, by the G20. To help implement these AI Principles, the OECD also announced the creation of an "AI Policy Observatory" which will provide evidence and guidance on AI metrics, policies and practices, and constitute a hub to facilitate dialogue and share best practices on AI policies.

Subsequently, France and Canada announced during the G7 meeting in August 2019 the launch of a "Global Partnership on AI" (GPAI) hosted by the OECD and which will operate in tandem with the "AI Policy Observatory". Envisioned initially as a sort of "IPCC[Intergovernmental Panel on Climate Change] for AI", GPAI aims to bring together many of the greatest AI scientists and experts globally to foster international collaboration and coordination on AI Policy development among link-minded partners. Both the observatory and GPAI will be launched in 2020. As a precursor to the GPAI multi-stakeholder plenary annual expert meeting, President Macron hosted end of October 2019 the first "Global Forum on AI for Humanity" in Paris. The second edition of the Forum will be held in Canada in the fall of 2020.

Finally, UNESCO General Conference voted unanimously in November 2019 asking the organization to develop, in the next two years, a standard-setting instrument on AI ethics. The process will include extensive multi-stakeholder consultations performed around the world in the frame of the "AI Civic Forum", a partnership between UNESCO, The Future Society, University of Montreal, and Mila.

Concretely, these and many other initiatives launched in 2019 (e.g. the report from the UN Secretary-General High Level Panel on Digital Cooperation; the Digital health & AI Research hub; AI Commons) demonstrate that more and more governments, experts and practitioners are shifting their focus on AI Governance away from just 'what is' or 'what should be' towards 'how to get there'.

Beyond policy-making, we have also seen this pivot from principles to practice happening on the ground, among companies and professional organizations. The IEEE "Global Initiative on Ethically Aligned Design of Autonomous and Intelligent Systems" released in March 2019 the first version of "Ethics in Action" intended to serve as a reference to guide engineers towards the responsible adoption of AI. Beyond, an increasing number of organizations and companies have started to work on translating international AI ethics principles into their respective practice and culture through codes of conducts and charters developed help guide digital transformation efforts towards a trustworthy adoption of AI. Finally, a number of government-backed or independent initiatives on the auditing and certification for AI systems have appeared on the horizon in 2019. The focus of such schemes is precisely to translate principles into practice, and to help shape the competitive race on AI adoption as a race to "the ethical top". As such, besides beefing up of regulatory capacities for example announced by the new European Commission, certification and auditing schemes have the potential to contribute massively to the establishment of the "infrastructure of trust".

ABOUT THE AUTHOR **Nicolas Miailhe**

Nicolas Miailhe co-founded The Future Society in 2014 and incubated it at the Harvard Kennedy School of Government. An independent think-and-do-tank, The Future Society specializes in questions of impact and governance of emerging technologies, starting with Artificial Intelligence through its "AI Initiative" launched in 2015. A recognized strategist, thought-leader, and implementer, Nicolas has lectured around the world, and advises multinationals, governments and international organizations. He is the co-Convener of the AI Civic Forum (AICF) organized in partnership with UNESCO and Mila, and of the Global Governance of AI Roundtable (GGAR) organized yearly during the World Government Summit in Dubai. He is also a Steering Committee member of the AI Commons partnership, a member of the AI Group of experts at OECD (AIGO), of the World Bank's Digital Economy for All Initiative (DE4ALL), and of the Global Council on Extended Intelligence (CXI). Nicolas teaches at the Paris School of International Affairs (Sciences Po), at the IE School of Global and Public Affairs in Madrid, and at the Mohammed bin Rashid School of Government in Dubai. He is also a member of three committees of the IEEE Global Initiative on Ethically Aligned Design of Autonomous & Intelligent Systems, a Senior Research Associate with the Program on Science, Technology and Society at Harvard, and a Fellow with the Center for the Governance of Change at IE Business School in Madrid.

A Global Reference Point for AI Governance

By Jessica Cussins Newman

At the end of 2018, Deep Mind co-founder Mustafa Suleyman predicted that 2019 would be the year we would build global arenas to support international and multistakeholder coordination that would facilitate the safe and ethical development of artificial intelligence (AI). Suleyman wrote that the arenas would need to be global because AI opportunities and challenges don't stop at national borders and don't respect organizational boundaries.

In many ways, Suleyman's predictions were realized; 2019 saw the emergence of several meaningful new global forums including the UN Secretary General's High-Level Panel on Digital Cooperation, the Global Partnership for AI, and the Organization for Economic Cooperation and Development (OECD) Principles and Policy Observatory.

The OECD AI Principles and Policy Observatory in particular represent significant progress in the global governance of AI. Released May 22, 2019, the principles and recommendations became the first intergovernmental standard for AI and a new "global reference point" for AI governance into the future.

All 36 OECD member countries signed onto the OECD AI Principles, as well as several non-member countries including Argentina, Brazil, Colombia, Costa Rica, Peru, and Romania. The European Commission additionally backed the Principles, and Ukraine was added to the list of signatories in October 2019. When the Group of Twenty (G20), released AI Principles one month later, it was noted that they were drawn from the OECD AI Principles. Notably, support from the G20 expanded the list of involved countries to include China.

The principles include detailed calls for inclusive growth, sustainable development and well-being; human-centered values and fairness; transparency and explainability; robustness, security and safety; and accountability. Moreover, the recommendations for national policies and international cooperation include investing in AI research and development; fostering a digital ecosystem for AI; shaping an enabling policy environment for AI; building human capacity and preparing for labor market transformation; and facilitating international cooperation for trustworthy AI. The OECD AI Principles represent widespread awareness of the need for global coordination and cooperation to facilitate trustworthy AI.

The OECD is additionally building on this momentum and aims to help countries implement the principles and recommendations. The OECD launched the AI Policy Observatory at the end of 2019 to facilitate dialogue among global multi-stakeholder partners and provide evidence-based policy analysis on AI. The Observatory will publish practical guidance to implement the AI Principles and a live database of AI policies and initiatives globally. It will also compile metrics and measurement of AI development, and use its convening power to bring together the private sector, governments, academia, and civil society.

The OECD AI Recommendation achieved a feat few would have thought possible just one year previously. The United States signed on at a time of relative aversion to international coordination in other policy arenas. China and Russia were part of a consensus agreement to support the effort more broadly. Other countries are welcome to add their support. While details regarding implementation are still being finalized, 2020 will likely see more substantive AI governance commitments and engagement from a broader range of actors.

ABOUT THE AUTHOR **Jessica Cussins Newman**

Jessica Cussins Newman is a Research Fellow at the UC Berkeley Center for Long-Term Cybersecurity, where she leads the AI Security Initiative, a hub for interdisciplinary research on the global security impacts of artificial intelligence. She is also an AI Policy Specialist with the Future of Life Institute and a Research Advisor with The Future Society. Jessica was a 2016-17 International and Global Affairs Student Fellow at Harvard's Belfer Center, and has held research positions with Harvard's Program on Science, Technology & Society, the Institute for the Future, and the Center for Genetics and Society. Jessica received her master's degree in public policy from the Harvard Kennedy School and her bachelor's in anthropology from the University of California, Berkeley with highest distinction honors. She has published dozens of articles on the implications of emerging technologies in outlets including The Hill, The Los Angeles Times, The Pharmaceutical Journal, and CNBC. Jessica is a member of the CNAS AI Task Force and a member of the Partnership on AI Expert Group on Fair, Transparent, and Accountable AI.

An Important Issue of the International Relations: AI Governance

By CHEN Dingding

With a new round of industrial revolution sweeping the world, artificial intelligence has become the core direction of industrial change. Artificial intelligence is a new engine of economic development, a new focus of international competition, and a new opportunity for social construction. In 2019, as the popularity of artificial intelligence continues to rise at the technological level, its urgency at the governance level is also emerging.

As the focal point of the fourth scientific and technological revolution, achievements in the field of artificial intelligence affect the overall national strength of a country. In 2019, countries have conducted a series of cooperation and competitive interactions around artificial intelligence. To ensure healthy competition in the field of science and technology and continuously stimulate innovation, global governance of artificial intelligence has become an important concern in international relations. Technology competition, trade conflict, information security, and ethical responsibility are all issues in the field of artificial intelligence. The absence of governance norms is not conducive to the positive effects of technology on human society and may even bring about disorder and chaos.

In 2019, countries strived to promote AI governance to keep pace with technological development by holding forums, publishing reports, and formulating specifications. But differences among countries in terms of governance philosophy, development stage, and technological development level pose numerous obstacles to consensus. As major powers in the world today, in 2020, China and the United States should play a leading role in shaping the international order, working with other countries to join the formulation of norms. The two powers are expected to lead the all-dimensional governance of artificial intelligence under the principle of "science and technology for good". Moreover, they should lead countries to jointly respond to the challenges in the development process, and promote the maximum application of technological achievements on a global scale. At the same time, the development of artificial intelligence is still at an unsaturated stage, and there is still much room for cooperation between China and the United States. The two countries should fully recognize the interdependence between the two sides in this industry chain and the broad future prospects of this field, and jointly promote the orderly development of the artificial intelligence industry.

ABOUT THE AUTHOR **CHEN Dingding**

CHEN Dingding is Professor of International Relations, Associate Dean of Institute for 21st Century Silk Road Studies at Jinan University, Guangzhou, China, and Non-Resident Fellow at the Global Public Policy Institute (GPPi) Berlin, Germany, Vice-President of International Studies Association (Asia Pacific region), senior research fellow of the center for global studies at Tsinghua University. He is also the Founding Director of Intellisia Institute, a newly established independent think tank focusing on international affairs in China. His research interests include Chinese foreign policy, Asian security, Chinese politics, and human rights.

PART 5 REGIONAL DEVELOPMENTS FROM POLICY PRACTITIONERS

European Parliament and AI Governance

By Eva Kaili

The value proposition of exponential technologies is compelling. It promises to reduce economic frictions and scarcity in the vital resources, streamline the function of market and public policy procedures, and create new social dynamics, wider inclusion and improved connectivity. Artificial Intelligence is in the core of this transformation.

AI though introduces us to new challenges. New sources of market failures emerge in the area of level playing field of global competitive forces, asymmetries in information possessing and processing, and new types of negative externalities.

In the field of competition, data become the central element of the new global leadership. The ones who can acquire and process data better and smarter, will be the winners. Access to data and technical quality of AI is the next big thing. In order to ensure a level playing field in the new era capacity building and regulatory frameworks will be instrumental in taming oligopolies generated by the prevailing digital platforms. New competition law rules should be designed to take into account not just the turnover of the digital companies but also the volume and quality of data they possess so that the value of their use will be fairly distributed to benefit our societies in respect to the individual rights.

In the same line, we need the development of high quality global technological standards in AI and an environment of research excellence through the development of strong innovation ecosystems linked in a global network. Bad quality of AI might deliver harmful results in the cause of economic development, social inclusion as well as the quality of our Institutions, our Democracy and the Media. High quality technical standards will reduce operational risks, provide legal certainty, improve the quality of options to the citizens, ensure interoperability and accelerate scalability.

European Union aspires to be the global leader in the space of AI, with systematic investments to AI-based innovative solutions, the acceleration of technology transfer mechanisms, a favorable regulatory environment, the strengthening of innovation ecosystems with digital innovation hubs and AI Centres of Excellence, and funding of high quality research projects. In addition, EU plans to develop AI-based pilot projects to experiment with applications of AI in large-scale initiatives, to gain operational experience and then trickle this experience and infrastructure design down to the national, regional and municipal levels of governance.

Artificial Intelligence without mission and social responsibility will end up being "artificial stupidity". High standards, ethical nudges and an enabling regulatory framework are essential. Putting the human in the centre of AI we need to address inequalities of skills, inequalities of access and inequalities to opportunities by planning strategies that improve connectivity and digital education. The quality and standards of AI should technically prevent exclusions and discrimination biases. GDPR set the basis by principles that would protect human rights, without the "one size fits all approach". Algorithms for AI that solve problems or take decisions, should be ethical by design, respecting privacy and the use of our data should be transparent.

As data is in the core of AI, digital platforms should require the consent of the citizens when they collect data and compensate them for the profit of the data they generate. Applications, cameras, microphones and any other way that is used to collect data, should be "by default off" unless citizens are aware of their use and have fair options. Similarly, for example, AI processed targeted messaging should be prevented in the new Media for certain content that is promoted, deep fakes should be flagged, while alternative propositions should be available in order people to have access to balanced information, avoid misperceptions and manipulation of their will.

Finally, the need of a European AI Adjustment Fund so that no-one is left behind, will be my flagship for 2020.

These principles and views epitomize my approach to this challenging technology in these challenging times. I share them with you in hope they can form the basis for a global approach of democracies and a cooperative technological regime between Europe Asia and America, with the good of the citizens and the prosperity of the societies in the core of our strategy for the future.

ABOUT THE AUTHOR **Eva Kaili**

Eva Kaili is a Member of the European Parliament, elected in 2014.

In her capacity as the Chair of the European Parliament's Science and Technology Options Assessment body (STOA) she has, been working intensively on promoting innovation as a driving force of the establishment of the European Digital Single Market. She has been particularly active in the fields of blockchain technology, m/eHealth, big data, fintech, AI and cybersecurity.

Since her election, she has also been very active in the field of taxation, where she has been the Rapporteur of the ECON committee's annual tax report. As a member of the ECON committee, she has been focusing on EU's financial integration and the management of the financial crisis in the Eurozone.

Eva was the Rapporteur of the European Parliament of the Blockchain Resolution, the Legislative Opinion of the EFSI, the Annual Tax Report, and the negotiator of the Social-democratic party in the files of Capital Markets Union and Family Business.

Prior to her position in the European Parliament, she has been elected two times in the Greek Parliament (serving between 2007-2012), with the PanHellenic Socialist Movement (PASOK). She holds a Bachelor degree in Architecture and Civil Engineering, and Postgraduate degree in European Politics. Currently, she is conducting her PhD in International Political Economy.

The European Multi-Stakeholder Approach to Human-Centric Trustworthy AI

By Francesca Rossi

Set up by the European Commission in 2018, the independent High Level Expert Group on AI is composed of a broad spectrum of AI stakeholders, and was mandated to develop guidelines and policies for a European AI strategy. In 2019 the group published two documents: the AI ethics guidelines and the recommendations on AI policy and investment. Both these documents are focussed on the notion of trustworthy AI and are the result of thorough discussions within the HLEG and with the whole European AI ecosystem, and provide a comprehensive blueprint for developing a thriving AI environment in Europe that can have a positive impact across the world.

The AI ethics guidelines define the notion of human-centered trustworthy AI by starting for fundamental human rights, passing to principles, and then listing seven requirements: human control, robustness and safety, privacy and data governance, transparency, fairness and inclusion, societal and environmental well-being, and accountability. They also define an assessment approach that companies can adopt to develop a process for building trustworthy AI and evaluating the compliance of their products and services with these requirements. This is aligned with existing efforts in companies like IBM, where the notion of AI factsheet has been thoroughly evaluated, discussed, and tested.

The policy and investment recommendations are very timely, as governments around the world seek input and guidance to define their own AI strategies. They advocate for a risk-based precision-driven approach to possible regulations, that should adapt to the specific context. They also recommend that the public sector, including governments, serves as a catalyst for the update and scaling of Trustworthy AI. This is an important route to expand access to and familiarity with the technology among the individuals that governments serve. They also advocate for strengthening and uniting Europe's AI research capabilities and harnessing an open and innovative investment environment. Placing the human at the centre of AI was at the core of the AI Ethics guidelines and it rightly continues through the policy and investment recommendations. This includes also ensuring that all sectors of the population have the skills to benefit from AI, which leads to the recommendation to redesign the education system from preschool to higher education.

While this effort is focused on a specific region of the world, the independent nature of the group, as well as it multi-disciplinary and multi-stakeholder composition, may and should serve as a leading example where a multilateral approach can bring successful results. The HLEG brings together not just technology experts but representatives of many different sectors, including multiple academic fields, industries, human and consumer rights associations. This is what allowed this process to deliver guidelines and recommendations that are both ambitious and feasible, and thus with high potential of deep, broad, and enduring impact in AI governance.

ABOUT THE AUTHOR **Francesca Rossi**

Francesca Rossi is the IBM AI Ethics Global Leader and Distinguished Research Staff Member at IBM Research.

Her research interests focus on artificial intelligence and the ethical issues in the development and behavior of AI systems. On these themes, she has published over 200 scientific articles, she has co-authored two books, and she has edited about 20 volumes, between conference proceedings, collections of contributions, special issues of journals, and a handbook. She is a fellow of both the worldwide association of AI (AAAI) and of the European one (EurAI). She has been president of IJCAI (International Joint Conference on AI), an executive councillor of AAAI, and the Editor in Chief of the Journal of AI Research. She is a member of the scientific advisory board of the Future of Life Institute (Cambridge, USA) and a deputy director of the Leverhulme Centre for the Future of Intelligence (Cambridge, UK). She serves in the executive committee of the IEEE global initiative on ethical considerations on the development of autonomous and intelligent systems and she is a member of the board of directors of the Partnership on AI, where she represents IBM as one of the founding partners. She is a member of the European Commission High Level Expert Group on AI and the general chair of the AAAI 2020 conference.

The European Union's Governance Approach Towards "Trustworthy AI"

By Charlotte Stix

Over the last two years, the European Union (EU) emerged as a key player in the field of artificial intelligence (AI) governance. Building on the European Commission's 2018 AI strategy, the EU is demonstrating the possibility of an ethically informed, fundamental-rights approach towards AI. In particular, the Ethics Guidelines for Trustworthy AI played a predominant role in this development. *The Ethics Guidelines*, drafted by the High Level Expert Group on AI (AI HLEG), an independent group set up by the European Commission in 2018, took a novel approach to what ethics guidelines can aim to do. Three aspects of the document are particularly noteworthy: (i) it demarcated 'what' AI Europe should strive towards; (ii) it is based on fundamental rights; and (iii) it provides a method to operationalise its suggestions. This piece will briefly highlight each of these aspects, and discuss how they move the European AI governance discussion forward.

The concept of 'trustworthy AI', as introduced by the AI HLEG, quickly became a red thread throughout European policy making. Trustworthy AI is defined as AI that is "lawful, complying with all applicable laws and regulations; ethical, ensuring adherence to ethical principles and values; and robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm." Trustworthy AI, as the type of AI that Europe strives towards, was subsequently picked up and reiterated in the European Commission's Communication: Building Trust in Human-Centric Artificial Intelligence (2019), and has since been a core idea underpinning multiple AI strategies from European Union member states.

A fundamental rights based approach formed the foundation of the entire document, supporting a human-centric and trustworthy route towards AI. By way of in-depth examination, this perspective yielded four Principles: 'respect for human autonomy, prevention of harm, fairness, explicability'. In turn, these Principles formed the groundwork for the development of the 'seven key requirements' ranging from transparency to technical robustness and safety, simultaneously achieving trustworthy AI and an alignment with fundamental rights. This approach is unique, even in light of a current landscape of over 84 sets of AI Principles.

Finally, the Ethics Guidelines provided an assessment list, introduced to guide practitioners and other stakeholders during the implementation phase of the seven key requirements derived from the ethical principles. To ensure that this assessment list was of good use to the ecosystem, the European Commission conducted a large scale piloting process over several months, soliciting feedback from hundreds of stakeholders across Europe. As of this writing, the input received is analysed and will be translated into a revised version of the assessment list. A granular, expertled and principled approach based on fundamental rights and ethics as demonstrated by the processes undergone with the Ethics Guidelines, alongside Commission President Von der Leyen's proposal to establish "a coordinated European approach on the human and ethical implications of Artificial Intelligence" in the first hundred days of her office, puts the EU in a unique position to lead on governance measures for ethical AI in the coming years.

ABOUT THE AUTHOR — Charlotte Stix

Charlotte Stix is the Coordinator for the European Commission's High-Level Expert Group on Artificial Intelligence. Charlotte is pursuing a PhD at the Eindhoven University of Technology, researching the ethics and governance of artificial intelligence and serves as Expert to the World Economic Forum's Global Future Council on Neurotechnologies. She collates the European AI Newsletter, widely seen as the definitive resource for insights into developments in AI policy across the EU. She has been awarded as a Forbes' 30 under 30 in Technology in Europe in 2020 and collates the European AI Newsletter, widely seen as the definitive resource for insights into developments in AI policy across the EU.

Formerly, she was a Researcher at the Leverhulme Centre for the Future of Intelligence, University of Cambridge, a Fellow to the World Economic Forum's AI Council, and a Programme Officer at the European Commission's Robotics and Artificial Intelligence Unit, where she oversaw €18 million in projects and contributed to the formulation of EU-wide AI strategy. She was also an Advisor to Element AI, a Policy Officer at the World Future Council, and a Founder of an award-winning culture magazine, which she grew from scratch to a team of 15.

The Driving Forces of AI Ethics in the United Kingdom

By Angela Daly

The UK Government has linked AI development directly to its industrial strategy, and also seems to view this as giving the UK a potential competitive edge, especially in its post-Brexit trajectory.

Between 2017 and 2018 the UK Government placed increasing emphasis on the national importance of AI, naming it as one of the country's four Grand Challenges in the 2017 Industrial Strategy, and investing in an AI Sector Deal in 2018. The UK Government also envisaged a leadership role for the country internationally in safe and ethical uses of data and AI. It set up a Centre for Data Ethics and Innovation as an advisory body and committed to be an 'active participant' in standard setting and regulatory bodies especially for AI and data protection. Between 2017 and 2018 there was also activity in the UK Parliament, with an All-Party Parliamentary Group on AI set up in 2017 and a Select Committee on AI formed which issued a report in 2018. The Select Committee's report included 5 non-legally binding 'overarching principles', as the basis for a possible cross-sector 'AI Code' that it suggested be formulated and developed by the Centre for Data Ethics and Innovation.

In 2019, the Centre for Data Ethics and Innovation commenced its work. It has focused so far on online targeting and bias in algorithmic decision-making, producing two interim reports on these topics in July 2019, and a series of 'snapshot' reports in September 2019 on ethical issues in AI, focusing on deepfakes, AI and personal insurance, and smart speakers and voice assistants. The Centre for Data Ethics and Innovation is scheduled to deliver formal recommendation to the UK Government in early 2020 on online micro-targeting and algorithmic bias.

There has been significant political instability domestically in the UK during 2019 with a change of Prime Minister and then a General Election in December 2019 which has given the new Prime Minister, Boris Johnson, a large majority in the House of Commons.The UK formally left the European Union on 31 January 2020, and the government now commands a sufficient majority to make and implement law and policy, including on AI.

However, divergence may yet occur within the UK on AI. The autonomous Scottish Government (led by the Scottish National Party) launched its own initiative to develop an AI strategy for the Scottish nation in January 2020. It has since released a scoping paper for public consultation. On the basis of consultation responses, the Scottish Government aims to publish its own AI Strategy in September 2020. It remains to be seen how aligned this strategy will be with the UK's overall approach to AI.

ABOUT THE AUTHOR **Angela Daly**

Dr Angela Daly is Senior Lecturer (Associate Professor) and Co-Director of the Centre for Internet Law & Policy in Strathclyde University Law School (Scotland) and Visiting Professor at the Università degli Studi di Macerata (Italy). She is a socio-legal scholar of new digital technologies, with particular expertise in data protection, telecoms regulation, intellectual property, competition law and human rights in the European Union, the United Kingdom and Australia. She has previously worked at the Chinese University of Hong Kong, Queensland University of Technology, Swinburne University of Technology and the UK communications regulator OFCOM. She is the author of academic monographs *Socio-Legal Aspects of the 3D Printing Revolution* (Palgrave 2016) and *Private Power, Online Information Flows and EU Law: Mind the Gap* (Hart 2016), and the co-editor of *Good Data* (INC 2019). Her current research examines the emergence of law, ethics statements and policy from public and private actors in the EU, US, China and India on artificial intelligence (AI).

Localizing AI Ethics and Governance in East Asia

By Danit Gal

2019 marked the year of moving from AI Ethics and Governance principles to action. In 2017 and 2018, numerous countries, companies, and institutions rushed to publish AI Ethics and Governance principles. Unsurprisingly, we witnessed broad international alignment on core principles such as accessibility, accountability, controllability, explainability, fairness, human-centricity, privacy, safety, security, and transparency. Now we're moving to the implementation stage, as these entities explore what localizing globally shared principles means.

This is a critical rite of passage in AI Ethics and Governance. As we pursue the localization of these principles, we're beginning to see major points of contention between alternative interpretations as well as discover new implementation paths. This is a positive development. AI Ethics and Governance principles can only prove effective if they are put into practice, and that requires adapting them to local needs and realities. Perhaps most common in the localization process is consulting local cultural, religious, and philosophical traditions when defining one's ethics. This is particularly salient in East Asia, where Confucian philosophical traditions, technoanimistic Buddhist and Shinto inclinations, and rich cultural perceptions of technology play a key role in the localization of AI Ethics and Governance principles.

Another notable process of localization is found in the different approaches to the implementation of principles such as privacy and accountability. In the localization of privacy, we see different approaches to data ownership and protection, also critical to AI training, between the EU, US, and China. Championing the GDPR, the EU seeks to empower users and regain individual control over personal data. In the US we're still seeing data being regarded as proprietary by technology companies despite evolving data protection regulations, especially when transacting with third parties. In China, authorities raised the stakes and are actively warning and banning applications deemed to abuse, misuse, and excessively collect user data.

The localization of privacy also feeds into that of accountability, which is central to AI developers. In the EU, US, and China (alongside other countries) we see authorities holding companies responsible for the technologies they develop and distribute. The EU, for example, fines companies directly for misconduct. South Korea, in comparison, takes a different approach in its Ethics Guidelines by dividing responsibility between providers (companies), developers, and users. The South Korean model of accountability offers new challenges and opportunities that are worth exploring, especially as we strive to create more individual accountability by promoting the informed and consensual use of technology.

These are a few examples of the growing AI Ethics and Governance principles localization trend. More research is needed to better understand how these processes take place and how they affect domestic and international technology users. The next step in this process will be to feed instances of these localizations back to principle drafters to share best practices and identify what is still missing. Looking forward, 2020 promises another year of AI Ethics and Governance principles localization, with a proliferation of local interpretations and implementations to learn from.

ABOUT THE AUTHOR **Danit Gal**

Danit Gal is Technology Advisor to the UN Secretary General High-level Panel on Digital Cooperation. She is interested in the intersections between technology ethics, geopolitics, governance, safety, and security. Previously, she was Project Assistant Professor at the Cyber Civilizations Research Center at the Keio University Global Research Institute in Tokyo, Japan. Danit chairs the IEEE P7009 standard on the Fail-Safe Design of Autonomous and Semi-Autonomous Systems and serves on the executive committee of The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. She is an Associate Fellow at the Leverhulme Centre for the Future of Intelligence at the University of Cambridge, and an Affiliate at the Center for Information Technology Policy at Princeton University.

Social Concerns and Expectations on AI Governance and Ethics in Japan

By Arisa Ema

The government took the lead in discussions about AI governance and ethics in Japan. The Ministry of Internal Affairs and Communications (MIC), since 2016, has held the "Conference toward AI Network Society." The conference released the "AI R&D Guidelines" in 2017 and "AI Utilization Guidelines" in 2019. Culminating from inter-governmental and multi-stakeholder discussions, the "Social Principles of Human-Centric AI" was released from the Cabinet Secretariat in February 2019. The "Social Principles of Human-Centric AI" outlines AI governance, allowing industries and sectors to turn its principles into practice. For example, the Japan Business Federation (Keidanren) released the "AI Utilization Strategy: For an AI-Ready Society" that developed an AI use strategy framework in February 2019. Companies such as Fujitsu, NEC, and NTT Data also released AI principles in spring 2019. Both traditional companies and a startup company (ABEJA) organized ethics committees to begin discussions on AI governance and ethics.

While industries commenced the discussion, two incidents in 2019 caught the public's attention and accelerated the importance of discussing AI governance. First, there was a scandal involving a recruitment management company selling users'/students' data to client companies in August. Although the main problem was related to the illegality of using personal information and not the algorithmic bias of AI, this incident was almost the first case in the media involving ethical and legal issues around AI in Japan. The second incident occurred in November, where the Project Associate Professor at the University of Tokyo (a director of an AI company) tweeted racist opinions regarding the company's recruitment policy, and claimed his discriminatory comments were caused by machine learning. The University of Tokyo immediately released its official statement that his tweets contravene the ideals of the University of Tokyo Charter.

These incidents raised social anxieties towards machine learning. In response, three academic communities that were engaged in machine learning released the "Statement on Machine Learning and Fairness" in December, declaring that (1) machine learning is nothing more than a tool to assist human decision making, and (2) machine learning researchers are committed to improving fairness in society by studying the possible uses of machine learning. This research group will organize a symposium in January 2020 to open a dialogue on machine learning and fairness supported by various organizations.

Regarding AI governance and ethics, 2019 in Japan has shown that the lead role in these factors has shifted from the government to business. Simultaneously, the social implementation of AI progresses and, consequently, the ethical, legal, and social concerns regarding AI and machine learning have emerged in Japan. However, multi-stakeholder and inter-disciplinary networks on AI governance have been organized in Japan since 2016, and we will continue to tackle these issues and contribute to the world's AI governance discussions.

ABOUT THE AUTHOR **Arisa Ema**

Arisa Ema is a Project Assistant Professor at the University of Tokyo and Visiting Researcher at RIKEN Center for Advanced Intelligence Project in Japan. She is a researcher in Science and Technology Studies (STS), and her primary interest is to investigate the benefits and risks of artificial intelligence by organizing an interdisciplinary research group. She is a co-founder of Acceptable Intelligence with Responsibility Study Group (AIR) established in 2014, which seeks to address emerging issues and relationships between artificial intelligence and society. She is a member of the Ethics Committee of the Japanese Society for Artificial Intelligence (JSAI), which released the JSAI Ethical Guidelines in 2017. She is also a board member of the Japan Deep Learning Association (JDLA) and chairing Public Affairs Committee. She was also a member of the Council for Social Principles of Human-centric AI, The Cabinet Office, which released "Social Principles of Human-Centric AI" in 2019. She obtained a Ph.D. from the University of Tokyo and previously held a position as Assistant Professor at the Hakubi Center for Advanced Research, Kyoto University.

The Innovation of Singapore's AI Ethics Model Framework

By Goh Yihan and Nydia Remolina

*This research is supported by the National Research Foundation, Singapore under its Emerging Areas Research Projects (EARP) Funding Initiative. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore.

Since 2017, Singapore government identified Artificial Intelligence (AI) as one of the four frontier technologies that would further the groundwork infrastructure that underpins the country's ambitions for its Digital Economy and its Smart Nation ambition. On the one hand, 2019 was a period when fundamental policy initiatives were launched in Singapore. On the other hand, in 2019 the Government reaffirmed the importance of developing and using AI by implementing projects in key high-value sectors and building a holistic AI ecosystem.

The policy initiatives positioned Singapore as one of the leading voices in AI Governance worldwide. Indeed, on April 2019 the country won a top award at the World Summit on the Information Society Forum, a United Nations level platform. The initiatives that contributed to the win included: Asia's first model AI governance framework that was released in January; an international and industry-led advisory council on the ethical use of AI and data; and a research programme on the governance of AI, ethics and data use established through the SMU Centre for Artificial Intelligence and Data Governance that I lead and from where we contribute to the ecosystem by conducting academic research to inform AI and data governance in Singapore and beyond, with a particular focus on legislation and policy.

One of the most relevant cross-sectoral policy initiatives of this year is the Model Artificial Intelligence Governance Framework — or Model Framework — launched in January 2019 as a guide for organizations to practically address key ethical and governance issues when deploying AI technologies. The Singaporean approach helps translate ethical principles into pragmatic measures that businesses can adopt. It is the result of the collaboration between the private sector and regulators and the first attempt of a country in Asia to put together this type of framework. Other jurisdictions lead similar initiatives this year. For example, the European Commission announced its final set of AI and ethics guidelines by March 2019, an approach likely to complement the EU General Data Protection Regulations. On a more international scale, the OECD presented on May 2019 a set of principles on AI to promote the innovative and trustworthy use of AI that respects human rights and democratic values.

Additionally, Singapore launched in October 2019 the National AI Strategy (NAIS) that will see over S$500 million committed to funding activities related to AI under the Research, Innovation and Enterprise 2020 Plan, in hopes of furthering AI capabilities in these fields. Highlighted in the NAIS, Singapore will start by focusing on five key sectors to concentrate its efforts on - transport and logistics, smart cities and estates, safety and security, healthcare, and education. These National AI projects aim to channel investment for research and development, anchor talent and guide the development of supporting digital infrastructure in Singapore.

What do we expect for next year? We look forward to keeping consolidating the AI ecosystem in Singapore from the academia by publishing cutting-edge research that can contribute to convene and facilitate dialogue, across academic, industry and regulators, especially between organisations in the Asia Pacific region. We also expect that regulators will continue to develop their initiatives towards having trustworthy AI, such as the second version of the AI Model Framework from IMDA, and the Veritas initiative announced by the Monetary Authority of Singapore which will translate into practice the principles-based approach for AI that the financial regulator has adopted.

ABOUT THE AUTHOR

Goh Yihan

Professor Goh's research focuses primarily on the law of contract and torts, with a secondary interest in the principles of statutory interpretation and the legal process. He has published numerous books, chapters and journal articles internationally and in Singapore, which have been cited on multiple occasions by the Singapore courts and the Federal Court of Malaysia. He has been appointed amicus curiae before the Singapore Court of Appeal and the Singapore High Court. In recognition of his invaluable contributions to the development and advancement of Singapore law, he became the youngest recipient of the pentennial Singapore Academy of Law Singapore Law Merit Award in 2013. He obtained his LL.B. (First Class Honours) from the National University of Singapore on a University Undergraduate Scholarship, where he graduated as the top student in 2006. He subsequently obtained a LL.M. from Harvard University in 2010 on a NUS University Overseas Graduate Scholarship.

Nydia Remolina

Nydia Remolina is a Research Associate at the Singapore Management University´s Centre for AI and Data Governance. She holds a Master of the Science of Law from Stanford University and has more than ten years of experience in the financial services industry, currently acting as an advisor for financial regulation, digital transformation and Fintech for financial institutions. Nydia has also been the manager of policy affairs at Grupo Bancolombia, a financial conglomerate headquartered in Latin America, a senior advisor to the Organization for Economic Cooperation and Development (OECD), and Foreign Attorney at Sullivan & Cromwell LLP (New York Office). She has taught or delivered lectures at several academic institutions in the United States, Asia, Europe, and Latin America, and she has been invited to speak about fintech and financial regulation at various organizations, including the International Monetary Fund (IMF), the International Organization of Securities Commissions (IOSCO) and the U.S. Securities and Exchange Commission (SEC). Her main areas of work and academic research include financial and banking regulation, securities regulation, fintech, legaltech, and the intersections of law, finance and technology.

The Grand Indian Challenge of Managing Inequity and Growth in the AI Era

By Urvashi Aneja

Little progress has been made on the issue of AI governance in India this past year. Despite artificial intelligence being seen as a catalyst for economic growth and a solution for complex socio-economic challenges, India is yet to articulate a framework for how this technology should be governed. Much of the policy conversation has been informed by the private sector, with minimal consultation of civil society or academia. As a result, unlocking the potential of AI is seen primarily as a technical challenge, that can be addressed through the creation of a better innovation and start-up ecosystem, investments in skilled manpower, and creation of national data infrastructures. The societal challenges and risks have received comparatively little attention. To date, there is little meaningful conversation at the policy level on issues of access, equity, fairness and accountability. The data protection bill - yet to be finalised - also does not deal with the challenges posed by machine learning systems. The primary concern seems to be around finding ways to leverage personal data for public good and AI development, rather than privacy or social justice. The lack of governance frameworks is a critical concern, as AI is already being deployed in public systems. Police departments across the country are using predictive analytics as well as automated facial recognition systems. Plans are also underway to deploy AI based systems in both judicial and welfare delivery systems. India seeks to be a global AI leader, but this necessitates not just being at the forefront of innovation, but also developing normative frameworks and governance systems that align AI trajectories with societal needs. Blind technological optimism might entrench rather than alleviate the grand Indian challenge of managing inequity and growth.

At a global level, the past year has seen the proliferation of ethical frameworks for the governance of AI. But these are likely to be inadequate - they typically comprise of vague commitments by governments and technology companies, with no enforcement or accountability mechanisms. A more promising direction is to tether AI governance to already established and widely recognised international human rights frameworks. But, it is important to recognize that the issue of AI governance extends beyond the violation of specific human rights or individual harm. The growing use of AI can lead to increasing inequality, concentration of power, entrenchment of discriminatory and exclusionary systems, and even the creation of a surveillance society. Just as AI is not a silver bullet to address socio-economic challenges, neither is a single set of regulatory or governance frameworks adequate to address these societal harms. Governing AI will require a range of public policy interventions - from competition law to curb the powers of Big Tech to sector specific standards and risk assessments. India currently is yet to address these issues, with the few existing governance conversations limited to how Indian data can be leveraged to improve India's AI readiness and competitiveness.

AI presents a wicked problem for public policy - one that consists of multiple interacting systems, both social and technical; in which there is uncertainty about the impacts and risks; and in which the divergence between various stakeholders is one of competing values and world views. Addressing wicked problems requires engaging multiple stakeholders in iterative and adaptive strategies; enabling collaborative sense-making, experimentation, and learning; and building capacities for reflexiveness and foresight.

ABOUT THE AUTHOR **Urvashi Aneja**

Urvashi Aneja is CoFounder and Director of Tandem Research, an interdisciplinary research collective in India, that generates policy insights at the interface of technology, society, and sustainability. Her research focuses on the societal implications of data-driven decision making systems in the global south. She is also Associate Fellow at the Asia Pacific Program at Chatham House; a member of the T-20 Task Force on the Future of Work & Learning; and a regular contributor to national media publications.

PART 6 EMERGING INITIATIVES FROM CHINA

Benefit in Partnership

By FU Ying

Super-intelligent AI is still a way off but artificial intelligence already exceeds human capacity in many growing areas, sparking huge expectations as well as fear and concern. Both the United States, the AI leader, and China, which is rapidly creating massive applications, should shoulder the responsibilities for what needs to be done.

But before we can talk about the future, we need to consider whether we are going to do it together. Worsening US-China tensions cannot but have an impact on how we deal with the challenges down the road. Should we work to make technology symbiotic to mankind and ensure that the technological advances will make our civilisations prosper? Or would we go separate ways and use the technology to undermine, even hurt, the other side?

After three decades of rapid industrialisation, China finds itself among the top echelon in advancing AI technology and is aware of the needs of rule-making that comes with its advancement. China's AI governance expert committee, set up by the Ministry of Science and Technology in February 2019, has released eight AI governance principles. They include: harmony and human-friendliness, fairness and justice, inclusiveness and sharing, respect for privacy, security and controllability, shared responsibility, open collaboration, and agile governance. Efforts are also being made to put these principles into practice.

AI research is the product of global collaboration, with researchers sharing ideas and building on each other's work. With multinational AI platforms expanding globally, countries need to agree on ethical norms and industry rules. China is open to discussing and working with other countries on this. Our efforts in AI governance need to be connected to similar efforts in other parts of the world, the US in particular.

Neither China nor the US can monopolise the world's technological progress. If they complement each other, the prospects for AI technology will be brighter; if they stop working with each other, both will suffer and the general progress will pay a price. It would be self-destructive to allow geopolitical and a zero-sum competitive philosophy to dominate relations.

The US view of hi-tech as an area of strategic rivalry is not a perspective shared by China. While there is competition, the reality in the field is a kind of constructive and strategic mutual dependency. According to Clarivate Analytics, from 2013 to 2017, the number of AI-related papers co-authored by Chinese and Americans grew the fastest, reaching 4,000 in five years.

American companies lead the way in technologies, and American universities are ahead of the global pack. China has the largest user market and therefore provides faster iterative upgrading of algorithms. Both countries can benefit tremendously in a partnership, unless the US forces a decoupling and pushes China to find other partners or to develop its own solutions – which would also weaken US companies' position and influence.

For China, the preferred path is to encourage collaboration in developing common rules for safe, reliable and responsible AI.

ABOUT THE AUTHOR **FU Ying**

FU Ying is the Chairperson of the Center for International Security and Strategy, Tsinghua University (CISS). She is Vice-Chairperson of the Foreign Affairs Committee of China's 13th National People's Congress (NPC).

FU Ying started her career with China's Ministry of Foreign Affairs (MFA) in 1978 and had long engaged in Asian affairs. She served successively as Director of a Division in Asian Affairs Department of MFA and then was promoted to Counselor of the Department. In 1992 She joined UN peacekeeping mission in Cambodia. She was appointed Minister Counselor at Chinese Embassy in Indonesia in 1997, Chinese Ambassador to the Philippines in 1998, and Director General of Asian Department of MFA in 2000. She then was appointed Ambassador to Australia (2004-2007), and Ambassador to the United Kingdom (2007-2009). She served as Vice Minister of Foreign Affairs for European Affairs and then for Asian Affairs (2009-2013). FU Ying was elected deputy to China's 12th and then 13th NPC (since 2013) and served as Chairperson of the Foreign Affairs Committee and spokesperson of the 12th NPC (2013-2018). She took on her current NPC position in 2018.

Progress of Artificial Intelligence Governance in China

By ZHAO Zhiyun

China has always attached great importance to the governance of Artificial Intelligence (AI). On the ninth round group learning of the Political Bureau of the CPC Central Committee, which is the highest decision-making agency, the General Secretary Xi Jinping emphasized the demand to integrate multidisciplinary resources to strengthen the research on AI-related laws, ethics and social issues and establish and improve laws, regulations, systems and ethics to guarantee the healthy development of AI. The released national "Development Planning for a New Generation of Artificial Intelligence" has made clear deployments in following aspects, to conduct researches on AI relevant legal issues and regulations in such key areas as autonomous driving and robotics; to promote researches on AI behavioral science and ethics; to establish ethics and codes of conduct for R&D and designers; and to actively participate in the global AI governance.

On February 15, 2019, to strengthen the research on AI-related laws, ethics, standards, and social issues, and to get deeply involved in the international cooperation of AI governance, the Ministry of Science and Technology (MoST) initiated the establishment of the New-generation AI Governance Professional Committee consisting of experts from colleges and universities, research institutes and enterprises. On June 17, 2019, the Committee released the "Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence", which proposed eight principles, namely, harmony and human-friendliness, fairness and justice, inclusiveness and sharing, respect for privacy, security and controllability, shared responsibility, open collaboration, and agile governance. The eight principles gained profound echoes worldwide, of which partly due to its combination of global standards and Chinese characteristics. Subsequently, Beijing and Shanghai have released their own local AI governance principles or initiatives, such as "Beijing AI Principles", "Chinese Young Scientists' Declaration on the Governance and Innovation of Artificial Intelligence Shanghai, 2019" and "Shanghai Initiative for the Safe Development of Artificial Intelligence". Industries came up with governance principles based on their own, such as by Tencent and by MEGVII. All the above moves make a big impact.

In 2020, China's priority will be the implementation of the said eight governance principles. The aim will focus on accelerating the formulation and improvement of AI-related laws, standards and norms and making AI governance more legalized, more refined and more institutionalized. Given that AI governance is a global issue, international cooperation will be an important part for China's AI governance. In order to promote the healthy development of next-generation AI, China will always adhere to the cores of openness and cooperation in promoting the next-generation AI governance, to positively participate in the global AI governance agenda, to build international platforms including the World Artificial Intelligence Conference, and to keep communicating with the global players.

China is ready to work with any other countries or organizations around the world to promote AI which is good for all the human being.

ABOUT THE AUTHOR **ZHAO Zhiyun**

ZHAO Zhiyun, PhD in Economics, Professor, Doctoral Supervisor, the Party Committee Secretary of Institute of Science and Technology Information of China (ISTIC), Director of New-Generation Artificial Intelligence Development Research Center of Ministry of Science and Technology of the People's Republic of China (MOST). ZHAO Zhiyun is granted with the Special Government Allowance provided by the State Council, and selected for "New Century Million Talents Project", National Cultural Expert and Theorist of "Four Groups" and Leading Talent of the "Ten Thousands Talent Plan". She is well-known as a leading talent in economic theories and policies, and S&T management and policies. She especially has unique insights on emerging technology and industrial development. She pays great attention to the issue of AI governance, and focuses on promoting related research and cooperation between China and other countries. She has won outstanding achievements in the construction of theoretical system, in the promotion of technological progress, and in the related disciplinary construction. She has published more than 30 academic monographs, 4 Chinese translations, and more than 130 academic papers. As the Principal Investigator, she takes charge of nearly 30 national, provincial and ministerial research projects, including National Key Research and Development Project, National Sci-Tech Support Plan and National Soft Science Major Project.

From Principles to Implementation, Multi-Party Participation and Collaboration are Even More Needed

By LI Xiuquan

In 2019, the governance of AI has drawn wider attention from the international community. International organizations, governments, academia, and enterprises continue to explore values of new technological and publish their own principles for the development of AI. China also released "Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence" in 2019. The international community has formed a consensus statement around such key issues as people orientation, fairness, transparency, privacy, and security, reflecting that all parties have formed a universal value concept for the development of AI.

At the same time, the focus of global AI governance is moving from the formulation of principles to continuous refining and implementation of these principles and guidelines. In this process, it is more important to fully absorb the opinions of stakeholders. Compared with the previous stage, it will require more extensive multi-party participation and closer collaborative governance.

The application of AI will bring about various influences on the future society's economic activities, public management, travel, etc., and it will affect all walks of life and various groups. From governance principles to detailed rules and regulations, it is not enough to rely solely on government officials and experts. It requires the joint efforts and active participation of the government, academia, industry, and the public. China is continuously promoting the implementation of AI governance principles in the construction of AI innovation pilot areas and AI open innovation platforms, and put forward the governance rules in various fields through the exploration practice. It is particularly important to establish an effective opinion collection and feedback mechanism to enable all sectors of society to participate in the governance of AI, and thus to incorporate the appeals of different groups, especially vulnerable groups and other stakeholders, into the detailed rules.

Similarly, from a global perspective, different countries have different national conditions and different ethnic groups have different histories and cultures. The implementation of AI principles requires effective communication and coordination. It is helpful to establish a more diversified collaborative governance platform to strengthen dialogue and communication among countries and make differences fully collide and merge with each other in pragmatic communication, which will definitely help to form a broader consensus, and enable AI to better improve the people's livelihood and well-being in all countries.

ABOUT THE AUTHOR **LI Xiuquan**

Dr. LI Xiuquan is now Research Fellow of Chinese Academy of Science and Technology for Development (CASTED), and Deputy Director of New Generation Artificial Intelligence Development Research Center of Ministry of Science and Technology. He received his Ph.D. degree, in field of Computer Science, from Tsinghua University. He is also joint PhD in Information Science, University of Hamburg, Germany. He has many years of research experience in AI fields, such as multidimensional time series data modeling and prediction, and brain-controlled robot system based on EEG. His current research area is big data and AI technology foresight and evaluation, industrial technology roadmap and AI innovation policy research. He has strong interest in the study of the frontier trend of intelligent transformation, and the demands for innovative policies in various aspects of AI development such as research, industry and governance. He has presided over 10 research projects such as "Research on the Major Strategic Issues of Chinese Intelligence Economy and Intelligence Society development", "Research on the Leading Trends and Policies of Artificial Intelligence at Home and Abroad".

Towards Robust and Agile Framework for Ethics and Governance of AI

By DUAN Weiwen

In 2019, four aspects in AI ethics and governance in China deserve attention. Firstly, various principles, standards and declarations of AI ethics and governance were released. These include "Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence", the "Beijing AI Principles" released by Beijing Academy of Artificial Intelligence (BAAI), the artificial intelligence ethical principles in "AI Ethical Risks of AI Research Report" proposed by Artificial Intelligence Working Group, SAC, "Chinese prospects for the Standardization of Robot Ethics" (2019) by National Robotics Standardization Working Group and Peking University. Meanwhile, CCID and CAICT under the MIIT of China, respectively, have proposed the declarations or conventions of AI ethics, and Tencent also released its own AI ethical framework. Not only legal and philosophical scholars participated in related research, but researchers in the field of AI also shown great interest in the research of ethics system of AI and safe and reliable AI, etc. Secondly, certain progress has been made in the legal regulation of personal information protection and data rights, data governance, and data compliance. For example, the "Act on the Protection of Personal Information" and the "Data Security Law" has been included in the legislative plan for the next year; and MIIT has carried out the special rectification action against the APPs infringing on the rights and interests of users. It is worth mentioning that the revised draft of the Law on Protection of Minors emphasizing that informed consent is required to collect information about minors. Thirdly, AI applications such as face recognition are rapidly spreading and causing lots of ethical and legal disputes. Although the abuse of face recognition in classrooms, parks and other scenes has led to public criticism and even legal proceedings, its application in China seems unstoppable. In addition, AI companies have also conducted some ethical and governance practices. Leading companies such as Tencent have proposed Technology for Good as its goal, and applied AI to prevent game addiction and find lost children. Megvii, one of China's facial recognition giants, also released AI Application Criteria, which are used for internal review by its AI ethics committee. However, given that these efforts are far from being the basis, such as KPI, on which companies evaluate their products and services, they are inevitably criticized as flexible PR or some kinds of ethics washing.

All in all, China is generally more optimistic about the positive impact of AI on the economy, society, enterprises and personal well-beings. However, the ethical risks of AI are not fictitious. On the one hand, while enjoying the convenience of innovation, ordinary users will inevitably be concerned about the abuse of personal data and the opacity of algorithmic decisions. On the other hand, developers also worry that a lack of ethical regulation will make them pay a high price for the risks involved. In order to eliminate this double anxiety, it is necessary to carry out the ethical adjustment through ethical assessment of technology, "technology-ethics" correction and the construction of trust mechanism for AI. What's more important is to build a robust and practicable framework for ethics and governance of AI to achieve agile governance on the basis of full consideration of the social impact of AI, regional and global compatibility, and maintenance of the fundamental condition - world peace.

ABOUT THE AUTHOR **DUAN Weiwen**

DUAN Weiwen is the Director and Professor of the Department of Philosophy of Science and Technology in the Institute of Philosophy, Chinese Academy of Social Sciences (CASS), and he is also Distinguished Professor in University of CASS, and the Director of the Research Center for Science, Technology and Society, CASS. He holds a Bachelor of Science degree in Physics from Central China Normal University, and a Master of Philosophy and PhD degree in Philosophy of Science and Technology from Renmin University of China. He specializes in philosophy of science, philosophy of information technology, etc. In recent years, he has focused on the philosophical, ethical and social research of big data and AI. He was a visiting scholar in Oxford University (with Luciano Floridi), Colorado School of Mines (with Carl Mitcham), and University of Pittsburgh (with Edouard Machery). He is on the editorial board of the Journal of Information, Communication and Ethics in Society and Journal of Responsible Innovation, and he is one of the deputy chairmen of the Committee of Big Data Experts of China. He is now the chief researcher and project leader of several important and general social science fund research projects, including Philosophical Studies on Intelligence Revolution and Deepening Techno-scientific of Human Being (2017-2022), which is supported by the National Social Sciences Founding of China (NSSFC). He is the author of several books, including *Acceptable Science: Reflection on the Foundation of Contemporary Science, Ethical Reflection on Cyberspace*, and *Truss up Time: Technology and Life World*, etc.

Globalization and Ethics as the Consensus of AI Governance

By LUAN Qun

In 2019, AI governance is characterized by globalization and ethical integration. The major countries, economies and international organizations in the world have successively released documents on AI governance. The most representative ones are the EU Ethics Guidelines for Trustworthy AI (April 2019), the joint statement and "G20 AI Principles" (June) adopted by the G20 Digital Economy Ministers' Meeting and G20 Trade and Digital Economy Ministers' Joint Meeting held in Tsukuba, Japan; and, also in June, China's National New Generation AI Governance Expert Committee issued "Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence". China's AI governance, has also been transferred to ethical governance from the planning of the State Council and related departments in 2017, such as the "New Generation of Artificial Intelligence Development Plan" and the "'Internet+' Three Year Action Plan for Artificial Intelligence", as well as industry and domain plans such as the "Three-year Action Plan on Promoting the Development of A New Generation of Artificial Intelligence Industry (2018-2020), 2018 Intelligent Manufacturing Pilot Demonstration, and the "AI Innovation Action Plan for Universities", etc. This is highlighted by the emphasis on "responsibility" in the new generation of AI governance principles, which is the same meaning as the EU's emphasis on "trustworthiness". In August, the rule of law forum of Shanghai 2019 world AI conference released guidelines for AI security and rule of law (2019). The theme of the forum is "building the rule of law in the future and sharing the benefits of intelligence", so as to promote industrial development and follow-up of relevant systems, better serve and safeguard the overall situation of AI national strategy, and show the Chinese scheme of AI governance to the world.

As the industry management department, the Ministry of Industry and Information Technology mainly implemented the top-level industrial design plan in 2019, such as the "Three-year Action Plan for Promoting the Development of the New Generation of Artificial Intelligence Industry" (2018-2020), which mainly cover eight products and three technologies, the development plan and standards for key industries, such as the "Auto Driving Action Plan for the Development of the Internet of Vehicles (Intelligent Connected Vehicles) Industry", "Key Points for the Standardization of Intelligent Internet Connected Vehicles in 2019"; and, key work on joint promotion, such as joint efforts with the Ministry of Natural Resources and Beijing to carry out the pilot work of Internet of vehicles (Intelligent Connected Vehicles) and automatic driving map application; and industrial Internet work, such as the implementation of the Guide for the Construction of Integrated Standardization System of Industrial Internet. All of these new policy documents involve the related discussions on AI governance.

ABOUT THE AUTHOR **LUAN Qun**

Dr. LUAN Qun joined China Center for Information Industry Development in 2011 as the Director in the Institute of Policy and Law, holding a PhD in Civil and Commercial Law from the China University of Political Science and Law. He is an industry expert in the civil and commercial law and industrial economy and policy and leads the Legal Services Centre for Industry and Informatization. His recent consulting work has centered on industry strategy, business development and supervision, with a special focus on autonomous vehicles, industrial data and manufacturing. He has carried out successful projects for industrial development planning and interpretation of industrial policy in Nei Mongol, Henan and Shandong province. He has published more than 50 articles in "Learning Times", "China economy and Informatization", "Modern Industrial economy", "Economic Daily", "China Electronic Journal" and other magazines and newspapers.

The Principles of Well-being of Human Person and Accountability

By GUO Rui

In 2019, Artificial Intelligence (AI) affected every aspect of people's lives all around the world, with its increasing application in business, healthcare, transportation, financial services, education, and public safety. For the public and the policy makers, whether the negative impacts of AI will be properly handled, such as the leakage of personal information, the output of poorly-trained AI, and the misuse of AI, causes more and more concerns. The academia, the industry and the policy makers have actively joined the AI-ethics-related discussions and debates, making 2019 a critical juncture for the global community to move towards a consensus on AI governance.

Experts from industries, academia and civil societies have gradually come to a consensus that the negative impacts related to AI are best treated as risks, and could be identified, prevented and managed through a rigorous risk-management system. The insight has informed the standardization work, and much ethic-related standardization is steadily advancing and gaining momentum. This consensus is leading to a governance system that allows the world to reap the benefits and prevent the harms of AI. Although the concept of risk is helpful to deal with the known and immediate negative impacts of AI, it certainly does not exhaust all those AI brings, especially the uncertain and long-term ones. We should continue to explore ways that could help human society to deal with AI ethical issues.

In my capacity as the Lead Expert for the Research Group on the Ethics of Artificial Intelligence of the Artificial Intelligence Working Group, Standardization Administration of the People's Republic of China (SAC), I proposed that two principles need to be followed for Ethical and Responsible AI. First, Ethical and Responsible AI implies the principle of the well-being of human person. Promoting the well-being of human person should be the ultimate goal of AI research and applications. Second, Ethical and Responsible AI implies the principle of accountability. These two principals have informed the drafting of the AI Ethical Risk Research Report (published in May 2019 by Artificial Intelligence Working Group, SAC).

ABOUT THE AUTHOR **GUO Rui**

GUO Rui (Associate Professor of Law at Renmin University of China, researcher of RUC's Institute of Law and Technology, and Director of Center for Social Responsibility and Governance). Dr. GUO Rui researches on corporate law, financial regulations, human rights, and the ethics of AI. He graduated from China University of Political Science and Law (LL. B & LL.M) and Harvard Law School (LL.M & S.J.D). Professor GUO Rui is a member of the Sub-Committee of User Interface, National Standardization Committee of Information Technology, and the Lead Expert for the Research Group on the Ethics of Artificial Intelligence appointed by Artificial Intelligence Working Group, Standardization Administration of the People's Republic of China (SAC). He participated in the drafting of the first AI standardization white paper (published in 2018), and led the drafting of the AI Ethical Risks of AI Research Report (published in May 2019 by Artificial Intelligence Working Group, SAC).

Better AI, Better City, Better Life

By WANG Yingchun

AI research institutions, enterprises and application scenarios are mainly located in cities across the globe, thus cities are playing a prominent role in AI's development. As China's largest economic center, Shanghai is speeding up its march to become a global AI highland in terms of research and development of technology, application demonstration, institutional supports and talents attraction. Echoing "Better City, Better life", the theme of 2010 Shanghai World Expo, we need to seek paths and solutions for harmonious coexistence of human-AI to achieve the goal of "Better AI, Better City, Better Life"in the age of artificial intelligence.

Cities provide an experimental platform to promote AI development in a healthy way. In 2019, the Ministry of Science and Technology has issued the "guidelines for the construction of the national new generation artificial intelligence innovation and development pilot zone", which stress to take the city as the main carrier to explore replicable and generalizable experiences, and to lead the healthy development of artificial intelligence in China. On May 25, 2019,the Ministry of Science and Technology and the government of Shanghai Municipality jointly launched the" National New Generation of AI Innovation and Development Pilot Zone" in Shanghai. The pilot zone takes AI governance as one of the four core elements to promote scientific and technological innovation and institutional innovation. On the one hand, it supports to research and develop responsible artificial intelligence, and to encourage innovation in artificial intelligence applied in Shanghai; on the other hand, it strengthens the exploration in laws and regulations, ethical norms, safety supervision and other aspects of artificial intelligence, and contribute "Shanghai experience" in the artificial intelligence development in China and around the world. A focal concern is on how to provide citizens with higher quality medical care, more convenient transportation and safer and efficient urban services based on artificial intelligence technology.

Openness and collaboration are crucial in achieving *Better AI.* Shanghai has hosted the World Artificial Intelligence Conference for two years. In his congratulatory letter to World AI Conference 2018, Shanghai, President Xi Jinping pointed out that "we need to deepen cooperation and jointly explore the emerging issues of artificial intelligence". We organized the Governance Forum of World AI Conference 2019. At the Forum, dozens of international experts and participants from more than 200 government and industry attended. The involvement of global experts enhanced mutual understanding through open exchanges and has reached consensuses on some important issues. At the forum, the "Chinese Young Scientists'Declaration on the Governance and Innovation of Artificial Intelligence Shanghai, 2019 "was issued. It raised four major responsibilities to be followed in the development of artificial intelligence, namely, "Ethical Responsibility"," Safety Responsibility", "Legal Responsibility" and "Social Responsibility". Taking the forum as a starting pojnt, we hope to promote the formation of a global community of AI governance research and collaboration. We also aim to shed light on governance approaches.

Cities can play a vital role in the formation of global AI governance system. This system may consist of multi-subsystem programs and regional-programs on the basis of respecting cultural and institutional diversity. We need to ensure that these subsystems and regional programs are globally compatible and open-minded, and figure out the specific mechanisms for benefit sharing and security. Cities around the world can have more in-depth exchanges and cooperation on these aspects, and we have carried out relevant work in 2019.

We participated in the researching work for the construction plan of the Shanghai pilot zone, and are preparing to build Shanghai Academy of Artificial Intelligence Governance. We have gathered multi-disciplinary experts to work on systematic research on the ethical framework of artificial general intelligence and relevant legal, and social issues of narrow artificial intelligence. We hope to continue to work with friends at home and abroad on the path and scheme of harmonious coexistence of human and artificial intelligence.

ABOUT THE AUTHOR **WANG Yingchun**

WANG Yingchun, PhD, Head of Research Department of Science, Technology and Society at Shanghai Institute for Science of Science, areas of expertise include innovation transformation and innovation governance, and science, technology and society. He initiated and organized a multidisciplinary AI research group to conduct systematic research on AI. He has undertaken a number of consulting projects entrusted by the Ministry of Science and Technology and the government of Shanghai municipality, and has continuously participated in the research and policy drafting of the government's AI policy. He led the organizing work of the Governance Forum under World AI Conference 2019 in Shanghai. At the moment, he is also responsible for the running of the Secretariat of the Expert Advisory Committee of the National New-generation AI Innovation and Development Pilot Zone in Shanghai.