FOREWORD

Over the next decade, artificial intelligence is set to transform the global landscape with extraordinary force. Every industry, every aspect of life, and the functioning of all institutions of state and civil society will undergo a radical reimagining as machine-learning technology and other techno-sciences continue their rapid ascent.

In accordance with UAI, our national Artificial Intelligence Strategy launched in 2017, the United Arab Emirates has placed AI at the core of our vision for economic growth, innovation, development, and human well-being. We are making substantial investments in AI-enabled systems and in the creation of test beds for their evaluation, in order to provide our own country, and societies everywhere, empirical information that can beneficially serve public policy objectives.

Furthermore, we believe that harnessing the power of artificial intelligence for the betterment of humanity requires deliberate collective action. In order to facilitate such collective action, we launched the Global Governance of AI Roundtable at the 2018 World Government Summit in Dubai. The Roundtable offers a neutral forum where the international community - governments, multilateral organizations and civil society alike - can discuss and contribute in shaping global, but culturally adaptable, norms for the governance of artificial intelligence.

This report relates the proceedings of this inaugural Roundtable. It describes the current state of AI and offers a range of perspectives from the world’s foremost experts. With future editions of the Roundtable, we aim to develop, in collaboration with the international community, a consensus governance framework, guiding principles for policymakers, and implementation-level guidelines for industry and practitioners.

It is our hope that by establishing such a collective intelligence platform, we can serve as a catalyst in ensuring that humanity reaps the extraordinary promise of artificial intelligence while mitigating its risks. We invite you to join us in this ambitious and extraordinarily exciting, but also necessary, pursuit.

Omar bin Sultan Al Olama
Minister of State for Artificial Intelligence
CONTENTS

1. Background and Objectives
   Page 6  1.1 Introduction
   Page 7  1.2 Objectives
   Page 8  1.3 Methodology
   Page 10 1.4 Participants
   Page 11 1.5 Agenda

2. Roundtable Highlights
   Page 12  2.1 Insights: Key Elements in a Viable System of AI Governance
   Page 16  2.2 Model Practices: Examples of Approaches to the Governance of AI
   Page 18  2.3 Recommendations: Approaches to AI Governance to be Further Studied
   Page 21  2.4 Next Steps: Action Items for the 2019 Roundtable

3. Summary of Committee Proceedings
   Page 24  3.1 WORKING GROUP 1: Mapping AI and Global Best Practices, Reaping Benefits, Mitigating Risks
   Page 36  3.2 WORKING GROUP 2: The Governance of AI
   Page 48  3.3 WORKING GROUP 3: Impact on Cross-Sectoral Dynamics
   Page 56  3.4 WORKING GROUP 4: AI for Good

4. Appendix I
   Page 66

5. Appendix II
   Page 72

6. Acknowledgments
   Page 74
1.1 Introduction
The unrelenting ascent of artificial intelligence will transform the global economy, the functioning of all institutions of society, and the professional and personal lives of citizens. Uniquely in the history of technological progress, AI may also challenge our very conception of what it means to be human. The philosophical, moral, social, political, and security implications of the AI revolution loom large. Understanding and governing this extraordinary technology is imperative, if humanity is to reap AI’s benefits while mitigating its risks.

In 2017, in order to meet this challenge, the United Arab Emirates launched a strategic Artificial Intelligence Initiative as part of its Centennial 2071 project. Through its “UAI” vision – begun under the auspices of His Excellency Omar bin Sultan Al Olama, Minister of State for Artificial Intelligence – the UAE aims to:

• Facilitate the beneficial implementation of AI in government, industry, and civil society.
• Become a leader in AI research and investment.
• Establish itself as a neutral hub and global facilitator for the international community’s efforts to develop frameworks, public policy recommendations, and international accords for the governance of artificial intelligence.

In February 2018, in the context of the World Government Summit, the UAE hosted the inaugural Global Governance of AI Roundtable. Attended by more than 90 leading experts, scientists, and practitioners, the forum began the necessary dialogue on the pathways to the development of effective, yet culturally adaptable, norms that will assure the safe deployment of AI for the betterment of humanity.

This report offers a summary of the insights gleaned from the Roundtable, the most salient among which are:

• The need for continuity in the inclusive, deliberate, and systematic collective intelligence exercise that began at the first Roundtable.
• The development of a multilayered AI governance framework, with international legitimacy and broad applicability in advancing the well-being of all human societies.

1.2 Objectives
The Global Governance of AI Roundtable has three chief objectives:

1. Gathering information about the state of AI technologies, their socio-economic impact, and the state of AI governance practices and policies around the world.
2. Synthesizing that information into a governance framework, actionable public policy options, and implementation-level guidelines that can be implemented by the UAE and other governments around the world.
3. Serving as the world’s authoritative forum for AI governance.

The first objective (information gathering) stems from the recognition that AI research, development, and adoption is not evenly distributed around the globe and that there is a considerable amount of diversity and experimentation in the ways in which governments respond to the opportunities and challenges presented by AI. In such a circumstance, the first step in developing a governance framework that will be robust and effective is to gather information in order to develop a comprehensive knowledge base that can serve as the foundation for dialogue and decision-making.

The second objective (policy development) stems from the recognition that AI technologies are evolving rapidly in a wide range of sectors, bringing with them enormous potential for good but also risks to fundamental human values, including human dignity and privacy. In such conditions, it is essential that stakeholders in government, academia, civil society, and industry converge around a governance framework and set of policy options that facilitate the realization of the potential for good that AI holds while mitigating its risks.

The third objective (the establishment of an on-going forum for the discussion of AI) stems from the recognition that the development of effective policies for AI governance is a complex matter, involving a range of stakeholders and requiring a methodical approach that provides for on-going learning, adaptation, and evidence-based decision-making. As such, the development of effective AI policy requires more than one-off or even annual conferences; it requires the establishment of an inclusive, on-going forum defined by long-term objectives and by a methodical approach to realizing those objectives. The Global Governance of AI Roundtable is designed to serve as that forum.
1.3 Methodology

The Roundtable’s approach to achieving these objectives is shaped by an understanding that an effective system of AI governance must be anchored in current global governance realities, while accounting for changing power dynamics. It must consider the rising influence of new policy avenues (e.g., technical standards and soft law) and must be adaptive and designed to remain robust and relevant over time.

The Roundtable’s approach is also shaped by an understanding that the rise of AI is the result of a dynamic and complex sociotechnical system. In such systems, science, technology, and society are engaged in a continuous cycle of “co-production,” locally and globally: technological and scientific innovation, infused into societies through business, continuously impacts and redefines societal values and thus policies; changes in values and policy in turn continuously shape techno-scientific developments.

Premised on this perspective and oriented toward the specific objectives described above, the Roundtable’s approach to the development of policies for AI governance is characterized by the following essential features.

Strategic Approach

The Roundtable is not simply a conference or series of conferences. The government of the UAE and the Minister of State for AI, in particular, have a long-term strategic vision for the development of norms for the governance of AI. That vision includes the creation of a permanent forum for gathering information and perspectives from a wide range of stakeholders and systematically working toward a governance framework and actionable policy options. The AI Roundtable is the vehicle for the realization of this strategic vision.

Broad-Based

The Roundtable is an exercise in collective intelligence. It starts with a broad-based gathering of knowledge, perspectives, and experience and works toward norms via collaboration, dialogue, and consensus. Being broad-based means the Roundtable seeks input from individuals with a broad range of professional affiliations (representation from government, NGOs, industry, and academia), from individuals representing a broad range of fields of expertise (technologists, of course, but also ethicists, philosophers, political scientists, economists, lawyers, physicians, and individuals engaged in various fields of the humanities), and individuals representing a truly global range of cultures. The Roundtable brings together these views and perspectives for the purpose of developing norms for AI that are actionable and relevant across the globe.

Disciplined and Systematic Approach.

The Roundtable recognizes that discussions of the governance of AI, both because of the multifaceted nature of the topic and because of the diverse backgrounds of participants, can be wide-ranging and, at times at least, amorphous. For this reason, the Roundtable seeks to anchor the discussion in a foundation of shared core values and, based on that foundation, to follow a systematic approach to developing a coherent and generally applicable set of norms for the governance of AI.

Adaptive.

As noted above, the Roundtable’s approach to the development of norms is premised on an appreciation of the complex and dynamic relationships between science, technology, values, and society and on an expectation that policy will have to adapt to changing conditions. For this reason, the Roundtable is designed as a long-term initiative, capable of re-visiting and modifying its frameworks and policy recommendations.

Evidence-Based Decision-Making.

The Roundtable recognizes that policies for the governance of artificial intelligence, if they are to be effective, must be based upon, and continuously tested against, real-world conditions. For this reason, the Roundtable incorporates into its approach a wide-ranging and ongoing search for practices and relevant experience as it works toward actionable and effective norms.
1.4 Participants
On February 10, 2018, more than 90 of the world’s leading thinkers on AI and AI governance gathered in Dubai for the inaugural edition of the Global Governance of AI Roundtable. In keeping with the Roundtable’s broad-based approach, participants represented a wide range of countries, professional interests, and perspectives. Organizations represented included think-tanks and non-profit organizations, academic institutions, government agencies, international and multinational organizations, and private-sector enterprises. A full list of these thought leaders can be found in the Appendix section of this report.

"AI is set to transform the global landscape with extraordinary force. Every industry, every aspect of life, and the functioning of all institutions of society, will undergo a radical reimagining."

1.5 Agenda
In keeping with its objectives, the Roundtable was designed to facilitate information sharing and substantive dialogue among participants. Accordingly, the day’s proceedings took the form, not of formal presentations, but of several concurrent breakout sessions, each directed by a committee chair who led an interactive discussion on a specific topic of AI governance. The format enabled a rapid exchange of information among participants and a crystallization of issues for further study.

The 2018 Roundtable featured fourteen breakout sessions (or committees), organized into four sequential higher-level thematic working groups. The specific topics covered during the 90-minute sessions are as follows.

Working Group 1:
Mapping AI and Global Best Practices, Reaping Benefits, Mitigating Risks
• Committee 1:
  AI Technology Development
• Committee 2:
  Sectoral Development (Healthcare, Transportation, etc.)
• Committee 3:
  Public Sector Best Practices: Regulatory Frameworks, Implementation
• Committee 4:
  Judicial System, Access to Justice, and the Practice of Law

Working Group 2:
Governance of AI
• Committee 1:
  Empowering Innovation, Mitigating Risks: Norms, Metrics, Codes of Practice
• Committee 2:
  Government Objectives and Strategies
• Committee 3:
  International Cooperation on the Governance of AI
• Committee 4:
  Governance for the Transition from Artificial Narrow to Artificial General Intelligence

Working Group 3:
Impact on Cross-Sectoral Dynamics
• Committee 1:
  Impact on Workforce and Mitigation Strategies
• Committee 2:
  Cybersecurity
• Committee 3:
  Public Sector Best Practices: Regulatory Frameworks, Implementation Ethics, and Code of Practice

Working Group 4:
AI for Good
• Committee 1:
  Good Government and Public Sector Efficiency: Objectives and Strategies
• Committee 2:
  Private Sector, Academia, and Government Cooperation
• Committee 3:
  Metrics of Well-being
Over the course of the day’s discussions, a number of key themes and questions were voiced repeatedly, regardless of committee. In this section, we discuss these in three subsections:

- Insights into the elements essential to a viable system of AI governance.
- Model practices for the governance of AI.
- Questions for further study.

We conclude the section with an overview of steps to be taken in preparation for the 2019 Roundtable.

2.1 Insights: Key Elements in a Viable System of AI Governance

Participants in the committee discussions noted a number of considerations related to the creation and implementation of an effective system of AI governance. They are as follows.

2.1.1 Foundation in Values

A number of participants emphasized the importance of grounding the discussion of AI governance norms in a shared set of values. It is, after all, difficult to make meaningful progress toward a set of norms for the design, development, and use of AI technologies if there is not a clear, shared understanding of the values that those norms are supposed to protect and promote.

While the need for an anchor in values is clear, there are several challenges to articulating a set of values that would both win the assent of the wide range of parties interested in and affected by AI technologies and serve as an effective basis for the development of more specific norms. Among the chief challenges noted by Roundtable participants are the following:

Legitimacy

In identifying a set of foundational values, who has the authority to select and define the operative values? Can we cede the authority to a single actor or organization - or does that authority come from the nature of the process of arriving at the values (e.g., a collective intelligence process)? What are the characteristics (inclusivity, allowance for dialogue, identification of transparent criteria for decision-making, etc.) of a process that, by its nature, would impart legitimacy to its results? What are the criteria for selecting the set of values essential to the development of norms of AI governance?

One Set of Values or More

Do we need to arrive at one set of values that is a valid basis for developing the more specific norms needed to cover all forms of AI in all domains? Or might there be multiple sets of values, with each set functioning for a specific domain or sector (medicine, law, warfare, etc.)?

Cultural variation. In seeking to anchor globally applicable norms in a set of values, we must address the question of cultural variation. That variation extends not just to the values themselves, but also to how the values are prioritized and how claims of competing values are resolved. How do we account for cultural variation among values?

Variation Across Sectors

Just as there may be culture-specific variation in values, there may be sector-specific variation as well. How do we account for variation across sectors in operative values and their prioritization?

Changing Values

Values, and the way in which they are prioritized, can change over time. Such change can take place independently of technology but can also be driven by technological change itself. How do we articulate a set of values that allows for such change, change that may be driven by the very object of the governance norms (AI technologies)?

These challenges to identifying a foundational set of values are non-trivial, yet not insurmountable. What is required, if the challenges are to be addressed, is a framework that incorporates both universal (or near universal) values and flexibility in how those values are realized in different domains, sectors, and cultures. Such an anchored, but flexible, framework would meet the need for identifying a foundational set of values while also addressing the key challenges of variation and change.

One such approach, discussed in the Law Committee (Working Group 1, Committee 4), is the analytical model proposed by The Future Society. Under this approach, a small number of universal values are identified in a “values” tier, while variation and change are accounted for in tiers defined for more specific principles, policies, regulations, and implementation-level standards and codes of practice designed to realize those values in real-world circumstances. It is articulated as follows.

- Central value (human beings are equal in fundamental rights, dignity, and the freedom to flourish)
- General principles (ethics)
- Public policy Instruments (norms, legislation)
- Implementation-level instruments (standards, codes of practice, technical controls)
2.1.2 Accounting For Human Intervention

In a number of committee discussions, it was observed that in almost all AI systems currently deployed, some manner of human agency is essential to the functioning of the system, and will continue to be essential in the foreseeable future.

The precise locus and nature of the human intervention will vary considerably from system to system; in terms of locus, human intervention can take place anywhere from the design phase of the system to quality assurance of the system’s final results. In terms of nature, the expected roles of humans vary widely: selecting input data, configuring a system for a specific circumstance, selecting the optimal algorithm for a given set of conditions, measuring the quality of results, and so on. The role of humans may be simply making an informed use of the output of an AI application, a use that requires taking into account both the capabilities and the limitations of the application. Regardless of its specific locus and nature, human intervention is necessary in order to design, develop, control, explain, and measure the effectiveness of increasingly complex and independent AI systems.

The fact that the safe and effective functioning of AI systems requires – and will continue to require, at least for the near future – an element of human intervention has implications both for the development of AI governance norms and for how societies can prepare for the widespread deployment of AI technologies.

With respect to governance norms, it is essential, if they are to have their intended effect, that they account for human intervention in AI systems. They should, for example, require developers to specify clearly the points in the operation of a given system where human intervention is expected and the kind and level of expertise required of the individual making the intervention. With respect to the users or operators of AI systems, the norms should require that they have the skills and expertise expected at all the points at which human intervention is required. Only by bringing humans into the picture will norms intended to promote the safe, effective, and ethical use of AI systems achieve their goals.

With respect to practical readiness, it is important to recognize there are distinct skills required of operators of AI systems. As AI systems become more widely deployed, there will be a greater need for individuals with those skills. As societies seek to further their AI readiness, they would do well to consider methods for facilitating the informed adoption of a system. In the Law Committee, the example was cited of AI-enabled information retrieval systems used for legal discovery (fact-finding) in the United States. A key element in the informed adoption of such systems was the identification and use of easily comprehended measures of the effectiveness of such systems ("recall" and "precision"). Because the significance of the measures was understandable even to those with no familiarity with AI (good to be close to 100 percent; bad to be close to 0 percent), non-expert stakeholders (in this case, lawyers and judges, but also ordinary citizens involved in litigation) were able to assess the effectiveness of those systems, paving the way for their widespread adoption.

2.1.3 Metrics And Measurement

Measurement was another emergent theme in committee discussions. The ability to measure the effectiveness and impact of AI was viewed as important in engendering public trust and in enabling evidence-based decision-making both in public policy and in implementation-level standards and codes of practice.

With Respect To The Measurement of AI’s Effectiveness

In several committees, it was observed that the definition and use of meaningful, easily understood measures of the effectiveness of an AI system can be a key element in facilitating the informed adoption of a system. In the Law Committee, the example was cited of AI-enabled information retrieval systems used for legal discovery (fact-finding) in the United States. A key element advancing the adoption of such systems was the identification and use of easily comprehended measures of the effectiveness of such systems (“recall” and “precision”). Because the significance of the measures was understandable even to those with no familiarity with AI (good to be close to 100 percent; bad to be close to 0 percent), non-expert stakeholders (in this case, lawyers and judges, but also ordinary citizens involved in litigation) were able to assess the effectiveness of those systems, paving the way for their widespread adoption.

With Respect To The Measurement of AI’s Impact

In the Metrics of Well-Being Committee (Working Group 4, Committee 3), the focus was not on measuring the effectiveness of an AI system, but on measuring the aggregate impact of AI systems on the well-being of citizens. The limitations of Gross Domestic Product (GDP) as a proxy for well-being were discussed, and components of well-being overlooked by GDP were identified. The role of AI technologies and data analytics in gathering information on the factors that are overlooked by GDP but that contribute to well-being was also discussed. More work remains to be done on defining all the factors that would contribute to a fuller understanding of how established well-being metrics can best be utilized by AI technologies. However, the discussion highlighted the fact that, if we are to genuinely increase holistic prosperity for society, AI systems and practitioners must prioritize well-being metrics “Beyond GDP” in the design and distribution of what they build.

Measuring the quality of the results of an AI system and measuring the aggregate impact of AI systems on the well-being of citizens are two very different exercises. What they have in common, however, is the goal of grounding decision-making about the use of AI systems and the policies needed to guide their use on scientifically sound empirical measures. This emphasis on evidence-based decision-making is in keeping with the Roundtable’s overall approach to the development of policies and norms for the governance of AI.
2.2 Model Practices: Examples of Approaches To The Governance of AI

As noted above, one purpose of the Roundtable is to gather information about approaches to AI governance that have been proposed (and, in some cases, implemented) throughout the world. In this section, we look at five such approaches: regulatory sandboxes, data exchanges, the establishment of a digital identity, government-sponsored evaluations, and AI-enhanced government services.

2.2.1 Regulatory Sandboxes

A number of participants observed that one tool policymakers could employ to advance the safe and effective use of AI technologies was that of the regulatory sandbox: a narrowly defined zone in which a technology may be tested in a real-world environment, usually for a limited period of time, subject to rules that may be different from those that apply outside the testing zone.

Such sandboxes allow for closely monitored and controlled experimentation with new technologies in real-world conditions and thus provide information about the effectiveness of the technology and about the policy requirements needed to ensure its safe use. That information can then be used as a basis for decisions about whether to deploy the technology more widely and, if so, under what regulatory scheme. As an example of the use of such a regulatory sandbox, participants noted the government of China’s designation of four geographical districts to be zoned exclusively for autonomous vehicles.

2.2.2 Data Exchange

A data exchange proposal was the subject of much discussion in the Committee on Sectoral Development (Working Group 1, Committee 2). Under this proposal, a market would be created to match individuals willing to provide access to their data with vendors wishing to gain access to it. The exchange would be a transnational entity and would be subject to a strict regulatory regime. The potential benefits of such an exchange, for individuals, would be greater control over where individual data goes and the opportunity to monetize this data; for vendors, the potential benefits would be greater access to a more diverse population of data, thus accelerating development and innovation. Many practical and ethical questions would have to be answered before such an exchange could be created (e.g., how to put in place effective provisions to protect privacy and data security; how to protect the poor from exploitation). In the view of at least some participants, however, the proposal had sufficient merit to warrant further study.

2.2.3 Digital Identity

Estonia’s already-implemented program of digital identity was given some attention at the Roundtable. Under this program, every citizen has a state-issued digital identity, accompanied by a chip-enabled ID card. The digital identity enables rapid authentication of an individual’s identity without physical contact and improves access to government services. While such a state-issued digital identity will need to address valid concerns about security, privacy, and the potential for government abuse, its use in Estonia provides a real-world case for further study.

2.2.4 Government-Sponsored Evaluations

Some participants in the Roundtable observed that a potentially beneficial role a state can play in advancing the adoption of AI technologies is that of sponsoring scientifically sound evaluations of the effectiveness of those technologies. As an illustration, participants pointed to the series of evaluations sponsored by the U.S. Department of Commerce, and overseen by the National Institute of Standards and Technology, of the effectiveness of AI-enabled systems at meeting discovery (fact-finding) objectives in legal proceedings. The data and insights gained through these evaluations provided the crucial empirical support courts needed to give approval to the use of AI technologies in support of legal discovery. The empirical support, coupled with court approval, gave legal practitioners confidence they could defensively adopt the new technologies. As the technologies are more widely adopted, the legal profession is seeing increasing savings of cost and time in the conduct of legal discovery as well as improvements in the quality of the output. This model of state-sponsored evaluations of the application of AI technologies in real-world conditions is one that could be usefully replicated in many other domains.

2.2.5 AI-Enhanced Government Services

The benefits of bringing AI to basic government services were noted in a number of the breakout discussions. AI can improve citizens’ access to information about available government services, can help citizens identify and supply the information they need to provide in order to apply for and receive government services, and can allow greater and more rapid customization to the needs of a particular citizen or circumstance. Examples in which AI has been applied in improving government services include: assisting citizens with the procedures for formalizing a marriage (Estonia), helping to match citizens to appropriate job training and placement opportunities (Estonia), matching citizens seeking divorce to a lawyer appropriate to their circumstances (Denmark), and providing translation services for immigrants (Italy). Participants noted that AI technologies could also be used to improve the responsiveness of police and emergency personnel by predicting when and where they were likely to be needed (e.g., during a major public event, such as New Year’s Eve). Investment in bringing AI technologies to these sorts of government services could bring “quick wins” both for the state and for the citizens it serves.
2.3 Recommendations: Questions To Be Further Studied

The 2018 Roundtable was the inaugural event of what is to become an on-going initiative for the systematic development of policies for the governance of AI. As such, much of the focus of the 2018 Roundtable was on information gathering, taking inventory of policies already implemented, policies proposed, key issues to be addressed, and perspectives on those issues. Given this focus, the recommendations that emerged from the 2018 Roundtable take the form less of concrete policy recommendations and more of issues for further study. In this section, we summarize the key questions for further study.

Education

A recurrent theme throughout the day’s discussions was the need for more AI-centered education. Participants highlighted the need for both citizens and policymakers to be better educated about AI technologies; their capabilities and their limitations, as well as the inputs and skills needed to operate them effectively. Participants noted the need for new kinds of professional training if societies are to have a workforce capable of meeting the needs of an AI-enabled economy. What educational programs should be developed to promote better understanding of, and skills for, AI technologies?

Trust

A recurrent theme, related to that of education, was trust. In the current state of affairs, exaggerated claims (hype), misinformation, fear, and lack of trust all block a clear-eyed view of the capabilities and limitations of AI technologies and hence block an appropriate response to those technologies. What role can the state play in ensuring that AI-enabled systems, in particular in important societal domains (medicine, law, financial services, etc.) can be trusted by all institutions of society, as well as by ordinary citizens? And by what methods can it best fulfill that role?

Privacy

A recurrent theme was privacy and security of data. Participants raised concerns about increasingly invasive and irreversible privacy encroachment as the data requirements of AI systems increase and as firms continue to collect, combine, mine, and sell user data. The EU’s General Data Protection Regulation (GDPR) is one response to this concern. Are there other policies that will enable citizens to realize the benefits that AI systems offer without having to surrender their privacy in order to do so?

Competence

As noted above, most currently deployed AI systems are human-mediated. As such, the human’s role in enabling the successful operation of AI systems cannot be overlooked. What policies or implementation-level standards or codes of practice are needed to ensure that the humans involved in the operation of AI systems have the requisite skills and expertise and apply them correctly?

Source And Legitimacy of Norms

There is general consensus that norms for the governance of AI are needed, and must be regarded as legitimate. What are the sources of legitimacy for norms for the governance of AI? What are the shared values that could serve as the anchor for such norms? What is the appropriate framework for developing norms? What process of developing norms will provide legitimate results?

Empirical Evaluation

Over the course of the Roundtable, several discussions highlighted the value of empirical measures in addressing uncertainty and providing direction to policymakers. What role can the state play in promoting studies, evaluations, and analytical frameworks that are evidence-based?

Distribution

A number of participants drew attention to the fact that the availability and use of AI technologies was not evenly distributed around the globe or within societies. An effect of this uneven distribution could be to orient the design of AI systems toward their current user base and away from the segments currently without access, potentially leading to greater disparity over time. What role can the state play in promoting a wider distribution of access to AI technologies?

Bias

A number of committees noted the unequal access to and the flawed capturing of the data used to fuel the development of AI. Repeated oversampling of certain populations has created uneven representation and, in some cases, penalized certain populations around the world or within societies. In addition, algorithmic bias was recognized as an intricate topic, involving complex definitional issues, confusion around the nature, sources, and quality of bias (not all bias is necessarily bad), and even philosophical considerations. By what policies can governments promote greater clarity around the nature, sources, and effects of algorithmic bias? How do governments resolve the sometimes-competing goals of accuracy and avoidance of reinforcing disparities? What protocols might be implemented for the on-going monitoring and detection of algorithmic bias? How can governments promote greater clarity around the nature, sources, and effects of algorithmic bias? How can governments promote greater clarity around the nature, sources, and effects of algorithmic bias?
“The Fourth Industrial Revolution differs from previous ones in magnitude, velocity, and societal tolerance for sweeping change. It may also challenge our conception of what it means to be human.”

Flexibility
As noted earlier, a truly global and robust set of norms will be flexible enough to retain their validity across different cultures. Such a robust set of norms will also be flexible enough to retain their validity over time, as technologies, values, and the social mechanisms for responding to them change. By what means can the required flexibility be built into norms for the governance of AI?

2.4 Next Steps: Action Items For The 2019 Roundtable
The Global Governance of AI Roundtable is an ongoing initiative, systematically working toward meaningful and globally valid norms for the governance of AI. As such, the work of the 2018 Roundtable did not conclude with the February event; it continues as it lays the groundwork for the 2019 Roundtable. Next steps include the following:

- Circulate the current report to all 2018 participants and to other interested parties.
- Receive commentary from 2018 participants (and other interested parties) on the themes and issues discussed in the report.
- Develop an action plan for preparations for the 2019 Roundtable.

“*The system dynamics of the AI Revolution, within societies and across countries, are extraordinarily complex and unpredictable.*

Introductory Note To Summary of Committee Proceedings
Ai has widened the possible scope of outcomes for the way in which humanity develops. As AI technologies evolve, there is a wide range of potential paths for humanity, with both large scale upside and downside potential.

Our framework considers the co-evolution of technology and humanity in time space with three forces which influence each others: systems dynamics, values, and global governance.

**About The Model:**
System dynamics, values, and global governance affect and inform each other in a continuous feedback loop.

In system dynamics, technology raises new circumstances and enables new business models and social structures which fuel the re-definition of values (e.g. privacy or free-will). This in turn informs and guides advances of science, technology and business. Through this continuous socio-technical cycle, societies become more integrated and interdependent globally, leading to the emergence of regional, continental and global governance processes which need to be articulated one against each other.
Chapter 03

SUMMARY OF COMMITTEE PROCEEDINGS

The discussions in all the committees were lively and wide-ranging. While agendas established by the Roundtable guided the discussions, their observance was not so strict as to prevent participants from taking the discussions in the sometimes-unplanned directions that participants’ insights, creativity, and collective energy took them.

In this section, we provide summary reports for each committee discussion. These reports are not intended to be policy papers in which all uncertainty, disagreement, and inconsistency have been resolved. The reports are intended to serve rather as working notes, efforts to concisely memorialize the committee discussions for readers wishing to take a deeper dive in the proceedings on any given topic. Readers are invited to read all committee reports in sequence from first to last or to jump ahead to the specific committee discussions in which they have the greatest interest.

3.1 Working Group 1: Mapping AI and Global Best Practices, Reaping Benefits, Mitigating Risks

Working Group Coordinator: Professor Wonki Min, SUNY Korea.

Committee 1: AI Technology Development.
Committee Chair: Professor Jürgen Schmidhuber, University of Lugano.

Committee 2: Sectoral Development (Healthcare, Transportation, etc.).
Committee Chair: Rob McCargow, Program Leader of Artificial Intelligence, Technology & Investment at PwC.

Committee Chair: Leanne Fry, Acting Deputy CEO, Corporate, AUSTRAC.

Committee 4: Judicial System, Access to Justice, and the Practice of Law.
Committee Chair: Nicolas Economou, CEO, HS; Chair of the Science, Law and Society Initiative at The Future Society.

Working Group 1, Committee 1
AI Technology Development

Committee Chair: Jürgen Schmidhuber

Committee Outline
While AI can offer vast social benefits such as improved healthcare, sustainability, education, and public services, it also entails major risks that require precautionary action. Developers and industry players can help ensure responsible design and development by addressing issues of safety, collecting representative data, enabling greater explainability and transparency, and focusing on cyber security. What are the greatest challenges and global best practices to address these concerns?

The committee was asked to consider several topics in this arena, including the definition of safety in AI; the practical implementation of measures of safety during the development phase; the extent to which humans should remain “in the loop” to ensure safety and controllability; standards of trust and qualification for developers embedding safety measures; and practical policy suggestions to create safe and beneficial AI without curtailing innovation.

Topics Discussed By The Committee
The committee first considered questions of safety in AI development, particularly around creating definitional boundaries of what safety implies:

- Members explored a broad range of factors that would contribute to safe AI such as the reproducibility of technology, understanding whose safety a particular AI system is intended to protect (as in the case for military and combat applications), and the ability to predict and assess AI’s actions and outcomes as an indicator for safety.

- The ability to predict and assess the actions of AI was discussed as a two-step process of verification and validation: the first step addresses whether a particular AI system executed its task according to specification; the second step addresses whether the specification was correct. The latter raises complex philosophical and practical considerations of whether a certain goal was “good” and its underlying intent “right”. Some participants viewed it as difficult to operationalize because individuals and groups in society often have competing goals. It was observed, however, that societies have been able to set public policy objectives in spite of such competing goals.

- Transparency was viewed as an important enabling prerequisite for such verification and validation. Transparency, beyond facilitating human observation, can support the objective of determining whether an AI system meets its intended goal.
• Corrigibility—the ability to address incorrect or undesirable actions performed by AI systems—was viewed as an important corollary to the ability to predict and assess the actions of AI, including with respect to ensuring that machines, but also designers, developers, and operators of AI are accountable. It was observed that transparency, corrigibility, and accountability may rest on the ability to measure the effectiveness of AI systems in the real-world applications in which they are deployed.

• While Artificial General Intelligence (AGI)—if it were realized—would have profound implications beyond that of “narrow” AI, current AI systems entail significant safety challenges, as was illustrated in the use of social media in several national elections around the world.

The committee then discussed the question of value alignment.

• Value alignment was noted as a crucial component in creating safe AI systems. At the same time, the technical and theoretical difficulty of programming an AI machine to make decisions that align with human values was recognized. Some participants also raised the question of whether a common set of consensus moral norms on which to program AI could emerge.

• Some participants argued that an AI system does not need to have one set of universal goals and values. Instead, it could abide by different sets of values in different functional, social, or cultural contexts. At the same time, it was acknowledged that the dominant stature of certain social groups and cultures could amplify AI bias and induce conflicts.

• Some participants raised the possibility that future AI systems might become able to choose their own goals and values or learn the plasticity of human values. AGI systems, equipped with “artificial curiosity”, might develop entirely new value systems. Such capabilities might place AI systems at odds with humanity’s interests, or with national or international public policy objectives.

The Way Forward

The committee closed its proceedings by highlighting a range of desiderata, in support of the safe AI development. These are as follows.

• The privacy and dignity of each human being should be a paramount consideration in the development of AI. AI should be designed so as not to impede or undermine individuals’ decision-making authority over the use and sharing of their data.

• The risks and benefits of government-sponsored collective data lakes should be further assessed.

• The trade-offs involved in the sequencing of regulation should be further studied. (Should national governing bodies focus on domestic regulatory frameworks before, after, or in concert with internationally coordinated efforts?)

• Mechanisms for instantaneous corrigibility of AI systems in the event of deviation from the specified intent should be studied, including the potential for AI systems to instantly adapt to new statutes and regulations provided by trusted sources.

• Policy-makers and members of national and international legislative and regulatory bodies should take affirmative steps to enhance their awareness of the benefits, risks, and governance challenges presented by AI. (Organizations such as the IEEE can help support such endeavors.)

“Harnessing the power of Artificial Intelligence for the betterment of humanity requires an evidence-based approach and deliberate collective action.”
Working Group 1, Committee 2
Sectoral Development (Healthcare, Transportation, etc.)

Committee Chair: Rob McCargow

Committee Outline
The application of AI differs depending on sector, and with it, the role that government can play. For example, in the healthcare industry, data privacy remains an ongoing challenge. Companies seeking to maximize the benefits of industry-specific and cross-sectoral AI development; the need to break out of silos and foster trans-sectoral collaboration in data exchange; and the rapid growth of emerging technologies; the need to mitigate privacy risks in the healthcare industry; therefore issues of enforceability and ethics must be treated contextually within each sector.

The committee then discussed the challenges in establishing a global framework of AI governance. It first proposed that cultural issues at the state and community level are generally more impactful, and therefore more difficult, to resolve than establishing a global platform of universal values. It then discussed the possibly contradictory nature of establishing a global framework in a competitive market: some stakeholders in academia, non-profit, and government support open-data and sharing, while other stakeholders – private companies, government departments in security and enforcement - would prefer a “first mover” competitive advantage in developing their AI platforms. Those who create the global norms first could be incentivized to set the rules in their favor, thus gaining a competitive edge. There could be a power struggle between the “winners and the losers,” making the imposed regulations of a global framework restricting and challenging.

The committee debated issues of privacy within each sector, with strong opposing views. Some believed that privacy no longer exists, while others extolled the privacy protection measures taken in the European Union to protect citizens’ data. This led to a larger conversation on the nature of the development of AI and the inherent risks in imposing regulation too early on, especially in democratic countries. The tendency to over-regulate in matters of privacy and security could negatively impact growth, while similar constraints that do not exist in more authoritarian regimes would create a competitive edge for governments with access to citizens’ data.

Case Study: China
In China, the “Xue Liang” (“Sharp Eyes”) program seeks to use facial recognition and artificial intelligence to monitor the behavior and predict crimes in its populace of 1.4 billion. The program raises difficult questions on how vital, but sometimes competing, societal values such as public safety, the protection of private property, privacy, and dignity should be balanced. A case of using AI for behavioral recognition in Denmark was also discussed. The government is experimenting with predictive AI to match citizens with lawyers during divorce cases. In seeking to disrupt lengthy and often difficult legislative processes, the use of AI as “machine helping human” rather than “machine subverting human” was cited as a way in which AI can be used for public benefit.

Lastly, the committee discussed the need for a data charter. Citing the EU’s General Data Protection Regulation (GDPR), the committee debated the benefits and downsides of a charter that could mandate the free-flow of information among signatories, debating whether such an institution was too bureaucratic, or if it could exist in a virtual state for the quick dissemination of information.

The Way Forward
The committee closed its proceedings by suggesting that further dialogue is an essential component to supporting the sectoral development of AI. In particular:

- Future dialogue should involve all countries in Roundtable discussions, especially China, India, and Russia, in order to create a truly representative governing body.
- Solving local challenges at the city and state level remains the key to establishing a global framework, as the “bottom-up” approach to AI ensures a scalable and adaptable approach rather than a “top-down” imposition of a “one-size-fits-all” regulatory framework.
Customer service tools such as handover. Examples referenced include: considered in terms of a man-machine that good implementation should be avenues in which to deploy AI, noting The committee first identified imminent Committee Topics Considered By The Committee Chair: Leanne Fry Committee Outline Nested in digital transformations, AI has the potential to improve government and the delivery of public services by making processes more efficient and effective, and offering new and beneficial services to citizens. Radically enhanced planning, hyper-tailored education, demand prediction for utilities, autonomous transportation, and improved public health diagnostics represent select areas for greater public sector efficacy. Moreover, AI can reduce costs, natural resource consumption, time, errors, and waste in many administrative processes. What are the concrete strategies, cases, and best practices to help governments implement AI to deliver better services to citizens?

Topics Considered By The Committee
The committee first identified imminent avenues in which to deploy AI, noting that good implementation should be considered in terms of a man-machine handover. Examples referenced include:

- Customer service tools such as digital hotel or intercity concierge and chatbots increase staff productivity and customer satisfaction. For example, Go Moment is a Los-Angeles based Active Customer Engagement tool that provides real-time feedback for hotels, driving profitability and efficiency. Also, using AI for behavioral analytics can provide national services such as police and ambulance predictions on eventual casualties and crime.

The committee then reviewed public sector best practices for civic engagement on AI. The propagation of information through a Data Exchange platform can facilitate transparency and trust, and break the divide between government and its citizens. It could empower citizens to participate in creating solutions. Examples of potential scenarios referenced include hackathons that use government data to focus on people’s happiness and satisfaction, and start-up belt initiatives that provide data in order to help entrepreneurs start a business.

The committee suggested the public sector alter the education systems to adapt to emerging technologies. Introducing basics of AI and machine learning early in the curriculum will balance access to information, certify the public, and attract the right people to participate in positions of power. The committee acknowledged a need to reform the education structure where a reward is a degree, as it poses a barrier for non-traditional expertise. The identification of experts should change because degrees can narrow access of information and diversity of thought, and overlook the value of short courses and self-taught experts. Suggestions of an expiry date on degrees to drive continuity learning and knowledge intake were noted. AI knowledge hubs should include social sciences for an understanding of the societal implications of AI.

The committee then considered best practices for governments to manage and prepare for changes associated with AI. Government would benefit from a partnership with academia to ensure cutting-edge change and informed decision-making in relation to regulation. Educating government officials to have competence in AI builds trust from citizens. Governments can direct AI research towards an application that serves the public good through incentives, subsidies, grants, or prioritizing funding for AI research.

The committee also considered trusting academia to spearhead the beneficial development of AI. For example:

- Singapore allocates funds to researchers to identify common traits and frameworks in technology companies. Selection is based on groupings of common incentives rather than individual projects.
- Slovenia has achieved many successes in the field of AI and Big Data in the public sector, with the largest number of AI and robotics researchers per capita in the world (approximately one researcher per 7,000 inhabitants).

The Way Forward
The committee closed its proceedings by suggesting that:

- Government can play a role in proactively supporting beneficial innovation through policy and digital infrastructure. Government can create regulatory sandboxes and special zones for experimenting with AI applications.
- Governments should support endeavors and mechanisms designed to assess and quantify the risks and benefits of AI innovation, so as to further enable evidence-based public policy decision-making.

“Sound evidence of the real-world effectiveness of AI is critical. Test-beds envisioned in the UAE’s ‘UAE’ strategy and the work of measurement laboratories around the world, such as US NIST, can offer such evidence.”
Committee Chair: Nicolas Economou

Committee Outline

To what extent should societies delegate to artificial intelligence decisions that affect citizens? What is the role of the law and the justice system in seeing that our response to this question is applied in practice? How does our response to these two questions impact fundamental assumptions about the judicial system, access to justice, and the practice of law? AI holds the potential to improve access to justice for all citizens, and to advance the impartial, effective and speedy adjudication of justice. But it also brings the risk that the values that animate our legal systems, such as citizen participation, transparency, freedom from bias, dignity, privacy, and liberty will be lost in favor of practical objectives such as efficiency.

Recognizing the potential benefits and the potential risks presented by AI, the committee examined these questions with both a top-down and bottom-up approach. The top-down approach invited identification and conversation of emerging normative approaches, including consideration of general principles for the adoption of AI in the Law, taking into consideration the work of governmental and non-governmental organizations, think-tanks and standard setting bodies. The bottom-up approach invited identification and discussion of real-world implementations of artificial intelligence to the automation of legal decision-making, and sought to identify, from these implementations, emerging model best-practices that may serve as the basis for guidelines, codes of practice, and implementation-level standards.

Topics Considered By The Committee

The committee first reviewed the domains in which artificial intelligence is applied today, from judicial decision-making to electronic discovery (fact-finding). The committee then considered the benefits and risks of artificial intelligence to the functioning of the legal system.

• Salient potential benefits were deemed to be enhanced access to justice; acceleration of judicial decision-making; elimination of bias; uniformity of outcomes; ability to review and correct judicial errors; improvement of public confidence in the legal system; and democratization of legal expertise.

• Salient potential risks were deemed to be opaque decision-making; intentional or unintentional biases and abuse of power; emergence of non-traditional bad actors; sustaining inequality; reduction in public trust in the legal system; and an inadequacy of human capital active in the judicial system to manage and use AI. Also, a sacrifice of the spirit of the law in order to achieve the expediency that the letter of the law allows; the unanticipated consequences of the surrender of human agency to non-ethical agents; the loss of privacy and dignity; and the erosion of democratic institutions.

The committee then reviewed considerations that appear to be most overlooked in the current public dialogue. Salient among those cited were the competencies needed in order to effectively operate and measure the efficacy of AI systems in legal functions that affect the rights and liberty of citizens. In this context, reference was made to:

• The State v. Loomis matter in the United States, in which a man in Wisconsin was sentenced to a long prison sentence in part because an undisclosed algorithm deemed the defendant to be at high risk of recidivism. The topic of whether the judge had the requisite competencies to understand the performance and limitations of the algorithmic decision-making he partly relied on was discussed.

The question of whether the judge adequately balanced the competing societal values involved (protection of intellectual property rights vs. the interests of the individual) was also discussed.

• Electronic discovery, which typically involves an interwoven continuum of AI and HI (human intelligence), but where no standards of competence exist for the operator of the artificial intelligence.

The question was raised as to whether standards of scientific competence and credentialing were necessary for the operation and the assessment of the efficacy of AI in critical functions in the legal system, such as judicial decision-making and fact-finding, domains where AI is already in substantial use today.

The committee then considered emerging governance frameworks for artificial intelligence. Among those referenced in discussion were:

• The Future Society: an analytical model with four sequenced layers, general to specific:
  • Social; Legal (norms, regulations, legislation)
  • Technical controls
  • Implementation-level instruments (standards, codes of practice, technical controls)
  • Ethical (criteria and principles)

• Gasser & Almeida, Harvard University: an analytical model with three interacting layers:
  • Social; Legal (norms, regulations, legislation)
  • Technical foundations (data governance; algorithm accountability; standards)
  • General principles (ethics)
The Committee then considered emerging ethical principles for the governance of artificial intelligence. Reference was made to the working versions of those promulgated by:

- The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems’ General Principles:
  - Well-being: Prioritize metrics of well-being in the design and use of A/IS systems.
  - Accountability: Ensure that designers and operators of A/IS systems are responsible and accountable.
  - Transparency: Ensure A/IS systems operate in a transparent manner.
  - Awareness of misuse: Minimize the risks of misuse of A/IS systems.

- The Future Society:
  - AI shall not impair, and, where possible, shall advance the equality in rights, dignity, and freedom to flourish of all humans.
  - AI shall be transparent.
  - Manufacturers and operators of AI shall be accountable.
  - AI’s effectiveness shall be measurable in the real-world applications for which it is intended.
  - Operators of AI systems shall have appropriate competencies.
  - The norms of delegation of decisions to AI systems shall be codified through thoughtful, inclusive dialogue with civil society.

The committee then considered emerging best practices for the governance of artificial intelligence. The examples referenced included electronic discovery in the United States, where AI has been deployed since the early 2000s to conduct automated fact-finding in vast amounts of data such as e-mails and corporate document repositories. The model best practice cited was the progressive adoption by U.S. courts of the easy to understand, yet scientifically sound metrics of Precision and Recall. These metrics answer, on a scale of 0 to 100 percent, the question of the extent to which the evidence sought by an AI-enabled system was actually found (thus offering a gauge of effectiveness that stakeholders in the legal system can understand and assess).

The committee acknowledged both the commonality and differences of certain values among cultures in attempting to design common frameworks of governance. In particular, the European view emphasized the need for transparency into the algorithms themselves, as well as the potential need for public-private partnerships in deploying AI in critical societal legal tasks, such as judicial decision-making or law enforcement. Among the international perspectives discussed were South Korea’s Robot Act and Estonia’s proposed laws on governing robotics.

---

The Way Forward

The committee closed its proceedings by recognizing the need for continued dialogue, in the context of the Global Governance of AI Roundtable and beyond, focused specifically on achieving actionable recommendations for AI governance in the world’s extended legal systems, including:

- Achieve a consensus framework of governance that can be operationalized.
- Enhance the diversity of representation required in order to produce a framework that will have, and be perceived to have, legitimacy.
- Further the sustained process of collective intelligence required in order to produce such a framework, with public policy recommendations as well as practical guidelines, codes of practice, and technical safeguards.

“For AI to advance human well-being, within and beyond metrics of GDP, an effective, legitimate, agile governance framework is indispensable.”
3.2 Working Group 2: The Governance of AI

Working Group Coordinator: Wendell Wallach, Consultant, Ethicist and Scholar at Yale University’s Interdisciplinary Center for Bioethics

Committee Chair: Nozha Boujemaa, Research Director at the French Institute for Research in Computer Science and Automation (INRIA), Director of the DATAIA Institute (Data Sciences, Intelligence & Society), Professor, University of Louisville

Committee 1: Empowering Innovation, Mitigating Risks: Norms, Metrics, Codes of Practice
Committee Chair: Nozha Boujemaa

Committee Outline
As innovation continues to drive the development of AI, there is a need to ensure this rapidly evolving technology does not irreversibly encroach on citizens’ rights. How can policymakers further encourage socially positive innovation that upholds intangible values such as dignity and freedom without stifling technological development?

The committee was asked to examine a number of topics in this arena, including the alignment and adoption of universal norms and standards; the ability to balance fair access to data while mitigating privacy risks; and the use of market-oriented tools to incentivize AI development that can promote equality, diversity, and well-being.

Topics Discussed By The Committee
The committee first discussed defining and aligning universal standards for global governance. The question of effectiveness was raised, given that mitigating risk is often contextual. In this regard, some participants raised the concern that a universal code of ethics, in order to encourage broad adoption, would have to be vague, potentially rendering it ineffective.

The committee then considered that technical accountability might be a helpful element in developing a set of global norms. For example, cybersecurity standards are globally recognized due to their technical provability, which in turn confers a measure of accountability and integrity.

Agreed-upon codes of behavior, standards, and sound metrics might create a common basis for the governance of AI, engender societal trust, and produce broad international adoption. Furthermore, such codes of behavior, standards, and metrics can provide governments and civil-society alike with evidence-based instruments to reward beneficial innovation (as it may relate to public policy or other objectives), and thus reduce potentially detrimental calls to control or stifle innovation in the first place.

The committee discussed several factors that might support the emergence of a governance framework that can be operationally effective and trusted by civil society.

1. Any governance framework, in order to be effectively operationalized, needs to allow for mechanisms capable of deploying certain professional competencies (emanating from technical fields, but also social sciences, law, and the humanities), identifying or devising technical standards, and using scientifically sound metrics. Such competencies and instruments permit the reliable measurement of the effectiveness and impact of AI, so as to enable both accountability and evidence-based public-policy decision-making. The committee surveyed various approaches to operationalizing governance, such as the concept of a new international body, equipped with the appropriate instruments and expertise, or the concept of an IPCC-like initiative.

2. Advocate a better approach to the collection of and access to data to ensure privacy, the protection of citizens’ rights, and to mitigate adverse algorithmic bias. The overarching framework by which government may collect data or protects the rights and dignity of its citizens is a saliently important area of inquiry, in particular in light of the data-collection programs launched by China and India.

Some participants noted that monopolistic firms also present a challenge to data privacy. As global technological giants continue to acquire smaller but data-intensive companies, there is a risk of data accumulation in private firms with little public oversight or control. Yet regulating this data also presents an economic challenge, given that many firms rely on the use of this data to provide free services to sometimes billions of users (many of whom...
might not be able to afford to pay for such services). In this respect, the committee considered the idea of applying general competition laws within a defined “data market” in order to regulate the free-flow of information. Redefining the idea of anti-trust as a fragmentation of the data market might prove productive, given that consumers derive great benefits from platform technologies. The committee then considered options to enhance data protection within this market, such as paying for a no-ad Facebook. (The risk of creating a digital divide between those who would be able to buy privacy and those who could not afford to do so was noted.)

Case Study – Estonia

Estonia established a common data strategy to support data portability in 2001, as a way of mitigating data privacy risk. Estonia’s mydata.org is a legally mandated national data platform that seeks to advance “human-centric” data management, which includes provisions such as allowing citizens to granularly assign access rights to the doctors of their choosing, and also receive a record of law enforcement officials who access their data. Such measures have not only placed citizens in control of their data, but have also created a highly flexible, adaptive, and customer-centric digital e-government that does not undermine innovation and entrepreneurship.

Lastly, the committee considered the ongoing challenge of algorithmic bias within data collection and analysis. Some participants expressed the view that because human bias is difficult to remove from AI, the focus should be on “harm minimization” instead of “harm elimination.” In this regard, data fragmentation was discussed as a means to help minimize the risk of bias. Additionally, the committee noted that it is important to create inclusive conversations that are truly representative of societal demographics. In addition, it was observed that rules against discrimination and unfair market practices that exist for non-AI activities, could be beneficially used as part of the instruments deployed in addressing bias.

The committee then considered past global conventions as potential models for a collaborative framework. The UN’s 2017 treaty on the prohibition of nuclear weapons, the Chemical Weapons Convention, and the Geneva Conventions were discussed. Yet the challenge with AI remains its intangibility: as it is not yet a physical weapon, it is not clear whether a “Geneva Convention for Autonomous Weapons” could realistically be enforced. Analogous treaties on cyber warfare were also discussed, including the possibility that AGI could become an impenetrable system incapable of being regulated without a human-in-the-loop interface.

The committee concluded with a discussion of how public policy could be supported by standards and metrics that would enable the comparative assessment of the efficacy and impact of AI systems against non-AI-enabled alternatives (or the prevailing status quo).

The Way Forward

The committee closed its proceedings by suggesting several practices that should be further considered in mitigating AI-related risks.

- Develop (or identify when they are in existence) technical accountability standards, metrics, and credentials, in order to help operationalize public policy objectives and the societal and human values those reflect.
- Investigate data fragmentation and the creation of data markets as ways to put privacy back into the hands of citizens while maintaining the free flow of information.
- Create AI applications that require a measure of human intervention (keep “humans in the loop”) in order to mitigate risk as the technology increasingly approaches autonomy.
- Further develop codes of practice or credentials to ensure that the “human in the loop” has the appropriate competencies to perform the requisite tasks, from AI design to operation and quality assurance.
- Investigate data fragmentation and the creation of data markets as ways to put privacy back into the hands of citizens while maintaining the free flow of information.
- Create AI applications that require a measure of human intervention (keep “humans in the loop”) in order to mitigate risk as the technology increasingly approaches autonomy.
- Further develop codes of practice or credentials to ensure that the “human in the loop” has the appropriate competencies to perform the requisite tasks, from AI design to operation and quality assurance.
Working Group 2, Committee 2

Government Objectives and Strategies

Committee Chair: Antonio Samaritani

Committee Outline

Innovation lies at the heart of technological development, and yet can pose concerns for governments seeking to balance rapid advancements in technology with outcomes that optimize the public good. What models of governance can be adopted to encourage innovation while protecting citizens? The committee was asked to examine a number of topics in this arena, including elements of the innovation supply chain that could have constraint mechanisms imposed upon them: the critical factors that can strike a balance between enabling and regulating innovation, including regulatory sandbox; pilots, test beds and public-private partnerships; the constraints around industrialized research; sequencing regulation to support the development of AI; and the specific sectors that require regulation.

Topics Discussed By The Committee

The committee first debated the role of government in constraining or enabling innovation for the public good. Topics such as data privacy laws, consumer rights, preventing inequality, and access to data for a full range of firms and people were discussed. Some brought up the point that current sectoral regulations are imposed because of significant risks to human health and safety - in the areas of environment or finance, for example. Others noted the need to re-define the role of regulation, given that AI is not just about ensuring safety, but about promoting fairness and inclusivity.

The committee then discussed the challenge of policymakers lagging behind industry in innovation, and the cultural difference this problem entails. Entrepreneurs and major technology companies are in the business of producing innovation, sometimes of the disruptive kind, to which regulatory frameworks tend to not be well adapted. The legislative and policy-making cycle, by contrast, in particular in democratic societies, is deliberative by design and thus slow. The public sector more broadly, because of its fiduciary duties towards all members of society, tends to be far more cautious than, for example, Silicon Valley entrepreneurs. This lack of cultural alignment was perceived as a challenge both for policy-making and in the development of successful public and private partnerships.

The cultural challenge for policymakers seeking to regulate technical advancement was perceived by some participants as being rendered more difficult by the opacity that Big Tech firms operate under, both on intellectual property grounds and for fear of public scrutiny. These elements discourage larger firms from being transparent about their AI development. The committee then suggested the possibility of mitigating reputational damage by using subject matter experts to test newer, riskier technologies. This would give smaller firms the advantage of surviving in a monopolistic field while larger firms would minimize their public exposure.

The committee then discussed the challenges that arise in collecting accurate, representative data. Members expressed concern that the data feeding AI applications does not accurately represent many populations, since data collection can often exclude large segments of society. This has led, for example, to some algorithms penalizing certain languages in multi-lingual societies. This bias has tremendous consequence for diverse countries such as the UAE, which are highly multicultural and multilingual.

A number of further questions were raised, including how to ensure the transparency of data within companies; the ability for companies to ensure that data is representative through external or internal regulation; how to regulate technologies or data that are produced in another jurisdiction; and the difficulty of explainability and the responsibility of AI within certain sectors, such as transportation.

The Way Forward

The committee closed its proceedings by suggesting a number of ideas for further study for governance strategies and objectives to the questions raised, including:

- Promote a regulatory framework that is flexible, evidence- and values-based, and gradually sequenced to keep pace with the rapid advancements of AI. Adopting a model based on principles that promote the public good and evidence-based decision-making, rather than a rigid set of rules is key to creating this flexibility, especially in public-private partnerships (PPPs), or in collaborations where incentives and goals may be misaligned.

- Develop incentives for self-regulation within the technology industry in order to promote ethical and safe business practices. Creating a competitive market for reputational credibility and establishing a set of industry best practices might be effective complements to government intervention or international norms.

- Explore the idea of exporting regulations - say in the case of self-driving cars - to solve the problem of regulating technology that is imported from one jurisdiction to another. Such “scalable regulation” would be harmonized with international standards and guidelines.

- Support further research by neutral institutions who can focus both on developing standards and sound metrics, and on promoting the public good, without raising questions of legitimacy or partisanship.
Committee Outline

In the race to develop AI, it is difficult to place boundaries on safety given the current environment of “winner-take-all” in which developers can move to less-regulated countries or cities if they are found to be in breach of privacy and data regulation rules. How can we achieve global cooperation and governance to usher safe and beneficial AI given the incentives towards unconstrained development and deployment in a globally competitive landscape?

Topics Discussed By The Committee

The committee acknowledged that a salient obstacle to international cooperation on the governance of artificial intelligence is the fast rate of technological innovation, which outpaces policy-making and regulatory capabilities, at both the national and international level. The committee also recognized some of the fears and mistrust that the public associates with AI, in part because of its image in popular culture, and in part because of insufficient mechanisms on which evidence-based public trust could be engendered.

- Encourage the involvement of a diverse set of actors in addressing major challenges such as bias in data, which can produce superior and more legitimate outcomes with respect to fairness, accountability, and transparency. It can also potentially drive implementation assistance.

- Emphasize government-led technical support, such as regulatory sandboxes or the funding of initiatives that support sound empirical evidence to support decision-making.

- Encourage governments to support critical IT infrastructure, such as secure data storage.

- Promote greater cooperation between governments and the private sector through talent exchange and other mechanisms to promote knowledge sharing.

- Consider the potential of risk-sharing solutions that may mitigate the reputational risk of failure inherent to the AI innovation process.

“An AI governance framework, in order to be effective, must produce general principles, public policy consensus, and evidence-based implementation-level standards and codes of practice.”
Participants proposed mechanisms to coordinate different national activities for governance of artificial intelligence.

• Creation of an international body for comprehensive oversight, either new or embedded in a global organization such as the United Nations, where information is aggregated. Its role would be to engage in loose coordination with all AI stakeholders, underscore best practices and competing standards and metrics, monitor developments in the field of AI, and ensure challenges are addressed.

• Establish a global infrastructure of soft governance solutions for responsible technological development without inhibiting innovation, such as international industry standards and practical insurance frameworks. Simultaneously, national entities and governments could apply enforcement procedures and consider penalties for developments detrimental to society.

The committee then considered the question of the legitimacy of regulatory guidelines. The vital importance of inclusivity, in order to achieve broad international consensus, was acknowledged.

The committee discussed the dynamics of establishing a treaty to implement standards of responsible principles, and guidelines. (It was noted that the Japanese government and the Organization for Economic Co-operation and Development (OECD) are supportive of guidelines for AI research.) The committee consensus was that that a treaty may be a premature and unrealistic objective, absent certain steps precedent. Among those considered were the following.

• Establish a predictive timeline on the development of artificial intelligence and its potential effects, to help the international community assess anticipatory regulation.

• Seek to identify and document the sources and drivers of technological development.

• Develop the objectives, principles, and guidelines that may drive such international cooperation.

• Systematically document the positions and public-policy considerations of each government with respect to salient issues.

• Obtain public buy-in from the public and private sector, as well as civil society at large, as was the case with climate change. International cooperation in climate change was discussed as an example of a collective intelligence process, which involved a broad dialogue between experts, governments, and civil society, and resulted in a strong impetus for cooperation.

The Way Forward
The committee closed its proceedings by suggesting that its efforts focus on the development of a predictive timeline on the development of AI, drawing in part on parallels that may be drawn from prior waves of transforming technologies.

Working Group 2, Committee 4
Governance of the Transition from Artificial Narrow to Artificial General Intelligence

Committee Chair: Stuart Russell

Committee Outline
Artificial general intelligence (AGI), or AI with the full range of human abilities, can unlock unimagined benefits but also poses existential risk to humanity. If built using the standard methodology of optimizing a fixed objective, an AGI is incentivized to acquire unbounded resources to complete its objective and to disable its off-switch. Lacking full awareness of the balance of human values, an AGI can compromise humanity to achieve its aims. Moreover, the take-off period when an AGI can upgrade itself to super-human intelligence, after which it may be irreversibly outside of human control, is uncertain. Control and value-loading of human preferences are currently difficult to develop, and with a winner-takes-all market domination race, developers are not incentivized to embed safety and control precautions.

Although the AI research community has taken a proactive stance toward self-governance, global government coordination is necessary to ensure that developers manage the norms and guidelines with a more balanced interest and representation.

The committee was asked to examine a number of topics in this arena, including the lessons learned from history; issues in regulating hardware and software; and how to develop an AGI that is beneficial for humans.

Topics Discussed By The Committee
In seeking helpful precedent in assessing the governance of AGI, the committee discussed weapons of mass destruction. Salient views expressed by participants were as follows.

• In certain instances, the governance of nuclear weapons can be considered to have been successful. For example, the ban on testing curbed their development; arms reduction treaties were agreed; some governments agreed to relinquish their weapons altogether.

• Additionally, the advent of nuclear weapons helped shape a global moral norm: the use of nuclear weaponry is a powerful taboo around the world.

• That an industry should take ownership of the risks that its technology poses, rather than
The Way Forward
Committee participants closed the discussion by proposing practical policy suggestions for further consideration:

The deliberate (and potentially existentially threatening) misuse of AGI can be viewed as a magnified version of the challenge of controlling malware. Emphasis on addressing the threats related to malware might help inform a more broadly applicable technical and governance framework.

The concept of an internationally agreed principle on making AGI available to all nations, once achieved by any nation, should be discussed. It would be important to initiate this dialogue with actors from leading AI actors in the international community, including China.

Important precedent, such as the history of global climate change, should be considered.

Culture, imaginaries and the arts play an important role for informing citizens and could be serve beneficially in both informing citizens and obtain public support for governance.

The dialogue around governance would benefit from being centrally framed around the imperative to maximize the enormous potential benefits of AGI, while minimizing its risks.

engaging in a culture of denial, was identified as a critical to the fostering safe innovation. It was argued that the nuclear industry largely terminated itself by not owning up to the full risks of this technology.

- However, the committee also discussed examples where nuclear technology does not provide a suitable parallel to AGI development. For example, the expense of building nuclear weapons acted as a significant barrier to proliferation, which is not the case for AI.

- Some members expressed the view that, in some instances, nuclear war could be said to have been avoided by “sheer luck,” such that the governance of nuclear weaponry, while a useful example, was at risk of providing a false sense of comfort.

The committee then discussed issues of regulation pertaining to hardware as compared with software.

- It was agreed that, in order to enable corrigibility and accountability, AI should provide an adequate measure of transparency and be open to inspection.

- Hardware should always have an in-built “off-switch” to create controllability.

- One participant suggested that the international community would benefit from agreeing that AGI technology, once realized by any actor or nation, should be shared with all nations, in order provide broad access, enable substantial control, and mitigate catastrophic risk. Having such an agreement would reduce the incentive to engage in unsafe races.

- The committee also discussed whether it would be possible for humans to control AGI, a technology potentially extraordinary intelligence, capable of becoming free of any human control. In this respect, the committee discussed theoretical technical proposals that would prevent such impulses and eliminate incentive in this respect.

The committee also examined the question of how beneficial AGI might be developed.

- Industry should be encouraged to acknowledge and feel responsible and accountable for the downside potential to AGI development.

- Knowledge sharing between industry and policy makers should be encouraged.

- Progress from Artificial Narrow Intelligence (ANI) to AGI should be monitored closely, in order to manage the emerging characteristics of AGI, both positive and possibly malicious.
SUMMARY OF COMMITTEE PROCEEDINGS

3.3 Working Group 3: Impact on Cross-Sectoral Dynamics

Working Group Coordinator:
Professor Subbarao Kambhampati, Arizona State University

Committee 1: Impact on Workforce and Mitigation Strategies
Committee Chair: Calum Chace, Author of The Economic Singularity: Artificial Intelligence and the Death of Capitalism

Committee 2: Cybersecurity
Committee Chair: Robert Silvers, Former Assistant Secretary for Cyber Policy at the U.S. Department of Homeland Security

Committee 3: Ethics and Code of Practice (Privacy, Transparency, Accountability, etc.)
Committee Chair: Jaan Tallinn, Founder of the Centre for the Study of Existential Risk

Working Group 3, Committee 1
Impact on Workforce and Mitigation Strategies

Committee Chair: Calum Chace

Committee Outline
AI-enabled machines are becoming increasingly capable of undertaking tasks previously performed by humans. Workforce automation is already occurring and is predicted to increase across industries and up the skills ladder. Unlike past waves of technological automation, machines equipped with AI are increasingly capable of performing highly skilled cognitive and creative tasks. Jobs that previously required decades of education, such as medical diagnostics, music composition and legal research, are now the purview of AI. Even “human” jobs such as therapy and elderly care are witnessing this algorithmic encroachment.

Despite conflicting projections of the impact on employment, how can policymakers prepare society for a potential “unnecessary,” or mass unemployment?

The committee was asked to examine a number of topics in this arena, including if AI will actually lead to technological unemployment; which geographic or demographic sub-segments may be disproportionately affected by automation; what policy responses can manage inclusivity; how policymakers can prepare society for potential large-scale job losses and what economic, social, and educational policies should be prepared now versus 5-10 years from now; which conditions make redistributive policies, such as a UBI or negative income tax, more or less relevant, feasible, and beneficial to society; which societal objectives should policymakers safeguard in a possible transition to a “workless society”; and what standard of living, if any, should government ensure for the “unemployable.”

Topics Discussed By The Committee
The committee first discussed current views on the future of employment in an AI-driven world. Members were divided; most did not fear large-scale technological unemployment, namely because new job creation is hard to predict, and the capacity of humans to re-skill is a critical but necessary function of human development. Humans can and must adapt to new technologies and tasks continuously and quickly in anticipation of a “brand new world” every two years. A significant minority did think that cognitive automation will lead to a widespread, lasting loss of employability, but that it will take years to arrive, and the outcome can be wonderful if the transition is managed well.

The committee then debated the benefits and drawbacks of automation. Those who believe that automation is inevitable noted that humans would need to find a way for people to still create value, given that AI is now able to perform cognitive and creative functions, such as art or therapy. Some worried that the number of new jobs brought about by advancements in technology would be small. For example, the number of people employed as YouTube bloggers is miniscule compared to workforces in manufacturing, transportation, and other labor-intensive jobs that have existed since the Industrial Revolution.

Others within the committee argued that automation is not inevitable, but is merely a part of the “creative destruction” cycles that revolutionize economies every so often. They pointed to the increased prevalence of the short-term contract, or “gig economy,” and the inherent need for humans to create newer and better things, which subsequently leads to the creation of new jobs, companies, and products. Some argued that education will keep pace with technological development; others hypothesized that humans will create human-machine complementarities; and others promoted the idea that humans will have jobs taking care of one another.

The committee then discussed the challenge of unemployment, suggesting that there is a need to shift the paradigm of employability. Are jobs required for meaning and purpose? The independently wealthy or retired may yield lessons on creating meaningful life beyond work; demonstrating that a work-free life can produce positive opportunities to play, socialize, learn, and explore – as well as negative opportunities for unhealthy addictions.
The committee considered the role of a Universal Basic Income (UBI). Some noted it could act as a good redistribution mechanism, while others postulated there would be no need for money in the world of AI, as machines producing “radical abundance” could lead to massive price drops. Some noted that taxing machines to facilitate UBI would kill innovation, while others suggested implementing a negative income tax as an alternative means of redistribution, though this would assume people have jobs. Universal basic services was also proposed as a UBI alternative, as well as valuing other types of investment, such as time and purpose-driven activities, such as volunteerism.

The committee also considered the philosophical ramifications of a job-free future. Some noted that the wealthy, fearing revolution, reputational damage, or simply wanting to give back (i.e., current billionaires giving away their wealth), would prefer to distribute their gains. Others postulated that a labor-free future could cause a spiritual revolution or crisis, given that humans still have the desire to lead productive lives.

**The Way Forward**

The committee closed its proceedings by suggesting further study of the questions raised, including:

- The need for education, training, and occupational redesign to prepare people to be adaptable and flexible as jobs begin to automate.
- The need to value non-monetary time, such as volunteerism, and develop new metrics for measuring well-being.
- The need to mitigate the risks of populist politics. The loss of jobs - or the fear of it - by a critical mass of the population has the potential to incite populist regimes that will tend towards nationalism and protectionism.

**Working Group 3, Committee 2**

Cybersecurity

**Committee Chair: Robert Silvers**

**Committee Outline**

While AI provides relevant and powerful solutions, it also introduces new vulnerabilities to cybersecurity that can be exploited in malicious contexts. This is due to the fact that deep neural networks are “black boxes,” the granular decision-making process of which is not understood even by those who develop and operate them. AI can thus greatly improve cybersecurity through early detection of advanced threats, but also lends itself to new forms of digital insecurity by opening the way to new forms of attacks specific to machine-learning processes.

The committee was asked to examine a number of topics in this arena, including:

- The need for a Universal Basic Income (UBI). Some noted it could act as a good redistribution mechanism, while others postulated there would be no need for money in the world of AI, as machines producing “radical abundance” could lead to massive price drops. Some noted that taxing machines to facilitate UBI would kill innovation, while others suggested implementing a negative income tax as an alternative means of redistribution, though this would assume people have jobs. Universal basic services was also proposed as a UBI alternative, as well as valuing other types of investment, such as time and purpose-driven activities, such as volunteerism.

The committee also considered the philosophical ramifications of a job-free future. Some noted that the wealthy, fearing revolution, reputational damage, or simply wanting to give back (i.e., current billionaires giving away their wealth), would prefer to distribute their gains. Others postulated that a labor-free future could cause a spiritual revolution or crisis, given that humans still have the desire to lead productive lives.

**The Way Forward**

The committee closed its proceedings by suggesting further study of the questions raised, including:

- The need for education, training, and occupational redesign to prepare people to be adaptable and flexible as jobs begin to automate.
- The need to value non-monetary time, such as volunteerism, and develop new metrics for measuring well-being.
- The need to mitigate the risks of populist politics. The loss of jobs - or the fear of it - by a critical mass of the population has the potential to incite populist regimes that will tend towards nationalism and protectionism.

**The Committee Then Discussed**

managing the weaponization of AI in cyber warfare. On the one hand, there are no guardrails for AI. It’s democratized; actions are not clearly attributable to specific actors. Advanced forms of AI could potentially automate attacks and take humans out of the loop. However, the positive aspect of this also lies in automation: AI could detect and stop a threat on its own without human intervention. This checks-and-balances system could prove useful in disrupting asymmetrical warfare. Currently, sophisticated cyber attackers are unable to be stopped, creating an offensive superiority that promotes instability. AI machine learning could improve defenses, ostensibly lowering this asymmetry.
Yet such countermeasures also present a challenge, given that deterrence in cyber-warfare reveals a measure of capabilities, which risk diminishing effectiveness. How does a nation deter an attack without rendering its methods obsolete? The committee considered the ongoing challenge of talent retention in government cybersecurity. How can governments recruit the necessary technical talent when the best salaries are in the private sector? One possible solution might be to adopt a public-private partnership model that allows civilians to work with defense forces in countering cyber threats, similar to Estonia’s voluntary Cyber Defense Unit, which is made up of average citizens outside of government who are specialists in key cyber-security positions. Such collaborations could also help to counter the rise of disinformation campaigns aimed at eroding trust in institutions, and help to “check” AI programs that could potentially be corrupted without a human-in-the-loop interface.

The committee also discussed the importance of education in combatting disinformation and increasing the political leadership’s awareness of cyber threats. Most educational systems aren’t keeping up to speed with the rate of disinformation, necessitating an open education platform and governance. Attracting youth to roundtables such as this one was also raised: cyber-capable youth who care about the public good should be at the policy table to brainstorm creative solutions. The committee also questioned the cyber literacy of the general political leadership, citing examples of blatant misuse or misunderstanding of rudimentary technological functions. The group agreed that leaders must be fully aware of technology at every level to be able to craft proper policy.

The Way Forward
The committee closed its proceedings by suggesting two ways in which to shore up Cybersecurity as AI evolves:

- Create a global cybersecurity defense fund in which governments declassify and share attack attempts in order to create a deterrent and feed it to AI programs.
- Fund basic research at the academic and governmental level to counter the threat of existing cyber threats and expand the knowledge base of future scenarios.

Working Group 3, Committee 3
Ethics and Code of Practice (Privacy, Transparency, Accountability, etc)

Committee Chair: Jaan Tallinn

Committee Outline
Human societies have millennia of experience in the governance of human intelligence, as embodied in our traditions, institutions, laws, cultural norms, and moral codes. These codes reflect certain ethical aspirations or perspectives, such as the recognition that all human beings are equal in fundamental rights and dignity.

To empower innovation and competition while ensuring human dignity, we need common ethical norms and codes of practice. In the governance of AI, no such code exists. Its sweeping ascent in all aspects of life raises a number of ethical challenges. Some of those challenges precede AI, but are amplified by it; others are novel, central among which is the progressive and irreversible surrender of human agency to non-ethical agents. Such challenges may require entirely new normative approaches.

These challenges all compel careful consideration of the values and moral norms that will govern the delegation to machines (and indirectly to their designers and operators) of decisions that affect people.

Topics Discussed By The Committee
On the matter of ethics and codes of practice associated with AI, the committee agreed there are several angles to address the issues of ethics, namely, tensions between culture, tensions between core values and different speeds of development, and tensions in current universal values and those of the future.

Human evolution through adversarial environments and behavioral cognition research has proven that more differences exist than commonalities in human values. This will be problematic when codifying our value system and questions are raised as to whether ethics can be reduced to binary systems. Inherently, the different weights imposed on different values will create conflict and will translate into conflict. The committee suggested adopting certain broad values as a basic guideline regarding AI implementation.

“The AI governance framework, in order to be legitimate, must rest on a broad-based collective intelligence process. It must also be accessible to the ordinary citizen.”
of Artificial Intelligence suggested that deep ethical issues cannot be delegated to machines; there must be thresholds when AI impacts peoples’ lives.

“An AI governance framework, in order to be agile, must be adaptable to complex system dynamics, and must fit within and leverage existing international cooperation endeavors.”

The committee then discussed the trade-offs between automation and human agency. Potential elements for trade-off referenced human dignity, life or death, mortgage, and jail among the few.

Human agency would mean keeping an “unmanipulated human in the loop.” However, humans are influenced by thought and choice architecture, which could be exacerbated by AI.

Humans are willing to be judged by another human being but it is uncertain whether it is ethical to trust justice to a machine.

General consensus was that the concept of freedom should center on the direction for a good future with the presence of artificial intelligence.

The committee considered learning from the dynamics of the healthcare system. The healthcare system has a built-in code of ethics that must align with societal values. Depending on the situation, there should be agile adaptation of ethical limitation in the application of AI. Discussions touched on a shift in artificial intelligence to human intelligence where patients trust the AI doctors more than the human doctors. The availability of AI will raise difficulties when establishing causal connections and what constitutes malpractice. Comparison was made to human pilot failure versus autopilot failure and the conclusion was a machine cannot be accountable. The Association for the Advancement of Artificial Intelligence suggested that deep ethical issues cannot be delegated to machines; there must be thresholds when AI impacts peoples’ lives.

The Way Forward

The committee closed its proceedings by recognizing that recommendations usually fall into small, siloed subcategories:

- The need for inclusion and representation in the conversation and solutions. Management of these issues should be horizontal rather than vertical (i.e., in education) and focus on how AI will transform society.

- Creating Data Collectives as short- and mid-term solutions for developing countries lagging in information pace.
Topics Discussed By The Committee
The committee mainly focused on three questions during this session:

1. What are the top sectors that governments should focus on for AI development?
2. What are the biggest barriers to AI adoption by government?
3. What are the best opportunities that governments can seize during the AI revolution?

The committee identified education, public service, and transportation as the three sectors in which governments should support AI development.

- **Education**
  - Investments in higher education will yield long-term policies in training the youth of the future to understand and cope with this rising technology.
- **Public Services**
  - Choose for cost efficiency, as AI could theoretically replace high-skilled positions and decrease the burden of cost to taxpayers. Examples include performance management assessment in Estonia and translation for immigrants in Italy.
- **Transportation**
  - Increases mobility within cities, decreases accidents and lowers costs.

The committee also discussed existing barriers and opportunities, emphasizing the need for education programs to produce AI experts and the potential for AI to increase citizen engagement if given a positive narrative.
Topics Discussed By The Committee

The committee first discussed the greatest opportunities for multi-stakeholder cooperation in AI. Several technical initiatives were brought up that employ a consensus-building process and develop soft governance instruments that organizations can adapt to local law and governments, including:

• The British Standards Institution’s 2016 8611 guide to the ethical design and application of robots and robotic systems.
• The IEEE’s P7000 standard to enable the pragmatic application of Value-Based System Design methodology in identifying, analyzing, and reconciling ethical concerns of end users at the beginning of systems and software life cycles.

• In addition, the committee noted the need to create interdisciplinary conferences to address the ongoing development of ethics in AI, as well as the need to reward such interdisciplinary careers in academics.

The committee also discussed the challenges of cooperation between the public, private, and academic sectors, noting that governments have a role to play in driving and encouraging social impact ventures. Businesses geared toward the social good are not well-reflected in legal structures and are misunderstood by governments, yet must be incentivized by the public sector in order to mitigate the social risks that AI poses and be rewarded for promoting an ethical approach in technical development. The committee suggested that:

• Governments can lower the barriers to entry by allowing technologically driven social impact ventures to join the competitive market for free, giving them an edge in a Big Tech-driven environment. For example, Uber has been hugely successful, but has done so at the expense of the labor market, which could be corrected by a tax to level the playing field with smaller firms.
• Government can be strategic in encouraging academia to collaborate with the public sector by implementing a KPI approach for funding; incentivizing academic institutions to collaborate; and imposing incentives from the top down.
• Ministries of AI can help by building an ecosystem of support - convening people together, implementing templates of AI, incorporating neutral services, and educating entrepreneurs in how they can build a business. In addition to supporting a classic innovation model, governments must tailor their programs to address challenges specific to AI, including issues of data, privacy, etc.
• Governments can create new aspirations for businesses to become good corporate citizens

The committee considered if government will be needed to regulate AI in the future, given that the digitization of the economy has not been guided by policy, but by the tech world. The committee also raised the question of who would lead an international effort to establish norms - the United Nations? The G8 or G20? It was agreed it should be a neutral institution, such as the Association for Computing Machinery (ACM), which is a global scientific and educational organization representing the computing community. Such institutions can support practical solutions rather than create the redundancies of bureaucratic government systems.

The committee noted there is a need for governance to be done contextually, in consideration of culture and differences in values. For example, countries such as China place a high premium on harmony and community, while countries such as the U.S. place a premium on individualism and personal freedom – two potentially conflicting value systems. The committee drew upon Dr. Lan Xue’s thesis on global governance issues, which argues that not one single regime can have full authority of developing governance systems, and that there is a need to coordinate different regimes - perhaps through a global internet forum.

The committee suggested that a governance body can:

• Start by establishing basic trading standards for technical and institutional cooperation across borders.
• Create policies that imagine society in 5-10 years, rather than create reactive regulatory policies geared towards stopping AI today.
• Establish a blueprint for a global governance system to map all the necessary resources.
• Include all parties at the table in order to shape and guiding realistic smart city implementation.

The committee also discussed the need to re-evaluate the statistical measures by which societies define economic prosperity. Gross Domestic Product (GDP) as a metric is outdated, and governments must evolve and to use a triple bottom line mentality (fiscal numbers plus environment plus well-being or “people, planet, and profit”) in order to provide a sustainable platform for society and business to exist and flourish for the long term.

Indicators like the OECD’s Better Life Index and Bhutan’s Gross National Happiness are already providing multiple governments around the world with holistic indicators designed to increase citizen well-being. China also recently adopted their Green Development Index as an annual assessment that covers six categories, including resource utilization, environmental governance and quality, ecological protection, GDP quality, and green living, aimed at improving the country’s economic and social development for its citizens.

The committee then talked about the inherently outdated modes of most governmental systems and questioned their efficacy given the rapid pace of AI advancements, including:

• The need for government to be faster in the testing and application of privacy mechanisms such as GDPR in order to explain its algorithm and test its efficacy.
• Overhauling models of bureaucratic governance in order to keep up with the pace and flow of information, especially with data processing.
• Invest in collective human intelligence design.
• Focusing on positive opportunities rather than simply mitigating risk by creating positive, evidence-based applications for developing ethical AI, such as regulatory sandboxes.

The Way Forward
The committee closed its proceedings by suggesting further study of the questions raised, including:

• The need for governments to broadly govern by creating future narratives rather than “governing by crisis.”
• Create buy-in from academia and private firms by sharing information and data in order to create trust.
• Increase cooperation across institutions in order to encourage collective data sharing and the adoption of guidelines across local and national contexts.
• Explore governance models employed by the World Health Organization, Intergovernmental Panel on Climate Change, the Internet Governance Forum at the UN as interesting models to adopt.
• Further define the role of the UAE in the global conversation, advise them what role is best to play and advise them as a base to generate policy action.
Working Group 4, Committee 3
Metrics of Well-being

Committee Chair: John C. Havens

Committee Outline
Individual and aggregate citizen well-being are not accurately captured by traditional national performance indicators such as GDP. Societal well-being and progress depend on contributors beyond macro-economic Indicators which prioritize productivity as a primary metric of value.

While multiple global indicators of well-being are fairly robust, it is essential to grow awareness on their importance and use for policy makers and citizens alike. This is why highlighting respected indicators like The Organization for Economic Co-operation and Development that provides a framework for measuring well-being built around distinct domains of material conditions, quality of life, and sustainability is so important. It’s also important to highlight work like the UN’s World Happiness Reports and the recent Global Happiness Policy report released in Dubai in early 2018. These reports feature multiple global case studies showing how the inclusion of both subjective and objective well-being metrics about a country’s citizens can be utilized to create, change, or adapt policy. This becomes a much more direct way to show citizens a government is helping increase their well-being.

AI and associated technologies can support the measurement, analysis, and increase of positive well-being. For example, autonomous vehicles utilizing technology to minimize negative carbon emissions could provide data to well-being Indicators measuring environmental factors.

**Topics Discussed By The Committee**
The committee began the discussion with an exploration of the definition of well-being. A number of questions were raised by participants whose primary area of expertise did not relate to well-being indicators, including:

- “Is well-being focused on the experience of pleasure and sensation?”
- “Do purpose and direction in life provide a strong foundation for well-being metrics?”
- “Doesn’t the ability to have fun and to have options on the possible courses of life determine well-being? The greater the choices one has, the higher their well-being and ability to make decisions that lead to even more options, correct?”
- “Is it true that both quantitative and qualitative factors are important in measuring well-being?”
- “Considerations of the triple bottom line that ensure improvement for people, planet, and profit are important, correct?”

- “Isn’t it true there is no perfect singular metric of well-being, that it shouldn’t be measured in isolation, nor with a sense of geographic universality?”
- “How can we aspire to better metrics?”
- “How can AI help us achieve this?”
- “What kind of world do we want?”
- “How can citizens play a role and apply collective intelligence to define measures of well-being?”

The committee then considered key issues with using GDP as a proxy for measuring human well-being.

- The GDP is the primary measure for economic value, but was established as a global benchmark in the 1940s. It was noted, as an example of some its limitations, that GDP did not measure care work when it was created and that it still doesn’t today. It is noteworthy that care work is primarily performed by women around the world, which means that GDP can be said to not adequately measure women’s value.
- Conversely, it was noted, AI also holds the potential to diminish well-being by undermining fundamental values such as privacy, dignity, control, agency, and freedom to choose.

Members of the committee then discussed the reverse correlation - how well-being can impact AI development. This included:

- Having well-being measures built into the design of AI during the development phase to minimize downsides risks.
- Collecting data on citizens’ happiness and well-being from across the world and apply machine learning techniques to assess the global and regional state of well-being.

However, since environmental considerations were deprioritized as a result, humanity has severely depleted the planet’s resources, gravely threatening sustainable development. Measuring environmental Indicators reflects the vital need to manage the planet’s resources to the benefit of humans and the natural environment.

The committee examined the relationship between AI and well-being, and how the former can inform the latter:

- AI holds the potential to support the development of data to be utilized by happiness and well-being metrics.
- The greater the choices one has, the higher their well-being and ability to make decisions that lead to even more options, correct?

- Our environmental constraints are very different today from when GDP was invented. After the second World War, much of the world lay in ruin, so exponential growth was an appropriate objective and understandable value.
The Way Forward
The committee concluded the discussion by debating a series of pragmatic steps to achieve higher levels of well-being across the world through the AI revolution. Among the steps proposed were the following:

- Look beyond GDP and use established happiness and well-being indicators. Although countries aspire to have a high GDP, we must also prioritize the use of established metrics that consider the many other aspects of societal well-being.
- Embed an overarching principle for data governance that promotes human flourishing and prosperity.
- Implement happiness and well-being metrics that are complementary and have both objective as well as subjective measures.
- Apply an integrated approach that marries AI and well-being, whereby each informs the other in a positive way.

Further Resources
- The Well-being Committee report of The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
- The State of Well-being Metrics (general introduction to established Well-being Indicators from Ethically Aligned Design)
- OECD Better Life Index
- 2018 World Happiness Report

“AI must advance the well-being and freedom to flourish of all humans. How can this aspiration be realized without accountability? Without empirical evidence of the effectiveness of AI in vital domains such as medicine, financial services, or the law? Without ensuring that manufacturers and operators of AI in key societal functions are competent?”
4. APPENDIX I
2018 Roundtable Committee Chairs

ANDRÉ LOESEKRUG-PIETRI
founded A Capital in 2010 and serves as its Chief Executive Officer, President, Managing Partner, Managing Director, and Chairman. He has more than 10 years of private equity and merger and acquisition experience, among them: serving as Chairman and Managing Partner of A CAPITAL China Outbound RMB Fund; Managing Partner, Co-Founder and Chief Representative of CEL Partners; Deputy Managing Director and Partner of JACCAR Holdings SA; Investment Director at Burelle Participations, SA, SCR. He is the Chairman of the Private Equity and Strategic M&A Working Group of the European Chamber of Commerce in China. He is a graduate of the HEC School of Management, International M.B.A. Program of the Michigan Business School, and studied at the Global Leadership & Public Policy Program from Harvard Kennedy School.

FRANCESCA ROSSI
is a distinguished research scientist at the IBM T.J. Watson Research Center, as well as a professor of Computer Science at the University of Padova, Italy. She is a member of the Board of Directors of the Partnership on AI, where she represents IBM. She has been president of International Joint Conference on Artificial Intelligence (IJCAI), and an executive councilor of the Association for the Advancement of Artificial Intelligence (AAAI). She is the Editor-in-Chief of the Journal of Artificial Intelligence Research (JAIR) and a member of several editorial boards, including Constraints and Artificial Intelligence. She co-chairs the AAAI committee on AI and ethics and is also a member of the scientific advisory board of the Future of Life Institute, as well as the Leverhulme Centre for the Future of Intelligence.

ANTONIO SAMARITIN
joined AgID - Agenzia per l’Italia Digitale – the Agency for Digital Italy - in 2015 as Director General. Prior to that, he held different positions developing competencies in the industrial sector and skills in business development and in project and people management.

JAAN TALLINN
is a founding engineer of Skype and Kazaa. He is considered one of the foremost experts on P2P technologies. He founded the Centre for the Study of Existential Risk with Huw Price and Martin Rees. He is also a co-founder of the Future of Life Institute and supports other existential risk research organizations. He is a partner at Ambient Sound Investments (asi.ee), is an active angel investor, and has served on the Estonian President’s Academic Advisory Board.

CALUM CHACE
is the author of The Economic Singularity: Artificial Intelligence, and the Death of Capitalism (2016), Surviving AI: The Promise And Peril Of Artificial Intelligence (2015), and Pandora’s Brain (2014) – a novel examining the possible impact of super-intelligence. Prior to this, he served as a chairman, coach, and consultant (three Cs) to entrepreneurs based on 30 years’ experience as a CEO, strategy consultant, and marketer.

JOHN C. HAVENS
is the Executive Director of the Institute of Electrical and Electronics Engineers (IEEE) Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. He is the author of Heartificial Intelligence: Embracing Our Humanity to Maximize Machines and Hacking H(app)iness - Why Your Personal Data Counts and How Tracking It Can Change the World. Prior to this, he founded The H(app)athon Project, a non-profit foundation utilizing emerging technology and positive psychology to increase human well-being.
JOHN HIGGINS
CBE has been the public face of the digital sector in Europe since November 2011 and speaks regularly on platforms and at conferences throughout Europe and beyond. He was Director General of DIGITALEUROPE, the association for the digital technology industry in Europe, until Spring 2017, following nine years leading its UK member association, TechUK. In Spring of 2017 he became Chair of the Global Digital Foundation.

In 1995 he was appointed CEO of the Rocket Networks, a California based dot.com that provided the world's first online recording studios. He returned to the UK in 1998 and became Director General of the Computing Services and Software Association, one of Tech UK’s predecessors. He holds a degree in Mathematics from the University of East Anglia and then a post graduate Diploma in Accounting and Finance.

LEANNE FRY
is the acting Deputy CEO of the Australian Transaction Reports and Analysis Centre (AUSTRAC). She was formerly the Chief Innovation Officer and Chief Information Security Officer at AUSTRAC, and is also a Mentor at Tech Girls are Superheroes and WILES (Women in Law Enforcement Strategy).

JÜRGEN SCHMIDHUBER’S main scientific ambition has been to build an “optimal scientist” through self-improving artificial Intelligence. He has pioneered self-improving general problem solvers since 1987, and Deep Learning Neural Networks (NNs) since 1991. The recurrent NNs (RNNs) developed by his research groups at the Swiss AI Lab IDSIA & USI & SUPSI (ex-TU Munich CogBotLab) were the first RNNs to win official international contests. They recently helped to improve connected handwriting recognition, speech recognition, machine translation, optical character recognition, and image caption generation, and are now in use at Google, Microsoft, IBM, Baidu, and many other companies.

NICOLAS ECONOMOU
is the CEO of H5. He serves as the Co-Chair of the Law Committee of the IEEE’s Global Initiative on Ethics of Autonomous and Intelligent Systems. He is the Chair of the Future Society’s Science, Law and Society Initiative and Senior Advisor to the Future Society’s Artificial Intelligence Initiative. He has spoken on issues pertaining to artificial intelligence and its governance before policy makers and legal audiences at a wide variety of conferences and organizations, including the IMF Spring Meetings, the World Government Summit, and Harvard and Stanford Law Schools. He is the author of a number of published articles on issues relating to technology and the practice of law and has been featured in Forbes Magazine. He was a member of the Law and Judiciary policy committee for Barack Obama’s first presidential campaign, and was a pioneer in advocating for the application of scientific methods to electronic discovery.
NOZHA BOUJEMAA is the Research Director at the French Institute for Research in Computer Science and Automation, as well as the Director of the DATAIA Institute (Data Sciences, Intelligence & Society) and the Project Leader of TransAlgo scientific platform for algorithmic systems transparency and accountability. She is the Founding Director of the Digital Society Institute (ISN); President of the Scientific and Technological Council of IRT SystemX; Senior Scientific Advisor for “The AI Initiative”; and International Advisor for the Japanese Science and Technology Agency Program “Advanced Core Technologies for Big Data Integration”.

ROB MCCARGOW is the Programme Leader of Artificial Intelligence, Technology & Investment at PWC. He manages the company’s AI ecosystem and collaborates with partners across academia, government, technology vendors, start-ups, and other key stakeholders to drive innovation within the firm and develop new services for clients. In addition, he works with the All-Party Parliamentary Group on Artificial Intelligence, and he is also an advisor to the Institute of Electrical and Electronics Engineers (IEEE) Global Initiative for Ethical Considerations in AI and Autonomous Systems. McCargow is particularly focused on the issues relating to the impact of automation on the workforce, the future skills agenda, and ensuring that the benefits of AI are equitably spread across society.

STUART RUSSELL joined the faculty of the University of California at Berkeley, where he holds the Smith-Zadeh Chair in Engineering, and is the former Chair of Electrical Engineering and Computer Sciences. He is an Adjunct Professor of Neurological Surgery at UC San Francisco and Vice-Chair of the World Economic Forum’s Council on AI and Robotics. Russell is also the founder and Vice-President of Bayesian Logic, Inc., a data analysis start-up under contract with the UN to build a new Nuclear Test Ban Treaty global monitoring system. His books include The Use of Knowledge in Analogy and Induction, Do the Right Thing: Studies in Limited Rationality (with Eric Wefald), and Artificial Intelligence, A Modern Approach (with Peter Norvig).

ROBERT SILVERS is the former Assistant Secretary for Cyber Policy at the U.S. Department of Homeland Security. From 2014-2016, Silvers served as Deputy Chief of Staff at DHS, where he advised Department leadership on matters pertaining to cybersecurity, counterterrorism, aviation security, the immigration system, border security, trade and travel facilitation, and disaster response. Prior to that role, he served as the Senior Counselor to the Deputy Secretary of DHS, and before that as Senior Counselor to the Director of U.S. Citizenship and Immigration Services. Before joining DHS, Silvers was an attorney at the international law firm O’Melveny & Myers LLP, where he handled complex transnational disputes and investigative matters for his clients. Silvers also served as a law clerk to the Hon. Kim McLane Wardlaw of the U.S. Court of Appeals for the Ninth Circuit.
5 APPENDIX II

Working Materials used during the Law Committees deliberations

Proceedings of The First Global Governance of AI Roundtable
Dubai, 10 February 2018

Working Group 1: Mapping AI and Global best practices, reaping benefits, mitigating risks
Committee: Judicial system, access to justice, and the practice of law
• Committee Chair: Nicolas Economou nico.eco@thefuturesociety.org

Initial Considerations
1. Where is AI deployed in support of legal decision-making today?
2. What are the most significant benefits of AI for the justice system?
3. What are the most significant risks?
4. What considerations, if any, do you deem most overlooked in the current public dialogue?
5. What are the emerging “top-down” ethical and governance frameworks?
6. What are the “bottom-up” practical applications, pilot projects, or model best-practices we can beneficially draw on?

Example of Ethical Principles (1)
The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems:

General Principles
• Human rights: Ensure A/IS systems do not infringe on internationally recognized human rights
• Well-being: Prioritize metrics of well-being in the design and use of A/IS systems
• Accountability: Ensure that designers and operators of A/IS systems are responsible and accountable
• Transparency: Ensure A/IS systems operate in a transparent manner
• Awareness of misuse: Minimize the risks of misuse of A/IS systems

“It is our intention that by identifying issues and candidate recommendations regarding these principles they will eventually serve to underpin and scaffold future norms and standards within a new framework of ethical governance.”

Example of Ethical Principles (2)
The Future Society Science, Law and Society Initiative:

• AI shall not impair, and, where possible, shall advance the equality in rights, dignity, and freedom to flourish of all humans
 • AI shall be transparent
 • Manufacturers and operators of AI shall be accountable
 • AI’s effectiveness shall be measurable in the real-world applications for which it is intended
 • Operators of AI systems shall have appropriate competencies
 • The norms of delegation of decisions to AI systems shall be codified through thoughtful, inclusive dialogue with civil society.

Example of Models of Governance (1)
Gasser & Almeida, Harvard University:
Analytical model with three interacting layers:
• Social; Legal (norms, regulations, legislation)
• Ethical (criteria and principles)
• Technical foundations (data governance; algorithm accountability; standards)

Example of Models of Governance (2)
The Future Society Science, Law and Society Initiative:
Analytical model with four sequenced layers, general to specific:
• Central value (human beings are equal in fundamental rights, dignity, and the freedom to flourish)
• General principles (ethics)
• Public policy instruments (norms, legislation)
 • Implementation-level instruments (standards, codes of practice, technical controls, use of AI and blockchain)

Example of Models of Governance (3)
Wallach & Marchant:
Building global infrastructure for AI and Robotics:
• Comprehensive
• Agile
• Monitoring and flagging gaps
• Integrating soft law, hard law, and technological solutions
• International with complementary national and regional bodies
• Loose coordination of stakeholders

Some Entities With Research Interests At The Intersection of Law And AI
• IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
• The Future of Humanity Institute, Oxford University
• Center for Civil Justice, New York University
• The Future Society
• Berkman Klein Center for Internet and Society, Harvard
• Codex, Stanford
• Center for Law and Innovation at the Sandra Day O’Connor School of Law, ASU
• Berkeley Center for Law and Technology
• Information Society Project, Yale Law School
Example of ‘Model Best Practice’: Electronic Discovery in The Us

Legal Task:
- Fact-finding to enable the fair, speedy, and just adjudication of matters
- Involves the analysis of up to terabytes of data (e-mails, social media etc) to determine the truth of facts in controversy

Application of Artificial Intelligence:
- AI deployed since early 2000s to support or entirely replace human judgment in reviewing and analyzing the data

Challenge:
- While courts can demand transparency in training data or algorithms, and hold parties accountable, AI remained opaque

Model ‘Best Practice’:
- Emerging adoption by courts of measurement of efficacy in real-world applications of AI by real-world operators of AI
- Two metrics of efficacy: Precision and Recall (they answer the question: was the evidence sought actually found?)
- Benefits:
  √ Easy to understand for the ordinary citizen, judges and lawyers: 0% is worst possible; 100% is best possible
  √ No need to examine training sets or underlying algorithms, addressing powerful IP concerns
  √ Rests on studies conducted by the US National Institute of Standards and Technology
- Society can trust (and verify) that the vital fact-finding mission of the judicial process is beneficially served by AI

6. Acknowledgements
We would like to thank all of our participants for their engagement and leadership in the necessary dialogue on the governance of Artificial Intelligence.

Think Tanks And Non-Profit Organizations:
- Cyrus Hodes, Director & Co-Founder, AI Initiative and Vice President, The Future Society
- Nicolas Mialhe, President and Co-Founder, The Future Society
- Nicolas Economou, Chair, Science, Law and Society Initiative, The Future Society, CEO, HS
- Arohi Jain Rajvanshi, Head of Strategy, AI Initiative, The Future Society
- Yolanda Lannquist, AI Policy Researcher, AI Initiative, The Future Society
- Effie-Michelle Metallidis, Hollywood Representative, AI Initiative, The Future Society
- Shuang Lu Frost, China Representative, AI Initiative, The Future Society
- Josephine Png, Research Associate, AI Initiative, The Future Society
- Marie-Therese Png, Research Associate, AI Initiative, The Future Society
- Holley Chant, Sustainability Expert, AI Initiative, The Future Society
- Jessica Cussins, AI Policy Lead, Future of Life Institute
- Calum Chace, Author
- Dan Vogel, Executive Director, Center for Public Impact
- Jaan Tallinn, Founder, Centre for the Study of Existential Risk
- Jessica Cussins, AI Policy Lead, Future of Life Institute
- John Higgins, Chairman, Global Digital Foundation
- Joshua New, Policy Analyst, ITIF
- Kirsten Bound, Executive Director, Nesta
- Michael Page, Policy & Ethics Advisor, OpenAI
- Stephen Cave, Executive Director, Leverhulme Centre for the Future of Intelligence
- Subbarao Kambhampati, President, AAAI
- Academia:
  - Jonghwan Kim, Dean of the School of Engineering, Korea Advanced Institute of Science and Technology (KAIST)
  - Juergen Schmidhuber, University of Lugano
  - Lucas Kello, Director, Centre for Technology and Global Affairs, University of Oxford
  - Mitja Jermol, Co-Chair AI Lab, Jožef Stefan Institute, Slovenia
  - Nozha Boujemaa, Research Director at INRIA, Director of DATAIA Institute
- Marie-Therese Png, Research Associate, AI Initiative, The Future Society
- Holley Chant, Sustainability Expert, AI Initiative, The Future Society
- Jessica Cussins, AI Policy Lead, Future of Life Institute
- Calum Chace, Author
- Dan Vogel, Executive Director, Center for Public Impact
- Jaan Tallinn, Founder, Centre for the Study of Existential Risk
- Jessica Cussins, AI Policy Lead, Future of Life Institute
- John Higgins, Chairman, Global Digital Foundation
- Joshua New, Policy Analyst, ITIF
- Kirsten Bound, Executive Director, Nesta
- Michael Page, Policy & Ethics Advisor, OpenAI
- Stephen Cave, Executive Director, Leverhulme Centre for the Future of Intelligence
- Subbarao Kambhampati, President, AAAI
- Academia:
  - Jonghwan Kim, Dean of the School of Engineering, Korea Advanced Institute of Science and Technology (KAIST)
  - Juergen Schmidhuber, University of Lugano
  - Lucas Kello, Director, Centre for Technology and Global Affairs, University of Oxford
  - Mitja Jermol, Co-Chair AI Lab, Jožef Stefan Institute, Slovenia
- Nozha Boujemaa, Research Director at INRIA, Director of DATAIA Institute
- Marie-Therese Png, Research Associate, AI Initiative, The Future Society
- Holley Chant, Sustainability Expert, AI Initiative, The Future Society
- Jessica Cussins, AI Policy Lead, Future of Life Institute
- Calum Chace, Author
- Dan Vogel, Executive Director, Center for Public Impact
- Jaan Tallinn, Founder, Centre for the Study of Existential Risk
- Jessica Cussins, AI Policy Lead, Future of Life Institute
- John Higgins, Chairman, Global Digital Foundation
- Joshua New, Policy Analyst, ITIF
- Kirsten Bound, Executive Director, Nesta
- Michael Page, Policy & Ethics Advisor, OpenAI
- Stephen Cave, Executive Director, Leverhulme Centre for the Future of Intelligence
- Subbarao Kambhampati, President, AAAI

Government:
- Andre Loesekrag-Petri, Joint European Disruptive Initiative (JEDI)
- Alexander Kuleshov, Professor & President, Skolkovo Institute of Science and Technology, Russia
- Antonio Samaritani, Director General, Agency for Digital Italy, Presidency of the Council of Ministers, Italy
- Catherine Pugin, Scientific collaborator, TA Swiss
- Joel Martin, Director R&D, National Research Council, Canada
- Laurence Liew, Director for AI Industry Innovation at AI Singapore
- Leanne Fry, CIO AUSTRAC, Australia
- Marco Bani, Head of Technical Secretariat Agency for Digital Italy, Presidency of the Council of Ministers, Italy
- Marten Kaevats, National Digital Advisor, Estonia
- Yann Bonnet, General Secretary, French National Digital Council
- Yousef Alhammadi, Advisor (Analytics) at Office of Strategic Affairs, The General Secretariat of the Executive Council, Abu Dhabi
International And Multinational Organizations:
Anne Carblanc, Head of the OECD Information, Communications and Consumer Policy
Irakli Beridze, UN Center on AI, Head, UNICRI
John C. Havens, Executive Director, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
Reinhard Scholl, Deputy Director, ITU
Wonki Min, Chairman, Committee on Digital Economy Policy, OECD

Private Sector:
Ajit Jaokar, CEO Futuretext
Andrew Archer, CEO, Aitheon
Anima Anandkumar, Principal Scientist at Amazon AI
Carol Reilly, President of Drive-AI
Elizabeth Linder, CEO, Conversational Century
Francesca Rossi, IBM Watson, Partnership on AI
Gregory G. Curtin, CEO CivicConnect
Hazar Alzaki, Head of Government Affairs, Microsoft Gulf
Helen Liang, Managing Partner, FoundersX Ventures
James Hairston, Head of Public Policy for AI and Virtual Reality, Facebook
Kenneth Cukier, Senior Editor, The Economist, United Kingdom
Marc Teerlink, Global Vice President SAP Leonardo, New Markets and AI
Marek Havrda, Strategy Advisor, GoodAI
Michael Philips, Associate General Counsel, Microsoft
Natalie Kaspersky, CEO, InfoWatch Group

Pavel Abdur-Rahman, Associate Partner, AI and Big Data, IBM
Richard Stirling, CEO, Oxford Insights
Rob McCargow, AI Programme Leader at PwC
Ryan Burleson, Chief Strategic Officer, Aitheon
Ryota Kanai, CEO & Co-founder, Araya, Inc., Tokyo, Japan
Sui Yang Phang, Managing Partner, Nest
We would like to recognize the following organizations for their engagement and leadership in the necessary dialogue on the governance of Artificial Intelligence.