GLOBAL GOVERNANCE OF A.I. ROUNDTABLE
WORLD GOVERNMENT SUMMIT

GGAR Expert Group & Sub Committee Briefing Documents

This document provides an overview of each Expert Group and sub-committee topics that will be discussed throughout the day of the Roundtable. We give a short introduction to the discussion topic and highlight some key thought-starter questions that experts can use to frame and expand the discussions at GGAR.

WORKING GROUP 1: MAPPING THE RISE OF AI AND ITS GOVERNANCE

WORKING GROUP 2: GOVERNING THE RISE OF AI IN DIFFERENT CONTEXTS

WORKING GROUP 3: INTERNATIONAL PANEL ON AI & AI FOR THE UN SDGs

WORKING GROUP 4: MAKING THE AI REVOLUTION WORK FOR EVERYONE

Working Group 1: Mapping the Rise of AI and its Governance

February 10th, 9:45-11:00

Expert Group 1: Mapping AI Technological Development & Future Trajectories

Mapping AI technological development helps to set a foundation of understanding for all governance actors. This section briefly lists some major areas of recent research and advancements, serving as a starting point for discussion at GGAR. GGAR participants will take stock of the key trends and developments in AI technological progress within a rapidly changing context. This implies considering a methodology for mapping AI technological progress, and reviewing how progress is shared, including benchmarks, key indicators, or industry news that may be helpful to inform stakeholders. Although the course of technological development is unpredictable, anticipating future trajectories such as the impact of quantum computing and exponential technologies on the AI research & development landscape can provide important insights for governance.

Subcommittee A: Mapping AI Technological Development: Horizon Scanning

Chair: Gabor Melli

This group takes stock of the latest technological developments in Artificial Intelligence. 2018 saw progress in areas including computer vision, natural language processing and understanding, generative networks and adversarial examples, and transfer learning. In the path to human-level AI, notable gaps remain, particularly in common sense, model of the world, abstract reasoning, and meta-learning. Advances in hardware included growth in amount of compute used to train ML models[1] and a growing market for specialized hardware & semiconductors,[2] as well as use of new supercomputing[3] and quantum computing machines.[4]

Thought-Starters

- What are the most important aspects of AI development to map that could help the effective global governance of AI?
- What are key developments in AI research over the past year?
- What are key developments in hardware and compute for AI?
- Is it valuable to measure societal impacts from AI to help inform governance approaches?

[1] OpenAI Blog. 2018. *AI and Compute*.

[2] Investment and research in hardware for AI, including semiconductors or chips, are increasing among large technology companies (e.g. Intel, Qualcomm, NVIDIA, Samsung, AMD, IBM) and in new startups, primarily in the USA and China. Companies previously focused on software are also entering the market, including Amazon AWS, Google, Alibaba Group, Tencent Cloud, Baidu, Facebook, among others. Large technology companies such as NVIDIA and Google are making progress in specialized hardware such as TPUs and in GPUs (Ian Hogarth & Nathan Benaich, *State of AI 2018: A Good Old Fashioned Report*, June 28, 2018.) The market for AI chips could reach $30 billion by 2022. The Economist. 2018. *Artificial intelligence is awakening the chip industry's animal spirits.*).

[3] In October 2018, the German Research Center for Artificial Intelligence (DFKI) became the first institution in Europe to receive an NVIDIA DGX-2, considered to be the world's most powerful AI supercomputer. DFKI Press release. 2018. DFKI RECEIVES FIRST NVIDIA DGX-2 SUPERCOMPUTER IN EUROPE.

[4] In January 2019 IBM has unveiled an integrated quantum computer for scientific research. IBM Newsroom. 2019. *IBM unveils world's first integrated quantum computing system for commercial use.*

Subcommittee B: Mapping AI Technological Development: Methodology

Chair: Jack Clark, OpenAI

This group will discuss a relevant methodology or process for mapping the AI development landscape. What are the key criteria needed for an effective mapping process? What challenges are there in mapping technological developments that arise in existing exercises (e.g. AI Index 2019); and what are blind spots or areas of focus that are commonly missing (e.g. geographic representation, technical aspects)?

Thought-Starters

- What components/criteria should a good mapping methodology include?
- What could be good processes to map technological developments?
    - Who should be involved? Which industry, government, nonprofit, civil society and other actors are key?
    - How does one achieve global representation in mapping?
- What is missing and what are the usual blind spots when we do such mapping?[5]
- What are common challenges for mapping methodologies?
- How do we harmonize or standardize indicators (e.g. AI patents) to enable global mapping and comparisons?
- What are next steps to launch a mapping process? Who can support this?

Existing Initiatives

- AI Index Report
- EU Joint Research Center (JRC)
- China AI Index report
- Electronic Frontier Foundation AI Progress Measurement

[5] On GGAR preparatory expert calls, participants identified blind spots in developing countries & emerging markets, research undertaken by the military, and progress not published due to company secrecy/IP.

Subcommittee C: Mapping AI Technological Development: Key Indicators

Chair: Anima Anandkumar, NVIDIA

This subcommittee reviews how progress is shared, compared, and perceived across industry and the public. Methods and measures to separate hype from reality can support governance that is grounded in actual technological progress and can better inform stakeholders.

Mapping can include a review of the use of benchmarking,[6] publications in academic conferences, and other ways AI companies and researchers compare and share technological progress. In a competitive landscape where AI is a buzzword for investment and media attention, avoiding hype and identifying real technological progress is a challenge.

Thought-Starters

- What does AI technological progress mean? (E.g. Better data, compute, more accurate or complex models, capabilities in games or tasks)
- How should we measure progress? What is the role of benchmarks, indexing, or the reproducibility of results?[7]
- How can the research community be more effective in spreading news? (e.g. Communications from companies or individual researchers on social media and other platforms of engagement)
- Should we try to measure new AI technologies' impacts on societies? How do we measure positive and negative impacts?
- What role does governance and regulation have in certifying progress?

[6] Benchmarks are widely accepted indicators of progress in specific learning tasks, such as image classification (ImageNet) or natural language understanding (GLUE).

[7] Reproducibility of results in research is valuable to separate media hype from reality and to support governance that is grounded in actual technological progress. E.g. the 2019 ICLR Reproducibility Challenge can help researchers understand how reliable and reproducible their results currently are or are not.

Subcommittee D: Mapping AI Technological Development - Impact of Future Trajectories

Chair: Paul Epping, Philips Healthcare

In this roundtable, participants consider the impact of future technological or socio-economic trends on the course of AI development. Anticipating outcomes can be useful to support governance and policy-making, which should be flexible and adaptable to accommodate for fast-changing and unpredictable technological progress. Participants could consider the impact of advances in hardware, including quantum computing on AI research, or the intersection of AI with emerging and exponential technologies.

Thought-Starters:

- How can advances in hardware (e.g. quantum computing) impact deep learning or AI research more broadly?

- How can intersections with exponential or emerging technologies (e.g. nano, IoT, edge computing, robotics, blockchain) shape the future of AI?

- Which application areas may experience outsized investment or non-linear progress, and how can this affect technological progress in different societies (e.g. computer vision, natural language understanding)?

- How will public sentiment and socio-economic trends impact AI research & product development?
    - E.g. consumer demands for data privacy and security may drive shifts to decentralized ML; demands for energy efficiency may shape ML training; demands to monetize data assets may shape business models.

Working Group 1: Mapping the Rise of AI and its Governance

February 10th, 9:45-11:00

Expert Group 2: The Geopolitics of AI

The very rapid progress of AI technology means it has become a powerful tool on economic, political and military levels. Nested in the digital revolution, AI will help determine the international order for decades to come, accentuating and accelerating the dynamics of an existing cycle wherein access to power and technology mutually reinforce one another. The rising 'soft power' of private companies that currently lead AI research and deployment puts into question existing geopolitical frameworks, bringing into question the balance of power between states and corporations, and particularly the strategies that states can feasibly pursue to develop domestic innovation and influence the policies and practices of global corporations. AI could accelerate the concentration of resources and power in a small handful of large players, leading to a global, potentially lasting and destabilizing, consolidation of power. What are the implications of AI for geopolitics? What does it mean to talk of a "geopolitics of AI?" What should be the role of global and/or national government actors in establishing the norms and ethics of power differentials? How can global governance help promote shared prosperity and global stability while also allowing for pro-competitive and innovative business?

Subcommittee A: Digital Empires

Chair: Nicolas Miailhe, President, The Future Society; Co-Convener, Global Governance of AI Roundtable

This group takes stock of the latest geostrategic developments in Artificial Intelligence. AI, nested in the wider digital revolution, is impacting power dynamics across the world and at all levels. Given their leadership in AI technological development, China and the U.S. arguably form an AI 'duopoly', which may come to dominate international order for decades to come. Some analysts label the two nations as holding 'digital empires': their digital ecosystems are expanding rapidly, led by a handful of powerful multinationals, many of which have emerged as a result public-private partnerships emerging out of distinct political economies. These corporations deploy their products and solutions globally, and count hundreds of millions or even billions of users, which raises diverse concerns about strategic interests for other companies, states and regions.

Thought-Starters

- What are the implications of market concentration involving a few large firms? Are there counterexamples where lower market concentration has been equally beneficial/problematic?
- Can we speak of a US-China duopoly or is that fear over-inflated?
- What are the stakes for the Global South in the split of value extracted from AI?
- Should there be strategies deployed to help level the playing field for smaller countries and companies; and what fora might these be promoted in?
- How might global coordination and effective governance be used to prevent market dynamics which could drive a 'race to the bottom' in terms of ethical and safety standards?
- What steps should be taken now to enhance research intensity on the geopolitics of AI?

Working Definitions

Empire - Major political unit in which the metropolis, or single sovereign authority, exercises control over territory of great extent or a number of territories or peoples through formal annexations or various forms of informal domination.[8]

[8] O'Neill, D. 2016. *Empire/political science*. Encyclopedia Britannica. https://www.britannica.com/topic/empire-political-science

Subcommittee B: Exploring the geostrategic landscape of AI

Chair: Brian Tse, Policy Affiliate, Center for the Governance of AI, Future of Humanity Institute, Oxford University

This group explores the implications of AI on geopolitics and discusses policymakers' options for addressing potentially toxic power dynamics and to foster shared prosperity and stability. AI is becoming a strategic factor in international relations and will continue growing in prominence as its applications - particularly military and economic ones - develop. The dynamics of power and influence between nation states and multinationals affect the ability of states to develop successful strategies that support beneficial and safe domestic innovation while maintaining robust ethical and legal frameworks. Participants may wish to tackle topics such as the balance between domestic policies and regional and global collaboration in AI development. Participants may also want to devise possible strategies for the protection and advancement of smaller countries and companies and explore how global coordination and community building might prevent a "race to the bottom" in ethics and safety terms.

Thought-Starters

- What are some of the current and foreseeable trends at the intersection of international politics and AI development?
- What are the key disincentives and incentives that drive nation states to cooperate in AI development?
- What are the respective roles of private sectors, civil society and the research community in fostering global collaboration and cooperative norms in AI development?
- Is the prevailing discourse of an emerging 'AI arms race' misleading, self-fulfilling and/or suboptimal from a game-theoretic perspective?
- How promising are ideas such as "CERN for AI", an international AI mega-project for social good?
- What are some concrete and actionable approaches that can and should be taken now to prevent a 'race to the bottom'?

Working Group 1: Mapping the Rise of AI and its Governance

February 10th, 9:45-11:00

Expert Group 3: Agile Governance

A governance framework able to maneuver and manage the deep complexities of AI is agile, adaptive, credible, a good-faith broker, inclusive of multi-stakeholder input, comprehensive, and coordinated. At present in the governance of emerging technologies, traditional methods can lag and lack the flexibility needed to evolve at the same rate as technologies. Due to the rapid, complex, global and unpredictable nature of AI development, designing smart policies to mitigate these risks is particularly challenging. 'Agile governance',[9] or 'soft' governance tools, aim to address the shortcomings of standard policymaking processes and be more adaptive and responsive to fast-changing socio-technical trends as the AI and digital revolutions unfold.

Subcommittee A: Agile Governance: Multi-stakeholder Guidebook for Ethical and Safe AI

Chair: Andre Loesekrug-Pietri, Joint European Disruptive Initiative (JEDI)

This subcommittee aims to discuss the building of a handbook, guideline, or roadmap that will provide practical organizational steps for companies to integrate existing ethical principles within business activities. It will seek to establish a practical process to move from consensus-driven ethical principles, to standards and codes of conduct, to regulation and other approaches in agile governance, to ensure that AI adoption is ethical, safe, and benefits society broadly.

Thought-starters:

- Which ethical principles should be used as the foundation to an implementation guide book? What are existing examples of success?
- What are the organizational steps to build a practical guide for implementing ethical principles for AI?
- Which stakeholders should be involved to ensure that these practical guidelines are adopted?
- What steps are unique for different stakeholders or contexts? E.g. SMEs, technology companies, public agencies.

[9] For background on agile governance, see: Wallach, W. and Marchant, G. 2008. *An Agile Ethical/Legal Model for the International and National Governance of AI and Robotics.* AAAI.

Subcommittee B: Agile Governance: Decentralized approaches

Chair: Lawrence Lundy-Bryan, Outlier Ventures

This subcommittee will explore innovative, technology-based decentralized governance approaches. This can include use of blockchains or distributed ledger technologies for incentive mechanisms (cryptoeconomics), smart contracts, or technology-based governance approaches. In 'intelligent infrastructure,'[10] machine learning can be leveraged at systems levels to shift decision-making power towards local actors and improve efficiency for effective governance of AI.

At this prospective stage, discussions will consider the requirements, risks, and opportunities to operationalize such mechanisms for the governance of AI. Decentralized and technology-based governance approaches can be more agile and responsive and can transcend politics or national boundaries in some cases, to govern the societal risks presented in the digital economy. However, they carry important shortcomings and risks.

Thought-starters:

- What is required for effective AI governance? (e.g. justice, data quality, autonomy, explainability, transparency, accountability). Once we agree what characteristics are required, we can explore the range of technologies available to provide these.

- Who must be involved in governance decisions and at which levels? Which levels are most relevant: technological (DLTs, blockchains etc.), social & political (poli-centric decision-making), or socio-technical systems (e.g. Ostrom's principles)?

- What are the benefits, risks, and social concerns to manage? e.g. risks include technical risks, complexity for average person, security, privacy, human control, etc.

- Are decentralized approaches feasible at scale? What would be the requirements for successful implementation (e.g. broad cooperation on standardization and agreement to use open-source technologies)?

- What are initial steps or requirements to operationalize decentralized governance mechanisms? Which stakeholders must be involved?

[10] Jordan, M. I. 2018. *Artificial Intelligence - The Revolution That Hasn't Happened Yet*. Medium.

Subcommittee C: Agile Governance: Political Economy of Standardization

Chair: John C. Havens, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

Standards for AI are influenced by political, social and economic interests that surround the transformation of technology and market driven incentives. This subcommittee will explore the politics of setting standards to address the existing power dynamics and to level the playing field within the global standards community for AI.

Thought-starters:

- What are different standardization regimes according to source and legitimacy of (normative) power?

    1) Who is acting? Legislator/regulator or industry?
        - *A)* Top-down norm setting
        - *B)* Bottom-up consensus building

    2) What can we learn about the following two approaches for AI systems standardization?
        - *A)* National/states representation: (i) National Standards Body sets the standard which is then forwarded to (ii) ISO/IEC/ITU
        - *B)* Globally open, direct participation models (IEEE, IETF, W3C etc.)

- How do standard-setting regimes replicate existing concentrations of power, and how might an AI systems standardization learn from positive examples in other regimes?

- What are the consequences of these power dynamics for standard-setting regimes, and where have these been observed in previous case studies? (e.g. issues of inclusion, geographic diversity, market concentration, impact on public good, impact on innovation)

- How can we address these dynamics to meaningfully involve more stakeholders, organizations, companies and countries?

Subcommittee D: Agile Governance: Devising Innovative Regulation for AI

Chair: Isabela Ferrari, Federal Judge Brazilian Judiciary

This subcommittee will discuss the preconception that is held that regulation hampers emerging technologies like AI. It will ask questions about the legitimacy of this perspective and reflect on positive and negative examples of regulation and their impact on innovation. 'Hard' governance or law, including binding legislation and regulation, can help to support innovation by creating a level playing field and anchoring technological change in a given value system. Regulation can help to clarify rules of play and level the playing field. Additionally, regulation can be as holistic and inclusive as any other framework to support innovation.

Thought-starters:

- What are agile, adaptive policy or regulatory approaches that can support innovation? (E.g. Regulating the outcome rather than the process, clarifying rules, leveling the playing field)

- How might we reframe regulation as 'enabling' rather than hampering innovation both in public policy debates and within planning and implementation processes?

- What lessons can we learn from existing regulation such as the GDPR? How has GDPR supported, or hampered, innovation? Has GDPR met its intended objectives?

- How do we ensure legislators stay up to date or anticipate AI technological trends in order to prepare appropriate policy or regulation? What could be the role of different actors (e.g. NGOs, agencies) in helping legislators meet these challenges, and what techniques might be effective (e.g. timeline predictors of AI trends)?

Working Group 1: Mapping the Rise of AI and its Governance

February 10th, 9:45-11:00

Expert Group 4: Explainable & Interpretable AI

Deep learning neural networks are often labelled "black box" because, while their input and output are visible, the internal processes of getting from the input to the output remain opaque. Deep learning neural network architectures involve numerous "hidden" layers which are composed of linear and nonlinear functions. These functions are connected by weights which are adjusted in forward and back-propagation methods.

For some applications or sectors (e.g. healthcare, law, banking, HR), there is significant interest in being able to interpret and explain decisions made by AI systems. Therefore, explainability has become a topic for technical and academic research. This expert group explores how and in which circumstances explainability and interpretability are especially desirable, and why. It also pays particular attention to the issues of algorithmic bias and value alignment, for which a lack of explainability can exacerbate risks. Lastly, it aims at providing concrete policy recommendations for stakeholders that would enhance interpretability of AI decisions.

Existing Initiatives

- New York City task force to provide recommendations on addressing algorithmic bias in public services
- California law which requires companies to disclose whether they are using a bot to communicate with the public on the internet
- DARPA's research & innovation Explainable Artificial Intelligence program
- ACM Conference on Fairness, Accountability and Transparency (ACM FAT*)
- International Joint Conference on AI's Workshop on Explainable AI (IJCAI XAI)

Working Definitions

To avoid spending time discussing key terms & definitions, GGAR participants have drafted the following as working definition(s) for the purpose of discussion in this expert group:

Artificial Intelligence:

- A range of methods relying on algorithms at their core to learn and adapt, improving their models based on new data.

Explainability (in AI systems):

- "the information provided by a system to outline the cause and reason for a decision or output for a performed task – a 'post-hoc explanation."[11]

[11] Lipton, Z.C. 2017. *The Mythos of Model Interpretability.* https://arxiv.org/abs/1606.03490

- entails understanding the reasoning and justification of an AI system's decision

Transparency (in AI systems):

- "the level to which a system provides information about its internal workings or structure, and the data it has been trained with."[12]
- characteristic of AI systems where we can mechanistically understand what the AI does.[13]

Interpretability (in AI systems):

- "the level to which an agent gains, and can make use of, both the information embedded within explanations given by the system and the information provided by the system's transparency level."[14]
- "the degree to which an observer can understand the cause of a decision."[15]

[12] Lipton, Z.C. 2017. *The Mythos of Model Interpretability.* https://arxiv.org/abs/1606.03490

[13] Krones, J. 2018. *Ethics and Society* Presentation. Cognition, Microsoft.

[14] Tomsett, R. et al. *Interpretable for whom? A role based model for interpretable machine learning systems.* 2018 ICML Workshop on Human Interpretability in Machine Learning (WHI 2018), Stockholm, Sweden.

[15] Biran, O. Cotton, V.C. 2017. *Explanation and Justification in Machine Learning: A Survey.* Retrieved from : https://pdfs.semanticscholar.org/02e2/e79a77d8aabc1af1900ac80ceebac20abde4.pdf

Subcommittee A: What, Why and How?

Chair: Nozha Boujemaa, DATAIA Institute

This group takes stock of existing discussions on AI interpretability and explainability. In order to generate concrete recommendations for policymakers, it addresses the questions of what people mean by explainability, why explainability is desirable in some circumstances, sectors or applications and not in others, and how stakeholders can incentivize the implementation of explainability in AI systems being developed and deployed.

Thought-Starters

- In what circumstances, sectors or applications is it especially necessary to explain all the inner workings of an AI system? In what circumstances is it sufficient to explain a particular decision of an AI system?
- In what circumstances are explanations not necessary at all?
- What are the best guidelines available to policymakers trying to ensure that all AI applications are adequately explainable and interpretable?
- What role can counterfactual explanations play in making decisions more interpretable?

Subcommittee B: Algorithmic Bias - Value Alignment

Chair: Meeri Haataja, Saidot.ai

This group explores how to tackle algorithmic bias and how to align AI systems with intended goals and human values (the value alignment problem) This is a step beyond the general goal of AI interpretability and explainability. In order to generate concrete recommendations for policymakers, this subcommittee discusses policies and governance mechanisms as well as recent technical progress to facilitate the development and deployment of value-aligned, unbiased algorithms. It also considers how to operationalize these at scale.

Thought-Starters

- How can recent progress in explainable AI support work on algorithmic bias and value alignment?
- What policies or governance mechanisms could be assisting to reduce algorithmic bias in AI applications?
- What policies or governance mechanisms could work to address the problem of value alignment in AI applications?
- What are the government and non-governmental institutions that would be most suitable to incentivize better value alignment and prevent algorithmic bias at scale?

Subcommittee C: From Big Questions to Right Actions

Chair: Jim Dratwa, European Commission

This group addresses some of the fundamental questions in debates about AI explainability, interpretability, and algorithmic bias. It discusses the underlying reasons for the desirability of explainability and interpretability under certain circumstances, and for the aversion to algorithmic bias. In doing so, it will collect new insights for policymakers that are often forgotten in policy debates.

Thought-Starters

- Why do we want explainability and what do these reasons imply for the good governance of AI?
- Under what circumstances, sectors or applications is it ethical to trade off explainability for greater performance? How about algorithmic bias and performance?
- What kinds of regulatory frameworks and policies are required to align/orient market incentives towards the appropriate level of R&D funding to achieve better explainability of deep neural networks over the current techno-scientific cycle ?
- How do we ensure that demands for explainability are properly facilitated and meet the needs of more vulnerable groups?
- What kinds of discrimination (e.g. on the basis of wealth or gender) should be per se illicit in AI technology, and what forms could be deemed more legitimate in certain contexts?
- How can the answers to all of these questions inform near-term actions of stakeholders?

Working Group 2: Governing the Rise of AI in Different Contexts

February 10th, 11:30-12:45

Expert Group 5: Governance of the Development of AGI

Existing AI systems impact society in many ways and raise serious governance challenges. This pattern is likely to continue as AI systems with even greater learning and problem solving capabilities are developed. This expert group is split into three related subcommittees. Group A reviews current policies relevant to present-day AI that scale effectively to AGI. Group B investigates other governance mechanisms such as norms, education, science diplomacy and narrative building. Group C discusses the issue of stakeholder coordination.

Subcommittee A: Direct and Indirect Policy Recommendations

Chair: Jessica Cussins, UC Berkeley Center for Long-Term Cybersecurity

This subcommittee aims to identify concrete policy recommendations that could scale effectively through advances in AI and towards AGI. The objective is to generate actionable policy recommendations.

Thought-Starters:

- What are examples of current laws or regulations (proposed and implemented) that scale well to AGI?
    - Which ones arguably have a negative influence on long-term outcomes?
- What are the most promising regulatory approaches and institutions for governing AGI?
- If a technical solution to the value alignment problem were found, what policies would we want to see in place to ensure AGI is beneficial?

Subcommittee B: Other mechanisms for impact

Chair: Richard Mallah, Future of Life Institute

Governance goes beyond laws and policies. This subcommittee reviews governance mechanisms such as norms, ethical guidelines, education, science diplomacy, industry self-regulation, and narrative building.

Thought-Starters:

- Which mechanisms shaping the field today will scale well in the long-term? Are certain mechanisms overrated or underrated?
    - Which mechanisms will arguably have a negative effect on long-term outcomes?
- There are many ethical principles and guiding documents that exist today. What is the next step for producing cohesion and translating this into effective action?
- Given that AGI-related issues are more forward-looking and speculative than contemporary AI challenges and involve more uncertainty, what, if ever, is the right time to "convert" softer approaches to hard laws? What constellation of actors should be involved to decide if that shift is necessary?
- What can we learn from other areas of science and technology?

Subcommittee C: Stakeholder coordination

Chair: Seán Ó hÉigeartaigh, Cambridge Centre for the Study of Existential Risk

Ensuring that AI has positive impacts on society requires making progress in several domains, including ethics, technical alignment, and actor coordination. Improving coordination among actors will be essential if these positive impacts are to be harnessed and negative ones minimized.

Thought-Starters:

- How can we bridge near- and long-term concerns about AI?
- Given that AGI-related issues are more forward-looking and speculative than contemporary AI challenges and involve a lot more uncertainty, who are the key stakeholders now? What will be the right time to involve broader ranges of stakeholders, and how?
- What concrete steps can be taken to facilitate coordination on current issues that have lower stakes?
- What can we learn from past efforts to build a community of experts and decision-makers who trust each other and work together?
- What are the key failure modes to avoid in AGI stakeholder (non)-engagement?

Multi-stakeholder guiding principles

Name of initiative/conference	*Organizer(s)/Main actors*	*Guiding principles/output document*
The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems	IEEE	Ethically Aligned Design IEEE P7000 Standardization Projects
Beneficial AI	FLI	Asilomar Principles
Partnership on AI	Founding partners: Apple, Google/DeepMind, IBM, Facebook, Microsoft, Amazon.	Partnership on AI Tenets
EU High Level Expert Group on AI & European AI Alliance	EU Commission	Draft Ethics guidelines for trustworthy AI
G7 Summit 2018	G7 countries	Charlevoix Common Vision for the Future of AI
Forum on Socially Responsible Development of AI	University of Montreal	Montreal Declaration for Responsible AI

Single-actor guiding principles

- As of December 2018, 18 countries have established national AI strategies and guiding documents. 11 more have indicated that they are in the process of developing a national strategy. Overviews here and here.
- Companies including OpenAI, Google and Microsoft have released AI principles

AGI-specific initiatives

- January 2019 Beneficial AGI conference, Puerto Rico, Future of Life Institute
- Foresight Institute AGI strategy meetings. 2018 report here.
- Survey of AGI R&D projects by Seth Baum

Governance initiatives in other domains

- New York City task force to provide recommendations on addressing algorithmic bias in public services
- Washington Future of Work task force, which studies trends that might drive transformation, including automation
- The UK government is partnering with the World Economic Forum to develop a public procurement policy for AI

Policy and legislation specifically oriented towards governing AI-based technologies

Implemented

- California law which requires companies to disclose whether they are using a bot to communicate with the public on the internet
- State of California SB 10 law requires that every county uses algorithms to decide on bail, by Oct. 2019.
- State of California endorses Asilomar AI principles

Proposed

- The 'Self Drive Act' is a proposed law which will require US states to abide by federal safety standards for autonomous vehicles
- Export controls on AI and ML technologies as sensitive and dual-use technologies essential to US national security

Policies that affect technology in general, indirectly affecting AI-based technologies

Implemented

- GDPR

Proposed

- Compelling computing researchers to address the negative implications of their work through the peer review process
- The Data Care Act, which seeks to establish duties for online service providers with respect to user data
- California Data Privacy Act

Relevant working Definitions

Artificial Intelligence (AI)

- The designing and building of intelligent agents that receive precepts from the environment and take actions that affect that environment[16]

Artificial General Intelligence (AGI)

- AI with a wide range of intelligence capabilities, including the ability to achieve a variety of goals and carry out a variety of tasks, in different contexts and environments, including creative problem-solving and planning in new domains[17]
- AI system that equals or exceeds human intelligence in a wide variety of cognitive tasks[18]

[16] Russell S., Norvig P. 1995. *Artificial Intelligence: A Modern Approach*. Prentice Hall.

[17] Baum S. 2017. *A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy.* Global Catastrophic Risk Institute.

[18] Everitt T., Lea G., Hutter M. 2018. *AGI Safety Literature Review*. International Joint Conference on Artificial Intelligence (IJCAI).

Artificial Super Intelligence (ASI)

- Intellect that exceeds the best human brains in practically every field, including scientific creativity, general wisdom and social skills[19]

AGI governance

- Subset of AI governance, which "seeks to maximize the odds that people building and using advanced AI have the goal, motivation, worldview, time, training, resources, support, and organizational home necessary to do so for the benefit of humanity".[20] AGI governance can seek to, among other things, fund or otherwise support AGI R&D, or encourage certain ethical views to be built into AGI.

Unilateralist curse

- When acting out of concern for the common good in a unilateralist situation, the likelihood that a harmful action will be taken is higher the more actors there are who come to their decisions independently.

[19] Bostrom N. 2014. *Superintelligence: Path, Dangers, Strategies.* Oxford University Press.
[20] Dafoe, A. 2018. *AI Governance: A Research Agenda*. FHI. pp 5-6.

Working Group 2: Governing the Rise of AI in Different Contexts

February 10th, 11:30-12:45

Expert Group 6: Building Capability for 'Smart' Governance of Artificial Intelligence

AI adoption into public sector organizations and processes requires diverse and multi-level competence and capabilities. Governance is increasingly pertinent to striking and maintaining the right balance between maximizing the upside and minimizing the downside risks of AI systems. Designing and deploying smart governance for such systems implies the ability for policymakers to strike and continually maintain this equilibrium in a fast-paced and dynamic technological development environment.

While the stakes are high, policymakers are often less adept at managing AI technologies and systems and their adoption can often lag behind industry in terms of technical expertise. Knowledge gaps and talent shortages in this technology area and in cybersecurity hamper the capacity to formulate adequate technology policy and practices. For public sector organizations without prior expertise in managing AI systems or - more generally - digital technologies, the appropriate 'smart governance' strategy for their implementation is not immediately clear. Instead it is learned after experimental trials, and sometimes, errors. This expert group explores practical approaches to building capabilities for governing AI systems in the public sector.

Key definitions:

'Smart' Governance: this notion involves policymakers being well ***aware*** of technological developments (and the surrounding scientific, social, economic, industrial contexts) and ***anticipating*** their impacts on society before they are deployed at scale. Given the continually evolving nature of AI technologies and the uncertainty surrounding their future trajectories, smart governance implies having frameworks in place that are ***agile*** and ***adaptive***, depending on the course of technology and its deployment in society. This requires policymakers to develop a ***robust understanding*** of the global AI technologies and systems development landscape, granularly assess possible impacts on citizens, and quickly deploy control mechanisms through governance models that can intervene when technology deviates from ***societal values.***

The Virtuous Circle: In this 'learning-by-doing' smart governance approach, public organizations begin with small-scale projects involving adoption of AI systems and gradually scale up test projects in terms of size and scope. The process is iterative and based on feedback and learned expertise.

Subcommittee A: Building Competency for Governing AI in the Public Sector

Chair: Leanne Fry, AUSTRAC

This group defines and then outlines strategies to build capability and experience for 'smart' governance of AI systems and digital transformation. Whether implementing such systems within a public sector organization or setting a national data or innovation policy, managing AI adoption is complex and requires expertise. Approaches for public sector adoption can include gradually scaling up projects while building experience, forming multi-stakeholder partnerships with the private sector, and more.

Thought-Starters:

- What are the benefits and risks for the public sector to manage when governing adoption of AI systems?
- What are strategies or best practices for public offices beginning adoption of AI systems, when their technical expertise may be limited?
- How can the 'virtuous circle' for gradually scaling up projects help to build capacity? When is it especially appropriate or when might alternative policies be pursued?
- What is the role for multi-stakeholder partnerships, regulatory sandboxes, and other approaches to support innovation and build capability? What are successful examples of such strategies?
- What are effective national-level strategies to support an AI ecosystem (e.g. human capital, R&D, HPCs, social & labor policies)? What are unique variations across geographies and political economies?

Subcommittee B: How to Build Public Trust

Chair: Konstantinos Karachalios, IEEE

The public sector must gain general support and trust of citizens to ensure that digital transformation projects including AI technologies and systems succeed. This does not mean only "awareness or advertising campaigns" but also setting, ensuring and making visible adequate standards of trustworthiness of such technologies and systems. Without such substantial engagement, lack of support from the public is increasingly likely to emerge and this can threaten projects.[21] This group discusses how to build trust in public projects for AI systems and in digital transformation more generally.

Thought-Starters:

- What are the necessary conditions to build trust among publics (e.g. accountability, data governance, ethics, corporate governance)?
- How can governments demonstrate and prove trustworthiness (e.g. third party validators, communication strategies, showcasing case examples)?
- What is the role of public education and digital literacy, and how might that be better developed?
- What are examples of soft law (e.g. certification, voluntary standards, ethical frameworks) or hard law (e.g. legislation, regulation, top-down norms) that can build public trust while *enabling* innovation? Thinking specifically, is GDPR helping to build trust?
- How can the public sector engage citizens, industry, and other stakeholders in a bottom-up approach to building trust?

[21] As a motivating example, the city of Toronto's smart city project for Toronto's shoreline with help from Google's Sidewalk Labs has faced ongoing resistance and challenges from public concerns about privacy and data governance. Barth, B. 2018. *The fight against Google's smart city.* The Washington Post.

Subcommittee C: Lessons from Case Studies

Chair: Lord Tim Clement-Jones, UK House of Lords

Lessons from cases of adoption of AI into the public sector and public services can provide insights to enable less-experienced teams. Several countries have policies for AI adoption into public sector services for better efficiency and performance.[22] First, the UAE national AI strategy focuses on applications across public services and key sectors including energy, transport & traffic, health, environment & water, agriculture, security, education. Italy, China and Finland also plan to integrate AI into public administration and services to reduce costs and increase access and efficiency. What can we learn from cases?

Thought-Starters:

- Which cases of AI adoption in the public sector are pertinent to your country/context? In what ways is your country/context unique and what do those insights say about broader policy experimentation, learning and implementation? (See below)
- What are conditions or requirements needed to support AI adoption in the public sector or in the economy more broadly?

  Are there other case studies that you know of that offer unique insights about the emerging role of AI in public administration?

Cases:

- Finland[23]
- UK[24]
- UAE[25]
- Italy[26]
- China[27]
- New York City 'AI Sandbox' project, Vienna, Espo (Finland)

[22] See Dutton, T. 2018. *Building an AI World: Report on National and Regional AI Strategies.* CIFAR.

[23] Finland is leveraging partnerships with the IEEE to help build public sector capabilities. The IEEE is invited to act as 3rd party validator for systems in-country, and several members of Finland's public agency are in IEEE's Ethics Certification for Autonomous & Intelligent Systems (ECPAIS). Finland has a new strategy for training people in AI.

[24] See UK Sector Deal and UK House of Lords report: *AI in the UK: Ready, Willing and Able?*

[25] Strategy aims to make UAE government & public services more efficient and effective. Applications across public services and key sectors: energy, transport & traffic, health, environment & water, agriculture, security, education. The UAE Strategy for Artificial Intelligence, The Official Portal of the UAE Government, updated April 28, 2018.

[26] Task force studies how to implement AI to improve public services (e.g. reduce costs & increase access). White Paper identifies risks to manage (e.g. bias) and benefits (faster services, greater access through digital, cost & resource reduction). See: Artificial Intelligence at the Service of Citizens.

[27] Platform to integrate AI into government services. See: A Next Generation Artificial Intelligence Development Plan.

Subcommittee D: The case for Public-Private-People Partnerships

Chair: Ali Hessami, IEEE

Multi-stakeholder partnerships can build capacity and legitimacy by including a greater number of stakeholders. This can take several forms. First, partnerships with industry or companies with technical capabilities can boost capabilities.[28] Another form is regulatory sandboxes, involving support from a local municipality or government body to allow companies or other organizations testing and experimenting technologies in the real world. For example, several city municipalities have eased regulation to encourage testing and development of autonomous vehicles. Regulatory sandboxes are also an agile governance approach that allows for learning and capacity building for both public and private sector actors.

Thought-Starters:

- How can multi-stakeholder collaboration (municipal, national, individual, civil society, company stakeholders) support AI governance and adoption?
- What examples of PPPs or "public-private-people partnerships" (PPPPs) are relevant for building competence in governing AI?
- What are best practice models for PPP and PPPP initiatives that can be adopted in the context of AI?
- How to monitor, intervene and direct a PPP/PPPP towards achieving their stated goals?
- With AI research leadership residing in the private sector, how are governments best able to manage partnerships to benefit their societies at large?
- How are sandboxes or agreements best operationalized between regulators and companies to support innovation (e.g. autonomous vehicles)?
    - E.g. Autonomous vehicle zones

[28] For example, the Rwandan Health Ministry has partnered with Silicon-valley based drone manufacturing company Zipline to deploy drones that deliver medical resources including blood supplies in remote rural regions.

Similarly, the European Commission and individual EU member states have several initiatives aiming to better connect industry and government including innovation clusters and innovation hubs across countries. See the European Innovation Cluster for AI, Germany's Platform for Learning Systems and Germany's Cyber Valley, among numerous other examples.

Working Group 2: Mapping the Rise of AI and its Governance

February 10th, 11:30 - 12:45

Expert Group 7: Governing AI Adoption in Developing Countries

Developing countries have a distinct opportunity to harness emerging technologies to achieve inclusive growth and development. AI and other emerging technologies can empower people to improve their livelihoods through greater access to vital goods and services such as healthcare, education, food and energy, and help achieve the UN Sustainable Development Goals. However, governments also face considerable challenges such as policy and regulatory capability-building, as well as risks such as threats to employment, privacy, security, agency, inclusion and human dignity.

Given the pace and magnitude of the digital revolution, countries cannot afford to lag behind in leveraging emerging technologies such as AI. Emerging technologies offer developing countries immense potential to advance economic development and inclusive growth. While capturing opportunities from the AI revolution within the global digital ecosystem poses major hurdles, with significant room for error, embarking on a genuine digital transformation journey has become an urgent imperative for developing countries. To deal with the systemic complexity, velocity, and uncertainty surrounding the AI revolution, developing countries need to be able to rely on a clear governance framework which articulates policy and regulatory best practices, adoption/adaptation pathways and strategies, and to understand such policies and regulations in the regional and international context. Such governance frameworks and policy "play-books" are becoming crucial to guide developing countries in their work to strike a mature balance between capturing and maximising the upsides of the AI revolution in their context, while mitigating risks and minimizing downsides.

This expert group seeks to understand how developing countries can pursue pathways to becoming digitally mature and how they can foster AI adoption, while mitigating potential downside consequences that could come with the AI Revolution.

Working Definition

To frame the discussions of this Expert Group, we apply the 'developing countries' or 'developing areas' framing, which builds on low income (US$1,025 or less GNI per-capita) and lower middle income (US$4,026 - $4,035 GNI per-capita) countries, as set forth by the World Bank's World by Income 2017 framework.

Subcommittee A: Governing AI Adoption in Developing Countries: Building Capabilities while Avoiding Exploitation

Chair: Eileen Lach, IEEE

The lack of financial capital, technology, expertise and appropriate skills can put developing areas in a disadvantaged and vulnerable position, amplifying the risks of exploitation by large, foreign technology companies with global strategies that do not "think globally, but act locally". At the same time, developing areas need to be able to create and implement incentives and regulatory regimes that attract foreign investment and talent, which may involve short term imbalances in control and ownership. In the absence of smart governance mechanisms and capability-building pathways, rather than forging mutually beneficial partnerships, developing countries may fall into exploitative dynamics with actors offering capital, technical skills and exclusive access to technologies. This sub-committee takes stock of AI applications for developing areas through case studies and explores enabling conditions for local and regional AI capability building that minimize exploitation dynamics.

Thought-Starters

- What is the right balance between attracting foreign investment and expertise, and building local capabilities?
    - How can bilateral and multilateral agreements (e.g. trade agreements) be designed to enhance the possibility of this balance?
    - What roles do in-country education & incentives to study abroad and return play?
- How can we build robust Public-Private-People Partnerships for capability building?
- How is exploitation of local data assets avoided while investment from foreign technology companies is maintained and encouraged?
- How can we invest in education to build capabilities in AI and digital technologies?
- Case studies of developing countries where AI development and deployment has been actively promoted: India, Kenya, Rwanda, Ghana, Nigeria, South Africa

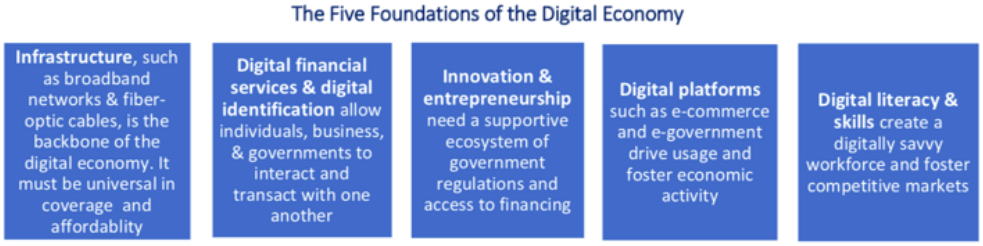
Subcommittee B: Governing AI Adoption in Developing Countries: Opportunities & Challenges

Chair: Zaki Khoury, World Bank

While the case for AI adoption in developing countries is clear, pathways towards safe and effective adoption remain uncertain. Several approaches can be taken to support AI and digital adoption in developing countries. While building fundamental capabilities for the digital economy in local contexts is crucial, given the several downside risks of the AI revolution, digital development must be coupled with robust policy/regulatory frameworks. This includes data and foreign direct investment governance to ensure that AI adoption takes place in legal and ethical ways. This group seeks to understand the various opportunities for adopting AI in developing countries, approaches that can be taken for digital and AI transformation (e.g. catch-up vs. leapfrogging strategies) and how the conditions for innovation might be best supported.

Thought-Starters

- From your experiences, what are the foundational nexus of elements (technical, economics, social, governance) needed to foster digital development in developing economy contexts?
- What are the unique differences (if any) between AI diffusion in 'developing areas' vs. 'advanced economies'?
- What layers of governance can be added to the "5 Foundations of Digital Economy" framing to enable safe and effective AI adoption in 'developing countries'? (*See below*)
- How can we leverage innovation, entrepreneurship and sustainable business models to help deploy AI and digital technologies in developing areas ?
- What is the right balance in leapfrogging vs. catch up to foster digital transformation? Is leapfrogging possible with AI?



Source: The World Bank Group Digital Economy for All Initiative

Subcommittee C: Managing Risks vs. Opportunities for Development

Chair: Stan Byers, New America, Policy Fellow on AI, Cybersecurity and International Development

This subcommittee explores governing AI in developing areas to mitigate risks while harnessing opportunities for innovation and development. AI offers unique opportunities for inclusive growth and improved access to healthcare, education and energy. Nonetheless, such opportunities are inextricably linked to risks arising from adopting digital and AI technologies (e.g. threats to data privacy, fairness, transparency, security), or from impacts on the economy. These include risks to international trade competitiveness, unemployment, inequality, and exploitation by other countries.

Thought-Starters

- What are the unique differences (if any) between AI diffusion in 'developing areas' vs. 'advanced economies'?
- How might we mitigate risks from AI adoption in 'developing countries' (e.g. unemployment or inequality from automation)?
- How can developing countries ensure labor competitiveness in international trade as manufacturing becomes much more technologically intensive?
- What strategies can developing countries use to invest in upskilling their workforce?
- How can developing countries ensure mutually beneficial public-private partnerships (PPPs) with foreign technology companies to support AI adoption? How can responsibility and enforcement mechanisms be better maintained to hold different actors to account?

Working Group 2: Governing the Rise of AI in Different Contexts

February 10th, 11:30-12:45

Expert Group 8: AI in the Judicial system, Access to justice, and the Practice of Law

Chair: Nicolas Economou, H5

Law (law-making; civil and criminal justice; law enforcement) is a social institution that both affects and is affected by technological change in fundamental ways. Although AI holds massive potential to benefit and improve access to justice for all citizens, and to advance the impartial, effective and speedy adjudication of justice, it also brings risks. The values that frame our legal systems, such as citizen participation, transparency, freedom from bias, dignity, privacy, and liberty may be compromised for practical objectives such as efficiency. AI, if properly governed, can enable the law to contribute more effectively to human well-being. The development of actionable, effective, and yet adaptable norms will help to ensure that the law's response to, and incorporation of, AI can be trusted by citizens, state institutions and civil society to enhance the functions of the Law and to protect and advance human well-being.

The Law Committee will be focusing on the IEEE principles of evidence of effectiveness (measurability), competence, accountability, and transparency to formulate policies and standards of practice (see below). These must have the ability to be operationalized in an effective but adaptable manner to foster trust.

Thought-starters:

- With respect to accountability of AI systems, what are the gaps for liability and responsibility, and which legal frameworks bridge these gaps?
- What metrics would we use to measure the effectiveness of AI and how do we make them implementable and practically usable for the public? What body or bodies should be responsible for the establishment and enforcement of these metrics?
- What standards should be used to ensure the effective competence of AI operators? What body should be responsible for the establishment and enforcement of these standards?
- How can we ensure transparency to understand AI systems in ways that do not compromise IP proprietary protection?
- What form should each of these principles take in practical terms and what issues should we bear in mind? What are the practical applications, pilot projects, or model best practices that we can draw on?

- How do we advance collaboration among relevant stakeholders to operationalize the Four Principles?

- What should an AI regulatory regime look like? Which expertise, standards and testing organizations have the authority and capabilities to create these standards and effectively enforce them?

Four IEEE Ethical Principles for Informed Trust

- **Evidence of Effectiveness:** Creators and operators shall provide evidence of the effectiveness (fitness for purpose) of an AI-enabled system.
- **Competence:** Creators shall specify, and operators shall adhere to, the knowledge and skill required for safe and effective operation of an AI-enabled system.
- **Accountability:** AI-enabled systems shall be created and operated such that it is possible to trace lines of responsibility among the agents involved in the creation and operation of the system for a given outcome.
- **Transparency:** The basis of any decision made (or to be made) by an AI-enabled system shall be discoverable.

Initiatives focusing on trust:

- **European Commission - High-Level Expert Group on AI**
  "Trustworthy AI will be our north star, since human beings will only be able to confidently and fully reap the benefits of AI if they can trust the technology."
- **OECD – Committee on Digital Economy Policy**
  "The OECD's Committee on Digital Economy Policy has established an Expert Group on AI in Society to scope principles that would foster trust in and adoption of AI and that could form the basis of a Recommendation of the OECD Council in the course of 2019."
- **US National Institute of Standards and Technology (NIST)**
  "NIST research in AI is focused on how to measure and enhance the security and trustworthiness of AI systems. This includes participation in the development of international standards that ensure innovation, public trust and confidence in systems that use AI technologies."
- **IEEE – Global Initiative on Ethics of Autonomous and Intelligent Systems**

Four of eight general principles have the objective of fostering an informed trust in AI in the Law (as stated above).

Working Group 3: AI for SDGs and International Panel on AI

February 10th, 14:00-15:15

Expert Group 9: From a Data Commons to an AI Commons

Chairs: Amir Banifatemi, XPRIZE & Don Gossen, Ocean Protocol

The AI Commons aims to bring together the key components for AI (data, compute, storage, interfaces, machine learning algorithms & talent) into a single platform. The objective is to connect problem owners with AI capabilities to address major global challenges. This involves bringing together diverse stakeholders to pool resources into a single platform that can be used to scale up use of 'AI for Everyone' and 'AI for Good'.

Notably, the AI Commons aspires to be a collaborative environment connecting problem owners with AI solutions. In its initial form, the AI Commons is proposed by several stakeholders and aims to launch in 2019. The 2019 GGAR Expert Group 'From Data Commons to AI Commons' will build upon the work already proposed by the group, as outlined in their website and paper titled 'AI Commons - Overview: Democratizing the Promise of Artificial Intelligence (AI).'[29] Specifically, it takes this framework as a starting point and aims to progress by exploring the main opportunities, requirements, challenges and next steps in making the AI Commons a practical reality. See **About the AI Commons** (backpage) for more information.

Key definitions:

AI Commons: an open contribution initiative that references problem solving approaches with AI. It contains a repository of AI models that have been used, a repository of usable and accessible data repositories and data commons, a directory of problems and AI specialists, and reference models for problem solving in the form of sandboxes that anyone can use.[30]

The AI Commons greatly widens the capacity of all Data Commons to serve as a platform for collaboration. Beyond data, the AI Commons expands the scope to include the key components necessary for AI: data, compute, machine learning algorithms & experts.

Data Commons: A curated data repository, organized by topic, community, or interest, that is usable for AI models and is accessible to anyone.[31] Open data has existed for many years, but usage is limited outside of specialist communities. Most open data is not data commons.[32]

[29] Banifatemi, A ; Bengio, Y ; Russell, Stuart; Rossi, Francesca et al. *AI Commons - Overview* (May 2018).
[30] Amir Banifatemi
[31] Amir Banifatemi, Alexandre Cadain
[32] Interview with Amir Banifatemi

Subcommittee A: Data Commons vs. AI Commons

Chair: Sarah Pearce, Paul Hastings LLP

This group outlines the scope and form of the AI Commons. The first step in shifting from the establishment of a Data Commons to an AI Commons is to identify the main commonalities, differences and boundaries between them. Expanding the type of resources beyond data and shifting towards a collaborative platform that connects problem owners with AI solutions raises new questions, requirements, and challenges.

Thought-Starters:

- What is the AI Commons and its scope?
- Where does a Data Commons end, and an AI Commons begin?
- What are the main opportunities? What are the main risks to mitigate?
- What is unique about an AI Commons compared to a Data Commons?
    - Which actors must be involved?
    - How should problem statements be framed and how might problem owners be matched with solutions?
    - What are the unique technical requirements for the AI Commons platform?
- What are the requirements or what is needed for an AI Commons?
    - How should data and AI services be governed on the platform to avoid disenfranchising certain domains or interest groups? (e.g. requirements for data & models, governing usage, permissions or limiting access to data)

Subcommittee B: Relevant Framework & Methodologies for Open Initiatives

Chair: Alpesh Shah, IEEE

This group explores ideas and lessons from other open initiatives that can help to inform the establishment of the AI Commons. Other open initiatives range from technical (e.g. Linux) to non-technical (Wikipedia) and their commonality is international participation in an open platform. What can we learn from open platforms and their collaborative frameworks?

Thought-Starters:

- What lessons can we learn from other open technology (e.g. Linux) or other initiatives (e.g. Wikipedia) to inform the AI Commons?
- How to bring international collaborators to build a common framework?
- How to raise awareness and build incentives to get diverse actors to join? How can a diversity of actors be achieved?
- What are relevant methods for identifying and prioritizing projects?
- How to frame problem statements in such a collaborative environment?
    - Might problem owners collaborate with technical experts to frame problems and possible solutions for AI?
- How can development best be accelerated, locally and/or globally? How best would an AI Commons be scaled up?

Subcommittee C: Building the AI Commons

Chair: Brent Barron, CIFAR

This group discusses approaches to practicably establish and implement the AI Commons. Considerations range from attracting and incentivizing participation to governing data usage. There is a need to identify next steps, partners, and participants from across sectors.

Thought-Starters:

- What are the steps and tasks required to establish and operationalize an AI Commons?
- What is needed to set up a regulatory sandbox or pilot that has value?
- How do we best frame problems and connect problem owners with solutions in a collaborative platform? How can match-making occur between problem owners and solutions?
- How do we incentivize participation and sharing data from diverse, international stakeholders?
- What should be the requirements or restrictions for joining? How can incentive mechanisms, rules or processes help to attract the “right” participants?
- What governance frameworks should be in place for execution?

Subcommittee D: Deploying the AI Commons

Chair: Ryan Budish, Berkman Klein Center for Internet & Society at Harvard University

This group discusses how to deploy and grow the AI Commons in the real world. Beyond scaling up and growing participation, this group will discuss important questions around governance and incentive mechanisms. Outcomes of the discussion can range from principles for guiding deployment to practical next steps and processes.

Thought-Starters:

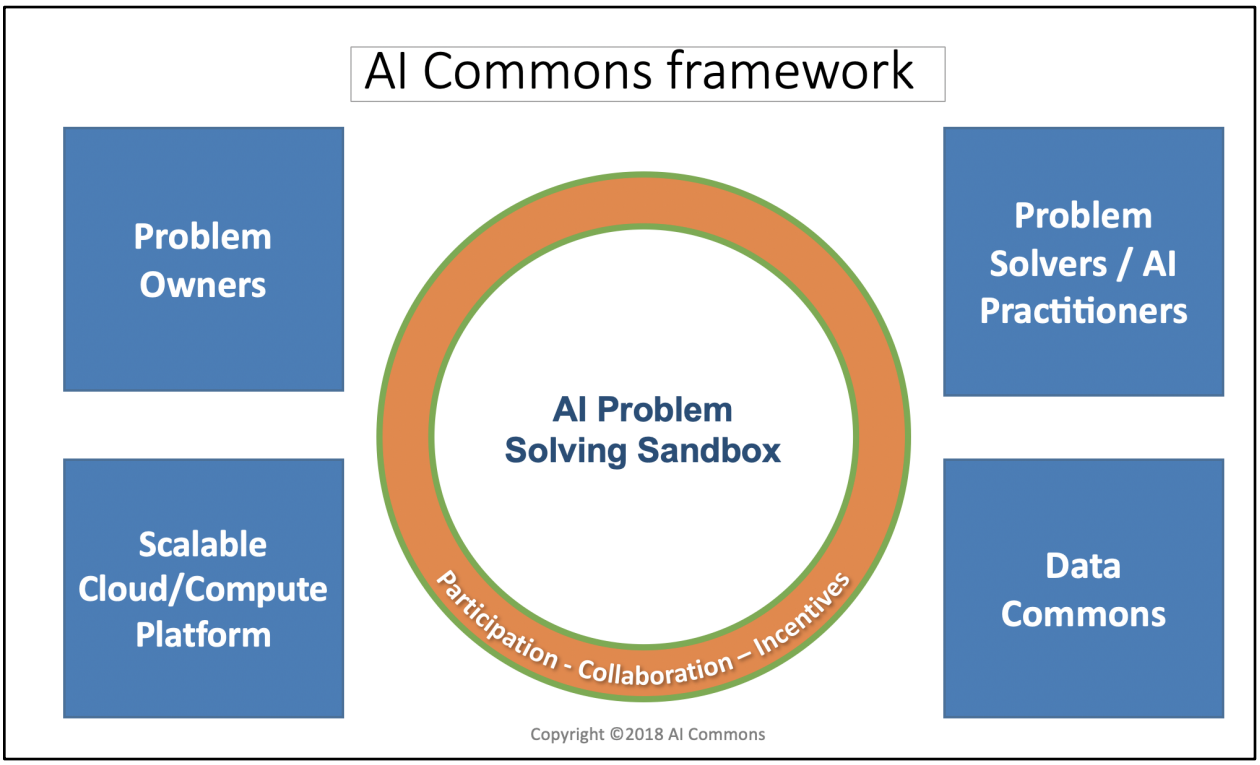
- Which stakeholders should be involved, and how might we attract them?
- What are the key governance issues?
    - How can we maintain security and privacy?
    - How can we best achieve global, inclusive, open participation?
- What incentives structures, rules and processes are useful to grow participation and govern behaviors?
- How can we raise awareness and foster stakeholder partnerships and support?
- What are the next steps for the development and partnership of the AI Commons? Who may be interested to support the project? GGAR participants?

About the AI Commons

According to the 2018 AI Commons initiative report, the AI Commons framework is designed to:

- Connect problem owners and the community of AI practitioners to collectively solve problems.
- Provide availability of trusted data repositories (data commons) and access to cloud and compute capabilities centers for problem solving to move forward.
- Create an "AI safe sandbox" for collaboration—a simple context with established standards for participation and incentives, as well as guidelines for safety, ethical consideration, data privacy, IP ownership, and project governance based on best practices.[33]

The report illustrates the AI Commons with the following framework:



Source: AI Commons - Overview (May 2018).

The AI Commons' framework can be visualized as a collaboration ring connected to four existing groups of (1) Cloud/Compute, (2) Data Commons, (3) Problem Owners, and (4) Community of problem solvers and AI Practitioners.

[33] Banifatemi, A ; Bengio, Y ; Russell, Stuart; Rossi, Francesca et al. *AI Commons - Overview* (May 2018).

Working Group 3: AI for SDGs and International Panel on AI

February 10th, 14:00-15:15

Expert Group 10: International Panel on AI

In December 2018, Canada and France raised the idea for the creation of an International Panel on Artificial Intelligence (IPAI). The new organization is to be modelled and adapted on the Intergovernmental Panel on Climate Change (IPCC). From what is understood of the current project configuration, which is still in the making, Canada and France are seeking the creation of an International Panel on AI that could become a global point of reference for understanding and sharing research on the issues arising from AI and best practices to cultivate positive outcomes from technological change, as well as convening international AI initiatives.[34]

This GGAR expert group seeks concrete proposals to make the International Panel on AI as viable, legitimate, and impactful as possible. These include recommendations concerning: pathways to incubation, establishment and expansion; memberships; governance, working processes and methodologies; and defining goals and objectives. This includes reflecting on a wider, inclusive framework for the governance of AI that has most potential for legitimacy, inclusivity and impact. This framework will need to borrow and adapt from other governance regimes including climate change, but also those pertaining to the internet, arms control, international trade and finance, and more. Government, industry, entrepreneurs, academia, and civil society will all need to be involved in the debate around values, ethical principles, the design of international agreements, and their successful implementation and monitoring.

Existing Initiatives on the governance of AI

- Global Governance of AI Roundtable (GGAR)
- IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
- ACM Conference on Fairness, Accountability, and Transparency (ACM FAT)
- AI4People
- UNICRI Centre for Artificial Intelligence and Robotics

[34] Prime Minister of Canada. 2018. *Mandate for the International Panel on Artificial Intelligence*. https://pm.gc.ca/eng/news/2018/12/06/mandate-international-panel-artificial-intelligence

Background

What the IPCC is.[35] The IPCC is a *process* for providing regular assessments of the scientific basis of climate change, its impacts and future risks, and options for adaptation and mitigation.[36] Created in 1988 by the World Meteorological Organization (WMO) and the United Nations Environment Programme (UNEP), the objective of the IPCC is to provide governments at all levels with a clear scientific view on the current state of knowledge on climate change and its potential environmental and socio-economic impacts. IPCC reports are also a key input for international climate change negotiations.[37] The IPCC is acknowledged by governments around the world as the most reliable source of scientific advice on climate change.

Members of the IPCC.[38] The IPCC is an organization of governments that are members of the United Nations and/or WMO. The Intergovernmental Panel on Climate Change is a panel of 195 member governments. Each IPCC member designates a National Focal Point, who prepares and updates a list of national experts to help implement the IPCC work programme. The IPCC admits Observer Organizations - the IPCC has at present 29 Observer Organizations from among UN bodies and organizations, and 87 non-UN observers. Representatives of observer organizations may attend sessions of the IPCC and the plenary sessions of the IPCC Working Groups.

Role of the IPCC. The IPCC reviews the most recent scientific, technical and socio-economic research produced worldwide relevant to the understanding of climate change.[39] The IPCC does not undertake new research, but examines published and peer-reviewed literature to develop a comprehensive assessment of scientific understanding which is published in IPCC Assessment Reports.[40] The scientific and consensus-based nature of IPCC assessments mean they provide vital and evidence-based common knowledge to underpin government policy decisions.

[35] IPCC (2019). *About — IPCC*. https://www.ipcc.ch/about/
[36] *Idem*.
[37] *Idem*.
[38] *Idem*.
[39] Australian Government - Department of the Environment. (n.d.). *The Intergovernmental Panel on Climate Change - Fact Sheet*. https://www.environment.gov.au/system/files/resources/7dc97a72-cc89-455c-a39a-0a2a4bc39e8d/files/ipcc-fact-sheet.pdf
[40] *Idem.*

Subcommittee A: International Panel on AI: Mapping and lessons from IPCC and other intergovernmental organizations

Chair: Francesca Rossi, IBM

This group takes stock of the lessons learned in the international governance of climate change and dual-use technologies. The Intergovernmental Panel on Climate Change (IPCC) sets a widely acknowledged example for a large multi-stakeholder platform based on science for international consensus-building on the pace, dynamics, factors, and consequences of climate change. Given the high systemic complexity, uncertainty, and ambiguity surrounding the rise of AI, its dynamics and its consequences - attributes similar to climate change - creating an IPCC for AI can help build a solid base of facts and benchmarks against which to measure progress. However, the IPCC is not without flaws, such as the difficulty to differentiate policy-relevant information from policy prescription, the challenge of dealing with uncertainty, and assessing so-called 'grey literature'. This group seeks to advance ways to avoid similar pitfalls. Other governance systems could also be relevant and discussed in light of shared features with AI.

Thought-Starters

- What are the main benefits and pitfalls of the IPCC?
- What are some key similarities and differences between climate and AI governance?
- What should be the role of industry and civil society besides scientific research organizations?
- What other domains might be relevant due to some shared features with AI, such as 'dual-use' potential (e.g. space, biotechnology, nuclear)? What are their benefits and pitfalls?
- What are the advantages and disadvantages of a multilateral framework for governance and what unique issues do you see AI facing in view of that?

Subcommittee B: International Panel on AI: Objectives & Approaches

Chair: Arisa Ema, University of Tokyo

This group seeks to establish the main objectives and approaches for the IPAI. A well-functioning IPAI would presumably help address a diverse range of ethical and safety risks, including the transition from Artificial Narrow Intelligence (ANI) to Artificial General Intelligence (AGI). This could be done by providing a continuous mapping of AI capabilities and orienting scientific research towards globally agreed goals that could include AI accountability, fairness, safety and control capabilities. Participants to this subcommittee will need to determine whether the IPAI should stick to the IPCC's approach on policy, which aims to make "policy-relevant" but not "policy-prescriptive" suggestions (which discusses associated gains and losses from climate change trajectories as well as policy implementation challenges). Once decided, an IPAI would also need to identify quantitative benchmarks and metrics for comprehensive reporting.

Thought-Starters:

- What would be the core motivations for international cooperation and global governance on AI?
- What should be the main objectives of IPAI?
- What would be the core functions of IPAI?
- Should the IPAI be "policy-relevant" or "policy prescriptive"?
- What would be the metrics and benchmarks used and set by IPAI?

Subcommittee C: International Panel on AI: Membership of IPAI

Chair: Anne Carblanc, OECD

This group will strive to make concrete proposals for participation in IPAI, in particular with regards to countries, industry and civil society organizations, and breadth of expertise. IPAI should include a sufficient breadth of expertise to adequately capture the diversity of AI impacts and applications. Sufficient expertise in computer science and machine learning is paramount. But given the systemic complexity, uncertainty, ambiguity and velocity surrounding the rise of AI, the work of a well-functioning IPAI will also tremendously benefit from inputs gathered from a diversity of disciplines (including the humanities) and professional fields. A key criterion for an impactful IPAI may also be to build a solid license-to-operate mechanism and to enhance representation from all sovereign nations, especially under-resourced and historically disadvantaged states.

Thought-Starters:

- Which countries should be part of IPAI? Should IPAI aim for a progressive inclusion of willing and able countries or universal participation from the beginning?
- Which kinds of expertise are most needed to govern the rise of AI and should be reflected in IPAI?
- How can fair and inclusive participation in IPAI be assured?
- How should participation from actors who might fear intrusion from a more powerful governing body be encouraged?
- How do we envision IPAI to interact with existing proposals and initiatives, such as Partnership on AI, IEEE, and many others?

Subcommittee D: International Panel on AI: Designing a global governance of AI framework

Chair: Raja Chatila, IEEE

This group should map and explore the place that the IPAI should have within a broader global governance of AI framework. The IPAI could be a crucial piece in the global governance of AI, providing evidence-based resources that inform various stakeholders, just like the IPCC was for climate change. Coordination between various stakeholders -- the UN and other international organizations, national and local governments, industries and smaller companies, academia, and civil society -- can ensure well-defined roles and improve legitimacy and viability.

Thought-Starters

- What would a global governance of AI framework look like?
- How would the IPAI coordinate with other organizations in the AI space?
- How should policy, governance dynamics and values at the national and international levels be articulated and coordinated?
- Is an IPAI the best institution to build now? If not, are there alternative venues and institutions which could form the centerpiece of the global governance of AI framework?
- What are the gaps in what your organization needs, or in the broader existing governance of AI, that IPAI could help address?

Working Group 3: International Panel on AI & AI for the UN SDGs

February 10th, 14:00-15:15

Expert Group 11: AI for Sustainable Development Goals

Artificial Intelligence ('AI') can be understood as a general purpose technology, which holds the potential to transform many of our current global challenges. Compounded with the urgency of issues faced by humanity today, such as rapid climate change, biodiversity degradation, and widespread global poverty, our problems are becoming increasingly complex in an interconnected global environment. AI technologies present many important existing and potential cases that showcase the power of AI to solve these enormous challenges.

Set by the United Nations, the Sustainable Development Goals ('SDGs')[41] aim to address a wide variety of global challenges faced by humanity such as poverty, climate change, human rights and inequality, among others. The SDGs framework is built upon key objectives of economic development, societal stability, and supporting the Earth's ecosystem over the long term.[42] Each high-level goal is associated with a set of sub-targets that provide greater detail on the indicators that need to be accomplished in order to achieve the overall goal. The SDGs provide a blueprint for governments, companies, and citizens to achieve a more sustainable future for all. Given the many applications of AI technologies, which for the framing of this Expert Group we define as "*big data driven, machine learning, algorithm-centric, socio-technical systems powered by supercomputing*," there are many use cases within current AI capabilities to further the SDGs. Using AI technologies, actors progressing the development goals can become better equipped to handle development challenges and bring amplified advances, at a greater magnitude, towards delivering the 2030 Agenda.

The objective of this Expert Group during the 2019 Global Governance of AI Roundtable is to understand how we can practically forge collaborations to deploy AI to advance the SDGs, how to do this in a safe and ethical manner, and what use-cases we can collectively devise for the specific areas of Education, Healthcare and Climate Change.

[41] Full list of the 17 SDGs available here: https://www.un.org/sustainabledevelopment/sustainable-development-goals/

[42] Griggs, D. 2013. *Sustainable Development Goals for the People and the Planet*. Macmillan Publishers Ltd https://sustainabledevelopment.un.org/content/documents/844naturesjournal.pdf

Subcommittee A: Preparing to Apply AI for SDGs

Co-Chair: Cyrus Hodes, AI Advisor to H.E. UAE Minister for AI

Within the realm of narrow AI technologies, the massive predictive power of these systems enables users to move from a position of analyzing historical data, towards assessing real term data, and increasingly to be able to predict future trajectories. With the rapid penetration of digital, IoT and satellite technologies, increasing reserves of data are being continually produced. Such significant pools of data can be used to train AI algorithms, which could help assess and predict progress towards each of the SDGs.

Nonetheless, to be effective in creating unified progress towards each of the SDGs in all parts of the world, there are several coordination and governance foundations that need to be put in place. Given the magnitude of the challenges covered by the SDGs, many stakeholders across society, such as governments, businesses, academia, think-tanks, and start-ups need to orchestrate robust coordination in using AI. The goal of this subcommittee is to explore what are the preparatory steps that need to be put in place before we can start applying AI for SDGs, and how to build global cooperation on this.

Thought Starters:

- What types of frameworks (legal, ethical, governance etc.) should we establish as a basis for using AI to progress on the SDGs?
- What are the key tensions in using AI to advance the SDGs? What are some practical mechanisms to minimize the downside risks of this?
- How do we build a collective and realistic understanding of where AI is currently at from a technical standpoint, and its abilities to help the SDGs?
- What are some current initiatives to foster global coordination in using AI for SDGs? How can this Expert Group amplify existing efforts?

Subcommittee B: Use-cases and Frameworks for Education

Co-Chair: Baroness Beeban Kidron, UK House of Lords

Existing AI capabilities, such as personalized learning and classroom teacher assistance, have significant implications for redefining our education models. Given the scale and magnitude of our world's growing educational needs, AI technologies can provide a valuable solution to ensuring that high-quality education is delivered to every human, consistent with the 'leave no-one behind' maxim of the SDGs. However, deploying AI for such a use-case also comes with challenges of data, privacy, bias, among others. This Expert Group seeks to understand how we can learn from existing use-cases of deploying AI in Education and simultaneously build governance mechanisms to safeguard citizens across the world.

Thought Starters:

- What are examples of successful cases in using AI to expand access to education around the world?
- What are the key ethical and societal concerns in deploying AI to advance development goals for quality education?
- Which stakeholders in society are responsible for delivering AI technologies to solve our educational challenges?
- How can we foster greater collaboration between the technical community and educators to develop AI solutions for this purpose?
- What is the role of governments in ensuring that AI solutions in education reach all citizens and that there are adequate legal and ethical frameworks for rollout?

Subcommittee C: Use-cases and Frameworks for Health

Co-Chair: Elizabeth Gibbons, Harvard FXB Centre

Artificial Intelligence (AI) and machine learning are becoming deeply embedded into the lives of citizens around the world. Various important use cases, such as medical diagnostics, showcase the many beneficial applications AI has for humanity through greater access to vital services. Through big data driven and machine learning intelligent systems that are capable of efficiently predicting, diagnosing and treating disease, with the same as, or greater than, human efficacy rate, AI beckons a step-change in addressing public health challenges, especially in terms of access, faced by large populations around the world and in emerging economies in particular. SDG3, which is "to ensure healthy lives and promote well-being for all at all ages", calls for reductions in maternal and child mortality, for universal access to sexual and reproductive health and healthcare, and numerous other advances,[43] which could all be accelerated by ethically designed and deployed AI.. However, such potential for innovation in progress toward SDG3 is inextricably linked to downside risks of medical irresponsibility, excessive privatization of value, fairness, trust, bias, ethics and many others. How AI technologies, systems and platforms are designed, developed, deployed, marketed and ultimately consumed will thus impact people around the world. The goal of this subcommittee is to understand how we can learn from existing use-cases of deploying AI for health and simultaneously build governance mechanisms to safeguard citizens across the world.

Thought Starters:

- What are examples of successful cases in using AI to expand access to health around the world?
- What are the key ethical and societal concerns in deploying AI to advance development goals for good health and well-being?
- Which stakeholders in society are responsible for delivering AI technologies to solve the challenges of enabling widespread access to health services and benefits?
- How can we foster greater collaboration between the technical community and the health community to develop AI solutions for this purpose?
- What global cooperation needs to be put in place to ensure that AI solutions in healthcare reach all citizens?

[43] See https://www.un.org/sustainabledevelopment/health/

Subcommittee D: Use-cases and Frameworks for Climate Change & Urban Development

Co-Chair: David Jensen, UN Environmental Cooperation for Peacebuilding Programme

Given the widespread use of satellites, mobile phones, sensors and financial transaction technologies, the digital revolution is generating more information than ever on the state of the planet.[44] It is estimated that there were 1,738 satellites in orbit in 2017, which generated 5,700 scenes per day. This wealth of real time data can have transformative impact on the management of Earth's natural resources and help achieve several SDGs, such as SDG 13: *Climate Action*, SDG 6: *Clean Water and Sanitation* and SDG 7: *Affordable and Clean Energy*.

Satellite imagery powered with AI capabilities can help "design, monitor, and evaluate effective policies that can achieve the SDGs."[45] To find the best responses to these climatic and natural disaster challenges, governments need to have an anticipatory view of disaster zones. Satellite imagery coupled with AI-based systems can enhance our ability to make quick and effective decisions in times of crisis, as well as redirect our resources to where they are most in demand.[46] The goal of this subcommittee is to understand how we can learn from existing use-cases of deploying AI in climate change and simultaneously build governance mechanisms to safeguard citizens.

Thought Starters:

- What are examples of successful cases in using AI to solve for climate change and to foster smart urban development?
- What are the key challenges, technical and social, of deploying AI to monitor and predict changes in our climate?
- Which stakeholders in society are responsible for delivering AI technologies to solve challenges of sustainable climate and sustainable cities?
- How can we foster greater collaboration between the technical community and climate change actors (governments, civil society, activists etc.) to develop AI solutions to curb negative effects of climate change?
- What global cooperation needs to be put in place to ensure that AI solutions for climate change and smart cities positively affect all citizens?

[44] *White paper: Digital Earth: Building, financing and governing a digital ecosystem for planetary data*, *UN Science-Policy-Business Forum on Environment,* Draft 1.2 2018.
[45] *Ibid*
[46] *Anatomy of a catastrophe: Using imagery to assess Harvey's impact on Houston*, retrieved from Planet.org

Working Group 4: Making the AI Revolution Work for Everyone

February 10th, 15:45-17:00

Expert Group 12: AI & Cybersecurity

Goal: To develop concrete governance recommendations about how to (1) ensure digital systems, including AI systems, are built to be resilient to and robust against cyber threats; and (2) manage the use of AI in cyber-attacks and defense.

Subcommittee A & D: Building Rules and Norms for Industry

Chair A: Roman Yampolskiy, University of Louisville
Chair D: Robert Silvers, Paul Hastings LLP

This group focuses on the role of the private sector — including companies, professional bodies, and individual engineers and researchers — in ensuring cyber security in a number of AI-driven contexts.

Thought Starters:

- What responsibility do companies have for protecting large pools of data they are using for AI purposes, for example to prevent data breaches or data poisoning? Who should set these norms and how should they be enforced to ensure accountability?

- How can we encourage responsible disclosure of vulnerabilities and breaches?
    - What responsibilities do industry participants have to share cyber threat information?
    - Case study: Microsoft called on governments to stop hoarding vulnerabilities in the aftermath of 2017 WannaCry cyber-attacks.[47]

- How can actors better identify and share best practices between different domains (e.g. between cybersecurity, computer science and AI)?
    - Are bug bounty programs appropriate for AI-enabled good and services?

- How do we prevent an arms race between offensive and defensive AI-based cyber tools? Or, if such an arms race is inevitable, how do we ensure that defensive capabilities prevail?

[47] Smith, B. 2017. *The need for urgent collective action to keep people safe online: Lessons from last week's cyberattack.* Microsoft.

Subcommittee B: Evaluating Policymakers' Options

Chair: Yann Bonnet, National Cybersecurity Agency of France (ANSSI)

This group addresses AI in the contexts of national and international cybersecurity. Specifically, it aims to help inform, discuss, and weigh options and strategies for policymakers in cybersecurity, noting that there is often a shortage of cybersecurity experts in the public sector.

Thought Starters:

- Is AI a serious threat to cybersecurity? How can actors foster the development of trustworthy AI?
    - What are the vulnerabilities of current AI systems that policymakers should take into account? How to quantify the cost of cyber-attacks?
    - How should explainability and the need to hold an algorithm accountable be balanced with safeguarding the security of the system against hackers?
    - What measures could be taken to maximize the benefits of AI while minimizing its risks? And, in what fora (e.g. standards bodies, UN, OECD)? What is the role of government policy or regulation in cybersecurity, compared to industry self-regulation?
- Is AI a "solution" for cybersecurity? How can we use AI to enable cybersecurity to scale up?
    - How might AI be expected to bring improvements in terms of cybersecurity?
    - How can collaboration be facilitated between industry practitioners, academic researchers and policymakers? What are some "red lines" for cyber-attacks and retaliation? How do we prevent an arms race?
        - Case study: Paris Call for Trust and Security in Cyberspace includes principles to avoid cyber-attacks on critical infrastructure such as electrical grids and hospitals.[48]

[48] France Diplomatie. 2018. *Cybersecurity: Paris Call of 12 November 2018 for Trust and Security in Cyberspace.*

Subcommittee C: Educating & Empowering Users: Individuals, Businesses and Governments

Chair: Lydia Kostopoulos, Digital Society Institute Berlin

This group addresses users of AI and cybersecurity products and services, i.e., the general public. It seeks to develop strategies to inform and equip the broader (non-technical) public about cybersecurity risks and strategies to mitigate them.

Thought Starters:

- What guidelines, advice, or rules of thumb to improve personal cybersecurity are actually effective? How do we know if they work?
    - Common existing advice includes using two-factor authentication, updating software, and not clicking on suspicious links.
- Once the right guidelines for the public have been identified, how do we raise awareness and encourage people to adopt them? What does success look like and how should non-compliance be dealt with?
- What are successful strategies for increasing consumer awareness of product security more generally? Are metrics and security standards for certain kinds of product a feasible idea?
- Are guidelines such as the NIST Cybersecurity Framework useful? How can they be improved?

Relevant working definitions

Cyber attack

- Attempts to "alter, usurp, deny, disrupt, deceive, degrade, or destroy computer systems or networks" and the artifacts connected to these systems and networks.[49]

Artificial intelligence

- The designing and building of intelligent agents that receive precepts from the environment and take actions that affect that environment.[50]

Some existing initiatives

Industry

- Cybersecurity Tech Accord
- IGF Best Practice Forum on cybersecurity
- Proposed *US Active Cyber Defense Certainty Act* which allows private companies to respond to attacks by accessing the attacker's computer to disrupt the attack, monitor the attacker, and delete or retrieve stolen information.
- Proposal by Microsoft for a Digital Geneva Convention, which calls on governments to report vulnerabilities to vendors rather than stockpiling them.

International law

- UN GGE on IT and International Security. Met for the last time in June 2017 and was unable to come up with a consensus final report.
- Tallinn Manual 2.0
- Paris Call for Trust and Security in Cyberspace

Public awareness and education

- NIST Cybersecurity Framework v1.1
- UK Government Cyber Aware program

[49] Lin, H. 2016. *Governance of Information Technology and Cyber Weapons* in *Governance of Dual-Use Technologies: Theory and Practice*. American Academy of Arts & Sciences.
[50] Russell S., Norvig P. 1995. *Artificial Intelligence: A Modern Approach*. Prentice Hall.

Working Group 4: Making the AI Revolution Work for Everyone

February 10th, 15:45-17:00

Expert Group 13: Managing the Economic and Social impact of the AI Revolution

AI-enabled automation of human labor has the potential to disrupt jobs that people previously thought were not automatable. Indeed, unlike past waves of technological automation, machines equipped with AI are increasingly capable of automating high-skill, cognitive, and even creative tasks. The transition will most likely create losers and winners. What approaches should policy-makers take? There are many proposals to reform national social security systems such as a universal basic income (UBI) or labor market policies including portability of benefits. Other proposals target education reform, including new curricula and new models for re-training and upskilling workers or for lifelong education. Given the myriad of ideas and the continued uncertainties about the net impact on jobs and timelines, what should policy-makers do? What are viable responses to this potentially transformative shift towards a society less dependent on human labor?

Existing Initiatives

NB: This is only a small selection of the many initiatives that exist already worldwide.

- Washington Future of Work task force, which studies trends that might drive transformation, including automation.
- Asilomar AI principle #15 of "Shared Prosperity": The economic prosperity created by AI should be shared broadly, to benefit all of humanity.
    - This is endorsed by the US State of California.
- OECD's various initiatives on AI and the Future of Work:
    - Expert group AIGO
    - Policy Observatory (2019)
    - Research on the Future of Work
    - Conference and report on "AI: Intelligent Machines, Smart policies" (October 2017)
    - Work on the future of education
- Initiatives to facilitate the digitalization of work
    - France's "Loi Travail" for modernizing social dialogue and securing professional careers.
    - United Kingdom's Taylor Review of Modern Working Practices

Working Definitions

To avoid spending too much time discussing key terms and definitions, GGAR participants have drafted the following as working definition(s) for the purpose of discussion in this expert group:

Artificial Intelligence:

- A range of methods relying on algorithms at their core to learn and adapt, improving their models based on new data.

Technological unemployment:

- "Unemployment due to our discovery of means of economizing the use of labor outrunning the pace at which we can find new uses for labor."[51]

Displacement effect:

- Substitution of technological means for workers in a company, an industry, or in general in the economy.
- Consequence of automation that have a negative impact on employment.

Productivity effect:

- The creation of novel occupations, jobs or tasks associated directly or indirectly with the use of new technological means in the company, industry or in general in the economy.
- Consequences of automation that have a positive impact on employment.

Universal Basic Income (UBI):

- A guaranteed and unconditional cash transfer paid to all citizens, regardless of employment status or wealth.
- Usually transfers are an amount that is enough to cover basic needs and ensure citizens can live above the poverty line.

Universal Basic Dividend:

- Collects returns from companies that use AI capital into a national fund and distributes it to all citizens.
- Similar to a sovereign wealth fund for AI revenues.

Portable Benefits:

- Social benefits (pensions, parental leave days, insurance, etc.) that accrue and carry across employers and employment status.

[51] Keynes, J.M. 1930. *Economic Possibilities for our Grandchildren*, in Keynes, J.M. 1931 *Essays in Persuasion.* London: MacMillan & Company

Subcommittee A: Radical Disruption

Chair: Calum Chace, Author

This subcommittee explores the scenarios wherein the AI revolution could result in most of the workforce being made redundant, presenting a radical disruption in our society. In such a scenario, AI and robotics enable automation of most human jobs. Among others, this subcommittee addresses questions of feasibility and the desirability of Universal Basic Income, how to fund it (e.g. taxes on capital or companies), 'radical abundance,'[52] and how stakeholders can best prepare for radical disruption in labor markets.

Thought-Starters:

- What are the key economic and social risks to mitigate in a situation of large scale jobs disruption? (e.g. inequality, poverty, political extremism, lack of social cohesion)
- What are policy options to mitigate this disruption? (e.g. UBI, radical abundance)
- Is there a smart way to manage periods of social transition?
- What are the benefits of and challenges for these policy options?
- Which policies or combinations of policy are feasible and could be most impactful? How does this vary across geographies and political economies? Are there countries in which a specific policy solution would be unfeasible?

[52] A significant increase in economic wealth enabled by the "radical" decrease in production costs for most goods and services due to technological progress.

Subcommittee B: Gradual Disruption

Chair: Marek Havrda, GoodAI

This subcommittee explores the scenario wherein the AI revolution could drive gradual change in the economy and in labor markets, and where structural unemployment is only moderately affected. In this case, AI and robotics enable automation of human jobs at a gradual or moderate pace. Among other topics, this subcommittee addresses policy responses including education reform and retraining of workers, adjustments in social and labor policy, and how the creation of novel occupations can best be stimulated.

Thought-Starters:

- What are the key economic and social risks to mitigate in a situation of gradual job disruption(e.g. inequality, poverty, political extremism, loss of meaning, social cohesion)?
- What are possible policy options to mitigate this disruption (e.g. education & re-training, labor market policies, social welfare policies, innovation & entrepreneurship)?
- Is there a smart way to manage a social transition period?
- What are the benefits of and challenges for these policy options?
- Which policies or combinations are feasible and could be most impactful? How does this vary across geographies and political economies? Are there countries in which a specific policy solution would be unfeasible?

Subcommittee C: Educating for AI

Chair: Priya Lakhani, CENTURY Tech

This subcommittee deep-dives into questions of education, retraining, and human capital for the AI revolution. Among others, it explores new models of retraining, education reforms and novel approaches to building human capital, that would enable the workforce and societies to benefit from the AI revolution. It aims to provide practical recommendations for how stakeholders can implement these changes.

Thought-Starters:

- What skills and competencies are especially needed for the age of AI?
- What is the role of STEM as opposed to soft skills, ethics and community building?
- How do we reform outdated education systems? What are the challenges and opportunities in educating for AI?
- What are strategies or collaborative approaches for governments, private sector, nonprofits and industry to improve education & skills training outcomes? What could be the role of public private public partnerships ("PPPPs")? For example:
    - German Baden-Wurttemberg Learning Factories 4.0 Initiative
    - The network of schools "Bridge International Academies in Kenya
    - Airbus Corporate Academy for Engineers of the Future
- What is a practical next step this committee can take to move forward the AI and Education agenda?

Subcommittee D: Mitigating Rising Inequality

Chair: Irakli Beridze, UN Centre on AI and Robotics

This subcommittee deep-dives into questions of economic inequality and social (in)cohesion resulting from the AI revolution. It discusses approaches to income redistribution and tackling consequences of inequality such as criminality, mass migration and political extremism. It also aims to provide practical recommendations for how stakeholders can implement these changes.

Thought-Starters:

- In which countries or contexts could AI-enabled automation lead to a particularly steep rise in inequality? How could this be seen within and across countries?
- What are consequences of rising inequality that may be unique to or especially endogenous to the AI revolution (e.g. crime, migration, political extremism, social cohesion)?
- What are policy strategies to mitigate inequality, at national or international levels?
- What reforms to social security systems are needed to alleviate the social distress from the AI revolution?
- How could AI itself be used, in positive or negative ways, to deal with inequality and its effects?
- What is a practical next step this committee can take to move forward the AI and Equality agenda?

Working Group 4: Making the AI Revolution Work for Everyone

February 10th, 15:45-17:00

Expert Group 14: AI Narratives

The definition and trajectory of Artificial Intelligence (AI) is deeply embedded in prevailing narratives and imaginaries of technology. The trust that citizens put in AI is also influenced by such narratives: shifts in our human values influences policy, which impacts how we manage developments in AI. There exists significant heterogeneity in perception and trust of AI technologies across the world: in some regions, AI is seen as an opportunity, while in others it is perceived with significant skepticism and fear. One narrative that has emerged frames AI as a tool for 'Good', which is reflected in the UN Sustainable Development Goals. Contrary to other narratives, this one claims a unifying force for the world's nations, including for developing countries, but like any narrative it carries shortcomings and potential for exclusions. The purpose of this Expert Group is to evaluate current narratives of AI, understand and build pathways to overcome underrepresented narratives, and assess how technology narratives can help inform the policymaking and technical community in AI.

Working Definitions

- **Narrative**: "narrative texts, images, spectacles, events; cultural artefacts that tell a "story"."[53]
- **Socio-technical imaginaries:** Collective, public and institutionally stabilized visions of possible futures, which are driven by shared understandings of social life and order. These common understandings are attained through advances in sciences and technology.[54] An imaginary frames a projection, symbol and associated belief about a technology, not only in an individual's mind but also across peoples and society. Such a framework is useful in analyzing and accounting for power dynamics and issuing ethical and inclusive policy.

[53] Bal, M. 2009. *Narratology: Introduction to the Theory of Narrative*. University of Toronto Press.

[54] Jasanoff S. and Kim, S.H. 2015. *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, University of Chicago Press.

Subcommittee A: AI Narratives: Underrepresented Narratives

Chair: Sarah Dillon, Leverhulme Centre for the Future of Intelligence

Socio-technical imaginaries, and the narratives built from these, have a real and profound influence over innovation and technology, as well as their perception, adoption and policy. As technology narratives are co-created based on public opinion, mutual understanding, and cultural values, they are influential in how the development of AI is imagined and framed, and in turn how it unfolds. As such, science, technology and society operate in a dynamic process of co-production and are continuously shaped and evolved by each other.[55] Therefore, individuals, groups, values, and geographies that shape prevailing technology narratives are key in understanding the essential values, or set of values, that humans want to preserve as technology progresses. This subcommittee interrogates some prevailing AI narratives, which are often over-influenced by Western dystopian or utopian imaginaries of AI and seeks to explore broader narratives about technology from diverse perspectives.

Thought-Starters:

- In what ways does technology influence narratives, and narratives influence technology, as a continuous feedback loop?
- How are narratives nested in geopolitical, economic, and political relationships?
- Are negative narratives of AI a concern? Why?
- What can we learn about underlying values or fears that are portrayed in popular stories projecting our technological futures?
- What are some underrepresented narratives of AI across people, cultures, geographies, and time, and how can we bring those to greater attention within the public sphere?

[55] Jasanoff, S. 2004. *States of Knowledge: The Co-production of Science and Social Order*. New York: Routledge.

Subcommittee B: AI Narratives: A tool for policymakers

Chair: Casper Klynge, Technology Ambassador of Denmark

Imaginaries, and the narratives that follow, serve as an important tool for policy-making and its legitimization because they structure a common and collective understanding between diverse stakeholders.[56] This can help build stakeholder buy-in for policies implemented by the government because such governance models are grounded in broader collective public perceptions about technology. Projections of societal futures have proved useful in identifying and deploying new investment by national actors in key areas of science and technology, which in turn solidify states' responsibility to act as stewards of their citizens' visions. The goal of this sub-committee is to unearth how policymakers can harness the power of AI narratives to inform policy-making that maximizes the upsides of technology, minimizes the downside risks, and fosters shared trust in AI among citizens.

Thought-Starters:

- How do policymakers currently use technology narratives to build governance mechanisms?
- Which narratives have been successful in garnering public trust in AI technologies?
- How is a 'technology ambassador' involved in shaping public narratives of AI?
- What lessons can we learn from Denmark on building trust between technology and citizens through deploying AI narratives?

[56] Taylor, C . 2003. *Modern Social Imaginaries*. Duke University Press.

Subcommittee C: A/IS Infrastructure and Ecosystem

Chair: Mark Halverson, Precision Autonomy

This subcommittee adopts a use case driven approach to identifying infrastructure that can accelerate the safe adoption of Autonomous Intelligent Systems. It seeks to draw corollaries with previous technology adoption cases to highlight infrastructure and ecosystem requirements that are essential to or beneficial for fostering safe AI adoption. The subcommittee will focus on drone taxi services[57] as the use case example for A/IS infrastructure and ecosystem.

Thought-Starters:

- Are policy-makers equipped to address risks posed by drone taxi services?
- What ecosystem participants need to be involved in order to accelerate safe adoption (e.g. the insurance industry, eVTOL Manufacturers, enforcement agencies)?
- Should infrastructure for drone taxi services be controlled by traditional government entities (e.g. Dubai Civil Aviation Authority, Dubai Police Force), or facilitated via public-private partnerships with greater expertise in A/IS?
- Where should infrastructure investment be channeled to accelerate?
- How can this case-driven narrative creation exercise help in policy-making for A/IS infrastructure and ecosystem?

[57] CB Insights Research. (2019). *How Drones Will Impact Society: From Fighting War to Forecasting Weather, UAVs Change Everything*.

Subcommittee D: Building Trust in AI Systems

Chair: John P. Sullins, Sonoma State University

Building trust is essential for safe adoption and deployment of AI technologies. Without building trust among stakeholders, be that citizens, consumers, government, technical community, academics, among others, the uptake of AI technologies will be severely hampered. As our world becomes more used to the rapid proliferation of technology in our daily life, the balance of trust and over-trust will continuously evolve. Trust between humans and machines and trust between humans as mediated through AI technologies change from a generational perspective and across different AI applications. This subcommittee will focus on examining the new kinds of trust that AI technologies ask us to consider.

Thought-Starters:

- How have we seen shifts in generational perspectives regarding human-machine interaction? What similarities and differences in perspective and cohesion have emerged between children, young adults and parents, or parents and grandparents?
- Does the younger generation hold too much implicit trust in technologies or do they hold less trust in technologies compared to their parents? How can we engender a balanced and critical perspective, while allowing the younger generation critical freedom and independence of thought in a technologically driven world?
- What are cultural differences in trusting AI technologies that we can observe in the world?
- What are different types of ‘trust’ we need to consider when it comes to AI technologies?
- Do we need a new ethical, legal, and/or policy definition of ‘trust’ to capture this new AI influenced global milieu? What would that look like?
- How can policy-makers manage the complexity of varying levels and types of trust for AI applications? What policy levers can be employed to increase human-machine trust? When should they or should they not be used?