OECD Expert Group on AI Futures – Meeting 1 (13 July 2023)

Context – Goal and Scope of the Expert Group

The OECD.AI Policy Observatory and OECD Strategic Foresight Unit convened the inaugural meeting of the Expert Group on Artificial Intelligence (AI) Futures on 13 July 2023, which is a core component of an emerging OECD workstream on AI Futures. The overall background and context for the group can be found in the expert group concept note.

The new Expert Group is composed of **54 experts** from at least **21 countries** representing government, private sector, academia, and civil society covering critical subject matter areas including healthcare, education, robotics, human cognition, defence, philosophy, and ethics. The group is composed of about 1/3 women, which the OECD will strive to strengthen over time. The group is unique in that, while there are many efforts focusing on the present-day opportunities and challenges of AI, including other expert groups of the OECD.AI Policy Observatory, there are few in an international context focused on the future(s) of the technology.

The Expert Group is led by three co-chairs:

- Stuart Russell, Professor of Computer Science at the University of California, Berkeley and Director of the Centre for Human-Compatible Artificial Intelligence.
- Francesca Rossi, IBM Fellow and AI Ethics Global Leader.
- Michael Schönstein , Head of Strategic Foresight and Analysis, German Federal Ministry of Labour and Social Affairs.

The full composition of the Expert Group is available here. The list of participants for the inaugural meeting can be found below.

Introduction

The session was run under the Chatham House Rule. It was kicked off by Karine Perset, head of the OECD.AI Policy Observatory, who provided a general overview and context regarding the expert group and the purpose of the meeting. The overview covered the role that the OECD plays in fostering beneficial progress in the AI policy field, including driving progress in the implementation of the OECD AI Principles.

As this was the first meeting of the expert group, the discussion focused on introductions and building rapport among the expert group members, as well as discussing members' thoughts on the initial vision for the group and what they would like to see it achieve. In addition, the OECD briefed the members on a potential questionnaire to help prioritise the group's future efforts, and to inform OECD research on AI Futures, including a stocktaking report currently under development. The co-chairs facilitated the discussion items listed below.

Proposed vision for the expert group

Experts were presented with a proposed vision for the expert group, which had been developed based on feedback from the OECD Working Party on AI Governance. In addition to the concept note, the vision for the Expert Group included:

- **Objective:** to assess potential future AI milestones, benefits, risks, and solutions to help policymakers take action today to shape them.
- **Next steps:** The Expert Group will undertake the following steps.
  - Stocktaking: Conduct a stocktake of potential future AI risks, benefits, and policy approaches.
  - Prioritisation: Prioritise which risks, benefits, and policy approaches from the stocktake will be the focus of the Expert Group for the coming year.
  - Research and scenario development: Summarise research and develop scenarios exploring the highest priority AI futures issues, to determine what desirable and undesirable futures may look like.
  - Assess governance: Consider the current state of managing high-priority risks, identify gaps, and propose solutions.
  - Recommendations: Identify priorities and recommendations for shaping positive AI futures, to be documented for governments' review.
  - Future priorities: Discuss next steps for action on these recommendations and identify future group priorities.
- **Outside engagement:** OECD will leverage **inclusive working methods** to provide a voice to other stakeholders, including those who may be impacted by AI, to feed into and supplement Expert Group discussions.

Experts' thoughts on the vision presented for the group

Experts showed significant general agreement and support for the proposed vision of the Expert Group, as outlined in the concept note and presentation. Specific points that were particularly praised consisted of:

- **The exploration of both positive and negative futures.** The analysis of future impacts of AI as one focus of the group was praised as being quite unique with regard to international engagement in the field.
- **The diversity of perspectives.** The group reflects heterogeneity by the nature of its composition as well as by its aims. Many experts highlighted how geographical, gender, and expertise-related diversity is key to achieving a comprehensive vision of the impacts of AI in the near to long-term future.
- **The use of scenario techniques.** Reflecting on issues via strategic foresight techniques, such as scenario exploitation, was seen as a useful methodology to broaden the perspectives of the Group and explore potential futures in new ways.

However, there have also been remarks about the vision and final goal of the group that are worth pointing out:

- **Balancing three different priorities.** Some speakers highlighted the importance of balancing three different priorities when working to produce significant output: the use of AI, its adoption by businesses and citizens, and the minimisation of its harms.
- **Ensuring any policy solutions are globally inclusive.** Others pointed out the importance of endorsing an inclusive vision and avoiding the creation of an asymmetry between the OECD as a

standard maker and developing countries as standard takers, as these might endorse different governance approaches which are worth being taken into consideration.

Experts' views on what the group should seek in order to positively impact AI futures

The co-chairs then asked members to present their personal views on the goal that the group should aim to achieve to positively shape AI's impact on the future of society. Experts shared insights that can be categorised by: (1) views characterised by general agreement, (2) topics characterised by divergence of opinion and (3) other significant insights.

Points of convergence

Experts' opinions tended to converge on some key issues which are reported below. Particular points of agreement that should characterise the Expert Group's aims concerned:

- **The focus on evidence to inform policymaking and demystify AI.** Experts generally agreed on the need to explore future plausible AI scenarios, as well as to create a vision of both the potential future risks and benefits arising from the increasing use of AI in society. In this context, given the fast pace of AI technology development, it is paramount to provide policymakers with potential future trajectories so that they can be prepared to react promptly and effectively in the present. Empirical evidence for the impacts of technology on societal values plays a pivotal role in creating a shared understanding and in dealing with different perspectives.
- **The recognition and respect for diverging views as opposed to the reach of a forced consensus.** Given the nature of the Expert Group, characterised by diversity in terms of expertise and points of view on AI-related topics, it is essential to recognise and respect the experts' individual insights and points of view, as opposed to limiting novel ideas to produce moderate unified consensus. Experts agreed that such sentiment should be recognised in both meeting discussions, as well as eventual products and outputs.
- **Accelerating the development of international agreements to manage current and future AI harms.** As risks in the field are unpredictable but projected to pose substantial harm, it is paramount to make significant investments in order to understand them and effectively mitigate them by means of international co-operation. The group should aim to create a set of future-focused principles, good practices, norms and standards that have practical and global support.
- **Developing potential technological trajectories to test regulation.** Contrarily to the case of the internet, advances in AI are being carefully monitored from a regulatory perspective. Some experts thought that there is a tendency to regulate too early and fast on hypothetical harms. Hence, society and policymakers need to expect possibilities and trajectories, to be ready to act promptly without preventing benefits from arising. In this context, plausible future scenarios could be developed by the group to test current and potential future regulations.
- **Addressing the lack of understanding of exponential situations and AI developments in the population**. Exponential trends such as the exponential increase in computational power used to produce leading AI systems are generally hard to manage and predict. Experts suggested that the group could come up with a set of questions to guide both the general public and policymakers into thinking more rigorously about the potential future AI benefits, challenges, and solutions. In this sense, it might be possible to derive lessons from other transformative technologies, such as the development of electricity, nuclear technologies, the internet, and synthetic biology. Public upskilling was also mentioned as an essential approach to tackling the issue.

- **Avoiding duplication of efforts**. A number of experts agreed on the existing plethora of groups advising governments and the resulting contrasting inputs produced. Consequently, the group should focus its efforts on facilitating good governance by collecting and synthesizing pre-existing work before taking action to pursue new workstreams and products. It is paramount to avoid duplication of efforts and use the best work developed so far as a robust starting point.
- **Focusing on social impacts**. Experts generally agreed to include a substantial focus on mapping out social and psychological changes that are arising from AI deployment in society. These are often overlooked in the literature, despite their crucial importance in defining wellbeing in our society.

Points of divergence

There were also some areas of divergence where experts expressed differing preferences on courses of action, or where their perspectives seemed to differ more generally. This consisted of:

- **Open source**. There were some differing opinions on open-source AI systems and models. While open source was highlighted as an effective way to counteract the tendency to lock up foundational models through proprietary protections, others highlighted the potential for misuse and accidents if there were insufficient controls on powerful models.
- **Policy recommendations**. On one hand, some experts believe the group should aim to work towards the development of a set of forward-looking policy recommendations, anticipating their concrete needs. On the other hand, others claim that before deciding on a set of policies, the group should focus on the big-picture values that will serve as a foundation for policymaking.
- **Conditional claims and scenario planning.** Some experts see scenario planning and conditional claims as valuable tools. However, others pointed out the importance of not limiting the exploration to just scenario planning. According to them, it might be interesting to look at particular narrative futures, for instance the regrowth of scientific productivity or the impact of artificial agency on human autonomy.
- **Greater divergence in the AI community**. Comments by some of the experts reflected broader divergence in the AI community, including differences in views related to the potential for Artificial General Intelligence (AGI), what impacts it may have (e.g., existential risks), and the extent to which research and dialogue should focus on such longer-term considerations relative to more immediate concerns on narrow AI impacts. As touched on in the convergent points above, however, the expert group is in agreement to approach, discuss, and cover all relevant points of view.

Other insights

Other important insights concerning the desired role of the group, shared by Expert Group members, are reported below. These represent views that were mentioned by individual speakers and that were not explicitly characterised by broader convergence or divergence among members.

- **Recognise the pivotal role of incentives.** Instead of focusing on funding and regulation alone, it will be important to develop other actionable incentives to achieve beneficial futures.
- **Consider regulation as an empowering tool.** Regulation should be treated as a useful tool to incentivise and foster positive AI developments, rather than just be seen as a “negative” instrument to forbid and slow down technological progress. It should be seen as a way to empower democracy to avoid a dependency on non-transparent and centralised large companies.

- **Focus on the social domain to find solutions.** Although the risks posed by AI systems can be technical, it will be valuable to explore solutions in the social domain (for instance, by analysing trust networks in the context of misinformation), instead of focusing on AI solutions alone.
- **Create a taxonomy to refine the scope of the discussion.** It was pointed out how the creation of a taxonomy to better categorise AI risks and benefits, along with the development of responsive mechanisms to adapt rapidly to technological changes, would represent a significant achievement in the field.
- **Articulate key areas of trade-off.** It was suggested to create a clear framework of analysis to better understand the trade-offs that can arise in different sectors of society concerning the beneficial and negative impacts of AI.
- **Narrow down existing divergencies in the existential risks field.** By closing the gap in opinion concerning the likelihood of materialisation of existential risks, it would be possible to drive effective policy action.

Expert Group members who were not able to join had a chance to send their written thoughts on the two questions. Important insights concerning this second question on the desired achievement of the Expert Group comprised:

- **Focus on tangible progress in the international policy domain**. To show the early significance of the group's effort is important to deliver tangible progress on a specific and achievable goal. With developments in foundation models/generative AI, it is essential to focus on some specific goals that will position the group to meet the future impacts that are already characterized in current scholarship.
- **Promote progress that is foundational in the near term and positioned for long-term success**. The group will need to update its work to the timeline of key events and progress that will characterise the policy ecosystem and reach a consensus about where its focus will have the most valuable impact.

List of Participants

Anthony Aguirre - Co-founder at Future of Life Institute

Markus Anderljung - Head of Policy at Centre for the Governance of AI

Rebecca Anselmetti - Senior Policy Advisor at Department for Digital, Culture, Media and Sport

Carolyn Ashurst - Senior Research Associate in Safe and Ethical AI at the Alan Turing Institute

Azeem Azhar – Founder of the Exponential View newsletter

Joscha Bach - Principal AI Engineer at Intel Labs

Amir Banifatemi – Chief Innovation & Growth Officer - XPRIZE

Yoshua Bengio - Founder and Scientific Director at MILA, Quebec AI Institute

Jamie Berryhill – AI Policy Analyst at OECD

Duncan Cass-Beggs - Strategic Foresight expert at Independent

Rumman Chowdhury - Responsible AI Fellow at Harvard University's Berkman Klein Center

Mariano-Florentino (Tino) Cuéllar - President and CEO at Carnegie Endowment for International Peace

Vilas Dhar - President and Trustee at Patrick J. McGovern Foundation

Virginia Dignum - Professor at Umea University

Pam Dixon - Founder and Executive Director at World Privacy Forum

Dexter Docherty - Foresight Analyst at OECD

Charles Fadel - Founder & Chairman at Center for Curriculum Redesign

Daniel Faggella - Head of Research, CEO at Emerj AI Research

Rebecca Finlay - CEO at Partnership on AI

Marko Grobelnik - AI Researcher at Artificial Intelligence Lab at Jozef Stefan Institute.

Sebastian Hallensleben - Head of Digitalisation and AI at VDE Association for Electrical, Electronic & Information Technologies

Juha Heikkilä - Head of Unit, Robotics, Directorate-General for Communication Networks, Content and Technology at DG CONNECT, European Commission

Hamish Hobbs - Policy Advisor at Longview and to the OECD Strategic Foresight Unit - Longview

Takayuki Honda - Assistant Director at Ministry of Internal Affairs and Communications

Eric Horvitz - Chief Scientific Officer at Microsoft

Holden Karnofsky - Director of AI Strategy at Open Philanthropy

Ziv Katzir - Head of the National Plan for Artificial Intelligence Infrastructure at Israel's Innovation authority

Rafal Kierzenkowski - Senior Counsellor for Strategic Foresight at OECD

Hiroaki Kitano - Senior Executive Vice President and CTO at Sony Group Corporation

Daniel Leufer - Senior Policy Analyst at Access Now

Jade Leung - Governance Lead at OpenAI

Nicklas Lundblad - Head of Global Policy and Public Affairs at Deepmind

Laurens van der Maaten - Senior Research Director at Meta AI

Aaron Maniam - Fellow of Practice and Director, Digital Transformation Education at Blavatnik School of Government, University of Oxford

Fernando Martínez-Plumed - Associate Professor at Technical University of Valencia

Jason Matheny - President & CEO at RAND

Sarah Myers-West - Managing Director of the AI Now Institute

Clara Neppel - Senior Director at IEEE European Business Operations

Sean Ó hÉigeartaigh - Interim Executive Director of CSER (Centre for the Study of Existential Risk) at University of Cambridge

Toby Ord - Senior Research Fellow at University of Oxford.

Asu Ozdaglar - Mathworks Professor of Electrical Engineering and Computer Science (EECS) at the Massachusetts Institute of Technology (MIT)

Karine Perset - Head of AI Unit and OECD.AI, OECD Digital Economy Policy Division - OECD

Andrea Renda - Senior Research Fellow and Head of Global Governance, Regulation, Innovation and the Digital Economy (GRID) - Centre for European Policy Studies

Francesca Rossi - IBM Fellow and AI Ethics Global Leader

Stuart Russell - Professor of Computer Science at the University of California, Berkeley and Director of the Centre for Human-Compatible Artificial Intelligence

Michael Schönstein - Head of Strategic Foresight and Analysis, German Federal Ministry of Labour and Social Affairs.

Osamu Sudoh - Professor at the Graduate School of Interdisciplinary Information Studies (GSII), The University of Tokyo

Helen Toner - Director of Strategy and Foundational Research Grants at Center for Security and Emerging Technology (CSET)

Conrad Tucker - Professor of Mechanical Engineering at Carnegie Mellon University

Effy Vayena - Professor of Bioethics and Deputy Head of Institute of Translational Medicine at Swiss Federal Institute of Technology Zürich (ETH Zürich)

Toby Walsh - Chief Scientist of UNSW.AI

Jess Whittlestone - Head of AI Policy at Centre for Long-Term Resilience

Rose Woolhouse - Head of International at Office for AI

Katharina Zweig – Professor and Researcher at TU Kaiserslautern