

OECD Expert Group on AI Futures – Meeting 5 (25 June 2024)

Background

Immediately following the ninth meeting of the OECD Working Party on Artificial Intelligence Governance ([AIGO](#)), the [OECD.AI Policy Observatory](#) and the OECD Strategic Foresight Unit ([SFU](#)) convened for the fifth meeting of the [Expert Group on AI Futures](#) on 25 June 2024. The expert group is a core component of the OECD workstream on [AI Futures](#).

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 General Digital Policy - Federal Chancellery of Germany.

For this meeting, [Sebastian Hallensleben](#); Head of Digitalisation and AI - VDE Association for Electrical, Electronic & Information Technologies, and co-chair of the [OECD Expert Group on AI, Risk & Accountability](#); took the place of Michael Schönstein, who was unable to attend.

The full composition of the Expert Group is available [here](#).

Introduction and context

This meeting was held in hybrid format immediately following two days of meetings of AIGO. While the meeting discussions were held mainly among members of the Expert Group on AI Futures, AIGO delegates and other members of the broader OECD Network of Experts on AI ([ONE AI](#)) were welcome to attend and participate. The meeting was held under the [Chatham House Rule](#).

Figure 1. Expert Group members, AIGO delegates and invited experts in attendance



[Karine Perset](#), head of the OECD.AI Policy Observatory, began the meeting by welcoming attendees and providing brief opening remarks regarding the context of the meeting, including its relation to the AIGO meetings that had just concluded and the blended composition of attendees, as discussed above.

Following Karine's introduction, Francesca Rossi took the floor to guide attendees through the meeting's agenda, which focused on:

1. A recap on the Expert Group's first year of activities ("year 1"), along with an update regarding ongoing work.
2. Discussion on potential approaches and priorities for the second year of Expert Group activities ("year 2").
3. Discussion on potential "deep dive" topics for future work.

Attendees were encouraged to share additional thoughts via email at ai@oecd.org if unable to contribute fully during the meeting or if they had new thoughts afterwards. This summary comprises both the meeting as well as any subsequent input received.

The co-chairs provided presentations and facilitated discussions on the agenda items, as detailed below.

Year 1 recap and update on ongoing activities

Francesca Rossi remarked that the Expert Group was approaching a significant milestone: its one-year anniversary. She also remarked that, since being [announced](#) in July 2023, the Expert Group has grown to **more than 65 leading AI experts** with diverse backgrounds, convened five official [meetings](#) (including the present meeting), and conducted a variety of asynchronous activities. With this, the time was right to reflect on accomplishments to date, lessons learned, and thoughts on the future of the group.

In providing a **recap of accomplishments to date**, Francesca provided a presentation and walked through the Expert Group's efforts to:

- Envision and articulate **desirable and undesirable futures** to orient Expert Group work, and engage in scenario exploration exercises to determine potential trajectories to achieving positive futures and potential actions to help steer away from negative ones.
- Through research, discussion, and a survey of Expert Group members, identified **21 potential future AI benefits, 38 risks, and 66 policy options** to help seize the benefits while mitigating risks.
- Develop a report that prioritises the **top 10 most important items each of identified future AI benefits, risks, and policy options**, as well as a gap analysis of current efforts to identify where policy actions have already begun and where potential gaps may remain.
- Conduct an effort to **analyse the wide variety of governance analogies** (e.g., governance of nuclear energy, biological weapons, etc.) often made about AI to help identify **which aspects of these analogies may or may not fit with AI** in order to derive lessons for AI governance.
- Contribute to efforts of the OECD SFU to conduct a **collective forecasting exercise** to identify the probability of potential future AI developments to help inform AI policymaking.

In providing an **update on ongoing activities**, Francesca invited Expert Group members to send their feedback and comments on the draft report, which the OECD Secretariat will incorporate prior to AIGO review and eventual publication. She also informed the members that a draft paper on AI governance analogies was underway, which they can soon expect for review.

Francesca then passed the floor to [Rafał Kierzenkowski](#), OECD Senior Counsellor for Strategic Foresight, to provide an update on SFU's collective forecasting exercise on AI Futures, developed with the support of the forecasting company [Hypermind](#). Informed by expert insights, the exercise aims at developing high-level probabilistic forecasts of potential future AI benefits and risks—as identified by the Expert Group—given significant uncertainties and complexity. The resulting estimates would help provide a basis to assess and understand the diversity of views among experts, with experts' results compared with a crowd of “superforecasters”—individuals who are not necessarily AI experts but have a solid track record in performing in this type of forecasting effort. As this methodology is new to the OECD, the exercise serves as an experimental initiative to evaluate the methodology's utility in informing AI governance approaches. Rafał informed Expert Group members that they would receive an invitation to participate in July.

Rafał passed the floor back to Francesca, who transitioned the meeting to the next agenda item.

Potential approaches and priorities for year 2

Presentation on initial ideas from Expert Group co-chairs the OECD

Francesca Rossi resumed by outlining preliminary proposals for the Expert Group's second year of collaboration from the co-chairs and OECD Secretariat. Building on the achievements of the group's inaugural year, which focused on mapping and prioritising the landscape of potential AI futures and policy actions, the proposals aimed to refine the focus of the Expert Group and deepen its impact in the second year. Proposals for future work included:

- **“Deep dives”** on specific topics of importance to group members, which could focus on a variety of different areas.
- **Assessing specific potential future technological developments.** For example, possible developments related to increasingly autonomous AI agents based on large language models (LLMs).

- **Future-proofing specific policies.** For example, identifying and implementing AI risk thresholds as a topic of interest.
- **Assessing specific potential future AI impacts.** For example, AI technologies' potential future impact on invasive surveillance.
- Examination of **areas of disagreement** among Expert Group members to help policymakers understand different angles. For example, shedding light on different sides of debate around the potential "loss of control" scenarios with future advanced AI systems.
- Diversifying the **types of outputs** from the Expert Group. For instance, formal OECD publications for areas where aligned thinking and the full buy-in of OECD members adds the most value, and other outputs such as blog posts or working papers when they can add the most value through rapid dissemination and discussion of exploratory or deliberative content.

Overall, Francesca emphasised that the proposed approach emphasised a strategy of "*narrow, deep, and unique*" – working in areas where the Expert Group and the OECD have a comparative advantage to add value in AI policy and governance. Francesca concluded by inviting feedback on these initial proposals, emphasising the openness of the co-chairs and OECD Secretariat to alternative approaches.

Francesca passed the floor to fellow co-chair Stuart Russell and acting co-chair Sebastian Hallensleben to facilitate discussion among meeting participants.

Discussion among meeting participants

[*Note:* The session agenda included separate discussions on "potential approaches and priorities for the second year" and "potential deep dive topics for future work". During the session, it was noted that there was some conceptual overlap in the agenda items and the subsequent discussion on them. As such, the summary below covers both agenda items.]

Expert Group co-chair Stuart Russell encouraged participants to consider areas where their expertise can uniquely contribute to AI policy and governance, building on year 1 discussions, surveys, and other activities on future risks, benefits, and policy options. He invited innovative ideas that had not yet been explored, emphasising the group's mandate to look beyond immediate issues and focus on future possibilities. Stuart and acting co-chair Sebastian Hallensleben then facilitated discussion among participants.

In general, participants expressed satisfaction with regards to Expert Group's first year of activities, as well as a desire to continue and complete the efforts underway that were presented by Francesca Rossi. Expert Group members also expressed appreciation with regard to the group's ability to discuss and collaborate on a variety of topics, including those that are controversial within the Expert Group and AI community alike, in a constructive and professional manner. There was also general agreement that that the Expert Group on AI Futures should stick to subjects and angles that are *future-oriented* and avoid focusing on areas that are set on the present or very near-term, which may fit better in other OECD expert groups and teams. Finally, participants also cautioned that the Expert Group should avoid overlap and duplication with the efforts of other organisations and groups, which may become more challenging with the advent of AI Safety Institutes in a growing number of countries.

Meeting participants also put forth a number of thoughts and suggestions that can be divided into two categories:

1. Inputs on Expert Group approaches and working methods.
2. Views on topics of importance for Expert Group analysis and discussion

Inputs on Expert Group approaches and working methods

Enhanced use of scenario exploration methods

Some Expert Group members suggested further use of scenarios exploration methods and exercises, as was done in some prior Expert Group [meetings](#), to further unpack potential impacts of conceived AI futures and policy options identified by the group.

One member suggested that instead of orienting such scenarios around dramatically and comprehensively altered futures (e.g., through transformative superintelligence), we consider “mediocre futures” where a few AI capabilities advance dramatically but other things remain the same (e.g., LLMs advance significantly but issues with hallucinations remain). This could assist in teasing out variables and impacts associated with specific changes and what policy actions may be needed to manage risks.

Developing Expert Group sub-groups

Several members of the Expert Group noted that, as the group has grown to 65 members, it can be difficult for everyone to provide thorough inputs at the regular group meetings. Thus, they suggested that the Expert Group form a few topic-specific sub-groups to work on various areas of high interest. One member added that these sub-groups could potentially hold broader workshops with topic-relevant experts to broaden inputs and help validate thinking and outputs.

Broader survey of other experts

One Expert Group member suggested conducting a survey of broader experts to build upon Expert Group surveys and discussions to get their views on AI safety topics such as responsible scaling policies (RSPs) and AI risk thresholds, topics discussed further below. This member emphasised surveying experts in academia and civil society in particular, as those in the private sector are already regularly consulted for advice.

Views on topics of importance for Expert Group analysis and discussion

Expert Group members discussed various specific topics of interest for potential focus over the coming year. Overall, the topics raised can generally be classified as either exploring **technology futures** – medium to long-term advances in AI technologies themselves; or the economic, societal and governmental **impacts and responses** that such AI advances may trigger or necessitate.

Exploring technology futures

Artificial general intelligence (AGI)

Overall, the topic of artificial general intelligence (AGI)—a concept that can be described as machines with human-level or greater intelligence across a broad spectrum of contexts—and Expert Group members’ view on the subject represented a continuous theme throughout the session.

Several Expert Group members suggested that the group should delve deeper into the potential ramifications of, and preparations and responses related to, AGI. Pertinent questions for potential exploration by the Expert Group were posed: What if AGI becomes a reality by 2027 or sooner, as predicted by some AI experts and industry leaders? How can governments and societies prepare for such transformative technologies? Several participants advocated for a dedicated workstream to explore various scenarios on this topic, including a focus on ensuring the alignment of AGI actions and outputs with human preferences, as well as mitigating the potential risk of humans struggling to retain control of AGI systems.

Some Expert Group members recognised that, mirroring the broader AI community, there are significant unknowns and disagreements among group members with regards to various aspects of AGI, including its coherence as a concept, the plausibility of its future development, and the pathways that could lead to achieving it. Others noted that a lack of an agreed-upon definition for AGI and hesitance among some in the AI community to engage in the topic may complicate matters. There was also recognition, however, that press coverage, statements from some experts and AI company leaders, and the articulated missions of some AI companies focus on the issue, and that this may warrant further coverage of the topic by the Expert Group to help governments and policymakers understand different angles of debate.

Other participants, however, cautioned that the Expert Group should exercise restraint on the topic. Several members stated their belief that there is too much focus on AGI—both from doomsday and utopian perspectives—in the AI community and coverage of AI in general. Another stated that AGI is worth discussing, but that it should not become the dominant focus of the group, as initial reactions to generative AI systems may have overestimated their capabilities, that narrow but increasingly capable AI systems not reaching the level of AGI may still pose tremendous challenges and opportunities, and that discussions on AGI are often driven by big tech companies to help secure additional funding. Another participant suggested that instead of jumping to a focus on AGI, the group should consider the evidence base of the potential future risks identified and prioritise those that may be most empirically supported and impactful to analyse.

Agentic AI

Some Expert Group members raised interest in extrapolating trajectories and impacts of “agentic AI”, systems that perform a range of functions with increasing levels of autonomy. Some see this as a potential pathway to AGI, but this is not necessary for agentic systems to pose significant potential benefits and risks.

Human-machine collaboration

A number of participants stressed that, when (or instead of) touching on the area of AGI, the Expert Group should include a focus on “hybrid systems” or “human-machine symbiosis” and how the benefits of advanced or highly advanced AI systems could be maximised through human-machine collaboration and synergies and for the direct purpose of supporting humans and helping them to enhance their performance and even creativity. These members believed that this subject deserved more focus in the Expert Group and broader AI community.

Potential new technological breakthroughs

Some participants suggested exploring other potential technological breakthroughs, both in algorithmic approaches and hardware, that could occur in the future, and what policies and investments may need to be made by governments to seize opportunities and manage challenges. This suggestion was underpinned by debates in the AI community about whether scaling LLMs and transformer models represents the future of AI and could overcome existing challenges with model generalisation, or whether there are other developments that represent new breakthroughs that may be on the horizon. Such discussions could then back into corresponding issues, such as what types of data would be needed to achieve such breakthroughs, and what types of analysis and measurements would be needed to ensure future systems are meaningfully solving problems (i.e., as opposed to regurgitating information found in training data).

Technological convergence

As related to the OECD's work on [synthetic biology](#), a couple Expert Group members raised the topics of infusing organic systems with AI, potentially on a molecular level. One of the members stated that they were not referring to implants, but introducing new information into organic systems through facilitation by AI. Another raised AI-enabled neurotechnology as a potential area of focus, including potential risks related to AI interfering with human minds and thinking either directly (through implants or on-body sensing) or indirectly through the supposed "understanding" of human thoughts based on input such as written or spoken text, for example. This participant noted, "this topic merits serious attention, because our thoughts are our ultimate barrier to the outside world. Our final barrier to privacy."

Another Expert Group member suggested further examination of potential trajectories for AI used in conjunction with quantum computing.

Exploring impacts and responses to AI advancements

Establishing and implementing AI risk thresholds

Some Expert Group members suggested additional efforts scoped on the potential approaches, opportunities, and limitations for establishing risk thresholds for advanced AI systems. With the concept becoming a key area of focus in policy and technical communities, AI risk thresholds have been introduced in both voluntary commitments and concrete policy instruments (e.g., enhanced reporting requirements for AI systems trained over a certain level of compute). However, external experts argue that current risk thresholds, often expressed as a specific level of compute power, are arbitrary and thus could result in unintended consequences.

[Note: Subsequent to the 25 June meeting, the Expert Group co-chairs and OECD Secretariat launched a [public consultation](#) on this topic, which provides further background and context.]

Concentration of power

Some Expert Group members suggested additional analysis and discussion focused on concentration of power with regards to AI inputs and R&D among companies and countries, including additional coverage of the current state of affairs and recommendations from Expert Group members on what should and should be done to address this issue.

Experts highlighted concentrations in compute power as an area exemplifying this challenge, wherein those who have and control access to compute in a position to have outsized influence on potential technology trajectories. They highlighted that this is a critical factor contributing to issues of inequity, as touched on as a separate item below. Such issues may involve the OECD's [Expert Group on AI Compute and Climate](#) and associated workstreams.

Expert Group members also raised the related concept of the commercialisation of AI research, following previous contributions to AI research being largely led by the more open academic community. To one member, this seems to be resulting in a decline in creativity in AI research and more unitary focus on building products that can compete for consumers. The question was raised about how to "reboot" academic interest and capacities for AI research, which was suggested for further Expert Group exploration.

The role of compute providers in AI governance

Some Expert Group members raised that AI regulations, policies, and voluntary commitments tend to focus on AI development companies. However, compute providers play a significant role in the AI ecosystem and are largely outside of the scope of such instruments. One expert made the analogy of compute providers being like financial services providers. For instance, when a credit card transaction raises a red flag for potential fraud, the bank can suspend the account until the issue is resolved. The member raised the question of whether compute providers could fulfil a comparable intermediary function if red flags were raised in AI development and deployment.

Issues of inequity

Expert Group members raised an interest in further exploring and extrapolating future impacts related to AI-related inequities within and among countries. For instance, with regards to compute power, one member raised that some geographic regions have tremendous compute resources, while others have practically none, questioning what future scenarios and associated ethical and geopolitical challenges could unfold if this inequity were not addressed.

Another Expert Group member stressed that the group should focus not solely on the technology, but also the broader systemic factors that may intersect with the technology, such as social gaps that already exist and inequitable allocations of resources globally.

Outdated and/or inflexible policy and governance

Some Expert Group members argued that current policy and governance norms and processes have become outdated and may be insufficient for rapidly advancing AI technologies. Privacy norms and rules were raised as a key example. For instance, one member raised that governmental privacy instruments (e.g., regulations, policies) were often originated in a time where technology was less complex and globally networked and may not serve as adequate governance for AI (e.g., they do not apply the *aggregated*—as opposed to individual—data used for developing AI systems, which can lead to identification and profiling of certain individuals and groups). Data governance and incentives around generating and using high-quality data while protecting privacy rights was raised as particularly important. This subject is being explored by the OECD [Expert Group on AI, Data and Privacy](#).

Global governance models

As related to the Expert Group's ongoing work on exploring analogies regularly made between AI and existing global governance processes, as discussed above, some Expert Group members suggested discussion on what governance models and mechanisms may be optimal for the potential AI futures conceived of by the group. Such discussion could include what may work and what may not, as well as the pros, cons, and trade-offs for different possibilities.

Addressing global challenges

Expert Group members suggested a number of global challenges for further analysis and discussion by the group, including climate change; supporting sustainable growth, such as improving capacities for food production and access and strengthening logistics and processes for travel and dealing with waste; treating major diseases; and managing risks brought about by increasingly interconnected and interconnected systems.

Impacts on labour markets

Participants suggested an enhanced focus on the future impacts of AI on the labour market as a whole. Instead of looking at task exposure and risks of job loss through automation, it was proposed that the focus be on how AI could and should be assimilated in labour markets, what skills will be needed, and understating the broader societal implications that could result. Potential policy discussions could centre on training and upskilling of the current and future workforce and special considerations for segments of society that may face the most disruption, those that may be least likely to adapt to AI-induced labour market shifts, and those who may have significant amounts of time freed up (e.g., through job displacement or a lack of a need to work). Efforts in this area may involve the OECD Directorate for Employment, Labour and Social Affairs ([ELS](#)), which conducts relevant analysis and issues related products, such as the OECD Employment Outlook and its [recent edition](#) on AI in the labour market.

Education

As related to impacts on labour markets, participants raised the importance of discussing future education needs and what children and youth should be taught given the impact AI. This would be useful both in enhancing readiness for labour market shifts, but also in helping society adapt to change, make informed decisions, and leverage AI in a trustworthy way.

Expert Group members also suggested additional work on the use of AI to transform the field of education, such as through virtual teachers, personalised educational aids and digital tutors, and personalised feedback.

One Expert Group member cautioned that these topics are already being explored by the [OECD Education Directorate](#), which should be consulted before conducting work in this area to avoid duplication.

Security

An Expert Group member raised that issues of national security have not been central to many of the discussions of the Expert Groups so far, and that additional focus on this area would be valuable. The OECD Secretariat noted that the remit of the OECD does not include issues of defence, so options for policy recommendations on this topic are limited.

Conclusions

The discussions underscored the necessity of a forward-thinking and multi-disciplinary approach to research and policy related to potential AI futures. The group identified key areas for deeper examination and called for robust frameworks to address ethical considerations, data privacy, and equity, among others. Moreover, the participants recognised the uncertainties and risks associated with AI futures, and noted their support for further exploring these areas for a second year.

The Expert Group co-chairs and OECD Secretariat committed to taking this discussion into account while developing a work plan for the second year of Expert Group activities, to be presented and discussed among the group.