LEVERAGING STRATEGIC FORESIGHT TO MITIGATE ARTIFICIAL INTELLIGENCE (AI) RISK IN PUBLIC SECTORS

INTRODUCTION

The Sustainable Development Goals Report 2024 warns that with less than one-fifth of targets on track, the world is failing to deliver on the promise of the Sustainable Development Goals (SDGs). Artificial Intelligence (AI) holds significant potential to accelerate the implementation of the SDGs by enhancing efficiency, fostering innovation, and improving decision-making across various sectors such as health, education, climate change, water, food, and energy. However, the unpredictable trajectory of AI development, coupled with its complex ethical, social, and political ramifications, necessitates a structured approach to anticipate and navigate its potential impacts. Strategic foresight[1] exercises are essential in this context, enabling stakeholders to proactively identify and address emerging challenges and opportunities associated with AI.

By leveraging collective intelligence and scenario planning, strategic foresight exercises can help ensure that AI technologies are developed and deployed responsibly, thereby increasing the likelihood of their positive contribution to sustainable and inclusive growth. Such forward-thinking methodologies are critical to mitigating risks and harnessing AI's transformative power in advancing the SDGs.

This policy brief explains how strategic foresight can inform and guide public sectors in anticipating unexpected challenges and effectively harnessing AI technologies.

EXPLOITING AI TO STRENGTHEN PUBLIC SECTORS AND ACCELERATE SDG PROGRESS

The integration of AI into public sectors offers a transformative opportunity to advance the SDGs by driving economic growth, boosting productivity, and improving living standards. AI can help enhance agricultural practices and address complex challenges like climate change. For instance, AI is used to address food security challenges exacerbated by climate change by making real-time crop-placement decisions, monitoring crop health and enhancing supply chain processes.[2] AI can support better decision-making through advanced analysis and forecasting, information production and sharing, and revolutionize healthcare and education through personalized interventions. For example,

Key Messages

» If deployed responsibly, AI holds immense potential to accelerate the implementation of SDGs by enhancing efficiency, innovation, and decision-making across critical sectors like health, education, and climate change.

» Strategic foresight can help navigate the uncertainties of AI's rapid advancements and societal impacts and ensure it aligns with ethical principles and societal values.

» Strategic foresight can enable the public sector to anticipate AI's unpredictable advancements, identify emerging risks, and foster responsible and impactful innovation.

» Public administrations must build robust foresight capabilities to anticipate AI developments.

a US-based medical imaging startup uses large learning models for early disease detection in stroke care, cardiology, and oncology.[3] AI can also improve job quality by automating dangerous or repetitive tasks. For example, France is currently experimenting with a generative AI tool called "Albert" to streamline the daily tasks of French advisors on public services.[4] AI can be used to empower citizens and civil society by fostering participation and strengthening institutional transparency and governance through robust monitoring and evaluation systems.

The responsible use of AI can also strengthen public sectors by increasing productivity, fostering inclusivity and responsiveness in public services, and enhancing accountability through improved oversight capabilities.[5] For instance, Korea's Disease Control and Prevention Agency developed an AI convergence system that analyses medical, quarantine, and spatial data to forecast and inform policy responses to emerging infectious diseases.[6] Public institutions can significantly improve public service delivery by using AI to analyse large datasets and better understand citizen needs and preferences, and tailoring services accordingly. For example, Finland is using the AuroraAI programme to identify public services that are overly cumbersome for the user.[7]

---

[1] Strategic foresight is an organized, systematic way of looking beyond the expected to engage with uncertainty and complexity. UN CEPA strategy guidance note on strategic planning and foresight.

[2] ITU (2024). AI for Good. Impact Report.

[3] Ibid.

[4] https://www.france-services.gouv.fr/actualites/experimentation-dun-modele-dassistance-france-services-IA.

[5] OECD (2024), "Governing with Artificial Intelligence: Are governments ready?", OECD Artificial Intelligence Papers, No. 20, OECD Publishing, Paris, https://doi.org/10.1787/26324bc2-en.

[6] Ibid.

[7] https://oecd-opsi.org/innovations/auroraai/.

Author: Prabin Maharjan, United Nations Project Office on Governance, Division for Public Institutions and Digital Government, UN DESA.

AI RISKS IN THE PUBLIC SECTOR

While AI technologies offer significant benefits, their development, deployment, and use also pose considerable risks that span multiple areas. Challenges such as bias in AI systems, AI-enabled infringement of data privacy, the digital divide, the threat of job displacement, growing data disparities, equity and inclusiveness in AI are becoming increasingly evident[8]—and may threaten progress if not proactively managed. Additional risks linked to advanced AI include "hallucinations" in large language models, excessive resource consumption, and threats to peace and security.[9] For example, if hallucinating news bots provide unverified information during a developing emergency, they can rapidly spread falsehoods that hinder effective response and undermine public trust in official communications.[10] AI's integration into military systems could lead to autonomous weapons that operate without human oversight.[11] This could accelerate the speed and scale of warfare in terms of inflicting harm on both civilians and the environment, and thereby compromise international peace and security.[12] Compounding these challenges, AI-generated misinformation and disinformation threaten the integrity of public institutions by eroding trust, distorting public discourse, and undermining democratic processes. These risks demand urgent attention to safeguard human rights and privacy, ensure algorithmic transparency and accountability, promote explainability, and prevent unfair or biased policy outcomes.

The UN Governing AI for Humanity - Final Report highlights varying levels of concern among experts regarding AI risks across multiple domains. Key areas of concern include:

- **Ethical and human rights risks:** AI could infringe on human rights, perpetuate biases, and exacerbate inequalities, particularly in decision-making systems like criminal justice, hiring, and social services.

- **Privacy and surveillance:** AI-enabled mass surveillance, data exploitation, and the erosion of personal privacy could undermine democratic freedoms.

- **Security risks:** AI's dual-use nature (civilian and military applications) could lead to misuse in cyberattacks, autonomous weapons, and other malicious activities.

- **Economic and labour disruption:** AI could lead to job displacement, economic inequality, and the concentration of power in the hands of a few tech giants.

- **Accountability and transparency:** A lack of transparency in AI algorithms and decision-making processes challenges accountability, especially when AI systems cause harm or make errors.

- **Global governance and cooperation:** The fragmented global governance of AI, risks of a "race to the bottom" in regulatory standards.

- **Environmental impact:** The environmental footprint of AI, particularly the energy-intensive training of large models, is a growing concern.

The findings (Figure 1) reveal concerns about AI-related harms in the coming year, emphasizing the urgency to address risks and vulnerabilities across various areas in the near future. These risks underscore the critical need for robust, comprehensive policies to govern AI responsibly and effectively mitigate its potential harms. While many governments acknowledge the potential benefits of AI, significant gaps remain in their preparedness. Strategic foresight can help anticipate potential AI futures and mitigate risks.

STRATEGIC FORESIGHT TO MITIGATE AI RISKS

The trajectory of AI innovation remains unpredictable due to the rapid pace of technological advancements, which continuously generate unforeseen opportunities and challenges. This uncertainty extends beyond technology itself, encompassing ethical, social, and political dimensions. Strategic foresight can play a crucial role in anticipating and navigating rapid change, preparing for diverse future scenarios and stress-testing current or proposed strategies.

The rapid advancement of AI technologies outpaces the development and implementation of regulations, creating a regulatory gap rather than an enforcement gap.[13] Existing laws are insufficient to address the unique and swiftly evolving challenges posed by AI, necessitating new regulations to ensure comprehensive and adaptive governance.[14] Strategic foresight is essential for understanding AI's societal impacts and proactively shaping responsive projects and policies. Public institutions can leverage strategic foresight to navigate the risks and maximize the benefits of AI for achieving the SDGs. By systematically analysing trends, forecasting outcomes, and aligning strategies with ethical principles and societal values, public institutions can guide AI's growth responsibly. This forward-thinking approach enables informed decision-making to mitigate risks, harness opportunities, and ensure AI evolves in ways that enhance global well-being and align with human values.

Public institutions can enhance their strategic foresight capacities by establishing a clear value proposition for integrating foresight into policy and decision-making processes.[15] This includes investing in strategic foresight initiatives and addressing barriers that limit its effectiveness, such as bureaucratic silos and challenges in fostering dialogue with non-governmental actors. By overcoming these obstacles, institutions can better understand and navigate the complex impacts of AI. For instance, Finland integrates strategic foresight into its national planning through the Government Foresight Group[16] and the parliamentary Committee for the Future.[17]

There is no one-size-fits-all approach to foresight. Foresight exercises can be highly structured or informal, depending on national or local circumstances.

---

8 Addendum on AI and Digital Government. UN e-Government Survey 2024.

9 UN (2024) Governing AI for Humanity - Final Report.

10 https://www.ibm.com/think/topics/ai-hallucinations.

11 United Nations General Assembly. (2024). Artificial intelligence in the military domain and its implications for international peace and security. A/RES/79/239.

12 UNEP (2024). Navigating New Horizons – A Global Foresight Report on Planetary Health and Human Wellbeing.

13 UNGA (2024). Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development. A/78/L.49.
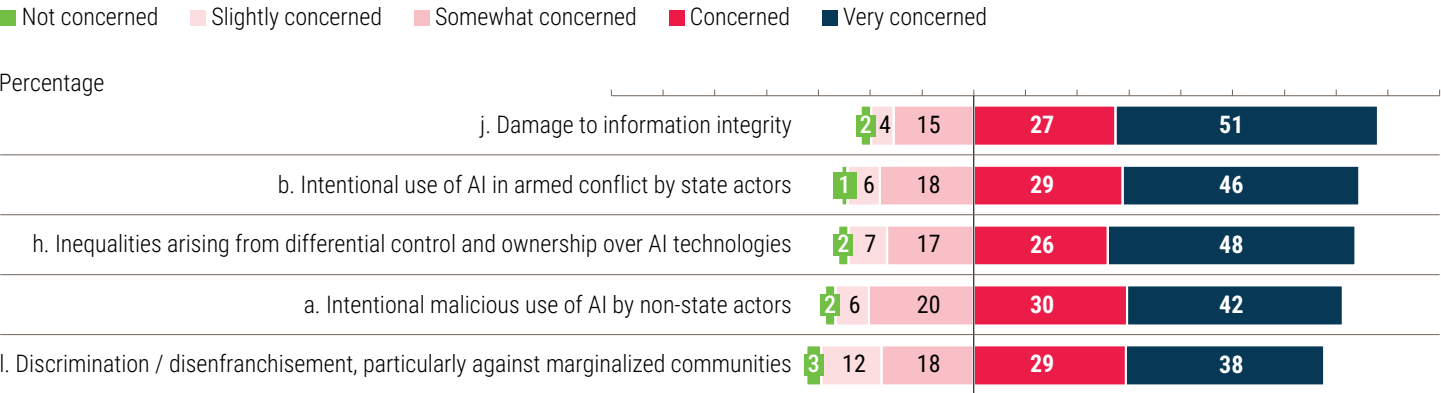
14 UN (2024) Governing AI for Humanity - Final Report.

15 OECD (2024), "Assessing potential future artificial intelligence risks, benefits and policy imperatives", *OECD Artificial Intelligence Papers*, No. 27, OECD Publishing, Paris, https://doi.org/10.1787/3f4e3dfb-en.

16 Finnish Government. Government Foresight Group.

17 The Committee for the Future is an established, standing committee in the Parliament of Finland.

Figure 1
Concern about AI risks across multiple domains[a]

■ Not concerned ■ Slightly concerned ■ Somewhat concerned ■ Concerned ■ Very concerned

Percentage

	Not concerned	Slightly concerned	Somewhat concerned	Concerned	Very concerned
j. Damage to information integrity	2	4	15	27	51
b. Intentional use of AI in armed conflict by state actors	1	6	18	29	46
h. Inequalities arising from differential control and ownership over AI technologies	2	7	17	26	48
a. Intentional malicious use of AI by non-state actors	2	6	20	30	42
l. Discrimination / disenfranchisement, particularly against marginalized communities	3	12	18	29	38

"Please rate your current level of concern that (existing or new) harms resulting from AI will become substantially more serious and/or widespread in the next 18 months for each area." (n = 348)[b]

Source: UN (2024). Governing AI for Humanity.

a For more concern about AI risk, please access: UN (2024). Governing AI for Humanity.

b The sampling was constructed from the Office of the Secretary-General's Envoy on Technology (OSET) and the Advisory Body's networks. Check UN (2024). Governing AI for Humanity for more information.

The UN Futures Lab recommends the following foresight tools to do three things:

1. **Make sense of change:** These are tools that help make sense of what is happening: they help to observe the world and to look out for signals of change–things that might be small today but could become big in the future, or vice versa.

   ▪ **Horizon scanning:** Identifies emerging changes that could have a big impact on a country or a specific sector. The Government of Rwanda used horizon scanning in its national development planning process by examining the future of urbanization, the future of rural sector development and large-scale public investment projects.[18] It can be used to develop a situation analysis to identify drivers of change brought by AI that could impact various industries, sectors, or societies.

   ▪ **Three horizons:**[19] A horizon scanning approach to understand societal transitions. Cabo Verde used this tool to explore and identify existing and newly required government structures to deliver collectively and coherently to strategic (i.e. sector transcending) objectives.[20] This tool can be used for situation analysis or mid-term review to identify forward-looking AI risks and opportunities. It can also be used to identify drivers of change—which ones might be holding back change and which ones might be advancing it.

   ▪ **Futures triangle:** An approach to engage people in conversation about broad forces that may shape the future. It can be used to analyse the potential future of AI by examining the interplay between three key forces: "the pull of the future" (desired vision), "the push of the present" (current trends and drivers), and "the weight of history" (past barriers and limitations), essentially mapping out the competing forces that will shape the development and impact of AI technology.

   ▪ **Futures wheel:** A tool that helps to explore direct and indirect consequences of trends, events, and emerging issues. It can be used to uncover the potential AI consequences of scenarios, events, or drivers during a situation analysis or strategic planning. It can be used to think through possible impacts of current disruptions, new changes, or trends in society such as the pace of AI development.

2. **Imagine possible futures:** These are tools for identifying new possibilities for the future, building scenarios, and identifying what a desired future might look like.

   ▪ **Scenario development:** An approach to broaden the understanding of how the future may evolve. Mauritius used this tool to construct generic scenarios of the country in 2025.[21] It can be used as a strategy to identify potential future AI risks and opportunities, particularly those that might emerge from interactions between drivers of change identified in the horizon scanning.

   ▪ **Desired future:**[22] A tool that helps to identify characteristics of the future. It helps us to think about a range of outcomes rather than one scenario only. It can be used to engage the public and community groups in developing more desirable AI futures.

18 UN Futures Lab (2023). UN Strategic Foresight Guide.

19 The first horizon represents the current stable situation, the second horizon encompasses emerging innovations that either sustain or challenge the status quo, and the third horizon highlights transformative changes and visions of potential future paradigms.

20 UNDP (2018). Foresight Manual - Empowered Futures.

21 Ibid.

22 While the Futures Triangle maps systemic tensions (pull/push/weight) to assess the feasibility of desired futures, and the Futures Wheel traces ripple effects of trends, whereas Scenario development and Desired Future take these imaginative insights and synthesize them into coherent, actionable narratives and strategic goals that inform policy decisions.

- **Matrix policy gaming:** A simulation exercise based on roleplaying. It combines the experience of roleplay gaming with strategic decision-making and policy. It is most useful in cases where the future will be defined by how different actors act and react with each other. It can be used to mitigate AI risks by simulating multi-actor interactions in high-stakes scenarios—such as misinformation campaigns or AI-driven conflict escalation—allowing policymakers to anticipate potential responses, stress-test governance strategies, and co-develop more adaptive, collaborative approaches to AI risk management.

- **Causal layered analysis:** A theory and methodology to explore the layers of change needed to truly transform and achieve the future that we want. It can be used to envision more desirable AI futures and develop a strategy that is focused on transformational change.

3. **Take action:** These are the tools that have to do with bringing the future back to the present. What transformations need to happen to bring about the desired future? What do we need to start doing now to move towards that future?

- **Backcasting:** A tool to develop pathways to the future, starting not from the present but from what we need to achieve. In the Dominican Republic, UN DESA held a workshop on Foresight and Systems Thinking for Strategic Planning for SDGs using a backcasting tool.[23] It can be used to envision and explore the future of AI starting from a desired future state and working backward to identify the steps needed to reach that state.

- **Change agenda:** It identifies the transformations needed to achieve the desired future. If you are using foresight to inform a set of decisions, the first step is to outline the change agenda. The change agenda plays a crucial role in answering the "so what?" of foresight. It can be used after a desired future and scenarios/horizon scan to identify what changes are needed to achieve the desired future. It can systematically mitigate AI risks by translating foresight insights into actionable strategies to align outcomes with a safer, equitable future.

- **Wind tunnel testing:** A process for stress-testing policies, plans, and strategies using scenarios. This method identifies potential challenges, missed opportunities, and areas to strengthen strategies and interventions. It can be used to test whether the strategy or theory of change may need to be updated. It also might be used when a strategy has been developed, and the team wants to understand how it will fare under different scenarios. By applying Wind Tunnel Testing to AI risk management, policymakers can rigorously simulate various disruptive scenarios to uncover vulnerabilities in existing strategies and adapt policies before adverse outcomes materialize.

Strategic foresight is essential for navigating the rapidly changing and complex world of AI. By actively examining potential future scenarios and their impacts, public institutions can guide AI development toward outcomes that maximize societal benefits, ensuring its transformative potential is utilized ethically and responsibly.

23 UN DESA (2024). Foresight and Systems Thinking for Strategic Planning for Sustainable Development Goals in the Dominican Republic.

Foresight tools are not confined to the above tools. A broad range of methods, techniques, and methodologies exists, each serving distinct purposes. A comprehensive foresight exercise rarely depends on a single method; therefore, it is essential to combine foresight tools tailored to the specific application and objectives of the foresight process. For instance, to develop a strategy that is more forward-looking and conscious of uncertainty, public institutions can utilize a combination of tools. These include visioning (desired future), strategic priorities and outcomes (change agenda), backcasting (outputs or interventions), and risk management (stress testing).

WAY FORWARD

The SDGs are severely off track, with only 17% of targets progressing as planned, highlighting the urgent need for innovative solutions. AI offers a powerful opportunity to accelerate SDG progress by enhancing decision-making, fostering innovation, and addressing global challenges across sectors like health, education, and climate change. However, the rapid advancement of AI also poses significant risks, including ethical concerns, privacy violations, and governance challenges. Strategic foresight helps navigate AI's unpredictable trajectory, enabling informed decision-making and fostering sustainable, equitable progress. It can serve as a bridge between innovative technologies—such as AI—and the SDGs by providing structured, evidence-based methods to plan and evaluate interventions across all the SDGs.

Governments should prioritize the establishment of comprehensive and adaptive regulatory frameworks to govern AI. Strategic foresight tools, like Wind Tunnel Testing, can simulate various policy scenarios under different future conditions. This helps governments stress-test regulatory frameworks for emerging AI technologies to ensure they remain robust against challenges like bias or privacy infringements, thereby supporting effective, accountable and inclusive governance at all levels.

Governments should leverage AI to enhance efficiency, innovation, and inclusivity in public service delivery. Using Horizon Scanning and the Futures Wheel, policymakers can identify emerging trends and innovative applications of AI in health, education, and social protection. These foresight tools facilitate early detection of opportunities to enhance public service delivery, promote innovation, and bridge disparities among different communities.

Investing in AI literacy and skills for policymakers, public sector employees, and citizens is essential. Scenario Development and Causal Layered Analysis enable the design of targeted capacity-building programs. By mapping future skill needs and identifying local priorities, these tools support the development of tailored educational initiatives for policymakers, public sector employees, and citizens—ensuring that workforce development aligns with future economic demands and equitable growth.

Public institutions should integrate foresight practices into policymaking to better anticipate AI-related challenges and opportunities. Methods such as Backcasting and Change Agenda guide the process of working backward from a desired sustainable future to articulate clear policy pathways. They align AI initiatives with societal values and environmental priorities, ensuring that investments in technology lead to resilient, sustainable cities and communities.

Governments should ensure AI benefits all societal segments, especially marginalized and vulnerable populations, to reduce inequalities. Foresight methods like the Futures Triangle and Three Horizons can be used to envision and plan for an AI-driven future that benefits all segments of society, including marginalized populations. These tools help articulate strategies to reduce inequalities by mapping out potential barriers and identifying opportunities for inclusive technological adoption.

Combining AI's analytical power with human critical thinking and emotional intelligence, strategic foresight can explore, anticipate, and prepare for future developments and uncertainties. By integrating these foresight techniques within national and local policy frameworks, public institutions can better anticipate challenges and opportunities, ensuring that AI innovations are not only optimized for efficiency and productivity but are also firmly aligned with the ethical and societal objectives embedded within the SDGs.

To strengthen government foresight capabilities, targeted capacity-building support is essential, especially in developing countries with limited resources and expertise. In several instances, UN DESA has supported Member States to build capacity on strategic foresight and resilience to address uncertainties, adapt to future demands, and facilitate sustainable development.[24]

24 Future in Focus: Preparing for What Lies Ahead, UN DESA Preparedness and Foresight Review.