# *Business at OECD* (BIAC) Webinar on Trustworthy AI

21 April, 2022

*Summary*

25 April 2022

## Background

*Business at OECD* (BIAC) held a webinar on trustworthy AI on 21st April 2022 to present and discuss the main findings of its forthcoming report on *"Implementing the OECD AI Principles: Challenges and Best practices"* to be published in May 2022.

The *Business at OECD* (BIAC) project launched in 2021, aims to support implementation of the OECD AI Principles, with a specific focus on the five values based principles to advance the development and implementation of trustworthy AI.

The project rests on two pillars: In-depth qualitative research around use cases in order to identify the learnings and challenges companies face when implementing the OECD AI Principles; and using the OECD's database of tools for trustworthy AI , the project evaluates the AI systems highlighted in our project case studies.

The project findings will also contribute concrete examples of good practice and lessons learned from private sector AI initiatives to the OECD.AI database of tools for trustworthy AI.

The webinar was an opportunity for multistakeholder dialogue around the key takeaways from the project. It included discussions with participating companies on the specific insights the project offered, and an exchange between stakeholders on implementation of the OECD Trustworthy AI Principles, including experts from OECD, business, civil society and the technical community.

The forthcoming project report will be published in early May.

The following is a summary of the webinar proceedings.

*BUSINESS AT OECD (BIAC) PROJECT ON TRUSTWORTHY AI IS TIGHTLY CONNECTED TO WORK LED BY THE OECD*

- The OECD set up its Network of Experts on AI, which includes over 250 experts from around the globe, and meeting monthly.
- In Spring 2022, the OECD Council and the Committee on Digital Economy Policy (CDEP) created a new Working Party on AI Governance, which will oversee the OECD AI work program.
- This special project on trustworthy AI is a collaboration between *Business at OECD* and the OECD Secretariat leveraging several OECD frameworks including the AI system lifecycle, the structure for tools for trustworthy AI as well as the framework for the classification of trustworthy AI.
- The results of the project will contribute to the OECD AI Catalog of Tools for Trustworthy AI aiming to be launched in the early fall 2022.

*THE BUSINESS AT OECD REPORT ON TRUSTWORTHY AI PRESENTS SEVEN BUSINESS USE-CASES[1], WHICH DEMONSTRATE A NUMBER OF COMMON PRINCIPLES:*

- Successful and strong AI governance is at the basis trustworthy AI; it requires full-time dedication to the issue, including to ensure good AI governance is at the core of the architecture of an organization.
- There is a need for customization of the AI tools depending on context.
- Dedicated and trained people play a key role within the organization's process and their digital literacy has to be ensured.
- The adoption of tools ensuring the implementation of trustworthy AI may have second order impact on the organization and stakeholders, and may lead to organizational and cultural change.

*RECOMMENDATIONS FOR REGULATORS EMERGED FROM THE STUDY, AMONG THEM:*

- There is a need for flexibility with respect to the implementation of principles for trustworthy AI.
- The tools developed by companies can be sources of information for regulators as they crystalize realities of practical implementation.
- The work of standards bodies can be further leveraged with respect to development of regulations.

*PRESENTATION OF DIFFERENT USE-CASES AND RELATED FINDINGS*

- For Boris Ruf, Research Scientist Lead Expert in Algorithmic Fairness from Axa Corporation, giving a clear definition of the "fairness objective" from the start is important to ensure the correct outcome is met. This point was illustrated in his presentation of the AXA "Fairness compass", which brings structure to the complex landscape of fairness definition.
- Michael Hind, Distinguished Research Staff Member from IBM Research, presented a "FactSheet" focusing on AI transparency, and which displays information for the understanding of AI models.  In presenting the IBM Factsheet use case, he highlighted the importance of customization, as not all customers will look for the same information.

---

[1] Axa, AWS, IBM, Meta, Microsoft, NEC, PwC

- The tool presented by Yasunori Mochizuki, NEC Fellow at NEC Corporation, addresses robustness, security and safety as well as accountability. By being both a guideline and a quality assessment tool, it ensures oversight for a solution delivery cycle to clients.
- The tools presented also highlighted a number of challenges related to implementation of trustworthy AI, including the need for flexibility, the difficulty to translate processes to achieve other outcomes (e.g., scoring; pricing), balancing transparency with intellectual property, burden on the development process, automation of collection of data and labor cost.

*PERSPECTIVES OF STAKEHOLDERS*

- Pam Dixon, Executive Director of the World Privacy Forum, highlighted a number of key issues for civil society related to the implementation of trustworthy AI:
  - Regulatory models for AI need to be comprehensive, similar to the regulatory model of toxic chemicals around the world.
  - Regulations should rely on administrative and procedural controls, such as labeling, documentation, facilities for redress and consumer complaint mechanisms.
  - Civil society is strongly interested in the diffusion of models for continuous improvement such as the one developed by Toyota.
  - The quality and methodology of the quantification of privacy, fairness and inclusion issues need to be assessed.
  - In addition, a whole of government approach presents interesting opportunities for inter-bodies dialogue, similar to New Zealand's algorithmic impact assessment led by experts from different authorities.
  - Moreover, the Sustainable Development Goals (SDGs) should be embedded in the research regarding trustworthy AI especially to harness the power of AI to advance the sustainable transition.

- Clara Neppel, Senior Director at IEEE, presented views of the technical community:
  - As much as regulation is important, there is a need to shift from focus on compliance to the business strategy itself.
  - Therefore, it is necessary to investigate how to increase value while minimizing risks. The latter being a compelling point for businesses, but also for investors looking to work with sustainable start-ups for example.
  - To do so, the development of values-based principles is essential and as illustrated by the work of the OECD as a testimony to this point.
  - Standards will play an important role in the redefinition of what stands as a risk and what is acceptable, which is why including all stakeholders in the conversation is important.
  - In the long run, it is important to have frameworks of governance of standards, best practices and conformity assessments in order to make trustworthy AI a reality.
  - Moreover, all research should be based on impact-assessment, not only risk-assessment, from which regulators would draw precise red lines.
- Norberto Andrade, Global Policy Lead for Digital and AI Ethics at Meta, presented main issues businesses face when implementing trustworthy AI and how companies can engage in a multi stakeholder process:
  - The process for going from principles to practice is not straightforward and involves trade-offs as well as it presents opportunities.

- o Definitions of what fairness entails and how it functions in practice with different stakeholders presents a challenge. To get to this level of granularity and complexity, shifting the debate from high-level principles to actual practice is essential.
- o Finally, frameworks, processes and tools should be following a holistic approach encompassing all principles and analyzing how they interact with one another. This approach would help understand how one principle can enhance another and be a vehicle to its implementation.
- o From a policy maker perspective, it is important to acknowledge the interdependence of different instruments and regulatory approaches and how they complement each other. The United States alone presents three different approaches, which could be made part of one single ecosystem: a precautionary approach to prohibited practices, a risk-based approach to high-risk systems and an experimental approach with regulatory sandboxes.
- o The iterative method used for experimentations can inspire policy and law makers to be more iterative and experimental to measure the impact of technologies, especially if they present a level of uncertainty.