Navigating Responsible AI

Christos Sarakinos

Director, Data Science and Innovation Division at Statistics Canada

Sprint on Artificial Intelligence and Data Science for Economic Statistics – December 12, 2024

Outline

Introduction

AI in the News – The Good, the Bad and the Ugly!

Artificial Intelligence's Promises and Risks

What is Responsible AI and Why is it Important?

Fundamental Principles of Official Statistics and Responsible AI

- Connecting The FPOS with the Core Principles of Responsible AI

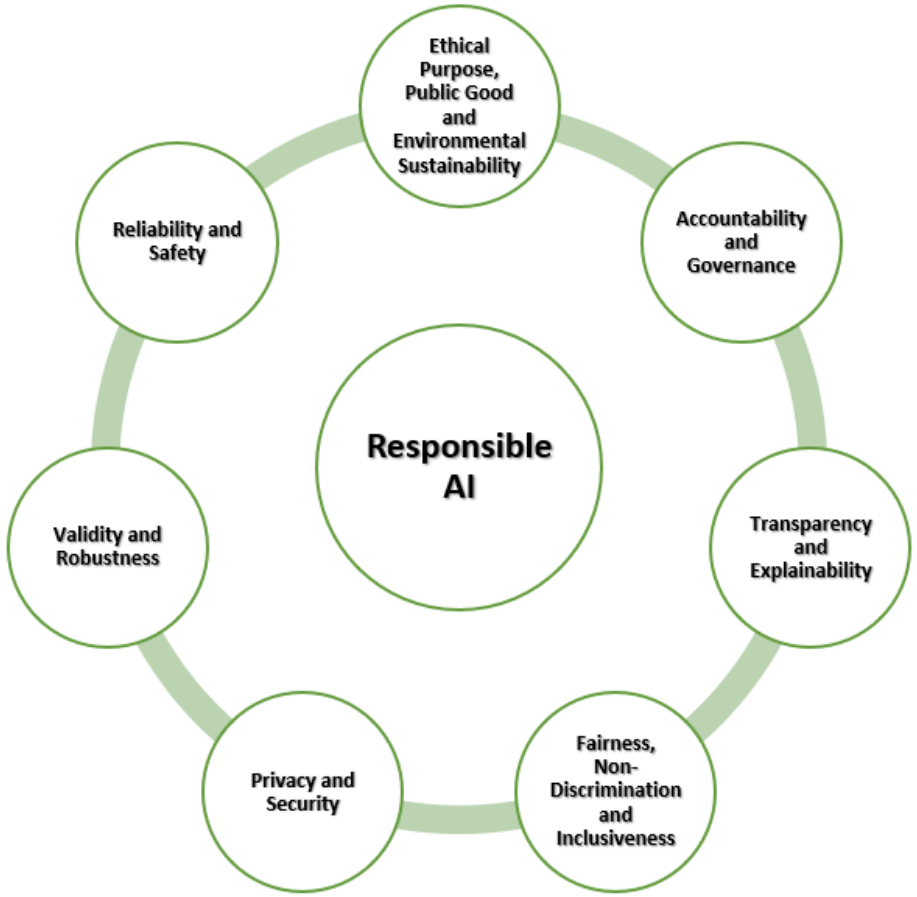
Current Landscape of AI Implementation and Governance

- Worldwide – Canada – Statistics Canada

Current Challenges and Opportunities

Establishment of Office of Responsible AI

- Key Responsibilities
- Key Activities
- Governance Products



AI in the News – The Good, the Bad and the Ugly!

Large Language Models for Official Statistics

HLG-MOS White Paper
December 2023

eurostat

An introduction to Large Language Models and their relevance for statistical offices

Harvard Business Review

Business And Society

AI Can Make Bank Loans More Fair

by Sian Townson

November 06, 2020

The New York Times

Opinion

OP-ED CONTRIBUTOR

When a Computer Program Keeps You in Jail

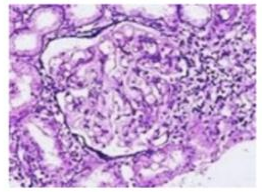
By Rebecca Wexler

June 13, 2017

Wrong decisions can be costly and dangerous

"Autonomous car crashes, because it wrongly recognizes …"

"AI medical diagnosis system misclassifies patient's disease …"



Credit: Samek, Binder, Tutorial on Interpretable ML, MICCAI'18

"The development of full artificial intelligence could spell **the end of the human race.**

–Stephen Hawking

May 01, 2024

With huge patient dataset, AI accurately predicts treatment outcomes

New model compares drug effectiveness – without a clinical trial

AI speeds up drug design for Parkinson's ten-fold

The Age of Secrecy and Unfairness in Recidivism Prediction

…ng, and Beau Coker

The New York Times

"AI has **hacked** the operating system of human civilisation.

– Yuval Noah Harari

Researchers develop AI-driven tool for near real-time cancer surveillance

by Mark Alewine, Oak Ridge National Laboratory

Ottawa mettra sur pied un institut de la sécurité de l'intelligence artificielle

Publié le 7 avril à 14 h 59 HAE
Mis à jour le 7 avril à 15 h 53 HAE

RADIO-CANADA

FORBES > INNOVATION > AI

AI For Smart Cities: It Could Be A Climate-Changer

Intelligence artificielle et santé: une révolution, mais aussi de nombreux risques, met en garde l'OMS

Par **Le Figaro Santé** et **AFP agence**

Publié le 18/01/2024 à 15:46, mis à jour le 18/01/2024 à 15:51

ECONOMIC VIEW

Biased Algorithms Are Easier to Fix Than Biased People

Racial discrimination by algorithms or by people is harmful — but that's where the similarities end.

"AI is far more dangerous than **nukes.**

– Elon Musk

MARCH 22, 2024

UK government's approach to realizing benefits of AI assessed in new report

by Russell Parton, University of Exeter

ACTUALITÉS

Intelligence artificielle

LA PRESSE

Quels dangers pour la démocratie ?

Artificial Intelligence's Promises and Risks

AI Promise and Risks

- AI has the potential to transform official statistics by enhancing data collection and analysis, leading to more accurate and timely insights. However, it also poses risks, such as data privacy concerns and the potential for biased outcomes. Balancing these opportunities and challenges is vital for the responsible use of AI in statistical practices.

The Montreal Declaration for a responsible development of AI (2017)

- *"Artificial intelligence (AI) is a major form of scientific and technological progress that can generate considerable **social benefits**. The development of AI, however, **poses ethical challenges and social risks**. It is incumbent on the various public and private stakeholders and policymakers, at the local, national and international levels, to ensure that the development and deployment of AI are compatible with the protection and fulfilment of fundamental human capacities and goals."* The Montreal Declaration

The EU AI Act (2024)

- The European Parliament has recently approved the so-called **AI Act**, where it is stated that *"Artificial Intelligence is a fast- evolving family of technologies that can bring a wide array of economic and societal **benefits** across the entire spectrum of industries and social activities. [...] However, the same elements and techniques that power the socio-economic benefits of AI can also bring about **new risks or negative consequences** for individuals or the society"*.

What is Responsible AI and Why is it Important?

Responsible AI – *framework safe, trustworthy and ethical implementation*

- "*Responsible Artificial Intelligence is about human responsibility for the development of intelligent systems along fundamental human principles and values, to ensure human-flourishing and well-being in a sustainable world. … **Responsible AI is not about the characteristics of AI systems**, but about our own role. **We are responsible** for how we build systems, how we use systems and how much we enable these systems to decide and act by themselves*." Virginia Dignum

Why Governance Matters

- AI's Rapid Evolution and Impact. AI is rapidly transforming industries and society.
- Risk Management: While AI offers immense benefits, it also poses risks and unintended consequences.

Potential AI Issues

- Bias and Fairness, confidentiality, Transparency, Explainability, algorithmic accountability, legality, security, safety, …

Our Shared Responsibility

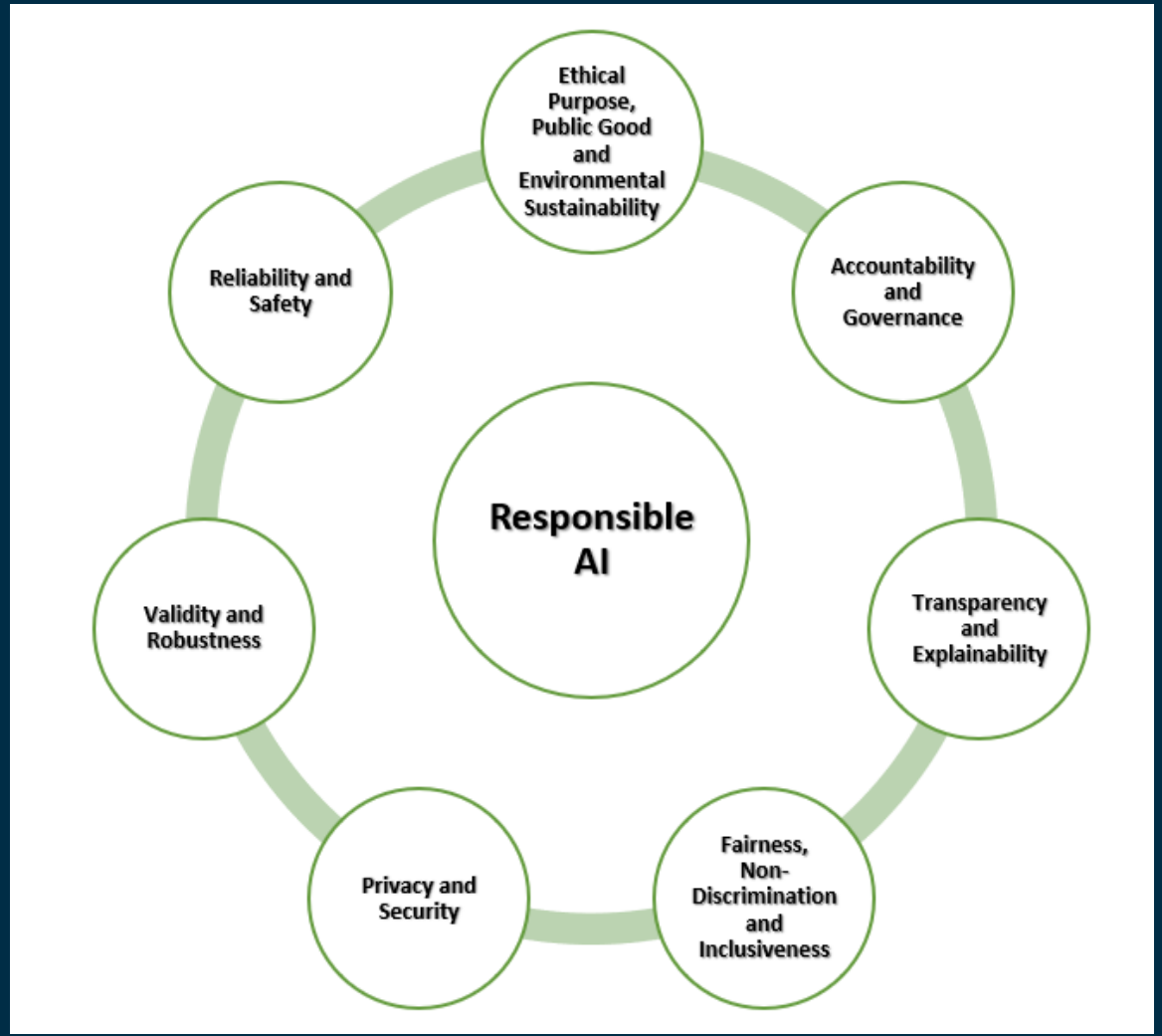
- **We are all responsible** for ensuring governance through policies, frameworks, guidelines, assessment tools, committees, audits ...

Fundamental Principles of Official Statistics and Responsible AI

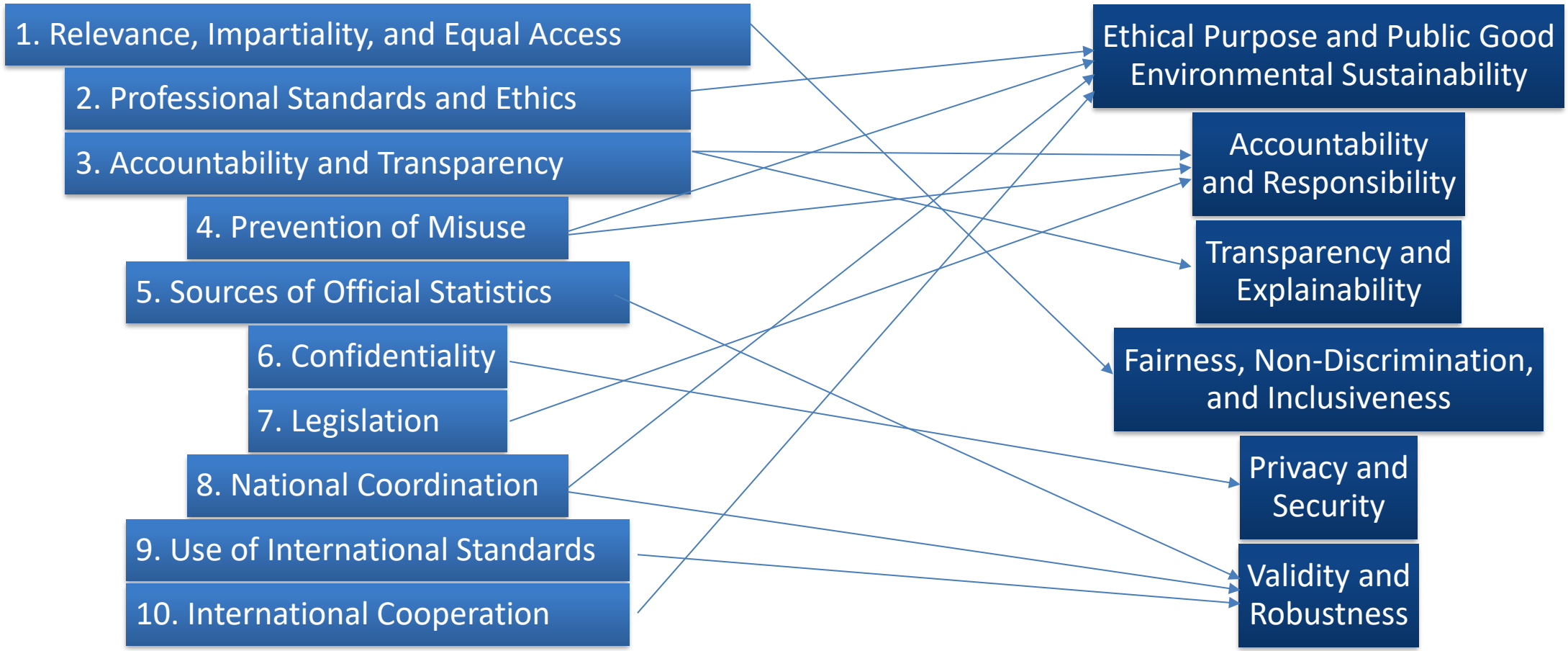
Responsible AI and The Fundamental Principles of Official Statistics

- To ensure the credibility and objectivity of official statistics, the **United Nations** Statistical Commission established the **Fundamental Principles of Official Statistics** (UN, UNECE) in 1994, endorsed by the UN General Assembly in 2014.
- The FPOS consists of ten key principles that **guide the responsible production and use of statistical data.**

Responsible AI Core Guiding Principles



Connecting The FPOS with the Core Guiding Principles of Responsible AI



AI Policy Milestones:

The Present and Progressive Landscape of AI Governance

Global Initiatives and the Canadian Experience

The Many Types of AI Governance

AI governance encompasses the frameworks, policies, and practices that ensure the ethical, transparent, and accountable use of AI technologies.





AI Principles

Guiding Concepts and Values

- **Montreal Declaration for a Responsible Development of Artificial Intelligence**
- OECD AI Principles
- Asilomar AI Principles
- IEEE's Ethically Aligned Design

AI Frameworks

General Operating Structures, objectives and definitions

- **Statistics Canada's framework for responsible ML**
- NIST AI Risk Management Framework
- OECD Framework for the Classification of AI Systems

Laws and Policies

Rules Enacted and enforced by Government

- **The Artificial Intelligence and Data Act (Canada)**
- **TBS Directive on automated-decision making**
- AI Act (EU)
- NYC Local Law 144 of 2021
- American Data Privacy and Protection Act
- Internet Information Service Algorithm Management Regulations (China)

Voluntary Guidelines

Practices, structures, and actions that are optional but encouraged

- **Canada's generative AI code of conduct**
- **Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems (ISED)**
- White House's voluntary commitments from leading AI companies

Standards and Certifications

Sets of practices and controls that demonstrate compliance with laws or otherwise provide assurance

- ISO/IEC JTC 1/SC 42
- IEEE P7000 series standards projects
- CEN/CENELEC standards development
- **RAII** Institute's Certification Program for AI Systems
- **Armilla's AI Governance Platform**

AI Governance Initiatives (Non-Exhaustive)

Global Initiatives

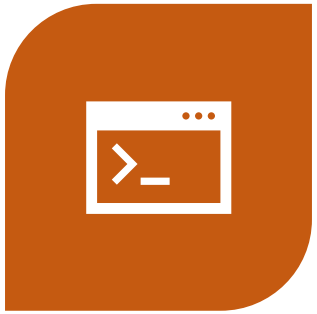
- 2017. Asilomar Conference on Beneficial AI (organized by the Future of Life Institute)
- 2019. Recommendation of the Council on Artificial Intelligence (OECD, 2019)
- 2019. OECD AI Principles
- 2019. Ethics guidelines for trustworthy AI (EU, 2019)
- 2021. Ethics, Transparency and Accountability Framework for Automated Decision-Making (UK, 2021)
- 2021. Recommendation on the Ethics of Artificial Intelligence (UNESCO, 2021)
- **2022. Algorithmic Accountability Act (USA, 2022) (Regulatory framework).**
- **2023. Interim Administrative Measures for the Management of Generative AI Services. CHINA. (Regulatory framework).**
- **2024. Artificial Intelligence Act (EU, 2024). (Regulatory framework).**

Canada's Initiatives

- 2017 - Pan-Canadian AI Strategy – CIFAR (Canada's national AI Strategy was the first in the world !)
- 2017 - The Montreal Declaration for a responsible development of Artificial Intelligence
- 2018 - The Toronto Declaration
- 2019 - Directive on Automated Decision-Making
- 2019 - Algorithmic Impact Assessment Tool
- 2023 - Guide on the use of Generative AI - Canada.ca
- 2023 - Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems
- 2023 - Principles for responsible, trustworthy and privacy-protective generative AI technologies
- Upcoming - The Artificial Intelligence and Data Act (AIDA) – Companion document
  - Bill 27 (NOTE. AIDA is part of Bill C-27 - Part III of Bill C-27)

Types of AI Solutions by Use (draft)

Algorithmic Applications



Productivity Tools & Software Acquisition

Citizen-Facing AI Applications

Corporate Tools for Operational Efficiencies

Systems for Statistical Production and Analysis

Inferential Applications

Algorithmic

Inferential

AI Governance Complexity Matrix (draft)

Application	Productivity Tools & Software Acquisition	Citizen-Facing AI Applications	Corporate Tools for Operational Efficiencies	Systems for Statistical Production and Analysis
Examples	automating repetitive tasks, enhancing data processing, improving document management	augmented website search, citizen help chatbots, automated dissemination / comms products.	e.g. workflow automation, human resource management systems, automated classification, data validation.	ML for data imputation, advanced statistical modeling, real-time data processing, combining administrative, census, and alternative data sources for new insights.
Delivery	Standard software procurement	In-house/custom development	In-house/custom development	In-house/custom development
Data	Agnostic/ Existing data	Agnostic/ Unstructured/Well-structured	Unstructured/Well-structured/primary/ admin data	Unstructured/Well-structured/primary/ admin data
Governance aspects	Standard software and product lifecycle, IT integration checks	Data privacy, security measures, accuracy verification, responsible AI	Cybersecurity assessments, light responsible AI	Data quality, methodological rigor, responsible AI

Establishment of Office of Responsible AI (ORAI)

ORAI - Key Responsibilities

AI Reviews

Playbooks and Guidance

Coordination with Legal, Privacy, Ethics, Cyber, SMs, DPMO, etc.

ORAI. Key Activities

Develop standards and best practices:

- Create high standards for AI governance and assess compliance tools.

Peer reviews & hackathons:

- Organize reviews (audits) and hackathons to evaluate AI systems, including generative AI.

Research:

- Conduct research on responsible AI/ML to anticipate emerging challenges.

Training and workshops:

- Offer courses and workshops to promote responsible AI practices.

Consultation services:

- Provide expert consultation on AI implementation across StatCan and beyond.

ORAI's AI Governance Products

Comprehensive Guidelines

Assessment Tools

Red-teaming & hackathons

AI Scorecards

AI Registry System

Brakes are there to let you go faster.

Guiding Principles in Action

- Grounding AI implementation in ethical governance and the Fundamental Principles of Official Statistics.

Shared Vision for Global Progress:

- By embedding accountability, transparency, and ethical considerations into every stage of AI adoption, organizations worldwide can drive innovation that is both impactful and sustainable.

Merci! / Thank You!

Questions ?

Contact:
Christos.Sarakinos@statcan.gc.ca

Avertissement/Disclaimer:
Les opinions exprimées par les auteurs de cette présentation ne reflètent pas nécessairement celles de Statistique Canada.
This presentation represents the views of the authors and does not necessarily reflect the opinions of Statistics Canada.