| Responsible AI/ML | | STC | UK | AUS | France | Swiss | EC | OECD | Other | Comments |
|---|---|---|---|---|---|---|---|---|---|---|
| **Why is responsible AI required (background and expectation)?** | | **2017 - The Montreal Declaration for a responsible development of responsible artificial intelligence**<br>"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." (https://mila.quebec/en/official-launch-of-the-montreal-declaration-for-responsible-development-of-artificial-intelligence/)<br><br>**2019 - The Government of Canada - Directive on Automated Decision-Making**<br>"The Government of Canada is increasingly looking to utilize artificial intelligence to make, or assist in making, administrative decisions to improve service delivery. The Government is committed to doing so in a manner that is compatible with core administrative law principles such as transparency, accountability, legality, and procedural fairness." (https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592)<br><br>**2019 - The Government of Canada - The Algorithmic Impact Assessment (AIA) tool** is a mandatory risk assessment tool intended to support the Treasury Board's Directive on Automated Decision-Making.<br>The tool is a questionnaire that determines the impact level of an automated decision-system. It is composed of 48 risk and 33 mitigation questions. Assessment scores are based on many factors, including systems design, algorithm, decision type, impact and data.<br>(https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html)<br><br>**2020 - Framework for Responsible Machine Learning Processes at Statistics Canada**<br>"The rapidly growing capabilities and increasing presence of ML processes in our work at Statistics Canada raise pressing questions about the impact, governance, ethics, and accountability of<br>these technologies. As indicated in the Government of Canada's Principles of Responsible Artificial Intelligence, (AI- which is inclusive of ML), the government will endeavor to:<br>1. **understand and measure** the impact of using AI by developing and sharing tools and approaches<br>2. **be transparent** about how and when we are using AI, starting with a clear user need and public benefit<br>3. **provide meaningful explanations** about AI decision making, while also offering opportunities to review results and challenge these decisions<br>4. **be as open as we can by sharing** source code, training data, and other relevant information, all while protecting personal information, system integration, and national security and defence<br>5. **provide sufficient training** so that government employees developing and using AI solutions have the responsible design, function, and implementation skills needed to make AI-based public services better<br><br>In order to put this in place, a Directive on Automated Decision-Making has been written and is based on the results from the Algorithmic Impact Assessment tool used to determined how acceptable are the AI solutions put in place from an ethical and human perspective.<br>To this end, a Framework for Responsible ML Processes has been developed at Statistics Canada. **The Framework consists of Guidelines for responsible ML and an accompanying checklist**, which are organized into four themes: **Respect for People; Respect for Data; Sound Methods; and Sound Application.** All four themes work together to ensure ethical use of both the algorithms and the results." (https://www150.statcan.gc.ca/n1/pub/89-20-0006/892000062021001-eng.pdf) | The Review into Bias in Algorithmic Decision Making (2020) provides a series of recommendations for government organisations and regulators on how to deal with risks generated by increased dependence on automated, algorithmic, decision making.<br><br>The overall plan on how to fairly govern AI as well as fairly reap its benefits on the medium and long term is captured in the National AI Strategy (2021). This is a 10 year strategy on how to grow the UK AI capability while taking into account risks and fair distribution of benefits.<br><br>The benefits and harms of algorithms: a shared perspective from the four digital regulators (2022) sets an overview of the benefits of ML / AI and related algorithms to the public. It takes into account the impact of both publicly and privately used algorithms, focus areas suggested by users, as well as the potential actions and activities that digital regulators can take. | AI ethics principles help (1) achieve safer, more reliable and fairer outcomes for all Australians, (2) reduce the risk of negative impact on those affected by AI applications, (3) businesses and government to practise the highest ethical standards when designing, developing and implementing AI.By applying the principles and committing to ethical AI practices, you can (1) build public trust in your product or organisation, (2) drive consumer loyalty in your AI-enabled services, (3) positively influence outcomes from AI, (4) ensure all Australians benefit from this transformative technology. | A recent report from the main court of administrative justice (Conseil d'Etat) emphasizes the need to build an AI of trust so that public decisions based on algorithms do not appear illegitimate to citizens. One of the administrations in charge of supporting administrations in the application of the new legal framework on public algorithms offers an online guide on the subject.<br><br>Since the subject of AI raises significant economic and geopolitical issues, to accelerate the development of AI in both the private and public sectors, the French government has adopted a national AI strategy consisting of investing 1.5 billion euros between 2018 and 2022.<br><br>This strategy is divided into three areas: (1) Adapting university training to the latest developments in the world of data science, and limiting the brain drain; (2) Accelerating the collection and use of data by public administrations. The quality of this data, rather than its quantity, has indeed been deemed essential to be able to adopt relevant algorithms by a report dedicated to data experts. (3) Promoting an ethical model that balances innovation and protection of fundamental rights.<br><br>The report of the Conseil d'Etat cited above emphasizes that the algorithms currently used in the administration are not in charge of decision-making but serve as decision support or as tools for extracting information from a large mass of data. For instance, once anomaly detection is implemented for control purposes, this method is used to identify suspicious cases where human control should be implemented.<br>It is important to distinguish the different use of algorithms depending on the purpose of the administration that implements it. The same data source can be used in very different business processes, which can affect citizens in different ways. For example, an algorithm built on tax data can be used both for public statistics, where prediction errors at the individual level are admissible as long as they are averaged out in a social group of interest, and for control, where errors can lead to actions by the administration that affect citizens. | **2020 : Guidelines on Artificial Intelligence for the Confederation** (General frame of reference on the use of artificial intelligence within the Federal Administration)<br>**2022: Federal data science strategy** (The strategy highlights how the administration uses data science. For example, rules of conduct for the application of data science are to be formulated for the federal government's administrative units. The federal administration serves as a role model for the responsible use of data science. | | The OECD has undertaken empirical and policy activities on AI in<br>support of the policy debate, starting with a Technology Foresight Forum on AI in 2016 and an international<br>conference on AI: Intelligent Machines, Smart Policies in 2017.<br><br>This work has demonstrated the need to shape a stable policy<br>environment at the international level to foster trust in and adoption<br>of AI in society. Against this background, the Recommendation on Artificial Intelligence (AI) – the first intergovernmental standard on AI – was adopted by the OECD Council at Ministerial level on 22 May 2019 on the proposal of the Committee on Digital Economy Policy (CDEP). The Recommendation aims to foster innovation and trust in AI by promoting the responsible stewardship of trustworthy AI while ensuring respect for human rights and democratic values (it covers 5 principles, **(i) inclusive growth, sustainable development and well-being; (ii) human-centred values and fairness; (iii) transparency and explainability; (iv) robustness, security and safety; and (v) accountability**, see here).<br><br>Trust is a key enabler of digital transformation. Although the nature of future AI applications and their implications may be hard to foresee, the trustworthiness of AI systems is a key factor for the diffusion and adoption of AI (Trustworthy AI refers to AI that respects the values-based principles). | | |
| **Principles** | **Sub-topics** | | | | | | | | | |
| **Governance** | Governance | **StatCan Data Science Strategy - Pillar: Governace** | | | | | | | | |
| | Data Governance | Statistics Canada Data Strategy : https://www.statcan.gc.ca/en/about/datastrategy | | | | | | | | |
| **Ethical purpose and Public Good** | Public Good | **This is one of the guidelines in the StatCan framework**: Values to Canadians (under theme: Respect for People). | Part of the UK Government Statistical Service Strategy.<br><br>The public good includes:<br><br>- informing the public about social and economic matters;<br>- assisting in the development and evaluation of public policy; and<br>- regulating quality and publicly challenging the misuse of statistics. | | From the Conseil d'Etat report to build trustable AI: "The administration must be able to demonstrate that its choice to use an AI system and the modalities of operation it has defined are guided by the concern to bring a benefit to the human community and, ideally, the greatest possible benefit." | | | | | |
| | Values to people/customers | **This is one of the guidelines in the StatCan framework**: Values to Canadians (under theme: Respect for People). | | | | | | AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and should include appropriate safeguards to ensure a fair and just society (see here). | | |
| | Means proportionate to needs | Principles of Necessity and Proportionality (statcan.gc.ca) | | | | | | | | |
| | Respect for human autonomy; societal and environmental wellbeing | | | This is one of the principles in Australia's Artificial Intelligence Ethics Framework, by the Department of Industry, Science, Energy and Resources. | | | | Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as augmenting human capabilities and enhancing creativity, advancing inclusion of underrepresented populations, reducing economic, social, gender and other inequalities, and protecting natural environments, thus invigorating inclusive growth, sustainable development and well-being (see here). | | |
| **Accountability** | Accountability | **This is one of the guidelines in the StatCan framework**: Accountability (under theme: Respect for People). | With regards to impact assessment, prevention of harm and means proportional to needs, see the UK Government Data Ethics Framework https://www.gov.uk/government/publications/data-ethics-framework | This is one of the principles in Australia's Artificial Intelligence Ethics Framework, by the Department of Industry, Science, Energy and Resources. | | | | Organisations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the OECD's values-based principles for AI (see here). | | |
| | Governance | Statistics Canada Data Strategy : https://www.statcan.gc.ca/en/about/datastrategy | | The organisation and individual accountable for a decision should be identifiable as necessary. | | | | | | |
| | Data Governance | Statistics Canada Data Strategy : https://www.statcan.gc.ca/en/about/datastrategy | | | | | | | | |
| | Impact Assessment | **1. Algorithmic Impact Assessment Tool** (GoC) - https://open.canada.ca/data/en/dataset/5423054a-093c-4239-85be-fa0b36ae0b2e<br>**2. Peer-Review Checklist** | Data Ethics Framework - GOV.UK (www.gov.uk) | | | | | | | |
| | Prevention of harm | **This is one of the guidelines in the StatCan framework**: Prevention of harm (under theme: Respect for People). | Data Ethics Framework - GOV.UK (www.gov.uk) | | | | | | | |
| | Means proportionate to needs | Principles of Necessity and Proportionality (statcan.gc.ca) | Data Ethics Framework - GOV.UK (www.gov.uk) | | | | | | | |
| | Fit for purpose | **Checklist Question**<br>How is it relevant to use this ML process in this context? Explain why this ML process is the right one to use in the current situation in terms of time to develop, required quality, type of data, question to be answered, etc.... | | | | | | | | |
| | Human oversight | **Checklist Questions**, eg.: "Is there appropriate human oversight, using good judgement and having control over the process?"<br>**StatCan Peer-review process**. Clients and subject matter experts involvment. | | There should be an appropriate level of human oversight for the particular AI system or use case. | Importance of human oversight is highlighted in the Conseil d'Etat report, drawing from the proposed European regulation on AI: "In this sense, the proposed European regulation intends to subject high-risk AIS to the principle of "human oversight", with the objective of preventing or minimising risks to health, safety or fundamental rights from AIS used for its intended purpose or under reasonably foreseeable conditions of misuse. In particular, humans must be able to fully understand the capabilities and limitations of the system, correctly interpret the results it produces, monitor the operation and detect anomalies, and decide to interrupt its operation." | | | | | |
| **Transparency** | Transparency | **This is one of the guidelines in the StatCan framework**: Transparency (under theme: Sound Application). | This is the *UK Government Algorithmic Transparency Standard* (July 2022).The standard is aimed at public sector organisations (not private sector) and is composed of the standard itself, as well as related guidance. The page also contains results from pilots with government organisations.<br><br>The standard includes capturing endpoints, SROs, organisations, models, frequency of usage, human involvement in decisions, etc. | This is one of the principles in Australia's Artificial Intelligence Ethics Framework, by the Department of Industry, Science, Energy and Resources. | | | | Transparency and responsible disclosure around AI systems to ensure that people understand when they are engaging with them and can challenge outcomes (see here). | | |
| | Explainability/Interpretability | **This is one of the guidelines in the StatCan framework**: Explainability (under theme: Sound Methods). | | Aim to be able to provide timely justifications for an AI's decision, including information like key factors used in decision-making. | | | | | | |
| | Communication to stakeholders | | | Different users have different reasons for wanting transparency. Users want to know what the system is doing and why. Creators and those deploying and operating the AI system need to understand the system's processes and input data. Regulators and accident investigators (if accidents occur) need the context on investigations. Those in legal processes need to inform evidence and decision-making. The public needs confidence in the technology. | | | | | | |
| | Disclosure of AI interaction | | | Aim to ensure that people have the ability to find out when an AI system is engaging with them. | | | | | | |
| | Sharing of methods and tools | **This is one of the guidelines in the StatCan framework**: Transparency (under theme: Sound Application). | | | Programs must be opened upon request in France: *"Any entity entrusted with a public service mission must publish any document produced or received within the framework of this mission, whatever the date, the place of conservation and the medium. Source codes, as administrative documents, fall under this obligation"* (source) | | | | | |
| | Reproducibility of process | **This is one of the guidelines in the StatCan framework**: Reproducibility of process and results (under theme: Sound Application). | | | European Statistics Code of Practice : Accuracy and Reliability (principle 12) | | | | | |
| | Reproducibility of results | **This is one of the guidelines in the StatCan framework**: Reproducibility of process and results (under theme: Sound Application). | | | | | | | | |
| **Fairness, Bias and non-discrimination** | Non-discrimination | **Checklist Questions**, eg. :<br>Discrimination is the unfair or prejudicial treatment of people and groups based on characteristics such as race, gender, age or sexual orientation. Should the resulting statistics suggest discrimination, did the team assess the source of the discrimination? If yes, explain how.? | | | Among the seven conditions listed by the Conseil d'Etat report to build trustable AI | | | | | |
| | Inclusiveness | | | | | | | | | |
| | Diversity | | | | | | | | | |
| | Accessibility | | | | | | | | | |
| | Bias | **Checklist Questions**, eg. :<br>What is your assessment of the level of bias in the learning data? | | This is particularly important given concerns about the potential for AI to perpetuate societal injustices and have a disparate impact on vulnerable and underrepresented groups including, but not limited to, groups relating to age, disability, race, sex, intersex status, gender identity and sexual orientation. Measures should be taken to ensure the AI produced decisions are compliant with anti-discrimination laws. | | | | | | |
| | Fairness | **This is one of the guidelines in the StatCan framework**: Fairness (under theme: Respect for People). | | This is one of the principles in Australia's Artificial Intelligence Ethics Framework, by the Department of Industry, Science, Energy and Resources. | | | | AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and should include appropriate safeguards to ensure a fair and just society (see here). | | |
| | Prevention of harm | **This is one of the guidelines in the StatCan framework**: Prevention of harm (under theme: Respect for People).<br>**Checklist Question :**<br>Should the model reveal specific and relevant information about vulnerable populations, is there a dissemination strategy put in place by the team? If yes, please provide an overview. | | | | | | | | |

| Responsible AI/ML | | STC | UK | AUS | France | Swiss | EC | OECD | Other | Comments |
|---|---|---|---|---|---|---|---|---|---|---|
| | Human-centred values and fairness | | | AI systems should be user-centric and designed in a way that allows all people interacting with it to access the related products or services. This includes both appropriate consultation with stakeholders, who may be affected by the AI system throughout its lifecycle, and ensuring people receive equitable access and treatment. | | | | | | |
| | Quality of service (for all) | StatisticsCanada Quality Framework | | | | | | | | |
| Privacy and Security | Privacy | This is one of the guidelines in the StatCan framework: Privacy (under theme: Respect for Data).<br><br>**Note.** "Personal information created, held, or collected by Statistics Canada is kept confidential and protected by the **Privacy Act** and by the **Statistics Act**. Data published by Statistics Canada are considered public data, are not sensitive in nature, and do not identify or provide ways to identify individuals or organizations."<br><br>**Note.** Statistics Act : https://laws.justice.gc.ca/eng/acts/S-19/FullText.html<br><br>**Note.** Statistics Canada has a **Directive on Conducting Privacy Impact Assessments** (https://www.statcan.gc.ca/en/about/pia/dcpia)<br>"Statistics Canada will conduct a specific privacy impact assessment for all new and significantly redesigned collections, uses or disclosures of personal information that raise privacy, confidentiality or data security risks if a generic, or already existing, PIA does not adequately address the risks."<br><br>**Privacy Act** (https://laws-lois.justice.gc.ca/eng/acts/P-21/)<br>"Statistics Canada is subject to the Privacy Act which means that it must ensure compliance to the privacy principles embedded in the Act, including the protection of personal information. As Statistics Canada's main legislation, the Statistics Act also provides for the protection of the information collected under its mandate, the interpretation of the two legislations for the protection of information can work together. The Privacy Act prohibits the disclosure of personal information without the consent of the individual unless specifically permitted by the Privacy Act." | See the UK Data Protection Act. It's the UK implementation of GDPR. A guide to data protection is provided by the Information Commissioners Office. | This principle aims to ensure respect for privacy and data protection when using AI systems. This includes ensuring proper data governance, and management, for all data used and generated by the AI system throughout its lifecycle. For example, maintaining privacy through appropriate data anonymisation where used by AI systems. Further, the connection between data, and inferences drawn from that data by AI systems, should be sound and assessed in an ongoing manner. | GDPR | | | | | |
| | Security | **This is one of the guidelines in the StatCan framework**: Security (under theme: Respect for Data).<br>**Note.** Statistics Canada has a Directive on the security of sensitive statistical information (http://icn-rci.statcan.ca/31/31b/31b_009-eng.html) | | This principle also aims to ensure appropriate data and AI system security measures are in place. This includes the identification of potential security vulnerabilities, and assurance of resilience to adversarial attacks. Security measures should account for unintended applications of AI systems, and potential abuse risks, with appropriate mitigation measures. | | | | | | |
| | Confidentiality | **This is one of the guidelines in the StatCan framework**: Confidentiality (under theme: Respect for Data).<br>**Note**. Statistics Canada has a **Policy on Privacy and Confidentiality** (http://icn-rci.statcan.ca/31/31a/31a_002-eng.html) | | | | | | | | |
| | Robustness, security and safety | **SECURITY COMPONENT - This is one of the guidelines in the StatCan framework**: Security (under theme: Respect for Data).<br>**Note.** Statistics Canada has a Directive on the security of sensitive statistical information (http://icn-rci.statcan.ca/31/31b/31b_009-eng.html) | | | | | | AI systems must function in a robust, secure and safe way throughout their lifetimes, and potential risks should be continually assessed and managed (see here). | | |
| | direct identification | **This is one of the guidelines in the StatCan framework**: Privacy (under theme: Respect for Data).<br>**Note**. Statistics Canada has a Directive on Conducting Privacy Impact Assessments (https://www.statcan.gc.ca/en/about/pia/dcpia) | | | | | | | | |
| | indirect identification | **This is one of the guidelines in the StatCan framework**: Privacy (under theme: Respect for Data).<br>**Note**. Statistics Canada has a Directive on Conducting Privacy Impact Assessments (https://www.statcan.gc.ca/en/about/pia/dcpia) | | | In the concept of statistical confidentiality, public statistics should not lead to direct or indirect identification of sensitive information | | | | | |
| | data security | **This is one of the guidelines in the StatCan framework**: Security (under theme: Respect for Data).<br>**Note.** Statistics Canada has a Directive on the security of sensitive statistical information (http://icn-rci.statcan.ca/31/31b/31b_009-eng.html) | | | | | | | | |
| | consent | **Privacy Act** (https://laws-lois.justice.gc.ca/eng/acts/P-21/)<br>Statistics Canada is subject to the Privacy Act which means that it must ensure compliance to the privacy principles embedded in the Act, including the protection of personal information. As Statistics Canada's main legislation, the Statistics Act also provides for the protection of the information collected under its mandate, the interpretation of the two legislations for the protection of information can work together. The Privacy Act prohibits the disclosure of personal information without the consent of the individual unless specifically permitted by the Privacy Act. | | | | | | | | |
| | permitted use of data | **Privacy Act** (https://laws-lois.justice.gc.ca/eng/acts/P-21/)<br>Statistics Canada is subject to the Privacy Act which means that it must ensure compliance to the privacy principles embedded in the Act, including the protection of personal information. As Statistics Canada's main legislation, the Statistics Act also provides for the protection of the information collected under its mandate, the interpretation of the two legislations for the protection of information can work together. The Privacy Act prohibits the disclosure of personal information without the consent of the individual unless specifically permitted by the Privacy Act. | | | Consent is one of the pillars of the GDPR. Some data sources, when processed by public authorities, benefit from an exception to the GDPR. For example, with regard to health data within the framework of the National Health Data System (SNDS): "[a person] may not, however, object to the processing of data necessary for the exercise of the missions of the State services and certain public establishments, such as, for example, monitoring an epidemic or health surveillance. " | | | | | |
| | Reliability | | | | | | | | | |
| | Safety | | | | | | | | | |
| Reliability and Safety | Reliability and Safety? | | | This principle aims to ensure that AI systems reliably operate in accordance with their intended purpose throughout their lifecycle. This includes ensuring AI systems are reliable, accurate and reproducible as appropriate.<br><br>AI systems should not pose unreasonable safety risks, and should adopt safety measures that are proportionate to the magnitude of potential risks. AI systems should be monitored and tested to ensure they continue to meet their intended purpose, and any identified problems should be addressed with ongoing risk management as appropriate. Responsibility should be clearly and appropriately identified, for ensuring that an AI system is robust and safe. | | | | | | |
| | Monitoring | | | | | | | | | |
| | Evaluation | | | | | | | | | |
| Validity | Validity of Model | **This is one of the guidelines in the StatCan framework**: Valid Inference (under theme: Sound Methods). | | | | | | | | |
| | Validity of process | **This is one of the guidelines in the StatCan framework**: Valid Inference and Rigorous modeling (under theme: Sound Methods). | | | | | | | | |
| | Technical robustness | **This is one of the guidelines in the StatCan framework**: Rigorous modeling (under theme: Sound Methods). | | | | | | | | |
| Others | | | | Contestability: This principle aims to ensure the provision of efficient, accessible mechanisms that allow people to challenge the use or output of an AI system, when that AI system significantly impacts a person, community, group or environment. The definition of the threshold for 'significant impact' will depend on the context, impact and application of the AI system in question. | | | | | | |
| **How to implement the framework (governance structure, review procedure, when to apply these principles)?** | | **Governance Structure :**<br>- StatCan Data Science Strategy - Pillar: Governance<br>- StatCan Responsible Machine Learning Centre of Expertise (RML-CoE) - Mandate<br><br>**Review Procedure:**<br>- StatCan Review Process of ML projects (transitioning to production)<br>- The review process consists in a technical/methodological review, the use of the checklist, a code review and a presentation to the (methodology) Scientific Committee. | | | | | | | | |
| **Challenges to implement the framework and solutions** | | | | | | | | | | |
| **Assessment tool / checklist / other tools** | | **Assessment tools:**<br>- Review Process includes technical/methodological review, code review and a checklist<br>- Checklist | | | | | | | | |
| **Results and latest status** | | Operational - The review process has been in place since 2020<br>Since 2020, we have reviewed ML projects for internal (StatCan) and external clients (other federal departments) | | | | | | | | |
| **Evolution of the framework inside the organisation** | | | | | | | | | | |
| **Lessons learned** | | | | | | | | | | |
| **Recommendations on how to implement the framework** | | | | | | | | | | |
| **Next steps** | | Improve and update the framework (guidelines), checklist and review process | | | | | | | | |
| **References** | | Framework for Responsible Machine Learning Processes at Statistics CanadaJuly 2020 : https://www150.statcan.gc.ca/n1/pub/89-20-0006/892020062021001-eng.pdf<br><br>Responsible use of machine learning at Statistics Canada : https://www.statcan.gc.ca/en/data-science/network/machine-learning | | | | | | | | |