



A REVIEW OF POLICY FRAMEWORKS ON THE USE OF ARTIFICIAL INTELLIGENCE IN POLICING AND THE PUBLIC SECTOR

Prepared for the
Thunder Bay Police Services Board
March 2024

Dr. Davut Akca
Assistant Professor
Lakehead University




TABLE OF CONTENTS

| | |
|---|-----------|
| Introduction..... | 2 |
| The Use of AI by Law Enforcement..... | 5 |
| Policy Frameworks | 7 |
| FEDERAL POLICY FRAMEWORKS AND REGULATIONS | 7 |
| Canadian Government’s Guiding Principles | 7 |
| Artificial Intelligence and Data Act (Bill C-27) | 8 |
| Joint Document on AI Principles by the Canadian Privacy Regulators | 9 |
| Voluntary Code of Conduct on the Responsible use of AI | 10 |
| PROVINCIAL POLICY FRAMEWORKS IN CANADA | 12 |
| Ontario’s Trustworthy Artificial Intelligence (AI) Framework | 12 |
| British Columbia – Yukon Joint Report | 13 |
| POLICIES SPECIFIC TO LAW ENFORCEMENT (IN CANADA) | 20 |
| The RCMP’s Digital Policing Strategy | 20 |
| Toronto Police Services - AI Policy | 22 |
| Privacy & Technology at the Waterloo Regional Police Service..... | 25 |
| FOREIGN AND INTERNATIONAL POLICY FRAMEWORKS | 27 |
| The U.S. Government’s Policy | 27 |
| EU’s Accountability Principles for Artificial Intelligence (AP4AI) in the Internal Security Domain..... | 29 |
| EU’S harmonized rules on Artificial Intelligence..... | 32 |
| UK Police Covenant on AI Use | 33 |
| INTERPOL’S Responsible AI Innovation in Law Enforcement Toolkit | 35 |
| REFERENCES..... | 38 |

INTRODUCTION

Artificial Intelligence (AI) is a field within computer science that aims to develop machines that imitate human intelligence. The utilization of AI in public services has been the subject of growing discourse in recent years. Law enforcement agencies are not exempt from this trend, as AI solutions can assist in enhancing efficiency, promoting data-driven practices, and expanding capabilities for specific tasks or decisions. Nevertheless, there are significant ethical considerations to take into account when implementing AI solutions, such as concerns regarding fairness, accountability, transparency, security, and privacy.

According to the [2021-2023 strategic plan](#) of the Thunder Bay Police Service (TBPS), one of the strategic objectives of the TBPS is to adapt and respond to the changing technology requirements of policing. In the strategic plan, the TBPS determined that investment in new policing technologies is one of their short-term priorities to support crime prevention, response times and solve rates. The service aims to identify and incorporate new technologies into operations to enhance police service delivery in keeping with emerging crime trends.

The recent technology initiatives of the TBPS outlined in their strategic plan include:

1. **Electronic Accident Reporting:** The service's migration to the electronic accident reporting system for its Primary Response Unit is currently in progress.
2. **BriefCam pilot project:** The service aims to use a video analytics tool that greatly speeds up the search of extremely large surveillance video files. This tool will assist investigators who are dealing with an increasing volume of digital video evidence.
3. **Next Generation 911 (NG 911):** The service will transit to the Next Generation 911 (NG 911) which offers text to 911 emergency services. This project also includes ongoing planning and coordination for the rollout of video and image services by 2023.
4. **Eye on the Street:** The service is in partnership with the City of Thunder Bay with the launch of the new Eye on the Street program using the latest technology which expands public space surveillance designed to enhance community safety and emergency response in critical situations.
5. **Digitization of data:** The service aims to improve the Administrative Services Unit through digitization of data and report processing. This includes information across court security, prisoner management, records, disclosure of evidence to prosecuting agencies, and criminal record checks.

6. **Crime mapping:** The service is transiting to the latest version of the City Protect public crime mapping tool which engages the residential and business community in crime prevention.
7. **Records Management System:** The service continues to work on the migration and integration of the Records Management System to the OPTIC collaborative in order to support information sharing with other police agencies including the O.P.P., Nishnawbe Aski Police, Anishinabek Police and Treaty Three Police.
8. **Camera Registry Program:** This program invites residents to register their home security cameras through the TBPS website, which will give the service access to their camera records and notify the service that the citizens are willing to assist in an investigation if something happens within view of their cameras.

These programs and services either currently integrate AI or have the potential to include AI-facilitated features in the near future. Particularly, the BriefCam project, which will be an extension of the Eye on Street project by creating a secure network connecting the cameras to a central control centre, integrates AI technology and requires special attention. The potential of the BriefCam software to adopt facial recognition features adds to the existing concerns regarding the ethical and responsible use of this technology.

There are serious concerns about the harmful use of AI technologies in policing and some examples of such uses have been witnessed so far. Recent [cases](#) from the U.S. showed that the irresponsible use of A.I.-driven technologies including facial recognition and automated license plate readers might lead to the [misidentification](#) of people as suspects, and thus, false arrests and seizures. Moreover, a recent [study](#) showed that the risk of being misidentified by facial recognition systems is higher for already disadvantaged groups such as people of colour, women, the elderly, and children.

To address the risks and concerns with the use of AI, the TBPS needs to have a comprehensive policy that will guide the current and future use of AI technologies to ensure effective, ethical, fair, accountable, and transparent services. Indeed, a recent public [survey](#) in Thunder Bay showed that the majority of residents support the use of this technology but also agree that the TBPSB must develop a policy to ensure proper oversight of its use.

This report provides a comprehensive review of existing artificial intelligence (AI) regulations both within Canada and internationally to guide efforts of the TBPSB to develop a policy framework for the AI use of the TBPS. The report provides a brief outline of how AI systems are used in the public sector and examines the overarching regulatory frameworks as well as those specifically tailored for law enforcement applications. The core principles and guidelines underpinning these regulatory measures will be outlined. Additionally, a

draft policy framework for the application of AI by the TBPS will be introduced, aiming to align with best practices and ethical considerations in the field.

THE USE OF AI BY LAW ENFORCEMENT

Artificial Intelligence (AI) technology has revolutionized the field of law enforcement, significantly improving the effectiveness of crime prevention, detection, investigation processes, decision-making capabilities, and training methodologies. By scrutinizing data for identifiable patterns, trends, and outliers, AI aims to facilitate the optimized distribution of resources and proactive measures against crime. Surveillance systems powered by AI have to ensure a balance between maintaining security and respecting privacy through the analysis of behavioral patterns.

To mitigate ethical concerns and prevent bias, it is crucial to develop impartial algorithms and implement effective regulatory measures. It is the responsibility of policymakers to enact careful regulation of AI applications in law enforcement to guarantee their ethical and advantageous usage. Policymakers must carefully regulate AI in policing to ensure responsible and beneficial implementation.

A review of the police practices by [the U.S. National Institute of Justice](#) yielded the following areas of AI use in law enforcement:

- 1. Automated License Plate Readers (ALPRs):** ALPRs, now enhanced by AI, are used extensively by law enforcement and private companies for surveillance, including creating "virtual fences" to identify which vehicles enter and exit a jurisdiction and automating traffic violation tickets.
- 2. Video and Photo Surveillance:** AI advancements allow cameras to run algorithms directly, facilitating real-time facial recognition and weapons detection without significant cost or bandwidth.
- 3. Redaction to Reduce Systemic Bias:** AI is employed to automatically redact personal information (race, ethnicity, religion etc.) in police narratives to mitigate bias in the criminal justice system, aiming for fairer prosecutorial decisions.
- 4. Gunshot Detection and Mapping:** Incorporating AI, systems like [ShotSpotter](#)[®] detect and locate gunshots, with new technologies enabling pre-shot detection. Research continues to evaluate these AI systems' impact.
- 5. Combatting Human Trafficking and Child Predators:** AI technologies, like facial recognition by [Thorn](#), assist in identifying missing children and combating child exploitation by scanning internet ads and the dark web for pictures of known missing children.
- 6. Video Redaction:** The need for video redaction in [police body cam footage](#) has led to the development of AI-driven solutions, significantly reducing the manual effort required.

- 7. AI-Enabled Transcription:** Automatic speech recognition software is revolutionizing law enforcement reporting, improving accuracy and efficiency in documentation and investigative interviews.
- 8. Computer-Aided Dispatch (CAD):** AI integration into CAD systems enables optimized resource allocation and decision-making, enhancing emergency response and law enforcement operations.
- 9. Hotspot Mapping or Predictive Policing:** AI augments predictive policing models, analyzing data to identify crime hotspots and individuals at risk, though concerns about systemic bias persist. Some examples of AI use in predictive policing include:
 - Bail algorithms that predict likelihood of being arrested or failure to appear;
 - Sentencing algorithms that predict likelihood of being arrested;
 - “Scoring at arrest” algorithms that advise how to charge an individual;
 - “Scoring suspects” algorithms that analyze an individual’s future behaviour;
 - “Scoring victims” algorithms that predict likelihood of being a victim of crime;
- 10. Improving Police-Community Relations:** AI-powered chatbots facilitate communication between law enforcement and communities, improving service and engagement.
- 11. Improving Case Clearance Rates:** Machine learning is applied to extensive [homicide databases](#) to improve investigation outcomes and case clearance rates, testing tools based on deep learning algorithms.
- 12. Social Media Monitoring:** The [online behaviours](#) of people (e.g., posts, emojis, friends) are monitored through AI to predict risky behaviour and cross-reference this information with private data to create a holistic profile of suspects.
- 13. Emerging AI Applications in Law Enforcement:** The evolution of AI applications in law enforcement continues with the development of technologies like [robotic officers](#), [surveillance systems](#), [DNA analysis](#), [gunshot detection](#), and [crime forecasting](#).

Taken together, AI technologies offer significant potential to enhance operational effectiveness, encourage the adoption of data-oriented strategies, and broaden the functional scope of law enforcement agencies. The core challenge lies in these agencies pinpointing scenarios where the integrity and accessibility of data, the advancement of technology, and moral considerations align with both their objectives and the expectations of the communities they serve. It is imperative for law enforcement entities, the communities they serve, and the judicial framework to actively engage in dialogues addressing the delicate balance between individual privacy rights and the collective need for safety and security, especially as AI technologies afford more intricate methods for monitoring and investigative work.

POLICY FRAMEWORKS

The recent developments in AI technologies and the ever-growing scope of the use of AI systems in policing underscores the urgent necessity for a well-rounded policy structure to govern the application of AI within law enforcement sectors, ensuring that technological deployment is both ethical and effective. This section outlines the current legislative and policy frameworks for the use of AI in the public sector that are either in effect or currently being developed.

First Canadian policy frameworks at the federal, provincial, and municipal (police forces) level are outlined. Then some best practices from the U.S. and Europe are provided as foreign policy frameworks. There are a very limited number of policy frameworks that are specific to law enforcement, which indicates the need for more research and conversation among stakeholders to identify the priorities and principles in this field.

FEDERAL POLICY FRAMEWORKS AND REGULATIONS

CANADIAN GOVERNMENT'S GUIDING PRINCIPLES

The federal government's guiding principles for the effective and ethical use of AI:

1. Promoting openness about how, why, and when AI is used;
2. Prioritizing the needs of individuals and communities, including Indigenous peoples, and considering the institutional and public benefits of AI;
3. Assessing and mitigating the risks of AI to legal rights and democratic norms early in the lifecycle of AI systems and following their launch;
4. Ensuring training or other input data used by AI systems is lawfully collected, used, and disclosed, taking account of applicable privacy and intellectual property rights;
5. Evaluating the outputs of AI systems, including generative tools, to minimize biases and inaccuracies, and enabling users to distinguish between AI and human outputs;

6. Publishing legal or ethical impact assessments, source code, training data, independent audits or reviews, or other relevant documentation about AI systems, while protecting privacy, government and national security, and intellectual property;
7. Explaining automated decisions to people impacted by them and providing them with opportunities to contest decisions and seek remedies, which could involve human review, where applicable;
8. Encouraging the creation of controlled test environments to foster responsible research and innovation;
9. Establishing oversight mechanisms for AI systems to ensure accountability and foster effective monitoring and governance throughout the lifecycle;
10. Assessing and mitigating the environmental impacts of the training and use of AI systems, and where appropriate opting for zero-emissions systems;
11. Providing training to civil servants developing or using AI so that they understand legal, ethical, and operational issues, including privacy and security, and are equipped to adopt AI systems responsibly; and
12. Creating processes for inclusive and meaningful public engagement on AI policies or projects with a view to raising awareness, building trust, and addressing digital divides.

ARTIFICIAL INTELLIGENCE AND DATA ACT (BILL C-27)

The Canadian federal government has introduced Bill C-27 which aims to regulate the use of AI in Canada. Bill C-27 is presently before the House of Commons and has passed a second reading as of April 24, 2023. If enacted, Bill C-27 would create the *Artificial Intelligence and Data Act* (AIDA). AIDA introduces a principles-based approach that is focused on ensuring that the use of AI is properly governed and controlled.

AIDA is primarily concerned with preventing harm to individuals, damage to property, and economic loss, including by preventing biased outputs of AI systems. AIDA targets “high-impact” AI systems and aims to mitigate risks involved with the use of such AI systems. The range of persons that are subject to AIDA compliance is broadly scoped to include developers, providers, and managers of AI systems. As a result, persons developing,

utilizing, and commercializing AI systems must be aware of the requirements set out by AIDA and the forthcoming regulations under AIDA.

[Canadian Civil Liberties Association \(CCLA\)](#) criticized Bill C- 27 by stating that the bill “inappropriately frames people’s privacy rights as something to be balanced against and placed below commercial interests.” According to CCLA, the bill “fails to capture the complexity of the harms and risks that AI can bring to bear on individuals, communities, and their fundamental rights.”

To address the gap in this bill, the CCLA recommended an amendment to ensure that the law will “recognize privacy as a fundamental human right, legislate stronger protections for personal information deemed sensitive, and improve concerning provisions that underplay individual consent and the harms that stem from reckless and non-consenting collection, use, and disclosure of personal information.”

JOINT DOCUMENT ON AI PRINCIPLES BY THE CANADIAN PRIVACY REGULATORS

Principles for Responsible, Trustworthy and Privacy-protective Generative AI Technologies

In December 2023, the federal, provincial and territorial privacy authorities of Canada announced that they have developed a set of principles to advance the responsible, trustworthy and privacy-protective development and use of generative artificial intelligence (AI) technologies in Canada. The principles address the developers and providers of these technologies as well as the organizations that will use them. The developers, providers, and users of AI technologies were invited to give particular consideration to identifying and preventing risks to vulnerable groups, including children and groups that have historically experienced discrimination or bias.

Key Principles

- **Legal Authority and Consent:** Ensure legal authority for collecting and using personal information; when consent is the legal authority, it should be valid and meaningful.
- **Appropriate Purposes:** Collection, use and disclosure of personal information should only be for appropriate purposes.

- **Necessity and proportionality:** Establish the necessity and proportionality of using generative AI, and personal information within generative AI systems, to achieve intended purposes.
- **Openness:** Be open and transparent about the collection, use and disclosure of personal information and the potential risks to individuals' privacy.
- **Accountability:** Establish accountability for compliance with privacy legislation and principles and make AI tools explainable.
- **Individual Access:** Facilitate individuals' right to access their personal information by developing procedures that enable it to be meaningfully exercised.
- **Limiting Collection, Use, and Disclosure:** Limit the collection, use, and disclosure of personal information to only what is needed to fulfill the explicitly specified, appropriate identified purpose.
- **Accuracy:** Personal information must be as accurate, complete, and up-to-date as is necessary for the purposes for which it is to be used.
- **Safeguards:** Establish safeguards to protect personal information and mitigate potential privacy risks.

VOLUNTARY CODE OF CONDUCT ON THE RESPONSIBLE USE OF AI

In September 2023, a voluntary code was developed and signed by the 22 developers and managers of the AI systems in Canada. The code aims to address and mitigate the risks of AI technologies and signatories to this code commit to adopting the identified measures that should be applied in advance of binding regulation pursuant to the *Artificial Intelligence and Data Act*.

The developers and managers who signed this code committed to working to achieve the following outcomes:

- **Accountability** – Firms understand their role with regard to the systems they develop or manage, put in place appropriate risk management systems, and share information with other firms as needed to avoid gaps.
- **Safety** – Systems are subject to risk assessments, and mitigations needed to ensure safe operation are put in place prior to deployment.
- **Fairness and Equity** – Potential impacts with regard to fairness and equity are assessed and addressed at different phases of development and deployment of the systems.

- **Transparency** – Sufficient information is published to allow consumers to make informed decisions and for experts to evaluate whether risks have been adequately addressed.
- **Human Oversight and Monitoring** – System use is monitored after deployment, and updates are implemented as needed to address any risks that materialize.
- **Validity and Robustness** – Systems operate as intended, are secure against cyber attacks, and their behaviour in response to the range of tasks or situations to which they are likely to be exposed is understood.

ONTARIO'S TRUSTWORTHY ARTIFICIAL INTELLIGENCE (AI) FRAMEWORK

The Ontario Government's Trustworthy Framework has been developed since 2021 and it consists of policies, products and guidance that aim to enable the transparent, responsible and accountable use of AI by the Ontario government.

The Framework will be grounded in three strategic priorities:

- 1. No AI in secret:** This means that we will provide a clear understanding of how and when AI is used.
- 2. AI use the people of Ontario can trust:** This means that we must clearly define the risks of AI use and work to prevent them to proactively protect the people of Ontario.
- 3. AI that serves all the people of Ontario:** This guarantees that the right processes will be in place to challenge decisions made with the use of AI.

The Framework proposes the [Beta Principles for Ethical Use](#) of AI that set out six points to align the use of data-enhanced technologies within government processes, programs, and services with ethical considerations and values:

- 1. Transparent and explainable:** There must be transparent use and responsible disclosure around data-enhanced technology like AI, automated decisions and machine learning systems to ensure that people understand outcomes and can discuss, challenge and improve them. This includes being open about how and why these technologies are being used. When automation has been used to make or assist with decisions, a meaningful explanation should be made available. The explanation should be meaningful to the person requesting it. It should include relevant information about what the decision was, how the decision was made, and the consequences.
- 2. Good and fair:** Data-enhanced technologies should be designed and operated in a way throughout their life cycle that respects the rule of law, human rights, civil liberties, and democratic values. These include dignity, autonomy, privacy, data protection, non-discrimination, equality, and fairness.

- 3. Safe:** Data-enhanced technologies like AI and ML systems must function in a safe and secure way throughout their life cycles and potential risks should be continually assessed and managed. Designers, policy makers and developers should embed appropriate safeguards throughout the life cycle of the system to ensure it is working as intended. This would include mechanisms related to system testing, piloting, scaling and human intervention as well as alternative processes in case a complete halt of system operations is required. The mechanisms must be appropriate to the context and determined before deployment but should be iterated upon throughout the system’s life cycle.
- 4. Accountable and responsible:** Organizations and individuals developing, deploying or operating AI systems should be held accountable for their ongoing proper functioning in line with the other principles. Human accountability and decision-making over AI systems within an organization needs to be clearly identified, appropriately distributed and actively maintained throughout the system’s life cycle. An organizational culture around shared ethical responsibilities over the system must also be promoted. Where AI is used to make or assist with decisions, a public and accessible process for redress should be designed, developed, and implemented with input from a multidisciplinary team and affected stakeholders. Algorithmic systems should also be regularly peer-reviewed or audited to ensure that unwanted biases have not inadvertently crept in over time.
- 5. Human-centric:** AI systems should be designed with a clearly articulated public benefit that considers those who interact with the system and those who are affected by it. These groups should be meaningfully engaged throughout the system’s life cycle, to inform development and enhance operations. An approach to problem-solving that embraces human-centered design is strongly encouraged.
- 6. Sensible and appropriate:** Every data-enhanced system exists not only within its use case, but also within a particular sector of society and a broader context that can feel its impact. Data-enhanced technologies should be designed with consideration of how they may apply to a particular sector along with awareness of the broader context. This context could include relevant social or discriminatory impacts.

BRITISH COLUMBIA – YUKON JOINT REPORT

In June 2021, the BC and Yukon Information and Privacy Commissioners and Ombudspersons published [a joint report](#) on the use of AI in the public sector called “Getting Ahead of the Curve: Meeting the Challenges to Privacy and Fairness Arising from the Use of Artificial Intelligence in the Public Sector.” In the report, the challenges of the AI

systems in the public sector and current regulatory measures and instruments developed in different jurisdictions were reviewed. Based on this review, they provided detailed and implementable guidelines on incorporating administrative fairness and privacy obligations across the different stages of the use of AI systems in the public sector.

Fairness-by-design framework

Ensuring that AI-enabled government decision-making is held to the same administrative fairness standards as human-based processes requires thoughtful consideration of how fairness-by-design principles should factor into the AI lifecycle. The goal of elucidating a fairness-by-design framework with respect to AI is to translate the established requirements of administrative fairness into the context of AI decision-making.

1. Fair procedure: Administrative decision-makers in government must follow a fair procedure in making decisions. Administrative decisions are decisions of government that are not legislative or broadly based on policy direction. Flowing from administrative decisions is a duty to act fairly and make procedurally fair decisions. This duty exists as a safeguard for people in their interactions with government, as decisions made by administrative bodies can have a serious and long-lasting impact on individuals' lives. Below are four requirements of a fair procedure that must be met in every case:

- **Adequate notice:** the person affected by the decision must be given adequate information to be able to participate meaningfully in the decision-making process (e.g., informed of the key issues in the decision process).
- **Fair hearing:** the person affected is given a reasonable opportunity to present their case or to respond to the facts presented by others. Moreover, the decision-maker has genuinely considered what the person has presented to them when making their decision.
- **Absence of bias:** the decision-maker decides with impartiality and independence. The term "impartiality" refers to the state of mind or attitude of the decision-maker and demands that there be no bias on this level, either real or perceived. Independence demands that the decision-maker not have ties with anyone that could lead to a reasonable doubt about their impartiality.
- **Justifiability:** the exercise of public power must be justified, intelligible and transparent, not in the abstract, but to the individuals subject to it. This does not always require formal reasons and may also be justified in relation to the constellation of law and facts that are relevant to the decision.

Integrating these fair procedure requirements in automated administrative decision-making (ADM) will require:

a) Publicly available, plain language descriptions and information about any ADS system that:

- Explains the organizational goal, purpose, or intent of the ADS, including the intended uses and out-of-scope uses as envisioned by the designers;
- Details what the system is doing as it interacts with persons (e.g., “looking” at our faces to gauge our expressions, pooling personal information from various sources, etc.);
- Explains how the components of the ADS work to enable or support lawful decision-making, including how criteria for automated processing and the processing itself are consistent with the decision-making criteria found in law and regulation;
- Provides a description of the data used to train and test the system (i.e., detailing the type of personal information being used and from what sources) and a link to the de-identified training and test data if the data is public domain so that users can understand the basis upon which decisions are reached; and
- Gives advance notice to individuals that an ADS will be used to render a decision, along with clear steps on how the decision will be made.

b) Giving users the means to appeal an ADS decision by:

- Providing users with a meaningful, plain language explanation of the steps and processes undertaken to arrive at a decision in their case; and
- Making publicly available, in plain language, reports, recommendations or other results arising from testing, monitoring, training, or auditing processes, so that people can contest an ADS decision with information regarding known or potential system issues.

c) Building confidence and trust in the quality of ADS decisions by:

- Ensuring that systems undergo periodic review, testing, and monitoring, and administrators undergo training as required:
 - Review: All ADS should be subject to risk assessments and systems deemed a substantial risk should, before implementation and mainstreaming, undergo peer review by several independent, well-positioned experts from independent oversight bodies, government ministries or agencies, academia, or NGOs with the relevant capacity and expertise;
 - Testing: The ADS, and its training and test data, should be fit-for-purpose and tested for relevance, accuracy and unintended data biases that may unfairly impact outcomes;
 - Monitoring: The processes and outcomes of an ADS should be periodically monitored to ensure compliance with applicable legislation, regulation and to safeguard against unintended outcomes;

- Education: The administrator of an ADS should be educated in the design and functionality of the system on a reoccurring basis;
- Evaluation and public reporting: Existing safeguards for the ADS, including the measures above, should undergo an independent and continuous evaluation and any findings of concern should be reported and made publicly available as soon as possible. Human control is critical to fairness in AI.

Human intervention during the AI design cycle and human monitoring of AI in its operation ensures the system is performing as anticipated (human-on-the-loop). Similarly essential is establishing what tasks or responsibilities humans transfer to AI and ensuring the ability to override a decision made by AI (human-in-command).

2. Fair decision: Case law also imposes an obligation on administrative decision-makers to give adequate reasons for their decisions, which is different from the procedural fairness requirement to give reasons. The reasons that underpin the decision must be based on “an internally coherent and rational chain of analysis that is justified in relation to the facts and law that constrain the decision maker.” In other words, there must be a rational connection linking the relevant evidence and the decision maker’s arguments and conclusions, including a clear explanation of how relevant legislation, regulation or policy was followed and applied. Decision-makers should also be able to explain that evidence was rejected and why it was rejected.

Integrating these fair decision requirements into autonomous administrative decision-making will require AI developers to design AI systems with an auditing function that is capable of:

- Identifying authorized decision-makers under the applicable legislation and the version of the system used to render the decision;
- Pinpointing all decision points or recommendations generated by the system;
- Linking decision points or recommendations within the system’s logic to relevant law or policy;
- Generating a notification of the decision, including a statement of reasons, where required; Integrating change control processes to track modifications to the system’s operations;
- Detailing the level and nature of human involvement in the decision-making process, logging instances where a human override of the system has occurred and identifying the natural person involved; and

- Incorporating the full discretion afforded to administrative decision-makers by law to leave an appropriate level of space for human judgment.

Imposing an obligation on organizations to track how their ADS works and maintain an audit trail of ADS decisions is a recommended measure for overcoming the lack of algorithmic transparency in closed-source, proprietary systems. This gives the individual impacted by ADS a better chance of appealing a decision made by ADS in a meaningful manner with sufficient precision. This also ensures that the bodies that review a decision can evaluate the fairness of the decision by examining the process used to arrive at the decision and outcome.

3. Fair service: Fair service regards how public bodies treat people who access their services. The principle of fair service poses the notion of ‘user’ interests as an obligation for the person responsible for providing the service. In the AI context, automated decision systems (ADS) should operate in a ‘human-centric’ manner. Developers and administrators of ADS should carefully consider and, where appropriate, integrate public feedback to manage continuous improvements as part of making sure that the service is fit-for-purpose, sufficiently individualized and does not produce uneven impacts or discriminatory outcomes.

a) In designing and developing an AI system:

- Ensure that algorithmic decision-making is appropriate for the proposed domain of application and does not run a foreseeable risk of producing bias, discriminatory outcomes, infringing on any other individual rights, negatively impacting public health or safety, or amplifying digital inequalities;
- Envision and design an ADS that can be more easily updated and maintained to facilitate continuous system improvement;
- Train and test AI systems using only quality data that is fit-for-purpose, and be transparent about the data’s accuracy, completeness, timeliness, update frequency, and uncertainty;
- Consider the use of synthetic data where possible to reduce privacy risks; and
- Engage with the public at the initial stages of the design and development of AI that is going to be used on the public as it is helpful for anticipating unintended consequences early on and builds trust with the public if they know how they will be impacted.

b) Before deploying an AI system that will be used by the public:

- Make sure that algorithms have gone through adequate training and tests to develop their predictive capacities

- Make sure that an ADS can perform its intended function with a high degree of predictive or explanatory power;
 - Have ready processes for genuine consultation with internal and external stakeholders to ensure that the AI system will meet the needs of target groups and garner a “social licence to operate”; and
 - Implement accessible mechanisms for people to raise concerns and appeal decisions made by AI.
- c) Once the AI system is deployed:
- Evaluate behaviours and outcomes as each new algorithm is introduced and continue to monitor them once a program is established to understand longer-term effects;
 - Introduce adequate mechanisms to collect, respond to and integrate critical feedback from users of AI systems for the purposes of ongoing quality review and continuous service improvement; and
 - Earmark resources for maintenance and improvement of the ADS.

Privacy rights and AI

In the BC-Yukon joint report, it is stated that the existing rules are not nimble enough to account for AI uses that would improve program and service delivery. They propose the following recommendations to improve the responsiveness of legislation and compliance tools to meet this challenge.

1. Rights-based approach to privacy: A robust rights-based approach to privacy is missing from Canada’s privacy laws at the federal, provincial and territorial levels. Unlike other jurisdictions where they have recently modernized privacy law (e.g., the GDPR and EU AI Regulation), there is currently no Canadian law in force that addresses rights or obligations relating directly to AI. Quebec’s Bill 64 goes further than Canada’s Bill C-11 in this regard. Neither law is in force yet. A modern interpretation of the right to privacy as a human right is necessary for the exercise of other fundamental rights. At a minimum, privacy legislation should be amended to include the right to notification that ADS is used, an explanation of the reasons and criteria used, and the ability to object to the use of ADS.

2. Adjusting compliance provisions and tools: For compliance purposes, government and the private sector should be required to assess the privacy impacts before implementing AI technology. This obligation should be ongoing and verifiable through proactive audits by regulators once the technology is deployed.

3. Standards for security safeguards, including third-party processing: AI can play a role in collecting, transmitting, processing, and destroying PI and needs to be designed

with adequate safeguards for processing PI. The current lack of explicit standards alongside the risk imposed by the use of third-party systems makes current requirements inadequate.

- The use of third-party solutions for ADS and other AI processing of PI must be balanced by requirements for transparency regarding this processing, including reporting on, and explicit standards for, security safeguards.
- This could include an obligation on public bodies to have the third parties they contract with prove compliance with their product or service with the security standard. For example, this could be done by means of demanding that third parties are (security standards) certified, and periodically validating the certification when these third parties process sensitive PI.
- Compliance with standards is no silver bullet, but it does provide a certain baseline and proof of due diligence. Compliance can be supported with proactive measures such as bug bounty programs and penetration testing of AI products or services.

4. Oversight of de-identified and synthetic data: With the compilation of massive amounts of data in recent years – some of which is publicly available – the de-identification of PI alone is an increasingly weak safeguard for the protection of privacy. Even when a name is stripped from a dataset, a combination of unique data points can be used to identify an individual with a high degree of certainty. If a dataset used for cross-reference contains a name, re-identification can be performed. Even if the dataset contains no name, the dataset still constitutes PI and is still a compliance risk to the controller because new datasets may become available that then enable re-identification.

POLICIES SPECIFIC TO LAW ENFORCEMENT (IN CANADA)

THE RCMP'S DIGITAL POLICING STRATEGY

The RCMP's use of facial recognition technology provided by a US company called Clearview AI has recently been a serious concern in the public agenda. The force stopped using this technology in 2020 upon [the investigation by the Office of Privacy Commissioner \(OPC\)](#). The OPC's investigation found that Clearview AI's technology allowed law enforcement and commercial organizations to match photographs of people against the company's databank of more than three billion images scraped from internet websites without users' consent. The OPC concluded that this represented mass surveillance and was a clear violation of the *Personal Information Protection and Electronic Documents Act* (PIPEDA), Canada's federal private sector privacy law.

Soon after the OPC's investigation, other police forces in Canada such as the [Ontario Provincial Police](#), [York Regional Police](#), [Waterloo Regional Police](#), [Peel Regional Police](#), and [Halton Regional Police](#) services also confirmed that they used the Clearview AI systems for facial recognition.

Currently, the RCMP does not have a specific and publicly available policy that regulates the use of AI by the force. Instead, the RCMP's official website provides a summary of their digital policing strategy called the Connected RCMP. This strategy does not provide a detailed explanation of how the force will address the concerns and risks posed by different types of AI technologies used by the force.

[The Connected RCMP](#)

In 2023, the RCMP created a digital policing strategy, called the Connected RCMP, which focuses on ensuring the organization has the right technology required to deal with the digital era's impacts on policing.

The core of the strategy deals with how the force will use digital services and technology to better and more quickly connect to the communities they serve and the stakeholders.

The Digital Policing Strategy aims to:

- provide new digital tools to enhance public and employee safety and service
- make better use of data to predict, prevent and fight crime
- provide new internal channels for information sharing within the RCMP

- introduce new digital channels for public engagement and service to the partners

The Connected RCMP is organized into four themes:

1. Our communities

Connecting to our communities is about how we serve and protect Canadians. People today expect to communicate digitally and from anywhere. To meet the needs of a modern public, we have to connect with our communities using modern methods. From online crime reporting to smartphone apps, the future of the RCMP is mobile and online. The targets of this theme are:

- Online crime reporting
- Next generation 9-1-1

2. Our partners

The borderless nature of modern crime means that cooperation and communication with partners is more important than ever. Under The Connected RCMP, we'll adopt more efficient digital methods of exchanging information with law enforcement partners and delivering services. The targets of this theme are:

- Modern electronic disclosure system
- Online law enforcement portal for collaboration and service delivery

3. Each other

A modern workplace should be efficient and flexible, allowing employees to work from anywhere, anytime. Mobile devices and apps will mean that employees can access RCMP systems no matter where they happen to be, without returning to a detachment or being physically in an office. Our employees will have instant access to key information during critical events, which will increase officer safety. The targets of this theme are:

- Situational awareness applications
- Android smartphones
- Wi-Fi in all RCMP facilities

4. Information

The RCMP collects and stores large amounts of information. The methods and systems used to do this often vary across the country. Through The Connected RCMP, we'll create better, more efficient processes for collecting, storing and using data to make better policing and business decisions. The targets of this theme are:

- Modern operational records management system
- Electronic document management system
- Digital evidence management

The RCMP webpage that describes this strategy ends with a note that the strategy must be updated as the technologies continually evolve. It states that the force will update this document annually to keep pace with how new and emerging technologies are shaping policing in the digital era.

TORONTO POLICE SERVICES - AI POLICY

The only Canadian policy framework that is specific to the AI use of law enforcement is the Toronto Police Service's "[Use of Artificial Intelligence Technology](#)" policy created by the Toronto Police Services Board in February 2022.

The purpose of the policy framework is "to establish Board governance for the consideration of the use of new or enhanced technologies using AI, or of previously approved AI technology that is to be used for a novel purpose or in a novel circumstance, and to establish an assessment and accountability framework"

The guiding principles of the policy are:

- **Legality:** All technology used, and all use of technology, must comply with applicable law, including the Police Services Act (and its regulations, as well as successor legislation), Ontario's Human Rights Code, and the Canadian Charter of Rights and Freedoms, and be compatible with applicable due process and accountability obligations.
- **Fairness:** Use of AI technology must not result in the increase or perpetuation of bias in policing and should diminish such biases that exist.
- **Reliability:** AI technology must result in consistent outputs or recommendations and behave in a repeatable manner. ☑ **Justifiability:** The use of AI technology must

be shown to further the purpose of law enforcement in a manner that outweighs identified risks.

- **Personal Accountability:** Service Members are accountable, through existing professional standards processes, for all the decisions they make, including those made with the assistance of AI technology or other algorithmic technologies. ²
Organizational Accountability: All use of AI technology must be auditable and transparent, and be governed by a clear governance framework.
- **Transparency:** Where the Service uses AI technology that may have an impact on decisions that affect members of the public, the use of that technology must be made public to the greatest degree possible. Where full transparency may unduly endanger the efficacy of investigative techniques or operations, the Service will endeavour to make publicly available as much information about the AI technology as possible, to assure the public of the reliability of the AI technology and the justifiability of its use. Where a decision assisted by AI technology may lead to the laying of criminal or other charges against an individual, the possible influence of the AI technology must be included in the disclosure provided to the Crown.
- **Privacy:** Use of AI technology must, to the greatest degree practicable, preserve the privacy of the individuals whose information it collects in line with ‘privacy by design’ principles.
- **Meaningful Engagement:** The adoption of specific AI technologies must be preceded by meaningful public engagement commensurate with the risks posed by the technology contemplated.

The policy consists of four sections:

- Review and Assessment of New AI Technologies
- Board Approval and Reporting Prior to Procurement, Utilization and Deployment
- Monitoring and Reporting
- Continuous review

A summary of these sections is provided below. The full policy can be found [here](#).

1. **Review and Assessment of New AI Technologies:** This section suggests that when a new AI technology is considered by the police service a comprehensive review and assessment should be conducted in collaboration with various stakeholders, including the Information and Privacy Commissioner of Ontario, the Ministry of the Attorney General, and the Anti-Racism Directorate. This initiative sets forth guidelines to ensure that AI technologies are adopted responsibly. The section also states that the members of the Toronto Police Service members are prohibited from

using new AI technologies without prior approval and training. A framework is established to classify AI technologies into risk categories, ranging from Extreme Risk to Minimal Risk, based on their potential to cause harm. It is stated that a detailed risk assessment, including a privacy impact analysis for each risk level, must be conducted. Based on the results of the assessment necessary risk mitigation measures, like training and contingency planning, are required for each level of risk. The procedures, including a detailed risk assessment tool, will be made publicly available on the Service's website.

2. Board Approval and Reporting Prior to Procurement, Utilization and

Deployment: This section mandates a thorough risk assessment for new AI technologies before procurement, use, or deployment. This includes securing funds, acquiring technology without financial exchange, novel usage or circumstances, or entering into agreements. Technologies identified as Extreme Risk will not be adopted, and those classified as High or Moderate Risk require Board approval. Low-risk technologies will be promptly reported to the Board with a justification of the risk level. The assessment process encompasses operational needs, intended use, risk level justification, legislative compliance, operation details including data management, vendor evaluation, privacy impact assessments, consultation feedback, legal and human rights analysis, risk mitigation, cost estimation, and tracking indicators for effectiveness and unintended consequences. Furthermore, a public engagement strategy and communication plan for judicial authorization or impact on criminal proceedings will be developed. The Board retains the authority to review, request independent evaluations, require additional analyses, approve pilot programs, or specify further requirements for AI technology deployment.

3. Monitoring and Reporting: This section mandates the Chief of Police to closely monitor and report on the deployment of new AI technologies based on their risk levels. For technologies deemed Moderate risk, monitoring will continue for 12 months post-deployment, and for High risk, for 24 months. Reports to the Board are due within 15 months for Moderate risk and 27 months for High risk AI technologies, covering deployment details, compliance with laws, performance metrics, public and internal concerns, and consultation outcomes. The Chief must also decide whether to continue using the AI, under which conditions, and which performance indicators will be tracked indefinitely to ensure quality and identify any unintended consequences. Additionally, a platform will be established for the public to raise concerns about AI technologies, with these concerns being reported to the Board

either as part of the scheduled reporting or annually. The Board will then review these reports to decide on the continued use of the AI technology and any further requirements needed. All reports and decisions are to be discussed in public Board meetings, except for confidential information, which will be shared privately with the Board.

- 4. Monitoring and Reporting:** In the last section of the policy the Chief of Police is directed to undertake a comprehensive risk analysis of all AI technologies currently utilized by the Service, with a deadline of December 2024 for completion and subsequent reporting of findings. There's an immediate requirement to publicly list all AI technologies in use, categorized by risk levels, and include detailed information about each technology, such as its name, manufacturer, purpose, usage, and data collection details. Any AI technology identified as Extreme risk and in use before this policy's adoption will be immediately discontinued, with the Board being informed about the decision, the technology's details, and the assessment of potential harms. Technologies deemed High or Moderate risk, also predating this policy, will be reported to the Board with a detailed action plan for pause, risk evaluation, mitigation, and possibly, continued use upon Board approval. The policy mandates regular reviews of the use of AI technologies based on their risk levels to ensure their ongoing necessity and alignment with the policy's objectives. Furthermore, the Board commits to reviewing this policy every three years to assess its effectiveness and adjust as needed to prevent risk misclassification.

PRIVACY & TECHNOLOGY AT THE WATERLOO REGIONAL POLICE SERVICE

On the [website](#) of the Waterloo Regional Police Service, a brief policy statement on the use of technology by the service is given. WRPS also uses the BriefCam software for video surveillance. There is also a brief explanation of how this software is used by the force and how its facial recognition system is different than the one used by the RCMP.

The exploration and use of technology are essential for WRPS to meet its obligations to the community regarding public safety, including the prevention and investigation of crimes, as well as to improve overall administration. Technologies are assessed to protect privacy and security while ensuring the public has access to police information as outlined in the [Municipal Freedom of Information and Protection of Privacy Act](#) (MFIPPA).

WRPS is committed to assessing the impacts of new and existing technology, procedures and programs with access and privacy at the forefront, as well as to ensure compliance with the Criminal Code of Canada, the Charter of Rights and Freedoms, the Police Services Act, the Youth Criminal Justice Act and any other relevant laws or legislation. As such, information is collected through lawful authority, judicial authorization or upon consent.

We continue our commitment to providing citizens with responsive policing services that foster a relationship of trust and transparency within our community.

Image Analytics

Image Analytics Technologies are utilized by authorized police officers to view, process and analyze lawfully obtained photographs, video footage, etc. for specific images that are relevant to law enforcement investigations or prosecutions. The purpose of using this technology is to expedite the process of locating objects or individuals within the lawfully obtained video.

BriefCam

BriefCam is a new program utilized by WRPS in 2022. BriefCam software can quickly search volumes of video that would otherwise be impossible to examine manually, providing investigative clues that create intelligence and operational information for officers. BriefCam does not expand the collection of personal information by investigators.

BriefCam has a module called "face recognition service"; however, it is important to recognize that it is not the facial recognition that has been the source of public scrutiny. This "face recognition service" is more akin to "object recognition." It allows an investigator to select an object, such as a red car, a specific licence plate, a black sweater, or a person deemed to be of interest, in the video being reviewed. The software will then find all instances of that "object" within the already lawfully obtained video in a couple of minutes rather than watching hours and hours of video footage.

FOREIGN AND INTERNATIONAL POLICY FRAMEWORKS

In this section, policy and regulatory framework examples from the U.S., EU, and the UK will be summarized. The frameworks are not specific to any individual police force. Rather, they provide general guidance to law enforcement agencies across their jurisdiction.

THE U.S. GOVERNMENT'S POLICY

In October 2023, President Biden issued an [executive order](#) on the use of AI in the public sector. The executive order focuses on establishing stringent standards for AI safety, security, and trustworthiness to ensure America's leadership in AI innovation while managing its risks. It mandates developers of significant AI systems to disclose safety tests and other vital information to the U.S. government, aims to protect Americans' privacy and advance equity and civil rights, and sets out to enhance consumer, worker, and student protections. Additionally, it seeks to promote innovation, competition, and international collaboration on AI, alongside ensuring responsible government use of AI technologies.

Following the executive order, in November 2023, the Office of Management and Budget (OMB) created [a draft policy](#) to guide federal agencies in the use of AI and is currently soliciting public comment on the draft guidance. The proposed guidance builds on the [Blueprint for an AI Bill of Rights](#) and the [AI Risk Management Framework](#) by mandating a set of minimum evaluation, monitoring, and risk mitigation practices derived from these frameworks and tailoring them to the context of the federal government.

In particular, the guidance provides direction to agencies across three pillars:

1- Strengthening AI Governance

To improve coordination, oversight, and leadership for AI, the draft guidance would direct federal departments and agencies to:

- Designate Chief AI Officers, who would have the responsibility to advise agency leadership on AI, coordinate and track the agency's AI activities, advance the use of AI in the agency's mission, and oversee the management of AI risks.
- Establish internal mechanisms for coordinating the efforts of the many existing officials responsible for issues related to AI. As part of this, large agencies would be required to establish AI Governance Boards, chaired by the Deputy Secretary or equivalent and vice-chaired by the Chief AI Officer.

- Expand reporting on the ways agencies use AI, including providing additional detail on AI systems' risks and how the agency is managing those risks.
- Publish plans for the agency's compliance with the guidance.

2- Advancing Responsible AI Innovation

To expand and improve the responsible application of AI to the agency's mission, the draft guidance would direct federal agencies to:

- Develop an agency AI strategy, covering areas for future investment as well as plans to improve the agency's enterprise AI infrastructure, its AI workforce, its capacity to successfully develop and use AI, and its ability to govern AI and manage its risks.
- Remove unnecessary barriers to the responsible use of AI, including those related to insufficient information technology infrastructure, inadequate data and sharing of data, gaps in the agency's AI workforce and workforce practices, and cybersecurity approval processes that are poorly suited to AI systems.
- Explore the use of generative AI in the agency, with adequate safeguards and oversight mechanisms.

3- Managing Risks from the Use of AI

To ensure that agencies establish safeguards for safety- and rights-impacting uses of AI and provide transparency to the public, the draft guidance would:

- Mandate the implementation of specific safeguards for uses of AI that impact the rights and safety of the public. These safeguards include conducting AI impact assessments and independent evaluations; testing the AI in a real-world context; identifying and mitigating factors contributing to algorithmic discrimination and disparate impacts; monitoring deployed AI; sufficiently training AI operators; ensuring that AI advances equity, dignity, and fairness; consulting with affected groups and incorporating their feedback; notifying and consulting with the public about the use of AI and their plans to achieve consistency with the proposed policy; notifying individuals potentially harmed by a use of AI and offering avenues for remedy; and more.
- Define uses of AI that are presumed to impact rights and safety, including many uses involved in health, education, employment, housing, federal benefits, law enforcement, immigration, child welfare, transportation, critical infrastructure, and safety and environmental controls.

- Provide recommendations for managing risk in federal procurement of AI. After finalization of the proposed guidance, OMB will also develop a means to ensure that federal contracts align with its recommendations, as required by the Advancing American AI Act and President Biden’s AI Executive Order of October 30, 2023.

EU’S ACCOUNTABILITY PRINCIPLES FOR ARTIFICIAL INTELLIGENCE (AP4AI) IN THE INTERNAL SECURITY DOMAIN

[The Accountability Principles for Artificial Intelligence \(AP4AI\)](#) initiative is a collaboration between the Centre of Excellence in Terrorism, Resilience, Intelligence and Organised Crime Research (CENTRIC) and the Europol Innovation Lab, Eurojust, the EU Agency for Asylum (EUAA), and the EU Agency for Law Enforcement Training (CEPOL). This initiative is part of the broader efforts under the EU Innovation Hub for Internal Security and aims to develop a hands-on toolkit that fosters accountability in AI applications within the realm of internal security. Since its inception in 2021, the project has engaged with experts from 28 countries, including law enforcement personnel, legal professionals, data protection and human rights specialists, and technical and industry leaders, to formulate principles of AI accountability.

The project has also consulted over 5500 individuals from 30 countries to gauge public opinion on AI accountability. Findings indicate broad public support for AI's role in enhancing internal security, especially in protecting children and vulnerable communities and in identifying criminal activities, despite existing reservations about its deployment by law enforcement.

The forthcoming stage involves transforming the AI Accountability framework into a toolkit for practical application across various AI uses within internal security, ensuring its alignment with European values and fundamental rights. This toolkit, which will be made available for free, aims to guide the responsible and transparent use of AI by police and security agencies, promoting a balance between technological advancement and accountability to both authority figures and the public.

As a result of consultations with experts and stakeholders, the initiative created the following 12 accountability principles for the use of AI in internal security:

- 1. Legality:** Legality means that all aspects of the use of AI should be lawful and governed by formal, promulgated rules. It extends to all those involved in building, developing and operating AI systems for use in a criminal justice context. Where any

gaps in the law exist, the protection and promotion of fundamental rights and freedoms should prevail.

2. **Enforceability and Redress:** Enforceability and redress requires mechanisms to be established that facilitate independent and effective oversight in respect of the use of AI in the internal security community, as well as mechanisms to respond appropriately to instances of non-compliance with applicable obligations by those deploying AI in a criminal justice context.
3. **Universality:** Universality provides that *all* relevant aspects of AI deployments within the internal security community are covered through the accountability process. This includes all processes, including design, development and supply, domains, aspects of police mission, AI systems, stages in the AI lifecycle or usage purposes.
4. **Compellability:** Compellability refers to the need for competent authorities and oversight bodies to compel those deploying or utilising AI in the internal security community to provide access to necessary information, systems or individuals by creating formal obligations in this regard.
5. **Pluralism:** Pluralism ensures that oversight involves all relevant stakeholders engaged in and affected by a specific AI deployment. Pluralism avoids homogeneity and thus a tendency or perception for the regulators to take a one-sided approach.
6. **Explainability:** Explainability requires those using AI to ensure that information about this use is provided in a meaningful way that is accessible and easily understood by the relevant participants/audiences.
7. **Transparency:** Transparency involves making available clear, accurate and meaningful information about AI processes and specific deployment pertinent for assessing and enforcing accountability. This represents full and frank disclosure in the interests of promoting public trust and confidence by enabling those directly and indirectly affected, as well as the wider public, to make informed judgments and accurate risk assessments.
8. **Constructiveness:** Constructiveness embraces the idea of participating in a constructive dialogue with relevant stakeholders involved in the use of AI and other interested parties, by engaging with and responding positively to various inputs. This may include considering different perspectives, discussing challenges and recognising that certain types of disagreements can lead to beneficial solutions for those involved.
9. **Independence:** Independence refers to the status of competent authorities performing oversight functions in respect of achieving accountability. This applies in a personal, political, financial and functional way, with no conflict of interest in any sense.

- 10. Conduct:** Conduct governs how individuals and organisations will conduct themselves in undertaking their respective tasks and relates to sector-specific principles, professional standards and expected behaviours relating to conduct within a role, which incorporate integrity and ethical considerations.
- 11. Commitment to Robust evidence:** Evidence in this sense refers to documented records or other proof of compliance measures in respect of legal and other formal obligations pertaining to the use of AI in an internal security context. This principle demonstrates as well as facilitates accountability by way of requiring detailed, accurate and up to date record-keeping in respect of all aspects of AI use.
- 12. Learning Organisation:** Learning Organisation promotes the willingness and ability of organisations and people to improve AI through the application of (new) knowledge and insights. It applies to people and organisations involved in the design, use and oversight of AI in the internal security domain and includes the modification and improvement of systems, structures, practices, processes, knowledge and resources, as well as the development of professional doctrine and agreed standards.

The AP4AI framework not only provides these accountability principles but also converts these principles into practical steps and guidelines, complete with legal and practical considerations for effective enactment. Central to this framework is the AI Accountability Agreement (AAA), a non-legal yet commitment-oriented document that delineates accountability measures for AI applications within the internal security sector. This agreement acts as a social pact, bolstered by legal responsibilities, involving internal security organizations and stakeholders like citizens, oversight entities, AI service users, and suppliers. Essentially, the AAA functions as both a guide and a blueprint, facilitating the real-world application of these principles in internal security operations and their broader network, including oversight and governmental bodies.

According to the framework, before initiating any AI-related project, an AAA must be established and approved, covering all phases of the AI lifecycle—from initial planning and research to development, deployment, and eventual decommissioning. This ensures each stage adheres to the 12 principles, balancing the need for strict compliance with legal and ethical standards against the operational flexibility required for AI's application in security tasks.

EU'S HARMONIZED RULES ON ARTIFICIAL INTELLIGENCE

On April 21, 2021, the European Commission published its proposal for a regulation on harmonized rules for artificial intelligence (EU AI Regulation).¹¹⁰

The European approach is a cumulation of years of consultation and research and complements the automated decision-making provisions already found in the EU data protection law, the General Data Protection Regulation (GDPR).¹¹¹

Much like the GDPR, the EU AI Regulation will undoubtedly become a benchmark for liberal democracies worldwide. The proposal regulates AI systems¹¹² through a classification model that rates them as “prohibited,” “high,” and “lower” risk.

Prohibited systems are those considered to be a clear threat to the safety, livelihoods and rights of people. These include systems which manipulate behaviour to circumvent users’ free will; social scoring systems by governments; and real-time biometric identification systems (except in extremely narrow and authorized circumstances).¹¹³

High-risk systems include those that use critical infrastructure; provide educational or vocational training, employment services and essential services; or conduct law enforcement, migration processes, and the administration of justice and democracy.¹¹⁴

These high-risk AI systems are subject to strict obligations before they can be put on the market, including:

1. Adequate risk management systems to continually evaluate the compliance;¹¹⁵
2. Requirements for high quality data and data governance for training, validation and testing data;¹¹⁶
3. Technical documentation and record-keeping requirements to ensure all necessary information is present to assess compliance, including the algorithm(s) used;¹¹⁷
4. Record-keeping and logs to allow for traceability of results;¹¹⁸
5. Transparent information as to allow users to interpret the system’s output;¹¹⁹
6. Human oversight sufficient to allow natural persons to effectively oversee the system;¹²⁰ and
7. Robustness, security and accuracy, including appropriate measures to protect against cybersecurity threats.¹²¹

The preamble to the EU AI Regulation clearly draws upon the potential harms that can arise from the unregulated use of AI in a free and democratic society, many of which are identified in this report, such as the risks of discriminatory outcomes due to bias,¹²² challenges of opacity in administrative justice, and government social credit scores.¹²⁴ Beyond the world of data protection, the EU has recognized through this regulation the need for standalone legislation to address the society-wide impacts of ADS.

UK POLICE COVENANT ON AI USE

In November 2023, British police chiefs signed a [covenant for using AI in policing](#) and agreed on a set of principles for the lawful and transparent use of AI systems across the country.

Principles of AI in Policing

1. Lawful: All use of AI will comply with applicable laws, standards, and regulations. This includes all users of AI and related data processing ensuring the use is recorded centrally in the National ROPA.

2. Transparent: All use of AI will be subject to “Maximum Transparency by Default (MTbD)

- Forces should ensure the public is aware of AI uses. This will typically include publishing an overview of the algorithms used and the known limitations of the training data used. The datasets will be present on the force IAR with allocated Information asset owners.
- Where operational or security requirements restrict the ability to share, the AI will undergo scrutiny by appropriate independent assessors (e.g., organised by the Chief Scientific Adviser).
- All AI projects must be able to allow a third-party to investigate the algorithmic workings, use scenarios, and underlying data from an ‘adversarial perspective.’ This might require the supplier to provide ‘expert’ witness/evidence of the tools’ operation. All third parties will have appropriate data protection and information security policies in place.

3. Explainable: The ability for any AI to provide an ‘explanation’ of its output will be a determining factor in its implementation.

- The level of explanation expected will be determined by (1) the function it performs (e.g., is it informing a high-impact decision about an individual); (2) the outputs

required of it (i.e., who needs to understand what regarding the output and how was this reached).

4. Responsible: All AI that affects the public will have responsible usage policies (i.e., intentions are defined before deployment so that outcomes and impact can be tracked) and procedures to ensure that users do not accept AI outputs uncritically.

- The ability of AI to make decisions without a human being part of that decision will be determined by the function that the AI performs.
- All AI that effects the public must have a human as the ultimate decision-maker.
- All AI will have a human or automatic means of being stopped if it displays unintended or undesired outputs.
- Those responsible for AI-enabled systems must proactively mitigate the risk of unintended biases or harms, during initial rollout and as they learn, change, or are redeployed.

5. Accountable: All AI will have a clearly identified individual accountable for its operation and outputs.

- All Accountable persons and end-users will be suitably trained in the use of the relevant AI.
- The use of AI in policing will be subject to proper governance and oversight at the relevant organisational level.
- AI enabled data sets and technology systems will be governed and assured under the same frameworks as wider data processing responsibilities, linking what is used and how it is used to the appropriate IAR and ROPA.

6. Robust: All data used to train, or that is analysed by, an AI will be robust and reliable enough for its intended purpose. This requires assessing, tracking and reporting on the quality of data, by way of recognising that the quality of data dictates the quality of the analysis.

- All AI in policing will be used only for the purpose it was designed, trained and authorised for.
- With regards to data usage, all data used in Police AI will be subject to a Framework outlined by a force governance board to guard against issues such as bias, unintended proxies, non-representativeness, unfairness, and untimeliness.
- The Government Office for Artificial Intelligence's Guidelines for AI procurement must inform contract implementation and management.

The use of AI in policing must also comply with established codes of practice, most notably the College of Policing’s Code of Ethics, which describes the standards of accountability, fairness, honesty, integrity, leadership, objectivity, openness, respect and selflessness that is expected of all in policing. All AI in Policing will also be subject to standard organizational technology, architectural, security and usage principles.

INTERPOL’S RESPONSIBLE AI INNOVATION IN LAW ENFORCEMENT TOOLKIT

In June 2023, INTERPOL and the United Nations Interregional Crime and Justice Research Institute (UNICRI) announced the *Toolkit for Responsible AI Innovation in Law Enforcement*, which is a practical guide for law enforcement agencies on developing and deploying AI responsibly while respecting human rights and ethics principles.

The AI Toolkit includes a comprehensive user guide that guides law enforcement executives and officers to navigate responsible AI innovation. It provides the technical foundations of AI, guiding principles for responsible use by the police, and organizational assessments on readiness and risks.

The toolkit provides the following principles for the responsible use of AI:

1. Lawfulness: Like any other activity that law enforcement agencies carry out as part of their mission to prevent, detect, and investigate crime, their engagement with AI systems needs to be lawful. This means that agencies must follow the applicable laws and regulations throughout the design, development and use of AI systems.

2. Minimization of Harm: Minimizing harm is a fundamental goal of policing. The essence of law enforcement is to protect people and society against illegal acts, including by preventing and combatting crime. The same principle is crucial in the context of responsible AI innovation. In this context, minimization of harm means that law enforcement agencies prevent, eliminate or mitigate the risk of harm to individuals and communities that can arise in the context of AI development, procurement and use.

The following principles are instrumental to minimization of harm:

- Robustness and Safety
- Accuracy
- Human and environmental well-being
- Efficiency

3. Human Autonomy: Respecting human autonomy means that law enforcement agencies engage with AI in a way that safeguards humans' capacity and right to self-governance, whether the law enforcement personnel using the tool, victims of crime, suspects, criminals, or the public in general. Human autonomy requires that any decisions that impact humans are ultimately taken by humans, especially in a high-stakes context such as law enforcement. Ensuring human control and oversight of an AI system is therefore essential to upholding human autonomy. However, safeguarding human autonomy goes further, entailing protecting the independence and dignity of every individual or group that interacts with or is affected by the use of an AI system. This principle is rooted in the idea that every human has an inviolable value simply by virtue of belonging to a species capable of rationality. It is the basis of globally recognized and valued concepts such as human dignity and human rights.

The following principles are instrumental to human autonomy:

- Human control and oversight
- Human agency
- Privacy
- Transparency and Explainability

4. Fairness: Fairness is a crucial principle for both AI ethics and criminal justice, and requires an equitable distribution of burdens and benefits, and resources and opportunities between individuals as well as across society.

In the context of responsible AI innovation, fairness means that law enforcement agencies should ensure, throughout their engagement with AI systems, a just and non-discriminatory treatment of individuals and groups and a contribution to a more equitable society. Stakeholder involvement is particularly relevant to achieving this kind of fairness. This substantive aspect of fairness is supplemented by a procedural aspect, which requires that agencies safeguard people's ability to contest decisions supported by AI systems and to be compensated if such decisions are harmful to them.

The following principles are instrumental to fairness:

- Equality and non-discrimination
- Protection of vulnerable groups
- Diversity and Accessibility
- Contestability and Redress

5. Good Governance: Good governance consists of establishing policies, processes, and structures within an organization that enable it to uphold human rights, adequately

manage collective resources, and respond to the needs of the people that the organization aims to serve. In the context of AI innovation in law enforcement, good governance means that agencies should aim to set up an overarching structure for audits and accountability and to foster a culture of responsible AI innovation.

Good governance, human rights and the rule of law are all mutually reinforcing: the principles of human rights and the rule of law serve as a guide for good governance, and good governance is essential to upholding human rights and the rule of law. The principle of good governance runs through the responsible AI innovation framework as it is essential to achieving the core principles of lawfulness, minimization of harm, human autonomy and fairness, and the respective instrumental principles.

The following principles are instrumental to fairness:

- Traceability and Auditability
- Accountability

Putting the Principles into Practice

The Principles for Responsible AI Innovation are relevant throughout the AI life cycle. They aim to provide law enforcement agencies with an ethical and human rights-compliant way to navigate the many complex and crucial decisions that need to be taken, from the conceptualization to use and monitoring – and, in some cases, the decommissioning – of an AI system. To put these principles into practice, it is helpful for agencies to follow a process of understanding and applying the principles, identifying and engaging with the relevant stakeholders, checking the results, and restarting if necessary. There is no set order, as the most appropriate way of performing each of these steps will vary depending on the circumstances. As illustrated in the figure below, this process should be followed throughout the AI life cycle and repeated cyclically

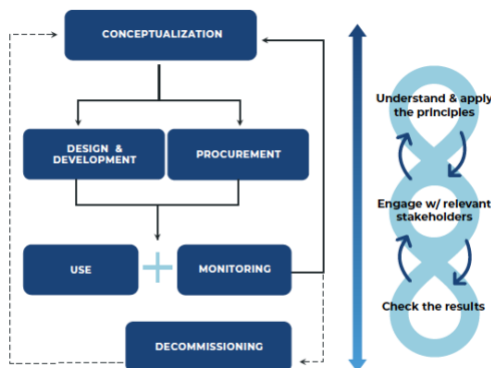


Figure 2 - Putting the principles into practice

REFERENCES

- Canadian Civil Liberties Association. (2023, September). Bill C-27 Submission to INDU-CCLA. <https://ccla.org/wp-content/uploads/2023/09/Bill-C-27-Submission-to-INDU-CCLA.pdf>
- CBC News (Feb, 2023). Thunder Bay police survey suggests public supports AI recognition software, but want to see oversight policy. Retrieved Feb 15, 2024 from <https://www.cbc.ca/news/canada/thunder-bay/briefcam-update-thunder-bay-1.6756761>
- Chiancone, C. (2023). The Role of Artificial Intelligence in Law Enforcement. Retrieved February 10, 2024, from <https://www.linkedin.com/pulse/role-artificial-intelligence-law-enforcement-chris-chiancone/>
- European Commission. (April, 2021). Proposal for regulation laying down harmonised rules on artificial intelligence. Retrieved February 11, 2024, from <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence>
- Europol Innovation Lab. (2022). Accountability Principles for Artificial Intelligence (AP4AI) in the Internet Security Domain. Retrieved February 11, 2024, from https://www.europol.europa.eu/cms/sites/default/files/documents/Accountability_Principles_for_Artificial_Intelligence_AP4AI_in_the_Internet_Security_Domain.pdf
- Global News. (Mar, 2020). OPP Clearview AI technology. Retrieved February 11, 2024, from <https://globalnews.ca/news/6616892/opp-clearview-ai-technology/>
- Government of Canada. (n.d.). Responsible use of AI. Retrieved February 15, 2024, from <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html#toc1>
- Government of Ontario. (2022, January 7). Principles for Ethical Use of AI. <https://www.ontario.ca/page/principles-ethical-use-ai-beta>
- Government of Ontario. (2023, September 14). Ontario's Trustworthy Artificial Intelligence (AI) Framework. <https://www.ontario.ca/page/ontarios-trustworthy-artificial-intelligence-ai-framework>
- Government of Ontario. (n.d.). Education Act, R.S.O. 1990, c. E.2. Retrieved from <https://www.ontario.ca/laws/statute/90m56>
- Government of Ontario. (n.d.). Ontario's Trustworthy Artificial Intelligence Framework. <https://www.ontario.ca/page/ontarios-trustworthy-artificial-intelligence-ai-framework>

- Innovation, Science and Economic Development Canada. (2023). Artificial Intelligence and Data Act. <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act>
- Innovation, Science and Economic Development Canada. (2023). Voluntary Code of Conduct for the Responsible Development and Use of AI Systems. <https://ised-isde.canada.ca/site/ised/en/voluntary-code-conduct-responsible-development-and-management-advanced-generative-ai-systems>
- INTERPOL. (2023). Artificial Intelligence Toolkit. Retrieved Feb 26, 2023, from <https://www.interpol.int/en/How-we-work/Innovation/Artificial-Intelligence-Toolkit>
- National Institute of Justice. (2018). Solve Crime. Build Trust? Investigating Homicides and Shootings in Los Angeles. Retrieved Feb 20, 2023, from <https://nij.ojp.gov/funding/awards/2018-75-cx-0003>
- National Institute of Standards and Technology. (n.d.). AI Risk Management Framework. Retrieved from <https://www.nist.gov/itl/ai-risk-management-framework>
- NPR. (2021, January 8). No charges for Colorado officers who held black children at gunpoint. <https://www.npr.org/2021/01/08/955165485/no-charges-for-colorado-officers-who-held-black-children-at-gunpoint>
- Office of the Information & Privacy Commissioner for British Columbia. (2021). Getting Ahead Of The Curve: Meeting The Challenges To Privacy And Fairness Arising From The Use Of Artificial Intelligence In The Public Sector. Retrieved Feb 20, 2023, from <https://www.oipc.bc.ca/special-reports/3546>
- Office of the Privacy Commissioner of Canada. (2021). Police use of Facial Recognition Technology in Canada and the way forward. Retrieved Feb 20, 2023, from https://www.priv.gc.ca/en/opc-actions-and-decisions/ar_index/202021/sr RCMP/
- Office of the Privacy Commissioner of Canada. (Dec, 2023). Guidance on Artificial Intelligence. Retrieved Feb 24, 2023, from https://www.priv.gc.ca/en/privacy-topics/technology/artificial-intelligence/gd_principles_ai/
- Police1. (Dec, 2023). Wrap Technologies Inc unveils cutting-edge AI functionality for Wrap Intrensic body-worn camera solution. Retrieved Feb 12, 2023, from <https://www.police1.com/police-products/body-cameras/wrap-technologies-inc-unveils-cutting-edge-ai-functionality-for-wrap-intrensic-body-worn-camera-solution>
- Redden, J., Aagaard, B., Taniguchi, T., & Criminal Justice Testing and Evaluation Consortium. (2020). Artificial Intelligence in Law Enforcement. U.S. Department of Justice, National Institute of Justice, Office of Justice Programs. <http://cjtec.org/>

- Reuters. (2022, April 8). China uses AI software to improve its surveillance capabilities. Retrieved Feb 15, 2023, from <https://www.reuters.com/world/china/china-uses-ai-software-improve-its-surveillance-capabilities-2022-04-08/>
- Rigano, C. (2018). Using Artificial Intelligence to Address Criminal Justice Needs. Retrieved February 12, 2024 from <https://nij.ojp.gov/topics/articles/using-artificial-intelligence-address-criminal-justice-needs>
- Royal Canadian Mounted Police. (n.d.). Connected RCMP. Retrieved February 28, 2024, from <https://rcmp-grc.gc.ca/en/connected-rcmp#a2>
- Science & Technology in Policing (2023). National Police Chiefs' Council Covenant for Using Artificial Intelligence (AI) in Policing. Retrieved Feb 20, 2023 from https://science.police.uk/site/assets/files/4682/ai_principles_1_1_1.pdf
- The Guardian. (2023, September 8). New York police tracking Voyager Labs meta contract. <https://www.theguardian.com/us-news/2023/sep/08/new-york-police-tracking-voyager-labs-meta-contract>
- The Star. (Feb, 2020). Peel and Halton police reveal they too used controversial facial recognition tool. Retrieved February 28, 2024, from https://www.thestar.com/news/gta/peel-and-halton-police-reveal-they-too-used-controversial-facial-recognition-tool/article_ce8ce53d-e4c2-5d6d-83da-c4c46602277b.html
- The Washington Post. (2021, April 13). Facial recognition false arrest lawsuit. <https://www.washingtonpost.com/technology/2021/04/13/facial-recognition-false-arrest-lawsuit/>
- The White House, Office of Management and Budget. (2023, November 1). OMB releases implementation guidance following President Biden's executive order on artificial intelligence. Retrieved from <https://www.whitehouse.gov/omb/briefing-room/2023/11/01/omb-releases-implementation-guidance-following-president-bidens-executive-order-on-artificial-intelligence/>
- The White House, Office of Science and Technology Policy. (n.d.). AI Bill of Rights. Retrieved from <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>
- The White House. (2023, October 30). Fact sheet: President Biden issues an executive order on safe, secure, and trustworthy artificial intelligence. Retrieved from <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/>
- Thunder Bay Police Service. (2022). Strategic Plan 2021-2023, Operational Progress Report Retrieved February 15, 2024 from

<https://www.thunderbaypolice.ca/sites/default/files/2022-06/TBPSB%20StratPlan%20Update%20FINAL.pdf>

Toronto Police Service Board. (2022). [Title of the document]. Retrieved February 28, 2024, from <https://tpsb.ca/jdownloads-categories?task=download.send&id=720&catid=5&m=0>

Waterloo Regional Police Service. (n.d.). Privacy and technology. Retrieved February 28, 2024, from <https://www.wrps.on.ca/en/about-us/privacy-and-technology.aspx>

Waterloo Regional Police Service. (Mar, 2020). WRPS internal review reveals use of facial recognition technology. Retrieved February 28, 2024, from <https://www.wrps.on.ca/en/news/wrps-internal-review-reveals-use-of-facial-recognition-technology.aspx>

Wired Middle East. (2023). AI-powered robot police take to the streets in Dubai. Retrieved February 21, 2024, from <https://wired.me/technology/dubai-police-ai-robocop/>

Yorkregion.com. (n.d.). York police officers used facial recognition technology without permission: YRP spokesperson. Retrieved February 21, 2024, from https://www.yorkregion.com/news/york-police-officers-used-facial-recognition-technology-without-permission-yrp-spokesperson/article_8b18d539-3fe1-5040-9d26-9b86aa7c4384.html