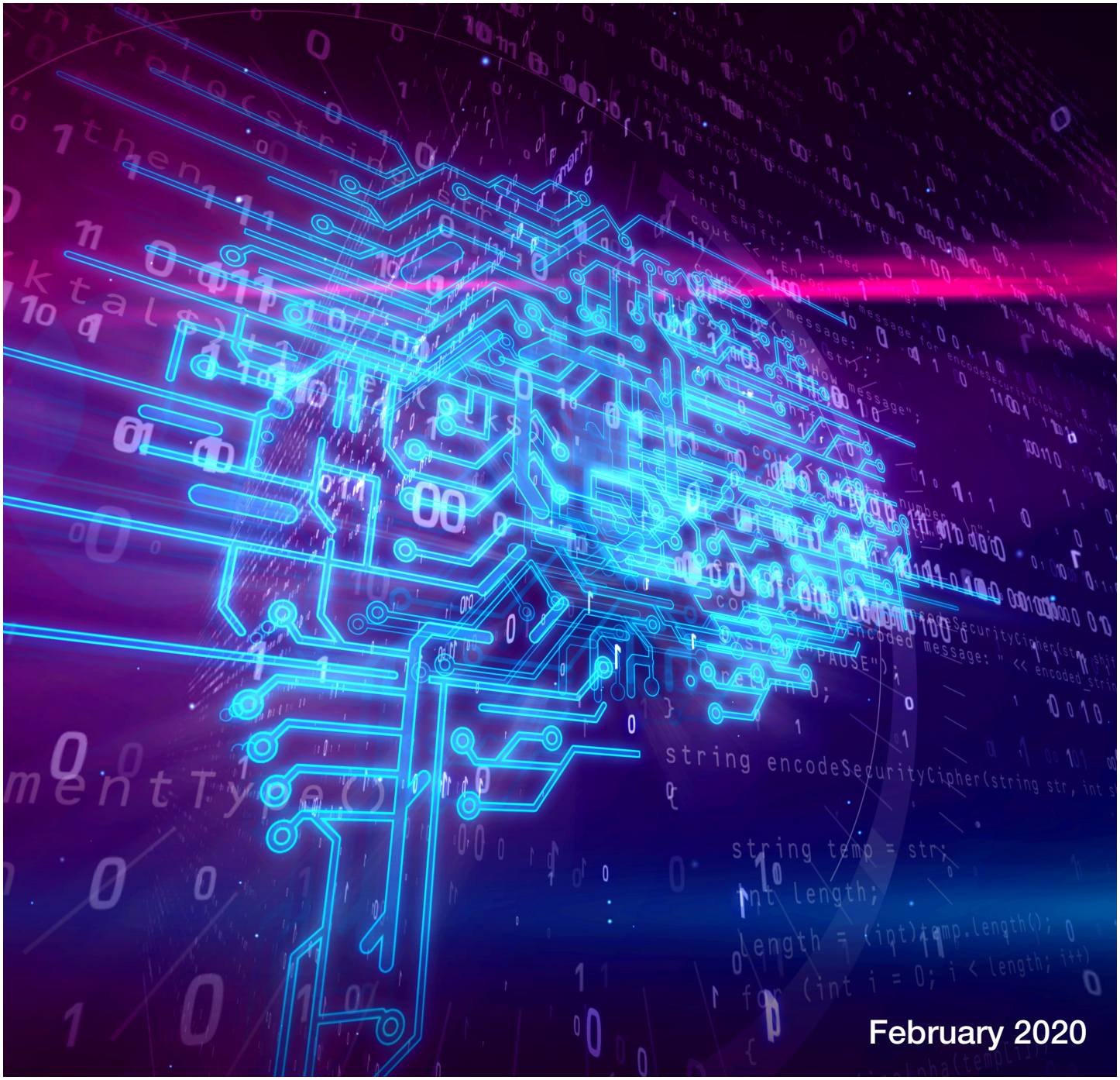
International Public Sector Fraud Forum

The use of Artificial Intelligence to Combat Public Sector Fraud

Professional Guidance



SERIOUS FRAUD OFFICE
Te Tari Hara Taware

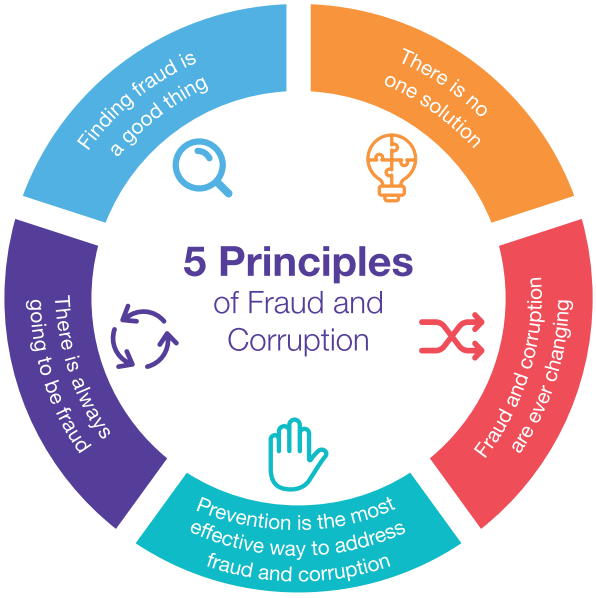
Produced in collaboration with New Zealand's Serious Fraud Office.

Contents

The International Public Sector Fraud Forum

The International Public Sector Fraud Forum (IPSFF) currently consists of representatives from organisations in the governments of Australia, Canada, New Zealand, the United Kingdom and the United States. The collective aim of the Forum is to come together to share best and leading practice in fraud management and control across public borders.

The Forum has established 5 principles for public sector fraud.



1. There is always going to be fraud

It is a fact that some individuals will look to make gains where there is opportunity, and organisations need robust processes in place to prevent, detect and respond to fraud and corruption.

2. Finding fraud is a good thing

If you don't find fraud you can't fight it. This requires a change in perspective so the identification of fraud is viewed as a positive and proactive achievement.

3. There is no one solution

Addressing fraud needs a holistic response incorporating detection, prevention and redress, underpinned by a strong understanding of risk. It also requires cooperation between organisations under a spirit of collaboration.

4. Fraud and corruption are ever changing

Fraud, and counter fraud practices, evolve very quickly and organisations must be agile and change their approach to deal with these evolutions.

5. Prevention is the most effective way to address fraud and corruption

Preventing fraud through effective counter fraud practices reduces the loss and reputational damage. It also requires less resources than an approach focused on detection and recovery.

Foreword

Paul O'Neil
General Counsel, New Zealand

SFO
SERIOUS FRAUD OFFICE
Te Tari Hara Tāware

The technology that exists to both counter and commit fraud does not stand still and the public sector has a responsibility to keep pace with developments in both regards.

In particular, the members of the International Public Sector Fraud Forum have recognised that the rise of Artificial Intelligence presents a huge opportunity for the public sector to detect and prevent fraud. At the same time, it is also accepted that its use presents significant strategic, operational and reputational risks if not employed appropriately. This discussion document is a response to that challenge.

There is already a huge amount of technical expertise in AI across the jurisdictions that make up the Forum and beyond. This is reflected in a number of excellent recent publications that reflect on the use of AI by both government organisations and the private sector. Rather than duplicating that work, Forum members have agreed that what we wanted to produce was a document that focused specifically on the issues raised by the use of AI in fighting fraud.

It was further agreed that the value of this document does not lie in its ability to serve as a technical guide or manual for the operational use of AI. Documents with this focus are of course invaluable, but the pace of technical development in this area, and the jurisdictional differences that exist, mean that their lifespan and scope is limited. Instead, as a group, we have agreed on a set of broad principles of general and hopefully enduring application that will assist leaders of public sector organisations to ensure they are thinking about the right issues (from both a risk and opportunity perspective) when it comes to using AI to combat fraud. In that sense, it should support the development of a framework that aspires to the highest standards in terms of legality, ethics, transparency, security, privacy and accountability, so that public trust and confidence in the use of this tool by governments is preserved.

These principles are supported by real life examples from the Forum members about how AI is already impacting on the work we are all doing to combat fraud. As well illustrating the challenges we all face in a practical and recognisable way, it again demonstrates the collective strength and depth of the counter fraud experience that the Forum has brought together.

We hope you will find this document to be an important addition to the growing suite of products the Forum has produced to assist in achieving our common goal of fighting fraud and corruption in the public sector.

Introduction

The International Public Sector Fraud Forum (IPSFF) is comprised of representatives from the governments of Australia, Canada, New Zealand, the United Kingdom and the United States. The purpose of the IPSFF is to share amongst its members best practice in combatting public sector fraud in order to reduce the harm caused by fraud and corruption.

As part of achieving this purpose, the IPSFF has set out to develop products that can be used by public sector agencies in the fight against fraud. This paper is intended to consider the appropriate elements of a framework for the use of Artificial Intelligence (AI) technology by public sector agencies in dealing with fraud and corruption.

AI has been described as the next great technological frontier and according to the World Economic Forum is the crucial component of the Fourth Industrial Revolution. AI is already enhancing the operation of commerce, government and numerous other aspects of our everyday lives including robotic process automations, virtual agents for improved customer service, as well as machine vision systems such as face recognition, other biometrics and driver assistance moving toward autonomous systems.

Of course, the use of AI within the public sector is not a new phenomenon and its application to fraud prevention and detection (particularly through predictive algorithms) has steadily grown in recent times and has become an essential tool. Examples include:

Canada

In Canada, they operate an Employment Insurance (EI) Sickness Programme. One of the integrity projects that utilize AI supports ongoing investigation into the abuses of the EI benefit program by focusing on identifying fictitious doctor's notes. Once such notes are discovered, they are associated with EI benefits to select cases for investigation. The project uses transcriptions and images and employs a variety of AI enabled technologies to extract relevant information from them. For example, Natural Language Processing (NLP) is applied to the transcripts to extract details about doctors. Optical character recognition (OCR) is used to extract that information from medical images while network analysis helps to identify claimants related to the known or newly identified cases of fraud.

Australia

One predictive policing tool has already been modelled to predict crime hotspots in Brisbane. Using 10 years of accumulated crime data, the system used 70% of the data to predict crime, with the researchers seeing if its predictions correlated with the remaining 30%. The results proved more accurate than existing models, with an improvement of 16% accuracy for assaults, 6% more accuracy for predicting unlawful entry, 4% better accuracy for predicting drug offences and theft, and 2% better for fraud. The system can predict long term crime trends, but not short-term ones. The Brisbane study used information from location-based app foursquare, and incorporated information from both Brisbane and New York.

Another driver for the use of AI in combating fraud is that the public sector must keep up with offenders. Those looking to engage in fraud are using increasingly sophisticated methods that rely on the systematic analysis of large amounts of data in an effort to identity and exploit weaknesses and vulnerabilities that might exist within the public sector. Letting fraudsters lead the way in the use of AI technology is not an option.

However, the use of AI and its increasing power and complexity, presents both opportunities and concerns. These concerns include: the collection, transmission, processing, storage, and curation of potentially vast amounts of information that can be factored into decisions; the potential lack of transparency around machine classification algorithms and/or decision making processes; and the critical need for the appearance of objectivity that must attach to the results.

This paper is not intended to address specific AI techniques, products or methodologies and is not a technical guide. Instead it is directed at providing leaders of public sector organisations with a resource to assist them in thinking about the right issues when it comes to using AI to combat fraud. It is also intended to give them direction in taking steps (consistent with the relevant conditions in their jurisdiction and organisation) to address those issues. In that sense, AI represents a strategic opportunity for public sector leaders to ensure the advantages of AI are maximised and the associated risks are either minimised or mitigated completely.

This should facilitate a use of AI by public sector agencies that best ensures:

- the highest standards of legality, ethics, transparency and accountability are met;
- evidential/admissibility and data quality requirements are fulfilled;
- public trust and confidence in the use of AI by the public sector is maintained;
- the data collected, transmitted, processed, and stored is secure; and
- personal privacy and civil liberties are maintained.

The paper draws on the experiences of other jurisdictions in terms of AI issues they have faced and responses they have formulated[1]. It then seeks to assimilate these experiences into a discussion of the relevant issues that arise in using AI to combat fraud.

1 The OECD recently published a list of the national public sector AI strategies being employed by 50 countries, which illustrates the global priority this issue is being given and the range of responses that exist: https://oecd-opsi.org/projects/ai/strategies/

Scope

The starting point for any discussion of AI is to define it. This is notoriously difficult to do as AI can range from predictive algorithms and machine learning all the way through to complex robotics.

From a substantive perspective, AI can be defined as the use of digital technology to create systems capable of performing tasks commonly thought to require intelligence. In terms of its relationship to us as humans, it can be regarded as a collection of interrelated technologies used to solve problems autonomously and perform tasks to achieve defined objectives without explicit guidance from a human being. It will involve some element of learning by that system, but that can be supervised or unsupervised.

In any event, while AI is constantly evolving, generally it involves:

- machines using statistics to find patterns in large amounts of data; and
- the ability to perform repetitive tasks with data without the need for constant human guidance.

To further assist in interpreting the concepts discussed in this paper we have included a Glossary of Key Terms at **Annex A.**

This paper is structured around the following five central issues raised by the use of AI[2] in combatting fraud:

- Accuracy
- Human control
- Transparency and explainability
- Fairness
- Privacy and civil liberties

This is not intended to be exhaustive and other areas may also be relevant. In particular, any assessment of the use of AI must take into account and be tailored to the legal, practical and social conditions present in the relevant jurisdiction. However, we believe if a public sector agency were to construct a framework that covered each of these areas, it would represent an appropriate level of consideration of the issues raised by using AI.

We note also that underpinning each of these issues is the concept of ethics. For AI to play a central role in combatting fraud, then its use must be both lawful and effective. However, fulfilling these two criteria is not enough. This is because (1) legal requirements are not always up to speed with either general practice or the expectations of society and (2) the mere fact that something produces a desired outcome does not necessarily justify the means used to achieve it.

2 We note that other more general AI discussion documents offer variations on these issues by selecting other key principles or themes to frame the discussion, such as data governance or security. There is no single correct method and for examples of other approaches note the additional AI publications listed at **Annex B.**

AI systems must also be ethical, ensuring alignment with ethical norms. Accordingly, in considering the above issues as they relate to fraud, this paper will also refer to relevant ethical principles and how they assist in ensuring that the goals of AI in combatting fraud are achieved and that AI is regarded as being trustworthy. This is broadly true of the use of AI in any context, but is absolutely paramount in the context of a public sector fight against fraud. If public sector organisations are to maintain credibility in countering fraud (which is, at its heart, unethical behaviour), then the ends cannot justify the means. Ethical AI should be the ambition and the highest ethical standards must be maintained.

At a practical level, given ethical compliance is often more difficult to assess than matters of accuracy or legality, we have also included as **Annex B** to this paper the Assessment List which is being piloted by the EU's High Level Expert Group on Artificial Intelligence in order to achieve Trustworthy AI. Completion of this list will of course not guarantee that an organisation's AI use is appropriate, but it will provide a useful cross-check that the substantive issues identified in this paper have been considered.

We note also that in analysing the risks and concerns presented by AI, it should not be assumed that the issues discussed do not also exist (to a certain extent) when human decision making is involved. The presence of biases, error, opacity and prejudice is certainly not unique to AI although it may be possible to detect and address these issues more efficiently when dealing with them in a more predictable digital context.

Finally, it is important to recognise that a significant amount of work has already been done by various jurisdictions and organisations in terms of identifying the key principles applicable to the use of AI (not just as it applies to fraud) and also navigating a path through the multitude of issues presented by its use. This paper is a thought piece intended to provoke discussion amongst and provide some direction to public sector organisation leaders in identifying potential pitfalls and solutions with the use of AI in their fight against fraud. It therefore should be read alongside the guidance that has been produced to date and we have identified some of the existing key guidance publications in **Annex C** to this paper.

Accuracy

One of the obvious perceived advantages of AI is that it should ensure more accurate outcomes. The use of AI allows the consideration of a vastly increased number of input variables across large data sets in a systematic way. It also allows variables that are not relevant to be disregarded in a way that humans can find hard to do.

In addition, AI decision making evaluates errors in a way that allows settings to be adjusted. For example, the view could be taken that in the context of detecting potential fraud, false positives are more tolerable than false negatives (or the reverse depending upon the stage or nature of the process) and this can be adjusted for in a way that a human led review cannot be.

Against this, this principal benefit of AI is also its biggest potential weakness. To the extent that reliance on the accuracy of AI tools is found to be misplaced because unacceptable errors have occurred, public trust and confidence in its use will be undermined in a significant way. In addition, in the context of detecting or preventing fraud, the stakes are high. The effects of being implicated in fraudulent conduct (or even simply behaviour lacking in integrity) can be severe and sometimes irretrievable. Society has always expected that the detection of criminal behavior by the state (in this case fraud) is inextricably linked with provable accuracy. The partial automation or augmentation of a decision-making process by an AI tool does not change that fact.

AI tools must be regularly tested and retrained to ensure their settings remain appropriate and that they continue to reflect ever-changing government priorities, policies, legislative settings and societal conditions.

In particular, the data used to train AI systems may itself contain and therefore be introducing inherent bias that affects accuracy and outcomes of AI systems. When data is gathered, it may contain socially constructed biases, inaccuracies, errors and mistakes. In this context, bias can include a predisposition towards or against a particular thing, person or group, such as an ethnic group, social class, political party, religion or other demographic (such as an age group). This issue needs to be addressed prior to training with any given data set. In addition, the integrity of the data must be ensured. Feeding biased data into an AI system may change its behaviour, particularly with self-learning systems. Processes and data sets used must be tested and documented at each step.

By way of example:

Canada

The Government of Canada has instituted a Directive on Automated Decision-Making Consulting which requires appropriate qualified experts to review the Automated Decision System for quality assurance in terms of the accuracy of outcomes. The Directive requires that they have at least one qualified expert from a federal, provincial, territorial or municipal government institution, one member of faculty of a post-secondary institution and at least two qualified experts from the National Research Council of Canada, Statistics Canada, or the Communications Security Establishment. The performance measures that are applied are selected on case by case basis to reflect the nature of the decision being made and the overall goals and expectations set for the particular AI application.[3]

AI programs can also be used to review other AI systems. Several companies have developed tools that may be able to effectively assess algorithms used by AIs and report on how the system is operating and whether it is acting fairly or with bias. IBM has released an open source, cloud-based software that creates an easy-to-use visual representation that shows how the algorithms are generating decisions. In addition, it can assess the algorithm's accuracy, fairness and performance. Microsoft and Google are working on similar tools to assess algorithms for bias. The use of such technologies could improve the ability to efficiently, effectively and objectively review the components of AI to ensure that they adhere to key ethical principles. Of course, these AI enabled technologies would require a significant degree of scrutiny to ensure that they did not have the same flaws that they were purporting to assess.

3 The Canadian Treasury Board Secretariat has also developed a draft Algorithmic Impact Assessment tool intended to help practitioners implement AI in an appropriate and ethical manner.

In addition to the scrutiny applied to the AI process, the following internal enablers will play a key role in controlling bias and accuracy risks:

- **Competencies:** Within many public sector organisations the pool of talent for technology professionals with AI expertise is small. Recruitment and training in key disciplines, including natural language processing, data analytics, computer vision and machine learning, will be vital.
- **Data governance:** A lack of meaningful data sets and benchmarks to validate real-world performance as well as insufficient volume of labeled data for machine learning could slow the adoption of AI within the public sector. Recognising that the power of current AI technology is in the data, the more abundant and clean data available on fraud cases, the better the AI will perform. Data sharing between different platforms also raise debates on privacy, security, trust and accountability.
- We note that some models for the use of AI suggest there should be independent and public oversight of the accuracy of the AI models being used in government. In the context of fraud, some form of general public assurance is appropriate, but (as noted below in respect of transparency) a detailed dissection of actual methodologies used by agencies may be counterproductive and could provide a roadmap for offenders to avoid detection.

In summary:

- Before it is deployed, an AI tool should be tested against independent and well understood data for accuracy.
- Post-deployment, it should again be periodically tested and trained using quality, unbiased data (in certain cases, it may even need to be retrained). As part of this process, the items used to train the AI tool should be representative of the data on which the AI tool will be deployed.
- There should be a process for regular gathering and curating of new training data so that the system does not become out of date or skew into unintended bias.
- Agencies should also not employ a single testing standard for different AI tools used in different circumstances.
- The level of scrutiny applied before AI is deployed should reflect the fact that in the context of fraud detection or prevention, a particularly punitive or intrusive intervention may follow. Agencies should ensure that they have appropriately qualified people to operate the AI tools and also, to the greatest extent possible, ensure that the data being analysed is meaningful.

Human Input

AI should be used to inform human decision making but should not entirely replace human oversight. The extent of oversight will depend on the significance of the decision and on other safeguards in place. Where a decision or selection being made about an individual is significant (its operation impacts benefits, freedom, or access to a service), careful consideration should be given to the level of human input required.

New Zealand

In the New Zealand Court of Appeal, the Court considered an appeal against the imposition of an Extended Supervision Order (ESO) for a prisoner about to be released. The original granting of the ESO had been supported by the results of complex algorithm-based instruments measuring the likelihood of the prisoner reoffending. In considering the relationship between the appellant's personal circumstances and the actuarial results, the Court made the following observation:

"Obviously factors which have arisen post-release must be allowed for in an ESO assessment. For instance, if the appellant had been rendered a tetraplegic as a result of a post-release accident, this would have presumably eliminated the likelihood of him reoffending and would undoubtedly have negated any adverse inferences which might otherwise have been drawn for actuarial assessments."

Of course, while the extent of human input may be an exercise of discretion in some cases, many jurisdictions also have legislation which proscribes the extent to which AI can be used independently of human decision making. For example, the General Data Protection Regulation 2016/679 is a regulation in EU law on data protection and privacy for all individual citizens of the European Union and the European Economic Area which sets out some key principles for the level of human input that is appropriate.

It is also important to ensure that human input is actually meaningful (as opposed to token) and also to acknowledge that it could differ from case to case. It could mean that the agency regularly checks the output of the AI against defined metrics, or that it runs a parallel human process and compares the conclusion with the machine. Meaning can also be added by assessing what skills the relevant humans might bring to an AI review process. For example, using multi-disciplinary teams to design AI systems that include computer scientists, policy experts and legal officers. Such teams might be better placed to ensure that the outcome of the AI is not ill-conceived or mismatched with the policy and legislative intention. These teams might also be able to identify relevant variables that should be taken into consideration.

In any event, automation complacency and automation bias are real phenomena which reflect the natural human tendency to overestimate the value of a machine's outputs and/or to trust an automated system so much that we ignore other sources of information, including our own senses. If human input is not meaningful, then its presence may only serve to reinforce the seemingly apparent, but mythological infallibility of AI.

Of course, if human intervention is too pronounced, then it could undermine the system's accuracy or efficiency and defeat the purpose of the AI tool. There are obvious situations where automated systems are not reliable enough to be left to operate independently, where factors need to be considered that are not readily automated, or in situations where a measure of discretion is desirable or required.

The above matters therefore require a careful balancing of the importance of human oversight against the efficiencies in operational delivery that AI can provide.

There are also legal obstacles presented by an absence of human control. Jurisdictions may have a prescribed delegation system within their respective public sectors whereby powers to make decisions are specifically granted to certain individuals or positions. Such powers could include the ability to withhold or grant a service or benefit, the compulsory acquisition of information, the commencement of an inquiry or investigation, or the application of sanctions.

Such decisions may or may not be able to be delegated, either directly where AI is literally making the decision, or even indirectly where a person is effectively (and possibly illegally) fettering their discretion by unquestioningly following the direction of an AI tool. A person must take responsibility for the decisions made by an automated system and each jurisdiction will need to satisfy itself that the use of AI is in accordance with the legislative framework and administrative law principles, including that each case is considered on its own merits.

In a criminal procedure context, prosecutors will also need to satisfy themselves that the obligations they owe to the relevant Court or tribunal around the decisions they make are being met. A direct or indirect delegation of these obligations is perhaps less likely in a prosecution context, but an overreliance on AI here could still manifest itself in an inability to adequately explain a decision. This issue will be considered in more detail below in relation to transparency.

Overall, it is an open question whether AI systems (at least in the fraud environment) will ever be trusted with full decision-making. It is perhaps more likely that AI continues to operate as a humandecision augmentation. This is to say, the machine will provide "tipping and cueing" functions to support the limited resources of the human decision-maker. As such, AI will inform decision-making, but it won't be the decision maker.

In summary:

- The use of AI should inform human decision making and should not entirely replace human oversight.
- Human oversight must be meaningful, or it will simply reinforce overreliance on automated decision making. However, the oversight should not be so pronounced that it undermines the system's effectiveness or efficiency.
- Careful consideration must be given to the impact of AI on the delegation of decision making in both a public sector and criminal procedure context.
- Consideration should be given to ways in which agencies can develop formal policies regarding the balance between automated and human decision-making. Demonstrating accountability at an organisational level regarding decisions that affect the public directly is key to maintaining public confidence in the work of the public sector.

Transparency

One of the more fundamental issues with the use of AI by public sector agencies is the potential for a lack of transparency, both perceived and actual.

True transparency requires accountability or answerability meaning a responsiveness to requests for information about the process, or a willingness to offer justification for actions taken or contemplated. Meeting this expectation lies at the heart of justifiable public sector use of AI.

At a systemic level this reflects the need to strike the delicate and difficult balance between complete transparency and obfuscating counter fraud activities to prevent fraudsters developing approaches to circumvent controls and protections.

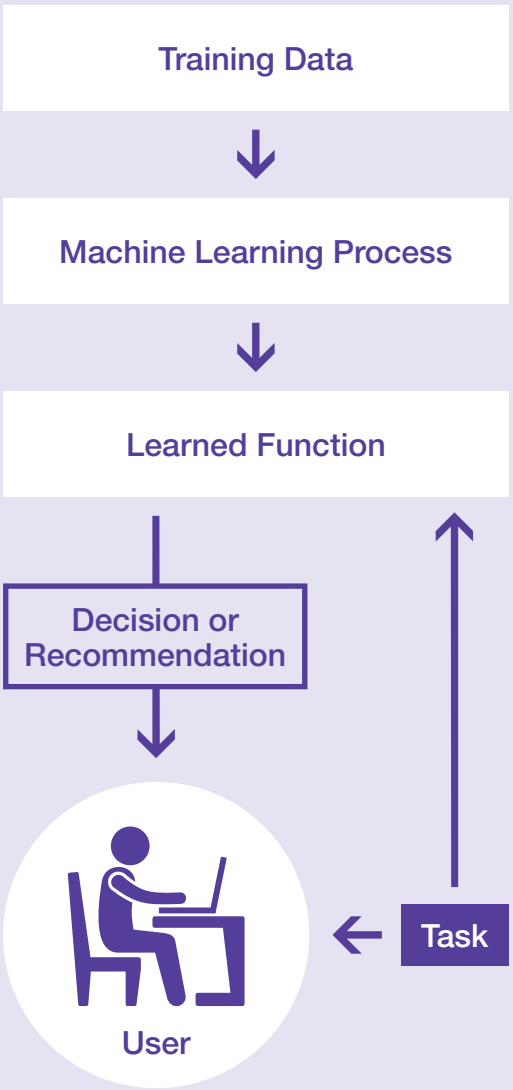
At a technical level, it reflects the fact that AI can be seen as a 'black box' process. Black box is a description applied to some deep learning systems which take an input and provide an output, but the calculations that occur in between are not easy for humans to interpret. Black box AI systems, often based on machine learning over big data, make decisions experimentally or intuitively, without the capability of exposing the reasons why. This is problematic not only for lack of transparency, but also for possible biases inherited by the algorithms from human prejudices and collection artefacts hidden in the training data, which may lead to unfair or wrong decisions.

In one sense, transparency may be associated with legal but also moral responsibility. This captures such familiar notions as blameworthiness and liability for harm. It also connects closely with the ethical use of AI which (as noted above) underpins each of the concepts discussed in this paper. Ethics both inform and are informed by laws and community values. In developing and governing AI technologies, neither over-regulation nor a completely hands-off approach is appropriate or workable.

Transparency also relates explicitly to the auditability of institutions, practices and instruments. Here transparency is about mechanisms: How does this or that tool actually work? How do its component parts fit together to produce outcomes like those it is designed to produce?

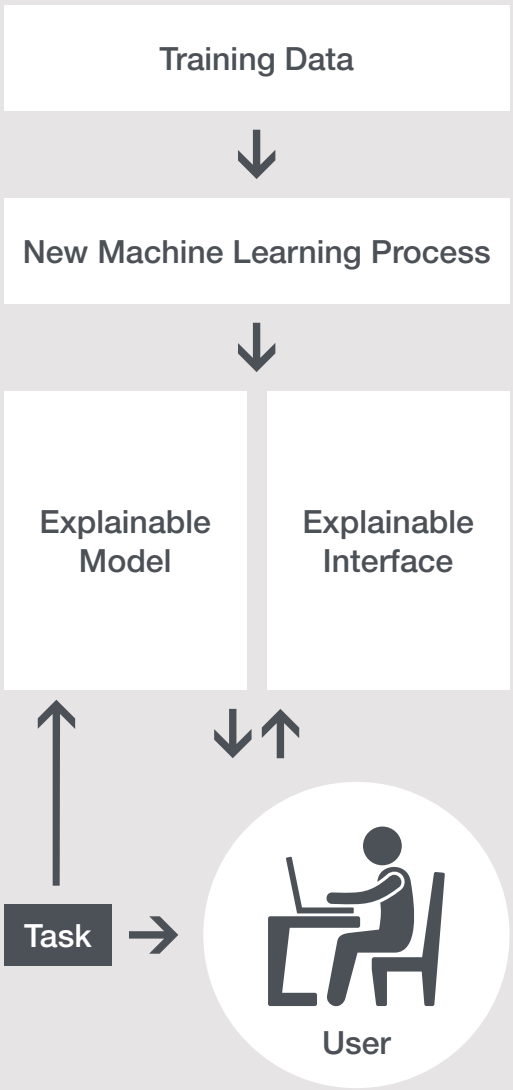
Finally, transparency also denotes accessibility. Meaningful explanations of an algorithm may be possible, but they may not be available. Intellectual property rights might prevent the disclosure of proprietary code, or preclude access to training data, so that even if it were possible to understand how an algorithm operated, a full reckoning may not be possible for economic, legal or political reasons.

Today



- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

Explainable AI



- I understand why
- I understand why not
- I know when you succeed
- I know when you fail
- I know when to trust you
- I know why you erred

To ensure public trust and confidence in the use of AI is established and maintained, it is important that some level of explanation as to how an AI tool operates is publicly available. Accordingly, an 'ability to explain' should be an important part of the selection/design process when an AI tool is acquired or developed. While there may be reasons that a detailed exposition of an AI system may not be possible (including maintenance of law, legal privilege, intellectual property rights or technical complexity), an inability to offer even a basic explanation will significantly impact on public trust in AI. Without that level of trust, the engagement of citizens in the process of government decision making will be lacking, which in turn undermines the legitimacy and effectiveness of the public sector generally.

Each jurisdiction will also have obligations placed on its public sector agencies to respond to requests for official information. The existence of such a right is fundamental and supports and promotes transparency and accountability by governments. It also allows those affected by the actions of public sector agencies to understand and, if they wish, to challenge decisions that are made. If agencies are unable or unwilling to adequately respond to requests for information around how a decision underpinned by AI was made, then legal challenges are almost certain.

Of course, in a fraud context, the stakes are high in that reputations, livelihoods and personal liberty may be at stake. Accordingly, transparency may itself lead to legal challenges through which the basis for action by the state (looking at either substance or process) will be scrutinised. This is not something to be avoided and legal actions which test the legality and trustworthiness of the use of AI in the fight against fraud will only strengthen its legitimacy in the long term.

Also, once Court processes are underway, disclosure or discovery obligations in each jurisdiction will inevitably require the basis for decisions around relevance, privilege status and admissibility of documents to be justified. It will not be acceptable in a prosecutorial context to justify decisions made in these critical areas purely by reference to the operation of an intelligent algorithm.

Against the above considerations, transparency for the use of AI is not straightforward. The reality is that the complex operation of some AI tools will defy easy explanation. There is also the consideration noted above that in the context of fraud offending, a well-intentioned explanation may inadvertently provide a roadmap for how to avoid detection. A possible solution to these issues lies in the fact that providing an explanation of the priorities or strategic direction behind an AI backed decision may actually provide more transparency than a scientific breakdown of the tool itself.

United States

A proprietary AI system was used by the Houston school district to assess the performance of their teaching staff. The system used student test scores over time to assess the teachers' impact. The results were then used to dismiss teachers deemed ineffective by the system. The teacher's union challenged the use of the AI system in court. As the algorithms used to assess the teacher's performance were considered proprietary information by the owners of the software, they could not be scrutinised by humans. This inscrutability was deemed a potential violation of the teachers' civil rights, and the case was settled with the school district withdrawing the use of the system. Judge Stephen Smith stated that the outputs of the AI systems could not be relied upon without further scrutiny, as they may be "erroneously calculated for any number of reasons, ranging from data-entry mistakes to glitches in the computer code itself. Algorithms are human creations, and subject to error like any other human endeavour".

Canada

The Government of Canada Digital Playbook Guide on automated decisions recommends making available to the public all of the source code used for their Automated Decision Systems and requires meaningful explanation to be given to affected individuals, including the variables in the decision, together with the decision itself.

United Kingdom

HM Revenue and Customs (HMRC) uses AI to support a number of activities including: identifying risks on some large-scale transactional services, such as repayment claims for Value Added Tax (VAT) and Income Tax Self Assessment; using analytics to help identify risks that need attention and building case packages that are passed to teams of investigators. AI also works well to assimilate large amounts of data – this is a newer implementation, important for compliance casework where HMRC are using AI alongside other tools like geo-mapping.

From a technical perspective, cloud computing is removing many of the barriers. However, there is a growing conversation around the ethical adoption of AI and what that means. HMRC, set up a working group to build greater awareness around the ethics issues and consider the governance needed. HMRC recognise that being able to explain how AI is used is very important in terms of maintaining the trust of customers. And, as AI technology matures further, it will undoubtedly bring different ways of working, which will bring different cultural and educational challenges.

In summary:

- The ability to explain the operation of an AI tool should be a key consideration in its selection and/or development.
- The legal right of the public to understand and potentially challenge government decisions through requests for information is important and must be preserved.
- Agencies should be prepared to explain their decision-making processes at a level that satisfies criminal procedure requirements.
- Where a technical explanation for an AI tool is not possible, practical or meaningful, an ability to explain the priorities or strategic basis for a decision may suffice and may even be more meaningful depending upon the context.

Fairness

From one perspective, the fact that for certain classes of problems AI tools can ostensibly make decisions without the need for human intervention can be perceived as a benefit in circumstances where impartiality is seen as important.

However, this ignores the reality that AI tools are programmed by humans and that data bias (as opposed to direct human operator bias) poses a significant challenge for effective AI use.

Even the best AI tools can perpetuate historic inequality if biases in data are not understood and accounted for and the priorities of the system are not aligned with expectations of fairness. In other words, where an AI tool is applied to an uneven playing field and this has not been allowed for, flawed outcomes can result. This is particularly relevant when considering vulnerable or disadvantaged members of society such as indigenous populations who have suffered historic injustices. In a law enforcement context (which includes but is obviously not limited to fraud), bias against minorities has been particularly pronounced and its impact can be devastating.

As a further example, where an AI tool directs resources to a particular issue (for example fraud that is occurring in a certain sector or demographic) and then receives its 'learning' from that same issue, then the conclusions it reaches can be self-reinforcing. Notably, anti-discrimination rights (for example in a Bill of Rights) will still apply in the context of a decision or action that is driven by an AI tool.

The risks outlined above can be mitigated somewhat where AI is used to support human decision making rather than replace it, but as noted above, the possibility of automation complacency and automation bias mean that human input needs to be real rather than token.

In addition, a process is required to ensure that irrelevant and/or unfairly prejudicial characteristics are specifically excluded from the information provided to program and train the AI tool. This process will require careful consideration of the context in which the AI tool will operate, as well as who may be most affected by its use, to determine whether principles of fairness (which will include an awareness of historic or systemic discrimination) are being observed.

In summary:

- AI tools still require human input and if this information is flawed then outcomes will be affected. Agencies should not assume a level playing field and an AI system should have to prove it is correct.
- Inequities or biases (whether overt, latent, or historic) can be reinforced through the use of AI and the data fed into an AI system unless they are taken into account and normalised or corrected.
- A review process of any AI tool should consider the context for its use and those who may be most affected by it before drawing any definitive conclusions about its fairness or objectivity.

Privacy

Legislative regimes differ across jurisdictions, but the underlying privacy and associated civil liberties issues presented by the use of AI are relatively universal.

New Zealand

In a 2010 case before the Privacy Commissioner, a complaint was made about a fully automated transfer process between a debt collector and a credit reporter. The process was deemed to breach Principle 8 of the Privacy Act 1993, which requires an agency holding personal information to take reasonable steps to ensure that "the information is accurate, up to date, complete, relevant, and not misleading". The Commissioner held that, to be compliant with the accuracy aspect of Principle 8, a manual notation had to be added to the record. In effect, a human had to be kept "in the loop". The complainant, however, was unable to show that breach of Principle 8 caused her any harm.

For some time, technology has allowed a vast amount of data (including personal information) to be collected and stored by public sector agencies. Notwithstanding this fact, resource constraints had previously meant that there was an incentive to only gather targeted and proportionate amounts of material because the ability to analyse it in a meaningful way was limited. Now, with the use of AI tools, there is almost no limit on the amount of information that can be reviewed. This has in turn led to risks of overreach, whereby agencies are less likely to take a focused approach to data collection and therefore may face accusations of taking more information than they need. The risk is that this leads to an unnecessary compromise of the public's rights to privacy.

As a consequence of its analysis, AI is also able to take primary personal information and effectively anonymise it and turn it into inferred personal information. Whether the same privacy protections apply to this new category is open to legal debate and the position will differ from jurisdiction to jurisdiction, but it may create a further class of information that needs to be held and dealt with appropriately from a privacy perspective.

The data used to feed and train AI systems is often a mix of non-personal and personal data. The basis upon which agencies obtain this information is often that it may be used for intelligence purposes because, to the extent it is personal, it will be anonymised. Ironically, the same technology that allows AI analysis of such information to occur, also now poses a risk in the form of an ability to re-identify formerly anonymised data, therefore cutting across the basis upon which the information was gathered. Risks of re-identification are also heightened when inter-agency sharing occurs.

Also, as anonymised information is typically able to be used for any purpose if the individual is not identifiable, the point at which an obligation to delete information arises is unclear. This creates a civil liberties tension between an individual's 'right to be forgotten' and an agency's interest in holding the information in case it may be able to be used (potentially in re-identified form) for a useful purpose in the future.

Ultimately, agencies collecting information should already be guided in their approach by what is reasonable, necessary and proportionate rather than what technology will allow. Of course, there are considerable nuances to such an approach including how to define what is 'reasonable, necessary and proportionate' and who determines those matters. Privacy regimes in the various jurisdictions are likely to already set out principles that assist in determining what is or isn't a reasonable, necessary and proportionate collection of information. While additional considerations will apply in a fraud prevention/detection context, the relevant privacy requirements will represent a useful starting point in deciding whether the collection of information represents an overreach by the state.

It is also noteworthy that in responding to privacy requests from members of the public, most jurisdictions will have a 'maintenance of law' type exception that allows them to refuse such requests on the grounds that the detection and prevention of crime is a purpose that justifies limiting privacy rights in some way. It is vital that this important exception isn't undermined by a lack of rigour being applied to the collection of information. In other words, if a public sector organisation is going to limit privacy rights in order to prevent fraud, it must be seen to be operating in an ethical and appropriate manner.

We note also that from an agency perspective, secrecy/confidentiality protections may apply to information that is collected. Indeed, as a general proposition, privacy legislation which unduly restricts the sharing of information between public sector agencies can often limit the ability to detect fraud. While privacy protections are important in terms of ensuring that certain types of information will only be used for a specific set of purposes, the reality is that secrecy creates more barriers for the effective use of AI. Limits on the use of information gathered will have an impact on the quality of decision making, but again this is a balancing exercise that each jurisdiction must undertake.

From an unforeseen consequence perspective, it is also the case that using AI may impact on other areas of fraud control or government action. For example, an AI tool leading to someone being removed from a program for non-compliance may impact on a larger criminal operation into that person by tipping them off that they have come to government attention.

In summary:

- AI creates a category of inferred information which (on a conservative approach) should be afforded the same levels of protection as primary information.
- There should be limited reliance on the anonymous nature of material as a basis for how the information is used, shared, or held, as this status may not be permanent.
- When agencies gather information, they should be guided by what it is reasonable, necessary and proportionate rather than by what technology will allow.
- An ethical approach to the collection of private information is vital to retaining public faith in the use of AI to review that information.

Annex **A**

Annex A - Glossary

Term	Definition
AI Ethics	A set of values, principles, and techniques that employ widely accepted standards to guide moral conduct in the development and use of AI systems.
Algorithm	A set of step-by-step instructions. Computer algorithms can be simple (if it's 3 p.m., send a reminder) or complex (identify pedestrians).
Auditability	The ability of an AI system to undergo the assessment of the system's algorithms, data and design processes. This does not necessarily imply that information about business models and Intellectual Property related to the AI system must always be openly available. Ensuring traceability and logging mechanisms from the early design phase of the AI system can help enabling the system's auditability.
Bias	An inclination of prejudice towards or against a person, object, or position. Bias can arise in many ways in AI systems. Bias can be good or bad, intentional or unintentional. In certain cases, bias can result in discriminatory and/or unfair outcomes.
Black Box	A description of some deep learning systems. They take an input and provide an output, but the calculations that occur in between are not easy for humans to interpret.
Ethical AI	Used to indicate the development, deployment and use of AI that ensures compliance with ethical norms, including fundamental rights as special moral entitlements, ethical principles and related core values. It is the second of the three core elements necessary for achieving Trustworthy AI.
Machine Learning	The use of algorithms that find patterns in data without explicit instruction. A system might learn how to associate features of inputs such as images with outputs such as labels.
Supervised and Unsupervised Learning	A type of machine learning in which the algorithm compares its outputs with the correct outputs during training. In unsupervised learning, the algorithm merely looks for patterns in a set of data.
Trustworthy AI	Trustworthy AI has three components: (1) it should be lawful, ensuring compliance with all applicable laws and regulations (2) it should be ethical, demonstrating respect for, and ensure adherence to, ethical principles and values and (3) it should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm. Trustworthy AI concerns not only the trustworthiness of the AI system itself but also comprises the trustworthiness of all processes and actors that are part of the system's life cycle.

Annex **B**

Annex B[4] - Trustworthy AI Assessment List

1. Human agency and oversight

Fundamental rights:

- ✓ Did you carry out a fundamental rights impact assessment where there could be a negative impact on fundamental rights? Did you identify and document potential trade-offs made between the different principles and rights?
- ✓ Does the AI system interact with decisions by human (end) users (e.g. recommended actions or decisions to take, presenting of options)?
  - Could the AI system affect human autonomy by interfering with the (end) user's decision-making process in an unintended way?
  - Did you consider whether the AI system should communicate to (end) users that a decision, content, advice or outcome is the result of an algorithmic decision?
  - In case of a chat bot or other conversational system, are the human end users made aware that they are interacting with a non-human agent?

Human agency:

- ✓ Is the AI system implemented in work and labour process? If so, did you consider the task allocation between the AI system and humans for meaningful interactions and appropriate human oversight and control?
  - Does the AI system enhance or augment human capabilities?
  - Did you take safeguards to prevent overconfidence in or overreliance on the AI system for work processes?

Human oversight:

- ✓ Did you consider the appropriate level of human control for the particular AI system and use case?
  - Can you describe the level of human control or involvement?
  - Who is the "human in control" and what are the moments or tools for human intervention?
  - Did you put in place mechanisms and measures to ensure human control or oversight?
  - Did you take any measures to enable audit and to remedy issues related to governing AI autonomy?
- ✓ Is there is a self-learning or autonomous AI system or use case? If so, did you put in place more specific mechanisms of control and oversight?
  - Which detection and response mechanisms did you establish to assess whether something could go wrong?
  - Did you ensure a stop button or procedure to safely abort an operation where needed? Does this procedure abort the process entirely, in part, or delegate control to a human?

4 Extracted from the EU High-Level Expert Group Report dated 8 April 2019 on Ethics Guidelines for Trustworthy Artificial Intelligence – see https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. Note that the questions posed in this EU assessment list are in many cases open ended and are not intended to provide complete or definitive guidance to someone seeking to implement AI tools or systems within their organisation without reference to other material. For further reading we note the resources set out in **Annex C.**

2. Technical robustness and safety

Resilience to attack and security:

- ✓ Did you assess potential forms of attacks to which the AI system could be vulnerable?
  - Did you consider different types and natures of vulnerabilities, such as data pollution, physical infrastructure, cyber-attacks?
- ✓ Did you put measures or systems in place to ensure the integrity and resilience of the AI system against potential attacks?
- ✓ Did you verify how your system behaves in unexpected situations and environments?
- ✓ Did you consider to what degree your system could be dual-use? If so, did you take suitable preventative measures against this case (including for instance not publishing the research or deploying the system)?

Fallback plan and general safety:

- ✓ Did you ensure that your system has a sufficient fallback plan if it encounters adversarial attacks or other unexpected situations (for example technical switching procedures or asking for a human operator before proceeding)?
- ✓ Did you consider the level of risk raised by the AI system in this specific use case?
  - Did you put any process in place to measure and assess risks and safety?
  - Did you provide the necessary information in case of a risk for human physical integrity?
  - Did you consider an insurance policy to deal with potential damage from the AI system?
  - Did you identify potential safety risks of (other) foreseeable uses of the technology, including accidental or malicious misuse? Is there a plan to mitigate or manage these risks?
- ✓ Did you assess whether there is a probable chance that the AI system may cause damage or harm to users or third parties? Did you assess the likelihood, potential damage, impacted audience and severity?
  - Did you consider the liability and consumer protection rules, and take them into account?
  - Did you consider the potential impact or safety risk to the environment or to animals?
  - Did your risk analysis include whether security or network problems such as cybersecurity hazards could pose safety risks or damage due to unintentional behaviour of the AI system?
- ✓ Did you estimate the likely impact of a failure of your AI system when it provides wrong results, becomes unavailable, or provides societally unacceptable results (for example discrimination)?
  - Did you define thresholds and did you put governance procedures in place to trigger alternative/fallback plans?
  - Did you define and test fallback plans?

Accuracy

- ✓ Did you assess what level and definition of accuracy would be required in the context of the AI system and use case?
  - Did you assess how accuracy is measured and assured?
  - Did you put in place measures to ensure that the data used is comprehensive and up to date?
  - Did you put in place measures in place to assess whether there is a need for additional data, for example to improve accuracy or to eliminate bias?
- ✓ Did you verify what harm would be caused if the AI system makes inaccurate predictions?
- ✓ Did you put in place ways to measure whether your system is making an unacceptable amount of inaccurate predictions?
- ✓ Did you put in place a series of steps to increase the system's accuracy?

Reliability and reproducibility:

- ✓ Did you put in place a strategy to monitor and test if the AI system is meeting the goals, purposes and intended applications?
  - Did you test whether specific contexts or particular conditions need to be taken into account to ensure reproducibility?
  - Did you put in place verification methods to measure and ensure different aspects of the system's reliability and reproducibility?
  - Did you put in place processes to describe when an AI system fails in certain types of settings?
  - Did you clearly document and operationalise these processes for the testing and verification of the reliability of AI systems?
  - Did you establish mechanisms of communication to assure (end-)users of the system's reliability?

3. Privacy and data governance

Respect for privacy and data Protection:

- ✓ Depending on the use case, did you establish a mechanism allowing others to flag issues related to privacy or data protection in the AI system's processes of data collection (for training and operation) and data processing?
- ✓ Did you assess the type and scope of data in your data sets (for example whether they contain personal data)?
- ✓ Did you consider ways to develop the AI system or train the model without or with minimal use of potentially sensitive or personal data?
- ✓ Did you build in mechanisms for notice and control over personal data depending on the use case (such as valid consent and possibility to revoke, when applicable)?
- ✓ Did you take measures to enhance privacy, such as via encryption, anonymisation and aggregation?
- ✓ Where a Data Privacy Officer (DPO) exists, did you involve this person at an early stage in the process?

Quality and integrity of data:

- ✓ Did you align your system with relevant standards (for example ISO, IEEE) or widely adopted protocols for daily data management and governance?
- ✓ Did you establish oversight mechanisms for data collection, storage, processing and use?
- ✓ Did you assess the extent to which you are in control of the quality of the external data sources used?
- ✓ Did you put in place processes to ensure the quality and integrity of your data? Did you consider other processes? How are you verifying that your data sets have not been compromised or hacked?

Access to data:

- ✓ What protocols, processes and procedures did you follow to manage and ensure proper data governance?
  - Did you assess who can access users' data, and under what circumstances?
  - Did you ensure that these persons are qualified and required to access the data, and that they have the necessary competences to understand the details of data protection policy?
  - Did you ensure an oversight mechanism to log when, where, how, by whom and for what purpose data was accessed?

4. Transparency

Traceability:

- ✓ Did you establish measures that can ensure traceability? This could entail documenting the following methods:
  - Methods used for designing and developing the algorithmic system:
    - Rule-based AI systems: the method of programming or how the model was built;
    - Learning-based AI systems; the method of training the algorithm, including which input data was gathered and selected, and how this occurred.
  - Methods used to test and validate the algorithmic system:
    - Rule-based AI systems; the scenarios or cases used in order to test and validate;
    - Learning-based model: information about the data used to test and validate.
  - Outcomes of the algorithmic system:
    - The outcomes of or decisions taken by the algorithm, as well as potential other decisions that would result from different cases (for example, for other subgroups of users).

Explainability:

- ✓ Did you assess:
  - to what extent the decisions and hence the outcome made by the AI system can be understood?
  - to what degree the system's decision influences the organisation's decision-making processes?
  - why this particular system was deployed in this specific area?
  - what the system's business model is (for example, how does it create value for the organisation)?
- ✓ Did you ensure an explanation as to why the system took a certain choice resulting in a certain outcome that all users can understand?
- ✓ Did you design the AI system with interpretability in mind from the start?
  - Did you research and try to use the simplest and most interpretable model possible for the application in question?
  - Did you assess whether you can analyse your training and testing data? Can you change and update this over time?
  - Did you assess whether you can examine interpretability after the model's training and development, or whether you have access to the internal workflow of the model?

Communication:

- ✓ Did you communicate to (end-)users – through a disclaimer or any other means – that they are interacting with an AI system and not with another human?
- ✓ Did you label your AI system as such?
- ✓ Did you establish mechanisms to inform (end-)users on the reasons and criteria behind the AI system's outcomes?
  - Did you communicate this clearly and intelligibly to the intended audience?
  - Did you establish processes that consider users' feedback and use this to adapt the system?
  - Did you communicate around potential or perceived risks, such as bias?
  - Depending on the use case, did you consider communication and transparency towards other audiences, third parties or the general public?
- ✓ Did you clarify the purpose of the AI system and who or what may benefit from the product/service?
  - Did you specify usage scenarios for the product and clearly communicate these to ensure that it is understandable and appropriate for the intended audience?
  - Depending on the use case, did you think about human psychology and potential limitations, such as risk of confusion, confirmation bias or cognitive fatigue?
- ✓ Did you clearly communicate characteristics, limitations and potential shortcomings of the AI system?
  - In case of the system's development: to whoever is deploying it into a product or service?
  - In case of the system's deployment: to the (end-)user or consumer?

5. Diversity, non-discrimination and fairness

Unfair bias avoidance:

- ✓ Did you establish a strategy or a set of procedures to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data as well as for the algorithm design?
  - Did you assess and acknowledge the possible limitations stemming from the composition of the used data sets?
  - Did you consider diversity and representativeness of users in the data? Did you test for specific populations or problematic use cases?
  - Did you research and use available technical tools to improve your understanding of the data, model and performance?
  - Did you put in place processes to test and monitor for potential biases during the development, deployment and use phase of the system?
- ✓ Depending on the use case, did you ensure a mechanism that allows others to flag issues related to bias, discrimination or poor performance of the AI system?
  - Did you establish clear steps and ways of communicating on how and to whom such issues can be raised?
  - Did you consider others, potentially indirectly affected by the AI system, in addition to the (end)-users?
- ✓ Did you assess whether there is any possible decision variability that can occur under the same conditions?
  - If so, did you consider what the possible causes of this could be?
  - In case of variability, did you establish a measurement or assessment mechanism of the potential impact of such variability on fundamental rights?
- ✓ Did you ensure an adequate working definition of "fairness" that you apply in designing AI systems?
  - Is your definition commonly used? Did you consider other definitions before choosing this one?
  - Did you ensure a quantitative analysis or metrics to measure and test the applied definition of fairness?
  - Did you establish mechanisms to ensure fairness in your AI systems? Did you consider other potential mechanisms?

Accessibility and universal design:

- ✓ Did you ensure that the AI system accommodates a wide range of individual preferences and abilities?
  - Did you assess whether the AI system usable by those with special needs or disabilities or those at risk of exclusion? How was this designed into the system and how is it verified?
  - Did you ensure that information about the AI system is accessible also to users of assistive technologies?
  - Did you involve or consult this community during the development phase of the AI system?
- ✓ Did you take the impact of your AI system on the potential user audience into account?

- Did you assess whether the team involved in building the AI system is representative of your target user audience? Is it representative of the wider population, considering also of other groups who might tangentially be impacted?
- Did you assess whether there could be persons or groups who might be disproportionately affected by negative implications?
- Did you get feedback from other teams or groups that represent different backgrounds and experiences?

Stakeholder participation:

✓ Did you consider a mechanism to include the participation of different stakeholders in the AI system's development and use?

✓ Did you pave the way for the introduction of the AI system in your organisation by informing and involving impacted workers and their representatives in advance?

6. Societal and environmental well-being

Sustainable and environmentally friendly AI:

✓ Did you establish mechanisms to measure the environmental impact of the AI system's development, deployment and use (for example the type of energy used by the data centres)?

✓ Did you ensure measures to reduce the environmental impact of your AI system's life cycle?

Social impact:

✓ In case the AI system interacts directly with humans:

- Did you assess whether the AI system encourages humans to develop attachment and empathy towards the system?
- Did you ensure that the AI system clearly signals that its social interaction is simulated and that it has no capacities of "understanding" and "feeling"?

✓ Did you ensure that the social impacts of the AI system are well understood? For example, did you assess whether there is a risk of job loss or de-skilling of the workforce? What steps have been taken to counteract such risks?

Society and democracy:

✓ Did you assess the broader societal impact of the AI system's use beyond the individual (end)user, such as potentially indirectly affected stakeholders?

7. Accountability

Auditability:

- ✓ Did you establish mechanisms that facilitate the system's auditability, such as ensuring traceability and logging of the AI system's processes and outcomes?
- ✓ Did you ensure, in applications affecting fundamental rights (including safety-critical applications) that the AI system can be audited independently?

Minimising and reporting negative Impact:

- ✓ Did you carry out a risk or impact assessment of the AI system, which takes into account different stakeholders that are (in)directly affected?
- ✓ Did you provide training and education to help developing accountability practices?
    - Which workers or branches of the team are involved? Does it go beyond the development phase?
    - Do these trainings also teach the potential legal framework applicable to the AI system?
    - Did you consider establishing an 'ethical AI review board' or a similar mechanism to discuss overall accountability and ethics practices, including potentially unclear grey areas?
- ✓ Did you foresee any kind of external guidance or put in place auditing processes to oversee ethics and accountability, in addition to internal initiatives?
- ✓ Did you establish processes for third parties (e.g. suppliers, consumers, distributors/vendors) or workers to report potential vulnerabilities, risks or biases in the AI system?

Documenting trade-offs:

- ✓ Did you establish a mechanism to identify relevant interests and values implicated by the AI system and potential trade-offs between them?
- ✓ How do you decide on such trade-offs? Did you ensure that the trade-off decision was documented?

Ability to redress:

- ✓ Did you establish an adequate set of mechanisms that allows for redress in case of the occurrence of any harm or adverse impact?
- ✓ Did you put mechanisms in place both to provide information to (end-)users/third parties about opportunities for redress?

Annex **C**

Annex C - Key Publications On The Use of AI

Jurisdiction	Publication title	Link
OECD	OECD AI principles	https://www.oecd.org/going-digital/ai/principles/
OECD	Recommendation of the Council on Artificial Intelligence	https://one.oecd.org/document/C/MIN(2019)3/FINAL/en/pdf
	Ethics guidelines for trustworthy AI	https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
	A guide to using artificial intelligence in the public sector	https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector
	Government use of Artificial Intelligence in New Zealand	https://www.lawfoundation.org.nz/wp-content/uploads/2019/05/2016_ILP_10_AILNZ-Report-released-27.5.2019.pdf
	AI Ethics Principles	https://www.industry.gov.au/data-and-publications/building-australias-artificial-intelligence-capability/ai-ethics-framework/ai-ethics-principles
	Directive on Automated Decision-Making	https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592
	Government of Canada – Digital Playbook Guide	https://canada-ca.github.io/digital-playbook-guide-numerique/views-vues/automated-decision-automatise/en/automated-decision.html
	Artificial Intelligence: Emerging Opportunities, Challenges, and Implications	https://www.gao.gov/products/GAO-18-142SP

Notes

Notes

Notes