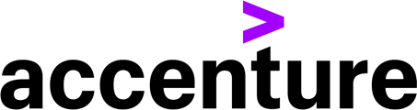
**A picture containing clothing, underwear, umbrella

Description automatically generated****Accenture response to Ontario’s Trustworthy AI Framework Consultation**

**Trustworthy AI:**

**Explainability,**

**Trust, and Inclusivity**

June 4, 2021

**A picture containing clothing, underwear, umbrella

Description automatically generated**

**Submitted by:**

Accenture Inc.

Shannon MacDonald

[shannon.macdonald@accenture.com](mailto:shannon.macdonald@accenture.com)

Graeme Larsen

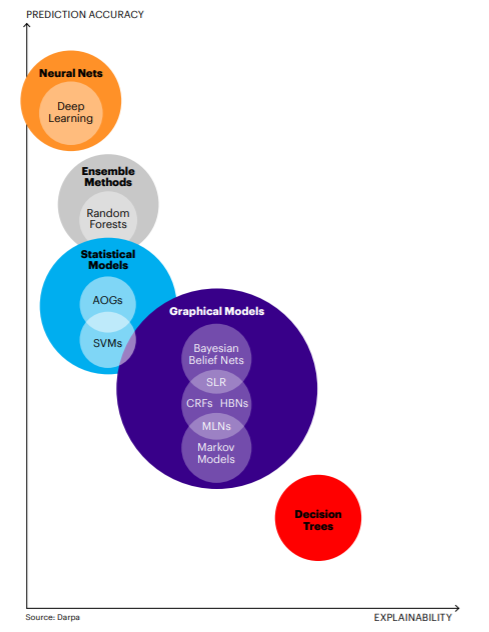
[graeme.larsen@accenture.com](mailto:graeme.larsen@accenture.com)

**Data Transparency and Explainability**

Any time data is collected from the public, governments have the responsibility to follow certain basic principles of data ethics to ensure the person’s safety and privacy is maintained. According to Accenture’s [*Universal Principles of Data Ethics*](https://www.accenture.com/_acnmedia/pdf-24/accenture-universal-principles-data-ethics.pdf), the top two priorities are to not only respect the persons behind the data, but also to attend to the downstream uses of datasets. This means not only providing transparency in how data is collected and stored, but also in the downstream uses of this data.

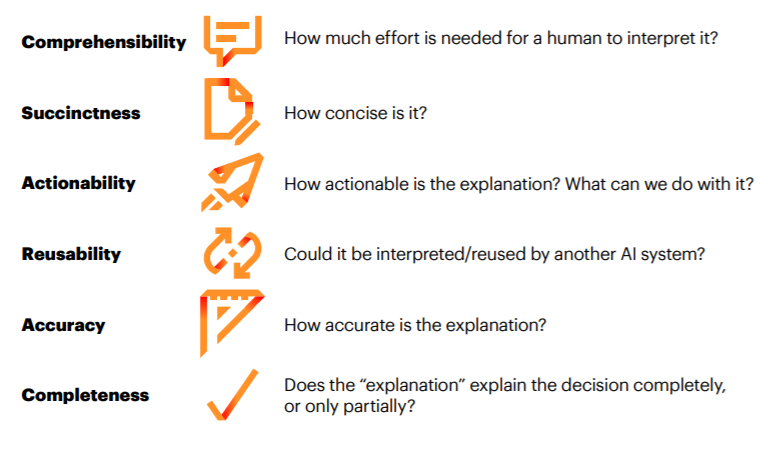
Many regulations govern datasets on the basis of status, labelling it as “public”, “private”, or “proprietary”. However, what is done with datasets is ultimately more consequential to subjects than the type of data or the context in which it is collected. Therefore, the collection of data from the public must be accompanied by information around how the data intends to be used. Such information should be easily accessible to the public and be communicated at the time of collection. This will be important later when discussing contextual data and its implications on AI decision-making.

The challenge of ensuring transparency in AI is inherently tricky. Compared to AI, the algorithms of the past were relatively simple. In most cases, researchers and data scientists were still able to determine how their models produced the answers they came up with. With the advent of deep learning and neural networks, however, taking a peek under the hood has gotten much more difficult.

While it may be acceptable to not understand the rational behind your Spotify recommendations, this lack of transparency can have far-reaching consequences in other applications. For example, in the US, black-box risk assessment tools are being used in the justice system to make sentencing recommendations. In more than one case, [this has led to defendants receiving longer sentences](https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/) based on the outputs of algorithms they were not allowed to review. Naturally, this doesn’t play well with the public.

As consumers and citizens, we want AI to deliver products, services, and information to provide us with an appropriate level of transparency, but the more complex algorithms are, the more difficult it becomes to understand their decisions. This trend is outlined in the chart to the right.

It is from this need that the concept of Explainable AI (XAI) is born. XAI is defined as a system with the ability to explain rationale for decisions, characterize the strengths and weaknesses of the decision-making process, and convey an understanding of how they will behave in the future. There already exists frameworks and techniques for ensuring that AI outputs are delivered in an interpretable and faithful manner. Most notably, a technique called [Local Interpretable Model-Agnostic Explanations (LIME)](https://homes.cs.washington.edu/~marcotcr/blog/lime/) has been developed by researchers at the University of Washington.

An ongoing debate for XAI is the question of to whom the explanations should be directed. It is one thing to explain a decision to advanced mathematicians or engineers, but it is a whole other challenge to break processes down for the common person to understand. Also, explanations can vary in complexity depending on the system and its underlying methods. In the future, Accenture is forecasting the emergence of a new “Explainer” role to fill the void, allowing for technical explanations to be translated into more common language. Accenture has outlined 8 measures to assess the value and effectiveness of any AI explanation. These measures are outlined below.

Going forward, AI promises to help us identify dangerous industrial sites, warn us of impending machine failures, recommend medical treatments, and take countless other decisions. But the promise of these systems won’t be realized unless we can understand, trust and act on the recommendations they make. Therefore, it is only once these requirements are also met that an AI system can be truly transparent to the population.

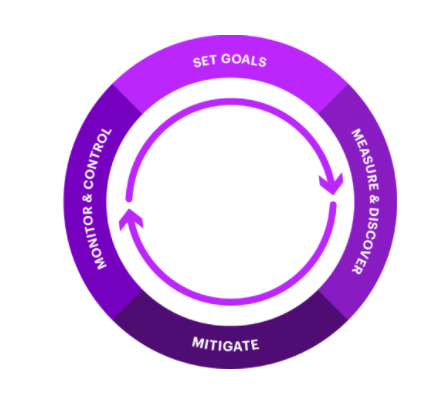
Transparency is a key component used to build trust in AI systems, however it is not the only one. Next, we will look at other factors that will enhance public confidence in AI, allowing us to leverage AI to the fullest extent of its capabilities.

**Building Trust in Automation**

Trust gives human interaction meaning. It helps us separate truth from fiction and distinguish good intentions from bad. This is by no means a trivial matter. As businesses continue to expand their use of artificial intelligence, consumers will increasingly interact with digital agents. They will need to be able to put their trust in these AI systems when they apply for insurance, student loans or a mortgage. But establishing that trust is easier said than done. There are significant challenges that must be addressed on the way to creating trustworthy, responsible AI, beyond what has already been examined.

[In a 2019 global survey of risk managers](https://www.accenture.com/us-en/insights/financial-services/global-risk-study), 58% identified AI as the single biggest potential cause of unintended consequences in the near future. Only 11% of the same risk managers described themselves as fully capable of assessing the risks associated with adopting AI. One of the main reasons for this crisis of confidence? The fact that AI shifts the risk landscape in new and unexpected ways for organizations. As AI increasingly interacts with people and influences their lives at scale, responsibility to manage the potential ethical and socio-technical implications of AI adoption also increase.

Accenture has outlined an algorithmic assessment framework that we believe should be considered in all AI applications. The **Algorithmic Assessment** is a technical evaluation that helps identify and address potential risks and unintended consequences of AI systems, creating trust and building supportive systems around AI decision-making. Use cases are prioritized to ensure that those with the highest risk and impact are being evaluated first. Once priorities are defined, they are evaluated using a series of qualitative checks to support various stages of AI development. The assessment consists of four steps:



1. **Set goals**around your fairness objectives for the system, considering different end users.
2. **Measure & discover**disparities in potential outcomes and sources of bias across various users or groups.
3. **Mitigate**any unintended consequences using proposed remediation strategies.
4. **Monitor & control**systems with processes that flag and resolve future disparities as the AI system evolves.

The Centre for Information Policy Leadership has provided suggestions for a more standardized approach to risk management. In its paper, [*The Role of Risk Management in Data Protection*](http://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/white_paper_2-the_role_of_risk_management_in_data_protection-c.pdf), the Centre concludes that risk management is “a valuable tool for calibrating the implementation of and compliance with privacy requirements, prioritizing action, raising and informing awareness about risks (and) identifying appropriate mitigation measures.”

Based on an earlier paper by the Centre titled [*A Risk-based Approach to Privacy: Improving Effectiveness in Practice*](https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/white_paper_1-a_risk_based_approach_to_privacy_improving_effectiveness_in_practice.pdf), Accenture has developed a risk assessment matrix to judge an application against generally accepted relevant concerns. While this matrix is in draft and does not purport to be an exhaustive analysis, it offers a helpful tool that can be customized to a wide variety of situations. The matrix can be seen in the table below.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Unjustifiable Data Collection** | | | **Inappropriate Use** | | | **Security Breach** | | | **Aggregate** |
|  | Sensitive Information | | | Inaccuracies  Biases  Poor Assumptions | | | Lost/Stolen Data  Access Violation | | |  |
| **Risks** | **Likeliness** | **Severity** | **Total** | **Likeliness** | **Severity** | **Total** | **Likeliness** | **Severity** | **Total** | **Risk Rank** |
| **Tangible Harm** | | | | | | | | | | |
| **Bodily Harm** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Loss of Liberty or Freedom** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Financial Loss** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Other Tangible Loss** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Intangible Distress** | | | | | | | | | | |
| **Excessive Surveillance** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Suppression of Free Speech** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Suppression of Associations** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Embarrassment or Anxiety** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Discrimination** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Excessive State Power** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Loss of Social Trust** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Legend:**

Rank “*Likeliness*” from **1 (low) to 10 (high)** based on the highest score for any component

Rank “*Serious*” from **1 (low) to 10 (high)** based on the highest score for any component

**Aggregate Risk Rate:**

Highest score is 300

Lowest score is 0

Using a risk assessment matrix like the one above, it is possible to gauge the potential risks associated with any AI system quickly and easily, and then make decisions around how to mitigate these risks. Using such a matrix, priority can also be given to assessing and managing those applications which stand to have the largest negative effect on the population, allowing regulators to maximize their resources and ensure their work has the biggest impact.

Furthermore, algorithms are constantly changing, and their risk scores may also change as they become more widespread and interact with a larger percentage of the population. The government should seek to implement a system to continuously reassess the risk associated with algorithms that are widely used. An Ethics Committee within government should also be established to oversee the continued monitoring of algorithms in the public domain, and this will be discussed more in the next section.

By assessing and continuously reassessing the risks associated with these systems, while also remaining transparent in their processes and procedures, the government will instill trust in the public, allowing it to benefit greatly from the resulting automation.

**Inclusive AI for All**

The biggest question anyone must ask when assessing automated systems is what happens if its decision outcomes turn out to be incorrect or even unlawful? The potential fallout for both the government and its citizens could be immense. And what about other unintended consequences of its decisions?

AI has already shown it can be biased in ways that weren’t anticipated and can be harmful to certain groups that interact with it. Amazon, for instance, had to scrap its AI-based recruiting tool that appeared to show bias against women. So the age-old question remains: how does a human know when to intervene in a process driven by a machine?

In part, this relates back to the person’s understanding of how the machine is actually making its decisions, as was discussed previously. However, it's not just about establishing the appropriate governance structures. It’s also important to translate those ethical and legal frameworks into statistical concepts that can be unambiguously represented in software.

To do so, considerations around the potential ethical consequences of AI must be built into core values and a robust compliance process should be established. We program algorithms to give us exactly what we ask for, and we shouldn’t be surprised when they do. The problem is that simple algorithms treat all data as absolute, excluding contextual data that often gives the information meaning. Because of this, AI can sow the seeds for systemic discrimination, leading autonomous systems to make decisions that unfairly disadvantage certain groups. This phenomenon is only exacerbated by the training process, which propagates and often amplifies any biases that are present in the data. The only way to combat this is by identifying and eliminating unconscious bias in the data before AI interacts with it.

Because this issue is so prevalent, [Accenture has developed a tool to help pinpoint and avoid sources of bias in datasets](https://techcrunch.com/2018/06/09/accenture-wants-to-beat-unfair-ai-with-a-professional-toolkit/), effectively making it easier to raise AI systems well. For the same reasons, we test all our solutions extensively and change our own datasets and algorithms whenever necessary.

We also recommend that all organizations build Data and AI Ethics committees to ensure they are using AI responsibly in their work, and [have created a roadmap for ethics oversight in AI](https://news.northeastern.edu/2019/08/29/northeastern-researchers-team-with-accenture-to-offer-a-road-map-for-ethics-oversight-for-artificial-intelligence/). After all, not using people’s information illegally is the minimum responsibility for organizations, and it will not be enough to sustain trust, manage risk or be responsive to stakeholder expectations. Such expectations should also be extended to government, and these committees should consider the voices from both industry, academia, and the communities that they serve. In doing so, they can provide oversight from a variety of perspectives, helping to propose legislative changes that will better regulate AI use in the province. They may also work towards developing a certification program through which AI applications may be vetted before being implemented at scale, creating a safer system that is less prone to bias and discrimination, either from the developer or the underlying data.

It is Accenture’s belief that all people stand to benefit from the use of AI, if implemented properly. Efforts should therefore be used to promote acceptable use, even in the populations that are the most vulnerable. However, in such situations extreme care should be taken to ensure that the system meets all necessary requirements to avoid unintended negative repercussions. Risk thresholds may need to be adjusted in these settings as a result, although a good risk assessment tool should account for these factors.

Artificial intelligence will empower mankind, enabling innovative technologies and solutions we can only dream of today. But that power must be balanced by responsibility, transparency, and fairness. Just as data privacy and cybersecurity are quickly becoming the responsibility of the government, Responsible AI must also be quickly elevated in importance as its effect on society will be immense. In many ways, the technology behind AI is well developed and accessible. We now have a unique opportunity to determine what the human side of the equation should look like, and the time to act is now.

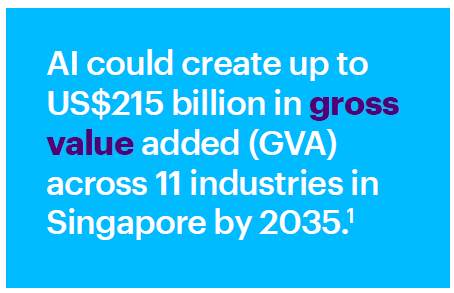
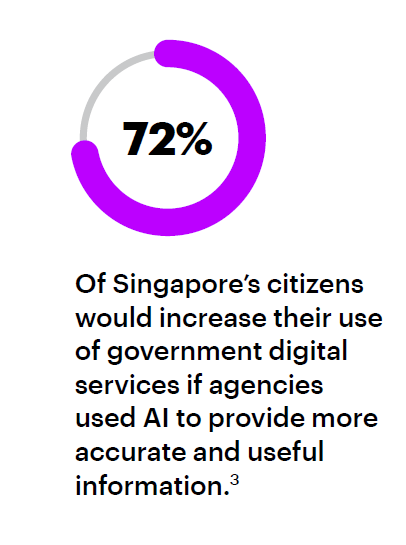
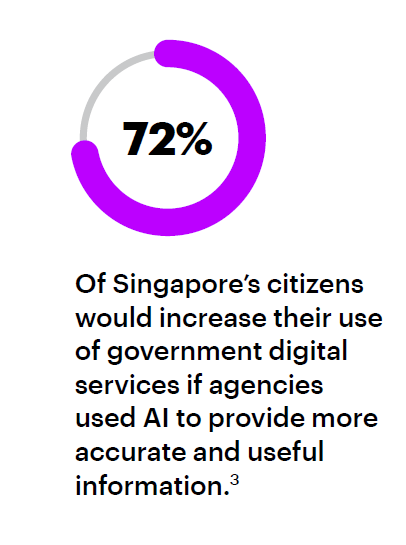
The largest challenge will be simple adoption and acceptance of AI in citizen services as employees tend to protect jobs and fear downsizing. The true human benefit of increased AI application in government will be a shift in people’s time from administrative tasks to increased connection time with citizens. The second most significant benefit is the enabling of self-service and control for accessing information and programs in the hands of the citizens, a consumer-like expectation that citizens increasingly expect. Acknowledging the change management challenges and creating compelling use cases to push the AI opportunity throughout government is essential to successful implementation.

**AI Public Sector Use Cases**

***Government of Singapore***

From cutting-edge labs to public agencies across the Lion City—healthcare, transportation, tourism, customs, public safety, education and more—the Singaporean government is investing in AI to improve the day-to-day lives of citizens and make public service delivery more effective, responsive and efficient. This investment is part of Singapore’s push to become an AI powerhouse in Southeast Asia—and the world. The government has named AI as one of the four key technologies for digital readiness, and it is also the backbone of its smart city infrastructure. In a first for the region, Singapore created a governance framework with the World Economic Forum’s Centre for the Fourth Industrial Revolution that addresses the responsible and ethical use of AI.

Here, AI is not coming. It has arrived. **The public sector in Singapore has moved past the AI experimentation phase to implementation.**

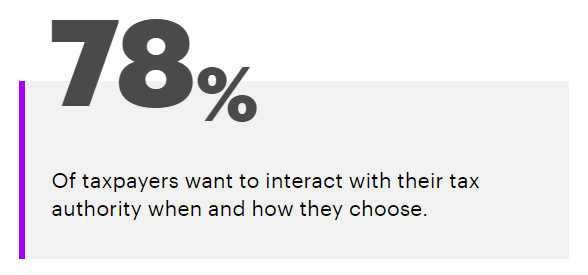
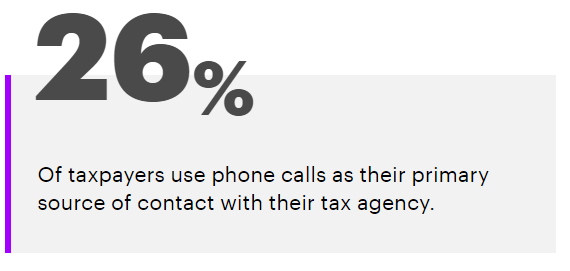


***Improving the Taxpayer Experience***

**Tax. Known for complexity over customer service, it’s not the first place you look for innovation inspiration. Unless you’re The Office of the Revenue Commissioners in Ireland (Revenue). In a world first, Revenue saw the opportunity to better meet customer needs using new technology, delivering game-changing outcomes.**

For John Barron, CIO at Revenue, this customer service demand represented the perfect chance to take advantage of automated services, as well as cutting-edge natural language processing. Revenue took a subset of calls related to tax clearance, to pilot a fresh way of managing customer calls.

The aim of this pilot was to prove the technology and offer a 24/7 automated service, providing the most efficient, effective experience for customers. **The result? A functioning system that both reduced the burden on call centre agents and gave the caller the freedom to make calls whenever they choose**.

With most taxpayers looking for more flexible interaction with their tax agency, and most being familiar with AI, a virtual agent is a suitable tool for Revenue agencies to meet customer demands.