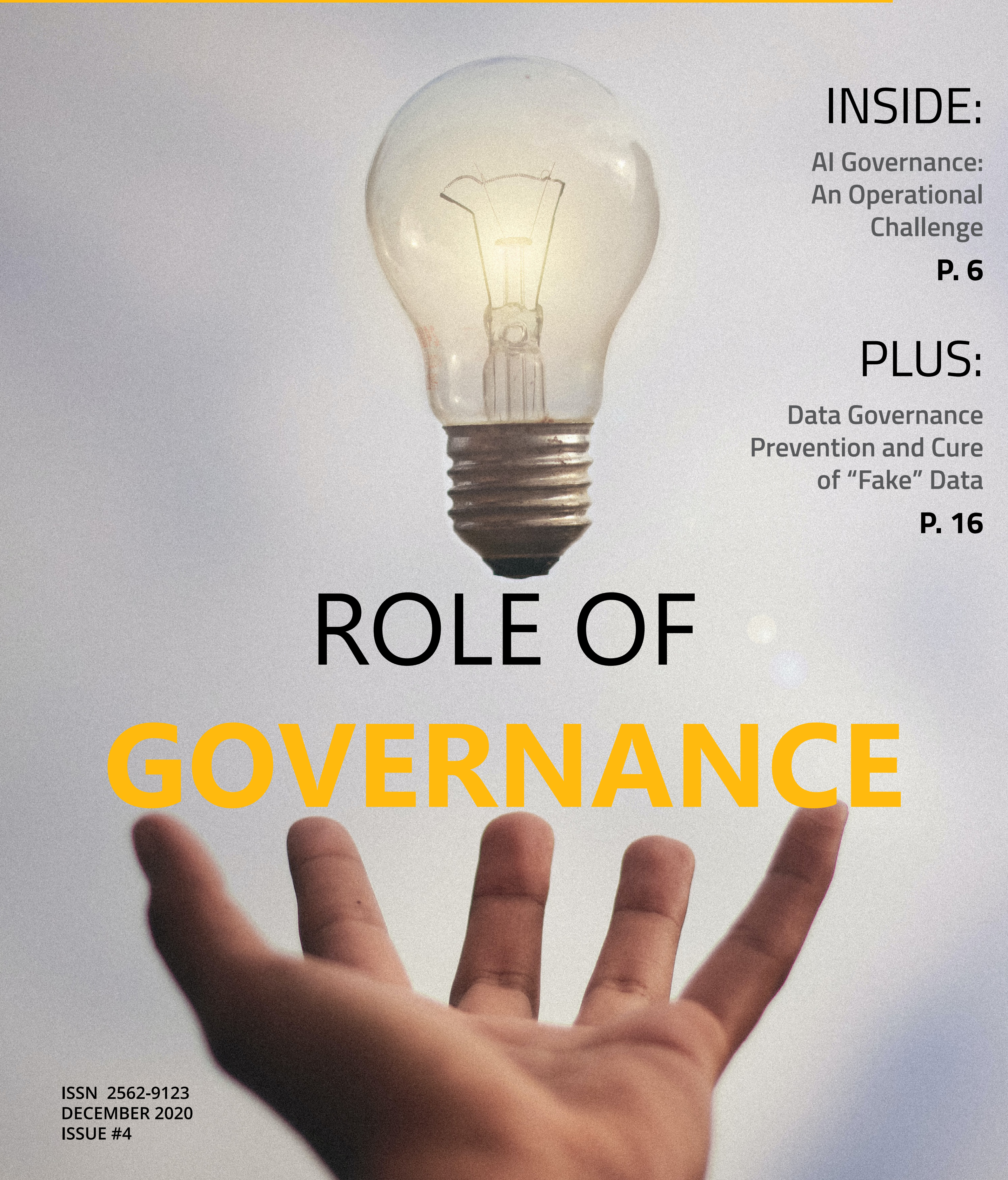
ANALYTICS IN GOVERNMENT QUARTERLY

FOR GOVERNMENT DECISION MAKERS



INSIDE:

AI Governance:
An Operational
Challenge

P. 6

PLUS:

Data Governance
Prevention and Cure
of “Fake” Data

P. 16

ROLE OF GOVERNANCE

ISSN 2562-9123
DECEMBER 2020
ISSUE #4

CONTENTS

FEATURES

6/Artificial Intelligence Governance: An Operational Challenge

By Hubert Laferrière

10/Institutionalized AI: A Governance Conundrum?

By Gregory Richards, Ph.D.

12/Program Integrity, Entity Resolution and A 360-degree View of the Citizen

By Amanda Holden & Dan Finerty

16/Data Governance Prevention and Cure of "Fake" Data

By Betty Ann M. Turpin, Ph.D.

COLUMNS

4/Data Science

Making the Business Case for Data Governance: Improved Public Service

By Kevin Kells, Ph.D.

19/Point of View

Governing Analytics to Promote Diversity and Inclusion

By Tara Holland

ANALYTICS IN GOVERNMENT QUARTERLY

Editorial

Managing Editor | Gregory Richards, Ph.D.

Editor & Design | Yvonne Leng

Columnists |
Boris Bogatirev
Kevin Kells, Ph.D.
Hubert Laferrière
Rami Abielmona, Ph.D.

EDITOR'S NOTE

Hello again. I hope everyone is managing well in our COVID-infused world. This issue continues our discussions related to the various uses of artificial intelligence and the challenges associated with ensuring transparency and preventing fraud.

Amanda Holden from SAS Canada provides an overview of how analytics can be used to create a 360 degree view of data for entity resolution—ensuring that the person with whom you might be working in a digital world is indeed the right person. Hubert Laferrière addresses AI governance and the need for the responsible use of AI, providing details about what this means in a government context and an approach used by IRCC to deal with the issues. Betty Ann Turpin discusses ways of preventing the propagation of "fake data" and for resolving the situation should data become corrupted. Kevin Kells follows up with guidelines and critical success factors for sound data governance. Tara Holland addresses the important issue of diversity and inclusiveness in analytics. Finally, I close this issue with a bit of an integrative overview of institutionalizing AI (integrating it into day-to-day operations) and the important role of governance in doing so. In this article, I suggest that government organizations already have tools in place that can be tweaked to ensure sound governance of AI initiatives.

You can see that the overall theme of this issue is about safety and security of your data in an increasingly complex world. It's a direct analogy to our world at the moment; please stay safe and help others to do so as we move into the 2020 holiday season.

Gregory Richards, Ph.D.
Managing Editor

Corporate

Analytics in Government Quarterly magazine is published 4 times per year by the Government Analytics Research Institute, a consortium of the University of Ottawa, Carleton University, the University of Quebec en Outaouais, SAS and the Institute on Governance. The institute conducts research with government organizations who are experimenting with the introduction of analytics of all forms. Professors and students work on proof of concepts, testing of algorithms as well as examining the organizational practices needed to fully integrate analytics into business processes.

All opinions expressed herein are those of the contributors and do not necessarily reflect the views of the publisher or any person or organization associated with the magazine.

General Inquires

Letters, submissions, comments and suggested topics are welcome, and should be sent to agq@governmentanalytics.institute or visit our website http://governmentanalytics.institute

Subscription Information

You can subscribe to the magazine online at http://governmentanalytics.institute/magazine

Reprint Information

Reproduction or photocopying is prohibited without the publisher's prior consent.

Privacy Policy

We do not sell our mailing list or share any confidential information about our subscribers.

ISSN 2562-9123

By Kevin Kells, Ph.D.

Making the Business Case for Data Governance: Improved Public Service

Demonstrating the potential value of data governance for the organization is key to obtaining buy-in from stakeholders organization-wide and making the introduction of a data governance program a success. A recent review[1] of data governance in the academic and industry literature summarized "critical success factors" (CSFs) for data governance.

One group of CSFs spoke to the importance of organizing the effort well to introduce data governance:

- Establish data governance team structure;
- Define roles and responsibilities;
- Develop processes, procedure guidelines, principles, policies, and standards to support the data governance;
- Assess the data governance situation;
- Define the sustaining requirements.

A second group of CSFs involved technology and data governance implementation:

- Monitoring tools and metrics;
- Technology;
- Accountability;
- Compliance monitoring; and
- Data governance tools.

But the largest group of CSFs centered around outreach to organizational stakeholders, organizational culture, and

[1] Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2018). A systematic literature review of data governance and cloud data governance. Personal and Ubiquitous Computing

communication:

- Develop a business case for data governance;
- Awareness of data stakeholders;
- Develop a communication plan;
- Develop an integration process;
- Develop a change management plan;
- Training and education.

An approach to introducing data governance that embraces the need to promote buy-in and participation begins with two-way communication with organizational stakeholders to co-author a business case for data governance. Input from key stakeholders should reflect stakeholders' positive future-picture of how a successful data governance program will help their functioning, improve their efficient use of resources, and ultimately allow them to provide a better service both internally and ultimately to the public. This input can also relate a few, practical data-quality "horror stories" to emphasize the contrast of this improved future compared with the status quo. Include a visual vignette as an example of how a department or team may use and present data to clarify complexity better and to provide improved actionable intelligence under a data governance program.

Compiled into a compelling story, this stakeholder input can be persuasive throughout the organization, avoiding resistance to governance policies when they are introduced, and inspiring confident championing of the data governance effort from the leadership suite.

The team that spearheads the planning and implementation of the data governance program should organize itself well and include the mentioned success factors. Due focus should also be placed on the technology and implementation of the data governance principles, policies, and procedures. And yet the most important consideration for organizational buy-in is to make the business case – and to have current data stakeholders help make this case – tying improved functioning, better use of resources, and ultimately improved public service to the data governance effort, in line with the organization's goals and vision.

About the Author

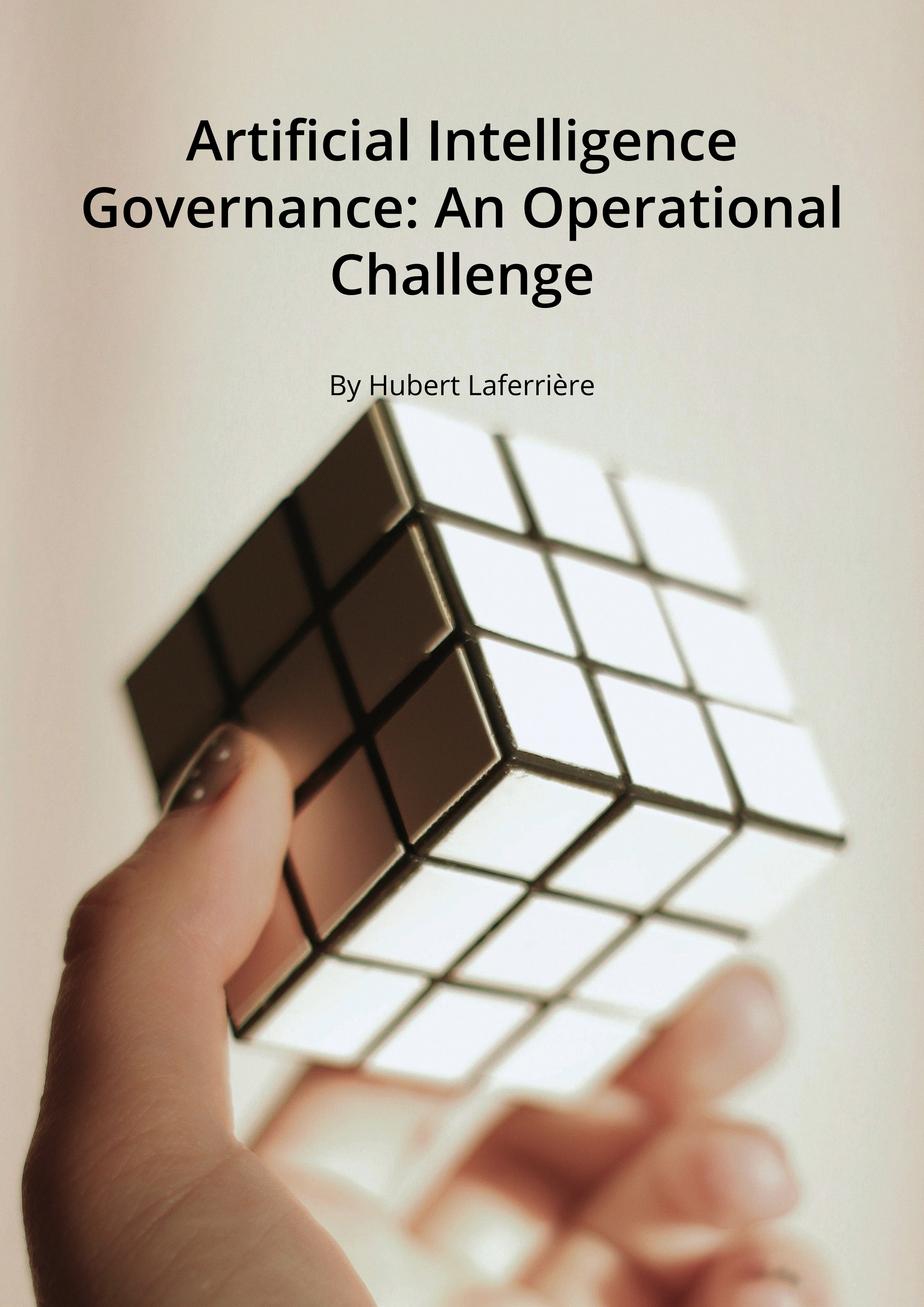
Kevin Kells, Ph.D., has worked as an R&D Engineer in software systems in the Financial and Semiconductor industries in Switzerland, Silicon Valley, and Ottawa, and currently works with real-time data and news feed systems at a major market news and data company in New York City.

He has extensive experience in non-profit management, both in the area of human systems and IT systems.

He received his Ph.D. from the Swiss Federal Institute of Technology (ETH), Zurich in computer simulation of semiconductor devices and holds an MBA with areas of focus in entrepreneurship and business analytics from the University of Ottawa, Telfer School of Management.

Artificial Intelligence Governance: An Operational Challenge

By Hubert Laferrière



Absence of AI Governance

In their 2020 State of AI report, " ... a compilation of the most interesting things ... about the state of AI and its implication for the future." N. Benaich and I. Hogarth, two AI investors, mentioned the only prediction among the six they made in 2019 that did not materialize is AI governance. For them, the necessity for governance is sine qua non given the increasing power of AI systems and interest of public authorities. Despite many actors' attempts to define principles for responsible use of AI, nothing substantial has been achieved, except loosely-stated fashion principles.[1] However, the authors did not identify the factors that may have contributed to the absence of AI governance framework nor what is required to establish such a framework. For example, should AI governance be underpinned by a body of law like the EU General Data Protection Regulation?

Responsible Use of AI

Yet, over the past years, many players have published proposals for the governance of AI, ranging from high-level principles to more down-to-earth directives. In Canada, the Toronto Declaration (Protecting the right to equality in machine learning) and the Montréal Declaration (For a responsible use of AI) were published in 2018 with the aim of developing a responsible use of AI.[2] Both declarations are the result of collaborative work between partners from different backgrounds.

The Government of Canada led the development of AI guiding principles, adopted by leading digital nations in the same year to ensure a responsible use of AI while supporting service improvement goals.[3] Five guiding principles were established: understand and measure the impact of using AI; be transparent about how and when AI is used; provide meaningful explanations about AI decision making; share source code, training data, while protecting personal information; and government employees have the skills to make AI-based public services better.

The principles intersect with the major trends identified by the Berkman Klein Center at Harvard University in a comparative analysis of 36 principles documents aimed at providing normative guidance regarding AI-based systems. Each document had a similar basic intent: to present a vision for the governance of AI. The documents were authored by actors coming from different sectors of the public and civil societies, such as governments and intergovernmental organizations, private sector firms, professional associations, advocacy groups, and multi-stakeholder initiatives. Although the goals sought among them were different, the Centre has identified eight key or common thematic trends that rally the stakeholders. The trends comprise ethical and human rights-based principles that could guide the development and use of the AI technologies (privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and promotion of human values).[4] For the Center, conversation around principled AI is converging towards a responsible development of AI. "Thus, these themes may represent the "normative core" of a principle-based approach to AI ethics and governance."[5]

Explaining the absence of AI governance, as Benaich and Hogarth assert, is not a straightforward undertaking since a wide range of factors could be considered. For instance, one may argue that this absence is the result of a divergence of views between the public and private sectors. Some AI leading players in the private sector prefer to adopt their own guiding principles rather than having to comply with AI principles endorsed by public authorities or governments. Others may bring forward the proliferation, in the public and civil sectors arena, of committees, working groups, boards and commissions on AI and ethics. Despite good intentions, this proliferation has

[1] N. Benaich, I. Hogarth (2020), State of AI Report 2020

[2] Amnesty International, Access Now (2018), The Toronto Declaration, Protecting the right to equality in machine learning and Montreal University (2020).

[3] Treasury Board of Canada Secretariat (2019), Ensuring responsible use of artificial intelligence to improve government services for Canadians.

[4] J.Fjeld, N. Nele, H. Hilligoss, A. Nagy, M. Madhulika, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI (January 15, 2020).

[5] Ibid., p. 5

not yet generated an enabling AI governance framework. Discussions are too often limited to insiders and have yet to come to fruition.

However, stating that there is an absence of AI governance does not entirely reflect the reality of organizations using AI technology.

Managing AI

Those adopting the AI technology, in particular predictive analytics, machine learning, neural network, and deep learning, implemented management frameworks to support the development and the production of AI solutions. The management of AI activities is based on specific practices, processes, methods, and procedures related to the technology. Tools, such as the data science lifecycle and standardized methodologies such as CRISP-MD and ASUM-DM[6], are assisting AI practitioners and management to shape activities and processes, including performance monitoring and assessment. The tools support efforts to address unique AI issues such as training data, algorithms bias, and the drifting of AI models, to name a few. As a result, organizational and managerial efforts focus on setting the conditions for facilitating the technology to operate and produce value. The very AGQ first issue was on "Responsible AI" and provided a series of considerations to move forward.[7]

Two Spheres, Two Solitudes

In this context, the AI governance seems split in two spheres: (1) managing per se the dynamics (development and production) of the new technology, what is unique and specific to it and, (2) managing the impact of the technology, in particular issues and concerns raised by stakeholders, including clients affected by this technology. This impact is calling for ethical and human rights-based principles and legal frameworks.

These spheres have their own distinct interest and activities. I suggest that recognizing this situation could help to shed some light on the absence (at least the slow pace) of AI governance as observed in the 2020 State of AI report.

The problem resides in these spheres evolving in parallel, having their own management processes and procedures and, accordingly, being developed separately as two solitudes. This leads to a lack of integration or what has been identified as an absence of AI governance. In addition, each sphere cannot by itself offer an AI governance framework that will support efforts in maximizing the benefits and minimizing the harms of AI. Bridging or interlocking both spheres becomes an imperative.

Efforts to bring together the two spheres could be a path for developing a sound AI governance framework. This path must enable the achievement of a precise objective: ensure guiding principles and directives are transposed into a modus operandi for AI practitioners, too often stuck with second guessing the principles' meaning and aim.

A Challenge

Meshing the two spheres is a challenge that must be tackled by the AI community, in particular the AI practitioners, and the key stakeholders of public and civil societies including governments.

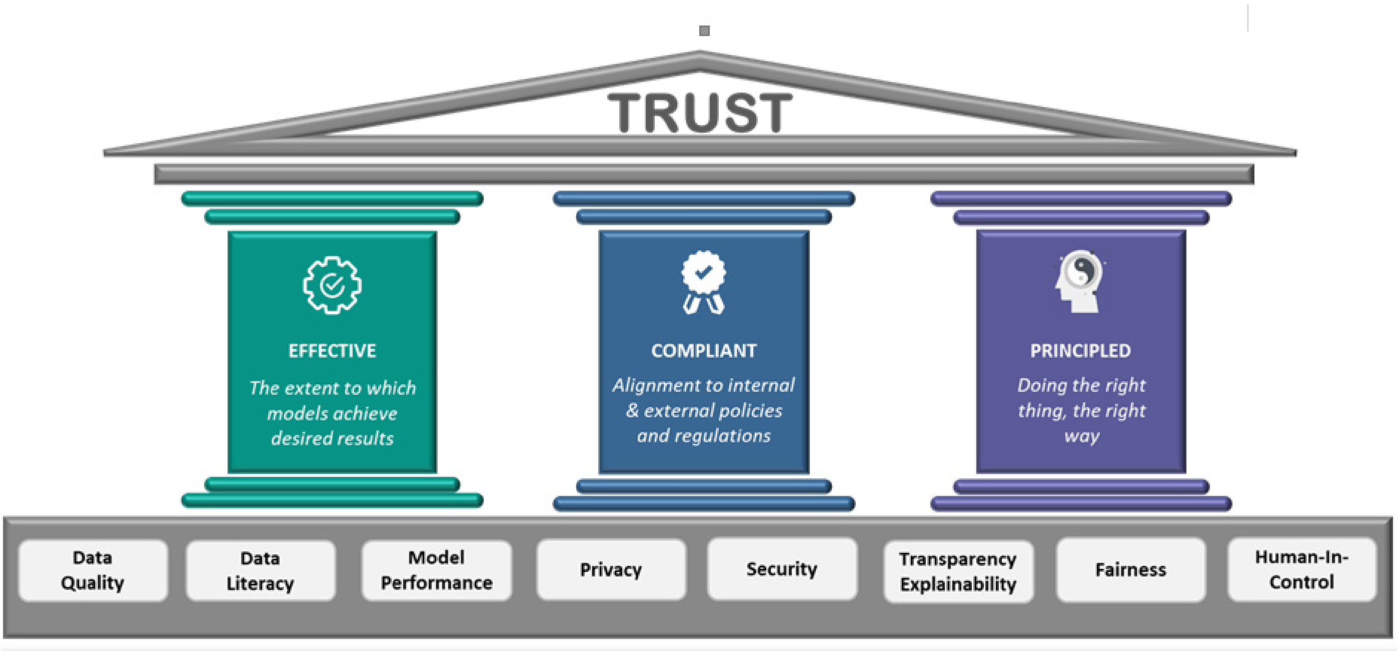
Guiding principles may provide a sense of direction but often they are too general, subjected to a plethora of opinions, while abstract by essence. This renders the efforts and attempts to embed the principles into AI processes laborious and inefficient. Building the bridge between the AI day-to-day practices sphere and the requirements of the guiding principles sphere is not a given. This even if some principles are already enshrined in laws and policies, for example the privacy laws, current requirements may be restrictive, potentially stifling the technology, and probably not adapted to the new reality of AI.

Embedding Operational Practices into AI Business Process

AI practitioners need to embed the principles into operational practices and into components that will make up the DNA of their operations. AI practitioners carry out complex tasks and operations and need to outline operational practices which allow the integration of a guiding

[6] "Cross-industry standard process for data mining, known as CRISP-DM, is an open standard process model that describes common approaches used by data mining experts. It is the most widely-used analytics model. In 2015, IBM released a new methodology called Analytics Solutions Unified Method for Data Mining/Predictive Analytics (also known as ASUM-DM) which refines and extends CRISP-DM." (extract from Wikipedia).

[7] Analytics In Government Quarterly For Government Decision Makers (2019), Government Analytics Research Institute (GARI)


TRUST
EFFECTIVE
The extent to which models achieve desired results
COMPLIANT
Alignment to internal & external policies and regulations
PRINCIPLED
Doing the right thing, the right way
Data Quality
Data Literacy
Model Performance
Privacy
Security
Transparency Explainability
Fairness
Human-In-Control

principle into the AI business process, in their AI day-to-day activities. The operational practice represents a roadmap for building the bridge that interlocks the two spheres. An operational practice contributes to create efficiencies, to ensure the AI solution is aligned with guiding principles and comply with the law. An operational practice brings consistency and reliability to produce an efficient and quality AI solution.

To illustrate that such an endeavour can be done, I am exhibiting a portion of the work undertaken by my ex-colleagues at Immigration, Refugees and Citizenship Canada and Prodago, a firm hired to assist them.[8] On the drawing presented above, the guiding principles form the foundation on which the three columns of the temple rest. For example, the privacy principle is selected. The team looked at The Privacy Act and related policies and directives to break them down into specific operational practices. The team adopted a Privacy by Design approach to carry out the efforts.

Outlined by the team at outset, these practices are implemented at the deployment phase of the AI business process. This means that an AI solution cannot be implemented for production if clear responses about compliance are not provided. The exercise must be done for each principle.

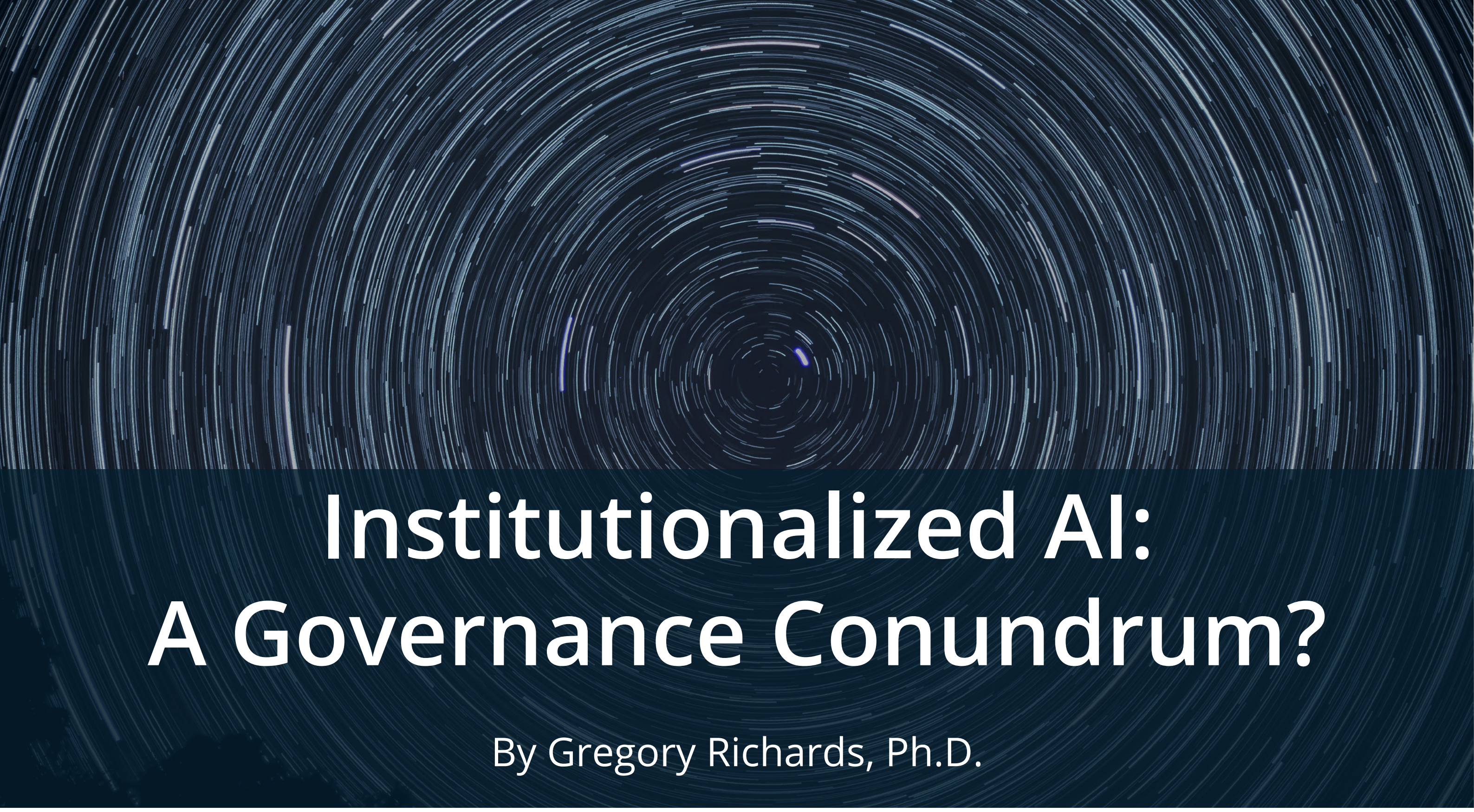
Establishing AI governance must be based on what characterizes the new technology and its own challenges. AI managers and practitioners cannot avoid the imperative of aligning their actions with key principles. In public service, conformity with laws and policies is a must. Lack of attention or care to this area will plague AI trustworthiness.

The establishment of operational practices that integrate AI procedures and methods with specific ethical (and legal) principles helps to build a sound governance framework. A bridge is built which is likely to pave the way for AI governance.

About the Author

Hubert Laferrière was the Director of the Advanced Analytics Solution Centre (A2SC) at Immigration, Refugees and Citizenship Canada. He had established the A2SC for the Department of IRCC and led a major transformative project where advanced analytics and machine learning were used to augment and automate decision-making for key business processes.

[8] Thanks to Na Guan, Jae-Jin Ryu and Wassim El-Kass, my ex-colleagues at Immigration, Refugees and Citizenship for the use of the drawings and to Mario Cantin from Prodago for his substantial contribution. Gartner just recently recognized the firm as a “2020 Cool-Vendor” in AI governance.



Institutionalized AI: A Governance Conundrum?

By Gregory Richards, Ph.D.

I am going to follow up on my colleague Hubert Laferrière's discussion on AI governance. Although it's a very complex area, I would argue that government organizations already have tools in place to enable sound governance of AI.

It's no secret that Artificial intelligence (AI) has captured the imagination of government with visions of streamlined business processes and decision-making aided by algorithms. There are, of course the usual caveats about potential bias and privacy issues related to AI. But a broader question that has so far not been addressed is this: what governance structures should you think about when institutionalizing AI—that is, when integrating AI into the day-to-day business of your organizations? Should you seek informed consent from those whose data you are using? To what degree should they be aware of what you are doing with data you collect? How would you guard against bias when it comes to decisions that affect people's work life? Although these are complex issues, the government already has tools in place to address them although a few tweaks are needed to tackle AI-related issues.

By way of context, research by McKinsey Global Institute suggests that the best approach for delivering value through AI is to integrate it into your day-to-day management processes. This notion is supported by the rationale advanced by Ajay Agrawal and his colleagues at the University of Toronto that AI is a sophisticated form of predictive analytics. Predictive analytics, like other forms of analytics, depends on data and data governance structures that have been well-established in the information systems field. But AI introduces some additional governance challenges that data governance structures did not anticipate.

Taking a broader view, we might ask ourselves, what is good governance? Many governance models exist, take for instance, The United Nations Economic and Social Commission for Asia and the Pacific that identifies the following eight characteristics:

- participatory;
- consensus oriented;
- accountable;
- transparent;
- responsive;
- effective and efficient;
- equitable and inclusive; and
- follows the rule of law.

Consider, for example, that you have integrated AI into your HR planning processes to identify which of your employees might retire or leave the organization in the next few years. To do so, you would use aggregate data to create the predictive model, but since the data are gathered from your employee base, do you need informed consent?

If you examine the eight characteristics above, the answer would be yes. Furthermore, as Hubert has mentioned elsewhere in this issue, the Berkman Klein Centre for Internet and Society at Harvard University summarized a variety of recommendations for principled AI in a report (published January 15, 2020) entitled Principled Artificial Intelligence: Mapping Consensus in Ethical Rights-based Approaches to Principles for AI. In this report (pg. 5), the authors conclude that the following three key themes are important for the use of AI:

"**Privacy.** Principles under this theme stand for the idea that AI systems should respect individuals' privacy, both in the use of data for the development of technological systems and by providing impacted people with agency over their data and decisions made with it. Privacy principles are present in 97% of documents in the dataset.

Accountability. This theme includes principles concerning the importance of mechanisms to ensure that accountability for the impacts of AI systems is appropriately distributed, and that adequate remedies are provided. Accountability principles are present in 97% of documents in the dataset.

Safety and Security. These principles express requirements that AI systems be safe, performing as intended, and also secure, resistant to being compromised by unauthorized parties. Safety and Security principles are present in 81% of documents in the dataset."

Every manager I've talked with about these issues agrees with the principles discussed above. The question is how to apply them, especially in a government context.

Fortunately, most government organizations already have the tools in place: business cases, performance measurement, and audit. These simply need to be modified to fit with an AI-driven organization.

The Treasury Board of Canada Secretariat (TBS) has published a Business Case Guide that outlines the basics: business need, options comparisons, and risk management. For an AI-driven organization, these sections would include and evaluate options that might not include the use of AI. If an AI approach is selected, then the section on risk would evaluate risks and potential costs related to data leakage, ensuring informed consent and managing security breaches. The overall cost-benefit analysis would consider total costs including those related to potential breaches and the cost of transparency.

TBS's Directive on Results provides guidance for performance measurement. In the case of AI, the process is as important as the results achieved, and thus process-level performance indicators would need to be included in the organization's results management framework.

Similarly, the TBS Policy on Internal Audit provides broad guidelines that would need to be expanded to include reviews of the operations of the AI algorithm including the "drift" associated with the algorithm over time and its potential for bias. The challenge here is the "black box" nature of most algorithms.

At the moment, AI adoption is in early phases, but as organizations move forward with integrating these tools into their day-to-day management processes, it's important to rely on established policies and frameworks to ensure sound governance. For example, research is being done on "white-box" AI systems that permit users insight into the data being used and the way the algorithms treat the data. The consideration of these types of approaches would start with the business case and flow through performance measurement to audit.

Although the integration of AI is indeed complex, if we consider it to be an advanced form of predictive analytics, we can find ways to leverage current governance tools to better institutionalize AI into the day-to-day work of managing government organizations.

About the Author

Gregory Richards, MBA,Ph.D., FCMC, is currently the Director, Executive MBA & Interim Vice-Dean, Undergraduate and Professional Graduate Programs at the Telfer School of Management. He was a visiting professor at the Western Management Development Centre in Denver, Colorado and a member of Peter Senge's Society for Organizational Learning based at MIT. His research focuses on the use of analytics to generate usable organizational knowledge.

Program Integrity, Entity Resolution and a 360-degree View of the Citizen

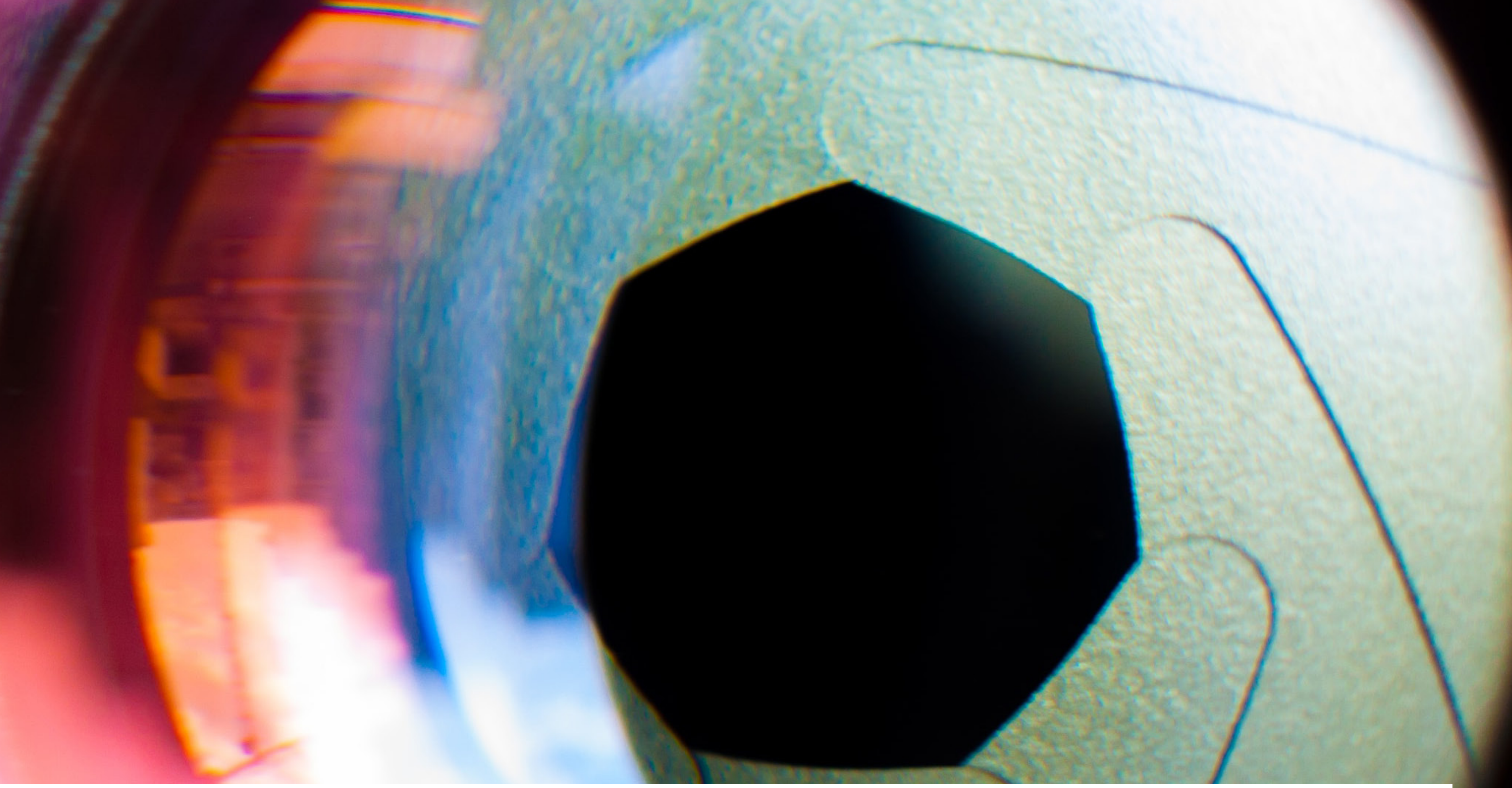
By Amanda Holden & Dan Finerty

In late January, the first Canadian was diagnosed with a novel coronavirus. Scientists had been tracking the spread of the virus, but despite previous experience with outbreaks like SARS and swine flu, the scope and depth of the impact of COVID-19 was unprecedented.

Within weeks, business closures and quarantines would throw millions of Canadians into precarious financial positions. All levels of government scrambled to put together massive social programs on an unprecedented timeline.

Speed of delivery is always a prime consideration with any government programming. But responsible use of government resources is often a competing interest—how do we ensure services are delivered only to those who qualify?

This problem is exacerbated when multiple departments and levels of government are involved in the delivery. Canada's response to the pandemic on a program delivery basis has been enviable, but not without hiccups. By June, 190,000 Canadians had to return payments made under the Canadian Emergency Response Benefit (CERB) program, often because they were unknowingly covered under another Covid-19 program—both Service Canada (through Employment Insurance) and the Canada



Revenue Agency (CRA) are administering COVID-19-related programming. In addition, the Ministry of Employment, Workforce Development and Disability Inclusion estimated that one to two percent of claims filed were fraudulent.

These issues highlight fundamental challenges in program integrity—the lack of a citizen identity consistency and the integrity of such data has created challenges, and techniques like entity resolution can help.

Entity Resolution

The goal of entity resolution seems simple: Make sure you are dealing with the person you think you're dealing with. It isn't. The nightmare scenario for entity resolution is Robert James Smith.

First, there are countless Robert Smiths. Are we dealing with Robert Smith the professor, the politician, the war hero, the plumber?

Second, in various contexts, Mr. Smith could be identified as Robert, Rob or Bob. Suppose the Robert Smith in question prefers to go by his second name. He could now also be James, Jim or even Jack. These identities could emerge variously in passport applications, driver's licenses, criminal records, employment records, professional registrations, business ownership records, property tax records... the list goes on. Some even use multiple identities intentionally, creating confusion so they can take advantage of the system.

Entity resolution is important because it is about mitigating the risks of improper or incomplete identification. And it's more than just name checking.

The 360 View

Suppose for the sake of argument we've resolved the identity of the person of interest as Robert James Smith, the plumber. (We'll go into detail about how later.) We know who we're dealing with. But do we know what we're dealing with?

The holistic, or 360-degree, view of the citizen comes from all the representations of identity, across departments and levels of government. Financial institutions have pioneered this approach to customer identity in the name of risk management. Is it safe to give this person a $50,000 loan? Consider his repayment record on credit cards. Does he have a mortgage, and thus security? Student loans? Savings accounts? Insurance?

Similarly, by pulling together Robert Smith the plumber's interactions with vehicle licensing agencies, the CRA, municipal tax departments, EI, provincial and federal service agencies, etc., we can glean a more complete view of the citizen—and the likelihood of his eligibility for government assistance, or to attempt to defraud the program. Thus, the data needs to not only be managed, but requires a governance model.

Data Architecture

The integrity of this data is paramount. For example, in a pilot project that SAS Institute participated with in Ontario, some of the children were older than their mothers. That's a rather serious data integrity issue.

At this stage of the process, we're not concerned with what the data means, only that it can be trusted. It's not about what the data says, it's about whether it's speaking the right language.

Data cleansing highlights anomalies (like children older than their mothers) and incomplete identities. It also prepares data to be integrated across systems. Fields that co-relate can be matched. One system's Surname is another system's Last Name, to cite a very simple example; phone numbers can be collected in myriad formats. Varying taxonomies can be reconciled for consumption across systems.

Citizen Protection

With data in a more manageable state, protecting the information—and the citizen—can be a more focused effort. It's a three-legged stool:

Security. Tales of customer information being lost, leaked and stolen are nothing new. Robust data security measures must be in place to protect and secure data for appropriate access and use.

Privacy. Much data in the individual datasets, for example sensitive health information, shouldn't be shared with other systems. That doesn't mean data can't be shared; insights and risks in the data can be shared in appropriate ways while still respecting privacy requirements.

Authentication. There is myriad data about Robert J. Smith in his online interaction (IP address, biometric behaviors, etc.) that can safely be used to confirm and protect the real Robert J. Smith. A consistent and deep 360-degree view of the citizen can protect the individual and government in many ways:

- With appropriate security, privacy and authentication controls, ministries can feel more comfortable sharing data across services and create a better view of what the individual needs and should have access to. For instance, an AI model might predict that the individual would benefit from a particular training program if they've recently applied for EI.
- Sadly, identity theft and falsified information are key criminal tools to defraud government programs. Accurate and a deep understanding of a citizen's interactions can help AI to predict identity theft, protecting the citizen and the government from fraud.
- Entity resolution can be complemented with techniques like network analytics, which use even broader aspects of data to connect individuals across a network. This perspective allows ministries to see who's connected to high-risk situations, creating a proactive view of interactions that can predict and prevent fraud and abuse.

Technology can be an impediment or an enabler. Rules baked into the data structure are needed to serve these three goals. With well-organized and quality data, policies, rules and practices can be automated to reduce risks.

Resolving Robert Smith

So how do we know we're dealing with Robert James Smith (the plumber)? There is some golden entity resolving data, though some of it can lead us into the woods.

The gold standard is the social insurance number (SIN). If Robert Smith and Jack Smith have the same SIN, it's almost indisputable that we're dealing with the same entity. Also, other government departments reuse SIN numbers in their own systems; for example, the Ontario Health Insurance Plan (OHIP) bases its registrations on a variation of the citizen's SIN. If these numbers don't match, we may be dealing with two different people—or someone concocting an identity.

Addresses are also compelling evidence, but not as foolproof. Perhaps Robert Smith pays property taxes on a residence at 123 Any Street. To his customers and business partners, he's Jack Smith, and applies for business relief under that name listing his shop address. This takes some untangling of the data. This is also true of phone numbers, perhaps more so—home, office, mobile phones credited to Robert, Bob and Jack could belong to the same person.

Beyond these basics, there are many other data points to help resolve Robert Smith: birth dates, other registration data, digital authentication data, and more.

How do analytics and data science streamline service delivery? Consider the case of CERB. As of September 28, the federal government had received 27,570,000 CERB applications from 8.9 million unique applicants. Entity resolution can pre-screen applicants for fast-tracking, while flagging anomalies for investigation—a triage, essentially. The 360 degree view of the citizen that cross-checks Jack Smith's employment data, tax filings, business data, could flag him for further investigation.

Doing Even More

Deep analytics and artificial intelligence offer fertile ground for even further streamlining government programming. There is such a huge body of data to train artificial intelligence to discover and predict anomalies based on behaviour in a citizen interaction, and further focus investigative resources while speeding delivery. Better citizen outcomes with efficient and effective use of resources is the mandate of every government depart. Getting control of your data and exercising deep, program-related analytics based on input from program experts can make it happen.

About the Authors

Amanda Holden is the National Executive - Fraud & Security Intelligence at SAS Canada and brings 15+ years experience in payments and 25+ years experience in financial services. Amanda is focused on finding solutions to customers' financial crimes, loss and AML problems. She is passionate about data & analytics and the role they play in reducing financial crimes in Canada.

Dan Finerty is a Data Scientist for SAS Canada specializing in Data Management. Dan is responsible for helping customers to improve their returns on their existing technology investment and to create the roadmap that enables better performance in the future.

Data Governance Prevention and Cure of “Fake” Data

By Betty Ann M. Turpin, Ph.D.

You have likely heard the expression "garbage in = garbage out". Data are facts, figures, images, etc. at their lowest level of unit that are then aggregated to conduct analysis, visualization, and reporting – the latter from which interpretation follows – thus creating information. It is this information that is used for decision-making, program design, communication, learning, etc. Hence the importance of "good" data cannot be over-stated.

Within the parameters of a governance framework organizations should have a Data Governance Committee. The authority of this committee should include oversight regarding the operationalization of data management and data integrity. Obviously, a committee should not manage the operations, but Data Stewards, for example, within each business unit can. Fast forward to how to prevent and cure fake data, which hopefully provides useful insights for Data Stewards and organizations.

Typical reasons for fake data are:

1. Deliberate misleading, often undertaken for personal /organizational gain of some sort.[1,2]
2. Lack of knowledge of data techniques and skills, such as tidy data, analysis methods, and data understanding.[3, 4]
3. Poor data credibility. Credibility[5] refers to the quality of data being believable or trustworthy, the research methodology, and data sources. It includes the following data elements: integrity, quality, reality, context, and probity. These are affected by systems and humans.
4. Poor due diligence, such as not fact checking or lack of concern about credibility. Sometimes this can be a function of "rushing" to get the information out, hence unintentional errors in the data.

So, what are some tips on how to prevent and cure fake data?

Suggestions for Prevention

Governance committees can guide the develop of policies and procedures to ensure fake data is minimized.

Data protection[6] and privacy[7] is a growing concern today. Protection is usually managed through network security protocols designed to prevent any unauthorized user, application, service or device from accessing network data such as secure file transfer protocol (SFTP), secure hypertext transfer protocol (HTTPS), firewalls, email/spam protection, secure socket layer (SSL), etc. Organizations do/can adopt security strategies, particularly important when data is shared between business units, and between government bodies such as the extensive collaboration within the Canadian Federal Government.[8] This helps keep unwanted hackers out and ensures that unauthorized users cannot change the data.

Data privacy is about ensuring that the personal information collected on individuals is secure and accurate. The latter is increasingly a concern, because as individuals attempt to ensure their privacy is maintained they may intentionally provide wrong data. This leads to erroneous information. In some situations, offering the individual an exchange of value (e.g., white paper, free access to trial software) for their information can prompt more accurate data provision.

Credible data practices go a long way toward producing and using "good" data. Data

[1] Kupferschmidt, Kai. (2018). Tide of Lies. Science.

[2] Reyes-Velarde, A. (2020) Beware the coronavirus scams: Colloidal silver, herb remedies and fake test kits. March 22.

[3] Harris, Jeanne. (2012). Data Is Useless Without the Skills to Analyze It. Harvard Business Review. September 13.

[4] Wickham, Hadley. (2014). Tidy Data. The Journal of Statistical Software, vol. 59.

[5] Credibilty is a term is often mistakenly used interchangeably with integrity or evidence. The former is an element of credibility, the latter refers to information (means information bearing on whether a belief or proposition is true/false, valid/invalid, warranted/unsupported. Evidence alone is not sufficient to determine truth, it must be interpreted.

[6] AGJ Systems & Networks. (2020). Beyond Foundational Network Security.

[7] Pritchard, K. (2017) How do marketers solve the problem of fake data? Global Marketing Allicance.

[8] Public Safety Canada. (2020). Cyber Security in the Canadian Federal Government.

credibility can be supported by regularly screening the data to ensure data records are as accurate as possible. Consistency in implementing data standards, procedures and protocols is vital.

Machine learning can expedite the data checking process by automating the often-time-consuming task.

Suggestions for a Cure

Assume your data is untrustworthy until this assumption is proven false or you can trust the source or system from which the data is derived.

Be more discerning about the data and information you are using or relying. Look to trusted sources.

Be critical of the data you use or receive. Check the source, who else is reporting on this data or information, and be sure the data is credible. Use common sense, if it "looks" odd or seems unlikely, this is often a good clue as the potentiality of fake data – investigate and do not hesitate to ask the tough questions.

Theories provide testable hypotheses on which to assess the information produced by the data. Where they do not formally exist, only those with alternative explanations can challenge the data facts or information. Both forces enable us to look deeper into the data to look for inadequacies or inaccuracies. Artificial intelligence modelling using advanced data analytics can support this depth probing too, by specifying data rules and conditions.

Data, as a qualitative and quantitative unit by itself is not meaningful, until it is analyzed and interpreted, but even then it must usually be compared to something else, or historical data of the same units.

The more systematic and prior knowledge organizations have regarding the data in question, the better. This will enable data comparison to historical data or similar data.

Upskill by ensuring all your employees have a good understanding of data and the skills required to use data credibly. This means building a data culture within the organization. In today's world, this is an imperative.

In closing, for a myriad of reasons, "fake" data is a growing concern for organizations, businesses, and citizens globally. This paper has not dwelled on the consequences or treatment of fake data or those who generate it. But I do leave you with a few questions: Should the consequences for creating and/or presenting FAKE data be regulated? What consequences can organizations impose in a timely, equitable, and legal manner?

About the Author

Betty Ann M. Turpin, Ph.D., C.E., President of Turpin Consultants Inc., is a freelance management consultant, practicing for over 25 years, has also worked in the federal government, in healthcare institutions, and as a university lecturer.

Her career focus is performance measurement, data analytics, evaluation, and research. She is a certified evaluator and coach.

By Tara Holland

Governing Analytics to Promote Diversity and Inclusion

How important is it to consider diversity & inclusion when it comes to analytics? How important is diversity and inclusion in society? The charter of rights states that "Every individual is equal before and under the law and has the right to the equal protection and equal benefit of the law ..., without discrimination based on race, national or ethnic origin, colour, religion, sex, age or mental or physical disability". As governments modernize the delivery of services to their citizens through digital transformation efforts, analytics will be at the heart of the decision making.

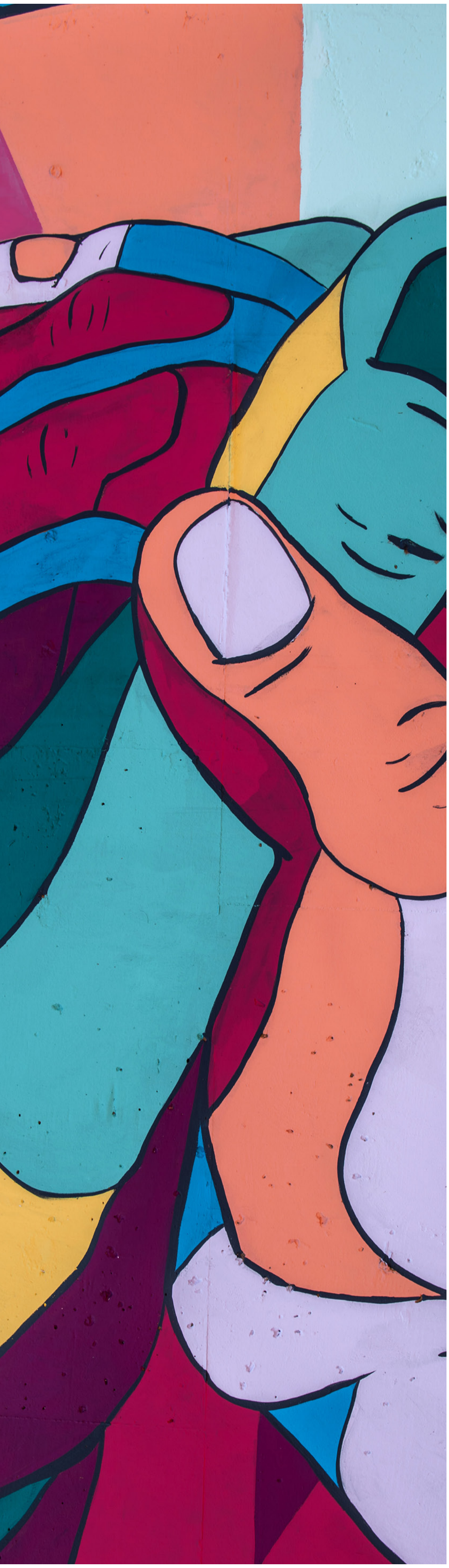
Decision-making about how the digital solutions are designed and using artificial intelligence, more and more of the day-to-day decisions in government will also be driven by analytics. The protections offered by the charter of rights is not a standard to live by but more so represents the minimum requirements that need to be met, and despite these clear instructions from the charter there is still overwhelming systemic inequality. From a moral and ethical perspective, we owe it to society to consider and include the perspectives of all individuals as we build analytic solutions, and to treat individual differences as assets rather than outliers.

The Diversity & Inclusion Imperative

If not for the betterment of society or ethical and moral reason, how about for a sound business decision? Research has consistently shown the business benefits of diversity. An article from McKinsey shows that "Companies in the top quartile for gender diversity outperform their competitors by 15% and those in the top quartile for ethnic diversity outperform their competitors by 35%"[1]. These numbers cannot be ignored and quantify the importance of diversity in the workplace; it is in everyone's best interest that policies and bodies are put in place to move in this direction. Furthermore, Canadian laws are requiring companies to be transparent and answer for a lack of diversity as summarized in this Library of Congress Article "Canada: Higher Standards Set for Workplace Diversity"[2] and the federal government is

[1] Vivian Hunt, Dennis Layton, and Sara Prince (2015), Why diversity matters, McKinsey Company

[2] Tariq Ahmad, 2019, Canada: Higher Standards Set for Workplace Diversity



ensuring this with Bill C-25[3] in effect as of January 1, 2020. Government agencies must hold themselves to the same or higher standard.

A Starting Point for Analytics Governance

How do we approach diversity and inclusion in the field of analytics? We start by challenging the status quo. Are we solving the right problem? Do we have the right data? What is the limitation of our findings? Are we having the right conversations? When it comes to the governance of analytics, the processes and decision makers must assume the answer to these questions is 'no' and the next step after acknowledging and accepting this assumption is to address these issues. We must question things we assumed to be fact; things that may have seemed to be absolute in the past.

Unique Challenges: Culture, Methods, Data

The field of analytics must address and overcome some inherent challenges to build diverse and inclusive solutions. The disciplines that lead individuals toward a career in analytics and the skills, training and communication tools are built on a premise of objective quantification. This creates a culture in the field of analytics which is based on "finding the right answer" and "trusting the data", which in a perfect world, should be within reason. However, when problems are complex, the parameters set by analytical techniques, laws, regulations and cultural norms are often too limited or turn out to be exclusive, therefore missing certain perspectives and considerations. This can have serious impacts on the results. During a recent webinar on "Building a Diverse and Inclusive Government with Analytics", one of the panelists shared an example of how "A survey conducted came back with strange results because the questions though translated properly had different than intended meaning in varying cultures which in the end resulted in publishing misleading results and ultimately the retraction acknowledging the data was incorrect". Even with the best intention, mistakes and missteps will happen along the way, but we must learn from these mistakes and do our best mitigate them moving forward.

The field of analytics serves the purpose of simplification, taking vast amounts of data and seeking to categorize and define, and has developed methods to do just that. When it comes to diversity, one of the most miscommunicated, misused and misclassified terms is the "average". When it comes to human experiences there is no such thing as an average human and creating a target or measurement to this end should not be an analytic endeavour. Historically it has also been shown that systems in our society are built around

[3] Parliament of Canada, 2018, BILL C-25

the concept of "average". This has proven to marginalize many groups as discussed in this article[4] from Harvard Graduate School of Education. As we are in the middle of a digital revolution, similar in scale to the industrial revolution, we cannot afford to make the same mistakes and further marginalize the groups already suffering.

How do we Address the Challenges?

Some recognized strategies to proactively address these challenges are having diverse teams, employing a diverse set of analytic techniques, and a governance structure that intentionally looks for gaps.

Diversity is a strength and a strategy of trying to get individuals to fit a certain mold diminishes that strength. The traditional method of hiring for specific technical skillsets needs to be examined. Skills can be taught, but there are qualities, perspectives and characteristics that can only be acquired through diverse hiring. Lack of diversity in a team was the acknowledged factor in a well-publicized misstep by Google[5]. After launching the YouTube for iOS, the functionality did not work for left-handed people simply because there was no one on the development or testing team who was left-handed. This example is amusing but enlightening. Given the multiple facets of diversity and the fact that most aspects of diversity are not as visible as being left-handed, hiring and staffing practices must intentionally seek out diversity. Attracting different mindsets and disciplines to the field of analytics and broadening the perspectives of teams through continued learning is a key to limiting mistakes. We should use analytics to access what talents we have and to maximize those skillsets rather than to force individuals and teams to fit an "established norm".

Analytics teams that employ a broad set of disciplines and techniques are also more likely to spot the gaps and unintended bias. This may take a turn to different strategies such as mixed methods and greater emphasis on the qualitative and interpretative side of data. Sophisticated machine learning methods are capable of modelling complex scenarios, but we as decision makers and governors of analytics must provide these algorithms with complete and inclusive data. In order to do so we must have a diverse team, mindset and culture.

Individual analytics professionals must be willing to have uncomfortable discussions that question the validity of our assumptions and results. A governance body that has D/I built into its terms of reference will have the authority and responsibility to ensure these discussions take place.....If we continue to address diversity and inclusion gaps head on, there will come a point where these become a key part of our analytic strategy and will no longer feel uncomfortable. We must acknowledge that our perceptions and assumptions about specific groups is typically wrong. If we put in place governance practices that challenge our teams, if we learn basic skills to recognize and embrace diversity, and if we build on that knowledge over time, the solutions we build will be more inclusive. To make a difference it is important to feel some discomfort. If, as government leaders and analytics professionals, we aren't feeling the growing pains or making some mistakes, it is likely that nothing is changing, and we should re-examine our approach.

About the Author

Tara Holland has been with SAS for over 20 years and has specialized in Analytics, Artificial Intelligence and Visualization solutions. Prior to joining SAS, Tara led the Data Mining team at Canada Revenue Agency and lived the real-life challenges faced by organizations wanting to be more data and analytics driven. A part of a global team, Tara is able to bring the best solutions from around the world to government customers, prospects and partners.

[4] Lory Hough, 2015, Beyond Average

[5] Sean Buckley, 2014, Unconscious bias is why we don't have a diverse workplace, says Google

ANALYTICS IN GOVERNMENT QUARTERLY