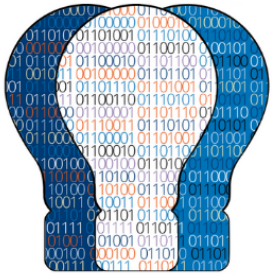
ADVISORY • ENGAGEMENT • TRAINING

THINKDIGITAL

DIGITAL TRANSFORMATION FOR PUBLIC GOOD



Generative AI in the Public Sector: Review of Existing Employee Guidelines

Environmental Scan & Analysis

Think Digital | May 2024

Table of Contents

Preface

This document was created by Think Digital and commissioned by The Regional Municipality of York with the intention of contributing to what is an increasing body of knowledge on the use of Generative AI in public sector organizations. We thank York Region for their support of this important work and their willingness to share it publicly for the benefit of others.

All content in this document is based on our gathering and analysis of original source material. It is important to note that for this environmental scan, source materials were limited to whatever data was publicly available at the time of writing, and was, for the most part, published by the relevant implementing organizations. Because of its limited scope, and the emergent nature of Generative AI technologies, this scan does not offer a complete representation of existing guidance either published or still underway. Readers are encouraged to consult the source material for additional details on the specifics of each case study, which can be found in the included links throughout the document.

Introduction

This environmental scan of Generative AI guidelines showcases efforts across jurisdictions to navigate the promising yet complex landscape of Artificial Intelligence (AI) technologies. The featured guidelines all aim to help their readers better understand Generative AI, guide project teams building Generative AI-based solutions, and detail how to use and govern Generative AI safely and responsibly in their respective jurisdictions and organizations.

For context, Generative AI is a type of advanced AI that can "generate" new content by learning from the patterns in their training data. Large Language Models (LLMs), first widely popularized by OpenAI's ChatGPT, are an increasingly common type of Generative AI tool that have become widely available for free or low-cost on a consumer basis. Generative AI systems, tools, and models can often decipher and generate human language, as well as other forms of content like images, video, and audio. They are trained on massive amounts of data from the internet, books, images, and other sources from which they can learn patterns in language, visual data, and other modalities. Their knowledge can subsequently be leveraged to complete tasks like answering questions, generating images/video/audio, analyzing data, and generating or reviewing computer code. Generative AI has already proved useful across a wide range of industries, and recent developments in the field have the potential to unlock significant productivity benefits, and to change the way we approach content creation for both industry and the public sector alike.

The thrust of this environmental scan suggests that while the potential of Generative AI to transform public service delivery is widely recognized, its deployment in public sector contexts must be carefully managed to balance innovation with socio-ethical considerations, security concerns, and legal compliance. The resounding attitude of the guidelines examined is to proceed – but with caution, and with emphasis on ensuring that the benefits are harnessed by the public sector without compromising ethical standards and societal values, or risking non-compliance to laws and legislation governing their jurisdictions.

Included in the scan below are nine case studies of internal governance approaches for Generative AI adoption that are being used by various government bodies, public sector institutions, and municipalities to govern their own use cases. We have attempted to summarize each of the case studies to present their key characteristics, and to indicate what is unique about each of them as well as common elements identified across the

approaches from different jurisdictions. For each case, we have included a highlight box that zeroes-in on key points, followed by their longer summarization.

The nine case studies covered in this scan are listed below, and have been grouped according to their prospective audience and intended scale:

Municipal/Local

1. *Generative AI Guidelines*, The City of San Jose.
2. *GenAI Guidance for Local Authorities*, London Office of Technology and Innovation (LOTI).
3. *Interim Guidelines for Using Generative AI*, City of Boston.

Canadian

4. *Guide on the use of generative artificial intelligence*, Treasury Board of Canada Secretariat (TBS).
5. *Generative artificial intelligence (AI) - ITSAP.00.041*, Canadian Centre for Cyber Security (CCCS).
6. *Principles for responsible, trustworthy and privacy-protective Generative AI technologies*, Office of the Privacy Commissioner of Canada (OPC).

International

7. *Initial advice on Generative Artificial Intelligence in the public service,* New Zealand.
8. *Guidelines For Staff on The Use of Online Available Generative Artificial Intelligence Tools*, European Commission.
9. *Gen AI framework for HM Government*, Central Digital and Data Office, United Kingdom (CDDO).

Overall, for the cases researched, a principle-based approach emerges as the foundational strategy that can best align public sector Generative AI deployment with ethical standards and societal values. The universal emphasis of the guidelines on ethical and responsible use, particularly the focus on safeguarding vulnerable populations and mitigating biases, reflects consensus around adopting Generative AI in a way that supports social responsibility and equity.

In our analysis, we have extracted 12 central themes from the case studies and highlight four additional themes as still emergent governance considerations for Generative AI usage in the public sector. These themes are captured in the table below:

Central Themes	Other Emergent Themes
• Understanding the Risks and Benefits • Principle-based Guidelines • Interim, Adaptive Guidelines • Ethical and Responsible Use • Data Security and Privacy • Accuracy and Accountability • Testing • Training and Education • Oversight and Governance • Legal Compliance • Transparency and Disclosure • Risk Mitigation	• Use of Generative AI for Coding • Sandboxing and Experimental Spaces for Testing • Engagement with Indigenous Communities • Environmental Impact Considerations

Mapping Generative AI Guidelines to Central Themes

	Generative AI Guideline	Understanding Risks & Benefits	Principle-based Guidance	Interim, Adaptive Guidelines	Ethical & Responsible Use	Address Security & Privacy Risks	Testing	Training & Education	Oversight & Governance	Legal Compliance	Risk Mitigation
Municipal/ Local	City of San Jose	x	x	x	x	x	x			x	x
	London Office of Technology and Innovation	x	x		x	x	x	x	x	x	x
	City of Boston	x	x	x	x	x					x
Canadian	Treasury Board Secretariat (TBS)	x	x	x	x	x	x	x	x	x	x
	Canadian Centre for Cyber Security	x	x		x	x		x	x		x
	Office of the Privacy Commissioner (OPC)	x	x	x	x	x	x		x	x	x
International	New Zealand	x	x	x	x	x	x		x		x
	HM Government (UK)	x	x	x	x	x	x		x	x	x
	European Commission	x	x	x	x	x					x

Central Themes and Analysis

Taken together, the nine case studies offer input and experience on a broad range of governance considerations around government employee usage of Generative AI, which are summarized below into twelve central themes. Officials, staff, and senior decision-makers can better navigate the complexities of Generative AI by adhering to the principles laid out in the guidelines below. Moreover, by incorporating existing best practices and governance considerations into their AI ecosystems, public sector employees should be able to integrate these technologies more confidently into their workflows to enhance services and operations, while upholding the highest standards of responsibility, transparency, and public trust.

In the subsequent section, four noteworthy approaches or considerations that were unique to a few or only one of the case studies are flagged as “other emerging themes” and discussed briefly.

Understanding the Risks and Benefits

All guidelines articulate Generative AI’s ability to revolutionize service delivery, enhance operational efficiency, and foster innovation. However, enthusiasm is always tempered by an acute awareness of the associated risks like data privacy breaches, misinformation, and the amplification of biases. This balanced viewpoint encourages organizations to approach Generative AI adoption with both optimism and caution, ensuring that the benefits are harnessed without compromising ethical standards or societal values.

Principle-based Governance

There is a unanimous call for principle-based governance of Generative AI. A common thread is the reliance on core principles (fairness, accountability, security, transparency) to guide the ethical deployment of Generative AI. Values-driven approaches to technology adoption are popular because they are adaptable and enduring, which is particularly useful considering the complex and emergent nature of Generative AI technologies, and the increasing demand for foundational Generative AI strategies across governments and jurisdictions.

Interim, Adaptive Guidelines

Due to the dynamic nature of the technology, where rapid advancements and ongoing research are continually shaping our understanding and attitudes, many of the guidelines state that they are "living" documents that will evolve and be revised over time. Moreover, guidelines from the Office of the Privacy Commissioner and the Canadian Centre for Cyber Security recommend that organizations develop their own internal policies beyond their guidance for appropriate use and governance that are context specific. San Jose and Boston suggest their guidelines should (eventually) be replaced with more substantial policies and standards, and that their guidelines are an initial part of an "evolving governance structure around responsible AI usage."[1]

Ethical and Responsible Use

The guidelines collectively highlight the importance of deploying Generative AI in a manner that does not perpetuate or exacerbate societal inequities. Special attention to protecting vulnerable populations and mitigating biases in Generative AI outputs reflects a broader commitment to social responsibility and equity. This emphasis on ethical use is a call to action for organizations at large to prioritize the societal impact of Generative AI applications, ensuring that government use of Generative AI benefits the public good.

Data Security and Privacy

Unanimously, the guidelines prioritize protecting sensitive and personal information, reflecting the critical importance of data security in the age of Generative AI. They all recommend, to the point of prohibiting, inputting sensitive and/or personal information (internal and client/external data) in publicly available Generative AI systems. Other guidelines (Treasury Board of Canada Secretariat, New Zealand, UK Central Digital and Data Office) clarify that public servants may input personal data into systems that are controlled and configured by the relevant government agency, and when appropriate privacy and security controls are in place. San Jose's approach on this is to recommend public servants not use any information, including both the inputs to and outputs of official decisions, that isn't ready for public disclosure.

Recommendations for robust data protection measures and adherence to privacy laws are also common across the guidelines. In addition to proactive security practices to protect

[1] https://www.sanjoseca.gov/home/showpublisheddocument/100095/638314083307070000, 5.

against data-related risks, many of the guidelines encourage individual user vigilance around the data they use with Generative AI systems to ensure safe use in general, while AI-specific regulations continue to catch up and evolve.

Accuracy and Accountability

The guidelines stress the necessity of checking the accuracy of AI-generated outputs and maintaining human oversight throughout the AI deployment process. This theme underscores the importance of not only verifying the factual correctness of Generative AI outputs, but also *retaining* human judgment and accountability in decision-making processes influenced by AI. While reviewing and fact checking all Generative AI outputs is the commonest recommendation for meaningful control and human intervention, its ubiquity as a recommendation reminds us that user overreliance on AI-generated outputs is a dynamic risk factor with these technologies specifically. Moreover, several of the guidelines remind us that these technologies should augment and support, not replace, human expertise and ethical judgment. And finally, in the view of the OPC, accountability is established at the legal level for all parties/stakeholders, with "compliance with privacy legislation and principles."[2]

Testing

Several of the guidelines recommend setting up a testing process, and that Generative AI solutions or systems and their results be continually tested throughout its use. The TBS, OPC, and LOTI suggest more in-depth testing methods to identify and mitigate system vulnerabilities like penetration testing, adversarial dataset testing, and "red teaming".

All of the guidelines state, in one form or another, that Generative AI tools are not guaranteed to be accurate because their outputs are generated based on the most likely pattern of text or images relative to the input and their training data, and this doesn't directly take into account what the text or image means. They are designed to produce highly plausible and coherent results based on the data that they have access to and processed at a given time. This means that they can, and do, make errors.

In addition to recommended techniques for ensuring reliable results, many of the guidelines recommend putting methods in place to test these results. Project teams need to understand how to monitor and mitigate common Generative AI output errors like drift,

[2] https://www.priv.gc.ca/en/privacy-topics/technology/artificial-intelligence/gd_principles_ai/

bias, and hallucinations. It is recommended by many guidelines to have robust testing and monitoring processes in place to catch these problems. Various and common testing methods include:

- Regular and iterative system testing before and during operation to assess the functionality and effectiveness of the system and its results.
- Validating the model's outputs against human judgment or ground truth and obtaining user feedback whenever possible.
- Closely reviewing the outcomes of the technical decisions made, the system and decision infrastructure, running costs, and environmental impact. This information should be used to continually iterate your solution.
- Conducting independent, third-party audits for assessing Generative AI systems, including but not limited to risk, impact, and privacy assessments.

Training and Education

All of the guidelines featured recognize that Generative AI, and our understanding of it, is rapidly advancing and continually evolving. Therefore, many of them advocate for continuous education and skill development for all stakeholders involved in AI processes and deployment. For example, the first recommendation of the LOTI guidelines for local authorities looking to enable successful Generative AI use, is to "develop staff training to ensure a baseline understanding of GenAI to boost productivity and minimize potential risks of inappropriate use by staff."[3] Training is highlighted as essential for equipping individuals with the knowledge and skills required to navigate the ethical, legal, and operational complexities of Generative AI, while ongoing education of all staff can foster an organizational culture of informed and responsible AI use.

Oversight and Governance

The call for establishing governance frameworks and oversight bodies, such as AI Working Groups, emphasizes the need for strategic oversight and accountability in Generative AI deployment. This theme highlights the role of governance in ensuring that AI technologies are deployed in alignment with organizational goals and ethical standards, promoting transparency and fostering public trust in government AI applications.

[3] https://drive.google.com/file/d/1kHfm5KTjaRHeLcZrpFjIAT83X11g8BRq/view?usp=sharing

Legal Compliance

Amidst the evolving regulatory landscape of AI governance, the guidelines underscore the importance of legal compliance and readiness for future legislative developments. Navigating the legal complexities associated with Generative AI use typically looks like remaining compliant with existing laws while being prepared to adapt to new regulatory requirements. While many of the guidelines recommend including legal counsel as a stakeholder in project teams, the OPC clarifies that the legality of certain practices, "through investigative or legal findings," have not yet been made in the context of Generative AI. There is, therefore, an operative lag between the law (Canadian) and Generative AI use that organizations should consider when adopting these technologies. For the OPC, this gap can be mitigated with adherence to existing data privacy law and the anticipation of certain "no-go zones" such as using AI-generated content for malicious purposes, and the use of chatbots to coerce people into divulging personal information.[4]

Transparency and Disclosure

Advocating for openness about the use of Generative AI, including the disclosure of AI's role in content creation and decision-making processes, the guidelines promote transparency as a cornerstone of ethical AI deployment. Transparency in this context goes beyond system transparency and explainability to the need for clear communication with all stakeholders around when Generative AI is used, how it is used, and why it was used, to build trust and foster informed engagement with these newer-to-market AI technologies. Sufficient documentation can boost this kind of transparency, with San Jose mandating that their staff document when they use Generative AI to support their work by filling out an online form that allows for central tracking of usage, while the UK's CDDO recommends creating an AI/Machine Learning inventory to catalogue all Generative AI use including inputs, prompts, and outputs collected within an organization.

Risk Mitigation

The repeated emphasis on risk mitigation strategies across the guidelines demonstrates the importance of proactively addressing the potential negative impacts of Generative AI. Risk mitigation commonly includes implementing security measures, ethical reviews, and conducting privacy impact assessments to manage the risks associated with AI

[4] https://www.priv.gc.ca/en/privacy-topics/technology/artificial-intelligence/gd_principles_ai/

technologies effectively. By prioritizing risk mitigation, organizations can navigate the challenges of Generative AI deployment while safeguarding public interests and societal values. For more detailed reference, the guidelines from the UK's CDDO and San Jose provide useful, risk-based scenarios or case studies of some of the most common vulnerabilities, putting them in context of how they could apply to Generative AI applications in government.

Other Emerging Themes

Use of Generative AI for Coding

The guidelines from New Zealand, Boston, and San Jose all include small sections addressing the specific risk of using Generative AI to generate code, cautioning against deploying AI-generated code without thorough review due to potential vulnerabilities. As per San Jose's guide: "Code generated by an AI may be outdated, copyrighted, have identified vulnerabilities, or rely on other code that no longer works."[5] The UK CDDO's extensive guidance sporadically includes more technical/practical recommendations for developers respecting Generative AI tools for coding, and around the use of code assistance tools.

Sandboxing and Experimental Spaces for Testing

New Zealand and LOTI guidelines uniquely advocate for creating "sandboxes" or safe environments for testing and experimenting with Generative AI, emphasizing the importance of beginning with, and learning from, low-risk deployments. San Jose also recommends to "test out [a Generative AI] system in a risk-free environment."[6] Municipal authorities could benefit from virtual or closed testing environments where staff can experiment with different Generative AI functions, using dummy data sets and fully synthesized data, to understand its impacts fully and ensure its reliable deployment in future live settings.

---

[5] https://www.sanjoseca.gov/home/showpublisheddocument/100095/638314083307070000, 24.
[6] Ibid., 25.

The TBS and New Zealand guidelines also recommend exploratory, stratified approaches to Generative AI adoption. By this, they mean organizations should begin with safe or low-risk applications to learn and trial safe and ethical use of these tools, gauging data quality and testing outputs before adding complexity or deploying them in real world scenarios.

Engagement with Indigenous Stakeholders

Unique to New Zealand's guidelines is the emphasis on engaging with Iwi Māori and respecting Te Tiriti o Waitangi principles, ensuring that Generative AI use considers indigenous perspectives and rights.

Environmental Impact Considerations

Reflecting growing concerns around sustainability and the energy needed for AI computation and data storage, the TBS and CDDO guidelines specifically address the environmental costs associated with running large AI models. Both guidelines stress the importance of sustainable Generative AI solutions, and recommend teams check the environmental credentials of potential model providers, using net-zero or carbon-neutral data centers, and conducting life-cycle analysis to assess the carbon footprint of AI systems to mitigate their climate impact.

Environmental Scan

Municipal/Local Case Studies

1. *Generative AI Guidelines*, The City of San Jose

Version Accessed: September 23rd, 2023
Link:
https://www.sanjoseca.gov/home/showpublisheddocument/100095/638314083307070000

Case Highlights: San Jose
• Guidelines are to be read as an initial part of an evolving, more robust AI governance strategy for the City. • Encourages participation in "AI Working Groups" and the creation of advisory task forces to further develop organization-wide AI strategy. • Recommends citation and reporting each use of Generative AI to achieve optimal transparency.

The City of San Jose's guidelines on the use of Generative AI articulate a cautious yet forward-thinking approach towards integrating these technologies within municipal operations. The City recognizes the dual-edged nature of Generative AI, weighing the efficiency gains in public service delivery against the risks of misinformation, privacy breaches, and cyber-attacks, and so aims to craft in their guidelines a governance structure that balances innovation with responsibility. Moreover, the guidelines serve as an integral component of San Jose's efforts to develop broader policies around AI usage, marking a step towards creating their own responsible AI governance framework.

Notably, the guidelines are applicable to all City staff, contractors, and volunteers in their professional capacities, but does not extend to personal or non-City related business uses. The principles laid out in San Jose's guidelines apply to the use of Generative AI in the

above capacities, with the caveat that departments may provide additional rules. The guidance is boiled down to six core principles, which are summarized as follows:

Privacy: Ensuring that any information submitted to Generative AI tools is suitable for public disclosure, reflecting a bottom-line, proactive stance on safeguarding privacy.

Accuracy: Maintaining the City's credibility by verifying the factual accuracy of Generative AI outputs, to maintain trust and reliability inside and outside the organization.

Transparency: Mandating clear disclosure when Generative AI is utilized, fostering openness and honesty in public service delivery. This includes citing that these tools were used when "a substantial portion of the content used in the final version comes from the Generative AI." Moreover, "*all images and videos* must cite any AI used in their creation, even if the images are substantially edited after generation."[7]

Equity: Acknowledge and address potential biases inherent in Generative AI that reflect existing societal disparities, to promote fairness and inclusivity.

Accountability: "Use Generative AI with a healthy dose of skepticism" – Users should engage critically with Generative AI tools and will be held accountable for the content produced by them.

Beneficial Use: Integrate Generative AI in ways that enhance public services, efficiency, and justice. Use of these tools should be limited to situations where they can "make services better, more just, and more efficient."

To put these principles into action, the guidelines offer practical steps for safe and responsible Generative AI use, including:

- Adherence to public records laws (the California Public Records Act), which emphasizes the public nature of Generative AI interactions in the municipal context. Prompts, outputs, and other information used in relation to these tools may be released publicly.
- Creation of dedicated accounts for City-related Generative AI use, ensuring a clear separation between public and personal data, and enhancing data security and record-keeping.

[7] Our emphasis.

- Compliance with terms and conditions of Generative AI services used, which may not have official agreements with the City, placing the onus of legal compliance on individual users.
- Opting out of data collection by Generative AI services wherever possible, to minimize data privacy risks.
- Verifying the copyright compliance of Generative AI-generated content to avoid legal infringements.
- Encouraging participation in "AI Working Groups" and the creation of advisory task forces to share knowledge, develop best practices, and shape the City's AI strategy.

Respecting the last practical recommendation, the City has three such groups: the City AI working group, comprised of City staff who discuss AI policy, use cases, and guidelines; the Digital Privacy Advisory Taskforce, comprised of external experts around digital privacy and AI who advise and recommend on the City's privacy practices, including responsible AI; and the GovAI Coalition, which represents the City's collaboration with other government agencies across the country on things like AI governance, vendor accountability, and knowledge-sharing, to ensure that the AI systems they use serve their communities.

It is worth noting that San Jose's guidelines also include a section on how to cite Generative AI use. As mentioned above, when "a substantial portion of the content used in the final version comes from Generative AI", the City mandates that users cite their use. What the City defines as "substantial" will be further defined in future working group discussion. In addition, the City expresses their need to understand how Generative AI tools are being used by city staff in their work. Therefore, "to track usage in aggregate," the guidelines mandate that staff document when they use Generative AI to support their work by filling out a Microsoft form (which is publicly accessible here: https://forms.office.com/g/3Znipym4k5). This level of reporting sets a clear standard for documentation as a mechanism of transparency and explainability around AI usage. Notably, staff do not need to wait for a response after filling out the form to use Generative AI, unless required by their department or manager.

The guidelines also delve into risk assessment for Generative AI use cases, which for the authors is determined by two factors: information breach risks and adverse impact risks (essentially, that is "data" and "context" respectively). In an appendix, the City breaks these two risk factors down into three categories: Mid-risk, High-risk, and Prohibited risk. For risk of information breach, the three categories refer to the type of data that would be used

when interacting with Generative AI; and for the risk of adverse impact, the categories refer to the use case or scenario in which these tools would be used.

The City has a bottom-line approach to Generative AI use, which is summarized by "anything that would not be released or shared with the public should not be input into the AI."[8] Operations can include Mid to High-level risk in either case, however High-risk scenarios require special consideration, including careful review of AI outputs for tone and accuracy consistent with the City and its values, citing verifiable sources, and avoiding the submission of sensitive or confidential information into Generative AI systems.

[8] https://www.sanjoseca.gov/home/showpublisheddocument/100095/638314083307070000, 12.

2. *GenAI Guidance for Local Authorities*, London Office of Technology and Innovation (LOTI) and Faculty

Version Accessed: July 2023
Link: https://drive.google.com/file/d/1kHfm5KTjaRHeLcZrpFjIAT83X11g8BRq/view

Case Highlights: LOTI
• Offers "hard-rules" as bottom-line requirements for risk mitigation when using Generative AI. • Includes a short-list of detailed use-cases and a long-list of other possible use cases for Generative AI at the local authority level and frames them in terms of three guiding principles. • Includes a helpful list of questions staff can ask when identifying use cases, and to distinguish when a more tailored solution than "off the shelf" options may be needed.

The London Office of Technology and Innovation (LOTI) has issued guidance to steer the ethical and responsible deployment of Generative AI within local government operations. LOTI teamed up with Faculty, an official partner of OpenAI, to deliver guidance for local authority employees to use as they explore and navigate Generative AI adoption. The LOTI guidelines are broken down into several standalone pieces, each with the purpose of providing local authorities a baseline for the different stages of Generative AI adoption, including guidance for responsible use, a technical overview, use case development, governance, and a list of key considerations and activities.

A significant portion of LOTI's guidance is dedicated to articulating the risks associated with the use of Generative AI including inaccuracies, biases, and privacy concerns. To mitigate these risks, the LOTI guidelines recommend several "hard rules" that serve as a bottom-line or non-negotiables for the use of these tools in local public service, all of which align with what is suggested by the other guidelines in this scan. For example, one of LOTI's 'hard rules' is prohibiting the input of personally identifiable information into Generative AI tools, personal data about local residents/citizens even if it is not personally identifiable, and commercially sensitive local authority data. Other hard-rules include not using a personal account for work related purposes (i.e., shadow IT); verifying AI-generated content before

use; following relevant legal and regulatory requirements including policies set by local authorities; and disclosing Generative AI use internally with other staff and externally with citizens to ensure public trust, especially when directly quoting or using a "significant portion" of a tool's output, or using an output to meaningfully inform a decision.

To harness Generative AI's potential responsibly, LOTI recommends a structured approach for local authorities, focusing on:

Education and Training: Providing staff with the knowledge to utilize Generative AI tools effectively, emphasizing ethical use, data security, and privacy considerations.

Infrastructure Readiness: Preparing the digital and IT infrastructure to support Generative AI applications and incorporating necessary safeguards for responsible use.

Governance and Oversight: Establishing governance frameworks to oversee Generative AI use, ensuring alignment with ethical principles, legal compliance, and organizational goals.

LOTI's guidelines also highlight several practical applications and use cases of Generative AI for local authorities that demonstrate the potential for these tools to improve administrative efficiency, service delivery, and creative problem-solving. Each use case is evaluated against LOTI's principles for responsible Generative AI use that are, when compared to the other case studies in this scan, seemingly broad and all encompassing. Their three guiding principles are:

1. Use of Generative AI must comply with applicable legal and regulatory requirements.
2. Use of Generative AI must protect data privacy and security.
3. Employees and end users must always be accountable.

When thinking about potential Generative AI solutions, the LOTI guidelines include a helpful list of questions local authority staff can ask when identifying use cases, which can help distinguish when a more tailored solution may be needed than what is offered by "off the shelf" applications such as ChatGPT and MS Copilot. The final two sections of the guidance include three in-depth use cases that demonstrate the nuances of developing Generative AI solutions tailored for local government, followed by a longlist of more potential use cases. The guidelines are careful to frame all current use cases with reference to future technological advancements and regulatory changes. LOTI therefore advocates

for local authorities to remain agile, updating policies and practices in response to new insights, evolving technological capabilities, and societal expectations regarding ethical AI use and Generative AI.

3. *Interim Guidelines for Using Generative AI*, City of Boston

Version Accessed: May 18th, 2023
Link: https://www.boston.gov/sites/default/files/file/2023/05/Guidelines-for-Using-Generative-AI-2023.pdf

Case Highlights: Boston
• Emphasizes human accountability/empowerment, and that Generative AI tools are meant to support human judgment, not replace it. • Guiding principle of inclusion extends to the use of Generative AI to 'repair damage done' to marginalized peoples. • Includes the caveat that guidelines should be eventually replaced by policies and standards. • Do's and Don'ts section.

The City of Boston's *Interim Guidelines for Using Generative AI* provides guardrails for City staff to leverage Generative AI technologies responsibly. Notably, from the top, the guidelines are careful to stress that these tools are not actual intelligence in the human sense; rather, they are statistical models that use complex probabilities to predict what language, text, or image satisfies inputs. With this as its backdrop, one of the fundamental purposes of Boston's guidelines is to instill a sense of empowerment among employees regarding the use of Generative AI. By emphasizing that AI is a tool and not a replacement for human judgment, the guidelines underscore the importance of maintaining organizational and individual accountability for the outcomes generated by AI systems. This empowerment is not just about using AI tools but also about exercising judgment, ensuring that the benefits of AI are realized while mitigating potential risks and negative impacts.

To do this, the guidelines offer a principles-based approach to Generative AI governance. The principles outlined in the document revolve around user empowerment, inclusion and respect, transparency and accountability, innovation and risk management, privacy and security, and public purpose. Using a principles-based approach allows Boston to lay the overall groundwork for what is commonly referred to as "responsible" use of Generative AI, which from their perspective would enhance services for residents while upholding ethical

standards and the principles above. It is worth noting that Boston's principles of inclusion and respect involves mitigating discrimination and bias, and couches "the use and development of AI" at their organization in work "that *repairs damage done* to racial and ethnic minorities, people of all genders and sexual orientations, people of all ages, people with disabilities, and others."[9]

Boston's principles offer a light-touch form of "guidance as governance", insofar as they operate like a code of conduct without any enforcement mechanisms, and there are currently no penalties for noncompliance. The guidelines, therefore, finish with the explicit caveat that "they should be replaced in the future with policies and standards," which are typically weightier, compliance-driven forms of governance.

Beyond their operating principles, the guidelines also provide specific recommendations and best practices for using Generative AI effectively and responsibly. These include fact-checking and reviewing AI-generated content, disclosing the use of AI in content creation, and avoiding the sharing of sensitive information in prompts. These practices highlighted by the City are common across the guidelines covered in this scan, and similarly aim to ensure the accuracy and reliability of AI-generated content, promote transparency and trust, and protect sensitive data. To illustrate their recommendations and best practices, the guidelines include brief rationales why each practice should be adopted, clarifying examples and suggestions, and a helpful Do's and Don'ts section. Here, it's also worth noting that Boston's Dos and Don'ts speak specifically to City coders and programmers in addition to more general knowledge-worker employees.

Lastly, the guidelines encourage continuing education and training. The document substantiates the support of Boston's Department of Innovation and Technology for ongoing learning and collaboration opportunities through events, workshops, and knowledge sharing platforms, which would enable employees to further explore the capabilities of Generative AI in a responsible and productive manner. Boston's guidelines emphasize the rapidly evolving nature of Generative AI tools and so highlight the need for ongoing research to understand their functioning and societal impacts.

[9] Our emphasis.

Canadian Case Studies

4. *Guide on the use of generative artificial intelligence*, Treasury Board of Canada Secretariat, Government of Canada

Authors: Treasury Board of Canada Secretariat
Version Accessed: March 21st, 2024
Link: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/guide-use-generative-ai.html

Case Highlights: TBS
• Supports adjacent, existing government policies around Automated Decision-Making and Privacy laws. • Offers risk-based approach to Generative AI through the adoption of the "FASTER" principles (Fairness, Accountability, Security, Transparency, Educated, Relevant). • Essentially a risk mitigation document, that encourages public sector employees to explore with caution. • Suggests phased integration – a stratified approach that should begin with low-risk, interior use cases before high-risk, public facing/external use cases. • Final section details potential issues around Generative AI deployment and associated best practices.

In September 2023, the Treasury Board of Canada Secretariat (TBS) issued their *Guide on the use of generative artificial intelligence*, which is designed to help federal government officials use Generative AI tools responsibly. The document provides preliminary, overarching guidance for all federal departments on how AI can and cannot be used and identifies best practices and the risks and opportunities of day-to-day use of Generative AI. The new guidelines are also designed to enhance and work in conjunction with the government's *Responsible use of artificial intelligence* and *Directive on Automated Decision-Making*.

The primary objective of the *Guide* is to address instances where the use of Generative AI constitutes malpractice, potentially creating cybersecurity risks and misinformation, or when Generative AI tools produce discriminatory or biased results. The new guide stresses that the use of Generative AI be "governed with clear values, ethics, and rules."[10] While the authors recognize the value propositions of improved efficiency for service delivery in government, they also point to the associated risks and that these tools should not be used in all cases. Like the other guidelines in this scan, the risks highlighted by the TBS include concerns over data privacy, bias amplification, intellectual property rights, and the potential for generating inaccurate or misleading content.

In the TBS's own words, government officials should *"explore"* with caution, evaluating and mitigating "certain ethical, legal and other risks" prior to use, and that the use of Generative AI tools should be limited to "instances where they can manage [these] risks effectively."[11] That is, the associated risk of using Generative AI tools depends on what task they are used for, and on what risk mitigation measures are in place. The guidelines therefore advocate for an exploratory, stratified approach to AI deployment, that distinguishes between low and high-risk use cases.

They also recommend a phased integration of use cases that allows for accumulated experience and a greater understanding of the technology's benefits and limitations as it evolves. This type of strategy, combined with an emphasis on stakeholder engagement and multidisciplinary oversight, underscores the importance of adaptability and continuous learning in the governance of AI technologies. Relevant stakeholders include an organization's legal team, privacy, and security experts; chief information officers or data scientists; and diversity and inclusion specialists.

At the heart of the TBS guidelines is a risk-based approach to Generative AI through the adoption of the "FASTER" principles (Fairness, Accountability, Security, Transparency, Educated, Relevant), which serve as a principle-based foundation for the responsible deployment of Generative AI. These principles should help federal institutions maintain public trust and ensure that current and future innovations in AI are matched with ethical, enduring governance. The FASTER principles are detailed as follows:

---

[10] https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/guide-use-generative-ai.html

[11] Ibid.

- **Fair:** Ensures that AI deployments do not perpetuate biases or inequality, advocating for AI outputs that are equitable across diverse Canadian demographics.

- **Accountable:** Emphasizes the importance of ownership and oversight for AI-generated outcomes, insisting on accuracy, legality, and ethical integrity.

- **Security:** Ensures robust measures to protect sensitive information and uphold privacy standards.

- **Transparent:** Calls for clear disclosure about the use of AI, the nature of AI-generated content, and information about institutional policies to foster public trust and informed engagement with AI technologies. Documentation and explainability are required when AI tools are used to support decision-making.

- **Educated:** Stresses the need for ongoing AI literacy for both developers and users, enabling them to navigate AI tools responsibly and effectively.

- **Relevant:** AI applications should align with organizational and societal needs, ensuring that AI tools are deployed to meaningful ends and contribute to better outcomes, without undue environmental or social costs. Appropriateness of use, and whether AI should be used at all, are important considerations.

To bolster FASTER principle adoption, there are responsibilities for deploying institutions, the first of which is to "evaluate Generative AI tools for their potential to help employees, not replace them." Other responsibilities include providing access to ongoing training and education for employees; providing access to secure tools that meet information, privacy, and security requirements, and implementing oversight and performance management processes to monitor impact, and to make sure tools and their use comply with applicable laws and policies.

In addition, the *Guidelines* include an in-depth section detailing potential issues around Generative AI deployment and associated best practices. Many of these best practices are common across public sector guidelines and frameworks for Generative AI use, such as prohibiting the use of sensitive/personal information, understanding how a system uses input data (ex. whether it's used for training data; regular systems testing before and after deployment); and developing bias mitigation strategies from the planning and design stages including using GBA+ (Gender Based Analysis plus).

Due to the thoroughness of the document, the TBS guidelines are unique for their added emphasis on consent and disclosure, the need for public servant autonomy, and the environment/climate impact. For example, regarding transparency and disclosure, they demand that government departments always make it known to citizens if a service uses AI to interface with a user. This applies to any service that uses Generative AI to respond to a citizen's query, create a document, or make a decision.

Moreover, respecting the common problem associated with equating machine learning with objectivity, the *Guidelines* encourage digital literacy training, and extra consideration around what organizations are choosing to optimize, their reasons for doing so, and the inherent limits of Generative AI. Regarding the potential for overreliance on AI that could unduly interfere with human judgment, stifle creativity and workforce capability, the guide reifies the need for public servant autonomy, and an understanding of these tools as enabling, not substituting human efforts. And lastly, the guide includes a specific section around the environmental costs of Generative AI systems and suggests, among other practical options, using tools hosted in net-zero and carbon-neutral data centres.

It is worth noting that these best practices and recommendations are not accompanied by any enforcement mechanisms, and that there are currently no penalties for violating the guidelines, although they are based on existing pieces of legislation like the Privacy Act. So, while compliance is not enforced, violations could lead to penalties or legal action depending on their nature.

The TBS states that because AI technologies and their environment change rapidly, the guidelines are expected to evolve and be revised over time. Given the emergent nature of these technologies and their governance, the *Guidelines* also suggest that more experimentation, paired with ongoing performance measurement and analysis, will lead to a better understanding of potential gains and trade-offs and should inform future approaches.

5. *Generative artificial intelligence (AI) - ITSAP.00.041*, Canadian Centre for Cyber Security, Government of Canada

Authors: Canadian Centre for Cyber Security
Version Accessed: July 14th, 2023
Link: https://www.cyber.gc.ca/en/guidance/generative-artificial-intelligence-ai-itsap00041

Case Highlights: CCCS
• Emphasizes dual nature of Generative AI as a tool to enhance public service delivery and a vector for risk. • Offers various risk mitigation tasks and approaches at the organizational and individual levels. • Suggests developing organizational policies beyond these guidelines to govern corporate use of Generative AI tools.

The Canadian Centre for Cyber Security's (CCCS) publication on Generative AI offers an overview of both the transformative potential and inherent risks associated with this emerging technology. As one part of the Canadian government's interim suite of guidance on corporate use of Generative AI, this guideline stands out in its emphasis on the dual nature of Generative AI as a tool for innovation and a vector for potential threats – and, on how to mitigate these threats.

The guidelines detail the many applications of Generative AI, distinguished by its ability to create new content through models trained on extensive datasets, with use cases spanning sectors like healthcare, software development, online marketplaces, business, publishing, media, education, and cybersecurity. The authors illustrate the technology's broad impact while pointing out the significant risks associated with Generative AI like the propagation of misinformation, privacy breaches, biased content generation, and the potential for facilitating cyberattacks.

The CCCS' recommendations for mitigating these risks are various. At the organizational level, they recommend the adoption of strong network authentication mechanisms, regular updates and patching of vulnerabilities, and training for employees on the

recognition and management of social engineering attacks. For individuals, they recommend reviewing and verifying the authenticity of AI-generated content, practicing basic cybersecurity hygiene, and implementing online safety practices to protect against phishing and other forms of cyber threats. The CCCS thinks both organizations and individuals both should have access to verified channels and should know how to identify and report network abnormalities or malpractice, and to whom.

CCS recommends establishing clear organizational policies governing employee use of Generative AI tools. These policies should dictate the types of content permissible for generation, outline the technological safeguards to protect sensitive data, and describe the oversight and review processes necessary to maintain accountability. Additionally, the guidelines stress the importance of selecting training datasets with care to avoid the perpetuation of biases, and it advocates for the engagement with security-focused vendors to ensure robust data protection measures are in place.

In summary, the CCCS guidelines outline some of the actions organizations and individuals can take to mitigate threat and promote safe and responsible integration of Generative AI into various sectors. By balancing the technology's innovative capabilities with stringent security measures and ethical considerations, CCCS suggests that organizations and individuals can harness the benefits of Generative AI while minimizing its risks.

6. *Principles for responsible, trustworthy and privacy-protective Generative AI technologies*, Office of the Privacy Commissioner, Government of Canada

Authors: Office of the Privacy Commissioner of Canada
Version Accessed: December 7th, 2023
Link: https://priv.gc.ca/en/privacy-topics/technology/artificial-intelligence/gd_principles_ai/

Case Highlights: OPC
• Privacy-forward approach to Generative AI use: identifies considerations for the application of key privacy-centered principles to the use of Generative AI technologies. • Intended audience is twofold: AI developers and providers; and organizations using Generative AI for internal and external use-cases. • Proposes Fairness as a framing principle for the application of the OPC's nine guiding principles. • Privacy-forward means data-forward: suggests responsible use of Generative AI is based on the legal imperative to safeguard individual privacy/data rights.

In December 2023, the Office of the Privacy Commissioner of Canada (OPC) published their *Principles for responsible, trustworthy and privacy-protective Generative AI technologies*. From the unique perspective of Canada's privacy regulator, their document offers another principles-based approach to Generative AI adoption, but further identifies considerations for the application of key privacy-centered principles to the use of these technologies. There are nine principles in total, and the OPC states that the intended audience for the document is twofold: the developers and providers of foundational models and Generative AI systems; and organizations using Generative AI systems, including both public-facing and private use cases.

Controls for bias prevention are found throughout the OPC's principles, however the top of the document includes a special consideration for the unique impact Generative AI systems can have on vulnerable groups. In sum, the principle of Fairness is meant to frame the application of the rest of the principles below, insofar as developers, providers, and

organizations using Generative AI should give particular consideration to their “mutually-shared responsibility to identify and prevent risks to vulnerable groups, including children and groups that have historically experienced discrimination or bias.” From the perspective of the OPC, a privacy-first approach to data collection and use, when combined with robust privacy safeguards and mitigation measures, can act as a first principle way of ensuring the use of Generative AI protects vulnerable groups and does not create or amplify historical and present biases.

Beyond Fairness, the OPC outlines 9 other, privacy-centered principles for the use of Generative AI:

1. **Legal Authority and consent**

Given the regulatory and legal nature of the OPC’s duties, their guidelines are centered around an imperative to ensure meaningful consent for the collection and use of personal information within Generative AI systems. Their first order of business is legal authority and consent, emphasizing the necessity for all involved parties to clearly understand and document their legal basis for handling personal data throughout an AI system's lifecycle, from training and development to deployment, and eventual decommissioning. The guidelines also address the need for consent to be specific and freely given, avoiding manipulative designs, and for third-party data to be lawfully acquired with proper disclosure authority.

2. **Appropriate purposes**

The OPC document advocates for the collection, use, and disclosure of personal information only for purposes deemed appropriate and reasonable. It calls for a thoughtful consideration of the appropriateness of Generative AI applications, including a staunch avoidance of "no-go zones" — areas of use that could lead to unethical, unfair, or discriminatory outcomes. The document clarifies that, while there are not yet any firm rulings on the legality of such practices in the context of Generative AI, the OPC anticipates various ‘no-go zones’ to include AI content created for malicious purposes (including deep fakes), coercive chatbots, and the publication of defamatory information or mis/disinformation. Developers and providers of Generative AI systems are encouraged to proactively identify and mitigate potential misuses through adversarial testing and policy development, while organizations are encouraged to only use systems that respect privacy laws and best practices.

3. **Necessity and proportionality**

Guiding organizations should prefer anonymized or synthetic data over personal information. Related to the concept appropriateness, organizations should consider whether the use of a Generative AI system is necessary and proportionate, particularly where it may have a significant impact on individuals or groups.

4. **Openness**

The principle of openness extends through the lifecycle of a Generative AI system, requiring organizational transparency about how personal information is collected, used, and disclosed. It calls for clear communication with individuals about the use of AI-generated outputs, especially those with significant potential impacts.

5. **Accountability**

Accountability for the OPC is framed as essential for ensuring compliance with privacy legislation, necessitating clear internal governance and the ability to demonstrate adherence to privacy obligations. The guidelines suggest regular assessments, such as Privacy Impact Assessments (PIAs) and Algorithmic Impact Assessments (AIAs), to identify and mitigate privacy and fundamental rights impacts associated with Generative AI use.

6. **Individual access**

The OPC stresses that individuals should be able to "meaningfully" access and correct any personal information collected or generated by AI systems. In the context of Generative AI systems that are used for administrative decision-support, this principle relates to what is commonly known as "explainability," so that all parties impacted can access adequate documentation and information about that decision.

7. **Limiting collection, use, and disclosure**

Advocates for minimizing the personal information footprint of AI systems, utilizing data only as necessary for explicitly specified, legitimate purposes, and avoiding indiscriminate data collection. Building off the principle of appropriateness, this principle also includes cautioning "function creep" with respect to data collection and use beyond the original

purposes at the time of collection, i.e., based on the breadth of potential purposes for a Generative AI system. They also state that the *public accessibility of data does not mean that it can be/should be indiscriminately used within a Generative AI system.*

8. Accuracy

Accuracy demands that personal information used in training Generative AI models be precise and up-to-date, necessary for the intended purposes, and includes provisions that developers and providers correct or update the AI system when inaccuracies in the training data are identified. Organizations need to take reasonable steps to ensure that any outputs from a Generative AI tool are accurate and appropriate, especially if outputs will be used in high-risk contexts (i.e., released publicly, or used to make or support decisions made about an individual or individuals). Organizations should also be aware that issues regarding the accuracy of training data and the potential for bias outputs, either in general or respecting a specific group of people, may make a Generative AI system inappropriate for use.

9. Safeguards

Additionally, organizations should employ robust safeguards to protect personal information and address privacy risks specific to Generative AI technologies, including measures against exacerbating biases and potential security threats like prompt injection attacks, jailbreaks, and model inversions.

In essence, these guidelines encapsulate a data-forward, principled approach to navigating the complex interplay between Generative AI innovation and privacy protection. By focusing on the "raw material" of what goes into, and what comes out of Generative AI systems, the OPC couch responsible use of these tools in the legal imperative to safeguard individual privacy rights, and the ethical imperative to maintain public trust.

International Case Studies

7. *Initial advice on Generative Artificial Intelligence in the public service,* Government of New Zealand

Authors: National Cyber Security Centre in New Zealand, Te Tari Taiwhenua (Department of Internal Affairs), and Statistics New Zealand
Date Accessed: July 2023
Link: https://www.digital.govt.nz/assets/Standards-guidance/Technology-and-architecture/Generative-AI/Joint-System-Leads-tactical-guidance-on-public-service-use-of-GenAI-September-2023.pdf

Case Highlights: NZ
• Recommends that officials at each adopting organization collaborate with diverse stakeholder groups to develop their own, context-specific policies and use standards. • Addresses common risk of "Shadow IT" – the unsanctioned use of Generative AI within an agency's environment on personal or internal devices. • Notes that Generative AI governance requires alignment with Te Tiriti-based principles and engagement with Indigenous stakeholders. • Explicitly mentions the risk of using Generative AI to code. • Recommends creating dedicated testing spaces, like sandboxes, for teams to trial safe Generative AI use.

The National Cyber Security Centre in New Zealand, in collaboration with Te Tari Taiwhenua (Department of Internal Affairs) and Statistics New Zealand, has formulated a set of guidelines aimed at guiding federal agencies and officials across the public sector in the responsible use of Generative AI. The authors state that the document is meant to serve as interim guidance and initial advice, and that although it could find broader application beyond the public service, the specific and intended audience are Public Service AI practitioners and decision-makers. In their intended audience they also include public

service procurement, data, digital, privacy and security leaders, for whom the guide should provide iterative “guardrails” for safe learning and use of these technologies.

Their approach begins by recommending that these officials at each organization work together to create their own, context-specific policies and standards for experimenting, testing, and using Generative AI. They also begin by acknowledging the transformative implications of Generative AI across various aspects of public service, including process improvement, service delivery, cybersecurity, and policy development, while also cautioning against its potential risks, particularly concerning data sensitivity and personal information security.

The structure of the document features key recommendations, followed by risk management and mitigations strategies. Their first recommendation is to avoid using Generative AI tools with data classified as sensitive or higher, as defined by the New Zealand Protective Security Requirements, and references the severe consequences that a breach could entail for society, the economy, and public services. The guidelines urge extreme caution regarding personal information, advising against its use with tools inside and outside internal networks, to protect individual privacy and maintain public trust in government. The broader directive is to minimize the use of personal data in Generative AI applications, resorting to non-personal or synthetic data wherever possible, and ensuring any use of personal information is necessary and safe. This recommendation also includes avoiding using any information that would be withheld under the Official Information Act.

The guidelines also address the common risk of Shadow IT — the unsanctioned use of Generative AI within an agency's environment on personal or internal devices — which could compromise security and data privacy, adding unnecessary complexity to technological ecosystems. Moreover, sanctioning Generative AI use behind firewalls, and paying for Generative AI tools/services does not do away with their risks, either. The NZ guidelines point out that while free Generative AI tools may lack robust privacy and security controls, paid tools and services also carry their own risks. The authors therefore recommend that, to make informed decisions about what tools to use, agencies carefully evaluate them based on several factors including cost, functionality, and privacy and security maintenance and support.

In their section on understanding and managing Generative AI risk, the guidelines recommend that agencies “robustly govern” their internal use of these technologies. For them, this should include senior approval for Generative AI related decisions and the

development of an AI policy in collaboration with the Government Chief Privacy Officer. Governance establishes the foundation for risk mitigation/management and should pre-emptively ensure that public service Generative AI applications are safe, ethical, transparent, and unbiased. In addition, in the context of New Zealand, this means government applications should be based on the principles of security and privacy by design and should be aligned with Te Tiriti-based principles and with Government's Procurement Rules.

Like other jurisdictions, privacy and security risks are a top priority for NZ, and so the guidelines recommend conducting privacy impact assessments and strong cybersecurity practices to manage these risks effectively. Privacy-based risk mitigation strategies include applying “privacy by design” principles, data anonymization, and encryption to ensure that outputs are only accessible to those authorized. The guidelines also recommend being open and transparent with stakeholders and the public about how personal data, if any, will be collected, stored, and used in general and in relation to Generative AI use.

The New Zealand guidelines are unique because they explicitly mention the risk of using Generative AI to code. Respecting access and control for security risks, the guidelines caution agencies against using code generated by any generative tool or putting this code in their production systems without robust review. Code generated by these tools can be vulnerable and can include mistakes, resulting in potential security vulnerabilities and system compromise.

The NZ guidelines are also unique insofar as they highlight the importance of engaging with Iwi Māori and other indigenous stakeholders. As part of their risk mitigation recommendations, the guidelines stress respecting the Te Tiriti o Waitangi (the Treaty of Waitangi), Māori data governance, and working with Māori representatives, especially when Māori data or outcomes could be impacted by Generative AI use. This level of engagement invites diverse views and concerns about discrimination and bias in Generative AI outputs and promotes mutual benefit around Generative AI use that is safe, value-adding, and simple for indigenous participants.

The NZ guidelines also advocate for ethical Generative AI use that understands the inherent limitations of these tools and validates the accuracy of its outputs, emphasizing the need for human oversight and accountability in decision-making processes supported by Generative AI at any point. Agencies are also encouraged to be transparent about their use of Generative AI, and to proactively assess and mitigate potential social, security, and

intellectual property risks associated with publicly available AI tools, even those that are part of government assurance and procurement processes. Agencies are therefore encouraged to apply the NZ government's strict procurement principles when thinking about Generative AI vendors.

Lastly, the guidelines include the somewhat nuanced recommendation that agencies create guardrails and dedicated spaces for public servants to safely test and experiment with Generative AI. According to the guide, "safely trialling [Generative AI] is important to ensuring that AI systems and their outputs are as expected, and do not cause unintended harm to people, communities, society, the economy and/or the environment." To do so, the guide recommends choosing low-risk datasets to learn and trial safely using AI tools and to gauge the nature and accuracy of outputs before deployment scenarios. This would include testing under various conditions to identify common issues, and to validate that models and training data are appropriate for use in local contexts.

8. *Guidelines For Staff on The Use of Online Available Generative Artificial Intelligence Tools*, European Commission

Date Accessed: May 24th, 2023
Link:
https://www.asktheeu.org/en/request/13063/response/45877/attach/3/guidelines%20on%20the%20use%20of%20online%20generative%20artificial%20intelligence%20tools.pdf?cookie_passthrough=1

Case Highlights: European Commission
• Outlines an approach to Generative AI that emphasizes critical assessment, ethical usage, and cautions against dependency on Generative AI outputs for sensitive tasks. • Emphasizes understanding the limitations of publicly available Generative AI tools like ChatGPT and DALL-E. • Copyright and IP concerns are highlighted.

The European Commission's (the Commission) approach to the incorporation of Generative AI technologies into the workflow of its staff is encapsulated in a set of guidelines focused on assessing risks, understanding limitations, and setting conditions for safe usage. Notably, these guidelines cater specifically to third-party tools that are publicly available online and distinguish between these and internally developed tools, which are subject to separate assessments under existing IT governance structures. Acknowledging the dynamic landscape of Generative AI, the Commission treats this document as a living entity, adaptable in response to technological progress and legislative developments, notably the ongoing negotiations of the EU's Artificial Intelligence Act.

The Commission mentions popular examples of Generative AI technologies such as ChatGPT and Dall-E that underline Generative AI's potential to assist staff in diverse tasks, from drafting documents to creating visual content. However, the Commission emphasizes the importance of understanding the intrinsic limitations of these technologies. Notably, these tools lack the personal experience and emotional depth of their human counterparts, and while some models can access the internet, this is typically bound to a

limited timeframe. As a risk, this limit is compounded by the fact that models do not inherently understand context, which can lead to errors in output.

The Commission is also concerned with the potential for unauthorized disclosure of sensitive information when interacting with generative models, prompting a strict directive in the guide against sharing non-public or personal data with these AI tools. Moreover, the guide highlights the inherent risks of biases and inaccuracies in outputs, driven by potentially flawed or biased training data and algorithmic predispositions. Commission staff are thus advised to critically evaluate any generated content for biases and factual correctness.

Intellectual property rights are also a significant concern, given the opacity surrounding the materials and data used to train Generative AI models. The potential for copyright infringement necessitates scrutiny and review of their outputs to ensure they do not unlawfully replicate third-party intellectual property. Furthermore, the reliability and of Generative AI tools are called into question, with a strong cautionary note against depending on these technologies for critical and time-sensitive operations.

To navigate these challenges, the guidelines offer some practical directives for staff, and underscores the need for a critical and responsible approach to using Generative AI tools. Directives include avoiding verbatim replication of AI outputs in public documents, and ensuring any use of such technologies is in line with established intellectual property laws and organizational security protocols.

In conclusion, the Commission's guidelines provide an interim, albeit limited framework for the adoption of Generative AI tools, balancing the exploration of these technologies' potential benefits against a backdrop of ethical, legal, and operational considerations. Their focus on human oversight and critical assessment, ethical usage, and caution against dependency on Generative AI outputs for sensitive tasks, the Commission aligns itself with the EU’s broader governmental efforts to harness AI's potential responsibly. The Commission’s initiative exemplifies the kind of preliminary move organizations can make towards more structured frameworks that govern the use of emerging AI technologies.

9. *Gen AI framework for HM Government*, Central Digital and Data Office, United Kingdom

Authors: Central Digital and Data Office
Date Accessed: January 18th, 2024
Link:
https://assets.publishing.service.gov.uk/media/65c3b5d628a4a00012d2ba5c/6.8558_CO_Generative_AI_Framework_Report_v7_WEB.pdf

Case Highlights: CCDO
• Of the cases researched, one of the most comprehensive and detailed approaches to Generative AI adoption in the public sector. • Outlines practical steps public servants need to take when building Generative AI solutions, including goal definition, team building, creating an AI support structure, and procurement. • Intended audience includes a broad spectrum of public service staff across areas of interest and expertise.

The Central Digital and Data Office's (CDDO) guidelines on Generative AI usage is one of the most comprehensive and resourceful frameworks in the field today. Like the others in this scan, the UK's guidelines have as their backbone 10 principles to help UK officials adopt and deploy Generative AI across government services in a way that is lawful, ethical, responsible, and in alignment with the public interest. However, this is an exceptional case study because its bulk is dedicated to supporting officials, from developers to policy makers, in the practical application of these principles. The guidelines echo others in the space while outlining the practical steps public servants need to take in building Generative AI solutions, including defining their goals, building their team, creating the Generative AI support structure, and procurement.

The guidelines speak to a broad spectrum of staff involved in the development, deployment, and oversight of Generative AI tools, suggesting a collective responsibility across different levels of expertise and decision-making authority within the UK government. Therefore, CDDO's guidelines are an essential document for team leaders

and higher-level decision-makers, and a potentially useful document for all levels of staff involved in the ecosystem around Generative AI and technology, from procurement to deployment and day-to-day use. That said, the sheer breadth of the document and its highly technical language may present barriers to access.

The CDDO's guidelines are structured around 10 core principles, which are read from the perspective of the user, and could almost function like a checklist/assessment for use:

1. You know what Generative AI is and what its limitations are.
2. You use Generative AI lawfully, ethically, and responsibly.
3. You know how to keep Generative AI tools secure.
4. You have meaningful human control at the right stage.
5. You understand how to manage the full Generative AI lifecycle.
6. You use the right tool for the job.
7. You are open and collaborative.
8. You work with commercial colleagues from the start.
9. You have the skills and expertise needed to build and use Generative AI.
10. You use these principles alongside your organization's policies and have the right assurance in place.

Over the course of the document, the CDDO offers a wealth of practical advice on how to apply and support these principles. Due to the lengthiness of their guidance, it is summarized below into key categories:

Defining Goals and Identifying Use Cases

The guidelines stress the importance of setting clear Generative AI usage goals aligned with organizational objectives and measurable outcomes. Identifying specific use cases where Generative AI can offer substantial benefits is critical, so that technologies are adopted to addresses real challenges rather than for its novelty. The CDDO's guide identifies a long list of promising use cases, and those to avoid, as a framework for current and potential government Generative AI projects.

Building the Team and Acquiring Skills

A multidisciplinary team encompassing various expertise areas is crucial for developing technically sound and responsible Generative AI governance ecosystems and solutions. For

the CDDO, expertise areas could include business leaders, data scientists, software developers, user researchers, and support from legal, security colleagues, as well as ethics and data privacy experts. Members of the team should be provided opportunities for continuous skill development and education in emerging areas to keep pace with Generative AI advancements. Notably, whatever an organization's learning plan, the CDDO suggests education plans should meet the needs of five "groups of learners:" Beginners, Operational delivery and policy professionals, Digital and technology professionals, Data and analytics professionals, and Senior civil servants.

Creating Support Structures and Making Informed Decisions

For Generative AI to be responsibly adopted, the CDDO recommends establishing appropriate organizational support structures. This includes the development of formal and informal AI strategies, governance boards, communication strategies, and sourcing strategies (i.e., definition of which capabilities are built internally, and which would be sought from partners). Decision-making on tools and technologies must consider the organization's specific needs and existing infrastructure.

Ensuring Security and Generating Reliable Results

Like all other guidance in this scan, security is paramount for the CDDO, who recommend practical strategies to mitigate risks associated with Generative AI. Achieving reliable results requires careful model selection, clear interfaces, input and output evaluation for accuracy and bias, and maintaining human oversight in all automated processes. Continuous evaluation and feedback are essential for refining Generative AI solutions, and for keeping up with changing fairness considerations and societal expectations.

Incorporating Governance

The guidelines recommend strong governance processes to help navigate the risks associated with security, bias, and data management. They advocate for continuous improvement, stakeholder engagement, and long-term planning for sustainable AI initiatives in the public service. Recommendations include establishing an "AI Governance Board" or ensuring various AI representation on existing boards or working groups who can provide strategic oversight and accountability. Meanwhile, something like an "Ethics Committee" could focus on the ethical implications of decisions, emphasizing values like fairness and privacy.

They also recommend creating an AI/ML systems inventory that catalogues all deployed AI systems within an organization. An inventory could help organizations avoid the duplication of efforts while enhancing oversight, knowledge-sharing and awareness of AI use and potential risks across programs and projects. Moreover, programme governance within teams should record and detail model maintenance, knowledge transfer, and establish clear accountability for AI systems.

The CDDO makes several practical recommendations for overall AI governance that would include Generative AI usage. These general governance recommendations include:

- Engage with assurance teams and consider setting up an AI governance board or including AI experts on existing boards.
- Establish an ethics committee with a broad representation to focus on ethical implications.
- Create an AI/ML systems inventory for a comprehensive overview of all AI deployments within the organization.
- Ensure program teams have clear governance structures that promote diversity within project teams for a range of perspectives.

Risk Mitigation

Following the LLM AI Cybersecurity & Governance Checklist approach of the Open Worldwide Application Security Project (OWASP) to identify the unique risks posed by LLMs, the CDDO lists risk-based scenarios of some of the most common vulnerabilities, and puts them in context of how they could apply to LLM applications in government. These scenarios serve as an extremely helpful example of common development and deployment options for governments of all levels, and although they focus on the use of LLMs specifically, many of them would also apply to other types of Generative AI models. Notably, one of the cases speaks directly to developers on the use of LLMs to generate code, why this is a high-impact risk, and how to mitigate this risk if pursued.

In summary, the CDDO guidelines provide a detailed roadmap for adopting Generative AI in government operations, focusing on well-defined goals, targeted use cases, skilled teams, supportive structures, and strong governance practices. This methodical approach supports the application of the guideline's 10 core principles and goes to much length to ensure Generative AI's beneficial, ethical, and responsible use in the public service.

Annex: The Team

This report was produced by Think Digital and commissioned by the Regional Municipality of York with the intention of contributing to what is an increasing body of knowledge on the use of Generative AI in public sector organizations. The team of key contributors to the research and writing of this report are as follows:

Lead Researcher / Writer:

Jacob Danto-Clancy, Digital Policy Analyst, Think Digital

Jacob is a multi-disciplinary researcher and writer with experience consulting for non-profit and public sector clients on projects centred around AI, GovTech, and digital infrastructure. As a digital policy analyst and consultant, Jacob helps federal agencies better understand and respond to opportunities and risks that emerge from technological change. Jacob received his Master of Public Policy in Digital Society in 2022 from McMaster as part of the program's first cohort. He is also a co-founder of Boon, a research and public policy consultancy based in Toronto, ON.

Project Coordinator:

Ryan Androsoff, CEO and Founder, Think Digital

Ryan Androsoff is the Founder and CEO of Think Digital, a consultancy focused on helping public sector organizations to adapt and thrive in the era of digital disruption. He is also the host of the Let's Think Digital podcast that explores how technology is transforming government. Ryan is an international expert on digital government with a passion for public sector entrepreneurship and more than two decades of experience working with governments and international organizations both at home in Canada and around the world. Ryan also serves as an Associate with the not-for-profit Institute on Governance where he leads the Digital Executive Leadership Program and related digital government training and advisory practice. Ryan is a graduate of the Harvard Kennedy School of Government in Cambridge, Massachusetts where he earned a Master in Public Policy degree, with research focused on the impacts for governments of new digital technologies. Ryan also has an Honours degree in Public Affairs and Policy Management from Carleton University in Ottawa.

Subject Matter Experts:

Jen Schellinck, Associate, Think Digital

Jen Schellinck's goal as a data scientist and AI technologies specialist is to help organizations understand the value that cutting-edge data technology can bring to their work and success. She uses her knowledge of artificial intelligence, machine learning and data science to help organizations achieve their greater potential. For each project, she draws from a pool of experts to provide clients with the most valuable information they need, through consulting, workshops and data solutions. She received her PhD in Cognitive Science in 2009 and has been active in the AI field for over a decade. She is currently an adjunct researcher at the Institute of Cognitive Science at Carleton University and continues to be an active researcher.

John Stroud, Associate, Think Digital

John is a strategic adviser to leaders on linking people with technology. His vision opens people's minds to new possibilities, and he challenges them to consider creative options. People turn to John for plainspoken, easy-to-understand explanations. John is a certified OpenExO consultant in exponential technologies. Prior to launching his company AI Guides, John served as Vice President, Strategy at a federal crown corporation ($600M budget and 8000+ workforce) with responsibilities for Governance, Human Resources, Communications, Legal, Performance Measurement and Risk.

John obtained his Master of Philosophy from Oxford, his law degree and Master of Public Administration from the University of Victoria, and his BA from the University of Toronto. He also obtained his ICD.D for completing the Director's Education Program at the Institute of Corporate Directors.

THINKDIGITAL

DIGITAL TRANSFORMATION FOR PUBLIC GOOD

CONTACT THINK DIGITAL

@ThinkDigitalCA

contact@thinkdigital.ca

/think-digital-ca

https://thinkdigital.ca/

/ThinkDigitalCA

https://thinkdigital.ca/blog/

Adapt and Thrive in the Era of Digital Disruption