MacEwan UNIVERSITY

# AI BEST PRACTICES GUIDELINES

OFFICE OF GENERAL COUNSEL

September 2024

# Table of Contents

**How to read this document**[1]**:** There is a lot of genuine excitement and curiosity about AI systems. To help better understand how to use AI systems responsibly, including generative AI, we have prepared this document that outlines a series of best practices for you to incorporate into your day-to-day work when using AI systems.

## Introduction

Artificial intelligence is changing the world. For many of us at MacEwan University, there is a desire to embrace the use of emerging technologies within a framework of informed and responsible innovation. In addition, the Teaching and Learning Committee of the General Faculties Council formed a working group, Artificial Intelligence/Academic Integrity (AI2). This working group published its final report on November 15, 2023. This report included 12 recommendations for the University's approach to generative AI.

## What is AI?

> ***An artificial intelligence system (or "AI system") refers to a technological system that, using a model, makes inferences in order to generate output, including predictions, recommendations, or decisions.***[2]

New technologies such as AI systems have the potential to significantly improve the world. However, it is also important to be aware of the risks involved in using these systems. This document offers guiding principles for responsible AI use, outlines known AI-related risks, and provides helpful guidance for using AI responsibly. This document contains the following:

A. **Responsible AI principles:** This section summarizes a set of guiding principles that should be considered when designing, developing, and using AI solutions.

B. **Risks relevant to the use of AI:** This section summarizes the key risks when using AI solutions. It covers risks such as accuracy, bias and non-discrimination, liability, intellectual property, privacy, and explainability.

C. **Responsible AI guidelines:** This section provides guidance in the following four contexts to help operationalize responsible AI:

1. the use of publicly available AI solutions;
2. use of third-party vendor systems (including both AI procurement and embedded AI systems);
3. designing AI systems; and
4. third-party consultants using AI.

---

[1] This document was prepared with assistance from INQ Consulting.

[2] Champagne, F.-P. (2023). Letter to Mr. Joël Lightbound, M.P. regarding proposed amendments to the Digital Charter Implementation Act, 2022. In *Letter*. https://www.ourcommons.ca/content/Committee/441/INDU/WebDoc/WD12751351/12751351/MinisterOfInnovationScienceAndIndustry-2023-11-28-Combined-e.pdf

# A. RESPONSIBLE AI PRINCIPLES

AI systems are not inherently neutral, perfect, or fair. To reduce harm and boost the social benefits of AI, we need to adopt ethical practices when we design, develop, and use these systems.

There are several ethical frameworks already available to help understand and think through AI ethics and impact, such as the EU's Ethics Guidelines for Trustworthy AI[3], IEEE Ethically Aligned Design[4], and NIST's AI Risk Management Framework[5]. Other global standards organizations such as the International Organization for Standardization (ISO) have also developed standards for the responsible management of AI systems, as demonstrated by ISO/IEC 42001[6] (Artificial Intelligence Management System).

There is no 'one-size fits all' ethical framework, but a good starting point is the OECD's Responsible AI Principles.[7] These principles serve as the backbone for many legislative efforts worldwide, including Canada's Bill C-27 Part III[8] and the European Union's *AI Act*[9]. These principles have been adapted from the OECD principles.

## **Principle 1**: Inclusive growth, sustainable development, and well-being

Individuals should responsibly manage AI to benefit people and the planet. This includes enhancing human capabilities, promoting creativity, including underrepresented groups, reducing inequalities, and protecting the environment, fostering overall growth, well-being, and sustainability.

## **Principle 2:** Human-centered values and fairness

AI developers must respect laws, human rights, and democratic values throughout the AI lifecycle. This includes ensuring freedom, dignity, privacy, equality, diversity, fairness, and labour rights. Appropriate mechanisms should be implemented to maintain these values.

## **Principle 3:** Transparency and explainability

AI developers should commit to transparency and responsible disclosure about AI systems. They should provide clear information to:

---

[3] *Ethics guidelines for trustworthy AI*. (2019, April 8). Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[4] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2018). Ethically aligned design. In *A Vision for Prioritizing Human Well-being With Autonomous and Intelligent Systems* (Version 2-For Public Discussion). https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

[5] Raimondo, G. M., Locascio, L. E., U.S. Department of Commerce, & National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). In *NIST AI 100-1* [Report]. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

[6] *ISO/IEC 42001:2023*. (n.d.). ISO. https://www.iso.org/standard/81230.html

[7] *OECD AI Policy Observatory Portal*. (n.d.). https://oecd.ai/en/ai-principles

[8] *Bill C-27, An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts*, 1st Sess, 44th Parl, 2021 (second reading April 24, 2023). https://www.parl.ca/legisinfo/en/bill/44-1/c-27

[9] *Regulation - EU - 2024/1689 - EN - EUR-LEX*. (n.d.). https://eur-lex.europa.eu/eli/reg/2024/1689/oj

- foster a general understanding of AI;
- inform individuals about their interactions with AI;
- help those affected by AI understand its outcomes; and
- enable those negatively impacted by AI to challenge its outcomes with understandable information about the decision-making process.

### **Principle 4:** Robustness, security, and safety

AI systems must be robust, secure, and safe throughout their lifecycle, functioning correctly under normal and adverse conditions without posing unreasonable risks. Developers should ensure traceability of datasets, processes, and decisions to analyze outcomes and respond to inquiries. A continuous risk management approach should be applied to address risks like privacy, security, safety, and bias.

### **Principle 5:** Accountability

AI developers should be accountable for the proper functioning of AI systems and for upholding these principles, based on their roles and the context in which they operate.

Guiding questions to evaluate risk:

1. How accurate are the AI system's predictions compared to the accepted true values?
2. What methods are used for continuous testing and monitoring of the AI system's performance?
3. How does the system maintain its performance under various conditions?

## B. UNDERSTANDING THE RISKS WITH AI

The use of AI systems comes with risk, including legal, reputational, operational, financial and ethical concerns, due to the potential scale and complexity of harm. Not all AI systems are the same. Some pose minimal risk, while others, like facial recognition or biometric processing, carry higher risks and need more oversight. Here are some risks associated with AI systems:

### Accuracy and Effectiveness

Accuracy is defined as the "...closeness of results or observations, computations, or estimates to the true values or the values accepted as being true," while robustness is described as "...the ability of a system to maintain its level of performance under a variety of circumstances."[10] Continuous testing and monitoring are essential to evaluate AI system validity and ensure consistent performance. The accuracy and robustness of AI systems are important considerations for assessing the reliability of outputs and usefulness of their application. Some applications of AI need a high degree of accuracy whereas others can support lower levels.

[10] International Organization for Standardization & International Electrotechnical Commission. (2022). *ISO/IEC TS 5723:2022(EN) Trustworthiness — Vocabulary*. Online Browsing Platform (OBP). Retrieved July 5, 2024, from https://www.iso.org/obp/ui/#iso:std:iso-iec:ts:5723:ed-1:v1:en

## Bias and Non-discrimination

Bias can manifest without any deliberate intention to be prejudicial or discriminatory in a harmful or unlawful way. Within the context of an AI solution, bias can be categorized into at least three types:[11]

1. Systemic bias which arises from datasets, organizational norms, processing, and broader societal factors.

2. Human-cognitive bias which refers to how individuals or groups perceive and process information from AI systems.

3. Computational or "algorithmic bias" refers to the "…application of an algorithm that compounds inequalities in socioeconomic status, race, ethnic background, religion, gender, disability, sexual orientation, and amplifies inequalities…"[12]

Guiding questions to evaluate risk:

1. How are datasets evaluated to ensure they are representative and unbiased?
2. What steps are taken to mitigate human-cognitive biases in AI system use?
3. How is the AI system tested and adjusted to prevent algorithmic bias?

## Safety and Security

AI systems are tailored to adapt to the data environments to which they are exposed. This adaptability means that alterations to these environments can considerably impact the behaviour of AI systems, occasionally leading to unforeseen outputs. Additionally, AI systems, especially those exposed to the public, can be targets of cyber attacks, potentially leading to harmful consequences.

Guiding questions to evaluate risk:

1. What measures are in place to protect the system for cyber threats?
2. How is the AI system monitored for unexpected or harmful outputs?
3. How does the system adapt to changes in its data environment?

[11] Raimondo, G. M., Locascio, L. E., U.S. Department of Commerce, & National Institute of Standards and Technology. (2023b). Artificial Intelligence Risk Management Framework (AI RMF 1.0). *In NIST AI 100-1* [Report]. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

[12] Panch, T., Mattie, H., & Atun, R. (2019). Artificial intelligence and algorithmic bias: implications for health systems. *Journal of Global Health*, 9(2). https://doi.org/10.7189/jogh.09.020318

## Privacy

The privacy implications of AI has been raised by several privacy commissioners in Canada:

> *"...uses of AI that are based on individuals' personal information can have serious consequences for their privacy. AI models have the capability to analyze, infer, and predict aspects of individuals' behaviour, interests, and even their emotions in striking ways. AI systems can use such insights to make automated decisions about individuals...Such decisions have a real impact on individuals' lives, and raise concerns about how they are reached, as well as issues of fairness, accuracy, bias, and discrimination."*[13]

AI systems are often trained on large datasets containing personal information[14] implicating various privacy laws and practices. AI systems may also collect and use personal information to deliver and improve the product, raising concerns about secondary uses of personal information.

Throughout the assessment process, it is important to understand how personal information is handled and protected and to ensure that any personal information used to train AI models is as accurate as necessary for the purposes for which it is to be used. Transparency is essential for understanding the decision-making process, especially when the decision directly impacts an individual. Finally, it is crucial to comply with Alberta's *Freedom of Information and Protection of Privacy Act*[15], the *Personal Information Protection Act*[16] and all other relevant privacy laws.

Guiding questions to evaluate risk:

1. How is personal information protected and managed?
2. How can we ensure compliance with privacy laws and regulations?
3. How are individuals informed about the AI system's data collection and use practices?

## User Experience

If an AI solution does not perform as expected, it could result in inappropriate solutions, decisions, or recommendations resulting in an overall poor experience for those using or interacting with the AI system. Such errors may result in harm, cause friction in the user experience, and exacerbate an erosion of trust.

Guiding questions to evaluate risk:

1. How frequently does the AI system perform as expected without error?
2. What support is provided to users to address errors or issues?
3. How is transparency maintained to build and maintain user trust?

---

[13] Office of the Privacy Commissioner of Canada. (2020, November 12). *A Regulatory Framework for AI: Recommendations for PIPEDA reform*. https://www.priv.gc.ca/en/about-the-opc/what-we-do/consultations/completed-consultations/consultation-ai/reg-fw_202011/

[14] *Ibid.*

[15] *Freedom of Information and Protection of Privacy Act*, RSA 2000, c F-25.

[16] *Personal Information Protection Act*, SA 2003, P-6.5.

## Explainability

The inner workings of an AI system or model can be hard or impossible to explain. When a system's internal workings are not transparent, it's called a 'black box'. Some AI systems may not need explainability if they don't make decisions requiring human understanding. However, the inability to interpret how an AI makes decisions or recommendations should be weighed against its use context.

Guiding questions to evaluate risk:

1. How are the AI system's decision-making processes made understandable?
2. In what contexts is explainability most critical, and how is it ensured?
3. What methods are used to interpret and explain the AI system's outputs?

## Knowledge Transfer & Transparency

Users need to understand AI systems' limitations and best practices. AI can sometimes produce outputs that are fictional or incorrect but seem factual, known as 'hallucinations.' Proper guidance on usage is essential. Even if not legally required, organizations might choose to inform individuals about AI's data use as an ethical practice, including notices about data collection and usage.

Guiding questions to evaluate risk:

1. What training and resources are provided to users to understand the system?
2. How are the AI system's limitations and best practices communicated?
3. How are users educated about potential inaccuracies or hallucinations in outputs?

## Ethics

Ethical considerations for AI include data ethics and AI ethics. Data ethics involve the respectful and secure handling of personal and confidential information. AI ethics focus on the responsible use of data for training AI systems and selecting systems that are explainable, traceable, and secure. AI systems should be rigorously tested to avoid discrimination and ensure consistent, unbiased performance.

Guiding questions to evaluate risk:

1. How is sensitive information handled by the AI system?
2. What ethical guidelines govern the use and training of the AI system?
3. How is the AI system tested and monitored to ensure it remains non-discriminatory?

## Intellectual Property and Liability

Generative AI poses risks related to intellectual property. AI-generated content, such as text, images, or music, may unintentionally infringe on copyrights, leading to legal issues and financial liabilities. Inadequately trained AI systems, like chatbots, can spread incorrect information, risking reputational damage and legal liability if users rely on this misinformation.

Guiding questions to evaluate risk:

1. How is AI-generated content validated to ensure it does not infringe on intellectual property rights?
2. Who is liable for errors or misinformation generated by the AI system?
3. How is the training data sourced and verified for compliance with intellectual property laws?

## Regulatory Uncertainty

The changing landscape of AI regulation presents challenges for organizations. While the EU has made strides with the *Artificial Intelligence Act*[17], other regions lack comprehensive frameworks. This uncertainty complicates compliance and investment decisions, leading to hesitation in adopting AI solutions due to potential future legal issues.

For instance, in anticipation of the above-noted *Artificial Intelligence Act*[18], many companies with European clients must prepare for extensive regulatory requirements, including comprehensive compliance audits of their existing AI systems. Should Canada's regulation take a similar approach (in the way that Canadian privacy law has been affected by the European Union's *General Data Protection Regulation*[19]), Canadian entities using AI systems will likely have numerous regulatory compliance requirements, including the establishment of governance systems, and the conduct of artificial intelligence impact assessments. In addition, as with the GDPR, Canadian legal entities with European clients may have to contend with compliance with the *Artificial Intelligence Act*.

Guiding questions to evaluate risk:

1. How does our organization stay up to date on AI laws?
2. What strategies are in place to ensure we comply with current and future regulations?
3. How does regulatory uncertainty impact the adoption and investment in AI solutions for our organization?

[17] *Supra* note 9.
[18] *Supra* note 9.
[19] *General Data Protection Regulation (GDPR) – legal text.* (2024, April 22). General Data Protection Regulation (GDPR). https://gdpr-info.eu/

## C. GUIDELINES ON THE RESPONSIBLE USE OF AI

This section provides a series of guidelines for MacEwan University staff to incorporate into their day-to-day activities when using, interacting with, procuring, or designing AI systems. The guidance also extends to third-parties that are performing work in partnership with MacEwan University. This section is divided into the following four sections:

1. how to use publicly available AI (consumer or general-purpose AI solutions) responsibly;
2. how to responsibly procure and use third-party vendor systems, including software with embedded AI features;
3. how to design AI systems responsibly; and
4. guidelines for third-party consultants using AI.

## **CHECKLIST #1:** Use of publicly available AI (consumer or general-purpose AI solutions)

In this section, "consumer" or "general-purpose" AI systems refer to publicly available AI systems accessible online, such as ChatGPT, Copilot, Gemini, Bard, or DALL-E. Use caution with these systems, as they can produce incorrect information or infringe on intellectual property. This section offers guidelines to help use these systems safely and responsibly. While not exhaustive, these guidelines provide a solid foundation for their safe use. These guidelines have been adapted from the Canadian Government's Guide on the use of generative AI[20].

**Consideration 1:** Certain consumer or general-purpose AI systems do not meet adequate information security and data protection standards.

Organizations that supply AI systems may inspect input data or use it to train their algorithms, potentially leading to privacy or security breaches. These risks can arise when input data is stored on servers not controlled by our organization, where it may be retained longer than necessary.

Checklist:

- ☐ Do not enter sensitive information including personal information or confidential business information into any systems that are not managed by our organization.
- ☐ Wherever possible, use the "opt-out" feature or disable history to ensure that prompts are not used to further develop the AI system.
- ☐ Consult with Legal and Information Security to review a potential system and its terms of use, privacy policy, and other applicable legal documentation before using any system to process sensitive or proprietary information.

**Consideration 2:** Content generated by AI systems may amplify bias due to flawed or skewed training data.

These systems can produce content that is discriminatory, unrepresentative, or laden with biases and stereotypes. Many AI systems are trained on vast amounts of internet data, which is often rife with bias. The widespread or inappropriate use of these systems can amplify or reinforce these biases and dominant viewpoints, leading to a reduction in the diversity of ideas and perspectives.

Checklist:

- ☐ Review any content generated by an AI system to ensure that it aligns with MacEwan University's values and commitments.
- ☐ Review any content generated by an AI system to ensure it meets legal obligations.
- ☐ Learn about bias, diversity, inclusion, anti-racism, and values and ethics to improve your ability to identify biased, non-inclusive, or discriminatory content.
- ☐ Clearly indicate when content has been produced by an AI system.

[20] Secretariat, T. B. O. C. (2024, June 26). *Guide on the use of generative artificial intelligence*. Canada.ca. https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/guide-use-generative-ai.html

- ☐ Formulate prompts to generate content that provides holistic perspectives and minimizes bias, wherever possible.

**Example:** Anti-Cheating Software was Biased

AI bias is not limited to one institution. During the Covid-19 pandemic, schools used anti-cheating software for virtual exams, which relied on video analysis and facial recognition. However, this software often failed to detect non-caucasian faces and penalized students without a stable internet connection or a quiet, private space. It was also biased against students with various disabilities and increased anxiety in those with certain mental health conditions.

Source: VOX

**Consideration 3:** Generated content may be inaccurate, incomplete, or hallucinated.

AI technologies, especially those that generate content, can produce content that appears to be well-developed, credible, and reasonable but is, in fact, inaccurate or hallucinated. Also, content provided by an AI system may not provide a holistic view of an issue, instead, it may focus on the perspectives featured solely in the training data.

Checklist:

- ☐ Clearly indicate that you have used AI to develop content where applicable.
- ☐ Don't consider AI content as authoritative. It must be reviewed for factual and contextual accuracy by checking against trusted sources.
- ☐ Think about the impact associated with using AI in a given context and the risk of inaccurate outputs. Don't use these solutions when factual accuracy is needed.
- ☐ Consider your ability to identify inaccurate content. If you cannot confirm the quality of the content, do not use it.
- ☐ Do not use generative AI systems as search engines unless sources are provided so that you can verify the content.

**Example:** Air Canada Held Liable for Chatbot Giving Bad Advice

In 2022, Air Canada's chatbot promised a discount to a passenger who was assured that he could book a full-fare flight for his grandmother's funeral and then apply for a bereavement fare after the fact.

In early 2024, the British Columbia Civil Resolution Tribunal ruled in favour of the passenger. "It should be obvious to Air Canada that it is responsible for all the information on its website," read tribunal member Christopher Rivers' written response. "It makes no difference whether the information comes from a static page or a chatbot."

Source: BBC

**Example:** AI Hallucinations and Fake Legal Cases

In January 2024, two lawyers discovered fake case law submitted by the opposing lawyer in a civil case in the B.C. Supreme Court.

The lawyer allegedly used ChatGPT to prepare legal briefs in support of a father's application to take his children to China for a visit – resulting in one more or more cases that do not actually exist being submitted to the court.

The lawyer told the court that she was unaware that chatbots like ChatGPT can be unreliable and did not check to see if the cases existed.

Source: Global News

**Consideration 4:** Learn to use AI systems intelligently without stifling creativity or eroding capabilities.

Overreliance on AI solutions is a significant issue that may cause users to uncritically accept system recommendations or outputs, which could be incorrect. Furthermore, excessive dependence on AI solutions can impede employees' ability to develop and maintain the skills needed to perform their tasks, potentially eroding overall workforce capabilities. That said, AI systems are likely here to stay, and when used intelligently, can potentially result in significant increases in productivity and creativity.

Checklist:

- ☐ Consider using AI systems as aids, not substitutes, for skill development.
- ☐ Consider seeing what other AI systems are out there, especially if other AI systems may be better suited to the task at hand (i.e. more task-specific, or more explainable results).
- ☐ Before accepting an AI system's recommendation or output, try using other AI systems and compare outputs.
- ☐ Consider training on using AI systems, including learning how to use prompts to reduce the likelihood of hallucinations.
- ☐ Consider forming your own views before seeking ideas or recommendations from AI systems.
- ☐ Consider whether the outputs of the AI could be improved.

**Consideration 5:** AI systems may infringe on intellectual property or privacy rights.

At this stage, it is unclear whether using copyrighted works as training data for AI falls under fair use. AI systems trained on copyrighted materials might be found in violation of copyright law and personality rights. Often, AI-generated works are not protected by copyright, but their outputs could infringe on these rights if they closely resemble copyrighted content. Additionally, ownership of content created by or with the assistance of generative AI remains uncertain. Privacy rights are also at risk because the data used to train AI systems may include unlawfully collected personal information, such as data from publicly accessible online sources.

Checklist:

- ☐ Consult with Legal about the risks of using generative AI solutions. This could include a review of the supplier's terms of use, copyright policy, privacy policy, and other legal documents.
- ☐ Evaluate system outputs for factual inaccuracies or biases.
- ☐ Consider using generative AI systems for tasks that have a lower likelihood of outputting copyright-protected information, such as revising content that you have already created.
- ☐ Consider using materials where copyright issues are unlikely to be present instead of using an AI system. For example, instead of asking a system to generate an image, source an image from the public domain or pay for a stock image, rather than using a system.

**Example:** Sony Music Warns AI Companies

In May 2024, Sony Music Group warned AI companies and music streaming platforms to not use the company's content without explicit permission.

Sony Music sent letters to more than 700 companies to protect its intellectual property, which includes album cover art, metadata, musical compositions and lyrics from being used for training AI systems.

Source: Bloomberg, STP News

**Example:** NYT and OpenAI Lawsuit

In late 2023, the NYT sued OpenAI over the AI use of copyrighted work. Specifically, the lawsuit claims that millions of copyrighted articles from The New York Times were used to train chatbots that now compete with it.

The Times said that it approached OpenAI and Microsoft in April to raise concerns about the use of its intellectual property, but the talks did not lead to a resolution.

This lawsuit could carry major implications for the news industry.

Source: The New York Times

## **CHECKLIST #2**: Use of procured or embedded third-party vendor systems

This section offers a checklist for those looking to procure a new AI solution or use an existing AI system within an enterprise product (e.g., Microsoft Copilot). Generally, *Checklist #1 will apply here as well*. This section builds on those guidelines with key questions to consider before procuring and using a third-party AI system.

Checklist:

| Dimension | Questions to consider[21] |
|---|---|
| Privacy | ☐ Will you provide any data that could contain personal information or sensitive business information to the AI supplier? If so, have you performed a detailed review of the supplier's privacy program and data governance policies?<br>☐ Does the supplier make any representations about the use of personal information collected during the use of the AI system or feature?<br>☐ How will personal information collected or used throughout the lifecycle of the AI system or feature be safeguarded? |
| Bias and Non-discrimination | ☐ Has the supplier described possible sources of bias in the data used to train and develop the AI system or feature? If so, have they clearly articulated the steps taken to remediate the identified bias?<br>☐ Has the supplier adequately described bias checking procedures, including requirements for periodic bias reviews of the AI system or feature to identify and remediate potential abnormalities?<br>☐ Has the supplier conducted a third-party audit of their AI system or feature to ensure that they are free from bias? |
| Explainability | ☐ Are the explanations provided by the AI system or feature consistent with MacEwan University's expectations?<br>☐ Has the supplier provided guidance and explanations as to how the results from the AI system or feature should be interpreted?<br>☐ Has the supplier highlighted the key inputs to their AI system or features and how they affect the outputs?<br>☐ Is the supplier willing to use additional technology solutions (sometimes referred to as XAI "explainable AI tools") to increase explainability of the AI system or feature in question? |

[21] Adapted from the World Economic Forum Centre for the Fourth Industrial Revolution, Project Fellows From the UK Government's Office for AI, Deloitte, and Salesforce, World Economic Forum Centre for the Fourth Industrial Revolution, & Splunk. (2020). *AI procurement in a box: AI Government Procurement guidelines*. https://www3.weforum.org/docs/WEF_AI_Procurement_in_a_Box_AI_Government_Procurement_Guidelines_2020.pdf.

| | |
|---|---|
| Transparency and Knowledge Transfer | ☐ Is the supplier willing to provide us with documentation on any of the following aspects?<br>○ How to correctly implement and use the AI feature or system.<br>○ Intended, unintended, or restricted uses of the AI feature or system.<br>○ Relevant performance criteria, definitions, or metrics.<br>○ Specifications for (in)appropriate inputs and data for use by the AI system or feature.<br>○ Installation and use instructions.<br>○ Known risks and limitations.<br>○ Malicious or inappropriate uses that may emerge under conditions of foreseeable misuse.<br>○ Risk mitigation procedures, systems, or practices that can be made available to our organization. |
| Safety and Security | ☐ Has the supplier included human oversight and intervention mechanisms? If not, is our organization equipped to perform these duties?<br>☐ Has the supplier obtained any cybersecurity certification (e.g., ISO 27001, SOC 2)?<br>☐ Has the supplier assessed the AI system or feature to ensure that are reliant against malicious actors?<br>☐ Has the supplier described a disaster recovery and continuity plan should the AI system or feature be rendered unavailable? |
| Accuracy and Effectiveness | ☐ Has the supplier described the key performance indicators and performance metrics that the AI system must meet?<br>☐ Has the supplier described an approach to actively monitor and track performance of the AI system or feature to identify degraded performance, accuracy, or effectiveness?<br>☐ Has the supplier provided evidence that relevant experts were involved in the development and testing of the AI system or feature?<br>☐ Has the supplier conducted a third-party audit of the AI system or feature to validate its accuracy end effectiveness? If so, obtain and analyze a copy of the findings report. |
| User Experience | ☐ Has the supplier clearly articulated how the proposed AI system or feature will lead to benefits for users?<br>☐ Does the supplier allow for redress in instance where individuals may be negatively affected?<br>☐ How are updates to the algorithms and data handling practices conducted and communicated to users? |

| | |
|---|---|
| Ethics | ☐ Does the supplier have a track record of providing reputable services?<br>☐ Has the supplier implemented safeguards or mechanisms to prevent users from becoming overly reliant on the system, which could lead to complacency?<br>☐ Has the supplier implemented any of the following?<br>○ AI ethics policy or principles.<br>○ AI risk management policy or procedure.<br>○ Ethical AI development procedures.<br>○ AI impact/risk assessment procedures.<br>○ AI acceptable use policy.<br>○ AI incident response or management procedure. |
| Intellectual Property and Liability | ☐ Does the supplier have a copyright policy to review? If so, obtain a copy and consult with Legal. |
| Regulatory Uncertainty | ☐ Does the supplier make any representations about whether their AI system or feature is compliant with any data, privacy, or AI legislation? Which ones? |

## **CHECKLIST #3:** Designing AI systems

When designing AI systems, project teams should consider both the positive and negative impacts. A formal impact assessment process removes guesswork, ensures accountability, and helps determine the necessary measures, controls, and oversight for AI systems. Identifying potential impacts early in the design stage gives developers time to adjust their approach and address issues proactively.

Risk assessments should include technical, social, ethical, and legal perspectives. This assessment should be conducted when evaluating any new AI system. Here are some suggested questions to consider, aligned with various assessment categories:

**Category 1:** System Assessment

Assess the system's intended use and functionality. Certain applications may carry more inherent risk than others and may warrant specific mitigation measures and considerations.

Questions to consider:

- ☐ Will the AI system replace human decision-making?
- ☐ Will the AI system interact directly with an individual?
- ☐ Will the AI system be used to render decisions autonomously, without human oversight?
- ☐ Will the AI system operate in a sensitive domain (e.g., employment decision-making, biometric information processing, etc.)?

**Category 2:** Transparency and Explainability

Ensure that user-centric considerations are embedded into the design of the AI system. Support internal and external users with explanations to understand how the AI system operates.

Questions to consider:

- ☐ What techniques will be used by the AI system (e.g., rules-based, machine learning, statistical approaches)?
- ☐ What explanations will be given to individuals (i.e., both internal and external to MacEwan University)?
- ☐ Is the system's technique compatible with the required level of explainability (e.g., if MacEwan University is using a black-box approach, will it be able to meet explainability requirements)?
- ☐ Where necessary, can system outputs be translated into plain-language explanations?
- ☐ What will we be communicating externally to impacted users of the AI system?

**Category 3:** Accountability and Organizational Readiness

Deploying AI systems almost always has organizational impacts that need to be considered early in the process. These impacts range from developing new policies and procedures, personnel training and re-skilling, and governance requirements.

Questions to consider:

- What human oversight and override mechanisms will be implemented?
- Is it clearly understood who is responsible for the safe and continuous maintenance, operation, re-training, and decommissioning of the AI system?
- Will the AI system affect current employee roles and responsibilities? Will it lead to redundancies?
- Will MacEwan University staff need additional training to use and/or evaluate outputs from the AI system?
- Are MacEwan University's existing policies and procedures adequate to authorize and regulate the safe and effective operationalization of the contemplated AI system? If not, do existing policies need to be modified?
- Does the project team have the right talent mix to manage the system internally?

**Category 4:** Positive Impacts

Assessing positive impacts will enable more effective cost-benefit analysis. If the positive return is low and the risks are high, organizations should question whether deploying such a system is warranted.

Questions to consider:

- How will the AI system improve our operations? Will it enhance MacEwan University's competitive advantage?
- How will this AI system help MacEwan University's employees do their jobs better?
- Is the value that the AI system intended to provide clearly understood and defined?
- What are the expected efficiency or productivity gains?
- Will the AI system be reviewed at a later date to determine whether the efficiency and/or productivity gains were achieved?

**Category 5:** Reputational Impact

There are ongoing public distrust issues with the use of AI. If an AI system either does not operate as intended, or is perceived not to operate as intended (e.g., biased outcomes or high error rates) this can lead to significant reputational harm.

Question to consider:

> If the AI system does not operate as intended, could it lead to negative media coverage or impact the trustworthiness of MacEwan University?

**Category 6:** Legal Impacts

There are often regulatory issues that must be considered, such as data protection legislation, privacy laws, and liability risks. Often these, along with inappropriate operations of AI systems can lead to significant reputational impacts.

Questions to consider:

- What laws and regulations apply to this AI system? Are there any compliance requirements that must be met and documented?

- Has a legal assessment been conducted to identify liabilities and other legal issues, such as intellectual property, privacy, employment, and human rights concerns?

**Category 7:** Risk of Harms

It is critical to analyze and identify the known and foreseeable risks associated with the system. When looking at harm, consider the risks and impacts across a broad spectrum of possibilities and groups.

Questions to consider:

- If the system does not operate as intended, could it negatively impact end-users? If so, how? What safeguards should MacEwan University be putting into place to mitigate this?
- Could the AI system result in harm or damage to the physical or psychological well-being of individuals or society?
- Could the output from the AI system result in denying services to an individual?
- Could the output from the AI system result in discriminatory or biased outcomes?
- What is the likelihood of harm from an error in system output?
- What is the severity of harm from an error in system output?

**Category 8:** Data Quality

AI systems are a by-product of the data they are exposed to and trained on. A critical component of designing robust AI systems stems from ensuring the data used to develop them is of sufficient quality.

Questions to consider:

- Have data quality standards been established and documented?
- Have potential data gaps and shortcomings been identified and remediated?
- Are there audit logs in place to ensure traceability of the system design decisions and functionality throughout its lifecycle?
- Are there automatic triggers in place to alert system operators of misbehaviour or erratic performance?
- Are system tests performed at regular intervals?

## **CHECKLIST #4:** Guidelines for hiring third-party consultants that use AI

In general, the guidelines discussed throughout this section should also apply to third-party consultants that wish to use AI systems to provide services to our organization. Specific emphasis should be given to Sections 1 and 3 (i.e., "Use of publicly available AI consumer or general-purpose AI solutions" and "Designing AI systems").

Third-party consultants are permitted to use AI systems, including general-purpose AI systems, in compliance with the following criteria outlined below (illustrative, not exhaustive):

- ✓ Using an AI system to proofread or conduct an initial draft of a social media message or blog/news post, provided that the draft has been checked for accuracy and approved by the appropriate MacEwan University personnel.
- ✓ Using an AI system to summarize non-confidential information that is free from any personal information.
- ✓ Using an AI system to create a transcript of a video call or discussion, provided there is appropriate disclosure and that applicable privacy laws are respected.
- ✓ Using an AI-powered feature made available through an existing system or software solution (e.g., Microsoft Office, Adobe) that has been already approved for use by MacEwan University.
- ✓ Using an AI system to analyze de-identified datasets to uncover trends and potential areas of need or improvement.

Questions to consider:

- ☐ Will the Consultant take responsibility for any AI-generated content used in service delivery (i.e., hallucinations, IP infringement, etc.)?
- ☐ Will the Consultant ensure that it verifies any facts, conclusions or decisions made with the assistance of AI systems?
- ☐ Does the Consultant have sufficient insurance to cover any damages that might result from their use of AI systems?
- ☐ Will the Consultant ensure that no third parties retain personal or other confidential information due to the Consultant's use of AI systems?