Generative AI Guidelines for institutional use

1 Introduction

The use of Generative AI at the College is rapidly evolving, and its application use cases reaches all areas of the College. At the same time, many risks are associated with using Generative AI (GenAI).

In response to the rapidly evolving use of GenAI tools at the College, the College Technologies Committee (CTC) has approved the development of guidelines specifically focusing on the institutional use of Generative AI tools. The development of this guideline aligns with the approach to GenAI guidelines by multiple higher education institutions in Canada and the United States.

The guidelines also align with the risks associated with the use of GenAI as identified by the Canadian Centre for Cyber Security.

2 Purpose

The purpose of Algonquin College's Generative Artificial Intelligence guideline is to ensure the legal, ethical, and secure institutional use of GenAI technology by the College.

The guidance below outlines best practices when using or developing GenAI models and applications by college employees for business purposes. The use of Generative AI for educational and academic purposes by faculty and learners will be addressed through a specific guideline published at a later date.

The use of Generative AI at the college is rapidly evolving, and so are the available tools. This guideline is expected to enable and raise awareness regarding the safe use of Generative AI within the College.

3 Definitions

Term	Description
Artificial Intelligence	The theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.
Generative AI (GenAI)	Artificial intelligence capable of generating text, images, videos, or other data using generative models, often in response to prompts. Generative AI models learn the patterns and structure of their input training data and then generate new data that has similar characteristics.

<table>
<tr><th>Term</th><th>Description</th></tr>
<tr><td>Data</td><td>Facts and statistics collected for reference or analysis.</td></tr>
<tr><td>Information</td><td>Facts provided or learned about something or someone.</td></tr>
<tr><td>Information Sensitivity Levels</td><td>Information Sensitivity Levels categorize data based on its value and importance to the college. The level of classification of information per the College standard includes labels “Public”, “Internal”, “Confidential” or “Restricted”.

Two widely used models are shown below.

<table>
<tr><th>SENSITIVITY</th><th>MODEL 1</th><th>MODEL 2</th></tr>
<tr><td>High</td><td>Confidential</td><td>Restricted</td></tr>
<tr><td>Medium</td><td>Internal Use Only</td><td>Sensitive</td></tr>
<tr><td>Low</td><td>Public</td><td>Unrestricted</td></tr>
</table>
</td></tr>
<tr><td>Information Technology (IT) Resources</td><td>These resources may be broadly classified into four categories, namely, hardware, soft-ware, data and human resources. It is essential for a manager to be aware of the features and significance of each of these resources. Such awareness will not only help the manager to plan the IT infrastructure but also to assess the cost of the IT infrastructure.</td></tr>
<tr><td>Information Technology (IT) System</td><td>An information technology system (IT system) is generally an information system, a communications system, or, more specifically speaking, a computer system — including all hardware, software, and peripheral equipment — operated by a limited group of IT users, and an IT project usually refers to the commissioning and implementation of an IT system.</td></tr>
<tr><td>Personal Health Information (PHI)</td><td>Identifying information about an individual relating to the physical or mental health of the individual or the provision of health care. Where it is held for purposes related to the provision of health care, the Ontario Personal Health Information Protection Act (PHIPA) governs the College's collection, use and disclosure of the information.</td></tr>
</table>

Term	Description
Personally Identifiable Information (PII)	Any data that could be used to identify a specific individual, including but not limited to the individual's name, home addresses and email addresses, telephone numbers, age, sex, marital or family status, identifying number, race, national or ethnic origin, color, religious or political beliefs or associations, educational and medical history, disabilities, blood type, employment history, financial history, criminal history, anyone else's opinions about an individual, an individual's personal views or opinions, and name, address and phone number of parent, guardian, spouse or next of kin. Also called personal information.
Payment Card Industry (PCI) Compliance	The technical and operational standards that businesses follow to secure and protect **credit card data** provided by cardholders through card processing transactions.
Application Programming Interface (API)	A set of rules or protocols that enables software applications to communicate with each other to exchange data, features and functionality.
User	A person who uses or operates something, especially a computer or other machine.
Login Credentials	A set of unique identifiers–such as a username and password–that enables a user to verify identity to log in to an online account.
User ID	A logical entity used to identify a user on a software, system, website or within any generic IT environment.

4 Risks of Generative AI Use

Here are five of the top risks of GenAI. (Dilmegani, 2024)

Accuracy risks of generative AI

GenAI tools rely on large language models that are trained on massive data. To answer a question or to create a response to a certain prompt, these models interpret the prompt and induce a response based on their training data. Although their training data sets consist of billions of parameters, they are finite pools and the generative models may "make up" responses from time to time.

There can be many potential accuracy risks caused by GenAI models:

- **Generalization over specificity:** Since generative models are designed to generalize across

the data they're trained on, they may not always produce accurate information for specific, nuanced, or out-of-sample queries.

- **Lack of verification:** Generative models can produce information that sounds plausible but is inaccurate or false. Without external verification or fact-checking, users might be misled.

- **No source of truth:** GenAI doesn't have an inherent "source of truth". It doesn't "know" things in the way humans do, with context, ethics, or discernment. It's generating outputs based on patterns in data, not a foundational understanding.

Bias risks of generative AI

GenAI's potential for perpetuating or even amplifying biases is another significant concern. Like accuracy risks, as generative models are trained on a certain dataset, the biases in this set can cause the model to also generate biased content.

Some bias risks of GenAI are:

- **Representation bias:** If minority groups or viewpoints are underrepresented in the training data, the model may not produce outputs that are reflective of those groups or may misrepresent them.

- **Amplification of existing biases:** Even if an initial bias in the training data is minor, the AI can sometimes amplify it because of the way it optimizes for patterns and popular trends.

Data privacy & security risks of generative AI

GenAI technology, especially models trained on vast amounts of data, poses distinct risks concerning the privacy of sensitive data. Here are some of the primary concerns:

1. **Data leakage:** Even if an AI is designed to generate new content, there's a possibility that it could inadvertently reproduce snippets of training data. If the training data contained sensitive information, there's a risk of it being exposed.

2. **Personal data misuse:** If GenAI is trained on personal customer data without proper anonymization or without obtaining the necessary permissions, it can violate data privacy regulations and ethical standards.

3. **Data provenance issues:** Given that generative models can produce vast amounts of content, it might be challenging to trace the origin of any specific piece of data. This can lead to difficulties in ascertaining data rights and provenance.

Intellectual property risks of generative AI

GenAI poses various challenges to traditional intellectual property (IP) norms and regulations. Also, there are concerns around the eligibility of the AI generated content for copyright protection and

infringement.

Some of the primary risks and concerns of GenAI around intellectual property are:

- **Originality and ownership:** If a GenAI creates a piece of music, art, or writing, who owns the copyright? Is it the developer of the AI, the user who operated it, or can it be said that no human directly created it and thus it's not eligible for copyright? These are problematic concepts when talking about AI generation.

- **Licensing and usage rights:** Similarly, how should content generated by AI be licensed? If an AI creates content based on training data that was licensed under certain terms (like Creative Commons), what rights apply to the new content?

- **Infringement:** Generative models could unintentionally produce outputs that resemble copyrighted works. Since they're trained on vast amounts of data, they might inadvertently recreate sequences or patterns that are proprietary.

- **Plagiarism detection:** The proliferation of AI-generated content can make it more challenging to detect plagiarism. If two AI models trained on similar datasets produce similar outputs, distinguishing between original content and plagiarized material becomes complex.

Ethical risks of generative AI

Over the years, there has been a significant discourse on AI ethics. However, the ethical debate specifically surrounding GenAI is comparatively recent. This conversation has gained momentum with the introduction of various generative models, notably ChatGPT and DALL-E from OpenAI.

- **Deepfakes:** The biggest ethical concern around GenAI is deepfakes. Generative models can now generate photorealistic images, videos and even sounds of persons. Such AI generated content can be difficult or impossible to distinguish from real media, posing serious ethical implications. These generations may spread misinformation, manipulate public opinion, or even harass or defame individuals.

- **Erosion of human creativity:** Over-reliance on AI for creative tasks could potentially diminish the value of human creativity and originality. If AI-generated content becomes the norm, it could lead to homogenization of cultural and creative works.

- **Unemployment impact:** If industries heavily adopt GenAI for content creation, it might displace human jobs in areas like writing, design, music, and more. This can lead to job losses and economic shifts that have ethical implications.

- **Environmental concerns:** Training large generative models requires significant computational resources, which can have a substantial carbon footprint. This raises ethical questions about the environmental impact of developing and using such models.

5 Guidelines and recommended practices

For any use of GenAI applications, employees are recommended to adhere to the following:

5-1 Adherence to AC's existing policies

5-1-1 All employees should comply with applicable laws, regulations, and College Policies, directives and guidelines while using GenAI tools and generated content.

5-1-2 All employees use of GenAI tools should comply with College Business Code of Conduct and Non-Discrimination Policies.

5-1-3 The College will continue to refine and enhance its policies, guidelines and processes as the comprehension deepens regarding the use of GenAI applications and related impacts.

5-2 Transparency

5-2-1 Any use of GenAI tools for generation of (1) publicly facing content or (2) strategic or important business decisions, and (3) that is not modified prior to its use should be:

a. Disclosed through appropriate transparency disclosures, including but not limited to formal footnote citations, and acknowledgment of GenAI use statements…

b. Logged in an updated record of instances of GenAI use for work purposes and be able to share those records upon request. This record should at least include the date and time of use, GenAI tool used, disclosed input, and output, instances of content used such as website, courses, or deliverables.

5-3 Ethical use and information accuracy

5-3-1 All employees should always review content generated though GenAI tools is accurate through additional due diligence and verifications.

5-3-2 All employees should consider the ethical implications of using GenAI applications by thoroughly assessing potential biases, ethical and legal aspects, and data collection practices.

5-3-3 All employees are responsible for any content generated by GenAI tools used for College business. Recommended verifications include:

a. Reviewing output of GenAI applications to make sure it meets College standards for principles of equity, ethics, and appropriateness.

b. Ensuring that the output does not discriminate against individuals based on race,

colour, religion, sex, national origin, age, disability, marital status, political affiliation, or sexual orientation.

c. Preventing the use of GenAI applications to create text, audio, or visual content for purposes of misrepresenting an individual's identity or to commit fraud.

5-4 Generative AI Literacy

5-4-1 All employees are encouraged to commit to the ongoing development of their critical GenAI literacy. This entails critically assessing the capabilities and constraints of GenAI applications and making informed decisions when integrating them into their processes and practices.

5-4-2 All employees should continue to stay aware of GenAI applications risks and take proactive measures to mitigate them.

5-5 Data Privacy and Security

5-5-1 To maintain the security of our data and IT systems, all employees are required to prioritize using the approved GenAI tool(s) when using College systems or networks if possible. The College approved GenAI tools list is available via this link.

5-5-2 To avoid additional exposure of the College to cybersecurity risks, all employees are recommended to:

a. Not use College credentials, email addresses, or telephone numbers as a login to publicly available GenAI tools that are not approved by the College.

b. Review and approve code generated by GenAI via the appropriate process before using it on College Information Systems.

c. Not install non-College approved GenAI Application Programming Interfaces (APIs), plug-ins, connectors, or software on college equipment.

5-5-3 To maintain the confidentiality of the College's sensitive information, including but not limited to all employees, faculty and learners' information, intellectual property, copyrighted material:

a. Apply the principle of privacy-first when deciding what information is entered into GenAI applications.

b. Do not input College intellectual property into generative AI applications.

c. Do not enter Personal Health Information (PHI), Personally Identifiable

Information (PII) and Payment Card Information (PCI) in GenAI tools.

d. Adhere to section 5 regarding the recommended use of GenAI tools per information classification levels.

5-5-4 All employees should ensure to change the default data controls of GenAI Tools so that data is not archived and saved in their databases. Information shared with Generative AI tools using default settings could expose proprietary or sensitive information to unauthorized parties.

5-6 Consulting College stakeholders on GenAI use

5-6-1 All employees are encouraged before procuring GenAI tools to comply with the College's procurement processes and practices. As a result, college stakeholders should be consulted before acquiring GenAI tools, including Procurement, ITS, and Risk Management.

5-6-2 Furthermore, the College is working to ensure that GenAI Applications acquired uphold the necessary protections and controls.

6 Recommended usage of Generative AI Tool per information classification levels

6-1 The table below shows the recommended usage of Generative AI tools, both college-offered and publicly available tools, including the level of information classification (initial guidance accessible here: Link) they are approved for use.

GenAI Tools	Description	Availability	Approved information classification levels
College Approved GenAI Tools	GenAI tools that have been reviewed and approved by ITS and made available to the College. The list of college approved GenAI tools is available via this link.	College-offered using Algonquin College login credentials	• Public • Internal • Data not containing PII, PHI or PCI.
OpenAI tools (ChatGPT, Dall-e, Sora…)	AI chatbot primarily text-based, with additional functionalities for generating content such as images, audio, and videos.	Publicly available using personal login credentials	• Public

GenAI Tools	Description	Availability	Approved information classification levels
Other GenAI Tools	Other Generative AI Tools that are publicly available and not approved by ITS.	Publicly available using personal login credentials	• Public

7 Examples of GenAI Use Cases

Below are examples of use cases, which are subject to modification. The list below does not reflect all possible use cases and scenarios. If your use cases meet the criteria of more than one classification, follow the principles outlined in this guideline.

7-1 Unacceptable use cases

7-1-1 Generating or sharing any form of sensitive data, including but not limited to personally identifiable information (PII), Personal Health Information (PHI), and financial information, regardless of whether it pertains to learners, employees, research participants, or patients.

7-1-2 Submitting information classified Confidential or Restricted as input in any GenAI Tools.

7-1-3 Employing generative AI to impersonate college officials or departments in email communications or social media posts with the intent to disseminate false information or perpetrate scams undermines trust and credibility within the college community.

7-1-4 Generating academic papers, assignments, or research publications without complying with specific College policies, directives, and guidelines.

7-1-5 Generating code or scripts designed to exploit vulnerabilities in college systems or launch cyberattacks, such as malware or ransomware, poses severe security threats and could result in significant disruption and harm to college operations.

7-1-6 Using GenAI tools to interpret the college's policies and directives.

7-2 Acceptable use cases

7-2-1 Generating content using publicly available College information, including the information available via the public college website.

7-2-2 Generating and outlining, through college approved GenAI tools, course syllabi and lesson plans, receiving suggestions for learning objectives, teaching strategies, and

assessment methods.

7-2-3 Generating drafts of correspondence using approved GenAI tools by using fabricated information (such as an invented name for the recipient of an email message), while adhering to this guideline.

7-2-4 Drafting initial materials for potential professional development opportunities, including workshops, conferences, and online courses relevant to their field.

7-2-5 Drafting event and project plans by identifying themes, activities, tasks, timelines, and checklists.

RELATED DOCUMENTS

AA35 – Confidentiality of Student Records
HR18 – Employee Code of Conduct
IT01 – Information Security Policy
IT01 – Appendix 1 – Algonquin College Confidentiality Agreement for Employees and Contractors

References

Dilmegani, C. (2024, January 2). *5 Risks of Generative AI & How to Mitigate Them in 2024.* From AIMultiple Research: https://research.aimultiple.com/risks-of-generative-ai/

DOCUMENT CONTROL

Changed by	Date of change	Description of change
A. Lefriyekh	June 18th 2024	• Added Unacceptable use case 7.2.6