Artificial Intelligence Policy Discussion Briefing Report

Executive Summary

This report aims to inform stakeholders about the responsible use of Artificial Intelligence (AI) within the City of Fredericton. It provides an overview of AI technologies, their potential benefits, and the associated risks. Additionally, it outlines existing federal, provincial, and municipal regulatory frameworks and policies relevant to AI, and presents draft policy statements to guide the ethical and effective implementation of AI in municipal operations.

Key Points:

1. **AI and LLM Overview**:
    - AI refers to the capability of machines to perform tasks that typically require human intelligence, such as problem-solving, learning, reasoning, perception, and language understanding.
    - Large Language Models (LLMs) are a type of AI designed to understand and generate human language, trained on vast amounts of text data.
2. **Opportunities**:
    - AI and LLMs can transform municipal operations by enhancing document generation, policy analysis, internal knowledge management, data analysis, automated customer service, public engagement, crisis management, translation services, and accessibility.
3. **Risks**:
    - AI presents risks such as generating inaccurate information, quality control issues, bias and discrimination, cultural insensitivity, privacy breaches, security vulnerabilities, dependency, ethical and moral concerns, transparency issues, environmental impact, public perception, and copyright risks.
4. **Regulatory and Policy Frameworks**:
    - **Federal Regulations**: The Consumer Privacy Protection Act (CPPA) and the Artificial Intelligence and Data Act (AIDA) aim to ensure ethical AI use, enhancing transparency, accountability, and privacy protection.
    - **Provincial Regulations**: The New Brunswick Right to Information and Protection of Privacy Act (RTIPPA) provides a framework for safeguarding personal information and ensuring privacy.
    - **Municipal Policies**: Existing policies address various AI-related risks, but a clear and comprehensive AI policy would further enhance the City's ability to manage AI technologies responsibly.
5. **Draft Policy Statements**:
    - These include guidelines for the permitted use of AI, privacy and data security, AI-assisted translation, human review of AI-generated content, development of municipal AI, prohibition of social credit scoring, adoption of third-party AI technologies, and quality assurance.
6. **Responsible Use of AI**:

    - Defines responsible use and outlines principles such as ethical use, transparency, privacy protection, bias mitigation, and accountability.
    - Emphasizes achieving responsible use through training, stakeholder engagement, continuous monitoring, feedback mechanisms, and enforcement.
7. **Roles and Responsibilities**:
    - Outlines the responsibilities of the Municipal Council, Chief Information Officer (CIO), Chief Administrative Officer, department heads, and municipal staff in ensuring the ethical use of AI.

Introduction

This report aims to inform stakeholders about Artificial Intelligence (AI) and Large Language Models (LLMs), providing sufficient information on context, benefits, and risks to make informed policy decisions. The goal is to aid in developing a municipal policy for the responsible use of AI. By understanding the intricacies of AI and LLMs, stakeholders can ensure these technologies are integrated in a manner that enhances municipal operations while safeguarding public trust and compliance with ethical standards.

What is Artificial Intelligence (AI) and How Does it Work?

Artificial Intelligence (AI) refers to the capability of machines to perform tasks that typically require human intelligence, such as problem-solving, learning, reasoning, perception, and language understanding. AI systems can recognize patterns, make decisions, and predict future outcomes based on the data they process. Over the past decade, AI capabilities have significantly advanced, allowing machines to perform complex tasks like image recognition, translation, and creative content generation—tasks that previously required human intelligence.

AI systems are increasingly used to make critical predictions and decisions about people in areas such as credit evaluation, hiring, and digital services. In Canada, AI is being developed and utilized across various applications that contribute to the economy and enhance the lives of Canadians. Technologies that once seemed unthinkable are now part of everyday life, offering numerous benefits including:

- **Healthcare**: Advancements in cancer screenings and improvements in at-home healthcare services.
- **Agriculture and Energy**: Enhanced precision harvesting and improved efficiency in the energy supply chain.
- **Smart Products and Services**: Introduction of new smart products and personalized services.
- **Language Processing**: Increased capabilities in translation and text-to-speech technologies.
- **Information Access**: Improved ability for citizens to find and process information.

AI's rapid development and integration into various sectors demonstrate its transformative potential, making it an essential component of modern technology and daily life.

What are Large Language Models (LLMs)?

Large Language Models (LLMs) are a type of AI designed to understand and generate human language. Examples include ChatGPT and CoPilot. These models are trained on vast amounts of text data from books, websites, and other written sources to learn the patterns and structures of human language. They can generate coherent and contextually relevant text based on the input they receive.

How are LLMs Trained and How Do They Work?

LLMs are trained using a method called "deep learning," which involves feeding them massive datasets containing diverse language examples. This training helps the models learn grammar, facts about the world, reasoning abilities, and some level of common sense.

LLMs ***predict*** what comes next in a sentence by analyzing the context provided by the preceding words. For example, given the phrase "The weather today is," an LLM might predict the next word to be "sunny," "rainy," or "cold" based on patterns it has learned from its training data.

Definitions

- **Artificial Intelligence (AI)**: The simulation of human intelligence processes by machines, especially computer systems.
- **Large Language Models (LLMs)**: Advanced AI systems trained on large datasets to understand, generate, and interact in human language.
- **Bias**: Systematic error in AI and LLMs that results in unfair outcomes, particularly those that disadvantage certain groups.
- **Transparency**: The degree to which the operations and decision-making processes of AI and LLMs are open and understandable to stakeholders.
- **Accountability**: The responsibility of individuals and organizations to ensure that AI and LLMs are used ethically and in compliance with relevant laws and standards.
- **Personally Identifiable Information (PII)**: Any data that can be used to identify an individual, including but not limited to names, addresses, phone numbers, email addresses, social insurance numbers, health records, financial information, and any other information that alone or in combination can identify a specific individual.

Public AI (Open AI)

Public AI, also known as Open AI, refers to AI tools and models that are accessible over the internet and operated by third-party providers, such as ChatGPT and CoPilot. These models are trained on large datasets collected from public sources and are available to a broad audience, including external entities. Key characteristics include:

- **Accessibility**: Available to anyone with internet access, making it easy for a wide range of users to utilize these AI tools.

- **Data Sources**: Trained on publicly available data, which is not controlled by the end user. The specifics of the training data are typically not disclosed.
- **Data Privacy**: Information and chat inputs entered into these systems may become part of the training data, potentially being used in future iterations and becoming part of the public domain.
- **Third-Party Management**: Operated by external organizations, which means the end user has limited control over the data and model management.

Private AI

Private AI refers to AI and Large Language Models (LLMs) that are developed, managed, and operated exclusively by an organization, such as the City of Fredericton, using internal data and information sets. Examples include the use of Microsoft Azure AI services for private organizational applications. Key characteristics include:

- **Accessibility**: Restricted to the organization and its authorized users, ensuring controlled access.
- **Data Sources**: Trained on data that is internally controlled by the organization, allowing for compliance with privacy regulations and tailored relevance to organizational needs.
- **Data Privacy**: Inputs and outputs are contained within the organization's data systems and controls, ensuring that they do not become part of the public domain and are kept private and secure.
- **Internal Management**: Managed by the organization itself, providing full control over the data, training processes, and usage of the AI models.

Third-Party AI Applications

Third-Party AI Applications are software applications or services that use AI within the application or service, which are not open-source AI models but specific to an application or service. Key characteristics include:

- **Accessibility**: These applications are provided by external vendors and are used within specific software or service contexts.
- **Data Sources**: The data on which the AI is trained is not disclosed to the end user, and the municipality does not have insight into the training data or the model's workings, often referred to as a "black box."
- **Data Privacy**: The responsibility for data privacy and compliance lies with the external vendor, who must ensure adherence to relevant regulations and best practices.
- **Third-Party Management**: The AI models and applications are managed by external vendors, limiting the municipality’s control over the AI’s training and operations.

Opportunities

Artificial Intelligence (AI) and Large Language Models (LLMs) can be transformative for municipalities, offering a range of opportunities to enhance services, improve operations, generate content and insights, and increase efficiency for staff.

Document Generation and Management

- **Application**: Use AI tools like CoPilot and ChatGPT to draft routine documents, reports, and correspondence such as meeting minutes, agenda items, newsletters, resolutions, regulations, policies, and grant applications. LLMs can reference previously written documents or those produced in other comparable municipalities, which is particularly useful for creating consistent and accurate formulaic texts.
- **Benefits**: Streamlines document creation processes, saves time for staff, ensures consistency in formatting and language, and enhances the quality and accuracy of documentation.

Policy, Plan, and Strategy Analysis and Development

- **Application**: Employ AI to analyze existing policies, plans, and strategies by training it on the City's current documents. AI can compare these with best practices, suggest improvements, and evaluate new initiatives in the context of existing plans. Additionally, AI can summarize key points of existing corporate strategic documents, providing concise and comprehensive overviews.
- **Benefits**: Enhances the quality and effectiveness of policy-making, ensures policies are up-to-date with current standards, supports evidence-based decision-making, and streamlines the analysis of new initiatives in alignment with existing strategies.

Internal Knowledge Management

- **Application**: Implement private AI to organize and manage internal knowledge bases, making it easier for staff to find, understand, and share information and resources. LLMs can be trained to organize and present municipal information such as policies and procedures, contracts, records, property files, bylaws, resolutions, and zoning regulations in a user-friendly manner, assisting in intelligently navigating, searching, and summarizing this information.
- **Benefits**: Enhances collaboration and knowledge sharing among departments, reduces time spent searching for information, improves overall efficiency, and provides staff with easy access to well-organized municipal information.

Data Analysis and Reporting

- **Application**: Leverage AI for data analysis tasks such as compiling statistics, generating insights from large datasets, and creating visual reports. LLMs can analyze real-time data, enabling public officials to make informed decisions, perform predictive analysis, and develop evidence-based policies.
- **Benefits**: Improves the accuracy and efficiency of data analysis, aids in strategic planning, supports data-driven decision-making, and enhances the effectiveness and transparency of public administration.

Automated Customer Service

- **Application**: Implement ChatGPT as a virtual assistant on the City's website and social media channels to handle common inquiries, provide information, and guide residents to appropriate services. LLM-powered virtual assistants can provide round-the-clock support, addressing citizens' common queries and guiding them through various processes.
- **Benefits**: Reduces the workload on customer service representatives, provides instant responses to public queries, and ensures 24/7 availability. This improves citizen satisfaction and reduces the workload on government staff. Additionally, LLMs can offer accurate and timely information on services, regulations, permits, and public resources, providing personalized guidance and simplifying interactions between the government and citizens.

Public Engagement and Participation

- **Application**: Use AI to analyze public feedback from surveys, social media, and other channels to gauge public opinion and identify key issues.
- **Benefits**: Provides valuable insights into community needs and concerns, supports more responsive and inclusive governance, and helps prioritize city initiatives.

Crisis Management and Communication

- **Application**: Utilize AI to draft and disseminate emergency communications and updates in multiple languages quickly during crises such as natural disasters or public health emergencies.
- **Benefits**: Ensures timely and accurate information dissemination, enhances public safety, and supports coordinated emergency responses.

Translation Services

- **Application**: Use AI to translate documents and communications, reaching a broader audience, including non-English speaking residents. LLMs can support multilingual communication, enabling the quick translation of working documents and correspondence, and the preparation of drafts of official City documents in French.
- **Benefits**: Improves accessibility of city services, ensures compliance with bilingual requirements, and enhances community inclusiveness. The ability to quickly translate documents facilitates efficient communication and ensures that all residents, including those from marginalized communities, have access to necessary information.

Accessibility and Inclusivity

- **Application**: Leverage AI to describe photos in detail, enhancing accessibility for the visually impaired. AI can generate accurate and comprehensive descriptions of images, which can be implemented in various applications such as social media, educational materials, and public information services.

- **Benefits**: Provides visually impaired individuals with a clearer understanding of visual content, ensuring they can access and interpret visual information more independently and effectively. This enhances inclusivity and ensures that all citizens, regardless of physical ability, can benefit from city services and information.

RTIPPA (Right to Information and Protection of Privacy Act) Handling

- **Application**: AI offers the potential to transform the handling of RTIPPA requests by automating the time-consuming tasks of document processing and redaction. Utilizing Natural Language Processing (NLP) and machine learning, AI can swiftly identify, classify, and extract relevant information from large volumes of documents while accurately redacting personal data.
- **Benefits**: This automation speeds up response times, reduces the workload on staff, enhances the accuracy of redactions, and strengthens the protection of personal information, ultimately making the process more efficient and reliable.

Risks

Artificial Intelligence (AI) leverages advanced algorithms, including Large Language Models (LLMs), to predict text, generate content, and make context-based recommendations. While AI offers substantial opportunities, it also presents various risks.

Inaccurate Information ("Hallucinating" or "Confidently Incorrect")

- **Risk**: LLMs can generate information that seems plausible but is factually incorrect. This can lead to misinformation if the generated content is relied upon without verification.

Quality Control

- **Risk**: Maintaining consistent quality in AI-generated content can be challenging, particularly as AI systems may produce variable outputs depending on the data and context.

Bias and Discrimination

- **Risk**: LLMs learn from the data they are trained on, which can include biased or discriminatory content. This can result in the models perpetuating and amplifying these biases in their responses.

Cultural Sensitivity

- **Risk**: AI systems may not adequately account for local cultural norms and values, leading to insensitive or inappropriate outputs.

Privacy and Data Security

- **Risk**: Public LLMs continuously train on new data, which can include sensitive or personal information. Using these models without careful management can lead to privacy breaches and unauthorized use of data.

Security Vulnerabilities

- **Risk**: AI systems can be targeted by cyber-attacks. Hackers could exploit vulnerabilities to manipulate AI outputs or access sensitive data, leading to potential security threats.

Legal and Regulatory Compliance

- **Risk**: The rapidly evolving nature of AI technology often outpaces existing regulations, creating uncertainty and potential legal challenges regarding the appropriate use of AI.

Dependency and Loss of Skills

- **Risk**: Over-reliance on AI tools can lead to a decline in critical thinking and decision-making skills among staff, reducing their ability to function effectively without AI assistance.

Ethical and Moral Considerations

- **Risk**: AI systems can be involved in making decisions that have significant moral implications, such as allocating resources or prioritizing services, which can raise ethical concerns.
- **Social Credit Scores**: AI systems are increasingly used to make important predictions or decisions about people, such as credit, hiring, and digital services. There is a risk that these systems could be used to create social credit scores, evaluating individuals based on their behavior and interactions. This can lead to ethical concerns, discrimination, and a lack of transparency in how these scores are calculated and used.
- **Deepfakes and Misinformation**: AI can be used to create deepfakes, which are highly realistic but fake audio and video content. These can be used to spread misinformation, deceive the public, and manipulate opinions, posing significant risks to trust and security.

Transparency and Accountability

- **Risk**: AI systems often function as "black boxes" with decision-making processes that are not easily interpretable by humans, making it difficult to understand and justify AI decisions.

Environmental Impact

- **Risk**: Training large AI models requires significant computational power and energy, contributing to environmental concerns such as an increased carbon footprint.

Public Perception and Trust

- **Risk**: Public misunderstandings about AI capabilities and limitations can lead to unrealistic expectations or mistrust in AI systems, affecting public perception and trust.

Copyright Risks

- **Risk**: AI-generated content might inadvertently reproduce copyrighted material without proper attribution, leading to potential legal issues. This risk is heightened because LLMs are trained on vast datasets that may include copyrighted works.

Specific Risks for Municipalities

Inaccurate Data Provision

- **Risk**: Providing incorrect information to residents and businesses can have legal and economic repercussions, such as misguided decisions based on faulty data.

Reputation Damage

- **Risk**: Biases and inaccuracies in AI-generated content can harm the municipality's reputation, leading to public distrust.

Privacy Breaches

- **Risk**: Unauthorized disclosure of personal data can cause significant harm to individuals and result in legal consequences for the municipality.

Lack of Cultural and Linguistic Nuances

- **Risk**: LLMs may not fully capture the cultural and linguistic nuances of the region, leading to miscommunications.

Inconsistent Branding

- **Risk**: AI-generated content might not align with the city's branding guidelines, affecting the consistency and professionalism of official communications.

Legal Implications of Copyright Violations

- **Risk**: The use of AI-generated content that infringes on copyrighted material can lead to legal disputes and financial penalties, impacting the municipality’s legal standing and finances.

Trust and Transparency

- **Risk**: The use of AI to evaluate residents based on their behavior and interactions could lead to the development of social credit scores. This raises significant ethical concerns, including discrimination, lack of transparency, and potential misuse of power.
- **Deepfakes and Misinformation**: The creation and dissemination of deepfakes can undermine public trust, spread false information, and cause harm to individuals and the community.

Screening Systems Impacting Access to Services or Employment

- **Risk**: AI systems making decisions related to credit, employment, or other services can produce discriminatory outcomes and economic harm, particularly affecting women and historically marginalized groups.

Biometric Systems Used for Identification and Inference

- **Risk**: AI systems using biometric data to identify individuals or infer characteristics, behaviors, or psychological traits can significantly impact mental health and autonomy.

Systems That Can Influence Human Behavior at Scale

- **Risk**: AI-powered content recommendation systems can influence human behavior, expression, and emotion, potentially causing harm to psychological and physical health.

AI systems are inherently socio-technical in nature, meaning they are influenced by societal dynamics and human behavior. AI risks and benefits can emerge from the interplay of technical aspects combined with societal factors related to how a system is used, its interactions with other AI systems, who operates it, and the social context in which it is deployed. AI systems are increasingly being used to make important predictions or decisions about people, such as with regard to credit, hiring, and digital services.

These risks make AI a uniquely challenging technology to deploy and utilize both for organizations and within society. Without proper controls, AI systems can amplify, perpetuate, or exacerbate inequitable or undesirable outcomes for individuals and communities. With proper controls, AI systems can mitigate and manage inequitable outcomes.

Regulatory and Policy Frameworks

The regulatory landscape for Artificial Intelligence (AI) is evolving, with new regulations being developed to address emerging challenges, while existing regulations and policies are being adapted to include AI-specific considerations. Municipalities like the City of Fredericton must continuously adapt to existing regulations and evolve new policies to ensure ethical and effective AI use. This section outlines the federal, provincial, and municipal regulatory and policy frameworks relevant to AI implementation.

Federal Regulations

Canada is advancing its regulatory framework for AI and data protection through initiatives such as Bill C-27, which introduces:

- **Consumer Privacy Protection Act (CPPA)**: Strengthens personal data rights by enhancing transparency, accountability, and privacy protection.
- **Artificial Intelligence and Data Act (AIDA)**: Ensures ethical AI use by increasing transparency, accountability, and mitigating potential harms.

In addition to national initiatives, global efforts include:

- **EU's AI Act**: Aims to regulate AI systems to ensure they are safe, fair, and trustworthy.
- **US NIST AI Risk Management Framework**: Provides guidelines for managing AI risks to promote innovation while safeguarding public interests.
- **Industry Standards**: Various industry standards are being developed to balance technological innovation with the protection of individual rights and societal values.

The evolving federal regulatory framework, including the Consumer Privacy Protection Act (CPPA) and the Artificial Intelligence and Data Act (AIDA), aims to ensure AI technologies are used ethically and responsibly. These laws enhance transparency, accountability, and privacy protection, directly impacting third-party AI products. The Artificial Intelligence and Data Act (AIDA) proposed as part of Bill C-27 establishes regulatory requirements for companies operating in Canada that develop, deploy, or use high-impact AI systems. The Act mandates that these companies implement measures to identify, assess, and mitigate risks associated with AI technologies, ensuring ethical and responsible AI usage.

Additionally, industry guidelines like the US NIST AI Risk Management Framework will help guide third-party AI, promoting safe and trustworthy use. By ensuring third-party AI products comply with these federal regulations and industry standards, municipalities like the City of Fredericton can safely integrate these technologies, balancing innovation with the protection of individual rights and societal values.

Provincial Regulations

New Brunswick Right to Information and Protection of Privacy Act (RTIPPA) At this time, no new AI-specific regulations are anticipated at the provincial level. However, the existing RTIPPA has provisions that should be considered in the context of AI implementation:

- **Privacy Protection**:
    - **Collection Limitation**: Public bodies can only collect personal information necessary for their functions.
    - **Use and Disclosure**: Personal information must only be used and disclosed for the purpose it was collected, unless consented by the individual or authorized by the Act.

    - **Data Security**: Public bodies must take reasonable measures to protect personal information from unauthorized access, use, disclosure, or destruction.
- **Disclosure of Personal Information**:
    - **Access Requests**: Individuals have the right to request access to their personal information held by public bodies.
    - **Transparency**: Public bodies must be transparent about how personal information is collected, used, and disclosed.
    - **Correction Requests**: Individuals can request corrections to their personal information if it is inaccurate or incomplete.

The New Brunswick Right to Information and Protection of Privacy Act (RTIPPA) provides a robust framework for safeguarding personal information and ensuring privacy. While no new AI-specific regulations are anticipated at the provincial level, RTIPPA's provisions on privacy protection and data security are critical. Municipalities must ensure that personal information (PII) is not disclosed to public or open AI systems, as doing so would violate RTIPPA's requirements. By adhering to RTIPPA, municipalities can responsibly manage AI technologies, protecting individuals' privacy and maintaining compliance with provincial regulations.

Municipal Policies, Procedures, and Standards

The City of Fredericton's existing policies and procedures provide a robust foundation for addressing AI risks. Here is an analysis of relevant policies that could mitigate specific AI risks:

COR-OP-142, Internet Usage

- **Bias and Discrimination**:
    - **Provision**: Prohibits accessing or distributing discriminatory material.
    - **Code of Conduct**: Equal Opportunity and Non-Discrimination - Employees must avoid any form of discrimination.
- **Privacy and Data Security**:
    - **Provision:** The policy outlines expectations regarding privacy, including the potential for computer identities to be recorded by websites visited and the need for encryption techniques for higher levels of privacy. It also warns that electronic records created by employees may be accessible to the public under the Access to Information Act.
- **Dependency and Loss of Skills**:
    - **Provision**: Encourages professional development and training.
- **Transparency and Accountability**:
    - **Provision**: Emphasizes transparency and accountability in online activities.
    - **Code of Conduct**: Accountability - Employees must be accountable for their actions.
- **Copyright Risks**:
    - **Provision**: Users must comply with copyright laws.

COR-OP-159, Official Languages Act Guidelines

- **Legal and Regulatory Compliance**:
  - **Provision**: Ensures compliance with language laws.
  - **Code of Conduct**: Compliance with Laws - Employees must comply with applicable laws and regulations.
- **Cultural Sensitivity**:
  - **Provision**: Ensures respect for cultural and linguistic diversity.
  - **Code of Conduct**: Respect for Diversity - Employees must promote inclusivity.

COR-OP-170, Remote Network Access

- **Privacy and Data Security**:
  - **Provision**: Mandates strict security measures and access controls.
  - **Code of Conduct**: Confidentiality - Employees must safeguard confidential information.
- **Security Vulnerabilities**:
  - **Provision**: Details security protocols and monitoring practices.
  - **Code of Conduct**: Security - Employees must follow security procedures.

COR-OP-173, Use of Social Media

- **Ethical and Moral Considerations**:
  - **Provision**: Mandates ethical conduct and appropriate use of social media.
  - **Code of Conduct**: Ethical Behavior - Employees must act ethically in all professional activities.
- **Public Perception and Trust**:
  - **Provision**: Ensures professional handling of social media communications.
  - **Code of Conduct**: Public Trust - Employees must uphold public confidence.

COR-OP-141, Use of Electronic Mail

- **Inaccurate Information ("Hallucinating" or "Confidently Incorrect")**:
  - **Provision**: Emphasizes the importance of accurate and reliable communication.
- **Quality Control**:
  - **Provision**: Requires professional and error-free communications.
  - **Code of Conduct**: Quality of Work - Employees are responsible for the quality of their work.
- **Transparency and Accountability**:
  - **Provision**: Emphasizes transparency in communications.
  - **Code of Conduct**: Accountability - Employees must be accountable for their actions.

COR-OP-154, Wireless Device and Cellular Service Usage

- **Privacy and Data Security**:

    - **Provision**: Mandates strict security measures and access controls.
    - **Code of Conduct**: Confidentiality - Employees must safeguard confidential information.
- **Security Vulnerabilities**:
    - **Provision**: Details security protocols and monitoring practices.
    - **Code of Conduct**: Security - Employees must follow security procedures.

While the City of Fredericton's existing policies and procedures address many aspects of AI-related risks, including data security, privacy, and ethical behavior, a clear and comprehensive AI policy would be beneficial. Such a policy would provide specific guidelines and frameworks for the ethical and effective use of AI technologies, ensuring that all potential risks are mitigated and that AI is integrated into municipal operations in a responsible and transparent manner. By establishing a dedicated AI policy, the City can enhance its ability to manage AI technologies while maintaining public trust and regulatory compliance.

Summary of Regulations and Policy Review

The regulatory and policy frameworks at federal, provincial, and municipal levels provide comprehensive guidelines for the ethical and effective use of AI. The City of Fredericton's existing policies and the Employee Code of Conduct already address many AI-related risks, ensuring compliance, security, and ethical behavior. By continuously adapting these policies and integrating new regulations, the City can ensure that AI technologies are managed responsibly, benefiting the community while maintaining public trust and regulatory compliance.

Draft Policy Statements

1. Introduction

1.1 Purpose

The purpose of this policy is to provide a framework for the ethical, responsible, and effective use of Artificial Intelligence (AI) and Large Language Models (LLMs) within the municipality. This policy aims to ensure that the use of AI and LLMs aligns with the municipality's strategic objectives, enhances service delivery, and protects the interests of residents and stakeholders.

1.2 Background

AI and LLMs offer significant potential to improve municipal operations, enhance decision-making, and provide innovative services to residents. However, their use must be carefully managed to mitigate risks related to privacy, bias, transparency, and accountability.

2. Policy Statements

2.1 Permitted Use of LLM AI

Municipal staff are permitted to use ChatGPT, CoPilot, and other open and public Large Language Models (LLMs) to assist with their work. These tools can enhance productivity, improve decision-making, and support innovative problem-solving. Staff are encouraged to use these AI tools responsibly, ensuring their use aligns with the municipality's ethical standards, privacy regulations, and data security protocols, including the Employee Code of Conduct and other Policies and Procedures.

2.2 Applications of LLM AI

Municipal staff are encouraged to use ChatGPT, CoPilot, and other LLMs for various tasks to enhance efficiency and productivity. The following specific applications are permitted:

- **Document Generation and Management**
  - **Application:** LLMs, such as ChatGPT, can assist municipal workers in drafting a wide range of documents, including resolutions, regulations, policies, grant applications, routine documents, reports, correspondence, meeting minutes, agenda items, newsletters, job advertisements, internal working documents, drafts of policies, bylaws, procedures, and draft communications plans and strategies.
  - **Benefits**: This comprehensive use of LLMs ensures accuracy and consistency in official documents, streamlines the document creation process, saves time for staff, and maintains consistency in formatting and language
- **Policy Analysis and Development**
  - **Application**: Employ AI to analyze existing policies, compare them with best practices, and suggest improvements or new policies based on data-driven insights.
  - **Benefits**: Enhances the quality and effectiveness of policy-making, ensures policies are up-to-date with current standards, and supports evidence-based decision-making.

2.3 Privacy and Data Security in LLM AI Use

Municipal staff must adhere to strict privacy and data security protocols when using ChatGPT, CoPilot, and other open public LLMs. To protect sensitive information and comply with the New Brunswick Right to Information and Protection of Privacy Act and its regulations, the following guidelines must be observed:

- **Prohibited Data**: Staff must not input any PII or any data that could be inferred as PII into public LLMs.
- **Examples of Prohibited Data**: Names, addresses, phone numbers, email addresses, social insurance numbers, health records, financial information, or any combination of data that could lead to the identification of an individual.

- **Compliance**: This restriction ensures compliance with privacy laws and protects the personal information of residents and stakeholders.

2.4 AI-Assisted Translation

Municipal staff are permitted to use LLMs and AI tools for drafting translations to enhance communication and efficiency. However, it is essential to ensure the accuracy and appropriateness of translated content. The following guidelines must be observed:

- **Draft Translations**: AI and LLMs may be used to create draft translations of documents, communications, and other materials.
- **Mandatory Review**: AI-generated translations must not be used directly for official purposes. All AI-generated translations must undergo a thorough human review to ensure accuracy, cultural appropriateness, and alignment with official standards.

2.5 Human Review and Editing of AI-Generated Content

Municipal staff are required to ensure that all AI-generated content undergoes a comprehensive human review, proofreading, and editing process before it is used for any official purpose or internal working document. This is to guarantee the accuracy, quality, and appropriateness of the content. The following guidelines must be observed:

- **Review Requirements**: All content generated by AI tools must be reviewed by a human to check for information accuracy, completeness, clarity, biases, discrimination, grammar, tone, and consistency.
- **Editing Process**: Make necessary edits to correct any inaccuracies, improve clarity, and align the content with the municipality's standards and values.
- **Approval**: The reviewed and edited content must be approved by a qualified staff member before it is used or disseminated.

2.6 Development and Use of Municipal AI and LLMs

The City of Fredericton may develop its own private AI and LLMs using tools such as Microsoft Azure AI Services. These models may be trained on municipal-specific data and information sets. The following guidelines must be observed:

- **Data Sources**: AI and LLMs may be trained on policies, bylaws, plans and strategies, electronic records, property files, PAC files and records, departmental documents, drawings, and knowledge base articles. This data may include PII.
- **Data Control and Privacy**: Information contained in, used to train AI on, and data input into the private AI does not leave the City of Fredericton's control and is not accessible by any outside parties.
- **Compliance**: Maintaining the AI and LLMs within the city's control ensures compliance with privacy regulations, protecting sensitive municipal and resident information from external access and misuse.

- **Internal Use and Knowledge Management**: These AI tools will serve as the basis for internal knowledge management, providing staff with quick and accurate access to municipal information.
- **Document Citation**: AI models must cite internal documents used in responses so that they can be referenced and verified by staff.
- **Quality Assurance and Testing**: Extensive testing for accuracy and quality assurance must be completed prior to any AI tools being made available for public use to ensure the risks of misinformation, inaccuracies, or biases are not perpetrated on the public.

2.7 Prohibition of AI for Social Credit Scoring

The City of Fredericton is committed to ensuring the ethical and fair use of Artificial Intelligence (AI) technologies. To uphold these principles, the following policy is established:

- **Prohibition of Social Credit Scoring**: The City of Fredericton will not use AI systems for the purpose of creating or maintaining social credit scores. This includes any evaluation or ranking of individuals based on their behavior, interactions, or other personal data.
- **Decision-Making on Service Provision**: AI systems will not be used to make decisions about the provision of services to individual residents and stakeholders. All decisions regarding access to municipal services will be made by qualified human personnel, ensuring fairness, transparency, and accountability.
- **Ethical Standards**: The use of AI within the municipality will adhere to high ethical standards, ensuring that all AI applications respect individual privacy, prevent discrimination, and promote fairness and transparency.
- **Oversight and Compliance**: The Chief Information Officer (CIO) will oversee the implementation and adherence to this policy. Regular audits will be conducted to ensure compliance and to address any potential issues promptly.
- **Transparency and Accountability**: The City will maintain transparency in its use of AI technologies. Any AI applications used by the municipality will be openly documented and their purposes clearly communicated to the public.

2.8 Adoption of Third-Party AI Technologies

In alignment with the Artificial Intelligence and Data Act (AIDA) and international best practices, the following policy is established for the adoption of third-party technologies that incorporate AI:

- **Compliance with AIDA**: The City may obtain and use third-party software and other services that incorporate AI. All third-party software and services providers must adhere to Canadian AI regulations as outlined in Bill C-27 and the AIDA, including the following principles outlined in the Act:
    - **Human Oversight & Monitoring**: Third-party AI systems must enable meaningful human oversight with appropriate interpretability and monitoring procedures in place.

    - **Transparency**: Vendors must provide clear and accessible information about their AI systems, including capabilities, limitations, and potential impacts.
    - **Fairness and Equity**: Vendors must ensure their AI systems are built to mitigate discriminatory outcomes and promote fairness and equity.
    - **Safety**: Vendors must proactively assess AI systems for potential harms and implement measures to mitigate these risks.
    - **Accountability**: Vendors must establish governance mechanisms to ensure compliance with legal obligations and maintain thorough documentation of their policies, processes, and compliance measures.
    - **Validity & Robustness**: Vendors must ensure their AI systems perform consistently with intended objectives and remain stable and resilient under various conditions.
- By adhering to these principles, the City of Fredericton ensures that the adoption of AI technologies through third-party applications is conducted responsibly and ethically, safeguarding the community and maintaining public trust.

2.9 Quality Assurance

All AI technologies, whether made available to the public through third-party applications or developed as private AI solutions, must undergo rigorous quality assurance testing before deployment. This testing is essential to ensure the accuracy of data and the elimination of bias and discrimination. Additionally, an AI Impact Assessment may be required by the Chief Information Officer (CIO) to evaluate potential impacts and ensure compliance with ethical standards and regulatory requirements.

3. Responsible Use of AI

In instances where this AI policy does not provide clear direction on a specific application or other existing policies or procedures do not specifically address a particular use of AI, all AI uses must adhere to the principles of responsible use. This ensures that AI technologies are deployed ethically, transparently, and in alignment with the municipality's values and strategic goals.

3.1 Definition of Responsible Use

Responsible use of AI involves deploying AI technologies in a manner that is ethical, transparent, and aligned with the municipality's values and strategic goals. This includes ensuring AI systems are used to augment human capabilities, protect privacy, mitigate biases, and maintain accountability.

3.2 Principles of Responsible Use

- **Ethical Use**: Ensure AI is used in ways that respect human rights, fairness, and non-discrimination.
- **Transparency**: Make the operations and decision-making processes of AI systems understandable to stakeholders.

- **Privacy Protection**: Adhere to privacy laws and regulations, safeguarding the personal information of residents and stakeholders.
- **Bias Mitigation**: Implement measures to detect and mitigate biases in AI systems, ensuring fair and equitable outcomes.
- **Accountability**: Maintain clear accountability mechanisms for AI usage, including regular audits and compliance checks.

By following these principles, the City of Fredericton commits to the responsible use of AI technologies, ensuring they are integrated into municipal operations in a way that upholds ethical standards, promotes transparency, protects privacy, mitigates biases, and maintains accountability. This approach ensures that AI technologies contribute positively to the community while aligning with the municipality's strategic goals and values.

3.3 Achieving Responsible Use

To achieve responsible use of AI within the City of Fredericton, it is essential to implement a comprehensive strategy that includes training, stakeholder engagement, continuous monitoring, and feedback mechanisms. This section outlines the necessary steps and measures to ensure ethical, transparent, and accountable AI deployment.

Training and Awareness

Providing ongoing training to staff on the ethical and responsible use of AI is crucial. This training should focus on:

- **Privacy Protection**: Ensuring compliance with privacy laws and safeguarding personal information.
- **Bias Mitigation**: Detecting and mitigating biases to ensure fair and equitable outcomes.
- **Transparency**: Making AI operations and decision-making processes understandable to all stakeholders.

Stakeholder Engagement

Engaging with residents, businesses, and other stakeholders is essential to gather feedback and address concerns related to AI use. Key activities include:

- **Communication Plan**: Develop a plan to inform stakeholders about the AI and LLMs policy and its benefits.
- **Feedback Mechanisms**: Establish public meetings and online surveys to gather input from residents and ensure that feedback is considered in ongoing AI initiatives.

Continuous Monitoring and Compliance

Regular monitoring and auditing of AI systems are vital to ensure they perform accurately and comply with ethical standards. This includes:

- **Performance Indicators**: Measure the number of AI and LLMs projects implemented, compliance with privacy and security standards, and stakeholder satisfaction.
- **Compliance Measures**: Conduct regular audits to ensure adherence to the AI policy. Non-compliance may result in disciplinary action, including termination of employment.

Feedback and Improvement

Establishing mechanisms for feedback and continuous improvement ensures AI systems evolve to meet community needs and expectations. Steps include:

- **Continuous Improvement**: Regularly update AI systems based on stakeholder feedback and technological advancements.
- **Review Process**: Annually review the AI and LLMs policy by the CIO and the AI Ethics Committee, making necessary updates to reflect new developments in AI technology and regulatory changes.

Enforcement

Effective enforcement of the AI policy is critical. The Chief Information Officer (CIO) is responsible for enforcing this policy. Any violations should be reported to the AI Ethics Committee for review and action.

By adhering to these guidelines, the City of Fredericton can leverage AI and LLMs to enhance internal knowledge management while ensuring the integrity and accuracy of information provided to the public. The use of private LLMs ensures compliance with privacy regulations and maximizes the benefits of models trained on specific corporate data.

4. Roles and Responsibilities

4.1 Municipal Council

- Approve the AI and LLMs policy.
- Ensure that the use of AI and LLMs aligns with the municipality's strategic objectives.

4.2 Chief Information Officer (CIO)

- Oversee the implementation of the AI and LLMs policy.
- Ensure that all AI and LLMs initiatives comply with this policy and relevant regulations.
- Report to the municipal council on the use and impact of AI and LLMs.

4.3 Chief Administrative Officer and Department Heads

- Ensure that departments comply with the AI and LLMs policy.
- Identify opportunities for the ethical use of AI and LLMs to improve service delivery.

4.4 Municipal Staff

- Adhere to the AI and LLMs policy and related guidelines.
- Participate in training and awareness programs on the ethical use of AI and LLMs

Conclusion

The City of Fredericton recognizes the transformative potential of AI and LLMs in enhancing municipal operations and service delivery. However, the responsible and ethical deployment of these technologies is paramount. By adhering to the outlined principles and policies, the City can harness the benefits of AI while mitigating associated risks, ensuring compliance with regulatory frameworks, and maintaining public trust. The proposed AI policy will provide clear guidelines for the ethical and effective use of AI technologies, reinforcing the City's commitment to transparency, accountability, and the protection of individual rights.