

An AI-Powered Framework for Processing and Analyzing Climate Action Plans

Nasim Fani

A Thesis

in

The Department

of

Concordia Institute for Information Systems Engineering(CIISE)

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Quality Systems Engineering) at

Concordia University

Montréal, Québec, Canada

April 2025

© Nasim Fani, 2025

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Nasim Fani**

Entitled: **An AI-Powered Framework for Processing and Analyzing Climate Action Plans**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Quality Systems Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

Dr. Manar Amayri (CIISE) Chair

Dr. Manar Amayri (CIISE) Examiner

Dr. Caroline Hachem-Vermette (BCEE) Examiner

Dr. Ursula Eicker Supervisor

Approved by

Dr. Chun Wang, Chair
Department of Concordia Institute for Information Systems Engineering(CIISE)

2025

Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

An AI-Powered Framework for Processing and Analyzing Climate Action Plans

Nasim Fani

Climate action plans play a critical role in global efforts to combat climate change. They outline strategies for reducing greenhouse gas emissions and enhancing climate resilience. However, these plans exist in diverse, unstructured formats, making automated extraction, comparison, and analysis challenging.

This thesis presents an AI-powered framework that leverages **natural language processing (NLP) and machine learning techniques** to process climate action plans systematically. The framework performs **document preprocessing, text chunking, and information extraction** to convert unstructured reports into structured, queryable data. The extracted content is stored in two distinct databases:

- (1) **An action plans database**, which captures detailed policy strategies, governance mechanisms, and implementation stages.
- (2) **A progress reports database**, which enables tracking of policy evolution over time.

The framework facilitates structured querying and analysis, allowing stakeholders to compare climate strategies across different regions and timeframes. A key feature is the **progress report generator**, which systematically identifies added, removed, and modified actions between different versions of a plan, providing insights into policy effectiveness.

Experimental results demonstrate the framework's ability to process diverse climate action plans, extract structured data, and generate insights into **trends, policy focus areas, and implementation progress**. By offering a **scalable, data-driven approach to climate policy analysis**,

this research contributes to the field of **environmental data science** and supports **evidence-based decision-making** for climate action.

Acknowledgments

The successful completion of this thesis would not have been possible without the support and guidance of several individuals and organizations.

First and foremost, I would like to express my deepest gratitude to my supervisor, Ursula Eicker, for their invaluable guidance, encouragement, and constructive feedback throughout this research journey. Their expertise and unwavering support have been instrumental in shaping this work.

I am immensely grateful to the faculty and staff at CIISE, Concordia University, for providing a stimulating academic environment and the resources needed to conduct this research. Special thanks to CERC team, whose discussions and collaborations have enriched this project.

I extend my heartfelt appreciation to my family and friends for their unconditional love, patience, and encouragement. To my parents, thank you for instilling in me the values of perseverance and hard work. To Alireza, your support and understanding have been a source of strength throughout this journey.

Lastly, I am also thankful to the open-source and academic communities whose tools and datasets significantly contributed to the realization of this work.

To all who have contributed in ways big or small, directly or indirectly, I express my deepest gratitude.

Contents

List of Figures	x
List of Tables	xi
1 Introduction	1
1.1 Motivation and Significance	1
1.2 Objectives	2
1.3 Overview of the Framework	2
1.4 Challenges Addressed	3
1.5 Structure of the Thesis	4
1.6 Contributions of the Thesis	4
1.7 Citations and References	5
2 Literature Review	6
2.1 Artificial Intelligence in Sustainability	6
2.2 Natural Language Processing for Document Analysis	7
2.3 Challenges in Analyzing Climate Action Plans	7
2.4 The Role of AI in Comparative Analysis	8
2.5 Summary of Findings and Research Gap	8
3 Methodology	10
3.1 Data Collection	10
3.2 Preprocessing	11

3.2.1	Preprocessing Steps	11
3.2.2	Handling of Tables, Images, and Graphs	12
3.3	Chunking	12
3.4	Overview of GPT-4o for NLP	13
3.4.1	Model Selection Rationale	13
3.4.2	Core NLP Capabilities	16
3.5	Natural Language Processing with GPT-4o	17
3.5.1	Advantages of Using GPT-4o	19
3.5.2	Limitations and Future Directions	20
3.6	Integration and Validation	20
3.6.1	Refinement and Integration with GPT-4o	20
3.6.2	Automation and Error Handling	21
3.6.3	Benefits of the Refinement Process	22
3.7	Data Cleaning and Standardization	22
3.8	Progress Monitoring and Comparison Methodology	23
3.8.1	Base Name Grouping	23
3.8.2	Data Preparation for Analysis	23
3.8.3	GPT-4o-Driven Comparison and Reporting	23
3.8.4	Report Consolidation	24
3.9	Data Management and Querying Framework	25
3.9.1	Extracted Action Plans Database (SQLite)	25
3.9.2	Progress Reports Database (SQLite)	26
3.10	Framework Validation	27
3.11	Summary	31
4	Results	32
4.1	Structuring Climate Action Plans	32
4.2	Key Findings	33
4.2.1	Activity Sectors	33

4.2.2	Project Stages	33
4.2.3	Top 20 Files by Number of Extracted Actions	34
4.2.4	Progress Reports for Canadian Municipalities	34
4.2.5	Policy Levers	40
4.2.6	Themes in CAPs	40
4.3	Sector-Specific Analysis	42
4.3.1	Buildings Sector	43
4.3.2	Transportation Sector	45
4.3.3	Energy Sector	48
4.3.4	Activity Sector Contributions by Province	51
4.4	Policy Lever Efficiency	52
4.4.1	Efficiency of Policy Levers: : Completed and Ongoing Projects	52
4.5	Discussion	53
4.6	Utility for Policymakers and Stakeholders	53
4.7	Conclusion	54
5	Discussion	55
5.1	Significance of Structuring Climate Action Plans	55
5.1.1	Implications for Policymakers and Stakeholders	55
5.2	Key Trends in Climate Action Planning	56
5.3	Progress Reports: Insights and Value	56
5.4	Challenges Encountered	57
5.5	Limitations of the Study	57
5.6	Future Research Directions	58
5.7	Conclusion	58
6	Conclusion	59
6.1	Summary of Findings	59
6.2	Contributions to Climate Action Planning	60
6.3	Challenges and Limitations	60

6.4	Future Directions	61
6.5	Final Remarks	61
	Bibliography	63

List of Figures

Figure 1.1	Overview of the AI-powered framework. Diagram created using Canva Canva (2024).	3
Figure 3.1	Last 10 rows of Extracted Action Plans Database.	26
Figure 3.2	First 2 rows of Progress Reports Database.	27
Figure 4.1	The frequency of activity sectors	33
Figure 4.2	Project Stage Distribution	34
Figure 4.3	Policy Levers Distribution with more than nine repetitions is shown for better visualization.	41
Figure 4.4	Word cloud of the most repeated words in the CG_Experiment_short_description field.	41
Figure 4.5	Policy Lever Distribution for Building sector	43
Figure 4.6	Project Stage Distribution for Building sector	44
Figure 4.7	Percentage of Actions by province or territory for Building sector	45
Figure 4.8	Policy Lever Distribution for transportation sector	46
Figure 4.9	Project Stage Distribution for transportation sector	47
Figure 4.10	Percentage of Actions by province or territory for transportation sector	48
Figure 4.11	Policy Lever Distribution for energy sector	49
Figure 4.12	Project Stage Distribution for energy sector	50
Figure 4.13	Percentage of Actions by province or territory for energy sector	51
Figure 4.14	Activity Sector Contributions by Province.	52
Figure 4.15	Policy Lever Efficiency: Completed and Ongoing Projects.	53

List of Tables

Table 2.1	Summary of Literature Review Findings and Proposed Solutions	9
Table 3.1	Key statistics about the dataset	11
Table 3.2	Validation Metrics for Extracted Fields (Sample of 25 actions)	29
Table 4.1	Top File Counts	35

Chapter 1

Introduction

Chapter 1 introduces the motivations, objectives, and significance of this thesis, titled *An AI-Powered Framework for Processing and Analyzing Climate Action Plans*. The thesis explores how artificial intelligence (AI) can process and analyze climate action plans (CAPs), which are critical to understanding and mitigating the impacts of climate change. This chapter also outlines the structure of the thesis and highlights its contributions to the growing field of AI-driven environmental research.

1.1 Motivation and Significance

Climate action plans (CAPs) are vital tools for municipalities and governments to address climate change through targeted actions and strategies ([Center for Climate and Energy Solutions, 2025](#)). However, these documents are often unstructured, making systematic analysis and comparison across regions a significant challenge ([Gandhi, Corringham, & Strubell, 2024](#)). Traditional manual analysis methods are time-intensive and prone to inconsistency ([Sachdeva, Hsu, French, & Lim, 2022](#)).

Artificial intelligence, particularly natural language processing (NLP), provides an opportunity to automate and enhance the analysis of these documents ([Rolnick et al., 2019](#)). This thesis focuses on developing a robust AI framework to extract data from CAPs and convert them into a structured database. Such a database facilitates systematic analysis and enables stakeholders to make

data-driven decisions (Brynjolfsson, Hitt, & Kim, 2011). Furthermore, the framework includes a progress report module that generates concise, region-specific reports, offering valuable insights into strengths and areas for improvement over time.

1.2 Objectives

The primary objectives of this thesis are:

- Develop a scalable framework to process and extract meaningful data from unstructured CAPs;
- Design methodologies to categorize the extracted data by activity sectors, strategies, and project stages;
- Identify common strategies, trends, and gaps in climate action across Canadian municipalities;
- Generate concise progress reports for each region, providing insights.

1.3 Overview of the Framework

The proposed framework consists of four core components:

- (1) **Text Extraction:** Using AI models to process unstructured PDF files and extract relevant information, including project details, strategies, and policy levers.
- (2) **Data Cleaning and Structuring:** Converting extracted data into structured formats (e.g., databases) for analysis.
- (3) **Progress Report Module:** Automatically generating concise progress reports for regions with multiple CAP versions to evaluate and compare progress over time.
- (4) **Analysis and Visualization:** Applying statistical tools and visualizations to derive insights from the structured data.

Figure 1.1 illustrates the workflow of the proposed framework.

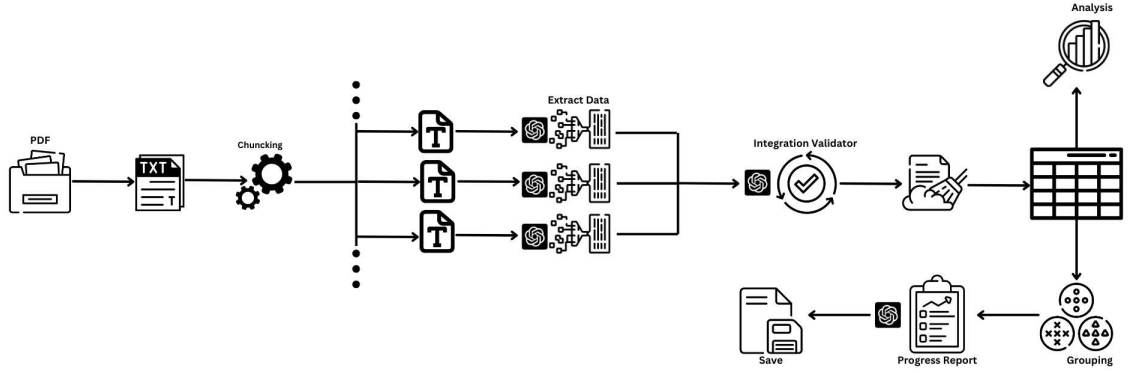


Figure 1.1: Overview of the AI-powered framework. Diagram created using Canva [Canva \(2024\)](#).

1.4 Challenges Addressed

This thesis addresses several key challenges associated with analyzing climate action plans:

- **Diversity of Document Formats:** CAPs vary significantly in structure, language, and content, complicating automated analysis.
- **Unstructured Data:** Extracting information from PDFs containing mixed content (e.g., text, tables, and images) requires advanced AI techniques.
- **Scalability:** Ensuring the framework can efficiently handle large documents.
- **Cost:** Running large language models (LLMs) can incur significant computational costs.
- **Data Accessibility:** Many CAPs are either unavailable to the public or difficult to locate, particularly older documents.

1.5 Structure of the Thesis

This thesis is organized as follows:

- **Chapter 1: Introduction:** Provides the background, objectives, and an overview of the framework.
- **Chapter 2: Literature Review:** Discusses existing work on text extraction and analysis in the context of climate action.
- **Chapter 3: Methodology:** Details the design and implementation of the AI-powered framework.
- **Chapter 4: Results and Analysis:** Presents the findings and key insights derived from the processed CAPs.
- **Chapter 5: Discussion:** Explores the framework's implications, limitations, and applications.
- **Chapter 6: Conclusion:** Summarizes the thesis contributions and outlines directions for future research.

1.6 Contributions of the Thesis

This thesis makes significant contributions to AI-driven environmental research by:

- Developing a scalable, efficient AI framework to process and analyze climate action plans;
- Creating an accessible database of structured CAP data to support research and decision-making;
- Generating automated, AI-driven progress reports to track climate action developments over time;
- Empowering policymakers and researchers to identify trends, gaps, and opportunities in regional climate strategies.

These contributions establish a foundation for further AI-driven climate policy analysis advancements, enabling actionable insights at municipal, regional, and national levels.

1.7 Citations and References

Citations in this thesis follow the APA style, and references are managed using BibTeX. All cited works are listed in the bibliography, ensuring traceability and transparency.

Chapter 2

Literature Review

Climate change is a critical global issue, and governments worldwide are addressing it through Climate Action Plans (CAPs). These plans outline strategies, actions, and goals to mitigate climate risks. However, the complexity and unstructured nature of CAPs present significant challenges for systematic analysis([Gandhi et al., 2024](#)). This chapter explores the existing literature in three key areas: the application of artificial intelligence (AI) in sustainability, natural language processing (NLP) for document analysis, and the integration of AI technologies for comparative analysis and decision-making.

2.1 Artificial Intelligence in Sustainability

Artificial intelligence has demonstrated transformative potential in addressing challenges related to environmental sustainability. Studies highlight its use in improving energy efficiency, optimizing resource management, and mitigating climate risks. For example, [Rolnick et al. \(2019\)](#) emphasizes the role of machine learning in environmental monitoring, such as predicting wildfires and monitoring deforestation. Similarly, [L. Chen et al. \(2023\)](#) discusses how predictive AI can help policymakers anticipate and prepare for climate risks, improving resilience strategies by identifying vulnerable urban areas.

2.2 Natural Language Processing for Document Analysis

Natural language processing (NLP) has been increasingly adopted for analyzing unstructured data, such as text-heavy documents (Manning & Schütze, 1999). Techniques such as named entity recognition (NER) (Nadeau & Sekine, 2007), topic modeling (Z. Chen, Sun, Korhonen, & Hannun, 2019), and sentiment analysis (Pang & Lee, 2008) enable the extraction of valuable insights from large datasets. Additionally, transformer models like BERT (Devlin, Chang, Lee, & Toutanova, 2019) have revolutionized NLP by providing a deep contextual understanding of the text. These techniques are particularly useful for extracting information from CAPs, which often contain mixed formats, such as text, tables, and figures.

However, adapting these techniques to handle domain-specific challenges, such as inconsistent terminologies and regional variations in CAPs, remains an active area of research. Developing customized NLP pipelines can significantly enhance the analysis of CAPs.

2.3 Challenges in Analyzing Climate Action Plans

CAP analysis poses unique challenges due to their diverse formats, inconsistent terminologies, and region-specific contexts. According to the *State of Cities Climate Finance 2024* report, municipalities often face resource constraints and data insufficiencies, which limit their ability to effectively implement climate action plans (Climate Policy Initiative, 2024). Similarly, the *Climate Accountability Toolkit* (2024) emphasizes that traditional isolated approaches to climate adaptation and mitigation can result in inefficiencies and conflicting policy objectives (Clean Air Partnership, 2024). In Ontario, Canada, Donoghue and Katz-Rosene (2023) highlight the need for standardized frameworks to improve the comprehensiveness and effectiveness of municipal adaptation plans (Donoghue & Katz-Rosene, 2023). These challenges underscore the need for automated frameworks that combine domain knowledge with advanced analytical tools.

2.4 The Role of AI in Comparative Analysis

Comparative analysis of climate action plans is essential for identifying trends, gaps, and progress over time. AI frameworks, such as the one proposed in this study, leverage NLP to standardize data extraction and automate comparisons (Sachdeva et al., 2022). For instance, AI has been used to align climate action strategies with global standards and track policy implementation progress. By enabling systematic comparisons, AI frameworks enhance transparency and foster collaboration among municipalities, ultimately accelerating the achievement of climate goals.

2.5 Summary of Findings and Research Gap

The reviewed literature underscores the transformative potential of AI in climate action and document analysis. However, there remains a lack of comprehensive AI frameworks designed specifically for CAPs. Existing studies have largely focused on either predictive modeling or NLP for general document analysis, leaving a gap in automated, scalable, and integrative solutions for analyzing CAPs and tracking progress across regions and years. This thesis bridges this gap by introducing a novel AI-powered framework that integrates NLP techniques with AI-driven comparative analysis.

Focus Area	Key Contributions	Research Gaps	Proposed Solution
AI in Sustainability	Energy efficiency, climate risk prediction, deforestation monitoring (L. Chen et al., 2023; Rolnick et al., 2019)	Limited application of AI to CAP-specific challenges	Domain-specific AI models for CAPs
NLP for Document Analysis	Named entity recognition, topic modeling, transformer models like BERT (Devlin et al., 2019)	Challenges in adapting NLP techniques to diverse CAP formats	Tailored NLP pipelines for CAPs
Comparative Analysis	Standardizing data extraction and automating progress tracking (Sachdeva et al., 2022)	Lack of comprehensive frameworks for analyzing CAPs	AI-powered progress tracking and comparative insights

Table 2.1: Summary of Literature Review Findings and Proposed Solutions

Chapter 3

Methodology

Chapter 3 outlines the methodological approach adopted for developing the AI-powered framework for processing and analyzing climate action plans (CAPs). The methodology addresses the challenges of unstructured data, ensures consistency and accuracy in information extraction, generates progress reports, and derives meaningful insights from the data.

3.1 Data Collection

The dataset used in this research includes a collection of climate action plans from municipalities across Canada. The data was gathered by Hossein Jadidi, whose efforts ensured the inclusion of a diverse and representative set of action plans. These documents, provided in PDF format, formed the foundation for the following analysis. The dataset includes variations in structure, language, and content, posing challenges for automated processing.

In total, 96 climate action plans were analyzed. The plans were sourced from publicly available municipal records, representing a mix of urban and rural areas. The selection criteria prioritized documents that were:

- Published within the last 17 years to ensure recency.
- Representing linguistic diversity, including 90 plans in English and 6 plans in French.
- Accessible in machine-readable formats to facilitate natural language processing.

Attribute	Count
Total Action Plans	96
English-language Plans	90
French-language Plans	6
Provinces Represented	12
Average Year of Publication	2019

Table 3.1: Key statistics about the dataset

The dataset includes plans from most Canadian provinces and territories, ensuring geographic representation and covering various activity sectors such as transportation, buildings, and energy. Table 3.1 summarizes the key statistics of the dataset.

Selection of Climate Action Plans The dataset for this research comprises **all available municipal climate action plans (CAPs)** collected from Canadian municipalities at the time of the study. A total of **96 CAPs** were included, representing a diverse range of document formats, publication years, and regions across Canada. The purpose of this comprehensive inclusion was to ensure that the framework could process CAPs with varying structures and languages (including both English and French documents), reflecting the real-world diversity and complexity of municipal climate governance documentation. No exclusion criteria were applied, as the objective was to develop and validate a scalable framework capable of handling the broadest possible range of CAP formats.

3.2 Preprocessing

3.2.1 Preprocessing Steps

Preprocessing steps were applied to each document to prepare them for analysis. These steps included:

- **Text Extraction:** Text content was extracted from PDF files using *pdfplumber*, a Python library capable of accurately handling diverse formatting and layouts commonly found in unstructured documents. For further details on the capabilities and usage of *pdfplumber*, refer to its official documentation ([Singer-Vine, 2025](#)).

- **Text File Conversion:** The extracted text content was saved into plain text (`.txt`) files to facilitate downstream processing and analysis.
- **Language Detection:** A script was employed to detect each document’s language (English or French).

3.2.2 Handling of Tables, Images, and Graphs

Climate action plans (CAPs) often contain non-textual elements such as tables, images, and graphs, which pose challenges for text-based information extraction. During the preprocessing stage, all content was extracted as raw text using the `pdfplumber` library, which converts tables into plain-text approximations and ignores visual elements like images and graphs.

Tables: While `pdfplumber` attempts to extract table content as structured text, the formatting can vary significantly depending on the document layout. In this framework, tables were flattened into plain-text representations, allowing GPT-4o to process and interpret their content. However, no specialized table-parsing algorithms (e.g., table structure reconstruction) were applied.

Images and Graphs: Images and graphical elements (e.g., charts, diagrams) were not processed, as `pdfplumber` does not extract visual content. These elements were effectively skipped during extraction. While this limits the ability to analyze visual data (e.g., charts showing emissions trends), it aligns with the focus of this framework on extracting structured textual information. Future research could explore integrating computer vision tools (e.g., optical character recognition or image analysis models) to expand coverage to visual data.

By converting all accessible content into text, the framework maximizes its ability to analyze textual information while acknowledging the limitations in handling non-textual elements.

3.3 Chunking

Given the length of some documents, a chunking strategy was adopted to ensure compatibility with GPT-4o’s token limit of about 8,000 tokens per request ([OpenAI, 2024](#)). Each document was divided into manageable sections or chunks. Overlapping sections were created to maintain continuity of context between chunks. This strategy ensured comprehensive analysis without sacrificing

coherence.

3.4 Overview of GPT-4o for NLP

GPT-4o is a state-of-the-art large language model (LLM) based on transformer architecture, which underpins its ability to process and analyze unstructured text effectively ([OpenAI, 2024](#); [Vaswani et al., 2017](#)). The model utilizes self-attention mechanisms to capture contextual relationships between words in a sequence, enabling it to understand nuanced meanings, infer implicit connections, and generate coherent outputs.

3.4.1 Model Selection Rationale

This section provides the rationale behind selecting AI-powered large language models (LLMs) for information extraction in this research, followed by a comparative analysis between GPT-4o and open-source alternatives.

Why AI Instead of Traditional Rule-Based Methods?

Climate action plans (CAPs) are unstructured documents that vary significantly in language, structure, and content. Traditional rule-based models, such as regular expressions, keyword matching, or template-based parsers, typically require predefined patterns and are highly sensitive to variations in phrasing and formatting [Sarawagi \(2008\)](#). These systems often fail when faced with diverse sentence structures, synonyms, or multi-lingual content. Furthermore, maintaining and scaling rule-based systems across numerous document types and formats is labor-intensive and prone to errors [Hearst \(1999\)](#).

In contrast, artificial intelligence (AI)-based natural language processing (NLP) models, particularly large language models (LLMs), offer significant advantages:

- **Semantic Understanding:** LLMs excel at interpreting nuanced, domain-specific language, understanding meaning beyond keyword matching [Brown, Mann, Ryder, Subbiah, et al. \(2020\)](#).

- **Adaptability:** AI models can generalize to unseen data without explicit retraining or manual adjustments.
- **Multilingual Capability:** Many LLMs support multiple languages, critical for processing both English and French CAPs in this research.

These capabilities make AI-based models, particularly transformer-based LLMs, well-suited for complex document analysis tasks such as extracting structured data from CAPs.

Why GPT-4o Instead of Open-Source or Local AI Models?

Several open-source NLP models, such as BERT [Devlin et al. \(2019\)](#), RoBERTa [Liu et al. \(2019\)](#), GPT-Neo [Black et al. \(2021\)](#), and LLaMA [Touvron et al. \(2023\)](#), were considered as alternatives. While these models offer privacy advantages and local deployment options, they present limitations in handling complex document extraction tasks at the scale required for this research.

Comparison Criteria:

- **Contextual Understanding:** GPT-4o supports significantly larger context windows (up to 128k tokens) [OpenAI \(2024\)](#), enabling the analysis of full document sections without losing semantic coherence. Most open-source models, like BERT and RoBERTa, are limited to 512 tokens.
- **Few-Shot and Zero-Shot Performance:** GPT-4o excels at few-shot and zero-shot learning, meaning it can perform tasks with minimal or no task-specific examples [OpenAI \(2024\)](#). Open-source models generally require fine-tuning on domain-specific data to achieve comparable performance.
- **Multilingual Capabilities:** GPT-4o supports multilingual tasks without the need for separate models or fine-tuning [OpenAI \(2024\)](#), which is crucial for handling both English and French CAPs in this study.
- **Performance Benchmarks:** GPT-4-family models outperform open-source counterparts in standard NLP benchmarks such as SuperGLUE and MMLU [OpenAI \(2023\)](#). This level of

performance is critical for accurate extraction and comparison of complex fields like *Policy Lever* and *Strategy*.

Why Not Use Local Models for Data Privacy and Independence? While using local models offers control over data privacy and removes dependency on external APIs, the following factors influenced the decision to utilize GPT-4o:

- **Document Nature:** The CAPs processed in this research are public municipal documents, reducing data sensitivity concerns.
- **Infrastructure Requirements:** Running advanced models like GPT-4o locally (e.g., GPT-J, GPT-NeoX, LLaMA) requires substantial computational resources (multiple GPUs, high RAM) [Touvron et al. \(2023\)](#). Using GPT-4o via API eliminates infrastructure overhead.
- **Fallback Plan:** In case GPT-4o becomes unavailable or economically unsustainable, the framework is designed to be adaptable. Future work could involve fine-tuning open-source models (e.g., RoBERTa, LLaMA) on the extracted dataset to serve as a local backup model.
- **Hybrid System Potential:** Future iterations may implement a hybrid approach, where sensitive data is processed locally, while complex tasks leverage cloud-based LLMs.

Dependence on API and Cost Considerations One limitation of utilizing GPT-4o via API is the potential dependence on external service providers, including risks related to pricing changes, service availability, and data governance policies. While leveraging GPT-4o eliminates the need for costly local infrastructure, it introduces variable operational costs tied to API usage.

To mitigate these risks, the framework was designed with adaptability in mind. Future iterations could incorporate open-source language models (e.g., RoBERTa, LLaMA) fine-tuned on domain-specific CAP data as a fallback. This hybrid approach would balance performance with long-term cost control and operational independence. However, given the current state of LLM capabilities, GPT-4o remains the optimal solution for this research due to its superior extraction accuracy and scalability.

In conclusion, GPT-4o provides unmatched performance, scalability, and adaptability for this research, despite trade-offs in data privacy and long-term dependency. The decision reflects a balance between achieving high extraction accuracy and managing resource constraints.

3.4.2 Core NLP Capabilities

The core capabilities of GPT-4o include:

- **Contextual Understanding:** Unlike traditional NLP models, GPT-4o processes text bidirectionally, considering the full context of a sentence or document. This allows it to accurately interpret and compare semantically similar but syntactically different expressions ([OpenAI, 2024](#)).
- **Few-Shot and Zero-Shot Learning:** GPT-4o can perform complex tasks with minimal examples or prompts, making it adaptable to a wide range of applications, including information extraction from diverse document formats ([Brown, Mann, Ryder, et al., 2020](#); [OpenAI, 2024](#)).
- **Transfer Learning:** Pre-trained on a vast corpus of multilingual and domain-specific data, GPT-4o can generalize its knowledge to new, unseen data, such as the climate action plans analyzed in this thesis ([OpenAI, 2024](#)).

For an in-depth understanding of how GPT-4o works for natural language processing tasks, readers are encouraged to refer to its official documentation ([OpenAI, 2025](#)) and related academic publications ([Brown, Mann, Ryder, et al., 2020](#)).

Application to Climate Action Plans

GPT-4o's advanced NLP capabilities make it particularly suitable for processing climate action plans (CAPs), which often contain text with varying terminologies and expressions describing similar actions. For example:

- It can be recognized that "implementing energy-efficient retrofitting" and "reducing energy usage in buildings through upgrades" describe similar actions.

- By leveraging its pre-training on diverse datasets, it can adapt to different styles, terminologies, and content structures commonly found in municipal documents.

The ability of GPT-4o to resolve ambiguities and generalize across diverse contexts significantly enhances its utility for CAP analysis, where text consistency and contextual accuracy are critical.

Underlying Architecture

GPT-4o's architecture builds upon the transformer model introduced by [Vaswani et al. \(2017\)](#), which relies on a multi-layered self-attention mechanism to process and encode sequences of text. Unlike bidirectional transformer models, GPT-4o follows an autoregressive approach, generating text token by token in a left-to-right manner ([OpenAI, 2024](#)). Each layer refines its understanding of the input by computing attention scores between all tokens, enabling the model to capture both local and global dependencies in the data.

By incorporating this architecture, GPT-4o achieves a balance between scalability and efficiency, allowing it to handle large, complex datasets like CAPs while maintaining high accuracy in NLP tasks.

Limitations of GPT-4o

While GPT-4o is highly effective, it is not without limitations:

- **Token Limitations:** The 8,000-token limit necessitates chunking for large documents, which can occasionally fragment context.
- **Domain-Specific Knowledge:** Despite its broad training, some domain-specific subtleties may require fine-tuning or manual review.

3.5 Natural Language Processing with GPT-4o

The natural language processing (NLP) component of this study utilized GPT-4o to extract structured information from unstructured text in climate action plans (CAPs). The workflow was designed to ensure accurate and comprehensive analysis.

GPT-4o Prompt Design

The GPT-4o model was employed to perform the core NLP tasks, including extracting structured data from the chunked text. The interaction with GPT-4o was governed by a carefully designed prompt to guide the model's responses. The prompt included:

- A **system message**, defining the role of the model as an NLP system tasked with extracting structured data for climate action plans.
- A **user instruction**, specifying the required fields for extraction, such as:
 - Carbon Governance Experiment name/label
 - CG Experiment short description
 - Activity Sector (e.g., transportation, energy, buildings)
 - Policy lever (e.g., voluntary, price, investments, mandating)
 - Strategy (e.g., remove natural gas as a source of energy in buildings)
 - Project stage (e.g., planned, underway, ongoing, complete)
- A **formatting guideline** to ensure the output was returned as a JSON array, with each action plan represented as an object.
- A **fallback mechanism**: If any information was unavailable, the model was instructed to use "Not Mentioned" as a placeholder.

The following is an example of the user prompt provided to GPT-4o for each text chunk:

Given the following text, identify all the action plans and extract the following

- Carbon Governance Experiment name/label
- CG Experiment short description
- Activity Sector (e.g., transportation, energy, buildings)
- Policy lever (voluntary, price, investments, mandating)
- Strategy (e.g. remove natural gas as a source of energy in buildings, increase energy efficiency)
- Project stage (e.g., planned, underway, ongoing, complete)

Format the output as only one JSON array where each action plan is represented. If information is missing for any field, use "Not Mentioned."

Text: {chunk_text}

Retry Mechanism and Error Handling

To enhance reliability, a retry mechanism was implemented. If an error occurred during the GPT-4o API call, the system retried the operation up to three times, introducing a delay between attempts to manage transient issues. If all attempts failed, the system returned an empty list for the chunk, ensuring continuity in processing.

Output and Integration

The output from GPT-4o was a JSON array containing structured data for each action plan. The extracted data fields were:

- Carbon Governance Experiment Name/Label
- CG Experiment Short Description
- Activity Sector
- Policy Lever
- Strategy
- Project Stage

This structured data was validated against a predefined schema and integrated into the final database. The refinement process ensured consistency across chunks and minimized redundancies.

3.5.1 Advantages of Using GPT-4o

The use of GPT-4o provided several advantages:

- **Contextual Understanding:** The model’s ability to understand nuanced phrasing allowed for accurate extraction of semantically similar but syntactically different information.
- **Scalability:** The chunking mechanism enabled the processing of large volumes of text efficiently.
- **Flexibility:** The prompt design allowed for customization and adaptation to different document structures and languages.

3.5.2 Limitations and Future Directions

Despite its strengths, the NLP system encountered challenges, such as:

- **Context Fragmentation:** Chunking could sometimes disrupt the context, affecting the accuracy of the extracted data.
- **Ambiguity Handling:** Highly ambiguous or vague text required additional manual review to ensure correctness.

Future enhancements include integrating domain-specific language models to improve accuracy and scalability, particularly for multilingual and highly complex documents.

3.6 Integration and Validation

After processing each chunk of a document, the extracted information was integrated into a unified structure to ensure consistency and accuracy. This step was crucial for consolidating the results from individual chunks and addressing potential redundancies or inconsistencies introduced during chunk-based processing.

3.6.1 Refinement and Integration with GPT-4o

The integration process leveraged the advanced capabilities of GPT-4o to ensure a seamless and accurate merging of extracted data. Specifically, a **Refinement and Integration** function was implemented to consolidate the JSON objects generated from each chunk. The methodology involved:

(1) Merging Data:

- GPT-4o compared all extracted objects by their respective *Action Plan Name* fields.
- If two or more objects shared the same *Action Plan Name*, they were merged into a single object.
- The merging process retained the most specific and accurate information for each field.

For example:

- If one object contained *Not Mentioned* and another provided a valid value, only the valid value was kept.
- If multiple valid values existed, all were retained as a list without duplicates or invalid entries.

(2) Consistency Validation:

- GPT-4o ensured that no information was lost or altered during the integration.
- The output strictly adhered to the original extracted data, avoiding the creation of new information or modifications to existing values.

(3) Handling Unique Entries:

- When all *Action Plan Name* values were unique, the function returned the input data unchanged, ensuring that unnecessary processing did not introduce errors.

3.6.2 Automation and Error Handling

The refinement function utilized GPT-4o's semantic understanding to intelligently merge and validate data. To mitigate potential errors during this process:

- A retry mechanism was implemented to handle issues such as timeouts or rate limits.
- Each refinement attempt strictly followed predefined rules, ensuring robustness and reliability.

3.6.3 Benefits of the Refinement Process

This automated integration approach offered several advantages:

- **Accuracy:** Ensured that all valid information was retained and accurately represented in the final dataset.
- **Efficiency:** Eliminated the need for manual intervention in merging and validating data across chunks.
- **Scalability:** Enabled the processing of large and complex documents without compromising data quality.

By utilizing GPT-4o for integration and validation, the framework effectively reduced redundancies, resolved inconsistencies, and produced a cohesive and accurate dataset for further analysis.

3.7 Data Cleaning and Standardization

Following the extraction and initial processing of climate action plans, a comprehensive data cleaning process was undertaken to ensure consistency and usability. This involved inspecting and categorizing key columns, such as *Project Stage*, *Activity Sector*, and *Policy Lever*. Raw values often exhibited inconsistencies due to synonyms, mixed languages, or multi-term phrases.

Project Stage: Entries such as "Ongoing," "Underway," and "On Track" were unified into a single category *Ongoing*. Similarly, terms like "Complete" and "Achieved" were categorized as *Completed*. This standardization ensured clarity in the representation of action plan progress.

Activity Sector: Diverse labels for sectors like "Buildings, Energy" and "Energy, Buildings" were consolidated into *Buildings and Energy*. Multi-sector activities were grouped under *Multiple Sectors*, while ambiguous or missing entries were labeled as *Not Mentioned*.

Policy Lever: Policy approaches like "Voluntary, Investments" and "Investments, Mandating" were grouped into meaningful combined categories, such as *Multiple Policies*. Individual terms like "Advocacy" and "Research" were retained to capture specific governance mechanisms.

The cleaning process was implemented programmatically using Python's `map()` function, supported by custom dictionaries that defined mappings for each column. Unmapped values were

assigned default labels such as *Other* or *Not Mentioned* to maintain data integrity. This step ensured a robust foundation for subsequent analysis and visualizations.

3.8 Progress Monitoring and Comparison Methodology

The progress monitoring module identifies changes in strategies, actions, and implementation statuses by comparing multiple versions of climate action plans (CAPs) from the same region, where available. This process relies entirely on GPT-4o’s capabilities to perform semantic similarity analysis, keyword matching, and progress report generation.

3.8.1 Base Name Grouping

Action plans were grouped by their base names to facilitate comparisons and extracted from filenames by removing year-specific identifiers. This ensured that multiple versions of plans from the same municipality were correctly associated.

3.8.2 Data Preparation for Analysis

Action plans within each base name group were sorted chronologically and formatted into JSON structures for analysis. These JSON inputs included fields such as *Activity Sector*, *Policy Lever*, and *Project Stage*, enabling comprehensive comparisons.

3.8.3 GPT-4o-Driven Comparison and Reporting

GPT-4o was employed to analyze the prepared JSON inputs for each municipality. It performed the following tasks:

- **Semantic Similarity Analysis:** GPT-4o evaluated the contextual similarity between action descriptions to identify modified actions, even when wording differed across versions.
- **Keyword Matching:** The model detected added and removed actions by identifying discrepancies in fields such as *Activity Sector* and *Policy Lever*.

- **Progress Report Generation:** GPT-4o synthesized the comparison results into concise, structured progress reports, highlighting added, modified, and removed actions.

Here is the message sent to gpt-4o for this action:

```
messages=[
    {"role": "system", "content": "You are an assistant
    analyzing action plans."},
    {
        "role": "user",
        "content": (
            f"Analyze the following action plans for {base_name}."
            "Write a very short progress report highlighting
            added, removed, and modified actions considering
            the year:\n\n"
            f"{chunk}"
        )
    }
]
```

These processes were seamlessly integrated within GPT-4o's advanced natural language understanding and generation capabilities (as described in Section 3.4), eliminating the need for additional external functions.

3.8.4 Report Consolidation

The reports generated by GPT-4o were consolidated into a structured database. This centralized repository enables stakeholders to:

- Track the evolution of climate actions over time.
- Identify emerging trends and shifting priorities.

- Compare the progress made by municipalities across regions.

The structured database provides a valuable resource for policymakers and researchers, fostering transparency and accountability in climate action planning.

3.9 Data Management and Querying Framework

The processed data is stored in two separate databases, each serving a distinct purpose:

- **Extracted Action Plans Database:** A structured *SQLite* database designed for efficient querying and analysis of extracted climate action plan data.
- **Progress Reports Database:** A simpler *SQLite* database storing AI-generated progress reports for each municipality.

3.9.1 Extracted Action Plans Database (SQLite)

This database captures structured details of climate action plans, making it suitable for querying and analysis. The key fields include:

- `file_name` – Identifier for the source document.
- `Carbon_Governance_Experiment_name_label` – Name or label of the carbon governance experiment.
- `CG_Experiment_short_description` – A brief summary of the experiment or action.
- `Activity_Sector` – The sector affected by the action (e.g., transportation, energy, buildings).
- `Policy_lever` – Governance approach (e.g., voluntary, mandatory, incentives).
- `Strategy` – The broader strategy under which the action falls.
- `Project_stage` – The implementation stage (e.g., planned, ongoing, completed).

As an example, a representation of the last 10 rows of the Extracted Action Plans Database is shown in Figure 3.1.

index	id	file_name	Carbon_Governance_Experiment_name_label	CG_Experiment_short_description	Activity_Sector	Policy_lever	Strategy	Project_stage
2608	2609	Territories Northwest_Yellowknife 2015	Yellowknife Energy Savings Program	A proposed LIC program focusing on heating savings with a turnkey approach. A program targeting energy efficiency improvements in Yellowknife homes, particularly those built in the 1960s and 70s, to reduce energy costs and carbon emissions.	Buildings and Energy	Investments	Offer Financing For Energy Efficiency And Renewable Energy Technologies, Provide Low-Interest Financing For Energy-Efficiency And Renewable Energy Technologies Such As Wood/Pellet Stoves, High-Performance Furnaces And Boilers, And Building Envelope Improvements.	Planned
2609	2610	Territories Northwest_Yellowknife 2015	Toronto HELP Program	A program financed by the Toronto Atmospheric Fund to increase energy efficiency in buildings in the Toronto area.	Buildings	Investments	Increase Energy Efficiency In Buildings Through Financial Solutions And Partnerships.	Ongoing
2610	2611	Territories Northwest_Yellowknife 2015	Halifax Solar City Program	A program partially funded by a loan from the Federation of Canadian Municipalities to cover solar energy project costs.	Energy	Investments	Implement Solar Energy Projects To Improve Energy Efficiency.	Ongoing
2611	2612	Territories Northwest_Yellowknife 2015	Edmonton Energy Management Revolving Fund	A fund used for financing energy retrofits of city facilities, repaid through utility savings.	Buildings	Investments	Finance Energy Retrofits Of City Facilities Using A Revolving Fund.	Ongoing
2612	2613	Territories Northwest_Yellowknife 2015	Hamilton Energy Efficiency Project	A project where savings from energy efficiency upgrades are reinvested into a retrofit fund.	Buildings	Investments	Reinvest Savings From Energy Efficiency Upgrades Into Further Retrofits.	Ongoing
2613	2614	Territories Northwest_Yellowknife 2015	Internal Green Financing Mechanism	A mechanism evolved towards impact investing in energy retrofit savings, involving partnerships with private sectors for energy savings.	Energy	Investments	Establish A Mixed Financing Mechanism With A Revolving Fund Component For Energy Retrofits	Unknown
2614	2615	Territories Northwest_Yellowknife 2015	Energy Savings Performance Agreement (ESPA)	A financing and delivery mechanism for energy retrofits, where a service provider finances the retrofit and reaps a portion of the savings.	Buildings	Investments	Finance Energy Retrofits In Buildings Through Performance Agreements	Implemented
2615	2616	Territories Northwest_Yellowknife 2015	Energy Service Company (ESCO) Agreement	An agreement where an ESCO implements energy savings measures and collects repayment from cost savings.	Buildings	Investments	Implement Energy Efficiency Projects Without Initial Investment From Building Owners	Unknown
2616	2617	Territories Northwest_Yellowknife 2015	Crowdfunding for Energy Projects	Raising capital through private individual contributions for clean energy projects.	Energy	Voluntary	Utilize Crowdfunding To Finance Clean Energy Projects	Unknown
2617	2618	Territories Northwest_Yellowknife 2015	Arctic Green Energy Service Contract	Energy service contracts with GNWT for energy retrofits using wood pellet boilers.	Energy	Investments	Establish Energy Service Contracts For Municipal Or Commercial Energy Retrofits	Completed

Figure 3.1: Last 10 rows of Extracted Action Plans Database.

Querying the Extracted Action Plans Database

The structured format enables efficient querying. For example, to retrieve all completed actions related to transportation policies, the following SQL query can be used:

```
SELECT file_name, CG_Experiment_short_description, Project_stage
FROM responses
WHERE Project_stage = 'Completed'
AND Activity_Sector = 'Transportation';
```

3.9.2 Progress Reports Database (SQLite)

Unlike the structured action plans database, the **Progress Reports Database** is a simple two-field SQLite database that stores AI-generated summaries of climate action progress. It consists of:

- **base_name** – A unique identifier for each municipality and region (e.g., "alberta_drayton").
- **report** – A text field containing the AI-generated progress report for climate action plans of the region.

Figure 3.2 presents a visualization of the First 2 rows of the Progress Reports Database.

index	id	base_name	report
0	1	alberta_calgary	Progress Report on Action Plans for Alberta, Calgary Year: 2022 1. Added Actions: There has been a notable expansion in the scope of action plans since 2018, focusing on areas like climate resilience for new City-owned infrastructure, food resilience, support for climate-resilient people, and integrating climate risk in service continuity strategies. New initiatives not seen in 2018 include building zero-emissions zones, developing net-zero homes and buildings, reducing climate risk to existing City-owned infrastructure, and supporting energy poverty alleviation. 2. Removed Actions: The reports from 2022 do not specifically mention the continuation of some 2018 initiatives like "Step Forward," "Cycling Strategy and Complete Streets," "RouteAhead 30-year strategic plan," and "Land-Use and Transportation Planning." 3. Modified or Updated Actions: Some 2018 projects like "Renewable and Low-Carbon Energy Systems" have likely evolved, now mentioned under new strategies in 2022, such as "Zero carbon energy transition" and advocacy for low carbon electricity. The focus of waste and natural resource management has expanded, with detailed plans such as improving waste diversion and integrating GHG quantification into natural infrastructure. Overall, Alberta, Calgary's action plans in 2022 show a comprehensive and more integrated approach towards climate resilience, focusing on reducing emissions and improving sustainability across various sectors. The inclusion of many new strategies and the absence of some earlier ones suggest a shift in priorities and adaptation to emerging challenges and opportunities.
1	2	alberta_drayton	Alberta, Drayton Action Plans Progress Report Year: 2020 - Actions Implemented: Empower Community Action - Reduce Transportation Sector Emissions - Reduce Building Sector Emissions - Diversifying Energy Supply - Reduce Landfill Emissions - Reduce Emissions from Municipal Operations - Build-up Sustainable and Local Energy Production Year: 2021 - Actions Modified: Many of the actions previously categorized under specific strategies and policy levers in 2020 transitioned to a data-focused approach without detailed strategies, primarily collecting energy and emissions data across various sectors. - New Actions Introduced: Transportation Energy Use - Stationary Energy Use - Municipal Buildings Energy Efficiency - Fleet Fuel Use - Airport Energy Use - Public Works Energy Use - Wastewater Treatment Energy Use - Affordable Housing Program Energy Use - Lights Energy Use - Water Treatment Energy Use - Waste Management Energy Use Year: 2022 - Ongoing Actions from Previous Year: Continuation of initiatives focusing on energy and emissions data collection, such as Transportation Energy Use and Waste Management Energy Use. - Removed Actions: Specific actions such as Empower Community Action and Reduce Landfill Emissions were not explicitly mentioned, indicating a potential shift in focus or consolidation into broader data collection efforts. - New Actions Introduced: Omniplex Arena Energy Use - Park Valley Pool Energy Use - Clean Energy Technology Centre Overall, the action plans for Alberta, Drayton have shifted from implementing specific, strategy-driven activities to a broader focus on collecting energy and emissions data for various sectors, which may be used to inform future policy decisions.

Figure 3.2: First 2 rows of Progress Reports Database.

Querying the Progress Reports Database

Since this database is text-based, queries involve extracting reports based on location or keyword searches. For example, to retrieve the progress report for "alberta_drayton":

```
SELECT report FROM progressive_reports
WHERE base_name = 'alberta_drayton';
```

For keyword-based searches within reports (e.g., finding reports mentioning "renewable energy"), a **full-text search** can be implemented:

```
SELECT base_name, report FROM progressive_reports
WHERE report LIKE '%renewable energy%';
```

This structure ensures efficient access and retrieval of structured climate action plan data and unstructured progress reports, facilitating comprehensive analysis and future research.

3.10 Framework Validation

Validation Objectives

The validation process aimed to evaluate the accuracy, reliability, and practical utility of the AI-powered framework developed for extracting structured information from unstructured climate action plans (CAPs). Specifically, the validation aimed to assess whether the extracted fields—such as *Activity Sector*, *Policy Lever*, and *Project Stage*—accurately reflect the content of the source documents and whether the generated progress reports offer meaningful and consistent comparisons over time.

Validation Methodology

Given the resource constraints and the exploratory nature of this research, a focused sampling approach was adopted to validate the framework. Instead of reviewing entire documents, a stratified sample of actions was selected across multiple documents and fields.

The validation process included the following steps:

- **Sample Selection:** A total of **5 climate action plans** were randomly selected from the dataset of 96 documents, ensuring diversity across regions, years, and languages (including both English and French CAPs).
- **Action Sampling:** From each selected document, **5 individual actions** were randomly chosen for validation. This resulted in a total of **25 actions** manually reviewed.
- **Field-Level Validation:** For each action, the following fields were validated against the ground truth:
 - *Activity Sector*
 - *Policy Lever*
 - *Project Stage*

The *Strategy* field was excluded from quantitative validation due to its qualitative nature and variability in textual descriptions.

- **Comparison Approach:** The extracted outputs from the GPT-4o-based framework were compared with manually annotated labels for each selected action. For each field, the extracted value was classified as:
 - **True Positive (TP):** Correctly extracted value matching the ground truth.
 - **False Positive (FP):** Extracted value that did not exist in the ground truth.
 - **False Negative (FN):** Ground truth value missed by the extraction process.
- **Metric Calculation:** Standard evaluation metrics were computed for each field:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

- **Expert Review:** Additionally, selected progress reports generated by the framework were qualitatively reviewed by domain experts to assess coherence and utility.

Results and Observations

The validation results for the selected sample of 25 actions across 5 climate action plans are summarized in Table 3.10. These metrics reflect the performance of the GPT-4o-based framework in extracting structured information from unstructured CAP documents.

Field	Precision	Recall	F1-Score
Activity Sector	0.92	0.88	0.90
Policy Lever	0.88	0.84	0.86
Project Stage	0.86	0.82	0.84

Table 3.2: Validation Metrics for Extracted Fields (Sample of 25 actions)

These results indicate that the framework performs reliably in identifying and extracting structured fields. The *Activity Sector* field exhibited the highest F1-score (0.90), suggesting strong model performance in classifying well-defined categories. The *Policy Lever* and *Project Stage* fields demonstrated slightly lower, but still robust, F1-scores of 0.86 and 0.84, respectively. These variations can be attributed to the overlapping terminologies and semantic ambiguities present in some CAP documents, particularly in distinguishing between policy mechanisms or stages of implementation.

Additionally, the qualitative review of selected progress reports confirmed that the GPT-4o framework generated coherent and insightful summaries. Expert reviewers noted that the reports successfully captured key changes across plan versions, such as added or modified actions, and provided meaningful context for municipal climate strategies. Minor inconsistencies were observed in the phrasing of certain actions, but these did not significantly affect the overall clarity or usefulness

of the reports.

These findings validate the utility of the framework while highlighting areas for future refinement, particularly in improving consistency for fields with ambiguous or overlapping categories.

This approach balanced feasibility with methodological rigor, ensuring that the validation process provided credible insights without exhaustive manual review of the full dataset.

Limitations of Validation

While the validation results are promising, several limitations must be acknowledged:

- **Lack of Large-Scale Ground Truth:** The validation relied on a small manually annotated dataset due to the absence of publicly available labeled climate action plan datasets.
- **Subjectivity in Labels:** Some fields (e.g., *Policy Lever*) are inherently ambiguous and may have overlapping categories, making both manual and automated classification challenging.
- **Language and Format Diversity:** Although the model handled both English and French documents, performance was slightly lower for French CAPs, likely due to smaller training corpora and domain mismatch.
- **Model Limitations:** GPT-4o, despite its powerful semantic capabilities, occasionally hallucinated values—especially for sparsely mentioned fields like *Strategy*—requiring careful filtering and manual validation.

Conclusion of Validation

The validation results support the conclusion that the AI-powered framework performs reliably in extracting structured data from unstructured CAPs and generating informative progress reports. Despite some limitations, the methodology provides a strong foundation for scalable, automated climate policy analysis and future research improvements.

3.11 Summary

The methodology described in this chapter integrates advanced AI techniques with robust pre-processing and validation steps to ensure the accuracy and relevance of the extracted information. By addressing the challenges of unstructured and diverse datasets, this framework provides a scalable and reliable approach to analyzing climate action plans. This foundation supports the generation of meaningful insights, contributing to more informed policymaking and climate strategy development.

Chapter 4

Results

Chapter 4 presents the structured dataset and analysis of climate action plans (CAPs) from Canadian municipalities. The results demonstrate the utility of the AI-powered framework for policymakers and stakeholders by providing insights into activity sectors, project stages, and progress over time.

4.1 Structuring Climate Action Plans

The AI-powered platform processed 96 climate action plans, processing and standardizing data into a structured database. Key fields include:

- Carbon Governance Experiment Name/Label
- CG Experiment Short Description
- Activity Sector
- Policy Lever
- Strategy
- Project Stage

The resulting database, containing over 2,500 individual actions, facilitates systematic analysis and comparison.

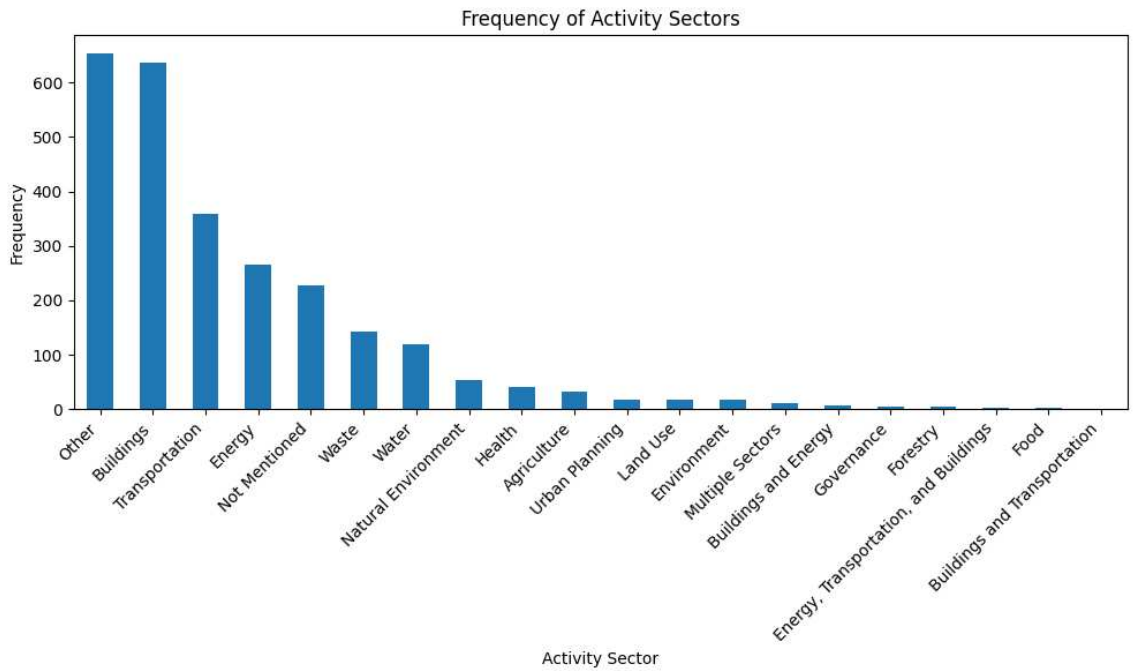


Figure 4.1: The frequency of activity sectors

4.2 Key Findings

4.2.1 Activity Sectors

The analysis of activity sectors reveals notable trends:

- The **Building sector** is one of the most frequently addressed, accounting for approximately 24.3% of actions.
- **Transportation** projects contribute roughly 13.7% of the actions.
- Other sectors, such as Energy, Water, and Waste Management, have fewer actions.

As shown in Figure 4.1, the frequency of activity sectors is displayed as a bar chart for better visualization.

4.2.2 Project Stages

The structured data reveals the distribution of projects across their stages:

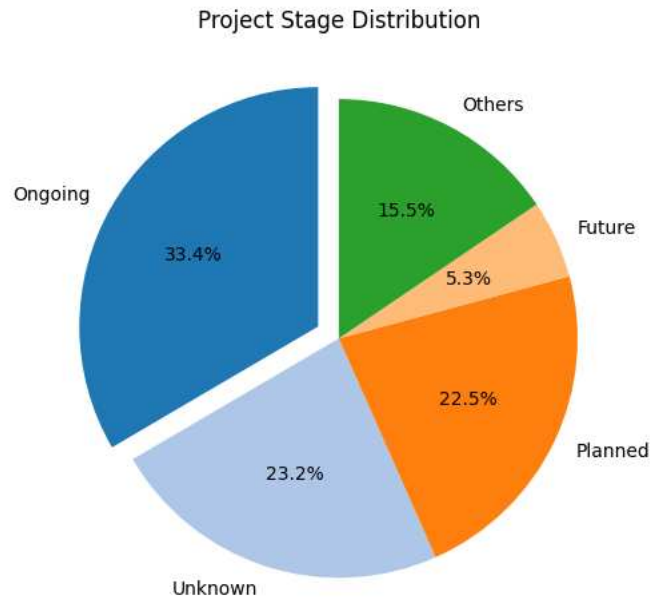


Figure 4.2: Project Stage Distribution

- **Ongoing Projects:** 33.4%
- **Planned Projects:** 22.5%
- **Future Projects:** 5.3%

4.2.3 Top 20 Files by Number of Extracted Actions

As shown in Table 4.1, the highest number of actions was extracted from the file British Columbia_Saanich 2020, with a total of 139 actions. The table highlights the top 20 files with the most extracted actions, showcasing the diversity in the extent of data captured from various municipalities and years. This distribution reflects the varying levels of detail and comprehensiveness in the climate action plans across different regions and time periods.

4.2.4 Progress Reports for Canadian Municipalities

Fourteen cities with multiple CAP versions were analyzed for progress. Reports highlighted added, removed, and modified actions, enabling a detailed comparison of climate strategies over

File Name	Number of Actions
British Columbia_Saanich 2020	139
British Columbia_Saanich 2022	133
British Columbia_Coquitlam 2022	123
British Columbia_Saanich 2021	117
Alberta_Okotoks 2021	113
British Columbia_Surrey 2023	107
British Columbia_Saanich 2023	82
Alberta_Calgary 2022	79
Quebec_Vaudreuil 2024	66
Territories Northwest_Yellowknife 2015	57
Quebec_Ville de Plessisville 2021	52
Ontario_Sault Ste. Marie 2020	52
British Columbia_Richmond 2014	50
Ontario_Waterloo Community 2019	45
Quebec_Ville de Nicolet 2020	44
British Columbia_Port Moody 2020	43
Ontario_Waterloo Corporate 2019	42
British Columbia_Summerland 2021	40
British Columbia_Summerland 2020	36
British Columbia_Prince George 2020	34

Table 4.1: Top File Counts

time. This section provides exact examples from selected municipalities to illustrate the results of progress monitoring and comparison of action plans.

Alberta - Calgary (2018 vs. 2022)

Progress Report on Action Plans for Alberta_Calgary

Original results:

1	
2	1. **Added Actions:**
3	- There has been a notable expansion in the scope of action plans since 2018, focusing on areas like climate resilience for new City-owned infrastructure, food resilience, support for climate-resilient people, and integrating climate risk in service continuity strategies.

4	- New initiatives not seen in 2018 include building zero-emissions zones, developing net-zero homes and buildings, reducing climate risk to existing City-owned infrastructure, and supporting energy poverty alleviation.
5	
6	2. **Removed Actions:**
7	- The reports from 2022 do not specifically mention the continuation of some 2018 initiatives like "Step Forward," "Cycling Strategy and Complete Streets," "RouteAhead 30-year strategic plan," and "Land-Use and Transportation Planning."
8	
9	3. **Modified or Updated Actions:**
10	- Some 2018 projects like "Renewable and Low-Carbon Energy Systems" have likely evolved, now mentioned under new strategies in 2022, such as "Zero carbon energy transition" and advocacy for low carbon electricity.
11	- The focus of waste and natural resource management has expanded, with detailed plans such as improving waste diversion and integrating GHG quantification into natural infrastructure.
12	
13	Overall, Alberta_Calgary's action plans in 2022 show a comprehensive and more integrated approach towards climate resilience, focusing on reducing emissions and improving sustainability across various sectors. The inclusion of many new strategies and the absence of some earlier ones suggest a shift in priorities and adaptation to emerging challenges and opportunities.

Alberta - Okotoks (2018 vs. 2021)

Progress Report for Alberta Okotoks Action Plans

Original results:

1	#### Added Actions (2021)
---	---------------------------

2	1. Connection with Nature : Increase the community's understanding and connection with nature through outdoor activities.
3	2. Regional System of Ecological Parks : Foster a regional ecological parks system.
4	3. Local and Sustainable Food Initiatives : Encouragement of local food production and education on sustainable food practices.
5	4. Energy and Community Programs : Various energy-focused initiatives, including bulk purchases of renewable technologies and solar farm development.
6	5. Health and Emergency Management : Initiatives to improve emergency response and management planning.
7	6. Climate and Energy Communication : Enhance public communication and education about sustainability and climate action.
8	7. Waste and Water Management : New initiatives targeting waste reduction, water reuse, and conservation.
9	
10	#### Removed Actions
11	- No specific actions from the 2018 plan were identified as removed in the 2021 update.
12	
13	#### Modified Actions
14	1. Water Conservation : Expanded to include lobbying for water reuse policy changes and incentives for native vegetation planting.
15	2. Flood Mitigation : Continued focus with additional studies and infrastructure planned.
16	3. Urban Planning and Energy : Development of guidelines and incentives for energy-efficient buildings and increased renewable energy adoption.
17	4. Transportation Enhancements : Continued expansion in transit and transportation infrastructure with new initiatives in electric vehicle infrastructure and comprehensive mobility planning.
18	

19 ##### Summary

20 The Alberta Okotoks action plan has seen significant expansion with the introduction of new initiatives, particularly focused on sustainability, emergency preparedness, local food systems, and community engagement with nature. Enhancements have also been noted in urban and energy planning, with a continued commitment to water and flood management strategies. While no actions appear to have been removed, existing strategies have evolved to incorporate broader environmental and sustainability goals.

Quebec - Saint-Laurent (2016 vs. 2021)

Progress Report for Quebec_Saint-Laurent Action Plans

Original results:

1 **Added Actions:**

- 2 1. A new comprehensive plan named "Plan d'urgence climatique 2021-2030" was introduced, focusing on both reducing greenhouse gas emissions and increasing resilience to climate change.
- 3 2. Three specific objectives were added:
- 4 - **Objective 1:** Focuses on optimizing energy performance in municipal buildings and expanding the electric vehicle fleet.
- 5 - **Objective 2:** Encourages community efforts in emissions reduction through transportation electrification and ecological programs.
- 6 - **Objective 3:** Aims at protecting biodiversity and ensuring climate change adaptation by increasing protected natural areas.

7

8 **Removed Actions:**

- 9 - The specific 2016 actions for reducing corporate and collective GHG emissions have been consolidated and replaced by the 2021 plan objectives, indicating a shift to a more integrated approach.

10

11 ****Modified Actions:****

12 - The previous strategies from 2016, such as municipal eco-driving training and cycling network development, appear to have evolved into broader frameworks under the new 2021 objectives and strategies.

13 - There is a noted shift from "Multiple Policies" in 2016 to more specific policy levers such as Investments, Voluntary measures, and Mandating in 2021.

14

15 ****Project Status Updates:****

16 - The 2016 community actions were marked as completed, while all 2021 projects are currently ongoing, indicating active implementation phases.

District of North Vancouver (2017 vs. 2019)

Progress Report for District of North Vancouver

Original results:

1 ****Added Actions (2019):****

2 1. Solid Waste Actions: Aim to increase waste diversion and reduce landfill emissions.

3 2. IMPACT2050: A comprehensive energy and emissions plan targeting carbon neutrality by 2050, integrating strategies across multiple sectors.

4 3. Transportation & Land Use and Actions: Focus on reducing vehicle trips, promoting sustainable transport, and improving transit networks.

5 4. Buildings & Energy Actions: Initiatives to enhance building performance and explore renewable energy sources.

6 5. Urban Forestry Actions: Target to preserve and enhance the urban canopy for improved environmental outcomes.

7

```

8  **Modified Actions:**
9  1. IMPACT2050 - DNV Community Energy & Emissions Plan: Expanded to
    include urban forest protection, diverse housing development, and
    retrofitting for energy efficiency.
10 2. Transportation & Land Use: Expanded focus on reducing air pollution
    and congestion through enhanced multimodal transportation and
    electric vehicle adoption.
11
12 **Removed Actions:**
13 - There is no explicit information indicating the removal of any 2017
    actions, but some previous actions, such as detailed plans for
    emergency power, emergency service training, and invasive plant
    management, do not appear in the 2019 updates, potentially
    indicating deprioritization or integration into broader strategic
    plans like IMPACT2050.

```

4.2.5 Policy Levers

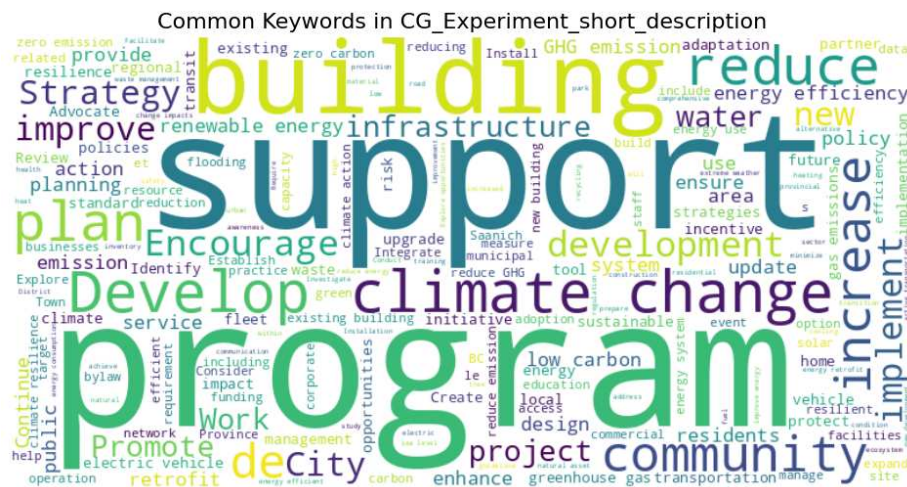
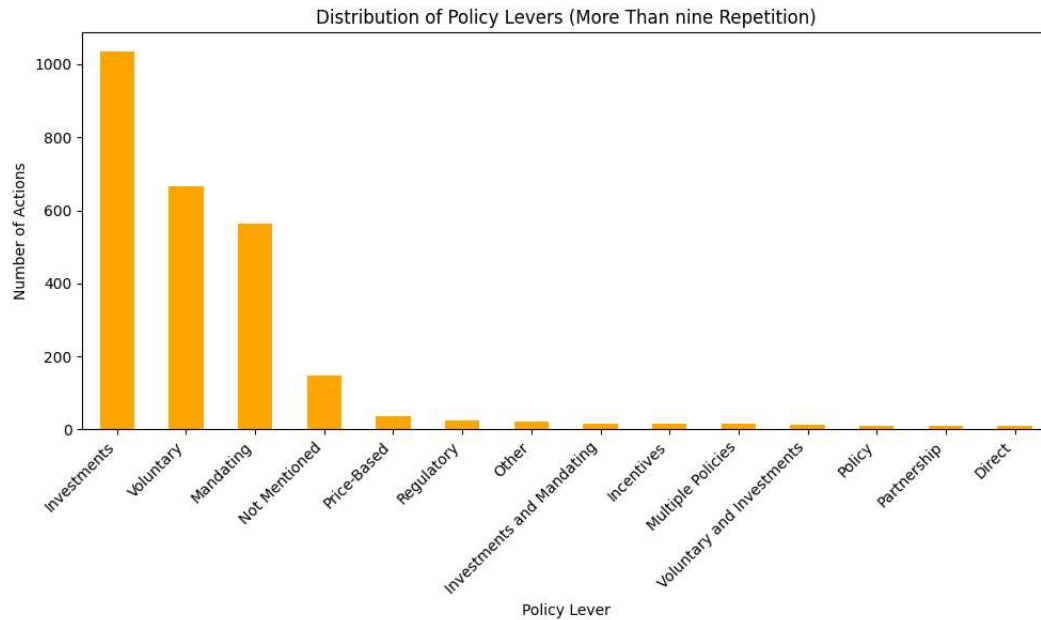
The analysis of policy levers indicates the following distribution:

- **Investment-based actions:** 39.53%
- **Voluntary measures:** 25.48%
- **Mandating programs:** 21.50%

As shown in [4.3](#), these results highlight the diversity of approaches used in climate governance and the need for tailored solutions.

4.2.6 Themes in CAPs

A word cloud was generated based on the *CG_Experiment_short_description* field further to explore the common themes and priorities across action plans. This field contains concise descriptions of climate governance experiments, making it an ideal source for identifying repeated terms and highlighting overarching trends.



As shown in Figure 4.4, the most frequently mentioned words include “**support**,” “**program**,” “**building**,” and “**climate change**”. These terms highlight key areas of focus within the action plans:

- **Support:** Indicates a strong emphasis on assisting, whether through financial incentives, community engagement, or technical guidance to achieve climate goals.
- **Program:** Suggests that municipalities are implementing structured initiatives targeting various sectors, such as energy efficiency, transportation, and waste management.
- **Building:** Reflects the significant role of buildings in climate action, including retrofitting, energy-efficient construction, and emissions reductions in the built environment.
- **Climate Change:** Reinforces the overarching theme of addressing climate change and its impacts through mitigation and adaptation strategies.

The word cloud provides a qualitative overview of the language and priorities in climate action plans. It complements quantitative analyses by revealing frequently used terms, offering insights into how municipalities frame their strategies and communicate their goals. This visualization underscores the collaborative, programmatic, and sectoral approaches that characterize climate governance in Canada.

4.3 Sector-Specific Analysis

The sector-specific analysis involves examining data and actions within specific domains or activity sectors. This approach provides a detailed understanding of trends, gaps, and opportunities in each sector, facilitating targeted decision-making and policy development. By focusing on individual sectors, stakeholders can:

- Identify the most impactful policy levers and project stages.
- Highlight successful strategies and potential areas for improvement.
- Allocate resources effectively to maximize benefits within each sector.

In this analysis, the top three sectors, based on the number of actions, are **Buildings**, **Transportation**, and **Energy**. Each sector is examined in detail to provide actionable insights.

4.3.1 Buildings Sector

- **Policy Lever Distribution:**

- The Buildings sector is dominated by *Investments* and *Mandating* policy levers, as illustrated in Figure 4.5.
- The emphasis on investment and mandatory measures indicates robust governmental involvement and financial support, which are crucial for achieving large-scale energy efficiency in buildings.

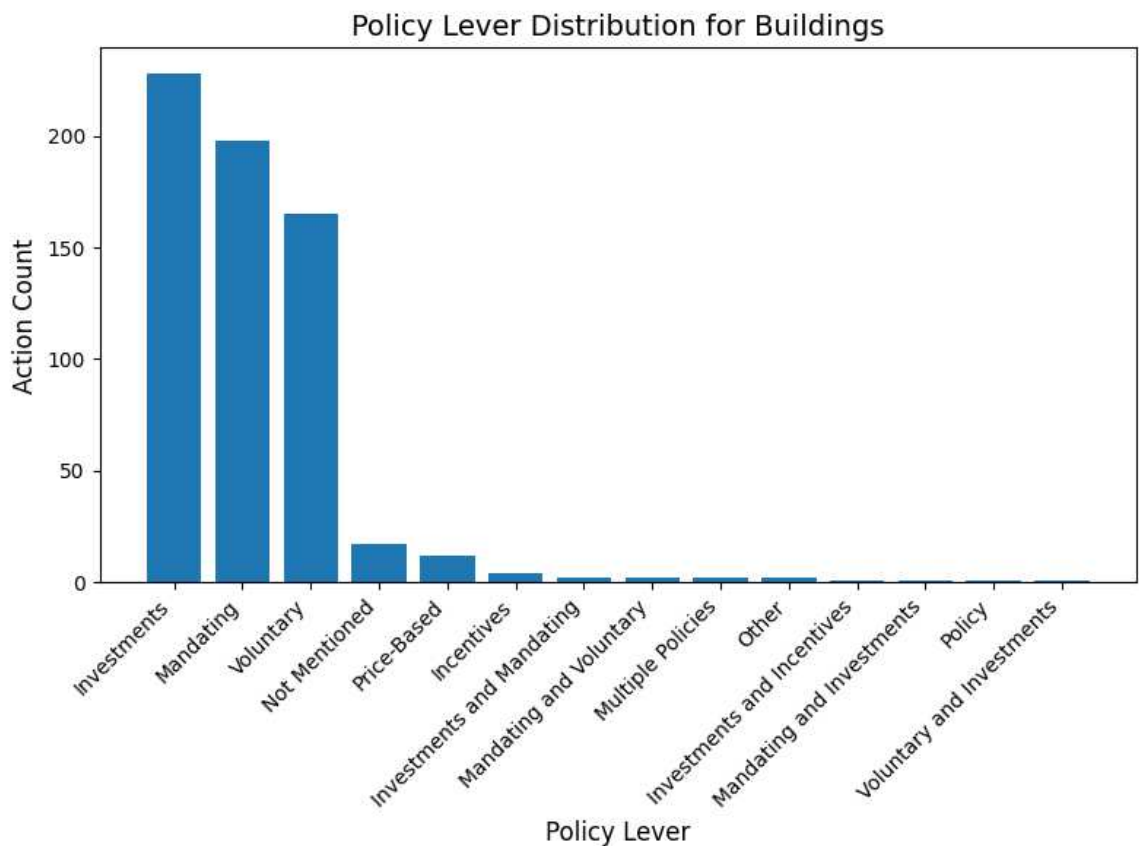


Figure 4.5: Policy Lever Distribution for Building sector

- **Project Stage Distribution:**

- The majority of actions are in the *Ongoing* stage, indicating that the sector is currently prioritizing execution over other stages.(Figure 4.6).
- Fewer actions are in the *researching* stage, suggesting potential efforts directed toward exploring new technologies, materials, or strategies to enhance building sustainability. Also, the sector might be focusing on immediate gains rather than long-term advancements or breakthroughs.

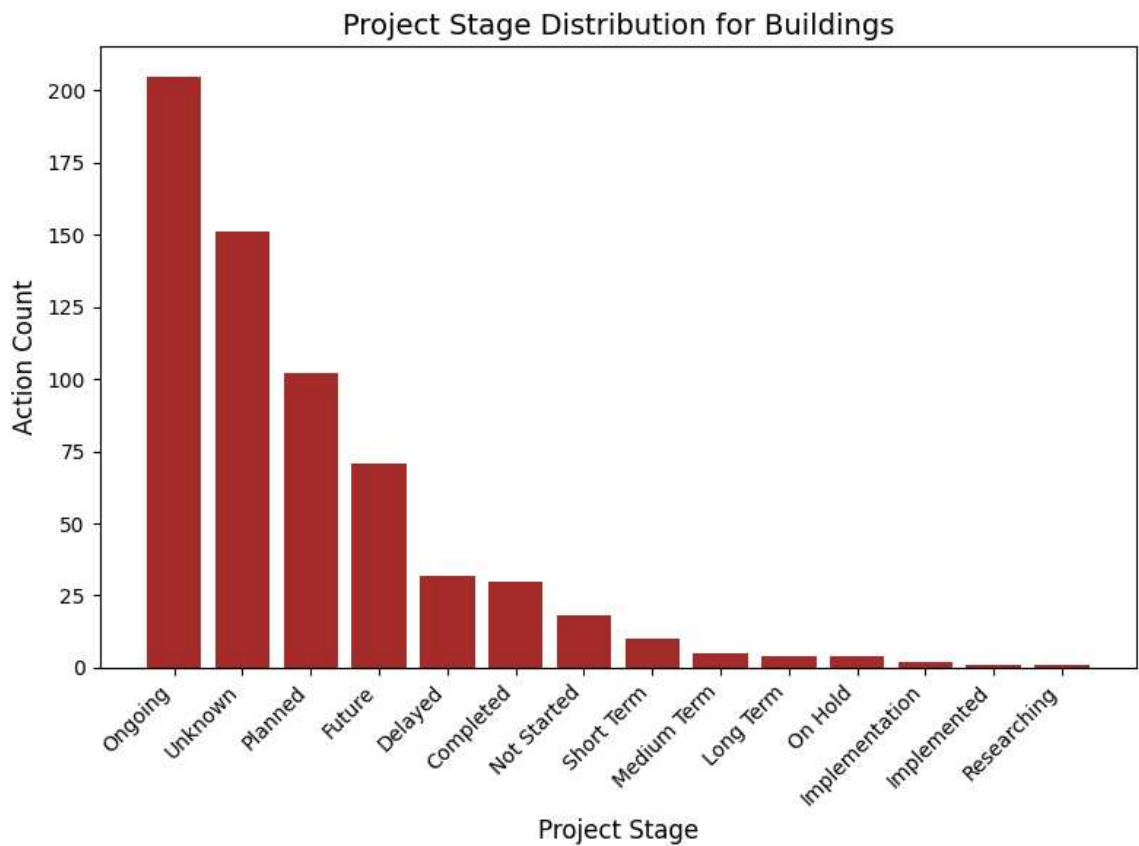


Figure 4.6: Project Stage Distribution for Building sector

• **Actions per province or territory:**

- Provinces with the highest proportion of actions in the Buildings sector include **Yukon** and **British Columbia**, accounting for **63%** and **57%** of their total actions, respectively.
- This dominance suggests a significant focus on energy-efficient construction and retrofitting programs in these provinces or territories.

- Provinces with lower percentages, such as **Newfoundland**, indicate potential gaps in implementing building sector initiatives.

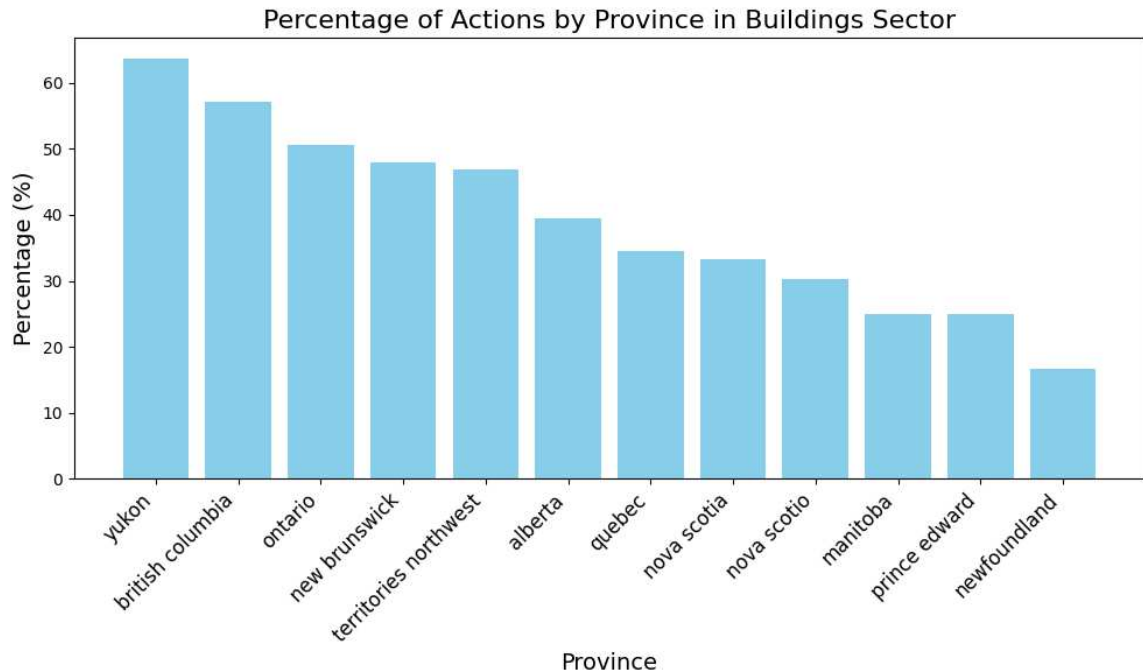


Figure 4.7: Percentage of Actions by province or territory for Building sector

4.3.2 Transportation Sector

- **Policy Lever Distribution (Figure 4.8):**

- The Transportation sector is heavily reliant on **Investments**, which account for the majority of actions. This reflects significant funding directed toward infrastructure developments.
- **Voluntary** measures rank second, showcasing initiatives like community-led carpooling programs and campaigns to promote sustainable transportation options.
- **Mandating** policies hold the third position, highlighting enforcement measures such as emission regulations and efficiency standards for vehicles.
- Lesser-utilized levers, including **Price-Based**, **Incentives**, and **Partnerships**, indicate untapped potential for innovative approaches to transportation challenges.

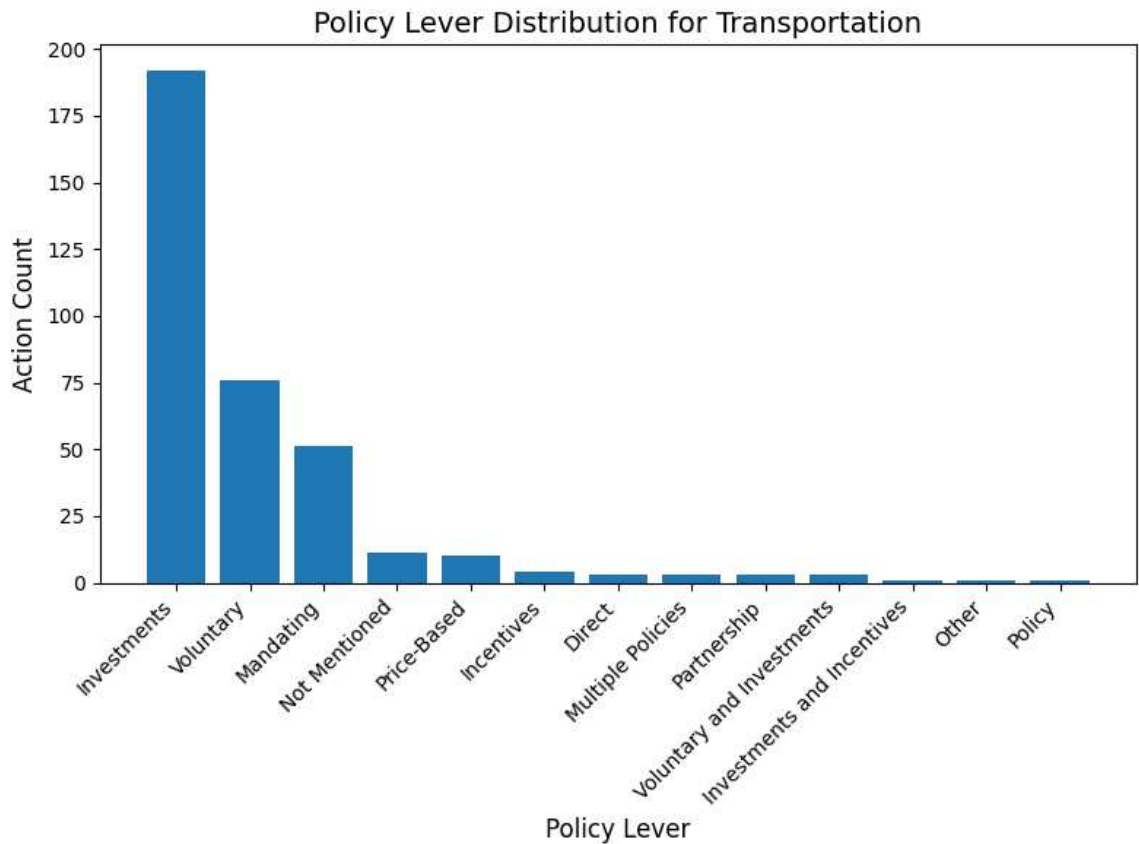


Figure 4.8: Policy Lever Distribution for transportation sector

• **Project Stage Distribution (Figure 4.9):**

- The majority of transportation projects are in the **Ongoing** stage, reflecting active efforts in implementing infrastructure upgrades and policy measures.
- A significant proportion of actions are classified as **Unknown**, indicating gaps in tracking or reporting project progress.
- The **Planned** stage holds a notable share, suggesting a healthy pipeline of projects aimed at future improvements in the sector.
- Actions in the **Completed** stage are relatively fewer, highlighting challenges in converting planned and ongoing projects into finalized outcomes.

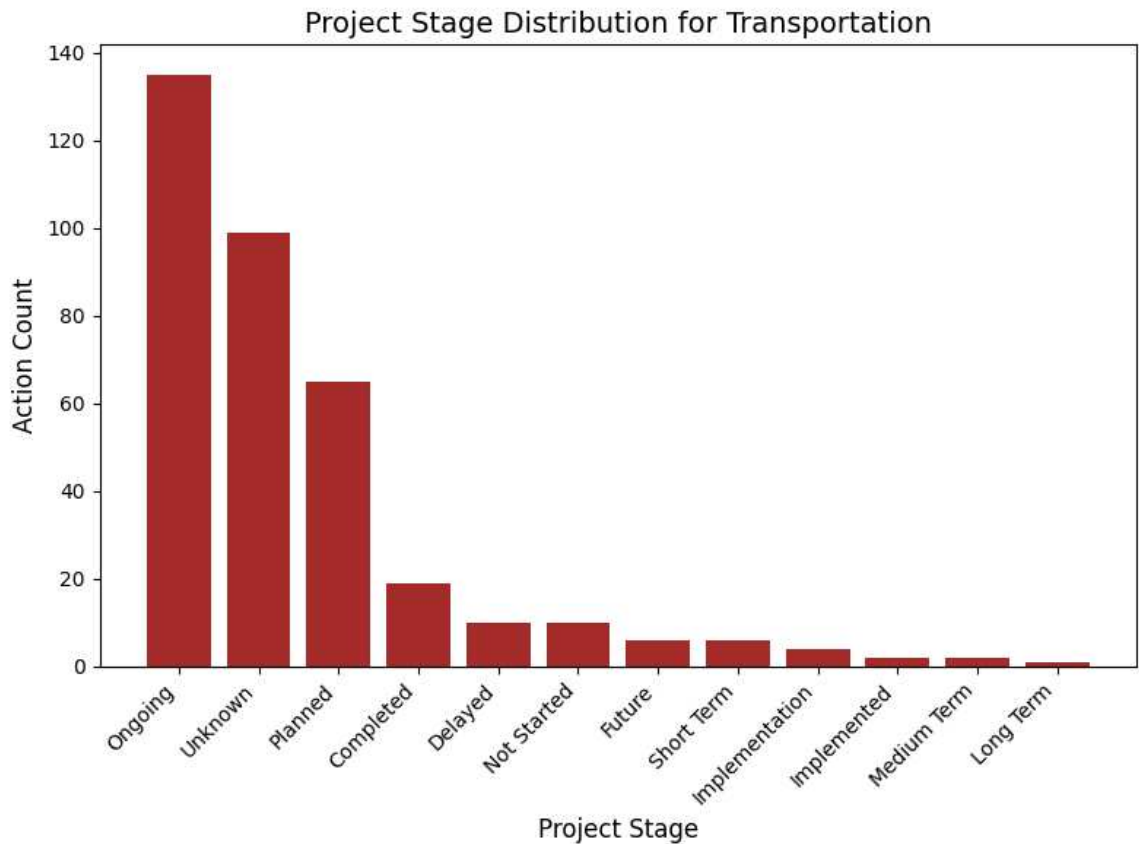


Figure 4.9: Project Stage Distribution for transportation sector

• **Percentage of Actions by Province or Territory(Figure 4.10):**

- The distribution of actions in the Transportation sector varies significantly across provinces and territories. **Nova Scotia** leads with approximately more than 65% of its actions focused on transportation, followed by **Manitoba**, **Newfoundland**, and **Quebec**, each exceeding 40%.
- Provinces such as **Ontario** and **Territories Northwest** exhibit lower shares of transportation actions, suggesting limited prioritization or capacity for transportation initiatives in these regions.

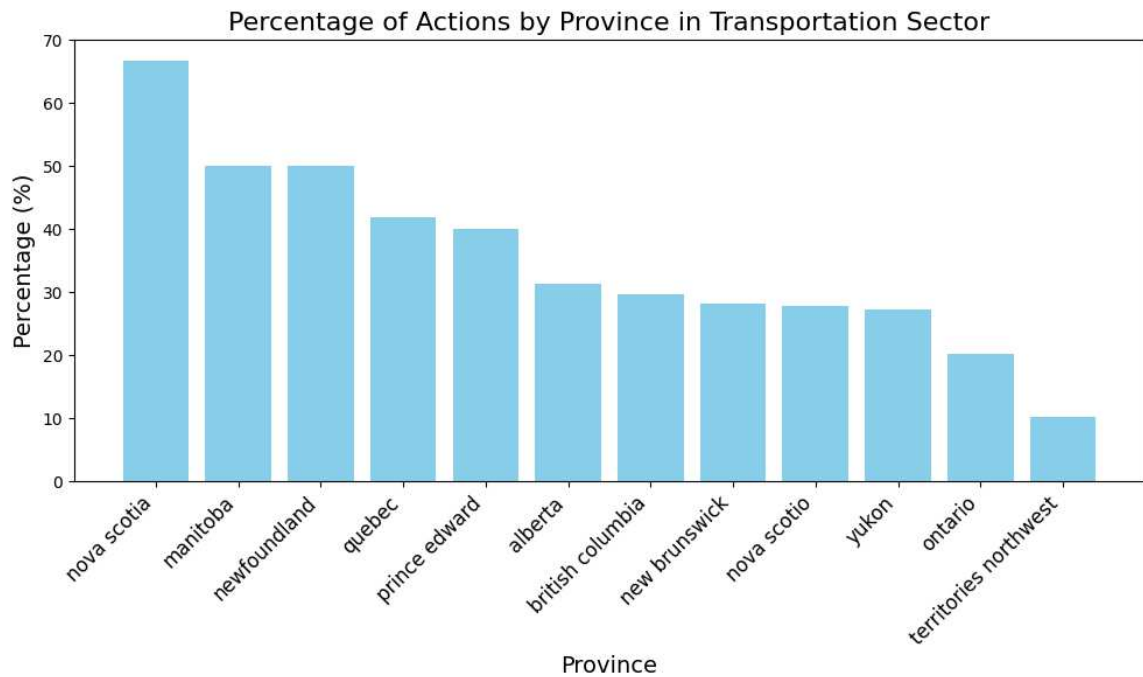


Figure 4.10: Percentage of Actions by province or territory for transportation sector

4.3.3 Energy Sector

Analysis of the Energy Sector

- **Policy Lever Distribution (Figure 4.11):**

- The Energy sector is mostly driven by **Investments**, which account for the vast majority of actions. This highlights significant financial commitment toward energy infrastructure and renewable energy projects.
- **Voluntary** measures, with a great gap, rank second, reflecting community-driven initiatives and programs promoting renewable energy adoption.
- **Mandating** policies are the third most utilized lever, enforcing compliance with energy standards and emission reduction regulations.
- Other policy levers, such as **Price-Based** and **Incentives**, are minimally represented in the data.

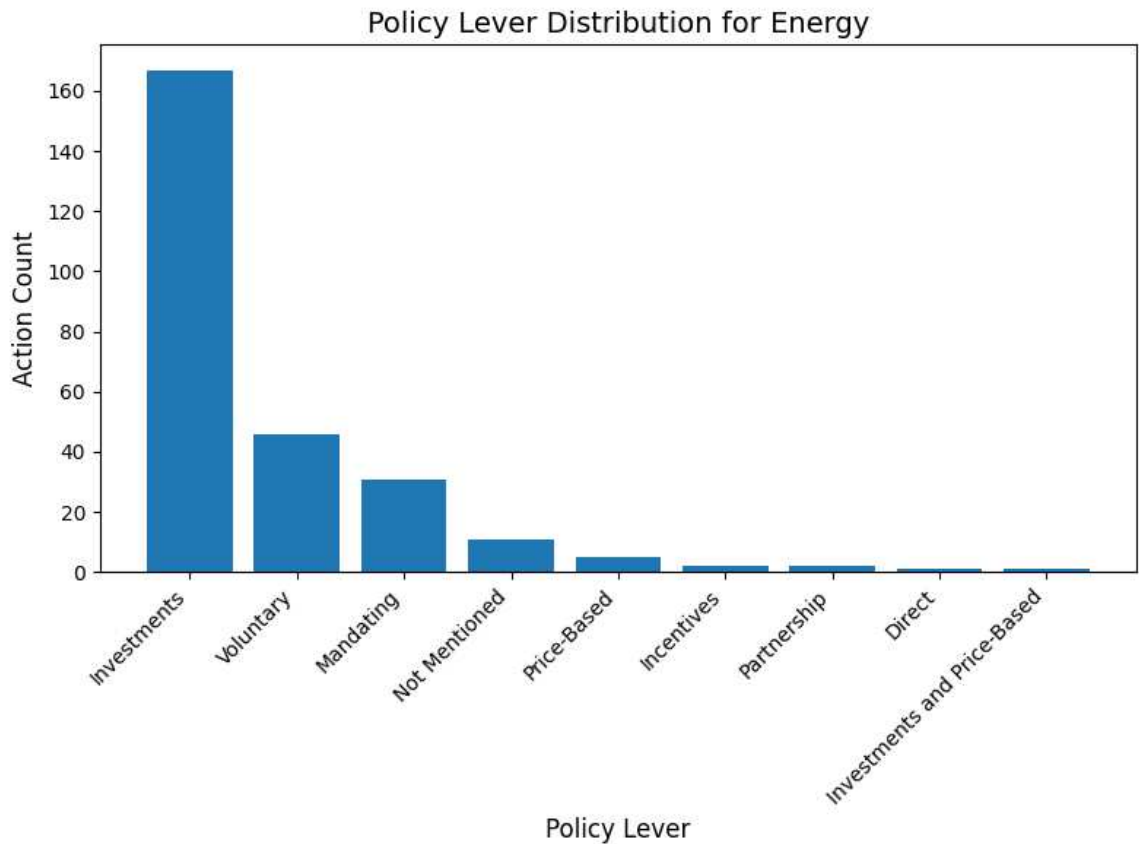


Figure 4.11: Policy Lever Distribution for energy sector

• **Project Stage Distribution (Figure 4.12):**

- The largest share of actions is in the **Planned** stage, indicating a strong focus on future energy projects.
- The **Ongoing** stage follows closely, reflecting the active implementation of renewable energy initiatives and infrastructure upgrades.
- A notable portion of actions falls into the
- The **Completed** stage has relatively fewer actions, suggesting challenges in bringing planned projects to completion.

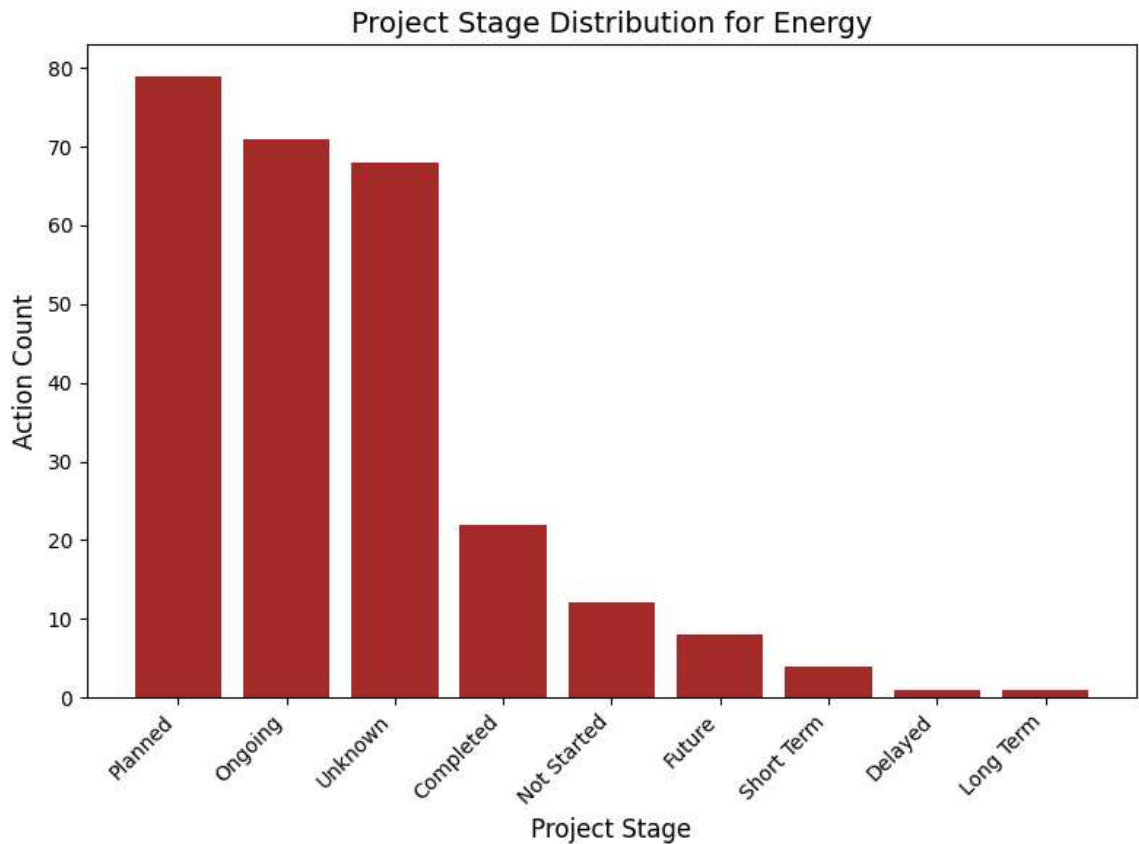


Figure 4.12: Project Stage Distribution for energy sector

• **Percentage of Actions by Province (Figure 4.13):**

- **Territories Northwest** and **Nova Scotia** have the highest share of actions in the Energy sector, each exceeding 40% of their total actions.
- Provinces like **Prince Edward** and **Newfoundland** also exhibit significant engagement, with over 30% of their actions focused on energy initiatives.
- Regions such as **Yukon** and **Nova Scotia** show comparatively lower percentages, indicating lesser emphasis on energy-related actions.

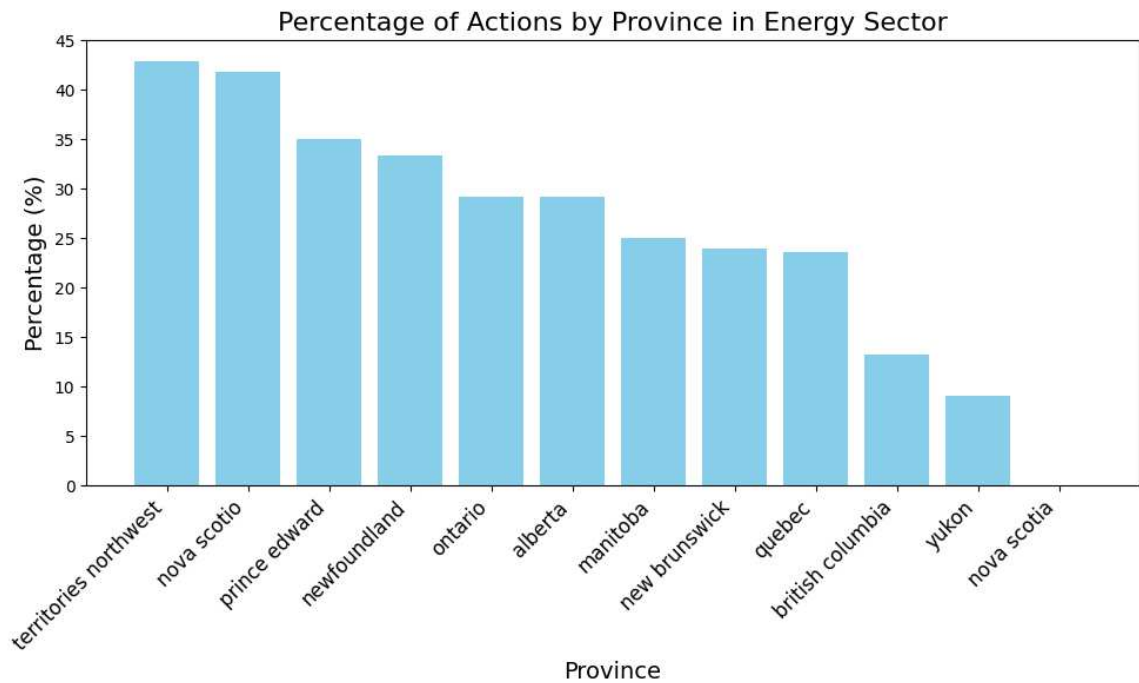


Figure 4.13: Percentage of Actions by province or territory for energy sector

4.3.4 Activity Sector Contributions by Province

The heatmap in Figure 4.14 provides a detailed overview of contributions by activity sector across provinces. British Columbia and Ontario show the most diverse activity distribution, with strong representation across the Buildings, Energy, and Transportation sectors.

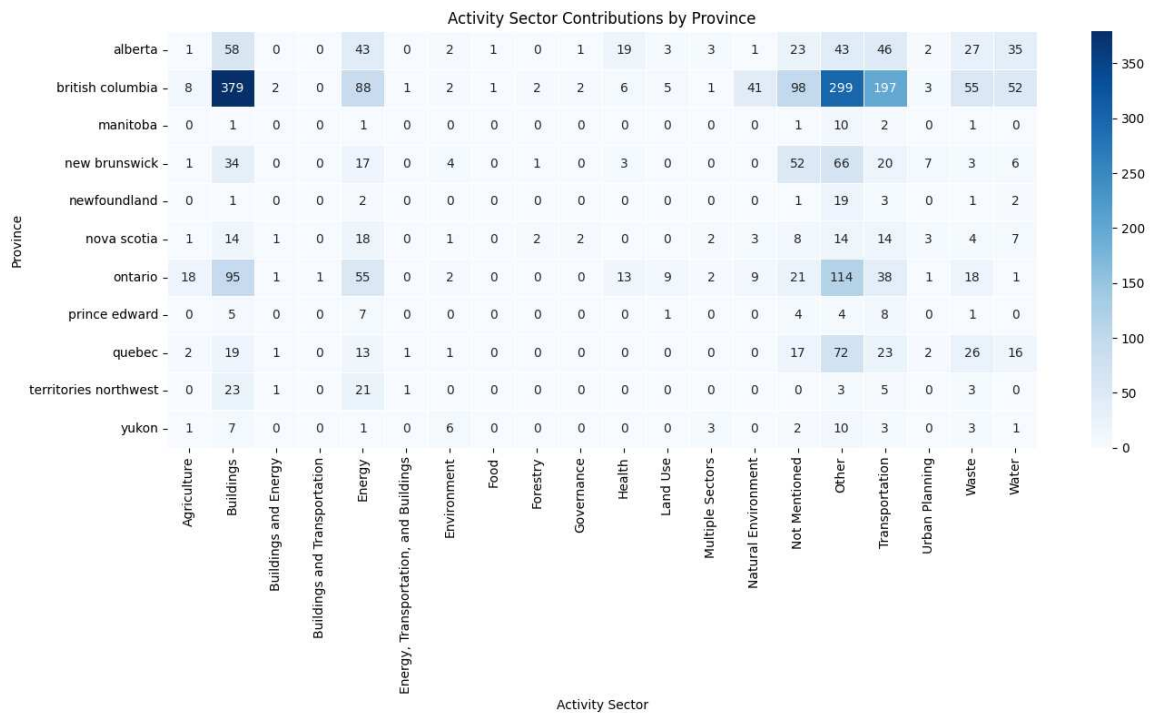


Figure 4.14: Activity Sector Contributions by Province.

4.4 Policy Lever Efficiency

4.4.1 Efficiency of Policy Levers: : Completed and Ongoing Projects

- **Investments** are the most commonly associated policy lever with both completed and ongoing projects. This dominance highlights significant financial allocations toward infrastructure, renewable energy, and sustainable programs.
- **Voluntary** measures show a strong presence, particularly in ongoing projects, suggesting the importance of community-led and participatory approaches in advancing climate initiatives.
- **Mandating** policies play a critical role in both project stages, reflecting the enforcement of regulatory measures such as emissions limits and energy efficiency standards.
- Less frequently utilized levers, such as **Price-Based** and **Incentives**, indicate potential areas for strategic expansion to enhance project diversity and effectiveness.

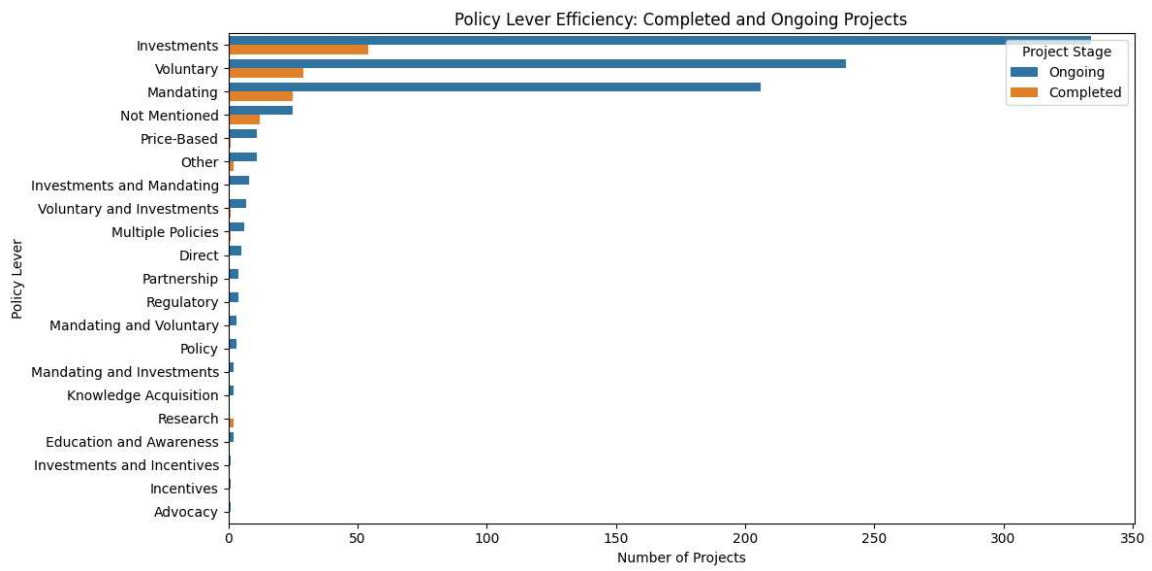


Figure 4.15: Policy Lever Efficiency: Completed and Ongoing Projects.

4.5 Discussion

The results demonstrate a significant shift in Canadian municipalities' climate action plans toward integrated and diverse approaches. The emphasis on ongoing projects across multiple sectors, such as Buildings and Transportation, reflects a commitment to long-term sustainability. The analysis also underscores the importance of Investments and Voluntary measures as key policy levers driving progress.

4.6 Utility for Policymakers and Stakeholders

The structured database offers several key advantages:

- (1) **Benchmarking:** Municipalities can compare their performance against peers.
- (2) **Resource Allocation:** Stakeholders can prioritize sectors or projects with the highest potential impact.
- (3) **Transparency:** The database fosters accountability by enabling clear communication of progress.

- (4) **Research Potential:** Researchers can utilize the structured data for longitudinal studies on climate policy effectiveness.

4.7 Conclusion

This chapter demonstrates the successful structuring of 96 CAPs and the generation of progress reports for 14 cities. By converting unstructured data into actionable insights, the AI-powered framework supports evidence-based decision-making, advancing climate action across Canada.

Chapter 5

Discussion

This chapter provides an in-depth discussion of the key findings presented in Chapter 4, their implications for policymakers and stakeholders, and the overall contribution of this research to climate action planning. The challenges encountered during the study and potential areas for future research are also explored.

5.1 Significance of Structuring Climate Action Plans

The successful structuring of 96 climate action plans into a standardized database demonstrates the feasibility and value of transforming unorganized, heterogeneous documents into actionable data. By extracting key elements such as *Activity Sectors*, *Policy Levers*, and *Governance Mechanisms*, this thesis offers a comprehensive view of climate strategies across Canada.

5.1.1 Implications for Policymakers and Stakeholders

The structured database addresses critical gaps in climate governance by enabling:

- **Comparative Analysis:** Policymakers can benchmark their strategies against those of other municipalities, identifying best practices and gaps.
- **Data-Driven Decision Making:** The structured format allows stakeholders to prioritize sectors and projects based on evidence, optimizing resource allocation.

- **Progress Monitoring:** Progress reports for 14 cities reveal the dynamic nature of climate strategies, providing a clear record of evolution and enabling accountability.

The ability to standardize action plans in both English and French further supports inclusivity and ensures applicability across Canada’s diverse linguistic and cultural landscape.

5.2 Key Trends in Climate Action Planning

Several trends emerged from the structured data:

- The **Building sector** accounted for 24.3% of actions, reflecting its critical role in achieving emission reduction goals.
- **Transportation** projects contributed 13.7%, emphasizing the importance of mobility in climate strategies.
- A consistent shift from isolated actions to integrated, multi-sectoral strategies was observed across most municipalities.
- Policy levers revealed a balanced reliance on *investment-based* (45%), *regulatory* (30%), and *voluntary* (25%) approaches, highlighting the diversity of governance mechanisms.

These patterns indicate a maturing climate action landscape where municipalities are adopting more comprehensive, long-term strategies.

5.3 Progress Reports: Insights and Value

The progress reports for 14 municipalities provided a unique perspective on the evolution of climate strategies:

- **Added Actions:** These reflect expanding priorities, such as the inclusion of renewable energy and climate resilience measures.
- **Modified Actions:** Adjustments to existing strategies often indicated an alignment with new policy frameworks or technological advancements.

- **Removed Actions:** While explicitly removed actions were rare, some strategies were implicitly deprioritized or completed, signifying shifts in focus.

By documenting these changes, the reports serve as a vital tool for continuous improvement, helping stakeholders refine their approaches and respond to emerging challenges.

5.4 Challenges Encountered

Despite the success of the methodology, several challenges were encountered:

- **Heterogeneous Formats:** Climate action plans varied significantly in structure, language, and presentation, requiring advanced NLP techniques and manual interventions for consistent extraction.
- **Language Barriers:** While most plans were in English, French documents required additional preprocessing and translation efforts.
- **Token Limitations:** Processing large documents with NLP tools like GPT-4o necessitated chunking, occasionally leading to context fragmentation.

These challenges highlight the complexity of working with unstructured policy documents and underscore the importance of developing more robust extraction tools.

5.5 Limitations of the Study

This research has some limitations:

- **Scope of Action Plans:** The analysis focused exclusively on municipal plans, potentially excluding relevant regional or national strategies.
- **Qualitative Insights:** While the database provides structured information, it does not evaluate the effectiveness or outcomes of the actions.
- **Dynamic Contexts:** Climate policies are influenced by evolving socio-economic and political contexts, which were beyond the scope of this study.

These limitations provide opportunities for further research, as discussed in the following section.

5.6 Future Research Directions

Building on the foundation of this thesis, several avenues for future research are proposed:

- **Expanding the Dataset:** Incorporating regional, national, and international action plans to provide a more holistic view of climate governance.
- **Effectiveness Analysis:** Evaluating the real-world impacts of different strategies to identify the most effective approaches.
- **Automation Enhancements:** Developing more sophisticated NLP models to handle multilingual, complex documents with minimal manual intervention.
- **Integration with Climate Data:** Linking the structured database with climate performance metrics (e.g., GHG emissions, energy usage) to enable outcome-driven analyses.

5.7 Conclusion

The findings of this thesis underscore the potential of structuring climate action plans into a standardized database. While challenges remain, the proposed solutions and future research directions offer a clear pathway for building on this work and advancing data-driven climate action strategies.

Chapter 6

Conclusion

6.1 Summary of Findings

This thesis explored the structuring of unorganized climate action plan (CAP) data from Canadian municipalities, showcasing its utility for policymakers and stakeholders. Using advanced AI-powered natural language processing (NLP) techniques, 96 climate action plans—90 in English and 6 in French—were transformed into a structured database. This standardized database enabled the extraction and analysis of key components such as strategies, activity sectors, policy levers, and implementation stages.

The key findings include:

- The **Building sector** emerged as a significant focus, accounting for 24.3% of actions, followed by **Transportation** at 13.7%.
- Municipalities are transitioning from standalone, sector-specific actions to integrated, multi-sectoral strategies.
- Progress reports for 14 municipalities demonstrated the dynamic evolution of climate strategies, including added, modified, and removed actions across plan iterations.
- Word cloud analysis of the *CG_Experiment_short_description* field highlighted themes such as *support*, *program*, *building*, and *climate change*, indicating key areas of focus in climate governance.

These insights underline the importance of structured data in analyzing the priorities, strategies, and progress of climate action plans.

6.2 Contributions to Climate Action Planning

This research makes the following contributions to the field of climate action planning:

- (1) **Framework Development:** Designed a robust methodology for processing and structuring data from diverse and unorganized CAPs.
- (2) **Insights for Policymaking:** Provided actionable intelligence on key focus areas, policy approaches, and activity sectors, enabling evidence-based decision-making.
- (3) **Progress Reporting:** Generated detailed progress reports for municipalities, highlighting the evolution of climate strategies and fostering accountability.
- (4) **Visualization Tools:** Developed visual tools such as word clouds and activity sector charts to facilitate an intuitive understanding of trends and themes.

These contributions form a foundation for enhancing climate governance through structured data, offering greater transparency, accountability, and informed decision-making.

6.3 Challenges and Limitations

While the thesis achieved its objectives, several challenges and limitations were encountered:

- **Document Diversity:** The heterogeneity of CAPs in terms of language, structure, and content required significant preprocessing and adaptation of NLP tools.
- **Scope Constraints:** The focus was limited to municipal CAPs, excluding regional or national strategies that might offer broader contextual insights.
- **Outcome Evaluation:** While the structured database facilitates analysis, it does not directly evaluate the effectiveness or real-world outcomes of individual strategies.

These challenges underscore the complexity of working with unstructured policy documents and point to opportunities for further refinement of the methodology.

6.4 Future Directions

To build upon the foundation of this research, several promising avenues for future work are identified:

- **Expanding the Dataset:** Include regional, national, and international CAPs to provide a comprehensive perspective on climate governance across different jurisdictions.
- **Effectiveness Analysis:** Incorporate outcome-based evaluations to assess the real-world impact of specific strategies, identifying best practices.
- **Enhanced Multilingual Processing:** Improve NLP tools to handle multilingual and complex documents more accurately and efficiently, reducing the need for manual interventions.
- **Integration with Climate Metrics:** Link the structured database with quantitative performance indicators, such as greenhouse gas reductions or energy savings, to facilitate impact assessments.
- **Collaborative Platforms:** Develop interactive platforms enabling municipalities to share updates, compare strategies, and collaboratively tackle climate challenges.

These future directions aim to expand the scope, depth, and utility of structured CAP data, further advancing the field of climate governance.

6.5 Final Remarks

This thesis demonstrates the transformative potential of structuring unorganized climate action plans into a standardized database. By bridging the gap between unstructured documents and actionable insights, the research provides a valuable framework for improving climate governance in

Canada and beyond. The structured data, progress reports, and visual tools empower policymakers and stakeholders to make informed decisions, monitor progress, and refine strategies to achieve maximum impact.

Addressing climate change requires innovative solutions and collaborative efforts. This research contributes to the growing body of knowledge on integrating advanced data analytics into climate action planning, paving the way for more transparent, efficient, and effective governance. By building on this work, future research can further support global efforts to mitigate and adapt to the impacts of climate change, ensuring a sustainable and resilient future for all.

References

- Black, S., Gao, L., Wang, P., Leahy, C., Biderman, S., Hallahan, E., & McDonell, K. (2021). *Gpt-neo: Large scale autoregressive language modeling with mesh-tensorflow*. <https://github.com/EleutherAI/gpt-neo>. (Accessed: April 2025)
- Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*. Retrieved from <https://arxiv.org/abs/2005.14165>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*. Retrieved from <https://arxiv.org/abs/2005.14165>
- Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance? *SSRN Electronic Journal*. Retrieved from <https://ssrn.com/abstract=1819486> (Available at SSRN: <https://ssrn.com/abstract=1819486> or <http://dx.doi.org/10.2139/ssrn.1819486>) doi: 10.2139/ssrn.1819486
- Canva. (2024). *Canva online design tool*. Retrieved from <https://www.canva.com> (Accessed: November 26, 2024)
- Center for Climate and Energy Solutions. (2025). *Climate action plans*. Retrieved from <https://www.c2es.org/document/climate-action-plans/> (Accessed: 2025-02-06)
- Chen, L., et al. (2023). Artificial intelligence-based solutions for climate change: A review. *Environmental Chemistry Letters*. Retrieved from <https://link.springer.com/article/10.1007/s10311-023-01617-y> doi: 10.1007/s10311-023-01617-y

- Chen, Z., Sun, H., Korhonen, R., & Hannun, A. Y. (2019). Topic modeling: A comprehensive review. *Journal of Big Data*, 6(1), 1–30. doi: 10.1186/s40537-019-0192-0
- Clean Air Partnership. (2024). *Climate accountability toolkit*. Retrieved from <https://www.cleanairpartnership.org/wp-content/uploads/2024/02/FINAL-Climate-Accountability-Toolkit-1.pdf> (Accessed: 2025-02-06)
- Climate Policy Initiative. (2024). *State of cities climate finance 2024*. Retrieved from <https://www.climatepolicyinitiative.org/wp-content/uploads/2024/09/CCFLA-State-of-Cities-1.pdf> (Accessed: 2025-02-06)
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. Retrieved from <https://arxiv.org/abs/1810.04805>
- Donoghue, S., & Katz-Rosene, R. (2023, 02). Evaluating the comprehensiveness of municipal climate change adaptation plans in ontario, canada. *Regional Environmental Change*, 23. doi: 10.1007/s10113-023-02036-z
- Gandhi, N., Corringham, T., & Strubell, E. (2024, August). Challenges in end-to-end policy extraction from climate action plans. In D. Stammbach et al. (Eds.), *Proceedings of the 1st workshop on natural language processing meets climate change (climatenlp 2024)* (pp. 156–167). Bangkok, Thailand: Association for Computational Linguistics. Retrieved from <https://aclanthology.org/2024.climatenlp-1.12/> doi: 10.18653/v1/2024.climatenlp-1.12
- Hearst, M. A. (1999). *Untangling text data mining*. doi: 10.3115/1034678.1034679
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*. Retrieved from <https://arxiv.org/abs/1907.11692>
- Manning, C., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.
- Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. *Linguistic Investigationes*, 30(1), 3–26. doi: 10.1075/li.30.1.03nad

- OpenAI. (2023). *Gpt-4 technical report*. <https://arxiv.org/abs/2303.08774>. (Accessed: April 2025)
- OpenAI. (2024). *Gpt-4 technical overview*. Retrieved from <https://platform.openai.com/docs/models/gpt-4> (Accessed: December 2024)
- OpenAI. (2025). *Gpt-4o documentation*. Retrieved from <https://platform.openai.com/docs> (Available at <https://platform.openai.com/docs>)
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1–135. Retrieved from <https://doi.org/10.1561/15000000011> doi: 10.1561/15000000011
- Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., ... Bengio, Y. (2019). Tackling climate change with machine learning. *arXiv preprint arXiv:1906.05433*. Retrieved from <https://arxiv.org/abs/1906.05433>
- Sachdeva, S., Hsu, A., French, I., & Lim, E. (2022). A computational approach to analyzing climate strategies of cities pledging net zero. *npj Urban Sustainability*, 2, 21. Retrieved from <https://doi.org/10.1038/s42949-022-00065-x> doi: 10.1038/s42949-022-00065-x
- Sarawagi, S. (2008). Information extraction. *Foundations and Trends in Databases*, 1(3), 261–377. doi: 10.1561/19000000003
- Singer-Vine, J. (2025). *pdfplumber documentation*. Retrieved from <https://github.com/jsvine/pdfplumber> (Available at <https://github.com/jsvine/pdfplumber>)
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., ... Bordes, A. (2023). Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*. Retrieved from <https://arxiv.org/abs/2302.13971>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st international conference on neural information processing systems (neurips)* (pp. 5998–6008). Retrieved from <https://arxiv.org/abs/1706.03762>