Using Large Language Models to Convert Documents to Knowledge Graphs to Check for Completeness and Consistency

by Michael Wacey

B.S. in Computer Science, June 1985, Washington University in St. Louis
M.S. in Computer Science, December 1990, Drexel University

A Praxis submitted to

The Faculty of
The School of Engineering and Applied Science
of The George Washington University
in partial satisfaction of the requirements
for the degree of Doctor of Engineering

March 8, 2026

Praxis directed by

Professor M. Elbasheer
Professorial Lecturer in Engineering and Applied Science

The School of Engineering and Applied Science of The George Washington University certifies that [Your Name] has passed the Final Examination for the degree of Doctor of Engineering as of December 18, 2025. This is the final and approved form of the Praxis.

Using Large Language Models to Convert Documents to Knowledge Graphs to Check for Completeness and Consistency

Michael Wacey

Praxis Research Committee:

Professor M. Elbasheer, Professorial Lecturer in Engineering and Applied Science, Praxis Director

XYZ, Professor of Online Engineering Programs, Chair

XYZ, Professorial Lecturer of Engineering and Applied Science, Committee Member

Dedication

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Acknowledgments

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Abstract

Using Large Language Models to Convert Documents to Knowledge Graphs to Check for Completeness and Consistency

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Table of Contents

List of Figures

List of Tables

Glossary of Terms

AI Artificial Intelligence

ANN Artificial Neural Network

API Application Programming Interface

GAT Graph Attention Networks

GCNN Graph Convolutional Neural Network

GNN Graph Neural Network

ML Machine Learning

NER Named Entity Recognition

SHAP SHapley Additive exPlanations

SVM Support Vector Machine

TN True Negative

TP True Positive

UBB User-Based Batching

UBS User-Based Sequencing

Chapter 1: Introduction

1.1 Background and Research Motivation

Ensuring document quality involves verifying completeness, consistency, and correctness (Zowghi & Gervasi, 2003). While evaluating correctness often necessitates access to knowledge external to the document and understanding the document's intent, completeness and consistency can be assessed within the document itself. This research focuses on developing automated methods using Large Language Models (LLMs) to address the latter two aspects. The specific focus is on converting a large document into a knowledge graph that can be used in future research to check the consistency and completeness of a document.

1.1.1 Background

The increasing complexity and scale of textual documents in various domains present significant challenges in ensuring consistency and completeness. Legal codes, technical documentation, and regulatory frameworks are often drafted collaboratively over extended periods, leading to inconsistencies, redundancies, and gaps in information. Traditional manual review methods, while necessary, are labor intensive and prone to human oversight, making automated solutions an attractive alternative. Advances in natural language processing (NLP) and artificial intelligence (AI) have introduced new methodologies for analyzing and structuring large bodies of text, with promising applications in document validation and knowledge extraction.

At the core of modern NLP advancements are Transformer-based models that rely on the Attention Mechanism to understand and generate text. LLMs, which build upon this foundation, can process and interpret vast amounts of textual data, though they are constrained by fixed context windows. To address this limitation, structured approaches such as knowledge graphs have emerged, enabling explicit representation of entities and relationships within documents. This research applies these technologies to Pennsylvania township laws, a domain where maintaining consistency is particularly critical. Given the size and complexity of municipal codes, inconsistencies in legal definitions, zoning regulations, and procedural rules can lead to legal disputes and financial losses. By leveraging AI-driven tools, this study aims to develop a framework for systematically analyzing and improving the consistency of legal documents.

Ensuring structural consistency and completeness in documents has been a longstanding challenge in various domains. Previous research has focused on methods to maintain internal coherence within documents (Laban et al., 2021), while other studies have explored domain-specific approaches to consistency checking (Tröls et al., 2022). In academic literature, the term coherence is often used interchangeably with consistency (Shen et al., 2021), reflecting the broader goal of ensuring logical and semantic alignment within textual content.

In 2017, a research team at Google introduced the Transformer model, a neural network architecture based entirely on the Attention Mechanism (Vaswani et al., 2017). Unlike previous sequential models, the Transformer processes all words within a given input simultaneously, allowing it to assess how each word influences others across the text. Using self-attention, this architecture captures long-range dependencies more effectively than previous models. However, despite advances in scaling Transformer-based models, they remain constrained by a limited attention window due to memory and computational efficiency considerations.

Large Language Models are built upon the Transformer architecture and inherit its fundamental attention-based mechanisms. However, LLMs are constrained by a fixed context window, limiting the amount of text they can analyze at once. As documents grow in length, they often exceed this window, preventing comprehensive processing in a single pass. Despite this limitation, document analysis does not necessarily require attending to an entire document at once. Instead, LLMs can be employed to extract key entities and concepts across different sections, enabling a more focused and structured approach to consistency checking. By identifying entities of interest and analyzing their relationships, LLMs can effectively navigate large documents while maintaining efficiency.

Knowledge graphs provide a structured, human-readable representation of information, serving as an alternative to the implicit encoding of knowledge found in neural networks. A knowledge graph is a directed acyclic graph in which nodes represent entities, and edges define relationships between them. Each node can possess attributes that enrich its descriptive properties. For instance, a node representing a car might include attributes such as color, model, or manufacturer. One useful way to conceptualize knowledge graphs is through the framework of frames, as described by Minsky (Minsky, 1974). Unlike LLMs, which rely on statistical inference, knowledge graphs offer explicit, interpretable relationships that can be leveraged for consistency and completeness verification in structured documents.

Pennsylvania is home to over 1,200 townships of the second class, each responsible for drafting and maintaining its own set of municipal laws. These laws regulate a wide range of local governance areas, including police services, fire departments, zoning, and land development. Over time, the cumulative nature of legal amendments introduces inconsistencies and gaps, which, if left unaddressed, can lead to legal ambiguities and enforcement challenges. While legal professionals and municipal officials work diligently to identify and resolve these issues, the complexity of these documents—often spanning thousands of pages—makes manual review error-prone and inefficient.

A key source of complexity is the interdependence of different sections within municipal codes. For example, early sections may define zoning regulations, specifying minimum frontage, setbacks, and other boundary constraints for different zoning districts. However, inconsistencies can arise when later sections introduce or reference zoning areas that were never formally defined. Similar discrepancies can emerge across other regulatory provisions, requiring careful synchronization of legal language and definitions. Ensuring consistency across these interconnected legal elements is a critical challenge that demands a more systematic and automated approach to legal document analysis.

1.1.2 Research Motivation

Despite extensive research on the analysis of small documents or specific sections of documents, there is a significant gap in addressing the challenges of comprehensive, large-scale document analysis. The need for automated consistency and completeness checks is critical in various industries. Currently, these tasks are often performed manually, requiring substantial time and resources while still potentially yielding suboptimal results. This research aims to bridge this gap by developing an effective and efficient automated solution.

Within the scope of this research, local regulations of townships in

Pennsylvania go through a time consuming and complicated process to be published. After the governing body enacts a law, it is sent to an organization to compile it into existing laws of the township. This is a manual and intensive process to determine if any of the existing laws are affected by the new law. Even with this, there are many cases of new laws that make a set of existing laws incomplete or inconsistent.

1.2 Problem Statement

Municipal laws in Pennsylvania Townships, authored by multiple people over time, develop inconsistencies and are incomplete (D. Curley, Easttown Supervisor, personal communication, September 16, 2024; A. Rau, Esq., Easttown Solicitor, personal communication, September 20, 2024; J. Sanders, personal communication, October 25, 2024), leading to annual revenue losses of hundreds of thousands of dollars. (A. Boscov, Easttown Supervisor, personal communication, September 23, 2024).

The complexity of municipal laws in Pennsylvania townships arises from their incremental development over time. Ordinances and regulations are often drafted by different individuals, including elected officials, legal counsel, and administrative staff, each contributing to the evolving legal framework. However, this decentralized process can lead to inconsistencies in language, overlapping provisions, and unintended gaps in regulatory coverage. As laws are amended or new ones are introduced, prior statutes may not be adequately reconciled, further exacerbating these inconsistencies. Without a systematic approach to maintaining legal coherence, townships face challenges in enforcing their laws effectively and equitably.

The consequences of these inconsistencies extend beyond legal ambiguity. Incomplete or conflicting municipal laws can create loopholes that hinder the township's ability to collect fees, fines, and other sources of revenue. For example, unclear zoning regulations may allow developments to proceed without appropriate permits or impact fees, and ambiguous tax ordinances may lead to disputes that reduce collections. When enforcement mechanisms are weak due to gaps in the legal framework, municipalities struggle to ensure compliance, leading to significant financial losses. These inefficiencies, compounded over time, place additional strain on local budgets, limiting resources for essential public services and infrastructure improvements.

Addressing these issues requires a structured methodology for analyzing, refining, and maintaining municipal laws. Traditional legal review processes, while valuable, are labor-intensive and reactive, often only identifying issues when disputes or financial shortfalls arise. Advances in artificial intelligence, particularly the use of LLMs, offer a potential solution by systematically identifying inconsistencies, redundancies, and gaps within legal texts. By applying LLMs to municipal laws, townships could proactively assess their legal frameworks, improving clarity, enforcement, and financial sustainability. However, implementing such an approach requires careful consideration of computational constraints, document formats, and the broader applicability of AI-driven legal analysis.

1.3 Thesis Statement

An LLM-based tool to convert a document into an attributed knowledge graph can be used to check for consistency and completeness will allow municipal lawyers to create consistent and complete law documents which prevent costly disputes and reduce revenue losses.

The application of LLMs in legal document analysis has the potential to revolutionize municipal law by providing an automated, systematic approach to ensuring consistency and completeness. Traditional legal drafting and review processes rely heavily on human oversight, which is inherently susceptible to errors, inconsistencies, and omissions—particularly in laws that have evolved over time through multiple amendments and contributors. By leveraging an LLM-based tool to convert legal documents into attributed knowledge graphs, municipalities can proactively identify gaps, redundancies, and contradictions before laws are enacted or enforced. This proactive approach minimizes ambiguity, strengthens legal clarity, and enhances the efficiency of legal review processes.

A knowledge graph-based representation of municipal laws enables a structured, machine-readable format that facilitates logical analysis. Unlike traditional text-based legal review, which requires extensive manual effort to trace dependencies and resolve conflicts, a knowledge graph explicitly maps relationships between legal provisions, definitions, and enforcement mechanisms. This allows municipal lawyers to assess the interconnectivity of legal clauses and verify their consistency against established legal principles and precedents. Furthermore, an attributed knowledge graph can highlight areas where laws are incomplete or misaligned with overarching governance policies, enabling timely revisions that improve legal coherence.

Beyond legal clarity, the ability to create consistent and complete municipal laws has direct financial implications. Inconsistent or incomplete regulations can lead to disputes over zoning, taxation, permitting, and compliance, often resulting in costly litigation or lost revenue due to unenforceable provisions. By employing an LLM-driven tool to detect and resolve these issues at the drafting stage, municipalities can reduce legal ambiguities that might otherwise be exploited, streamline enforcement mechanisms, and enhance overall regulatory efficiency. This, in turn, strengthens fiscal sustainability by preventing revenue leakage and ensuring that all applicable fees, fines, and taxes are properly assessed and collected.

The integration of LLM-based tools in municipal lawmaking represents a transformative step toward modernizing local governance. As artificial intelligence continues to advance, municipalities that adopt such technologies will gain a significant advantage in maintaining a legally sound, financially sustainable framework. Future research can extend this approach beyond municipal laws to other domains of legal and regulatory governance, demonstrating the broader impact of AI-driven knowledge representation in ensuring legal accuracy, reducing administrative burdens, and enhancing public trust in local government operations.

1.4 Research Objectives

The primary objective of this research is to develop a tool capable of automatically processing documents of any size into a coherent set of entities in a knowledge graph. This tool will leverage advanced techniques to analyze document content, identify potential entities, and provide access to the knowledge graph.

The created knowledge graph will be analyzed to determine whether it is appropriate to check the document for inconsistencies and incompleteness. This will include introducing issues in the source documents and then highlighting how easy they are to observe in the knowledge graph.

1.5 Research Questions

To achieve the research objectives, the following research questions will be considered.

RQ1: Can an LLM be used to convert a large document into a knowledge graph?

RQ2: Can an LLM be used to process multiple knowledge graphs into a typed cluster of knowledge graphs.

RQ3: Can a typed cluster of knowledge graphs be used to check the source document for consistency and completeness?

1.6 Research Hypotheses

Research will be conducted to test the following hypotheses.

H1: An LLM can be used to convert a large document into a knowledge graph.

H2: An LLM can be used to process multiple knowledge graphs into a typed cluster of knowledge graphs.

H3: A typed cluster of knowledge graphs can be used to check the source document for consistency and completeness.

1.7 Research Scope and Limitations

The subsequent sections outline the scope and limitations of this study, which employs Pennsylvania township laws as a case study to develop and evaluate an automated tool for the analysis of legal documents. These publicly accessible laws, having undergone extensive manual reviews for consistency and completeness, provide a rigorous benchmark for assessing the proposed methodology. The primary focus of this research is the construction of Knowledge Graphs that faithfully represent the structure and content of the documents, thereby laying the groundwork for future efforts in verifying legal consistency and completeness. Notwithstanding, this study acknowledges several inherent limitations, including computational constraints, challenges associated with specific document formats and linguistic nuances, and the primary emphasis on textual analysis. These limitations underscore the necessity for continued research to enhance and broaden the applicability of the proposed approach.

1.7.1 Research Scope

This study focuses on the use of Pennsylvania township laws as a case study for developing and testing an automated tool designed to analyze legal documents. These laws, which are publicly available in both PDF and Word formats, were selected due to their complexity, extensive length, and the fact that they have been authored by multiple contributors over time. Additionally, they have undergone rigorous manual reviews for consistency and completeness, making them an ideal benchmark for evaluating the effectiveness of the proposed approach. While the primary application is in the legal domain, the methodology is designed to be adaptable for broader use across various document types.

The core development in this research centers on constructing Knowledge Graphs that accurately represent the structure and content of the documents under review. These graphs will serve as a foundation for future work in verifying legal consistency and completeness. While the study will assess the suitability of the generated Knowledge Graphs for such tasks, the actual implementation of automated consistency and completeness checks will be left for future research. This approach ensures a focused and systematic exploration of Knowledge Graph generation while laying the groundwork for subsequent advancements in automated legal analysis.

1.7.2 Research Limitations

This study has several potential limitations. Computational constraints may affect the efficiency and scalability of processing large and complex legal documents. Challenges may also arise in handling specific document formats and language intricacies, particularly in ensuring accurate interpretation and structuring of legal text. Additionally, while this research focuses on leveraging LLMs such as Gemini and ChatGPT, it does not develop specialized models tailored for knowledge graph construction, consolidation, or query answering—an approach that could reduce computational costs and energy consumption.

This research does not perform direct testing for consistency and completeness. Instead, it utilizes Pennsylvania township laws, which are publicly available and have already undergone such validation. Future studies should explore the applicability of this approach to a broader range of legal and non-legal documents.

For document handling, this research mainly uses Word documents to facilitate modifications during testing. Although the methodology should also be compatible with PDFs, further research is needed to confirm seamless integration and processing across different formats.

Finally, this study is limited to textual analysis. Future research could expand upon this work by incorporating additional elements such as tables, formulas, images, and diagrams to improve document comprehension and analysis.

1.8 Praxis Organization

The remainder of this research is organized into several key chapters. Chapter 2 provides a comprehensive review of the relevant literature, focusing on the creation of knowledge graphs from documents by LLMs, the processing of multiple knowledge graphs into a combined knowledge graph by LLMs, the utility of knowledge graphs in representing the original document to ensure consistency and completeness, the process of creating and maintaining local laws in Pennsylvania, and background information on checking documents for consistency and completeness. Chapter 3 delves into the statistical and machine learning methodologies employed in this research, detailing the processes of data pre-processing, model selection, training, and evaluation. Chapter 4 presents and analyzes the results of the data analysis, addressing each research question and hypothesis while evaluating the performance of the proposed methodology and tool. Finally, Chapter 5 concludes the investigation with a discussion of the key findings, contributions to the field, recommendations for practical applications, and potential avenues for future research.

Chapter 2: Literature Review

2.1 Introduction

The landscape of Artificial Intelligence (AI), particularly Natural Language Processing (NLP), was significantly reshaped by the groundbreaking work conducted at Google Brain and documented in the seminal paper *Attention Is All You Need* (Vaswani et al., 2017). This paper introduced the Transformer architecture, leveraging self-attention mechanisms, which became the foundation for modern Large Language Models (LLMs). These models have demonstrated remarkable capabilities across a wide range of tasks, including text generation, summarization, translation, and question answering, often producing outputs nearly indistinguishable from human writing (Badshah & Sajjad, 2024).

Despite these advancements, LLMs possess an inherent architectural limitation: a finite context window. This window represents the maximum amount of text (measured in tokens) that the model can process simultaneously when generating a response or performing an analysis. Consequently, if critical information or dependencies exist within a document but fall outside this fixed window—separated by a larger span of intervening text—the LLM may fail to capture the relationship or address the query accurately. This limitation poses a significant challenge when dealing with large or complex documents where understanding relies on synthesizing information across distant sections.

One promising approach to mitigate this limitation involves transforming large, unstructured documents into structured representations using Knowledge Graphs (KGs). By extracting key entities, relationships, and attributes from the text and mapping them into a graph structure, it becomes possible to represent the document's core semantic content in a format amenable to computational analysis. This allows for querying and reasoning over the entire document's scope, independent of the LLM's context window constraints, potentially enabling a more focused and comprehensive analysis.

This chapter reviews the pertinent literature underpinning this approach. It begins by examining the development and characteristics of Large Language Models, focusing on their capabilities and limitations, particularly the context window constraint. Subsequently, it delves into the principles, construction, and application of Knowledge Graphs as structured knowledge representations. The challenges associated with processing large documents, especially within the legal domain, are then discussed. The role of Named Entity Recognition (NER) as a key technology for bridging text and KGs is explored. Finally, the chapter defines and discusses the critical concepts of Consistency, Completeness, and Coherence, particularly relevant for evaluating the integrity of legal document corpora, and how the proposed KG-based approach may aid in their assessment.

2.2 Large Language Models

The trajectory of modern NLP took a significant turn in 2017 with the publication of *Attention Is All You Need* by Vaswani et al. (Vaswani et al., 2017). This work introduced the Transformer architecture, which uniquely relies on self-attention mechanisms to weigh the importance of different words (tokens) in the input sequence, enabling superior handling of long-range dependencies compared to previous recurrent or convolutional architectures (Turner, 2024). This innovation paved the way for the development of increasingly large and powerful language models, such as Google's BERT (Koroteev, 2021) and OpenAI's Generative Pre-trained Transformer (GPT) series (Gao et al., 2023).

While research groups at numerous institutions continuously pursued improvements, the public release of OpenAI's *ChatGPT* (based on the GPT-3.5 architecture) on November 30, 2022, marked a pivotal moment, dramatically increasing public awareness and accelerating the deployment of conversational AI systems. This catalyzed the release and further development of competing models, including Google's *Gemini* (Team et al., 2024), Anthropic's *Claude* (Caruccio et al., 2024), and Meta's open-source *Llama* family (Grattafiori et al., 2024). The proliferation of models is evident on platforms like Hugging Face, a central repository for AI models and datasets, which reportedly surpassed one million hosted models by late 2024 (Edwards, 2024).

Functionally, LLMs process input text (the "prompt") by first converting it into numerical representations called tokens. Using the patterns learned during extensive pre-training on vast text corpora, the model then predicts subsequent tokens to generate a coherent and contextually relevant output. Prompts can be complex, potentially including substantial amounts of text for analysis. For instance, an LLM might be prompted with a company's annual report and asked specific questions about its contents. Many current LLMs can perform reasonably well on such tasks, provided the relevant information falls within their processing limits (Rzepka et al., 2023).

However, the fundamental limitation remains the context window size. This size, while increasing with newer model generations (ranging from a few thousand to potentially hundreds of thousands of tokens), is always finite (Liu et al., 2025) and (Kaplan et al., 2020). If a document exceeds this limit, the LLM cannot process it in its entirety in a single pass. Standard techniques involve processing chunks sequentially, but this can sever long-distance contextual links. For example, determining if a policy statement on page 1 of a lengthy legal code is adequately supported or contradicted by detailed regulations presented on pages 788-795 might be impossible if the intervening text exceeds the context window. The LLM would process the sections independently, unable to synthesize the relationship between them.

Furthermore, processing information up to the maximum context window incurs significant computational costs. The self-attention mechanism typically scales quadratically with sequence length, demanding substantial memory (RAM/VRAM), processing power (CPU/GPU), and energy resources (Cong, 2024). This quadratic scaling makes analyzing very large documents prohibitively expensive or slow, further motivating alternative approaches like KG-based structuring for comprehensive analysis.

2.3 Knowledge Graphs

Knowledge Graphs (KGs) provide a structured way to represent information and knowledge, drawing on concepts from semantic networks and earlier AI research in knowledge representation **[CITE - KG Overview/History]**. Formally, a KG represents knowledge as a collection of interconnected entities (nodes) and the relationships (edges) between them. Both nodes and edges can possess attributes or properties that store additional metadata.

The core components of a KG are:

- **Nodes (Entities):** Represent objects, concepts, or entities of interest (e.g., persons, organizations, locations, legal statutes, defined terms).

- **Edges (Relationships):** Represent the connections or relationships between nodes (e.g., 'works for', 'located in', 'cites', 'amends', 'defines'). Edges are typically directed and labeled with the relationship type.

- **Attributes (Properties):** Key-value pairs associated with nodes or edges, providing additional details (e.g., a 'Person' node might have an 'email' attribute; a 'cites' edge might have a 'date' attribute).

A simple example could involve representing that 'Section 5.A' of a 'Township Code' (entities) 'amends' (relationship) 'Section 3.B' of the same code, effective on '2024-07-15' (attribute of the 'amends' relationship).

Pioneering work in structured knowledge includes Minsky's concept of Frames (Minsky, 1974), which represented stereotypical situations using slots (attributes) and relationships, influencing subsequent knowledge representation formalisms.

Knowledge graphs can be implemented and stored using various technologies. Common approaches include:

- **RDF (Resource Description Framework):** A W3C standard based on triples (subject-predicate-object), often queried using SPARQL **[CITE - RDF/SPARQL]**.

- **Property Graphs:** A flexible model popular in graph databases (e.g., Neo4j, Neptune), allowing properties on both nodes and edges **[CITE - Property Graphs**

- **Graph Neural Networks (GNNs):** While not a storage mechanism, GNNs learn representations from graph structures, enabling tasks like link prediction or node classification within KGs **[CITE - GNNs]**.

- **Other Formats:** KGs can also be serialized using formats like JSON or XML, though these may lack the querying capabilities of dedicated graph databases.

KGs are employed in diverse applications, including enhanced information retrieval, semantic search, recommendation systems, data integration, and complex question answering **[CITE - KG Applications]**. Their ability to explicitly model relationships makes them potentially valuable for analyzing the structure and integrity of document collections.

2.4 Challenges in Large Document Processing

Research efforts in document processing are extensive, covering tasks like summarization **[CITE - Summarization]**, information extraction **[CITE - Info Extraction]**, document classification **[CITE - Doc Classification]**, and validating the faithfulness of summaries against source documents **[CITE - Summary Validation]**. Historically, much research focused on relatively small documents for several practical reasons. Smaller documents are computationally less demanding to process, and crucially, human evaluation and annotation are more feasible, allowing researchers to establish ground truth and verify system performance accurately.

However, many real-world applications involve documents that are significantly larger, presenting distinct challenges:

- **Computational Resources:** Processing large volumes of text demands substantial memory, storage, and processing time, scaling with document length.

- **Long-Range Dependencies:** Capturing semantic connections, references, or dependencies between sections that are far apart in the document is difficult for models with limited context windows (as discussed regarding LLMs).

- **Context Fragmentation:** Common techniques for handling large documents involve splitting them into smaller chunks **[CITE - Chunking Strategies]**. While necessary for models with fixed input sizes, this risks losing context that spans across chunk boundaries.

- **Evaluation Complexity:** Assessing the quality of processing (e.g., summarization, consistency analysis) for large, complex documents is inherently difficult for human evaluators, making ground truth creation and validation resource-intensive.

Techniques like Retrieval-Augmented Generation (RAG) **[CITE - RAG]** attempt to address some limitations by allowing LLMs to retrieve relevant snippets from a large corpus before generating a response. While effective, RAG typically retrieves discrete chunks rather than providing a holistic, structured view of the entire document's content and relationships, a gap potentially filled by KGs.

2.5 Challenges and Opportunities in Analyzing Legal Documents

Legal documents, particularly statutory or regulatory codes, represent a compelling domain for developing and evaluating advanced document analysis techniques. They possess several characteristics that make them both challenging and informative testbeds:

- **Complexity and Precision:** Legal language is often dense, employing specialized terminology (jargon), intricate sentence structures, and numerous cross-references. Ambiguity must be minimized, demanding high precision in interpretation and analysis **[CITE - Legal Language Complexity]**.

- **Volume and Interconnectedness:** Legal corpora can be vast (e.g., state statutes, federal regulations, municipal codes). Documents within these corpora are highly interconnected through citations, amendments, and definitions.

- **Semi-structured Format:** While often organized into articles, sections, and clauses, legal texts contain significant amounts of unstructured prose requiring sophisticated NLP techniques.

- **Critical Need for Integrity:** The consistency, completeness, and coherence of legal documents are paramount for their fair application, predictability, and enforceability. Flaws can lead to confusion, disputes, and litigation.

In this context, the specific focus on the codified ordinances (laws) of townships within the Commonwealth of Pennsylvania provides a valuable and concrete dataset. With over 1,200 townships (according to standard classifications, though the exact number of codified ordinances might vary), there is a substantial body of material available, exhibiting realistic complexity and evolution. These codes are often developed over many years, involving multiple authors, amendments, and review cycles.

The legislative process itself, while designed to ensure quality, highlights the potential for introducing errors. When a new ordinance is proposed—whether by elected officials or residents—it undergoes review by township staff and legal counsel (the solicitor) who drafts the formal language. A compiler may be engaged to integrate the new law and perform validation checks. Despite this multi-stage human review process involving various stakeholders, inconsistencies (contradictions with existing laws), incompleteness (missing definitions or procedures), and incoherence (unclear structure or references) can still arise, especially as the code grows over time **[CITE - Challenges in Legislative Drafting]**. The resource-intensive and fallible nature of manual review motivates the exploration of computational methods to assist in maintaining the integrity of these legal documents.

2.6 Named Entity Recognition

Named Entity Recognition (NER) is a fundamental task in information extraction that focuses on identifying and classifying named entities within unstructured text into pre-defined categories **[CITE - NER Definition]**. These categories typically include standard types like persons, organizations, locations, dates, and monetary values, but can be extended to domain-specific entities.

In the context of building knowledge graphs from text, NER plays a crucial role. It serves as the primary mechanism for identifying the potential **nodes** (entities) that will populate the graph. By extracting key actors, locations, concepts, and defined terms from the source documents, NER provides the raw material for structured representation. Furthermore, the identified entities often serve as anchors for identifying the **relationships** (edges) between them.

Various methods have been developed for NER:

- **Rule-based Systems:** Rely on hand-crafted grammatical rules and dictionaries (gazetteers). Often precise but brittle and difficult to scale **[CITE - Rule-Based NER]**.
- **Statistical Models:** Techniques like Hidden Markov Models (HMMs) and Conditional Random Fields (CRFs) learn patterns from annotated data **[CITE - Statistical NER]**.
- **Deep Learning Approaches:** Modern NER systems frequently utilize deep neural networks, such as Bidirectional Long Short-Term Mem-

ory networks (BiLSTMs) often combined with CRFs, or increasingly, Transformer-based models like BERT, which achieve state-of-the-art performance on many benchmarks **[CITE - Deep Learning NER]**.

Applying NER to the legal domain requires careful consideration of domain-specific entities. Beyond standard types, entities critical for legal analysis include: specific legal statutes or sections (e.g., '15 Pa.C.S.A. § 1502'), defined legal terms (e.g., 'applicant', 'nonconforming use'), roles (e.g., 'Township Supervisor', 'Zoning Officer'), specific dates or deadlines, and references to other documents or sections **[CITE - Legal NER]**. Training or fine-tuning NER models on legally annotated corpora is often necessary to achieve high accuracy due to the specialized vocabulary and context. The output of a robust legal NER system provides the essential building blocks for constructing a meaningful knowledge graph from legal texts.

2.7 Consistency, Completeness, and Coherence

When analyzing formal document corpora, particularly legal codes, evaluating their quality often involves assessing their Consistency, Completeness, and Coherence. These concepts are crucial for ensuring the documents are understandable, reliable, and effective in their intended function.

- **Consistency:** Refers to the absence of contradictions within the document set **[CITE - Consistency Definition]**. A consistent legal code should not contain provisions that assert mutually exclusive facts or prescribe conflicting obligations or permissions. For example, one section cannot mandate an action that another section explicitly prohibits under the same conditions. Detecting inconsistencies is vital for legal certainty.

- **Completeness:** Pertains to whether the document set contains all the necessary information required for its intended scope **[CITE - Completeness Definitio** In a legal context, this could mean ensuring that all relevant terms are defined, procedures are fully specified, criteria for decisions are enumerated, and potential scenarios are addressed. Gaps or omissions can lead to ambiguity and disputes. Assessing completeness is often challenging as it requires a clear definition of the intended scope and domain knowledge.

- **Coherence:** Relates to the logical organization, clarity, and interconnectedness of the information presented **[CITE - Coherence Definition]**. A coherent document flows logically, uses terminology consistently, ensures cross-references are accurate and meaningful, and maintains a clear structure. While related to consistency, coherence addresses the overall understandability and usability of the document, ensuring that the parts fit together meaningfully.

Ensuring these qualities in large, evolving legal codes through manual review is exceptionally difficult due to the sheer volume of text, the complexity of interdependencies, and the distributed nature of authorship over time. This is where computational approaches leveraging LLMs and KGs offer potential advantages.

A Knowledge Graph, by explicitly modeling entities (like defined terms, sections, obligations) and their relationships (like 'defines', 'cites', 'conflicts with'), provides a structured substrate for analysis. Queries can be designed to automatically detect certain types of inconsistencies (e.g., finding terms used before they are defined, identifying conflicting property values for the same entity under specific conditions, detecting circular references). While perfect completeness is often unattainable or ill-defined, KGs can help identify potential gaps, such as undefined terms or procedures referenced but not detailed, by analyzing the graph's structure for missing nodes or expected relationships. LLMs can play a role in the pipeline by aiding in the initial interpretation of text to populate the KG and potentially helping to formulate queries or summarize the findings from the graph analysis. However, the KG provides the persistent, queryable structure necessary for systematic checks that go beyond the limitations of an LLM's context window. Research exploring the use of KGs and related AI techniques for automated consistency and completeness checking in software requirements, logical formalisms, and legal texts provides a foundation for this approach **[CITE - Automated Consistency/Completeness Checking]**. This thesis aims to build upon such work, applying it to the specific context of municipal legal codes.

Chapter 3: Methodology

3.1 Introduction

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.2 Some Subsection

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel

leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.3 Some Subsection

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.3.1 Some Subsection

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.3.2 Some Subsection

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.4 Conclusion

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

At the end of Chapter 3, you must restate your research questions and hypotheses exactly as they are. The format should follow the structure shown below. Please Ensure all text coloring is removed.:

RQ1:

RQ2:

RQ3:

H1:

H2:

H3:

Example:

This methodological approach addresses the research questions and hypotheses outlined in Chapter 1 and repeated below:

RQ1: You will repeat your research question1 here?

RQ2: You will repeat your research question2 here?

RQ3: RQ2: You will repeat your research question3 here?

H1: RQ2: You will repeat your research Hyothesis1 here.

H2: You will repeat your research Hyothesis2 here.

H3: You will repeat your research Hyothesis2 here.

Leftover from chapter 2 - holding here

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida

mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.5 Some Section

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

This is how you insert an image by referencing it like this: Figure 3.1 shows XYZ.

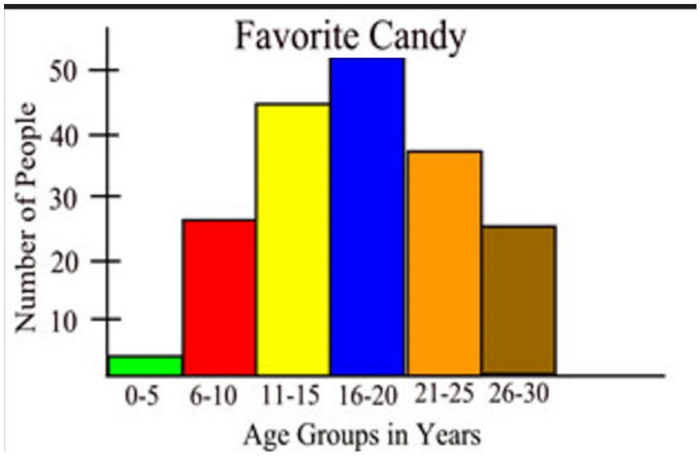

Figure 3.1: *Histogram of XYZ*

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.5.1 Subsection 1

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

This is how to enter an equation and reference it. Equation 3.1 and 3.1 show how XYZ is implemented.

$$TE_{(pos,2i)} = sin(pos/23^{2i/Lm}) \tag{3.1}$$

$$KN_{(pos,2i+1)} = cos(pos/453^{2i/Lm}) \tag{3.2}$$

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.6 Another Section

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

This is the correct format for a table, including the caption and proper alignment (centered and italicized).

Table 3.1: *My Table About Something*

Reference	**Technique**
something here	Technique here
something here	Technique here
something here	Technique here
something here	Technique here
something here	Technique here

3.6.1 Subsection

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.6.2 Subsection

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.6.2.1 Sub subsection

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.7 Summary

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Chapter 4: Results and Analysis

4.1 Introduction

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

4.2 Some Subsection

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel

leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

4.3 Some Subsection

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

This is how you build and format the table. The caption is centered and italized and in the top. The table number is bolded. Please remove the red coloring i use to illutrate this.

Table 4.1 presents an overview of xyz and zkm and xyz xyz. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit,vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris

Table 4.1: *Title of the table every first letter capitalized*

Factor1	Test 1	Test 2
Something here	123	123
Something here	123	123
Something here	123	1123
Something here	16	123
Something here	123	123
Something here	123	123

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

4.4 Some Subsection

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida

mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

This is the correct way to add and format an equation. Ensure the formatting is consistent, and remove any red coloring used for illustration.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4.1}$$

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum..

This is the correct way to add and format an equation. Ensure the formatting is consistent, and remove any red coloring used for illustration.

$$Precision = \frac{TP}{TP+FP} \tag{4.2}$$

4.5 Some Subsection

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

This is the correct way to build and format a table. The caption should be centered, italicized, and placed at the top. The table number should be bolded. In this example, the table is long, so ensure the table font matches the font of your document. It is not allowed to significantly reduce the table font size just to make it fit. In this case, I used a slightly smaller font, depending on the situation—only if necessary. However, try to avoid having too many columns to maintain readability. Make sure to remove the red coloring used for illustration.

Table 4.2 depicts the xxxx.

Table 4.2: *Test-2: Transformer vs. AutoTAB Performance Metrics*

Model	Method	A	P	R	F1	AUC	FNR	FPR
Mod1	Sub1	**0.0123**	**0.0123**	**0.0123**	**0.0123**	**0.0123**	**0.0123**	**0.0123**
	Sub2	0.0123	0.0123	0.0123	0.0123	0.0123	0.0123	0.0123
	Sub3	0.0123	0.0123	0.0123	0.0123	0.0123	0.0123	0.0123
Mod2	Sun1	0.0123	0.0123	0.0123	0.0123	0.0123	0.0123	0.0123
	Sub2	0.0123	0.0123	0.0123	0.0123	0.0123	0.0123	0.0123
	Sub3	0.0123	0.0123	0.0123	0.0123	0.0123	0.0123	0.0123

4.5.0.1 Sub Subsection

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

4.5.1 Subsection

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida

mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

4.6 Some Section

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

4.6.1 Conclusion

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

At the end of Chapter 3, you must restate your research questions and hypotheses exactly as they are. The format should follow the structure shown below. Please Ensure all text coloring is removed.:

RQ1:

RQ2:

RQ3:

H1:

H2:

H3:

Example:

The results from this Chapter address the research questions and hypotheses outlined in Chapter 1 and repeated below:

RQ1: How do Transformer encoders compare to Autoencoders in terms of accuracy, precision, and recall when detecting malicious insider threats?

RQ1: You will repeat your research question1 here?

RQ2: You will repeat your research question2 here?

RQ3: RQ2: You will repeat your research question3 here?

H1: RQ2: You will repeat your research Hyothesis1 here.

H2: You will repeat your research Hyothesis2 here.

H3: You will repeat your research Hyothesis2 here.

Chapter 5: Discussion and Conclusions

5.1 Conclusion

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

5.2 Contribution to the Body of Knowledge

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel

leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

5.3 Recommendations for Future Research

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

References

Badshah, S., & Sajjad, H. (2024, May). *Quantifying the capabilities of llms across scale and precision*. http://arxiv.org/abs/2405.03146

Caruccio, L., Cirillo, S., Polese, G., Solimando, G., Sundaramurthy, S., & Tortora, G. (2024). Claude 2.0 large language model: Tackling a real-world classification problem with a new iterative prompt engineering approach. *Intelligent systems with applications*, *21*, 200336. https://dx.doi.org/10.1016/j.iswa.2024.200336

Cong, Y. (2024, June). Research for enhancing processing and computational efficiency in llm. In *Proceedings of the 2024 2nd international conference on image, algorithms and artificial intelligence (iciaai 2024)* (pp. 970–980). Atlantis Press. https://www.atlantis-press.com/proceedings/iciaai-24/126004157

Edwards, B. (2024, September). *Exponential growth brews 1 million ai models on hugging face*. https://arstechnica.com/information-technology/2024/09/ai-hosting-platform-surpasses-1-million-models-for-the-first-time/

Gao, K., He, S., He, Z., Lin, J., Pei, Q., Shao, J., & Zhang, W. (2023, August). *Examining user-friendly and open-sourced large gpt models: A survey on language, multimodal, and scientific gpt models*. http://arxiv.org/abs/2308.14149

Grattafiori, A., Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Vaughan, A., Yang, A., Fan, A., Goyal, A., Hartshorn, A., Yang, A., Mitra, A., Sravankumar, A.,

..., & Ma, Z. (2024, November). *The llama 3 herd of models*. http://arxiv.org/abs/2407.21783

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020, January). *Scaling laws for neural language models*. http://arxiv.org/abs/2001.08361

Koroteev, M. V. (2021, March). *Bert: A review of applications in natural language processing and understanding*. http://arxiv.org/abs/2103.11943

Laban, P., Dai, L., Bandarkar, L., & Hearst, M. A. (2021). Can transformer models measure coherence in text? re-thinking the shuffle test. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2107.03448

Liu, J., Zhu, D., Bai, Z., He, Y., Liao, H., Que, H., Wang, Z., Zhang, C., Zhang, G., Zhang, J., Zhang, Y., Chen, Z., Guo, H., Li, S., Liu, Z., Shan, Y., Song, Y., Tian, J., Wu, W., ... Zhang, Z. (2025, March). *A comprehensive survey on long context language modeling*. http://arxiv.org/abs/2503.17407

Minsky, M. (1974, June). *A framework for representing knowledge*. http://hdl.handle.net/1721.1/6089

Rzepka, R., Muraji, S., & Obayashi, A. (2023, December). Expert evaluation of export control-related question answering capabilities of llms [ID: cdi$_{ieee_{p}rimary_{1}}$0487735]. In *Ieee asia-pacific conference on computer science and data engineering (csde)* (pp. 1–6). IEEE.

Shen, A., Mistica, M., Salehi, B., Li, H., Baldwin, T., & Qi, J. (2021). Evaluating document coherence modeling. *Transactions of the Association for Computational Linguistics*, *9*, 621–640. https://doi.org/10.1162/tacl_a_00388

Team, G., Anil, R., Borgeaud, S., Alayrac, J.-B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A. M., Hauth, A., Millican, K., Silver, D., Johnson, M., Antonoglou, I., Schrittwieser, J., Glaese, A., Chen, J., Pitler, E., ..., & Vinyals, O. (2024, June). *Gemini: A family of highly capable multimodal models*. http://arxiv.org/abs/2312.11805

Tröls, M. A., Marchezan, L., Mashkoor, A., & Egyed, A. (2022). Instant and global consistency checking during collaborative engineering. *Software and systems modeling*, *21*(6), 2489–2515. https://doi.org/10.1007/s10270-022-00984-4

Turner, R. E. (2024, February). *An introduction to transformers*. http://arxiv.org/abs/2304.10557

Vaswani, A., Shazeer, N., Brain, G., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Kaiser, Ł. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.

Zowghi, D., & Gervasi, V. (2003). On the interplay between consistency, completeness, and correctness in requirements evolution. *Information and Software Technology*, *45*(14), 993–1009. https://doi.org/10.1016/S0950-5849(03)00100-9