



UNIVERSITY OF BAHRAIN

College of Business Administration

Department of Management & Marketing

**Optimization of Decision-Making in Multi-Stakeholder
Architectural, Engineering, and Construction (AEC) Projects:
A Novel Data-Driven Framework**

A Thesis Submitted in Partial Fulfilment of the Requirements for the
Doctor of Philosophy Degree in Business Analytics

Submitted by

Entesar Jasim Mahdi Al Nama

University ID: 19990435

Supervised by

Dr. Maqsoud Mahmood

Supervisor

Dr. Huda Al Madhoob

Co- Supervisor

Kingdom of Bahrain

May 2025



Thesis Discussion Committee Decision

Form Code: UDNGS013

Student Name:	Entesar Jasim Mahdi Al Nama
Student ID. No.	19990435
Department:	Management and Marketing
College:	Business Administration
Program:	PhD in Business Analytics - Management
Student GPA:	3.5
Thesis Title:	Optimization of Decision-Making in Multi-Stakeholder Architectural, Engineering, and Construction (AEC) Projects: A Novel Data-Driven Framework

The Discussion Committee met on 25 May 2025 in S1B-218 upon discussing the student whose details are stated above in public. The discussion committee recommends the following:

The student is awarded the grade Pass in the thesis course MGT798 BAA.
BAB
BAC
BAD

☐ Without making any changes to the thesis.

☒ After making the modifications attached with this decision.

Discussion Committee Members Approval:

1.	Supervisor Name:	Dr. Maqsood Mahmud Dr Huda Almadhoob	Sign.:	
2.	Internal Examiner:	Prof Eslam Elghonaimy	Sign.:	
3.	External Examiner:	Chris Nugent	Sign.:	



Dissertation Deposit Form

Form Code: UDNGS025

Author Name:	Entesar Jasim Mahdi Al Nama
Thesis Title:	Optimization of Decision-Making in Multi-Stakeholder Architectural, Engineering, and Construction (AEC) Projects: A Novel Data-Driven Framework
Year of Publication:	2025
Department:	Management and Marketing
College:	Business Administration
E-mail:	entesarjasim@gmail.com
Contact No.	entesarjasim@gmail.com

Author's Declaration: I agree to the following Conditions:

The above thesis will be available (in the University of Bahrain Library and externally) and reproduced as necessary at the discretion of the University of Bahrain. It may also be digitized by the University of Bahrain and made available via the university's repository or via other parties on the Internet.

This condition shall apply to all copies including electronic copies.

The above thesis has been provided on the understanding that it is copyright protected material and that no quotation from this thesis may be published without paper acknowledgement.

Sign. of the Author:	Entesa	Date:	May 2025
----------------------	--------	-------	----------

Abstract

The complexity and dynamic nature of regulatory compliance in the Architectural, Engineering, and Construction (AEC) industry create significant challenges for stakeholders, often leading to delays, inefficiencies, and costly errors. Variations in interpretation, reliance on tacit knowledge, and evolving regulations further complicate decision-making. This study addresses these challenges by developing a Custom GPT model, identified as the best fit for small datasets and specialized Q&A applications, to provide instant, accurate, and context-specific answers to regulatory queries. The study is applied within the Kingdom of Bahrain, utilizing Chapter (14) of the Bahraini Unified Guidebook of Building Permit Regulations as its primary data source.

The study compares multiple LLMs to determine the most effective model for handling a small, domain-specific dataset related to regulatory compliance. The Custom GPT model achieved 100% across all performance metrics, while the base model, ChatGPT-4.0, recorded an accuracy score of 98.8%. These results demonstrate that CustomGPT outperforms other models by delivering more precise and contextually relevant answers, making it the most suitable choice for developing an AI-driven regulatory compliance assistant.

This study introduces and employs systematic 10-step methodology, that integrates human-AI collaboration, expert feedback, and domain-specific fine-tuning to develop a scalable compliance framework. The study is grounded in Knowledge Management (KM) Theory, facilitating the transformation of tacit expert knowledge into structured, actionable insights.

Expert interviews were conducted with four out of six domain specialists from the authority overseeing Chapter (14), representing a sampling rate of 66.7%. The findings confirm that Custom GPT significantly improves decision accuracy, accelerates access to regulatory knowledge, and fosters collaboration among AEC professionals. The model serves as a centralized knowledge hub, bridging the gap between complex regulations and practical implementation. Its scalability allows for potential integration across additional regulatory chapters and broader industry domains, supporting digital transformation in compliance management.

Despite its effectiveness, challenges remain in data availability, generalizability across jurisdictions, and handling complex interpretive queries. Future research should explore multilingual capabilities, adaptive AI frameworks, and integration with Building Information Modeling (BIM) systems to enhance real-world applicability.

This study contributes to AI-driven compliance solutions in the AEC industry, demonstrating the transformative potential of data-driven methodologies in regulatory adherence, decision-making optimization, and stakeholder collaboration.

Publications

- 1) Al Nama, E., & Alalawi, A. (2023). The Adoption of Automated Building Code Compliance Checking Systems in the Architecture, Engineering, and Construction Industry. *Proceedings of the 2023 International Conference on Cyber Management and Engineering (CyMaEn 2023)*, IEEE, Bangkok, Thailand, pp. 289.
- 2) Al Nama, E., & Mahmud, M. (2024). Building the Blueprint for AI-powered Compliance Checking: Analyzing ChatGPT-4 & Gemini by Question Category in Engineering Regulations. *Proceedings of the 2024 International Conference on Computing in Civil and Building Engineering (ICCCBE 2024)*, Montreal, Canada.
- 3) Al Nama, E., Mahmud, M., & Almadhoob, H. (2024). Leveraging Human Expertise and AI for Engineering Regulatory Data Expansion: A Case Study with ChatGPT. *Proceedings of the 2024 International Conference on Decision Aid Sciences and Applications (DASA 2024)*, IEEE, Bahrain.
- 4) Submitted Manuscript (Under Review): Al Nama, E., Mahmud, M., & Almadhoob, H. (2025). Enhancing Regulatory Compliance in the AEC Industry via LLM-Powered Decision Frameworks. *Submitted to Journal of Information Technology in Construction (ITcon)*, published by the International Council for Research and Innovation in Building and Construction (CIB).

Dedication

“To my beloved family, who planted the seeds of knowledge and nurtured my insatiable curiosity. Your unwavering prayers and steadfast support fuel my lifelong journey of exploration”

Acknowledgment

I extend my heartfelt appreciation to the following individuals and entities who contributed to my research journey:

My supervisors: *Their guidance, feedback, and unwavering support shaped the trajectory of my work.*

Experts from CRPEP: *Their engaging discussions and collaborative efforts enriched the research model.*

Recruited users - from AEO and Architects & Engineers from AEC

Industry: *Their participation enabled the practical testing of the research model in real-life scenarios.*

Their collective impact has been invaluable.

Table of Contents

List of Figures	xv
List of Tables	xvii
List of Abbreviations and Acronyms	xviii
<i>Chapter One</i>	<i>1</i>
1.1) Introduction.....	2
1.2) Background of the Study.....	5
1.2.1) Building Code Compliance	9
1.2.2) Challenges in Building Code Compliance	10
1.2.3) Building Code Interpretation Issues	11
1.2.4) Challenges Faced by AEC Stakeholders (Architects and Engineers)....	12
1.3) Within the Context of Kingdom of Bahrain.....	15
1.4) Research Problem	22
1.4.1) Challenges in Developing AI-Based Compliance Tools.....	23
1.4.2) Proposed Framework for Q&A Model.....	24
1.4.3) Key Variables Related to the Research Problem	25
1.5) Research Questions.....	26
1.5.1) Primary Research Question	26
1.5.2) Secondary Research Questions.....	26
1.6) Research Objectives.....	27
1.7) Theoretical Foundations.....	30
1.8) Research Significance	30
1.9) Research Contributions.....	32
1.10) Scope of the Study	36
1.11) Ethical Consideration	37
1.12) Organization of the Thesis	37
<i>Chapter Two</i>	<i>39</i>
2.1) Introduction.....	40
2.2) Optimization of AEC Stakeholder Decision-Making for Building Regulation Compliance	41
2.2.1) Integration of Advanced Technologies and Decision Support Systems.	43
2.3) Compliance with Building Regulations.....	44
2.3.1) Regulatory Bodies and Their Role	45

2.3.2)	Common Violations and Their Implications	46
2.4)	Adopted Solutions to Reduce Violations.....	47
2.4.1)	Enhanced Training and Education Programs	47
2.4.2)	Engagement of Compliance Consultants	48
2.4.3)	Adoption of Advanced Monitoring Technologies	48
2.4.4)	Strengthening Regulatory Frameworks and Penalties.....	49
2.5)	Building Information Modelling (BIM)	49
2.5.1)	Definition and Importance of BIM.....	49
2.5.2)	Challenges of BIM Adoption in the AEC Industry	50
2.5.3)	Identifying Future Research Directions BIM Adoption.....	51
2.5.4)	Challenges and Recommendations for Bahrain.....	52
2.6)	Q&A Models for Compliance Checking and Decision Making in the AEC Industry	53
2.6.1)	The Rise of Q&A Models and Improved Information Access in AEC (2020-Present).....	54
2.6.2)	Enhancing Regulatory Compliance and Informed Decision-Making through Q&A Models	55
2.6.3)	Focus on Recent Research Related to Functionalities of Q&A Models for AEC applications (2018-2022)	56
2.7)	The Role of FAQs in Regulatory Domains and Building Regulations...	57
2.7.1)	Types of FAQs in the AEC Industry	57
2.7.2)	The Importance of Categorizing FAQs for Q&A Model Development.	62
2.8)	Advancing Knowledge Management and Expert Systems through Experiential Learning.....	63
2.8.1)	Stages of Knowledge Management.....	65
2.8.2)	Knowledge Based Systems	68
2.8.3)	Expert System.....	69
2.9)	Evolution of Q&A Models.....	71
2.10)	Foundations of Language Understanding	73
2.11)	Methodological Advances in ML for Q&A Models	74
2.12)	Advancements and Applications of Large Language Models (LLMs) ..	76
2.13)	A Comparative Analysis of Embeddings	78
2.13.1)	Comparative Studies of LLMs (ChatGPT and Gemini) in Practice.....	81
2.13.2)	Comparing between ChatGPT and Gemini (Bard) Performance.....	82

2.13.3)	Human-AI Feedback Loops in Data Generation	86
2.13.4)	Future Trajectory of ML-Based Q&A Models	86
2.14)	Finetuning LLMs for Q&A Models	87
2.14.1)	Fine-Tuning Techniques for Q&A Model Generation	88
2.15)	Conclusion	97
Chapter Three		99
3.1)	Introduction.....	100
3.2)	Research Design	102
3.3)	Qualitative and Quantitative Methodologies.....	105
3.4)	Exploration Stage.....	107
3.4.1)	Language Considerations and Data Format Conversion.....	107
3.4.2)	Data Augmentation and Labelling	108
3.4.3)	Early Trials and Algorithm Selection	108
3.4.4)	Collaborative Effort and Expert Input.....	108
3.4.5)	Outcomes of the Exploration Stage.....	109
3.5)	Research Development Timeline	109
3.5.1)	During 2022-2023	110
1)	Observation of Building Code Violations:.....	110
2)	Building Information Modelling Familiarity Studies.....	110
3)	Transition Towards Q&A Model Development.....	111
4)	Navigational Challenges.....	111
5)	Historical Study of Q&A Systems.....	111
6)	Evaluation of Supervised and Unsupervised Approaches	112
7)	Comparative Analysis of LLMs	112
3.5.2)	During January to May 2024	112
8)	Transition to Advanced LLMs	112
9)	Focus on Specific Chapter.....	112
10)	Valuable Data Collection.....	112
11)	Additional Data Sources.....	113
12)	Human-AI Collaboration for Q&A Sets.....	113
13)	Evaluation of Gemini and ChatGPT 4.....	113
14)	Fine-Tuning of Gemma-2b.....	113
15)	Use of LoRA Keras Framework	114
16)	Development of Hybrid Model	114
3.5.3)	During June 2024	114
17)	Release of ChatGPT 4o.....	114
18)	Development of Custom GPT	114
19)	Continued Development and Fine-Tuning	115
20)	Expert Feedback Incorporation	115
21)	Real-World Testing and Feedback Collection	115
22)	Integration of Updates.....	115
23)	Comprehensive Model Development.....	115

3.6)	Overview of Methodological Steps	116
3.6.1)	Phase (1) Steps: Fine-tuning Pre-trained Models (LLMs).....	117
1)	Compile Data.....	117
2)	Evaluate and Compare LLMs	118
3)	Set Selection Criteria	118
4)	Fine-tune the Selected LLMs.....	118
3.6.2)	Models Phase 2 Steps: Evaluation and Testing of Fine-Tuned	119
5)	Evaluation of Fine-tuned Models by Researcher.....	119
6)	Select and Refine Best Model.....	119
7)	Expert Evaluation:.....	119
8)	Incorporate Expert Knowledge:.....	120
3.6.3)	Phase 3 Steps: Results.....	120
9)	Test with Recruited Users	120
10)	Analyse Impact on Decision Making:	120
3.7)	Research Approach for Phase 1 Steps: Fine-Tuning Pre-trained Models (LLMs) 122	
3.7.1)	Step 1: Dataset.....	122
3.7.2)	Step (2): Evaluate the Performance of LLMs	131
3.7.3)	Step (3): Establishing Selection Criteria for Optimal LLMs in Fine-Tuning 136	
3.7.4)	Step (4): Fine-Tuning LLMs.....	140
3.7.5)	Step (5): Evaluation of Fine-Tuned LLMs	149
3.7.6)	Step (6): Refine Fine-tuned Model and Add More Data	151
3.7.7)	Step (7): Evaluation and Continuous Improvement of the Fine-Tuned Model through Independent Expert	153
3.7.8)	Step (8): Incorporation of Experts' Feedback.....	155
3.7.9)	Step (9): User Testing and Feedback	157
3.7.10)	Step (10): Analyzing the Impact on Decision-Making.....	160
3.8)	Methodological Framework for Expert and User Involvement in Model Evaluation and Testing.....	163
3.8.1)	Respondent Selection Criteria	163
3.8.2)	Structuring Interviews and Questionnaires	164
3.8.3)	Sample Question Selection for Testing	164
3.8.4)	Interview Question Validation.....	165
3.9)	Conclusion	165
	<i>Chapter Four</i>	167
4.1)	Introduction.....	168

4.2)	Evaluation of LLMs Performance: ChatGPT-4 and Gemini (Step 2 of the Methodology).....	169
4.2.1)	Assessment of Overall Performance.....	170
4.2.2)	Analysis by Question Type.....	172
4.2.3)	Consistency Across Clause Titles	176
4.2.4)	Consideration of Additional Information	177
4.2.5)	Identification of Areas for Improvement.....	177
4.2.6)	Overview of Evaluating LLMs Performance: ChatGPT-4 and Gemini 177	
4.3)	Evaluation of Fine-Tuned Models by Researcher Gemma-2b vs Custom GPT (Step 5 of the Methodology)	178
4.3.1)	Comparison of Fine-Tuned Gemma and Custom GPT	178
a)	Gemma.....	178
b)	CustomGPT.....	179
4.3.2)	Comparison of Custom GPT against Base Pre-Trained Model ChatGPT-4o Performance	181
a)	ChatGPT4o Performance Evaluation.....	181
4.3.3)	Overview of Evaluating (CustomGPT).....	182
4.4)	Evaluation of the Fine-Tuned Model by the Expert - Custom GPT (Step 7 of the Methodology)	184
4.4.1)	Key Themes from Experts Feedback	185
a)	Accuracy and Performance	185
b)	Relevance and Coverage	186
c)	Clarity and Consistency	186
d)	Usability and Interaction	187
e)	Technical Accuracy	187
f)	Suggestions for Future Development	188
g)	Gaps in Handling Specialized or Internal Information	189
h)	Timeliness and Real-Time Updates.....	189
4.4.2)	Overview of the Experts' Feedback	190
4.5)	Analysis of the Recruited Users' Feedback (Step 9 of the Methodology) 192	
4.5.1)	Key Themes from Recruited Users Feedback.....	193
a)	Navigating Regulatory Information.....	193
b)	Impact of the Custom GPT Model.....	194
c)	Usefulness for Decision-Making and Compliance	195
d)	Suggestions for Model Enhancements	195
e)	Role in Project Phases and Communication	197
f)	Collaboration and Knowledge Sharing	197
g)	Concerns and Observations	198
4.5.2)	Overview of the Users' Feedback	198

4.6)	Analysis of the Impact on Decision-Making of AEC Stakeholders (Step 10 of the Methodology)	202
<i>Chapter Five</i>		206
5.1)	Introduction.....	207
5.2)	Answering the Research Questions	208
5.2.1)	Primary Research Question:.....	209
5.2.2)	Secondary Research Questions:.....	209
5.3)	Achieving Research Objectives through Iterative Methodology.....	210
5.3.1)	Development of a Data-Driven Q&A Model for Building Code Compliance	211
5.3.2)	Feasibility of AI Models with Limited Data	211
5.3.3)	Codification of Expert Knowledge	212
5.3.4)	Establishing a Framework for AI-Based Regulatory Compliance.....	212
5.4)	Capturing Tacit Knowledge and Enabling Knowledge Continuity	218
5.5)	Strategic Implementation, Adaptation, and Governance Considerations	219
5.5.1)	Strategic Alignment and Implementation	219
5.5.2)	Empowering Adoption through Training for all Stakeholders	219
5.5.3)	Adaptability of the Framework	220
5.5.4)	Ethical and Legal Implications	221
5.5.5)	Trustworthy and Sustainable AI Use.....	221
5.6)	Research Challenges and Limitations.....	222
5.7)	Recommendations and Future Directions	226
References.....		231
Appendices.....		247
Appendix (A):		248
LLMs Interrogation and Response Grading.....		248
Appendix (B):		249
Samples of Questions Generated through Human-AI Collaboration and Answers Used as Training Data		249
Appendix (C):		250
Samples of Questions Generated through Human-AI Collaboration and Answers Provided by CustomGPT, for Expert Feedback.....		250
Appendix (D):		251

Interview Questions for Experts.....	251
Appendix (E):	252
Meeting Questions for Users	252

List of Figures

Figure (1.1): Key stakeholders in a typical construction project in the Kingdom of Bahrain (Source: Student's work)	21
Figure (1.2): The impact of decisions made by different stakeholders involved along the life cycle of a typical construction project in the Kingdom of Bahrain (Source: Student's work)	22
Figure (2.1): Processes of knowledge management (adopted from (Marques et al., 2011))	66
Figure (2.2): Retrieval augmented generation with a knowledge graph (Adopted from (Bukowski, 2023))	95
Figure (3.1): Research development timeline	116
Figure (3.2): Methodological steps applied in the research (Source: Student's work)	121
Figure (3.3): Example of an instruction chat with ChatGPT-4 to generate a set of Q&A related to article (2) - (Source: Student's work)	126
Figure (3.4): Example of an instruction chat with Gemini to generate a set of Q&A related to article (2) – (Source: Student's work)	127
Figure (3.5): Example of instruction chat with ChatGPT-4 to generate an enhanced answer related to article (2) – (Source: Student's work)	129
Figure (3.6): Example of instruction chat Gemini to generate an enhanced answer related to article (2) – (Source: Student's work)	130
Figure (3.7): Evaluation of ChatGPT-4 and Geminin answers (Source: Student's work)	133
Figure (3.8): The interface for the developed CustomGPT (Source: Student's work)	145
Figure (4.1): Performance metrics analysis (Source: Student's work)	172
Figure (4.2): ChatGPT-4 vs. Gemini in answering 'Factual' questions (Source: Student's work)	173
Figure (4.3): ChatGPT-4 vs. Gemini in answering "Critical Thinking" questions (Source: Student's work)	173
Figure (4.4): ChatGPT-4 vs. Gemini in answering "Hypothetical questions" (Source: Student's work)	174
Figure (4.5): ChatGPT-4 vs. Gemini in answering "Open-ended" questions (Source: Student's work)	175

Figure (4.6): ChatGPT-4 vs. Gemini in answering “Comparative” questions (Source: Student’s work)	176
Figure (4.7): The start of ChatGPT4o evaluation process (Source: Student’s work)	182
Figure (4.8): Comparison of CustomGPT against base LLM performance (Source: Student’s work)	182
Figure (4.9): Experts’ evaluations of Q&A model performance across key metrics (Source: Student’s work)	191
Figure (4.10): Experts' recommendations for enhancing Q&A model capabilities (Source: Student’s work)	192
Figure (4.11): Users feedback themes (Source: Student’s work)	201
Figure (4.12): Suggested enhancements for CustomGPT by users (Source: Student’s work)	201
Figure (4.13): Key impacts of the CustomGPT model on AEC stakeholders decision-making (Source: Student’s work)	204
Figure (5.1): Key components of compiled data for model adaptability (Source: Student’s work)	215
Figure (5.2): Integrated framework for data-driven compliance model with iterative feedback mechanism (Source: Student’s work)	217

List of Tables

Table (1.1): Contributions of construction and real estate sectors to Bahrain's GDP in 2024	6
Table (2.1): Comparisons between BERT, SBERT and OpenAI embeddings	80
Table (3.1): Dimensions of research design	104
Table (3.2): Characteristics of the dataset for Q&A generation	124
Table (3.3): Comparison matrix between Custom GPT and Gemma (Source: Student's work)	138
Table (4.1): Results table - number of correct, partially correct, and incorrect answers provided by ChatGPT-4 and Gemini (Source: Student's work)	170
Table (4.2): Results table - assessment of overall performance of ChatGPT-4 and Gemini (Source: Student's work)	170
Table (4.3): ChatGPT-4 vs. Gemini performance per question (Source: Student's work)	176

List of Abbreviations and Acronyms

The definitions of the following words and phrases apply unless the context necessitate a different interpretation:

No.	Abbreviation	Description / Full Name
1	5Vs	Volume, Velocity, Variety, Veracity, and Value
2	ACC	Automated Compliance Checking
3	AEC	Architecture, Construction, and Engineering
4	AI	Artificial Intelligence
5	BERT	Bidirectional Encoder Representations from Transformers
6	BIM	Building Information Modelling
7	Building Code	Unified Guidebook of Building Permit Regulations
8	CAD	Computer-aided Design
9	CBR	Case-based Reasoning
10	CRPEP	The Council for Regulation the Practice of Engineering Professions
11	Custom GPT	Custom Generative Pre-Trained Transformer.
12	DoI	Diffusion of innovation
13	DSS	Decision Support Systems
14	EPC	Engineering, Procurement, and Construction
15	FAQs	Frequently Asked Questions
16	GCC	Gulf Cooperation Council
17	GPUs	Graphics Processing Units
18	HITL	Human-in-the-Loop
19	HSE	Health and Safety Executive
20	IFC	Industry Foundation Classes
21	IFT	Instruction Fine-Tuning
22	IoT	Internet of Things
23	KBS	Knowledge-based Systems
24	KM	Knowledge Management
25	Law No. (51) of 2014	Engineering Law
26	LLMs	Large Language Models
27	LoRA	Low-Rank Adaptation

No.	Abbreviation	Description / Full Name
28	ML	Machine Learning
29	NLP	Natural Language Processing
30	NLU	Natural Language Understanding
31	PEFT	Parameter-Efficient Fine-Tuning
32	Q&A	Questions and Answers
33	RAFT	Retrieval-Augmented Fine-Tuning (RAFT)
34	RAG	Retrieval-Augmented Generation
35	RetriKT	Retrieval-based Knowledge Transfer
36	SBERT	Sentence-BERT
37	SEFT	Self-evolution Fine-tuning
38	SFT	Supervised Fine-Tuning
39	VDC	Virtual Design and Construction

Chapter One

Introduction

1.1) Introduction

The Architecture, Engineering and Construction (AEC) industry is a multiple stakeholders' and multifaceted industry that encompasses a wide range of activities, from design and engineering to construction and maintenance. It involves, effective decision-making for essential project successes. Specifically, The AEC industry includes architecture, engineering, and construction activities, focusing on the design, construction, and maintenance of buildings and infrastructure (Rafsanjani & Nabizadeh, 2023). Within this sector, key stakeholders include owners who define project requirements and provide funding, architects and engineers who design and ensure the technical soundness of projects, contractors and subcontractors who execute construction tasks and manage on-site activities, and suppliers. Each stakeholder plays a specific role, such as architects designing structures, engineers ensuring technical feasibility, and contractors managing construction processes (Primožič & Kutnar, 2024; Yun-xi, 2008). To enhance project performance and reduce uncertainties, collaboration among stakeholders is essential for effective decision-making in the AEC industry (Yun-xi, 2008).

However, the AEC industry faces challenges such as ethical concerns, safety issues, and environmental impacts. Misrepresentations and lack of compliance with regulations can exacerbate these issues (Moodley et al., 2008). For example, non-compliance with regulations can lead to legal disputes, project delays, and increased costs. It can also damage reputations and lead to safety hazards (Moodley et al., 2008; Rubin, 2010). Furthermore, the integration of diverse perspectives and responsibilities among stakeholders makes the decision-making process both challenging and critical.

These challenges can be addressed through data analytics which plays a vital role in optimizing decision-making in multi-stakeholder projects. Data analytics can help identify and rank causes of claims, such as delays and contractual ambiguities, which are common in multi-stakeholder projects. By analysing these factors, stakeholders can reduce conflicts and ensure successful project completion (Ansari et al., 2022). By leveraging data analytics, stakeholders can gain insights into complex data sets, uncover hidden patterns, and identify trends and outliers. Moreover, Artificial Intelligence (AI) and big data technologies offer opportunities for improving efficiency, safety, and security in construction projects. These technologies can be applied to activity monitoring, risk management, and resource optimization, although challenges remain in their integration (Munawar et al., 2022).

In addition, organized datasets in the construction industry enhance data management and decision-making. Since big data requires innovative and cost-effective processing methods, its potential in construction is vast, enabling automation and removing project hurdles. Notably, big data and Machine Learning (ML) can help reduce challenges, improve project planning, and accelerate construction processes. As a result, automation in construction will require big data, deep learning, and ML tools. Additionally, technologies such as Building Information Modeling (BIM) and cloud-powered simulations can help minimize project waste and improve construction quality (Munawar et al., 2022).

Historically, during the 1980s, the AEC field witnessed the development of BIM which has become increasingly popular in recent decades, offering a comprehensive approach to generating and managing information for

constructed assets within the built environment (Autodesk, 2023). This method utilizes an intelligent model operating on a cloud-based platform. BIM incorporates organized and diverse data to produce a virtual representation of an asset throughout its entire lifespan, encompassing the planning and design phases, construction, and ongoing maintenance operations (Autodesk, 2023). While utilizing the native format generated by BIM software remains a viable option, a growing inclination exists among developers and users to embrace open standard formats. Within the context of BIM, the Industry Foundation Classes (IFC) standard has gained significant traction (Al Nama & Alalawi, 2023).

The IFC is a Computer-aided Design (CAD) data exchange standard specifically designed to represent data for the AEC industry. It functions as a conceptual model schema in a neutral file format where the data is stored in a format that is not tied to any specific software or vendor enabling seamless exchange of building-related information. This approach is widely regarded as more favourable and pragmatic (Al Nama & Alalawi, 2023).

The AEC industry is crucial in shaping our infrastructure and built environment. However, it faces significant challenges in compliance, decision-making, and technology adoption. These challenges highlight the need for innovative solutions based on data analytics and BIM to optimize stakeholder decision-making, address inefficiencies, improve collaboration, and ensure compliance.

The following section of this chapter delves into the background of the study, exploring the current status within the context of the Kingdom of Bahrain. It identifies the research problem, formulates research questions, and

outlines the research objectives. The theoretical foundations are discussed, along with the significance and contributions of the research. Additionally, the scope of the study is defined, research challenges and limitations are addressed, ethical considerations are considered, and an organization of the thesis is provided. This comprehensive approach ensures a thorough understanding of the research context and its implications.

1.2) Background of the Study

The AEC industry plays a crucial role in driving economic growth and societal development, both globally and locally. In 2023, the global AEC market was valued at \$10.05 billion, and it is projected to grow at a compound annual growth rate of 10.3%, reaching \$24.36 billion by 2032 (Benchmark International, 2024). Notably, the non-residential construction sector alone experienced nearly 20% growth in 2023, a rate that has not been observed since the pre-Great Recession boom (Benchmark International, 2024). Projections indicate that the global construction industry's market value will reach \$13.9 trillion by 2037, with significant contributions from China, the United States, and India, accounting for 57% of this growth by 2030. This growth is largely driven by the increasing demand for sustainable homes, comprehensive master planning, pedestrian-friendly communities, disaster-ready infrastructure, and flexible designs (Benchmark International, 2024).

The construction and real estate sectors are integral parts of the AEC industry, contributing significantly to Bahrain's GDP. Table below highlights their contributions throughout 2024:

Table (1.1): Contributions of construction and real estate sectors to Bahrain's GDP in 2024

Quarter	Construction Sector Contribution to GDP	Real Estate Sector Contribution to GDP	Total Contribution to GDP
Q1 2024 – (Ministry of Finance and National Economy, 2024a)	6.5%	4.0%	10.5%
Q2 2024 – (Ministry of Finance and National Economy, 2024b)	6.4%	3.8%	10.2%
Q3 2024 – (Ministry of Finance and National Economy, 2024b)	6.5%	4.1%	10.6%

Additionally, in 2023, the construction sector contributed 7.1% to GDP, and the real estate sector contributed 5.4%, totalling 12.5% (Ministry of Finance and National Economy, 2023).

According to the “Bahrain Economic Quarterly Q3 2024” report, the non-oil sector is anticipated to grow by 4.4% in 2025, driven by various non-oil activities, including construction and real estate. This projection underscores the importance of the construction and real estate sectors in driving economic growth, creating jobs, and supporting sustainable development in Bahrain (Ministry of Finance and National Economy, 2024b).

However, the AEC industry faces a multitude of challenges that impact its efficiency, productivity, and sustainability. The AEC industry is often characterized by its complexity and fragmentation, involving multiple stakeholders, including architects, engineers, contractors, and clients. This fragmentation can lead to significant communication barriers and coordination issues, which are major sources of inefficiency (Z. Wang et al., 2020). The traditional project management practices in the AEC sector are frequently outdated and unable to cope with the dynamic nature of modern construction projects (Azzouz & Papadonikolaki, 2020). These inefficiencies are further aggravated by the industry's reliance on manual processes and the lack of integration of digital technologies (Z. Wang et al., 2020).

Many recent studies have explored solutions to these challenges, particularly through the application of AI and data analytics. For example, the need for improved resource management, enhanced productivity, and better project coordination in the AEC industry led to the development of Virtual Design and Construction (VDC) which addresses these challenges by integrating BIM and Internet of Things (IoT) technologies to provide real-time data and visualization, enabling more efficient project planning and execution. Analytics play a crucial role in optimizing decisions by leveraging the spatiotemporal data generated by VDC to predict project performance, identify potential issues, and enhance overall project (Rafsanjani & Nabizadeh, 2023).

Another study by Adeloye (2023) reviewed the application of AI in the construction industry, highlighting that the sector faces significant challenges, including cost and time overruns, health and safety issues, and labour shortages, which have impeded its growth and productivity. The study emphasized that AI technologies, such as machine learning, robotics, and computer vision, are being adopted to tackle these challenges by enhancing automation, improving safety, and optimizing resource management. Furthermore, analytics play a crucial role in optimizing decisions by leveraging data from AI applications to predict project performance, identify potential risks, and enhance overall project outcomes (Adeloye, 2023).

The study by Yun-xi (2008) highlights challenges such as low efficiency, budget overruns, and frequent security incidents due to a lack of cooperation and effective decision-making among stakeholders. In particular, entropy theory quantifies uncertainty and disorder in stakeholder decisions, enabling a more structured approach (Yun-xi, 2008). These challenges can be

addressed through the adoption of AI technologies to enhance collaboration and decision-making processes. By integrating AI technologies like machine learning (ML) and data analytics with entropy theory, complex data can be analysed more effectively. This combination provides valuable insights that help stakeholders make informed decisions and ultimately improve project outcomes

One of the significant challenges in the AEC industry is the integration of advanced technologies such as BIM. BIM has the potential to revolutionize the industry by enhancing collaboration, improving project outcomes, and reducing costs. However, the adoption of BIM is hindered by several factors, including high implementation costs, the need for extensive training, and resistance to change within organizations (Z. Wang et al., 2020). Furthermore, the industry's digital transformation is often slow and fragmented, with many firms struggling to implement digital tools effectively (Bhattacharya & Momaya, 2021). While the integration of these technologies holds great potential, challenges such as data fragmentation, noise, and occlusions must be resolved. A unified platform is essential for effective construction management (Z. Wang et al., 2020).

By examining the challenges faced by the AEC industry in the Kingdom of Bahrain, particularly those related to optimizing stakeholder decision-making to ensure compliance with a selected chapter -concerning with the regulation of engineering practice- of the Unified Guidebook of Building Permit Regulations (Building Code), this research aims to bridge the gap between the potential of advanced technologies and the barriers to their effective implementation. It proposes a data-driven Q&A model designed to

overcome these obstacles by enabling real-time, accurate decision-making and reducing reliance on manual processes and expert consultations.

1.2.1) Building Code Compliance

Building code compliance is fundamental to the AEC industry, ensuring that buildings are safe, healthy, and accessible. These codes establish standards for the design, construction, and maintenance of buildings, addressing various aspects such as structural integrity, fire safety, energy efficiency, and accessibility. Compliance with building codes is essential for protecting the public and ensuring that buildings meet minimum safety standards.

Building codes are designed to safeguard the health, safety, and welfare of building occupants and the public. Compliance with these codes helps mitigate risks associated with structural failures, fire hazards, and other safety concerns. For instance, fire safety codes mandate the installation of fire alarms, sprinkler systems, and proper egress routes, significantly reducing the risk of fire-related incidents. Adherence to building codes ensures that buildings can withstand natural disasters such as earthquakes, hurricanes, and floods. This is particularly important in regions prone to such events, where non-compliance could lead to catastrophic consequences (Z. Wang et al., 2020).

Building codes also play a critical role in promoting energy efficiency and sustainability. Energy codes set standards for insulation, windows, heating, and cooling systems, helping to reduce energy consumption and greenhouse gas emissions. Compliance with building codes is not only a matter of safety but also has legal and financial implications. Non-compliance can result in

fines, legal liabilities, and increased insurance premiums. In severe cases, buildings that do not meet code requirements may be condemned, leading to significant financial losses for owners and developers (Nwadike & Wilkinson, 2022).

1.2.2) Challenges in Building Code Compliance

Despite the importance of building code compliance, stakeholders in the AEC industry, particularly architects and engineers, face numerous challenges in achieving and maintaining compliance. One of the primary challenges is the complexity and frequent updates of building codes. Building codes are often extensive and detailed, covering a wide range of technical and safety requirements. Keeping up to date with the latest codes and amendments can be challenging, especially for smaller firms with limited resources (Z. Zhang et al., 2023a).

Many researchers highlight the difficulty architects and engineers face in interpreting and applying building codes (Yun-xi, 2008). The technical language and detailed specifications in building codes can be challenging to understand, leading to potential misinterpretations and errors in design and construction. Different interpretations of regulation articles can also lead to disputes and inconsistencies in code enforcement, causing project delays, increased costs, and legal issues (Malsane et al., 2015). Moreover, the enforcement of building codes can be inconsistent across different regions and jurisdictions. Variations in code enforcement practices can lead to uncertainty and confusion among stakeholders. In some cases, local building officials may have differing interpretations of code requirements, leading to inconsistent application and enforcement of building regulations.

1.2.3) Building Code Interpretation Issues

The interpretation and application of building codes are crucial for ensuring compliance and safety in construction projects. However, the expertise required to comprehensively understand and interpret these codes, particularly complex or specific chapters, is often concentrated among a limited number of professionals. This scarcity of experts presents significant challenges to the AEC industry, necessitating strategies to preserve and disseminate their knowledge.

Interpreting regulatory clauses presents challenges due to implicit properties and complex computational logic. The LLM-FuncMapper approach, proposed by Zheng et al. (2023) uses Large Language Models (LLMs) to automate the identification and interpretation of regulatory functions. This method aims to enhance the understanding and application of building codes, potentially reducing the reliance on a limited number of experts (Zheng et al., 2023).

A study by Terenchuk et al. (2021) presents a methodological approach to creating knowledge-based systems for building-technical expertise. These systems use fuzzy logic and neural networks to capture and apply expert knowledge, ensuring that critical expertise is preserved and made accessible for future projects. This approach helps mitigate the risks associated with the limited availability of experts by providing a structured method for knowledge dissemination (Terenchuk et al., 2021).

1.2.4) Challenges Faced by AEC Stakeholders (Architects and Engineers)

Architects and engineers are pivotal in the AEC industry, tasked with designing structures that meet various functional, aesthetic, and regulatory requirements. However, these professionals encounter numerous challenges in ensuring compliance with building codes, which can complicate the design and construction process, some of these challenges are discussed below:

a) Complexity and Frequent Updates of Building Codes

One of the primary challenges faced by architects and engineers is the complexity of building codes. These codes are comprehensive documents that cover numerous aspects of construction, including structural integrity, fire safety, electrical systems, plumbing, and energy efficiency. The technical language and detailed specifications in these codes can be daunting, especially for professionals who may not have specialized training in all areas of code compliance.

The frequent updates and revisions to building codes add another layer of complexity. Keeping up to date with the latest changes requires continuous education and training, which can be time-consuming and costly for professionals. Smaller firms may struggle to allocate the necessary resources for ongoing education (Z. Zhang et al., 2023).

b) Interpretations of Regulation Articles

One significant challenge is the potential for different interpretations of building code articles. Building codes can sometimes be ambiguous, leading to varied interpretations by different professionals and even by different local building authorities. This ambiguity can result in inconsistencies in how codes

are applied and enforced, creating uncertainty and potential conflicts during the design and construction process (Malsane et al., 2015).

These differing interpretations can lead to disputes between architects, engineers, and code enforcement officials. Such disputes can cause delays in project approval and increase costs due to the need for redesigns or additional compliance measures. The lack of a standardized interpretation can also lead to projects being stalled or facing legal challenges, which can be particularly detrimental for large-scale developments.

c) Legal and Financial Implications

Non-compliance with building codes and regulations in the AEC projects can lead to severe legal and financial repercussions. Legal consequences include hefty fines, lawsuits, and mandatory project modifications, which can significantly delay project timelines and escalate costs. According to Guo et al. (2021), Automated Compliance Checking (ACC) has become crucial in mitigating these risks by ensuring that building designs comply with relevant laws and regulations (D. Guo et al., 2021).

The costs associated with rectifying non-compliant work, including redesign, rework, and additional inspections, can be substantial. As noted by (Varavenko et al., 2022), contractors often bear the financial burden of these risks, which can lead to increased project costs and reduced profitability.

d) Transforming Information Access

Questions and Answers (Q&A) models offer significant potential to enhance decision-making in the AEC industry by leveraging technologies like Natural Language Processing (NLP) and AI. These models expedite access to crucial information from BIM data, aiding in pinpointing feasible design

options within multidisciplinary design. Additionally, they automate data retrieval processes from BIM models and facilitate the assessment and comparison of design alternatives, enabling well-informed decision-making during the initial design stage. NLP techniques, such as semantic analysis and latent semantic analysis, empower these models to provide accurate and rapid responses to user queries. This significantly improves decision-making processes within the AEC industry.

The groundwork for Q&A systems in AEC was laid in the late 1980s and early 1990s. Pioneering research explored the potential of expert systems to facilitate knowledge sharing and enhance decision-making through interactive environments (Finn, 1989). Additionally, studies investigated automated question-answering methods to improve performance analysis and project management, highlighting the critical link between information accessibility and informed decision-making (Robert et al., 2006).

The 2000s witnessed a surge in research focused on harnessing NLP techniques for AEC-specific Q&A systems. Several studies introduced groundbreaking systems that allowed users to pose questions in natural language formats, significantly improving information retrieval compared to traditional web searches (Kovacevic et al., 2008). Advancements in NLP techniques like keyword extraction, question type classification, and answer retrieval strategies further enhanced the accuracy and user experience of these systems (Haitao et al., 2014).

The period from 2005 to 2020 saw the exploration of integrating Q&A systems with specific data sources within AEC. Research delved into leveraging BIM data for automated information retrieval, creating automated

question-answering system based on semi-supervised learning and Text-to-SQL model enhances information retrieval efficiency, reducing user effort in obtaining data, applicable in the AEC industry (Yuxia & Ruonan, 2021), and employing chatbots to utilize construction specifications for improved quality and employing chatbots to utilize construction specifications for improved quality (Yuxia & Ruonan, 2021).

The current era from 2020 to present is marked by the integration of AI paradigms into Q&A systems specifically designed for AEC applications. Studies propose frameworks utilizing advanced AI algorithms like NLP, Support Vector Machines (SVM), and BERT neural networks to automate information extraction from BIM models and construction documents (Nabavi et al., 2023; N. Wang et al., 2021; Wenwu, 2018). These advancements not only enhance information retrieval efficiency but also empower more informed decision-making by providing faster access to critical project data.

Looking towards the future, researchers envision a future where Q&A systems are seamlessly integrated with other emerging technologies. This convergence is expected to reshape the AEC landscape by equipping professionals with enhanced decision-making capabilities throughout the project lifecycle.

1.3) Within the Context of Kingdom of Bahrain

In the Kingdom of Bahrain, the building permits issued after checking the engineering design and completed structures compliance to local building

regulations. The building permits can be obtained from an electronic system called “Benayat”, the Arabic word for buildings. Benayat portal is a geographical information system (GIS) that is based on web application and integrated with permits system. The systems allow the users in the Kingdom of Bahrain, from licensed engineering offices, to apply for building permits, consult with concerned government entities, check status of their applications, pay building permit fees, and apply for other building-related services. Benyat enables users to access essential information for land parcels and consolidate all relevant regulations and requirements within a single system. By utilizing an interactive map, users can obtain the necessary data to prepare drawings and documents required for applying for a building permit (Benayat, 2025). Note that, most architectural and engineering firms prepare the design drawings in 2D format, and up to date there is neither a legislation that enforces the use BIM, nor a requirement to upload the drawings in 3D format in Benayat system. Thus, the use of BIM is subject to the decision of the architectural and engineering firms of which most of them are relying on the use of CAD/ 2D environment.

Issuing building permits, after checking the project’s compliance, through the Benayat system began in 2018, and before to this time, building permits were issued in the traditional manner, indicating that the usage of an electronic system is something new that began no more than five years ago. During this time, local newspapers informed that, notwithstanding the implementation of the system, building violations continued at a high rate. Furthermore, the Bahrain government program for 2023-2026 focuses on achieving four priorities that lead to several goals that the Bahraini government seeks to implement towards better, development, and prosperity for the country and

citizens. The plan aims to adopt programs and policies that aim to raise citizens' income and improve their standard of living while preserving their livelihoods. Sustainable development, and Quality and competitive government services are two of the four priorities included in the government plan (Government of Bahrain, 2025).

Government performance and digital transformation axis is one of the six axes included in the same plan, this axis concerns with of six items of which supporting efforts aimed at digitizing government services and transactions and gradually making them comprehensive electronically. Continue to develop systems and adopt modern technologies and AI that contribute to employing digital development in improving the quality and efficiency of government services and enhance the performance of government agencies in government digital systems (such as records, buildings, and communication systems) - (Government of Bahrain, 2025).

In 2024 the Kingdom of Bahrain has employed AI to detect changes and construction violations to Bahraini building code using satellite imagery. This system utilizes AI techniques for the early identification of violations related to construction quality and increased electrical loads. It generates clear periodic reports on construction changes throughout the Kingdom, aiming to reduce violations and prevent potential damages, thus enhancing the development of the Kingdom's construction sector (Information & eGovernment Authority, 2024). Furthermore, the article dated 19 November 2024 on (The Daily Tribune, 2024) reports that Bahrain is making it easier to secure building permits with the Ministry of Municipalities Affairs and Agriculture rolling out measures. The aim is to reduce rule-breaking, speed up approvals, and assist over 12,000 applicants recorded in that year. From

January to September 2024, the Capital Municipality and the municipalities of Muharraq, Northern, and Southern regions issued 12,205 permits for projects ranging from new builds and extensions to demolitions, fencing, and excavation. At the same time, 2,375 violations were identified which equal to 19.5% of the issued permits. The ministry is working with other government bodies to update construction rules and approvals, using aerial imagery to detect breaches early and ensure greater transparency.

This innovative approach reflects a broader governmental direction in Bahrain toward leveraging advanced technologies for improved regulatory practices. This research on creating a Q&A model based on a LLM for building regulations aligns with Bahrain's governmental directions. By providing instant answers to stakeholders' queries, the model can help reduce violations and enhance compliance, supporting the Kingdom's goal of leveraging advanced technologies and AI to improve regulatory practices and overall project outcomes.

Also, through searching in the local newspapers, it becomes clear that there are large percentages of building violations, with efforts by the competent authorities in the Kingdom to reduce these violations, examples of which are as follows:

- The article dated 17 January 2021 on [alayam.com], (BNA, 2021) reports that residential building permit violations in Bahrain are affecting privacy and the country's cultural heritage. The article notes that many residential buildings in Bahrain are being constructed without proper permits, which can lead to violations of privacy and other issues. The article also highlights the importance of preserving

Bahrain's cultural heritage by ensuring that new construction projects are in line with local architectural styles and traditions.

- The article dated 18 December 2021 on [alayam.com] by (Parliamentary Affairs Editor, 2021) reports that the Ministry of Works, Municipalities Affairs and Urban Planning has announced that it encourages violators to correct their violations.
- The article dated 21 January 2022 on [alayam.com] by (Sara Najeeb, 2022) reports that the Ministry of Works, Municipalities Affairs and Urban Planning has reduced building permit violations in Bahrain by increasing fines and penalties for violators. The Ministry has issued more than 1000 fines for building permit violations over the past three years. The article also notes that the Ministry has taken several measures to reduce building permit violations, including increasing awareness campaigns and inspections, and improving coordination between relevant authorities.
- The article dated 13 February 2022 on [alayam.com] by (Abrar Shukri, 2022) reports that the Public Prosecution has referred 89 out of 238 building permit violations in Hamad Town to the Public Prosecution for legal action. The violations were discovered during an inspection campaign carried out by the Ministry of Works, Municipalities Affairs and Urban Planning in cooperation with the Ministry of Interior and other relevant authorities. The campaign aimed to ensure compliance with building regulations and safety standards.

There is few research done in the area of building permits through Benayat, which include (Yasser & Abdulkarim, n.d.), this investigates

stakeholder satisfaction regarding the operational effectiveness of the transformational policy and the utilization of the Benayat system, an automated and re-engineered system that transitions the role of the Bahrain government from service provider to regulator. Aiming to gather feedback from diverse stakeholders in Bahrain, the research utilized an online survey. The results revealed that awareness of the transformational policy did not correlate with the operational effectiveness or expected benefits. However, a significant association was found between the technology used for implementing the policy (the Benayat system) and the anticipated benefits, as well as the operational effectiveness of the building permit process. Additionally, a strong correlation was observed between the expected benefits of the policy, operational effectiveness, and overall satisfaction. Based on these findings, it is recommended to improve the promotion of the Benayat system to enhance stakeholders' understanding and provide professional training for its effective utilization.

In various regions including the Kingdom of Bahrain, construction projects involve various stakeholders, including government authorities, contractors, consultants, and clients. Effective stakeholder management is crucial for the success of these projects, as it directly impacts cost, time, and quality. The Ministry of Works in the Kingdom of Bahrain emphasizes the importance of clear communication and collaboration among stakeholders to ensure project success (Ministry of Works, 2010; Mohammed, 2024). Additionally, the use of advanced technologies in the Kingdom, such as AI for early detection of construction violations, further supports the efficient management of construction projects (Information & eGovernment Authority, 2024).

Figures (1.1) and (1.2) provide a comprehensive overview of the key stakeholders involved in a typical construction project within the Kingdom of Bahrain and illustrate how their decisions impact the project's lifecycle. These figures highlight the interconnection of stakeholder roles and the cascading effects of their decisions on subsequent phases of the project.

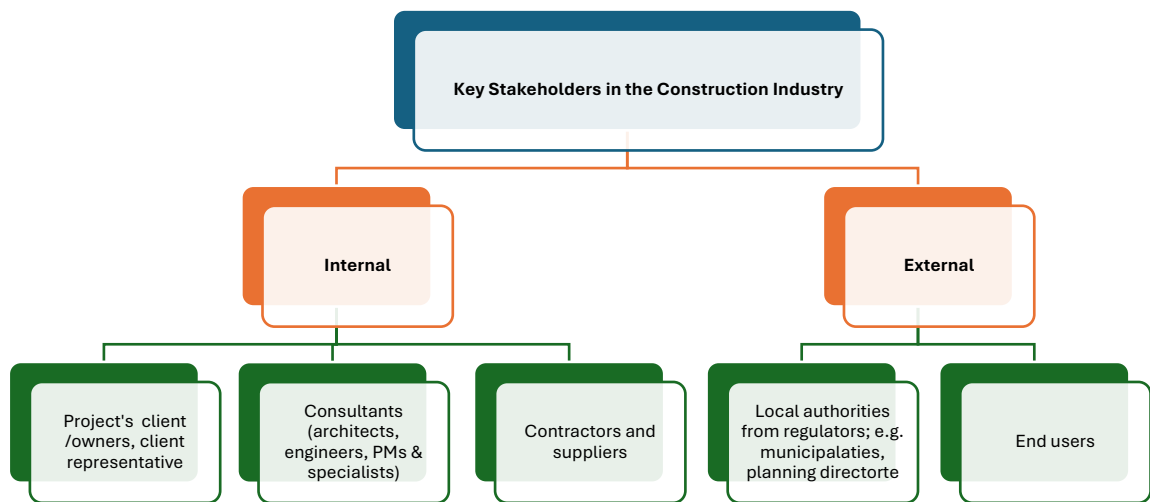


Figure (1.1): Key stakeholders in a typical construction project in the Kingdom of Bahrain (Source: Student's work)

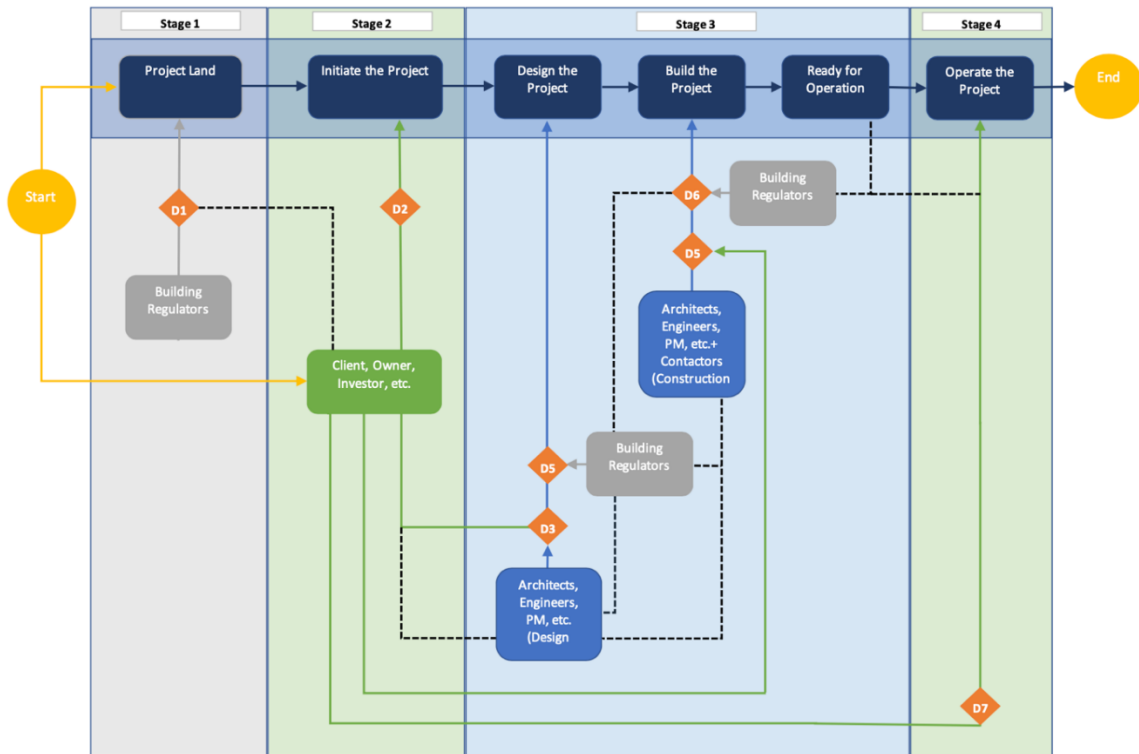


Figure (1.2): The impact of decisions made by different stakeholders involved along the life cycle of a typical construction project in the Kingdom of Bahrain (Source: Student's work)

1.4) Research Problem

As discussed in the previous sections the AEC industry is foundational to modern infrastructure, significantly influencing economic development and societal well-being. However, the complexity of building regulations presents substantial challenges to stakeholders such as architects and engineers. Ensuring compliance with these intricate regulations is crucial for the safety, sustainability, and legality of construction projects, yet the process is fraught with difficulties that impede timely and informed decision-making.

Stakeholders in the AEC industry in the Kingdom of Bahrain often struggle to navigate the complex web of building codes necessary for

regulatory compliance. These regulations are detailed and require precise interpretation to ensure that construction projects adhere to the required standards. A key challenge is the unavailability of critical information at the moment when design decisions need to be made. Architects and engineers frequently face delays because they must consult with experts who specialize in specific chapters of the Bahraini building code, such as Chapter Fourteen. This dependency not only slows the decision-making process but also increases the risk of non-compliance if the necessary expertise is not available or accessible in a timely manner.

Moreover, the limited number of experts with in-depth knowledge of these regulations exacerbates the problem. In many cases, stakeholders are forced to make decisions without the benefit of specialized guidance, leading to a higher likelihood of building code violations. These violations can result in project delays, increased costs, and legal disputes, all of which undermine the success and efficiency of construction projects.

1.4.1) Challenges in Developing AI-Based Compliance Tools

The second significant challenge pertains to the development of AI-based tools designed to assist in regulatory compliance. Traditional AI models typically require large datasets to be trained effectively. However, in the context of building regulations, such datasets are either not available or difficult to access, limiting the effectiveness of these models. The AEC industry, therefore, faces a critical barrier in developing AI tools that can accurately support decision-making processes.

Furthermore, the limited availability of experts familiar with specific chapters of the building code means that their knowledge is not sufficiently captured or preserved. This scarcity of expert knowledge presents a significant challenge for training AI models to provide accurate and contextually relevant guidance. As these experts retire or move on, the knowledge gap widens, making it increasingly difficult for stakeholders to access the information they need to comply with complex regulations.

The core research problem addressed in this thesis revolves around the need to optimize decision-making processes in the AEC industry in the Kingdom of Bahrain through the development of a robust Q&A model. This model must be capable of functioning effectively within the constraints of limited data availability and scarce expert knowledge. The challenge is twofold: how to ensure that stakeholders have access to the information they need when they need it, and how to develop an AI-based model that can operate within the data and expertise limitations inherent in the AEC industry.

1.4.2) Proposed Framework for Q&A Model

In response to these challenges, this research proposes a novel framework for developing a Q&A model that leverages advancements in AI technology while accounting for the specific constraints of the AEC industry. This framework aims to integrate the latest AI techniques, such as GPT models, to process and generate accurate, real-time responses to regulatory inquiries, even when large datasets are not available.

The framework is designed to be adaptive, incorporating human-AI collaboration to enhance the model's performance. By capturing and

preserving the specialized knowledge of the few available experts, the model ensures that this critical information remains accessible to all stakeholders. This approach not only mitigates the risks associated with expert scarcity but also provides a scalable solution that can be applied across various chapters of the Bahraini building code, with a particular focus on Chapter Fourteen.

1.4.3) Key Variables Related to the Research Problem

To fully address the research problem, it is essential to consider several key variables:

- **Complexity of Bahraini Building Codes:** The intricate and detailed nature of regulations, particularly in Chapter Fourteen, which stakeholders must navigate to ensure compliance.
- **Data Availability:** The limited scope and size of datasets available for training AI models in the context of building regulations.
- **Expert Knowledge:** The scarcity of experts who possess deep, specialized knowledge of specific chapters of the Bahraini building code, and the need to preserve and make this knowledge accessible.
- **Decision-Making Processes:** The methods by which stakeholders make regulatory decisions, which need to be optimized to handle the complexities of the AEC industry.
- **AI Model Development:** The challenge of function effectively with limited data and incorporate expert knowledge into its decision-making algorithms.

The research problem at the heart of this thesis involves addressing the dual challenges of optimizing decision-making in the AEC industry and developing an AI-based Q&A model that can function effectively given the constraints of limited data and scarce expert knowledge. By proposing a framework that leverages AI advancements and integrates expert knowledge, this research aims to create a tool that optimizes decision-making processes, ensures regulatory compliance, and ultimately leads to safer, more efficient construction practices.

1.5) Research Questions

To address the identified research problem, this study formulates specific research questions that guide the investigation. These questions are designed to explore the development and effectiveness of the AI-based Q&A model in optimizing decision-making processes within the AEC industry, particularly concerning regulatory compliance under the constraints of limited data and expert availability.

1.5.1) Primary Research Question

1. How can an AI-based Q&A model be developed to optimize decision-making processes for AEC stakeholders by providing instant, accurate answers to regulatory questions during the design stage?

1.5.2) Secondary Research Questions

1. What are the key factors contributing to the selection of a base AI model to create a customized Q&A model?

2. What methods can be employed to develop a robust AI model that operates effectively with limited data and integrates expert knowledge?
3. How can the tacit knowledge of Bahraini building code experts be codified into a structured, accessible system for use by stakeholders?
4. How effective is the AI-based Q&A model in ensuring compliance with Chapter (14) related to regulating the practice of engineering professions of the Bahraini building code?

These research questions aim to comprehensively address the challenges of regulatory compliance in the AEC industry and evaluate the potential of AI technologies, particularly the AI-based Q&A model, in overcoming these challenges. By systematically investigating these questions, the study seeks to provide actionable insights and practical solutions that can enhance decision-making processes and improve project outcomes.

1.6) Research Objectives

As this research seeks to address the challenges of ensuring regulatory compliance within the AEC industry, focusing on the complexities of Bahraini building codes, limited data availability, and scarce expert knowledge. The following objectives are derived from the research questions and are designed to systematically approach the key variables under investigation.

1. To Develop a Data-Driven Q&A Model for Bahraini Building Code Compliance:

This objective involves designing, evaluating and testing a sophisticated AI based Q&A model that assists AEC stakeholders by providing instant, accurate answers to regulatory questions.

The related variables are complexity of building codes, decision-making processes, AI model development.

The procedural point of this objective involves the development of AI based Q&A model with a focus on Chapter (14) of the Bahraini building code, ensuring the model's capability to handle the intricate nature of building regulations and enhance decision-making processes.

2. To Evaluate the Feasibility of AI Models with Limited Data:

This objective involves assessing the viability of developing an effective Q&A model using a limited dataset, augmented by human-AI collaboration.

The related variables are data availability, AI model development, decision-making processes.

The procedural point of this objective entails a systematic evaluation of the AI model's performance within the constraints of limited data availability in the context of Bahraini building code compliance, exploring the integration of AI techniques and expert input to mitigate data limitations.

3. To Codify Expert Knowledge into a Structured System:

This objective requires a systematic capture and structure of the tacit knowledge of Bahraini building code experts, making it accessible for broader utilization by stakeholders.

The related variables are expert knowledge, AI model development, decision-making processes.

The procedural point of this objective involves the creation of a framework that transforms implicit expert knowledge into explicit, actionable content, which is then integrated into the AI based Q&A model. The codified knowledge is validated for accuracy and relevance.

4. To Develop a Comprehensive Framework for AI-Based Regulatory Compliance Models:

This objective is to formulate a robust framework for the development of custom Q&A models for regulatory compliance, ensuring adaptability to various chapters of the Bahraini building code and potential integration with BIM systems.

The related variables are AI model development, complexity of Bahraini building code, decision-making processes. The procedural point of this objective includes the development of guidelines and methodologies for future AI model advancements, ensuring the adaptability and expansion of the AI based Q&A model framework across different regulatory domains, with a focus on integration into existing digital workflows such as BIM.

These objectives are designed to systematically address the research questions and are intricately connected to the key variables identified in the study. By aligning each objective with specific variables, the research seeks to comprehensively tackle the challenges associated with optimizing decision-making processes and ensuring regulatory compliance in the AEC industry through the innovative application of AI technologies.

1.7) Theoretical Foundations

The theoretical foundation of this research provides the conceptual framework for analysing the challenges and opportunities within the AEC industry. To address complexities in compliance, decision-making, and technology adoption, this study draws on knowledge management theory which guided the development and evaluation of the AI-based Q&A model, ensuring contextual relevance, operational efficiency, and alignment with the research objectives.

As the knowledge management theory focuses on the processes of creating, sharing, and managing knowledge within organizations to improve efficiency, innovation, and decision-making (Baskerville & Dulipovici, 2006a). This theory underpins the objective of codifying expert This theory underpins the objective of codifying expert knowledge into a structured system. By transforming tacit expert knowledge into explicit, accessible information, this codified knowledge forms the heart of the AI-based Q&A model. This ensures that stakeholders have access to valuable insights that improve compliance and support informed decision-making. The synergy between codifying expert knowledge and leveraging advanced AI technologies ensures that the research objectives are effectively met.

1.8) Research Significance

The significance of this research lies in its potential to revolutionize the AEC industry by addressing critical issues of regulatory compliance and decision-making inefficiencies. This study explores the innovative use of AI

based Q&A model, to provide immediate, reliable answers to regulatory questions, thereby optimizing the design and execution of AEC projects. Listed below the key significances:

1. Significance to the AEC Industry:

This research provides a novel solution to the persistent problem of violations to Bahraini building code, which can lead to costly delays and legal issues in construction projects.

The research is pioneering in its application of AI based Q&A model to Bahraini building code compliance, a topic that has not been previously explored in the Kingdom of Bahrain.

2. Practical Benefits:

The developed Q&A model will offer practical benefits by enabling architects and engineers to ensure compliance with building regulations during the design phase, thereby reducing project inefficiencies.

Similar studies have shown that providing instant access to regulatory information can significantly improve project outcomes.

3. Feasibility with Limited Data:

Demonstrating the feasibility of creating effective AI models with limited data will encourage further research and application of AI in regulatory compliance, even in data-scarce environments.

The approach of using human-AI collaboration to generate additional data is innovative and addresses a common limitation in AI research.

4. Framework for Future Developments:

The proposed framework for developing custom Q&A models for regulatory compliance can be extended to other chapters of the Bahraini building code and integrated with BIM systems, providing a comprehensive solution for the AEC industry.

There is a pressing need for replicating this study to explore the changes and advancements in AI technology and their application in regulatory compliance.

5. Theoretical Foundation for Innovation:

By leveraging knowledge management theory, this research not only addresses immediate challenges in the AEC industry but also provides a robust theoretical foundation for future innovations in regulatory compliance and decision-making systems.

The significance of this research is underscored by its potential to bring about substantial improvements in the AEC industry by leveraging advanced AI technologies to ensure regulatory compliance. By providing a practical, data-driven solution and a framework for future developments, this study addresses critical gaps and offers a path forward for optimizing decision-making processes in construction projects.

1.9) Research Contributions

The research reflects the innovative approaches and methodologies developed to address regulatory compliance issues in the AEC industry. After evaluating various Q&A model creation methods, the researcher identified AI

based Q&A model as the most effective solution. These contributions are categorized to highlight their technological, methodological, theoretical, and practical impacts as detailed below:

1. Innovative AI Based Q&A Model for AEC Industry (Technological Contribution):

The development of an AI based Q&A model tailored for the AEC industry that empowers architects and engineers to ensure compliance with building regulations during the design phase.

This tool reduces the occurrence of violations and enhances project outcomes by providing instant access to regulatory knowledge, optimizing the decision-making process.

2. Methodology for Developing AI Models with Limited Data (Methodological Contribution):

A pioneering methodology demonstrating the feasibility of developing effective AI models with small datasets, supplemented by human-AI collaborative data generation.

This approach broadens the applicability of AI in regulatory compliance, showcasing how limited initial data can be augmented to create reliable models.

3. Framework for Systematic Knowledge Codification (Theoretical Contribution):

Introduce a structured approach for codifying expert tacit knowledge into explicit, actionable information within the Bahraini building code context.

This framework facilitates improved access and decision-making for AEC stakeholders, ensuring that expert knowledge is systematically captured and utilized.

4. Development of a 10-Step Methodological Framework for Custom GPT Model Creation (Key Novel Contribution):

This research introduces a structured 10-step methodological approach that guides the creation, evaluation, and refinement of a Custom GPT model designed for regulatory compliance.

The approach includes exploration, data extraction, expert collaboration, model fine-tuning using small datasets, stakeholder testing, and iterative feedback integration.

This structured methodology not only supports model development but also serves as a replicable framework for future AI-driven compliance tools.

5. Integration Strategy with BIM Systems (Practical Contribution):

A proposal for integrating the AI based Q&A model as an add-in within BIM platforms.

This strategy streamlines regulatory compliance checks and enhances the digital workflows of the AEC industry, allowing for seamless incorporation of regulatory knowledge into existing project management tools.

6. Theoretical Integration:

This study contributes to the academic discourse by integrating the management theory. This framework underpins the development, adoption, and evaluation of the AI-based Q&A model, ensuring its alignment with stakeholder needs and industry challenges.

7. Human-AI Collaboration in Data Expansion (Empirical Contribution):

The process of data compilation involved human-AI collaboration to expand the dataset required to train and customize the model (Step 1 of the methodology) was elaborated in a separate study by the researcher (Al Nama et al., 2024), demonstrating the feasibility of combining human expertise and generative AI for engineering regulatory data creation.

8. Empirical Evaluation of LLMs for Engineering Regulations (Analytical Contribution):

The performance evaluation of two leading large language models—ChatGPT-4 and Gemini 1.5—was conducted and further expanded in an independent study (Al Nama & Mahmud, 2024). The study assessed LLM performance across regulatory question categories, providing comparative insights and validation of model selection decisions within this research.

The research contributions outlined provide a comprehensive view of the innovative advancements made in the field of regulatory compliance for the AEC industry. By developing an AI based Q&A model, enhancing AI

methodologies with limited data, codifying expert knowledge, and proposing integration with BIM systems, this research aims to significantly improve decision-making processes and project outcomes for stakeholders in the AEC sector.

1.10) Scope of the Study

This research is set within the context of the AEC industry, with a specific focus on regulatory compliance during the design stage of construction projects. The study is conducted in an environment where there is a pressing need to address the issue of violations to Bahraini building code. The setting involves collaboration with industry experts, regulatory bodies, and potential users to gather relevant data and insights.

The study leverages the advancements in AI technology, particularly CustomGPT, to develop a model that can provide instant answers to regulatory questions. The research setting includes various stages of construction projects, from initial design to final approval, where ensuring compliance with the Bahraini building code is crucial. Additionally, the study considers the integration of the developed model with BIM systems to streamline regulatory compliance checks within existing digital workflows.

From a topical standpoint, the study addresses key issues such as the need for efficient decision-making tools in the AEC industry, the challenges of ensuring regulatory compliance, and the potential of AI technologies to revolutionize traditional practices. The setting emphasizes the practical

application of the research findings to improve project outcomes and reduce the incidence of non-compliance in construction projects.

1.11) Ethical Consideration

This research considered ethical considerations such as informed consent, confidentiality and privacy. All participants will be given informed consent forms to read and sign. The forms will explain the purpose of the research, the risks and benefits of participating in the research, and the participants' rights. All data collected in this research will be kept confidential and stored in a secure location that is only accessible to the researchers. Participants were asked to provide their contact information for follow-up interviews, but this information will not be shared with any other individuals or organizations.

1.12) Organization of the Thesis

This research report is organized into five main chapters alongside the abstract as follows:

- **Chapter (1): Introduction to the Study:** This chapter introduces the research by covering the background, problem statement, research questions, and objectives. It also discusses the research significance, the study's setting, research limitations, and provides a brief introduction to the research structure.
- **Chapter (2): Literature Review:** This chapter provides a review of existing literature relevant to the study, including an overview of key

variables, the theoretical framework, and identification of research gaps. It concludes with a summary of previous studies and their implications for the current research.

- **Chapter (3): Research Methodology:** This chapter details the research methodology, including research design, sampling strategy, data collection procedures, and the statistical methods employed. It also discusses the research tools, their reliability and validity, and the ethical considerations associated with the study.

Chapter (4): Results: This chapter presents the findings of the study in relation to the research objectives. Results are presented through tables and figures, followed by a discussion that links the findings to the research questions, and existing literature.

Chapter (5): Discussion and Conclusions: This chapter concludes the thesis by summarizing the key findings and developed framework, discussing their implications, and offering recommendations for practice and future research. It also reflects on the limitations of the study and suggests areas for further investigation.

Chapter Two

Theoretical Framework and Literature Review

2.1) Introduction

The Architecture, Engineering, and Construction (AEC) industry is governed by complex regulatory frameworks that require stringent compliance to ensure safety, sustainability, and efficiency. However, navigating these regulations poses significant challenges for stakeholders, particularly in dynamic and resource-constrained environments (Malsane et al., 2015). This chapter provides a comprehensive review of the literature relevant to improving regulatory compliance and decision-making in the AEC industry, with a particular focus on leveraging emerging technologies. The review begins by examining traditional approaches to decision-making and regulatory compliance, highlighting their limitations in addressing the complexities of modern construction projects. It then explores advanced technologies such as Building Information Modeling (BIM) and Questions and Answers (Q&A) models to enhance compliance outcomes.

Additionally, the chapter delves into the theoretical foundations of knowledge management, expert systems, and Machine Learning (ML) methodologies that underpin the development of efficient Q&A systems. By addressing topics such as Frequently Asked Questions (FAQs) in regulatory domains, advancements in Large Language Models (LLMs), and fine-tuning techniques, this chapter aims to establish a robust foundation for understanding the potential of Q&A models in revolutionizing compliance practices. Through critical analysis and synthesis of past and current research, this chapter identifies existing gaps and emerging opportunities in the field. The insights derived here will directly inform the development of a tailored

Q&A model capable of supporting AEC stakeholders in navigating regulatory challenges both today and in the future.

2.2) Optimization of AEC Stakeholder Decision-Making for Building Regulation Compliance

Improving decision-making among stakeholders in AEC projects is critical, particularly in ensuring compliance with building regulations. Various methodologies that emphasize collaboration and address the complexities of multi-stakeholder environments have been shown to enhance this process. Digital technologies can significantly improve decision-making processes by promoting communication and coordination among stakeholder (X. Chen et al., 2025). Decision Support Systems (DSS), defined as computer-based systems that assist in decision-making by integrating data, analytical tools, and models, are particularly valuable in this context. These systems provide structured support for analysing complex problems and exploring alternative solutions, making them highly relevant to industries like AEC, where project success depends on balancing multiple variables and stakeholder needs (Power, 2002).

Moreover, understanding the critical factors that influence stakeholder engagement in design optimization—such as the role of subcontractors and the mechanisms for performance evaluation—is essential for achieving management goals in Engineering, Procurement, and Construction (EPC) projects (Y. Chen et al., 2023; Muthumanickam et al., 2023). A methodology that integrates stakeholder preferences early in the design process can lead to

outcomes that better reflect the needs and goals of all parties involved (Zhilyaev et al., 2022).

Incorporating Monte Carlo simulations into the decision-making process, as demonstrated by Prabatha et al. (2022), can help manage uncertainties and varying stakeholder priorities in energy performance contracts. Monte Carlo simulations aim to model the probability distributions of uncertain parameters, thereby predicting the overall uncertainties associated with the outcomes. This approach underscores the importance of flexible decision-making frameworks that can adapt to the diverse expectations of stakeholders. Collectively, these approaches highlight the need for integrating stakeholder perspectives and managing uncertainties to improve decision-making in AEC projects (Prabatha et al., 2022). However, their effectiveness in the Kingdom of Bahrain will depend on developing localized datasets, enhancing data integration capabilities, and fostering collaboration between regulatory authorities and industry stakeholders, to reduce uncertainties.

To enhance compliance with building regulations, integrating advanced DSS and performance-based metrics is crucial. Cano et al. (2017) propose a DSS framework that supports stakeholder dialogue and dynamic risk assessment, essential for managing building infrastructure complexities and ensuring regulatory compliance. The model uses a two-stage, dynamic stochastic optimization approach with moving random time horizons, allowing for the modelling of extreme events and structural changes from stakeholder dialogue. Implemented via an algebraic modelling language, this DSS framework facilitates robust planning and operation of building infrastructure, addressing systemic economic and environmental changes. It also enforces stakeholder dialogue throughout the decision-making process,

incorporating feedback into the DSS structure and decisions, and includes dynamic risk measures for strategic decisions (Cano et al., 2017), this model highlights the importance of making decisions in real-time, particularly under uncertain conditions. Similarly, (Mateus et al., 2024) emphasize structuring regulations through decision analysis practices and advocate for performance-based metrics that can adapt to regulatory changes. Their work, which was validated through a case study on light rights regulation, highlights the need for clear compliance thresholds and robust monitoring mechanisms. Additionally, Strobbe et al., (2012) discuss the benefits of integrating building simulation tools with heuristic design methods to improve compliance with energy performance regulations. This blend of stakeholder engagement, structured decision-making, and simulation tools provides a comprehensive strategy for optimizing compliance in building design and regulation processes (Strobbe et al., 2012).

2.2.1) Integration of Advanced Technologies and Decision Support Systems

While recent advancements in Automated Compliance Checking (ACC) for building regulations have shown potential for improving both the efficiency and accuracy of compliance processes, there are still significant challenges to address. (Z. Zhang et al., 2022) conducted a systematic review of ACC technologies and identified key areas needing further research, such as improving the representation of complex regulatory rules and developing real-time DSS for stakeholders. Their findings suggest that future research should focus on creating methodologies that can effectively interpret

ambiguous rules and automate the compliance process more comprehensively (Z. Zhang et al., 2022).

Moreover, there is a growing need for DSS that can provide immediate, accurate responses to inquiries from stakeholders such as architects and engineers. (Zhong et al., 2020) propose a question-answering system based on deep learning that is designed to retrieve regulatory information quickly and accurately, which is critical for rapid decision-making in compliance scenarios. By combining NLP with deep learning, this system could significantly enhance how regulatory compliance is managed in the AEC industry, offering precise, real-time answers to complex regulatory queries (Zhong et al., 2020).

These studies underscore the importance of continuing to explore the integration of advanced technologies and DSS within the AEC industry. Addressing the research gaps identified in these areas could lead to significant improvements in stakeholder decision-making and compliance with building regulations, making these processes more efficient and timelier.

2.3) Compliance with Building Regulations

Compliance with building regulations is a fundamental aspect of the AEC industry. These regulations are established to ensure that buildings are constructed safely, sustainably, and in accordance with national and international standards. They encompass various aspects, including structural integrity, fire safety, accessibility, and environmental impact. Adherence to these regulations is critical, as non-compliance can lead to severe legal, financial, and safety repercussions (Health and Safety Executive, 2025). The

following sections will discuss the role of regulatory bodies in enforcing these standards and examine common violations and their implications.

2.3.1) Regulatory Bodies and Their Role

Regulatory bodies in the AEC industry play a crucial role in ensuring that all construction activities adhere to established safety, quality, and environmental standards. These bodies operate at various levels—local, national, and international—and are responsible for setting and enforcing regulations that govern the construction process from design through to completion and occupancy.

In many countries, regulatory bodies such as the Health and Safety Executive (HSE) oversee workplace safety, including construction sites. These bodies enforce regulations that ensure safe working conditions and adherence to safety protocols during the construction process. HSE plays a crucial role in monitoring compliance, investigating incidents, and ensuring that safety standards are consistently met across the industry (Health and Safety Executive, 2025).

In the Kingdom of Bahrain, the regulatory framework is overseen by several key organizations, including the Ministry of Works, Municipalities Affairs, and Urban Planning, which is responsible for urban planning and infrastructure development. The Bahrain Electricity and Water Authority plays a significant role in regulating the utility aspects of construction projects, while the Civil Defence Directorate ensures compliance with fire safety regulations (Benayat, 2025).

2.3.2) Common Violations and Their Implications

Common violations in the AEC industry typically stem from non-compliance with safety standards, building codes, or environmental regulations. These violations can have severe implications, including legal penalties, project delays, financial losses, and, most critically, risks to human safety.

Structural violations are particularly concerning because failure to comply with structural codes can lead to building failures or collapses, posing significant risks to occupants and resulting in legal liabilities for contractors and developers (Occupational Safety and Health Administration, 2025). Fire safety violations are equally critical. Non-compliance with fire safety regulations, such as inadequate fire exits or the use of non-compliant materials, can lead to catastrophic fires, endangering lives and resulting in substantial damage to property (National Fire Protection Association, 2025). Environmental violations are also of great concern. Violations of environmental regulations, such as improper waste disposal or failure to implement energy-efficient systems, can result in significant fines and reputational damage (Environmental Protection Agency, 2025; National Fire Protection Association, 2025).

As introduced in Chapter (1), the violations of AEC projects in the Kingdom of Bahrain are mostly related to not abiding by the building regulations stipulated in the unified guidebook of building permit regulations, as elaborated in following sections of this research.

2.4) Adopted Solutions to Reduce Violations

To mitigate the risk of non-compliance, the construction industry has adopted several strategies aimed at ensuring adherence to regulatory standards and reducing the incidence of violations. These strategies are designed to address the root causes of non-compliance and promote a culture of safety and accountability across all levels of project execution.

However, while these traditional methods have shown effectiveness in addressing immediate compliance challenges, they often fall short in providing real-time, context-specific guidance to stakeholders. For instance, reliance on periodic training or external consultants may not adequately address the dynamic and complex nature of modern construction projects. This limitation underscores the need for innovative, tech-driven solutions.

2.4.1) Enhanced Training and Education Programs

One of the primary strategies involves the implementation of comprehensive training and education programs. These programs are tailored to keep all stakeholders, including workers, engineers, and project managers, informed about the latest regulatory requirements and best practices. By focusing on continuous education, these initiatives aim to reduce the knowledge gap that often leads to non-compliance. Recent studies highlight that regular and targeted training significantly improves compliance rates and reduces the likelihood of violations on construction sites (AMI Environmental, 2025).

Tamkeen, the Labour Fund in the Kingdom of Bahrain, plays a crucial role in providing specialized training across various industries, including

engineering. Their initiatives aim to equip Bahrainis with the skills needed to thrive in a competitive job market. Tamkeen collaborates with global training platforms and local institutions to offer programs in high-demand fields such as technology and digital marketing. These training schemes are developed collaboratively with stakeholders, ensuring alignment with market demands and industry standards. This unique opportunity can offer a chance to AEC stakeholders to get the required training in case any new strategy is adopted aiming at reducing the violations to building regulations (Tamkeen, 2025).

2.4.2) Engagement of Compliance Consultants

The engagement of specialized compliance consultants has become a widespread practice. These professionals bring expertise in navigating complex regulatory frameworks and provide ongoing oversight throughout the construction process. Their involvement helps in identifying potential compliance issues early and implementing corrective measures before they escalate into serious violations. According to (K2 Integrity, 2023), the use of independent third-party consultants not only enhances compliance but also reduces legal risks and project delays (K2 Integrity, 2023).

2.4.3) Adoption of Advanced Monitoring Technologies

The integration of advanced monitoring technologies is another key strategy. These technologies include the use of real-time data collection tools, automated inspection systems, and predictive analytics to monitor compliance in real-time. This approach enables project managers to detect and address potential violations swiftly, thereby minimizing the impact on project timelines and budgets. The implementation of such technologies is increasingly recognized as a best practice in the industry, as it allows for more

proactive management of compliance-related risks (Compliance Chain, 2024).

2.4.4) Strengthening Regulatory Frameworks and Penalties

In response to persistent compliance challenges, regulatory bodies have also taken steps to strengthen the existing frameworks and increase penalties for violations. For example, recent legislative changes, such as the introduction of Carlos's Law in New York, have significantly raised the stakes for non-compliance, with stricter penalties and enhanced enforcement measures. These changes aim to compel companies to prioritize compliance and adopt more robust safety protocols (K2 Integrity, 2023).

2.5) Building Information Modelling (BIM)

Building Information Modelling (BIM) represents a significant evolution in the field of AEC, by leveraging advanced technologies and cloud-based platforms, BIM enables stakeholders to visualize, simulate, and manage the entire lifecycle of a building project more effectively (Autodesk, 2025). The following sections will delve into the definition and importance of BIM, examining how it is reshaping the AEC industry and overcoming traditional challenges.

2.5.1) Definition and Importance of BIM

BIM is an advanced digital representation of a building's physical and functional characteristics. It serves as a shared knowledge resource, enabling stakeholders to make informed decisions throughout the entire lifecycle of a facility—from the earliest stages of conception through to demolition. BIM is

widely recognized for its ability to improve the accuracy of designs, enhance collaboration among various stakeholders, and increase overall project efficiency by integrating multiple dimensions such as time, cost, and sustainability into the management process (Sadek et al., 2020).

Moreover, BIM plays a crucial role in driving the digital transformation within the architect AEC industry. As a leader in BIM solutions, Autodesk has helped industry professionals optimize workflows and achieve superior outcomes for both businesses and the built environment. BIM operates on an intelligent model, enabled by a cloud-based platform, that integrates structured, multidisciplinary data to create a comprehensive digital representation of an asset across its lifecycle, ensuring that all aspects of the project are meticulously managed from planning and design to construction and operations (Autodesk, 2025).

2.5.2) Challenges of BIM Adoption in the AEC Industry

The adoption of BIM in the AEC industry is growing, yet it faces several challenges, particularly in regions like the Gulf Cooperation Council (GCC). Despite its potential to revolutionize the construction process, BIM adoption in the GCC remains slow due to factors such as a lack of trained professionals, resistance to change from traditional practices, and inadequate government support. In Bahrain, for instance, the implementation of BIM is hindered by organizational and technical challenges, including the absence of standardized protocols and a comprehensive legal framework (Umar, 2022). Moreover, the limited awareness of BIM's benefits among smaller contractors and the high cost of BIM technology further exacerbates these challenges (Olawumi et al., 2018).

2.5.3) Identifying Future Research Directions BIM Adoption

Despite the progress in ensuring compliance, there are several research gaps that need to be addressed:

1) Impact of Emerging Technologies

As new technologies like AI and IoT become integrated into construction processes, there is a need for research into how these can be used to enhance compliance with building regulations. For instance, the integration of AI-driven Q&A models within BIM frameworks could offer a real-time decision support system that enhances stakeholders' ability to make informed decisions during the early design stages (Fadoul et al., 2020).

2) Long-Term Effectiveness of Compliance Strategies

While current strategies are effective, more research is needed to assess their long-term impact on reducing violations and improving overall building safety. The use of Q&A models in BIM has shown promise, but more empirical studies are needed to validate their effectiveness over time (Villaschi et al., 2022).

3) Regulatory Harmonization

There is a lack of comprehensive research on the harmonization of building codes across different regions and how this could be streamlined to facilitate international construction projects. Harmonizing regulations and incorporating them into BIM tools could enhance global collaboration and ensure that projects meet diverse regulatory requirements (Sleiman & Robert, n.d.).

2.5.4) Challenges and Recommendations for Bahrain

Despite these advancements, the widespread adoption of BIM and Q&A models in Bahrain faces several challenges. These include a lack of trained professionals, limited awareness among smaller contractors, and the need for a robust regulatory framework that mandates the use of such technologies. For example, studies have shown that in the GCC region, including Bahrain, the implementation of BIM is still at a slow rate due to several barriers. These barriers include organizational resistance, technical challenges, and inadequate government support. The lack of standardization and the absence of a clear legal framework further complicate the adoption of BIM (Umar, 2022).

To overcome these challenges, Bahrain could benefit from developing targeted training programs, offering incentives for BIM adoption, and establishing clear guidelines on its use in ensuring compliance. Moreover, enhancing collaboration between the government, academic institutions, and industry stakeholders will be crucial in fostering a culture of innovation and continuous improvement in the construction sector. This is supported by research suggesting that collaboration between stakeholders and government incentives are critical to accelerating BIM adoption in the region (Olawumi et al., 2018).

As BIM adoption in Bahrain remains limited, implementing a Q&A model for compliance checking becomes an immediate necessity. Such a model will provide architects, engineers, and stakeholders with accurate, real-time information on building regulations, reducing the likelihood of violations.

2.6) Q&A Models for Compliance Checking and Decision Making in the AEC Industry

Stakeholder decision-making in the AEC industry is a complex process that relies on timely access to accurate information. Traditional methods of information retrieval can be slow and cumbersome, often hindering effective decision-making. However, recent advancements in Artificial Intelligence (AI) and NLP enable the development of Q&A models specifically designed for the AEC industry. These models hold immense potential to revolutionize stakeholder decision-making by empowering stakeholders with the ability to ask natural language questions and receive relevant answers instantly.

In the finance industry, Q&A models utilize domain-specific fine-tuning and iterative reasoning to deliver precise financial insights (Nguyen et al., 2024). Similarly, in healthcare, AI-powered Q&A systems support medical professionals by accessing relevant medical literature and patient records (IBM Watson Health, 2024). These instances illustrate the practicality and significance of developing a Q&A model for the AEC industry, showcasing how analogous sectors have effectively adopted these technologies to improve decision-making and operational efficiency.

Further detailed below are how Q&A models can enhance stakeholder decision-making in AEC projects. The focus is primarily on research conducted after 2020 to gain insights into the latest methodologies, findings, and recommendations in this field. However, earlier foundational studies are also acknowledged to provide context for ongoing advancements.

2.6.1) The Rise of Q&A Models and Improved Information Access in AEC (2020-Present)

Q&A models have emerged as pivotal tools within the AEC domain, addressing critical activities such as knowledge management and information retrieval. The AEC sector's data complexity, coupled with the challenges of non-standardized information models and natural language query ambiguity, necessitates the deployment of advanced algorithms and methodologies. These include machine learning, information extraction, and semantic parsing to navigate and resolve the sector's inherent complexities (Ren & Zhang, 2021; Zheng et al., 2022).

Early research laid the groundwork for contemporary Q&A models in AEC. Studies explored utilizing information retrieval systems (Nabavi et al., 2023) to navigate vast amounts of project data stored in BIM software. Other research investigated the potential of natural language interfaces for BIM (N. Wang et al., 2022a), paving the way for Q&A interaction with project models. These initial efforts established the importance of facilitating user-friendly access to project information.

Recent advancements in AI and NLP have led to the development of robust Q&A models specifically designed for the AEC industry. These models can access and process information from various sources, including BIM models, project documents, and industry regulations (J. Kim et al., 2022). This allows stakeholders, regardless of their technical expertise, to ask questions in natural language and receive relevant and timely answers (N. Wang et al., 2021). This democratization of access to information empowers stakeholders

to participate more actively in decision-making processes (Kovacevic et al., 2008).

In a similar vein, the Benayat website is Bahrain's official portal for issuing building permits for various types of projects, including residential, commercial, and industrial. It provides a centralized system for applicants to submit permit requests, track application statuses, and access essential information through an interactive map. The platform ensures compliance with regulations and streamlines the building permit process for engineering offices and property owners. The Benayat website also has a FAQ section where AEC stakeholders can refer to or contact officials to get answers to their queries (Benayat, 2025).

2.6.2) Enhancing Regulatory Compliance and Informed Decision-Making through Q&A Models

Q&A models serve a vital function in querying about AEC regulations, such as construction codes and environmental guidelines, which are essential yet complex and dynamic in nature. By providing precise and relevant answers, these models assist stakeholders in understanding and adhering to regulations, thereby ensuring project sustainability and safety. Overcoming challenges such as regulation formalization, data alignment, and model validation remains a critical area of focus for enhancing these systems' efficacy (European Commission, 2021, 2022).

Q&A models can significantly enhance stakeholder decision-making by providing real-time insights throughout the project lifecycle. For instance, architects can use Q&A models to explore the cost implications of design choices before finalizing plans (Yuxia & Ruonan, 2021). Similarly, engineers

can leverage the model to assess the structural feasibility of design variations (Robert et al., 2006). Stakeholders can also gain insights into potential environmental impacts or code compliance issues associated with design decisions (Haitao et al., 2014). By enabling stakeholders to explore “what-if” scenarios through Q&A, these models can foster informed decision-making and reduce the risk of costly errors later in the project.

2.6.3) Focus on Recent Research Related to Functionalities of Q&A Models for AEC Applications (2018-2022)

Recent research delves deeper into the specific functionalities of Q&A models for AEC applications. Studies explore the use of deep learning techniques to train Q&A models on domain-specific data, improving their accuracy and ability to understand complex technical inquiries. For instance, Wu et al. (2022) proposes a novel approach for BIM object classification using ML algorithms. This method achieves an accuracy of 99.6% F1-measure, significantly improving the ability of Q&A models to answer intricate questions about building components (Wu et al., 2022). Additionally, (Xu et al., 2018) investigates the use of rule-based systems and ML for automating complex tasks in BIM. These advancements pave the way for Q&A models with reasoning abilities, capable of not just retrieving information but also understanding the context and relationships between different data points within a BIM model.

Furthermore, research investigates the integration of Q&A models with existing AEC software, creating a seamless workflow for stakeholders to access information within their familiar design environments (Wenwu, 2018).

This integration is crucial for user adoption and ensuring that stakeholders leverage Q&A models throughout the project lifecycle.

Another area for further exploration is the integration of Q&A models with generative design tools. By combining the question-answering capabilities with the ability to generate new design options based on user queries, Q&A models could become even more powerful tools for creative exploration and decision-making in the AEC industry (Maureira et al., 2021).

2.7) The Role of FAQs in Regulatory Domains and Building Regulations

Frequently Asked Questions (FAQs) are integral to regulatory environments, particularly within fields such as building regulations and engineering licensure. These resources simplify complex regulatory frameworks into accessible formats, assisting professionals in the AEC industry in making informed decisions. The effective organization of FAQs enhances decision-making by offering clear guidance on compliance, a critical factor in the successful execution of projects.

2.7.1) Types of FAQs in the AEC Industry

FAQs within regulatory and building regulation contexts can be categorized into several distinct types, each serving a unique function:

- 1) Factual Questions:** Factual questions are designed to provide specific, concrete information grounded in established regulations. For example, an engineer might ask, “What are the required qualifications for obtaining a professional engineering license?” or “What are the fire

safety requirements for high-rise buildings?” The provision of accurate factual information is crucial for ensuring compliance with regulations across construction projects.

Accurate information is vital in the AEC industry, as it strengthens decision-making and fosters better collaboration among stakeholders. Although digital tools like BIM streamline information sharing, their effectiveness often depends on human factors such as teamwork and cognitive biases that influence decisions (S. Sujan, 2020; S. F. Sujan et al., 2019). Studies suggest that using structured methods for retrieving information, like conversational AI systems, can greatly enhance access to relevant data, aiding construction professionals in making well-informed decisions (Marc Robert et al., 2006; N. Wang et al., 2022b). Moreover, employing question-answering systems that deliver direct answers, rather than just references, helps bridge the gap between available information and its practical use on-site (Marc Robert et al., 2006). In summary, providing factual answers is key to reducing biases, improving collaboration, and enabling the AEC industry to fully utilize the abundant information available, thereby boosting project outcomes (S. F. Sujan et al., 2019; N. Wang et al., 2022b).

2) Theoretical Questions: Theoretical questions delve into the underlying principles or rationales behind specific regulations. Such questions might explore the reasoning for requiring particular materials in seismic design or the purpose of continuing education for licensure renewal. A deep understanding of these principles is vital for professionals to appreciate the importance of compliance. Theoretical knowledge in the AEC industry is crucial for understanding factors

affecting project collaboration and digitalization implementation, as highlighted in the holistic model developed in the research (S. Sujan, 2020). Theoretical knowledge in construction fosters a deeper understanding of sustainable practices and compliance with evolving regulatory standards.

3) Hypothetical Questions: Hypothetical questions, which explore “what if” scenarios, are crucial in allowing stakeholders to investigate potential outcomes and the effects of various actions. For example, an engineer might ask, “What would happen if new environmental regulations were applied retroactively to existing structures?” These types of questions are fundamental in strategic planning and risk assessment.

In the AEC industry, scenario-based planning has become increasingly important for managing uncertainty and enhancing adaptability. This method enables stakeholders to consider different “what-if” situations, deepening their understanding of potential results and how users might interact with changing environments (Eilouti, 2018). By using scenario planning, decision-makers can better foresee disruptive trends and craft innovative business models that are in line with future market demands (Kamers, 2018; Lavikka et al., 2018). Additionally, scenario-based design methods support the documentation and development of architectural decisions, ensuring that the final systems remain consistent with the original requirements. This approach not only helps in addressing the complexities of modern architectural challenges but also encourages a proactive approach to strategic planning, ultimately leading to more resilient and user-focused designs (Meiru Che, 2011).

Therefore, scenario-based planning is essential for driving innovation and ensuring the long-term success of AEC projects in a constantly changing environment.

- 4) Comparative Questions:** Comparative questions involve analysing differences or similarities between various regulatory processes, standards, or practices. For example, an engineer might inquire, “How does the licensure process in the United States compare with that in the European Union?” or “What are the differences between current building codes and those from a decade ago?”. Comparative analysis is crucial for professionals working in multiple regions or those seeking to align with international standards.

Comparative questions are vital in the AEC industry, as they enable the assessment of various practices, technologies, and organizational behaviours. Research shows that by understanding the difficulties linked to different types of innovations, AEC firms can improve their success rates in adopting new approaches. Notably, specialized firms often adapt more effectively to changes compared to general contractors (Maali et al., 2022). Additionally, comparative studies of digital skills among professionals reveal significant differences in the adoption of digital technologies, highlighting the critical need for targeted upskilling, particularly in smaller firms (S. F. Sujan et al., 2020). Moreover, comparing projects has proven invaluable in evaluating collaborative cultures, helping to identify key factors that contribute to effective teamwork (S. F. Sujan et al., 2020). Finally, assessing BIM practices across different countries exposes regulatory gaps and challenges, demonstrating the importance of comparative

analysis in raising industry standards and practices (Sarı & Pekerçli, 2020). In conclusion, comparative questions are crucial for fostering innovation and boosting operational efficiency within the AEC sector.

5) Open-Ended Discussion Questions: Open-ended discussion questions encourage broader reflection and dialogue, often delving into complex regulatory challenges without necessarily having one correct answer. These questions might explore issues like the obstacles to implementing sustainable practices or enhancing regulatory compliance. For example, stakeholders could discuss, “What are the biggest challenges in adopting sustainable building practices in urban settings?”

In the AEC industry, open-ended questions are crucial for sparking innovation, deepening learning, and boosting user engagement. Research shows that in architectural education, open-ended approaches push students to think critically and creatively, going beyond memorization to develop a deeper understanding of design principles and problem-solving strategies (Pott & Nortjé, 2021). This approach is closely tied to the concept of open architecture, which values user input throughout the design process, leading to more personalized and sustainable environments (Plummer, 2021). Additionally, integrating open design strategies into architectural education has proven effective in preparing students to tackle complex societal challenges like sustainability, as it encourages them to consider the dynamic nature of buildings and their contexts (Stam et al., 2022). However, the construction industry has been slow to adopt such innovative practices,

underscoring the need for a cultural shift toward embracing open-ended inquiry to drive significant advancements (Salvalai et al., 2023). In summary, open-ended questions are vital for creating a more adaptive and responsive AEC industry, engaging stakeholders, and improving policy outcomes within the construction sector.

2.7.2) The Importance of Categorizing FAQs for Q&A Model Development

Identifying and categorizing these different types of FAQs is critical for developing effective Q&A models, particularly in regulatory contexts. As regulatory frameworks grow increasingly complex, professionals in the AEC industry rely heavily on accurate and timely information to make informed decisions. Well-structured FAQs ensure that Q&A models can address a wide range of inquiries with precision, supporting compliance and facilitating informed decision-making.

Accurate question classification is essential for improving the retrieval of correct answers. Systems that effectively categorize questions can differentiate between subjective and objective queries, which leads to more relevant and precise responses. (Panicker & Venkitakrishnan, 2012) emphasize that proper classification directly influences the accuracy of the answers provided by these systems. A well-structured classification framework also significantly enhances the user experience. When questions are categorized accurately, users receive answers that closely align with their expectations, resulting in greater satisfaction.

Incorporating various techniques, such as combining rule-based methods with ML approaches, has been shown to improve classification accuracy.

Many studies highlight how integrating diverse categorization rules contributes to more accurate question classification, further validating the importance of a hybrid approach in these systems. In e-learning environments, categorizing questions based on concepts is essential for facilitating targeted learning experiences. This process helps in organizing educational content and guiding students more effectively, ensuring they receive the most relevant information for their needs. (Panicker & Venkitakrishnan, 2012) that contextual relevance, achieved through proper question classification, is crucial for effective knowledge dissemination in educational settings.

While the benefits of categorizing FAQs are clear, maintaining the accuracy of classification systems amid the growing complexity of natural language remains a challenge. Continuous advancements in ML and rule-based techniques are necessary to meet these challenges effectively. By staying at the forefront of these developments, classification systems can continue to improve and adapt to the evolving demands of natural language processing, particularly in regulatory and educational domains.

2.8) Advancing Knowledge Management and Expert Systems through Experiential Learning

Knowledge Management (KM) is the process of capturing, storing, sharing, and effectively using organizational knowledge. It involves both technological tools and organizational practices aimed at leveraging intellectual assets to improve efficiency, foster innovation, and maintain competitive advantage. KM involves systematically acquiring, organizing,

sharing, and utilizing knowledge within an organization. This concept has evolved significantly from its early days to the current AI age. There are two primary types of knowledge: explicit, which is codified and easily shared, and tacit, which is personal and context specific. Theories of KM emphasize its strategic importance in maintaining a competitive edge, especially in dynamic and volatile environments (Aktharsha, 2011). For instance, Syed Aktharsha's theory highlights the systemic approach of KM in organizational settings, underscoring its role in sustaining competitive advantage (Aktharsha, 2011).

The importance of expert knowledge management has grown, especially in fields requiring high levels of expertise, such as finance and water resources management. Expert systems, which mimic human decision-making, have become integral to KM, particularly in developing AI applications for knowledge acquisition and dissemination. For example, (Grljevic & Zita, 2011) discuss how expert systems facilitate KM in financial sectors by enhancing credit risk measurement and financial diagnostics.

Developments in KM have been marked by the integration of a technology, which allows for more sophisticated handling of large datasets and complex problem-solving. This is evident in a project for water resources management, which applies technology to effectively manage water-related challenges (M. Toman, 2007).

Furthermore, the theoretical foundations of KM are continually expanding, incorporating concepts from various disciplines such as information economics and strategic management, as detailed by (Baskerville & Dulipovici, 2006b). A study published in the *Journal of Knowledge Management* explores the latest trends in KM for 2025, highlighting the

incorporation of AI and data-driven methodologies. The study emphasizes the critical need to consider data as a valuable asset and establish strong data governance frameworks to maximize the utility of high-value information. This approach fits well with the need for continuous learning and hands-on training to fully use AI, making sure that knowledge management in the AEC industry is effective (Little, 2025; Onatayo et al., 2024).

By leveraging explicit and tacit knowledge, the Q&A model can provide accurate, reliable, and contextually relevant answers to user inquiries, making it a vital tool for improving decision-making in industries like AEC.

2.8.1) Stages of Knowledge Management

KM in the AEC industry is a multifaceted process that plays a crucial role in enhancing project efficiency and facilitating the reuse of valuable knowledge. The KM process can be broken down into several key stages, each contributing to the overall effectiveness of managing knowledge within the industry as illustrated in Figure (2.1).

KM in the AEC industry is a critical process that contributes significantly to enhancing project efficiency and the reuse of valuable knowledge. This process can be understood through several stages, each playing a pivotal role in the effective management of knowledge within the industry.

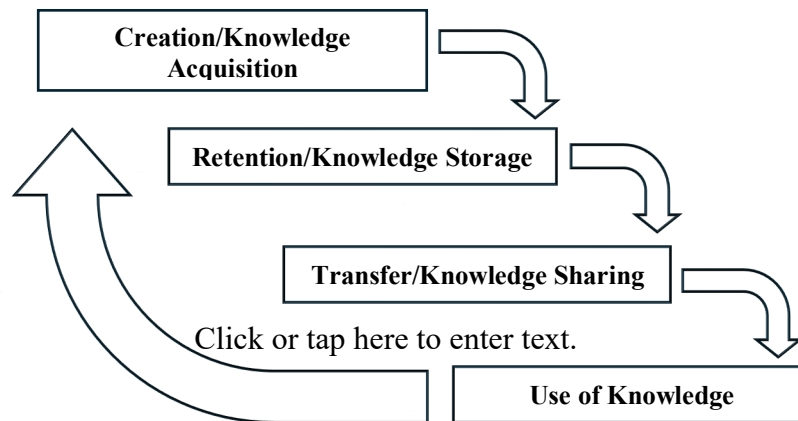


Figure (2.1): Processes of knowledge management (adopted from (Marques et al., 2011))

- **Knowledge Creation:** The first stage, knowledge creation, typically occurs during the planning and design phases of a project. During this phase, knowledge is generated through collaboration and innovation among stakeholders. BIM is particularly influential at this stage, as it facilitates the integration of knowledge from various disciplines, thereby improving design accuracy and overall project efficiency (Jasim et al., 2021; H. Wang et al., 2018). The collaborative nature of BIM allows for the pooling of diverse expertise, leading to more informed decision-making and better project outcomes.
- **Knowledge Storage:** Following knowledge creation, the next crucial step is knowledge storage. Systematic storage of knowledge is essential for its future use, involving the organization and classification of information in a way that makes it easily retrievable. The development of a building knowledge modelling library is one effective approach to this, as it enables the systematic storage and retrieval of design knowledge, ensuring that valuable insights from past projects are preserved for future reference (Jasim et al., 2021). Such a library not

only prevents the loss of critical information but also supports continuous improvement by enabling the reuse of knowledge in new projects.

- **Knowledge Sharing:** Knowledge sharing is the third stage in the KM process, is vital for disseminating lessons learned from past projects and for fostering continuous improvement within the AEC industry. Despite the common practice of conducting post-project evaluations and meetings, many organizations face challenges in effectively sharing this knowledge across teams. The integration of BIM has been suggested as a solution to this challenge, offering a collaborative platform that enhances knowledge exchange and ensures that insights gained are not siloed but are accessible across the organization (Debs & Hubbard, 2023; H. Wang et al., 2018). This enhanced sharing capability is crucial for ensuring that knowledge contributes to future projects, rather than remaining underutilized.
- **Knowledge Application:** Finally, knowledge application is the stage where stored knowledge is applied to new projects, driving continuous improvement within the AEC industry. However, this stage is often challenged by factors such as project complexity and communication gaps, which can hinder the effective use of past knowledge. Overcoming these barriers is essential to ensure that the knowledge captured during previous stages is not only stored but also effectively utilized in subsequent projects (Korkmaz & Bahidrah, 2018). Effective knowledge application is critical for translating the potential of KM into

tangible improvements in project outcomes and innovation within the industry.

These stages of KM provide a strong foundation for the development of a Q&A model. Each stage—creation, storage, sharing, and application—ensures that the knowledge powering the model is accurate, accessible, and actionable (Korkmaz & Bahidrah, 2018). By effectively managing knowledge, the model can deliver precise answers to stakeholders' inquiries, reducing risks and improving compliance outcomes.

2.8.2) Knowledge Based Systems

Knowledge-based Systems (KBS) represent a specialized branch of AI focused on facilitating machine decision-making by leveraging existing knowledge (Abioye et al., 2021). A KBS typically comprises three key components: a knowledge base, an inference engine, and a user interface. The knowledge base is built from the storage of domain-specific expertise, past cases, experiences, and other relevant sources. The primary advantage of KBS lies in its ability to enhance productivity and efficiency by providing easy access to and interaction with extensive domain knowledge. Through heuristic, flexible, and transparent inference processes, KBS can arrive at well-reasoned conclusions and offer logical advice when necessary (Abioye et al., 2021). Incorporating KBS into a Q&A model ensures that the system can dynamically access domain-specific knowledge and provide contextually appropriate responses. This integration enhances the model's capability to support decision-making in complex regulatory environments.

Knowledge-based systems can be broadly categorized into several types, each serving distinct purposes, including the expert systems which rely on a

knowledge base filled with task-specific information drawn from experts within a particular domain. The primary goal of expert systems is to mimic human decision-making processes to solve specific problems effectively (Salmanova, 2024). By replicating expert reasoning, these systems offer valuable support in complex decision-making scenarios.

Another type is the Case-based Reasoning (CBR) Systems that utilize a knowledge base composed of past experiences or old cases. These cases are employed to explain, interpret, or critique new situations. To be effective, especially in certain domains, CBR systems must incorporate expert knowledge to select cases that closely resemble the new project or problem at hand (Juell & Paulson, 2003). This approach allows for more accurate and contextually relevant problem-solving.

In summary, Knowledge-based Systems are integral to improving decision-making processes across various fields by leveraging AI to harness and apply expert knowledge efficiently. While each type of KBS has its unique strengths, all contribute to making complex information more accessible and actionable in their respective domains.

2.8.3) Expert System

Expert systems are sophisticated computer programs designed to replicate human expertise within specific domains, thereby enhancing problem-solving and decision-making processes (Salmanova, 2024). In the AEC industry, these systems are instrumental in improving efficiency and accuracy across a range of applications, including maintenance, rehabilitation, and energy efficiency.

Expert systems operate by utilizing a knowledge base and an inference engine to generate solutions and recommendations in specialized fields. These systems are particularly effective in addressing complex technical and economic challenges, as they offer consistent and reliable outputs that aid in decision-making (Salmanova, 2024; Węglarz & Gilewski, 2017). The structured approach of expert systems allows them to process large volumes of data and provide insights that are crucial for tackling intricate problems.

In the AEC sector, expert systems are applied in various areas to enhance operational effectiveness. Expert systems such as ExFMRAP assist civil engineers in the evaluation and repair of pavements, thereby streamlining the decision-making process (Musa, 2023). By providing precise assessments and repair recommendations, these systems help in optimizing maintenance strategies and extending the lifespan of infrastructure (Musa, 2023).

Expert systems also play a vital role in improving energy efficiency within buildings. These systems are capable of analysing building technologies and suggesting optimal solutions that integrate both financial and environmental considerations. By doing so, they help in achieving energy savings and reducing the overall carbon footprint of buildings (Węglarz, 2018).

In the realm of knowledge management, expert systems contribute by capturing and retrieving expert knowledge dynamically. This functionality is particularly valuable in enhancing the assessment of design performance in real-time, thereby supporting continuous improvement in project outcomes (Mathot et al., 2016). The use of expert systems within a Q&A model allows for the automated provision of expert-level advice, enabling stakeholders to address compliance issues and make informed decisions with minimal delay.

This functionality aligns with the broader goals of KM by leveraging stored expertise to deliver real-time, accurate answers.

2.9) Evolution of Q&A Models

In the early 1960s, Q&A Systems were created. The central database or knowledge system of BASEBALL and LUNAR, two of the first Q&A systems, allowed them to succeed. BASEBALL, which was created to provide answers to inquiries on the US baseball league for a whole year, handled inquiries like “where did each club play on July 7?” with ease. The geological examination of rocks recovered by the Apollo lunar mission was covered by LUNAR. These previous systems emphasized “closed domains”, in which all questions must pertain to the particular area or have a restricted vocabulary (Jessica Taylor, 2019). The knowledge that both BASEBALL and LUNAR systems possessed was a major factor in their respective successes. Over the following two decades, open domain systems with an emphasis on information retrieval were developed (Jessica Taylor, 2019). The development of Q&A models-based computer programs that try to provide users with answers to their inquiries, typically in natural language, began in the late 1990s with the introduction of internet services like Ask Jeeves and Google Answers. These services depended on volunteers or human specialists to respond to inquiries from users, frequently in exchange for payment or other benefits (“Q&A Software,” 2023).

Nowadays Q&A models are Natural Language Processing (NLP) systems designed to respond in natural language to questions from people. Open-domain and closed-domain are the two primary categories into which they can

be divided. Open-domain Q&A models can answer inquiries on any topic, using general knowledge sources such as the web or Wikipedia. Closed-domain Q&A models can use domain-specific knowledge sources, such as databases, papers, or ontologies, to provide answers to questions on certain subjects.

Later on, Q&A models developed into increasingly automated and complex systems that analysed questions and produced answers from a variety of sources, including databases, documents, and the internet, using methods like machine learning, information retrieval, natural language processing, and knowledge representation. These models include Yahoo! Answers, Quora, and Stack Exchange, where users can post questions and receive answers on a variety of subjects and domains (“Q&A Software,” 2023). Moreover, there are several sorts of Q&A models based on the structure of the questions and the sources of the responses. Factoid Q and A models, for instance, respond to queries like “What is the capital of France?” and “Who is the creator of Harry Potter?” that have a single factual response. In addition, non-factoid Q&A models address topics like “Why is the sky blue?” and “What is the best method to learn a new language?” that call for more in-depth analysis, justification, or opinion. In addition, there are closed-domain Q and A models that are focused on a particular domain or topic, like sports, history, or medicine, and open-domain Q and A models that can respond to queries from any domain or topic (Bhatia, 2023).

Many researchers suggest that Q&A models have made significant progress in recent years, especially with the advances in deep learning and natural language processing. However, there are still many challenges and opportunities in the field, such as data quality and availability, domain

knowledge and complexity, user needs and preferences, and technology integration and innovation.

2.10) Foundations of Language Understanding

Understanding the foundations of language processing, particularly in NLP, is essential for the development of advanced AI models in the AEC industry. Word embeddings are a powerful tool in NLP that can be used to represent words as vectors of real numbers. These vectors capture the semantic and syntactic meaning of words, allowing ML models to perform various NLP tasks such as sentiment analysis, machine translation, and question answering. (OpenAI, 2024) defines embedding as a list of floating-point numbers that represents a vector. The relatedness between two vectors is measured by their distance. When the distance between two vectors is small, it indicates high relatedness, while large distances suggest low relatedness.

The quest for a universal representation of text is central to the automated processing of natural languages. The development of pretrained text embeddings such as word2vec or GloVe has been a significant breakthrough in this area. While supervised models have consistently outperformed unsupervised models over the years, models based on unsupervised learning have become more prevalent in recent years. These models do not require specially labelled datasets but can learn from existing or automatically generated large corpora of texts, taking full advantage of deep learning (Koroteev MV, 2021; Ma et al., 2019). In Q&A models, word embeddings can be used to represent the question and the candidate answers, allowing the model to identify the most relevant answer. For example, in a typical Q&A

model, the input question is first converted into a vector using a word embedding model such as Word2Vec or GloVe. The candidate answers are then also converted into vectors using the same word embedding model.

Each word embedding algorithm has its own set of advantages and disadvantages, including early algorithms such as Word2Vec. As an illustration, Word2Vec is a well-known approach to word embedding generation that predicts the context of a word by analysing its adjacent words via neural network (Mikolov et al., 2013). Although these models perform well in NLP tasks, they ignore word order and lose some of the text's semantic information (Chu et al., 2023).

2.11) Methodological Advances in ML for Q&A Models

The application of ML methodologies in developing Q&A models for the AEC industry has significantly augmented project efficiency and efficacy. Key methodologies encompass:

- 1) Natural Language Processing (NLP):** Employed for the extraction and analysis of information from textual data, with applications demonstrated in automated construction specification reviews (Moon et al., 2022).
- 2) Feature Selection and Extraction:** This critical step in ML involves identifying and extracting salient features from data, as evidenced by (Nematzadeh et al., 2019) work on filter feature selection.

3) Model Training and Evaluation:

- The process involves training ML models on datasets to discern patterns, followed by evaluation to ascertain performance, as shown by (H. Guo et al., 2023).
- Post-training, models are optimized to enhance performance. (H. Guo et al., 2023) utilized genetic algorithms for this purpose in predicting construction project durations.

4) Model Deployment and Application: Many researchers, in their model deployment process, collected data from various sources, including BIM models, building regulations, and user queries. This data was processed using NLP techniques to extract relevant information, which was then stored in a knowledge base that integrated a relational database with an ontology for semantic representation (Lin et al., 2022). A bidirectional encoder, specifically a Bidirectional Encoder Representations from Transformers (BERT) model, was employed to handle user queries by matching them with relevant data from the knowledge base and providing natural language answers (Lin et al., 2022). This method aimed to improve conversational AI systems within the AEC industry, enhancing the accuracy and efficiency of information retrieval. Recent research in the AEC sector has further supported the integration of BERT models for query answering and information retrieval, highlighting the effectiveness of AI-driven systems in BIM-AIOT applications (Lin et al., 2022; N. Wang et al., 2022b).

2.12) Advancements and Applications of Large Language Models (LLMs)

In recent years, LLMs have witnessed significant advancements, with organizations such as OpenAI, Google, Meta, and Anthropic developing unique models tailored for diverse applications. Each LLM type brings distinct architectural features and capabilities, addressing various needs from general NLP tasks to more specialized, domain-specific applications.

1) OpenAI's GPT Series (Generative Pre-trained Transformer):

OpenAI's GPT series, most notably GPT-4 and GPT-4o, represent some of the most influential LLMs in the field. These models excel in generating coherent, human-like text and have been widely adopted across industries. Recently, OpenAI introduced Custom Generative Pre-trained Transformers (Custom GPTs), enabling users to fine-tune models for specific tasks by integrating proprietary datasets and APIs. This allows for high precision in specialized domains such as legal and medical industries. OpenAI has also launched the GPT Store, providing a platform for users to share and monetize their Custom GPT creations (OpenAI, 2024).

2) Google's Gemini: The Gemini series is Google's latest innovation in LLMs, succeeding earlier models such as BERT, T5, and PaLM. Unlike its predecessors, Gemini is a multimodal model, capable of processing not only text but also images, audio, and video, which broadens its applicability across various industries. These models are employed in systems like Bard and Duet AI, enhancing capabilities in reasoning,

complex problem-solving, and long-context comprehension Google Gemini (Google, 2024b). In addition, Gemma, a more lightweight and accessible version of Gemini, is designed for developers seeking customizable, smaller-scale models (Google, 2024c).

- 3) Meta's LLaMA (Large Language Model Meta AI):** Meta's LLaMA models are developed with efficiency in mind, offering high performance with smaller parameter sizes compared to larger counterparts like GPT-4. As an open-source model, LLaMA is particularly valuable for research and development purposes, allowing broader access to powerful LLMs without the need for significant computational resources. This efficiency makes LLaMA an ideal choice for resource-constrained environments while maintaining competitive performance across diverse tasks (AI at Meta, 2024).
- 4) Anthropic's Claude:** Claude models, developed by Anthropic, prioritize safety and alignment. Built with constitutional AI principles, these models are designed to generate controlled, reliable outputs, especially in contexts where precision and ethical considerations are paramount, such as legal assistance and customer support. Claude has gained recognition for its focus on mitigating risks such as bias and ensuring that outputs are interpretable and aligned with user expectations (Anthropic, 2024).

These LLMs are transforming various sectors, offering solutions that range from general-purpose conversational agents to highly specialized tools designed to meet the complex requirements of specific industries such as healthcare, legal services, engineering. Associated features with above LLMs

can support the development of specialized Q&A to support stakeholders in the AEC industry. By fine-tuning the right LLM with domain-specific data and incorporating iterative reasoning techniques, these models can provide instant and accurate answers to complex engineering queries.

2.13) A Comparative Analysis of Embeddings

Recent research has explored the use of more advanced word embedding models such as BERT and Sentence-BERT (SBERT) in Q&A models. In October 2018, the introduction of a new pretrained BERT text embedding model was the centre piece of the field of natural language processing. This model enables unprecedented precision results in many automated word processing tasks and is likely to replace the widely known word2vec model in prevalence, becoming the industry standard. Almost all scientific articles devoted to the problem of word processing in natural languages in 2019 were a reaction to the release of this new model, and the authors of this model have become some of the most cited researchers in the field of machine learning (Koroteev MV, 2021; Ma et al., 2019).

SBERT was first introduced in 2019. It is a language model that was designed to create sentence embeddings, which are vectors of numbers that represent the meaning of a sentence. SBERT uses the BERT model as its foundation, but it has been fine-tuned to be more effective at generating sentence embeddings. SBERT is a method for obtaining sentence embeddings that capture their semantic meaning. It is based on BERT, a pre-trained language model that can encode words in different contexts. Unlike BERT, which requires both sentences to be input into the network to measure their

similarity, SBERT uses a Siamese network structure to input different sentences into two BERT models and then averages the word vectors output by BERT to obtain fixed-length sentence vectors. This way, SBERT can reduce the computational cost and overhead of BERT and also make it suitable for semantic search and clustering tasks. SBERT was developed to improve BERT's performance on sentence-level tasks, such as clustering and semantic search (Chu et al., 2023). BERT and SBERT models have been shown to outperform traditional word embedding models on various benchmark datasets. For example, BERT-based models have achieved state-of-the-art performance on the Stanford Question Answering Dataset (SQuAD). SBERT has also been used to generate semantically meaningful sentence embeddings that can be compared using cosine-similarity, allowing for more accurate matching of questions and answers (Fernandez & Ackerson, 2023; Turing, 2023; Wilson, 2023).

The main difference between SBERT and BERT is that SBERT uses a Siamese network structure to encode sentences independently, while BERT uses a single network to encode sentences jointly. Another difference is that SBERT uses average pooling to obtain sentence embeddings from word vectors, while BERT uses the first token as the sentence embedding. This means that SBERT can capture the overall semantic information of the sentence, while BERT can only capture the information of the first token (Chu et al., 2023). SBERT has been shown to be more accurate than other sentence embedding models on a variety of tasks, including sentiment analysis, question answering, and natural language generation. It is also faster and more efficient than these models, making it a good choice for real-world applications.

OpenAI's embedding models are trained on large-scale text corpora to produce vector representations of texts that capture their semantic and syntactic features. The embeddings can be used for various tasks such as similarity search, clustering, classification, etc. (OpenAI, 2023a). OpenAI's fine-tuning service allows users to customize the behaviour of the original by training them on a new dataset. The fine-tuning process adapts the model's weights to the specific domain and task of the user. Fine-tuning can be used to make a model more focused, accurate, consistent, or creative. Users can fine-tune their models using OpenAI's Python library or web interface (API platform & OpenAI, 2023). OpenAI's embedding models are available through the OpenAI API. The API provides a variety of tools and resources for generating and manipulating embeddings. It also includes a number of pre-trained embeddings that can be used for a variety of applications.

The comparisons between BERT, SBERT and OpenAI Embeddings are depicted in Table (2.1) - (Devlin et al., 2019; Edwards, 2023; OpenAI, 2024):

Table (2.1): Comparisons between BERT, SBERT and OpenAI embeddings,
(Source: (Devlin et al., 2019; Edwards, 2023; OpenAI, 2024))

Model	Created in	Accuracy	Data Required	Pros	Cons	Developer
BERT	2018	High	Large amount of data	State-of-the-art performance on many NLP tasks	Computationally expensive	Google researchers
SBERT	2019	High	Small amount of data	Popular choice for semantic search and information retrieval tasks	Popular choice for semantic search and information retrieval tasks	Nils Reimers
OpenAI Embeddings	2020	High	Small amount of data	State-of-the-art in dense text embeddings	State-of-the-art in dense text embeddings	OpenAI

2.13.1) Comparative Studies of LLMs (ChatGPT and Gemini) in Practice

- 1) **Chat GPT:** ChatGPT is an AI language model created by OpenAI in 2022. It is based on neural network architecture and can process and generate responses for any sequence of characters that make sense, including different spoken languages, programming languages, and mathematical equations. ChatGPT is trained using a large dataset and fine-tuning techniques, which contribute to its impressive language generation capabilities. It can understand and respond to a wide range of user queries and prompts, making it versatile in conversational interactions. ChatGPT has the ability to handle complex and nuanced conversations, providing detailed and informative responses (I. Ahmed et al., 2023; Cretu, 2023).

ChatGPT-4 is a newer and improved version of ChatGPT, it was introduced in March 2023 with more advanced language generation capabilities and a larger number of parameters, it has the ability to generate high-quality and coherent responses to human queries. Previous research has demonstrated that ChatGPT-4 exhibits impressive generation capabilities when compared to other existing models. However, there has been limited quantitative analysis of Chat GPT-4's ability to understand the content it generates. Chat GPT-4 has limitations in handling paraphrase and similarity tasks but excels in inference tasks and outperforms all BERT models by a significant margin. Chat GPT-4 also performs comparably to BERT on sentiment analysis and question-answering tasks, indicating its competence in these domains (Jacobi, 2024; Qihuang et al., 2023).

2) Gemini (formerly known as Bard): Google's AI chatbot Bard underwent a significant transformation and was renamed to Gemini. This change occurred on February 7, 2024. Gemini is a conversational generative artificial intelligence chatbot developed by Google. Initially based on the LaMDA family of LLMs, it was later upgraded to PaLM and then to Bard. Bard was developed as a direct response to the meteoric rise of OpenAI's ChatGPT and was released in a limited capacity in March 2023 to lukewarm responses before expanding to other countries in May. Bard is an artificially intelligent LLM developed by Google AI. It is trained on a massive dataset of text and code, which allows it to generate human-quality text, translate languages, write different kinds of creative content, and answer questions in an informative way. Bard has the ability to handle complex and nuanced conversations, providing detailed and informative responses, it can understand and respond to a wide range of user queries and prompts, making it versatile in conversational interactions (I. Ahmed et al., 2023; Knight & Goode, 2023).

2.13.2) Comparing between ChatGPT and Gemini (Bard) Performance

A recent study by Ars Technica compared ChatGPT-4 and Google Bard in seven categories: dad jokes, argument dialog, mathematical word problems, summarization, factual retrieval, creative writing, and coding. The study found that ChatGPT-4 outperformed Google Bard on inference tasks, while Google Bard was better at handling back-and-forth voice dialogue on mobile devices (Edwards, 2023). Another article by Techopedia compared the two

chatbots in a head-to-head output test and explored their major differences (Keary, 2024). Zapier also compared the two models and found that Bard is built for research, while ChatGPT-4 is the better writer (Lau, 2023). While these studies provide some insight into the capabilities of each model, it's important to note that the specific strengths and weaknesses of each model will depend on the use case and the specific requirements of the bot. Both ChatGPT-4 and Bard can be used to create custom chatbots using their APIs, but the specific capabilities of each model will depend on the use case and the specific requirements of the bot (Edwards, 2023; Keary, 2024).

(López Espejel et al., 2023) examine the performance of GPT-3.5, GPT-4, and BARD models, by performing a thorough technical evaluation on different reasoning tasks across eleven distinct datasets. The paper provides empirical evidence showcasing the superior performance of ChatGPT-4 in comparison to both ChatGPT-3.5 and BARD in zero-shot setting throughout almost all evaluated tasks. While the superiority of GPT-4 compared to GPT-3.5 might be explained by its larger size and NLP efficiency, this was not evident for BARD. The study also demonstrate that the three models show limited proficiency in Inductive, Mathematical, and Multi-hop Reasoning Tasks.

ChatGPT4 and BARD have been compared in terms of their finetuning for a Q&A model. The understanding ability of ChatGPT4 was evaluated on the GLUE benchmark and compared with four representative BERT-style models. It was found that ChatGPT4 falls short in handling paraphrase and similarity tasks but outperforms all BERT models on inference tasks. Additionally, ChatGPT4 achieves comparable performance to BERT on sentiment analysis and question-answering tasks. By combining advanced

prompting strategies, the understanding ability of ChatGPT4 can be further improved. A comprehensive survey comparing the capabilities and features of ChatGPT and BARD positions ChatGPT as a leading model with exceptional performance (Qihuang et al., 2023). In a study evaluating the ability of LLMs to answer anaesthesia-related queries, ChatGPT was found to have lengthier, intellectual, and effective responses compared to Bard. However, Bard responses were more conversational and easier to read. Computational sentiment analysis showed that Bard had significantly greater polarity scores than ChatGPT (I. Ahmed et al., 2023).

Most recent papers suggest that ChatGPT-4 is a more capable LLM than BARD in terms of generating high-quality and coherent responses to human queries, but more research is needed to compare their performance on other NLP tasks and domains. (López Espejel et al., 2023) examine the performance of GPT-3.5, GPT-4, and BARD models, by performing a thorough technical evaluation on different reasoning tasks across eleven distinct datasets. (López Espejel et al., 2023) provide empirical evidence showcasing the superior performance of ChatGPT-4 in comparison to both ChatGPT-3.5 and BARD in zero-shot setting throughout almost all evaluated tasks. While the superiority of GPT-4 compared to GPT-3.5 might be explained by its larger size and NLP efficiency, this was not evident for BARD. The study also demonstrate that the three models show limited proficiency in Inductive, Mathematical, and Multi-hop Reasoning Tasks.

Other comparative evaluations of LLMs, particularly ChatGPT and Gemini, highlighted notable differences in their performance across a range of domain-specific tasks. In terms of factual accuracy, Gemini has demonstrated a slight advantage. For instance, in a healthcare-related

evaluation focused on dispelling myths about palliative care, Gemini achieved 100% accuracy, while ChatGPT closely followed with a 93.3% accuracy rate (J. O. Ahmed et al., 2025). When evaluated using a research-focused multiple-choice question set, ChatGPT scored 76%, marginally outperforming both Gemini and Perplexity, which each scored 72% (J. O. Ahmed et al., 2025).

Regarding conversational fluency, ChatGPT is often recognised for its originality in word generation, enhancing the perceived creativity of its responses. In contrast, Gemini has shown greater contextual consistency and adaptability, supported by more advanced learning strategies (Alsajri et al., 2024). This suggests that while ChatGPT may offer more engaging dialogue in open-ended settings, Gemini may better handle complex user intents and context retention.

In terms of practical application, both models have been successfully deployed across diverse domains, including education, healthcare, and customer service, reflecting their flexibility and relevance to real-world use cases (Rane et al., 2024). Moreover, their capabilities in managing complex linguistic phenomena underscore their value in interactive systems, with Gemini often demonstrating a slight edge in maintaining contextual coherence across longer exchanges.

Despite these advances, both models exhibit performance limitations when applied to highly specialised domains, underscoring the need for finetuning, ongoing refinement and domain-specific adaptation to enhance their reliability and precision.

2.13.3) Human-AI Feedback Loops in Data Generation

The use of LLMs for generating synthetic training data has gained significant traction, especially in specialized domains like engineering regulations where high-quality datasets are scarce. This research focuses on the role of Human-AI collaboration to overcome these challenges by leveraging the unique strengths of both AI-driven automation and human expertise. Recent studies demonstrate that involving human experts in the data generation process, known as Human-in-the-Loop (HITL) approaches, leads to enhanced data quality and performance, particularly for complex tasks that require deep domain understanding (Hemmer et al., 2022).

Human-AI feedback loops are crucial for refining AI outputs, particularly in applications requiring high accuracy. Recent studies highlight the significance of HITL methodologies. For example, active learning within HITL frameworks enhanced psychiatric models with less training data (Chandler et al., 2022). In the domain of text generation, generation negotiation interface allowed users to iteratively refine AI outputs, ensuring alignment with human standards (Strobelt et al., 2022). However, effective collaboration requires fair recognition of human contributions, as highlighted in (Tubaro, 2022), to maintain ethical and efficient systems.

2.13.4) Future Trajectory of ML-Based Q&A Models

The current GPT models have limitations in processing large and complex text documents, requiring careful prompt organization and fine-tuning (Liu et al., n.d.). However, looking forward, ML-based Q&A models are poised to

significantly impact AEC project management. By harnessing the capabilities of NLP and ML, these models provide extensive decision support across various AEC facets, from planning to safety and productivity. The continued evolution of ML technologies promises even more groundbreaking Q&A applications in the AEC industry.

2.14) Finetuning LLMs for Q&A Models

LLMs generate answers to questions, but their effectiveness is sometimes compromised by subpar quality responses and occasional inaccuracies. To tackle these issues, a fine-tuning process is employed, incorporating feedback and examples to improve the models in many research. In their research, (Shao et al., 2023) focus on enhancing AI models through continuous feedback loops, leveraging LLM evaluation metrics.

There are three prevalent learning models that demonstrate the adaptability of LLMs are zero-shot learning, which enables base models to respond broadly to different requests without specific training; few-shot learning, which enhances performance in a specific area by providing a small set of relevant examples; and fine-tuning, which allows data scientists to train the base model with additional data customized for the specific application, refining its parameters for optimal performance across various tasks. LLMs are powerful tools that may be customized to excel in specific areas and successfully suit diverse objectives due to their versatility (L. Zhang et al., 2024).

2.14.1) Fine-Tuning Techniques for Q&A Model Generation

There are many techniques currently used to improve the performance and accuracy of LLMs in extracting information and answering questions related to a certain topic. Supervised fine-tuning, unsupervised fine-tuning, instruction fine-tuning, prompt tuning, PEFT, adapter modules, and retrieval augmented generation have been compared in various research papers. One study found that for Natural Language Understanding (NLU) tasks, the parameter efficiency of adapter modules does not result in efficiency gains compared to full fine-tuning of models. Adapters are relatively expensive to train and have slightly higher deployment latency. It was recommended that practitioners rely on full fine-tuning or multi-task training for moderately sized models for NLU tasks (Mundra et al., 2023; Zhu et al., 2023). Another research paper focused on prompt-based fine-tuning and found that conventional continued pre-training does not consistently provide benefits. They proposed a prompt-based continued pre-training approach that combines task-related texts and prompt templates to improve the performance of prompt-based fine-tuning. This approach consistently improved the performance of state-of-the-art prompt-based fine-tuning approaches in both semi-supervised and fully supervised settings (Mundra et al., 2023).

Transfer learning from retrieval models and retrieval augmented generation both aim to improve the performance of large language models (LLMs) on downstream tasks. Transfer learning from retrieval models involves using external knowledge to ground model generation, while retrieval augmented generation actively retrieves relevant information during the generation process. In terms of performance, both approaches have shown strong results. Iter-RetGen, a retrieval augmented generation method, has

demonstrated superior or competitive performance on multi-hop question answering, fact verification, and commonsense reasoning tasks. Similarly, FLARE, another retrieval augmented generation method, has achieved superior or competitive performance on long-form knowledge-intensive generation tasks. These findings suggest that both transfer learning from retrieval models and retrieval augmented generation can effectively enhance the performance of LLMs on various downstream tasks (Shao et al., 2023).

A comparison in terms of fine-tuning techniques is carried out to understand the differences between the most up- to to-date techniques as detailed below:

1) Common Techniques

- **Supervised Fine-Tuning:** Supervised Fine-Tuning (SFT) plays an essential role ML by adapting pre-trained models to perform specific tasks using labelled datasets. By refining model parameters, SFT enhances a model's ability to deliver targeted and accurate results for specialized applications. SFT adjusts pre-trained models by incorporating labelled data, enabling them to retain prior knowledge while adapting to new tasks. This process ensures higher task-specific performance (Zhou et al., 2023).

The success of SFT depends heavily on the quality of data used. Recent research introduces the concept of “learnability”, which focuses on how well the fine-tuning data aligns with the model’s capabilities, ensuring efficient learning and adaptation (Zhou et al., 2023). Recent developments include integrating supervised contrastive learning, which helps improve a model's robustness and generalization by

identifying class similarities and differences. This approach enhances task performance without needing significant changes to the model architecture (Gunel et al., 2021a). Despite these advantages, some researchers warn of potential overfitting, especially when the dataset used for fine-tuning is too small or lacks diversity. Balancing data quality and quantity is therefore crucial to prevent overfitting and ensure reliable outcomes

- **Instruction Fine-Tuning:** Instruction Fine-Tuning (IFT) is a crucial method for improving LLMs to better interpret and execute complex instructions. Recent progress in IFT methodologies has targeted enhancing model performance across various languages and tasks. IFT datasets have traditionally been English-focused, which restricts performance in other languages. A new approach has introduced multilingual datasets that maintain linguistic authenticity, significantly improving non-English performance (Indurthi et al., 2024).

Phased IFT employs a gradual learning strategy by organizing instruction data based on complexity. This approach has demonstrated superior performance compared to conventional one-time training methods, improving the model's ability to follow instructions and overall performance (Pang et al., 2024). Incorporating coding data into IFT enhances reasoning abilities across multiple models. The proportion of coding data necessary varies by task, suggesting that careful consideration of dataset composition is key (B. Zhang et al., 2024).

According to (Gupta et al., 2023) instruction tuning can improve how well language models perform on new tasks, even when given only a few examples. This is because instruction tuning uses in-context learning. In the real world, there is often not a lot of data available for fine-tuning models. Instruction tuned models are more efficient at learning from data than standard models. This means they require less training data to perform just as well as other models. However, instruction tuning may not be effective for all tasks, such as question rewriting and title generation.

Despite significant advances, IFT still faces challenges in ensuring broad generalization and performance across diverse tasks and languages. Ongoing research will likely focus on further refining these techniques to overcome these limitations.

- **Unsupervised Fine-Tuning:**

Unsupervised fine-tuning has become an important method in machine learning, particularly for situations where labelled data is either limited or expensive to obtain. By utilizing unannotated data, it enables models to improve performance across a range of domains, from NLP to computer vision. Researchers experimented many unsupervised fine-tuning techniques, including Unsupervised feature transformation learning merges graph, contrastive, and generative learning techniques to develop new feature sets without relying on labelled data. This method successfully captures complex feature interactions, offering an alternative to traditional supervised approaches (Ying et al., 2024). Another technique is Self-Evolution Fine-Tuning (SEFT), SEFT leverages unannotated data to refine policy models, improving

response quality without needing vast labelled datasets. This method demonstrates enhanced stability and efficiency, positioning it as a viable alternative to supervised fine-tuning (R. Chen et al., 2024).

Although unsupervised fine-tuning presents substantial advantages, it may still fall short of the performance levels achieved by supervised methods, particularly in specialized tasks where labelled data is crucial for fine-tuning insights.

2) Advanced Techniques

- **Prompt Tuning:** Prompt tuning is an efficient method for fine-tuning LLMs by optimizing the input prompts instead of adjusting the model's internal weights. This technique allows for task-specific adaptations while preserving the model's generality, significantly reducing the computational cost of full-scale fine-tuning. Research has demonstrated its effectiveness, particularly in few-shot learning and cross-domain transfer tasks. For instance, (Lester et al., 2021) highlighted the scalability and robustness of prompt tuning in T5 models. Researchers also investigated the impact of various factors on LLM fine-tuning. These factors included the model architecture, training data, and methodology. Notably, (B. Zhang et al., 2024) discover that the size of the LLM model significantly influences fine-tuning performance, even more so than the amount of pretraining data. Additionally, their analysis revealed that scaling the prompt engineering template parameters generally does not yield substantial improvements in LLM fine-tuning outcomes. These insights provide valuable guidance for optimizing LLM-based applications and tasks.

These findings underscore prompt tuning's appeal as an alternative to traditional fine-tuning, especially when computational resources are limited.

▪ **Parameter-Efficient Fine-Tuning (PEFT) and Adapter Modules:**

Parameter-Efficient Fine-Tuning (PEFT) has emerged as a pivotal strategy for adapting LLMs such as GPT-3 and T5, significantly reducing the computational demands typically associated with fine-tuning. PEFT targets only a portion of the model's parameters, improving scalability and speed. A key technique within this framework is the use of adapter modules, which introduce lightweight layers into the model. These adapters are trained on task-specific data while the rest of the model remains frozen, preventing the loss of pre-trained knowledge. (Houlsby et al., 2019) demonstrated that adapter modules can achieve performance comparable to full-model fine-tuning with considerably fewer parameter updates. Building on this, (Pfeiffer et al., 2020) showed that stacking and combining adapter modules enables efficient multi-task learning, further advancing the flexibility of PEFT. These developments position PEFT, especially through adapter modules, as an essential method for efficiently fine-tuning LLMs in environments with limited resources.

3) Emerging Fine-tuning Techniques

- **Transfer Learning from Retrieval Models:** This involves Transfer learning from retrieval models has become a significant development in machine learning, particularly in NLP and information retrieval. This

technique enables models to utilize knowledge from large pre-trained models to improve performance, especially when data is limited.

Knowledge Transfer Mechanisms Retrieval-based Knowledge Transfer (RetriKT) draws knowledge from large LLMs into a knowledge store, allowing smaller models to retrieve relevant information, significantly boosting low-resource task performance (Liu et al., 2023). Graph neural networks enhance similarity-based retrieval in case-based reasoning by pretraining models on graph similarities, improving efficiency (Pauli et al., 2023).

Applications in Cross-Language Retrieval, the ColBERT-X model, showcases the success of transfer learning in cross-language dense retrieval, improving multilingual tasks using pre-trained encoders (Nair et al., 2022). While transfer learning has proven highly effective, challenges persist in applying this approach consistently across diverse domains, requiring further research.

- **Retrieval Augmented Generation (RAG):** Retrieval Augmented Generation (RAG) is an advanced framework designed to improve LLMs by incorporating external knowledge, enhancing both accuracy and relevance in generated responses. Recent developments in RAG have focused on modular enhancements, user-specific adaptations, and efficient processing methods. Modular Enhancements RAG systems now include features like the Query Rewriter and Knowledge Filter, which refine retrieval and address ambiguities (Shi et al., 2024). Additionally, Speculative RAG uses multiple LLMs to enhance both

accuracy and speed (Z. Wang et al., 2024). Figure (2.2) illustrates retrieval augmented generation with a knowledge graph.

User-Centric Adaptations, PersonaRAG, personalizes responses by integrating user data, tailoring the retrieval process to individual needs (Zerhoudi & Granitzer, 2024). Efficiency Improvements Techniques optimize efficiency by compressing input contexts, reducing processing time without compromising performance (Rau et al., 2024).

Despite these innovations, challenges remain, particularly in ensuring the reliability of retrieved information and managing the complexity of diverse knowledge integration.

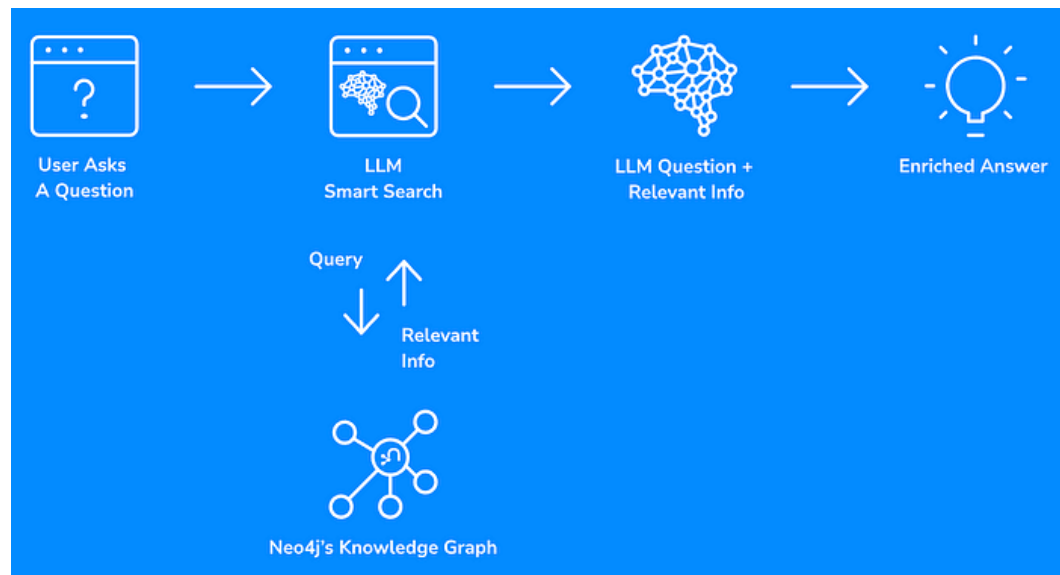


Figure (2.2): Retrieval augmented generation with a knowledge graph (Source: (Bukowski, 2023))

4) Other Techniques

- **Self-training models:** Self-training models have gained traction as an effective method for improving the performance of language and vision-language models by utilizing unlabelled data. This approach

allows models to autonomously generate labels, enhancing their capabilities without the need for extensive human annotation.

Self-training enhances arithmetic reasoning in language models through automated feedback, which validates predictions and improves performance both online and offline (Kadlčík & Štefánik, 2024). Additionally, LLMs can autonomously generate labels, reduce hallucinations and addressing catastrophic forgetting in out-of-distribution scenarios (Yeo et al., 2024).

The self-training on image comprehension framework allows vision-language models to generate image descriptions independently, leading to significant performance improvements while using fewer labelled datasets (Deng et al., 2024). In text classification, self-training enhances active learning efficiency by creating pseudo-labels from unlabelled data, facilitating model training with minimal labelled samples (Schröder et al., 2024).

Despite its advantages, self-training faces challenges in ensuring the quality of self-generated labels and balancing the use of labelled and unlabelled data. Further research is needed to refine these techniques and explore their application in diverse fields.

- **Advancing Large Language Models with Self-Updates and Logic Training:** Self-training models have emerged as a robust method to enhance machine learning performance, especially in contexts with limited labelled data. By leveraging automated feedback and pseudo-labels, these models reduce the need for extensive human-annotated datasets. In LLMs, self-training has proven effective, as models

autonomously generate labels, leading to reduced hallucinations and better performance on out-of-distribution tasks (Yeo et al., 2024). This approach also enhances arithmetic reasoning through online self-training without the need for new data (Kadlčík & Štefánek, 2024).

Self-training is particularly useful in handwritten text recognition, where models train on synthetic data, outperforming traditional methods (Wolf & Fink, 2024). Self-modelling aids in self-training by allowing models to predict internal states, leading to enhanced adaptability and performance (Premakumar et al., 2024).

Although self-training provides numerous advantages, there is a risk of propagating errors from pseudo-labels, which could affect model reliability. Ensuring a balance between self-training and validation is crucial for optimal results.

2.15) Conclusion

The Literature Review Chapter highlighted the challenges and opportunities involved in creating a Q&A model designed to assist AEC stakeholders in complying with building regulations. While BIM-based automated compliance checking has shown promise in other regions, its application is currently limited in the local context due to the widespread use of 2D designs. Therefore, the study emphasizes the need for alternative approaches that can leverage existing building regulation data. Various techniques have been explored, and the comparison shows that while some

approaches offer efficiency, they may lack the ability to handle the complexity of regulations in a dynamic environment.

Expert systems and knowledge management frameworks provide a promising path forward, particularly in preserving the knowledge of AEC experts, which is crucial given their limited availability in the building regulations domain. However, these systems face challenges in replicating human expertise and judgment in complex regulatory scenarios. Moving forward, the research identifies several gaps that future studies must address: improving the adaptability of the Q&A model, refining the knowledge capture process from domain experts, and enhancing the model's ability to handle updates in regulatory frameworks.

Overall, while the research offers valuable insights into current technologies and methodologies, further exploration is necessary to refine the Q&A model. In the next Chapter the methodologies used to create a custom Q&A model to optimize AEC stakeholders are discussed in detail.

Chapter Three

Methodology and Procedure

3.1) Introduction

The methodology chapter of this research outlines a comprehensive approach to developing and validating a Custom Generative Pre-trained Transformer (CustomGPT). The model is designed to optimize decision-making for stakeholders in the Architecture, Engineering, and Construction (AEC) industry. The methodological framework incorporates both quantitative and qualitative approaches to ensure a robust evaluation of Large Language Models (LLMs) and the practical applicability of the developed model in real-world scenarios.

The research begins with a quantitative phase, involving a systematic evaluation of various LLMs, including ChatGPT-4 and Gemini 1.5. This phase aims to identify the most suitable model for fine-tuning and further development into a comprehensive Questions and Answers (Q&A) system. The evaluation utilized a human- Artificial Intelligence (AI) generated dataset derived from Law No. (51) of 2014 (Engineering Law) that regulates the practice of engineering professions in the Kingdom of Bahrain, which regulates the practice of engineering professions (CRPEP, 2014a). Key performance metrics such as accuracy, precision, recall, and F1score were employed to assess the technical robustness of the models, beside human researcher evaluation (Jude Chukwura Obi, 2023; Riyanto et al., 2023; Silhavy & Silhavy, 2023). This comprehensive assessment ensured that the selected model was not only technically robust but also contextually appropriate for the specific regulatory environment.

The qualitative analysis then applied to validate the performance and utility of the Custom Generative Pre-trained Transformer (GPT) model. This

phase involved structured interviews with experts to gather critical insights and feedback on the model's outputs. The integration of expert's opinions was crucial for optimizing the model to accurately reflect regulatory requirements and support informed decision-making. The model was further validated through conduction of a real-world testing with AEC industry stakeholders. This step aimed to evaluate the efficiency, accuracy, and users' satisfaction associated with the final model.

The research tools employed included the LLMs themselves (ChatGPT-4, Gemini 1.5 and Gemma 2b), a CustomGPT model, structured interview questionnaires for expert feedback, and user feedback forms for real-world testing. The data collection process involved initial data exploration and preprocessing, human-AI collaboration to generate Q&A pairs, and iterative refinement based on expert feedback and real-world testing results. The research incorporated descriptive statistics to summarize the dataset, which including evaluation metrics such as accuracy, precision, recall, and F1 score for comparative analysis, and thematic analysis for qualitative feedback. These methods provide a comprehensive evaluation framework, ensuring both the technical and practical validity of the selected base model.

It worth mentioning that data compilation involved human-AI collaboration to expand the data required to train or customize the model (step one of the methodology). This approach was elaborated by this thesis researcher Al Nama et al. (2024), who leveraged human expertise and AI for engineering regulatory data expansion in their case study with ChatGPT (Al Nama et al., 2024). Furthermore, the evaluation of the Performance of LLMs; ChatGPT-4 and Gemini 1.5 (step two of the methodology), was elaborated by this thesis researcher Al Nama & Mahmud (2024), who analysed the

performance of ChatGPT-4 and Gemini by Questions Category in Engineering Regulations (Al Nama & Mahmud, 2024). Throughout this chapter the 10-step methodological approach is discussed

3.2) Research Design

The components of a research design encompass the various facets that necessitate consideration during the planning and execution of the research. Below are the details of dimensions of the research design. The details are further summarized in Table (3.1).

▪ Study Area

This research was set within the context of the AEC industry in the Kingdom of Bahrain, with a specific focus on regulatory compliance during the design stage of construction projects. The study was conducted in an environment where there is a pressing need to address the issue of violations to building code and relevant regulations. The setting involved collaboration with industry experts, from the regulatory body, and potential users from stakeholders to gather relevant data and insights.

▪ Unit of Analysis and Unit of Observation

The unit of analysis is the primary subject under study, while the unit of observation is what is being measured (Pressbooks, 2023). For this research, the units of analysis included the Unified Guidebook of Building Permit Regulations (Bahraini Building Code), Chapter (14) specifically. The units of observation were the feedback of the experts and potential users.

▪ **Sampling Design**

Sampling refers to the process of selecting a subset from a larger population (Cash et al., 2022). In this research, the sampling population comprised experts who serve as evaluators from the authority regulating engineering professions in the Kingdom of Bahrain, as well as users who are stakeholders from licensed architects and engineering governed by this authority.

▪ **Time Horizon**

This research adopted a cross-sectional time horizon, where data was collected at one specific point in time. The research was based on data and tools available up to June 2024.

▪ **Research Methodology and Data Collection**

A mixed-methods approach, combining qualitative and quantitative data, used. The research methodology included a quantitative phase, which systematically evaluates various LLMs, and a qualitative phase, which validated the custom GPT model through expert interviews and real-world testing with AEC industry stakeholders. This approach aimed to assess the model's efficiency, accuracy, and user satisfaction, providing insights into its practical advantages over conventional methods.

▪ **Variables**

Independent variables are predictors or treatments manipulated by researchers to observe effects (Flannelly et al., 2014). Dependent variables are the outcomes or responses measured in a study (Flannelly et al., 2014). Moderator variables alter the relationship between independent and dependent variables (Pokhariyal, 2019). The research variables can be defined as follows:

- a) The independent variables which include the building code, Chapter (14) specifically, and the maturity level of developers and experts.
- b) The dependent variables which is the model efficiency.
- c) The moderator variable which is the developed Q&A model for the purpose of this research.

▪ **Purpose of the Study**

This research is explanatory in nature, and one of its key objectives is to elucidate how a modeling framework can be developed and utilized to optimize decision-making in the construction industry, as further discussed in Chapter 1.

▪ **Validity and Reliability**

Reliability and validity are ensured by employing evaluation metrics like calculating accuracy, F1 score, precision and recall, along with experts' evaluation, and using a representative sample for users testing (Jude Chukwura Obi, 2023; Riyanto et al., 2023; Silhavy & Silhavy, 2023).

Table (3.1): Dimensions of research design

No.	Dimension	Details
1	Study area / setting	Non-contrived: the research (experiment) was conducted in the real world (AEC industry).
2	Unit of analysis, and unit of observation	Chapter (14) of the Bahraini building code, On the individual and organizational level (experts and users from stakeholders).
3	Sampling design	Stratified probability sampling involves the division of a population into distinct groups or strata; experts and users.
4	Time horizon	Cross-sectional. Data collection involved gathering information at a specific moment in time. Also, the choice of tools

No.	Dimension	Details
		depended on the available tools at a specific moment in time (available up to June 2024).
5	a) Research methodology b) Data collection method	Quantitative and qualitative; use of compiled data* to create a Q&A and then validate test the model efficiency through experts and potential users. *Compiled data: available documents, generated through human-AI collaboration, experts' knowledge, users feedback, etc.
7	Purpose of the study	Explanatory The research is trying to explain how a modelling framework can be developed and used to optimize decision-making in the construction industry.
8	Validity and Reliability	Accuracy, F1 score, precision and recall. Experts' evaluation and using a representative sample for users testing.

3.3) Qualitative and Quantitative Methodologies

Qualitative and quantitative analyses serve distinct yet complementary roles in research, each contributing to a fuller understanding of complex issues. Qualitative analysis focuses on capturing the depth and context of human experiences, often through methods such as interviews or focus groups and one on one meetings, allowing researchers to explore participants' perspectives and the meanings they attach to their experiences. This approach is particularly effective in revealing contextual factors and underlying themes that may not be evident through quantitative data, offering nuanced insights into social and behavioural phenomena (Halevi Hochwald et al., 2023; Wayessa & Nygren, 2023).

However, quantitative analysis emphasizes the measurement of variables and the identification of statistical relationships. By applying statistical tools, researchers can measure data systematically and identify patterns or trends that provide a broad overview of the phenomena under study. This approach is essential for generalizability, enabling researchers to make broader inferences about populations based on the data collected (Adu et al., 2022; Russell et al., 2022). Quantitative research is particularly useful in studies where precision, replicability, and statistical significance are critical.

The mixed-methods approach integrates the strengths of both qualitative and quantitative research, offering a more comprehensive understanding of research questions. By combining qualitative insights with quantitative data, researchers can validate findings and explore any discrepancies between the two methods. This approach is particularly valuable in complex fields where understanding both statistical trends and individual experiences is crucial for drawing meaningful conclusions (Adu et al., 2022; Tzagkarakis & Kritas, 2023; Wayessa & Nygren, 2023). Mixed methods allow researchers to address both the breadth and depth of a research problem, ensuring a more robust and nuanced interpretation of the data (Wayessa & Nygren, 2023). Mixed methods allow researchers to address both the breadth and depth of a research problem, ensuring a more robust and nuanced interpretation of the data.

In summary, this methodology chapter details a structured and rigorous approach to developing a Q&A model for regulatory compliance in the AEC industry. By integrating quantitative assessments with qualitative validations, the research aims to create a robust tool that enhances decision-making and compliance processes for AEC stakeholders.

3.4) Exploration Stage

Before delving into the specific methodologies employed, it is important to acknowledge the challenges faced during the initial exploration stage, which significantly impacted the research approach and shaped its direction. One of the primary challenges was constructing a Q&A model from a single, limited-size PDF document, primarily written in Arabic. This phase focused on navigating the complexities of working with a restricted dataset and the intricacies of the Arabic language, both of which influenced key decisions and required a collaborative approach. The insights gained from this initial exploration laid a solid foundation for the subsequent development and optimization of the research methodologies. The process unfolded as follows:

3.4.1) Language Considerations and Data Format Conversion

- **Addressing Arabic Text:** While Arabic was initially considered, the limited dataset and challenges in working with Arabic language processing libraries (e.g., “tnkeeh”) led to the decision to utilize the English version of the document. This enabled access to a wider range of resources and tools for data preparation and model training.
- **Conversion to CSV:** the PDF document was converted into a CSV format, which offered several advantages for subsequent analysis. The tabular organization facilitated easier manipulation and analysis of the data, enabling clearer insights to be drawn. Additionally, this format enhanced both the clarity and consistency of data presentation, ensuring that the information was well-structured and easier to interpret. Importantly, this transformation also made the dataset compatible with Machine Learning (ML) models.

3.4.2) Data Augmentation and Labelling

- **Generating Q&A Pairs:** To address the scarcity of examples, Q&A pairs were collaboratively generated with AI assistance and expert input. This integration of human expertise ensures that training data meets the required standards, making the model more robust (Leitão et al., 2022).
- **Incorporating into Original Dataset:** These newly created pairs were integrated into the initial dataset, significantly expanding the available training data.
- **Labelling:** Each question-and-answer pair was carefully labelled, providing essential guidance for model training.

3.4.3) Early Trials and Algorithm Selection

- **Supervised Algorithms:** Initial experimentation with supervised algorithms like KNN on the original dataset alone faced challenges due to its limited size, as evidenced by the formation of a single cluster, hindering evaluation.
- **Unsupervised Algorithms:** Attempts with unsupervised algorithms like DBSCAN and KMeans also yielded unsatisfactory results when applied to the dataset containing only engineering law articles, highlighting the need for diverse and labelled examples.

3.4.4) Collaborative Effort and Expert Input

- **Question and Answer Generation:** The generation of Q&A pairs involved a collaborative effort between AI and human expertise.
- **Expert Review and Refinement:** Expert input was crucial for ensuring accurate referencing of relevant engineering law articles, providing an essential layer of quality control.

3.4.5) Outcomes of the Exploration Stage

Creating a Q&A model for a specific domain, like the engineering industry, can indeed be challenging, especially with a small dataset. This initial phase focused on addressing the challenges of limited data, language considerations, and algorithm selection. By converting the PDF to CSV, generating and labelling additional Q&A pairs, and recognizing the limitations of early algorithm trials, a solid foundation has been established for further refining data preparation, finalizing model training code, and progressing towards a robust Q&A model.

The challenges encountered during this exploration laid the groundwork for the subsequent methodological refinements. With a clear understanding of the limitations and needs, the research approach evolved through the phases outlined in the development timeline.

3.5) Research Development Timeline

Building on the groundwork laid during the exploration stage, the research development timeline, as depicted in Figure (3.1), offers a detailed account of how the initial challenges were systematically addressed, refined, and overcome. The exploration phase, which focused on managing language complexities, limited datasets, and initial trials with ML algorithms, provided crucial insights that shaped the refinement process.

This development timeline captured the key phases in the evolution of the custom GPT model, designed to enhance decision-making within the AEC industry. It outlined the chronological progression of research activities, highlighting the methodological adjustments made to tackle obstacles and

integrate emerging technological advancements. Each phase was affected by literature review and contributed to the iterative refinement and validation of the model, ensuring its alignment with the goals of improving efficiency and compliance in the AEC field.

3.5.1) During 2022-2023

1) Observation of Building Code Violations

Numerous violations of the building code were observed across various projects, as further detailed in the literature review chapter. These violations significantly jeopardized the interests of AEC stakeholders, including clients, architects, engineers, contractors, and developers. The repercussions of such violations were multifaceted, affecting project timelines, financial investments, and overall quality. These violations often stemmed from non-compliance with the prescribed regulations, leading to project delays, increased costs, and compromised safety standards. The lack of adherence to the building code resulted in frequent disputes among stakeholders, contributing to a fragmented and inefficient project execution process. The need for a robust compliance mechanism became evident to mitigate these risks and ensure that all projects align with the established regulatory framework. The initial concept aimed to assist engineers and architects in performing self-compliance checks on their engineering design drawings utilizing Building Information Modelling (BIM).

2) Building Information Modelling Familiarity Studies

The studies revealed that most engineers and engineering firms lack familiarity with BIM. To implement this idea effectively, the use of BIM for

generating engineering design drawings should be mandated by law or regulation.

3) Transition Towards Q&A Model Development

Given the complexities of navigating building regulations and the need for a comprehensive compliance mechanism, the focus subsequently shifted to developing an advanced Q&A model. This model aims to enable architects, engineers, and other stakeholders to verify design compliance with building regulations. The building regulations comprise 17 chapters, each summarizing the regulations enforced by a specific organization such as the Electricity and Water Authority, Urban Planning Authority, Civil Defence, Road Directorate at the Ministry of Works, and the Council for Regulating the Practice of Engineering Professions, among others.

4) Navigational Challenges

Navigating the building code and relevant laws and regulations enforced by various regulatory bodies is challenging for architects and engineers. A system that allows them to resolve their queries before submitting engineering designs for building permits would significantly save time and resources.

5) Historical Study of Q&A Systems

The history of developing Q&A systems was studied, spanning from the 1960s to the present, when LLMs were introduced. This provided a foundational understanding of the evolution and advancements in Q&A systems.

6) Evaluation of Supervised and Unsupervised Approaches

Supervised and unsupervised approaches to creating Q&A models were explored but yielded suboptimal results, highlighting the need for more advanced methodologies.

7) Comparative Analysis of LLMs

A comparative analysis was conducted between different LLMs, primarily ChatGPT-4 by OpenAI and Gemini (formerly known as BARD) by Google, as these are the two leading LLM developers. This analysis aimed to identify the most suitable model for developing the custom Q&A system.

3.5.2) During January to May 2024

8) Transition to Advanced LLMs

In early 2024, BARD transitioned to Gemini 1.5, and ChatGPT-4 became accessible via subscription. This transition marked the beginning of advanced evaluations of these models.

9) Focus on Specific Chapter

Given the researcher's expertise in the engineering law and the need for creating and filtering data accurately, a decision was made to focus the research on the chapter relevant to regulating the practice of engineering professions within the building regulations.

10) Valuable Data Collection

The available data, though relatively limited, was highly valuable, primarily consisting of documents related to engineering law which contained 45 Articles in 11 Chapters.

11) Additional Data Sources

Additional available data included implementing regulations, other regulatory legislation, updates to these documents, circulars issued by the Council for Regulating the Practice of Engineering Professions (CRPEP), forms used for applying or renewing different engineering licenses, and other forms and data available on the CRPEP's website.

12) Human-AI Collaboration for Q&A Sets

In the absence of other data, the researcher developed sets of Q&A through Human-AI collaboration for training and testing purposes of the Q&A model. The most frequently asked question types (factual, hypothetical, comparative, theoretical, critical thinking) are selected to generate the Q&A set.

13) Evaluation of Gemini and ChatGPT 4

The generated Q&A sets were used to evaluate the performance of Gemini 1.5 and ChatGPT-4. The results indicated that ChatGPT-4 slightly outperformed Gemini 1.5. However, considering other criteria, features, accessibility, and ease of use, Gemini is deemed the more suitable option. Gemma-2b, a lighter-weight LLM based on Gemini by Google, is selected for fine-tuning using Google Colab.

14) Fine-Tuning of Gemma-2b

The generated Q&A set was further refined, and additional sets were created (around 250 Q&A per Article, totalling 11,250 – Available in Appendix (B)) to fine-tune Gemma-2b. Numerous challenges were encountered, including issues related to Graphics Processing Units (GPU) availability and accessible servers. The fine-tuned model yielded poor results

as the legal documents were not embedded in the model. The articles were incorporated in the form of Q&A, but the results remained unsatisfactory. LangChain, an advanced search framework, was utilized to create the Q&A model in Gemma and Colab, which also produce poor results.

15) Use of LoRA Keras Framework

Gemma was fine-tuned in Keras (API designed for human beings to build and deploy deep learning models) using Low-Rank Adaptation (LoRA) to create a Q&A model, which also produced poor results.

16) Development of Hybrid Model

Subsequently, a hybrid model, integrating Retrieval-Augmented Generation (RAG) with the fine-tuned model using LoRA, was developed, but it also exhibited some irrelevance in the answers.

3.5.3) During June 2024

17) Release of ChatGPT 4o

ChatGPT 4o was released by OpenAI, enabling the creation of custom GPTs. This advancement provided new opportunities for developing a more effective model.

18) Development of Custom GPT

A custom GPT related to the engineering law was developed, demonstrating promising results. However, it required additional information and data for further improvement. Other available data related to implementing regulations and other legislation were uploaded into the model, resulting in highly satisfactory answers.

19) Continued Development and Fine-Tuning

The model continued to be developed and further fine-tuned by the researcher, incorporating new data, and refining the model based on feedback.

20) Expert Feedback Incorporation

Structured interviews with experts from the CRPEP were conducted to gather feedback on the model's outputs. This feedback was integrated into the model to optimize its performance.

21) Real-World Testing and Feedback Collection

The model was tested in real-world scenarios by users from the AEC industry, primarily architects and engineers. Participants were introduced to using the custom GPT model, and their feedback was collected and analysed to assess the model's usability and effectiveness.

22) Integration of Updates

Continuous updates to the engineering law, relevant regulations, forms, and procedures were directly integrated into the model, ensuring it remains current and accurate.

23) Comprehensive Model Development

The model integrates the researcher's expertise, knowledge from other LLMs used to generate Q&A sets, other experts' knowledge, and updates to enforced laws and regulations. This iterative process was essential for developing an advanced model that enables architects and engineers to comply with engineering law. It also provides a roadmap and framework for creating models for other chapters within the building code to optimize AEC stakeholder decisions. Additionally, it is important to note that the number of experts embedding knowledge in each chapter of the building regulations is

limited. Developing and maintaining a Q&A model is essential for preserving this knowledge



Figure (3.1): Research development timeline

3.6) Overview of Methodological Steps

The development timeline offers a comprehensive overview of the research progression, providing broader context for the detailed methodological steps that follow. While some of these methodologies were briefly introduced in the timeline, they are explored in greater depth in the following sections, which delve into the specific approaches, techniques, and processes employed throughout the research. These steps build upon the foundations established earlier in the timeline, representing the practical implementation of decisions and refinements made at each stage to ensure the thorough and systematic development of the custom GPT model.

This chapter details each step involved in the model's development, covering the data used, the processes for gathering and generating this data, as well as the necessary cleansing and preparation steps. It also discusses the tools employed for both quantitative and qualitative methods, illustrating how these steps collectively form a roadmap for creating a robust model. This roadmap includes the shortlisting and evaluation of LLMs, the creation of required datasets from a limited initial dataset, the fine-tuning of models, and the subsequent evaluation of these fine-tuned models. Additionally, it explores the incorporation of expert knowledge by transforming tacit knowledge into explicit knowledge and engaging users to gather feedback on the model's efficacy in optimizing their decision-making processes.

Figure (3.1) outlines the methodological steps applied in this research.

3.6.1) Phase (1) Steps: Fine-tuning Pre-trained Models (LLMs)

1) Compile Data

Developing a comprehensive dataset of Q&A specifically related to engineering law. This dataset covers all listed articles and includes a variety of question types and complexities to ensure thorough coverage and robustness. The data compilation process involved leveraging available open data sources (law documents) and using this data to generate Q&A pairs through human-AI collaboration. This hybrid approach enriched the dataset with diverse perspectives and detailed information. The approach in this step utilizes explicit knowledge, combining human researcher (as an expert) insights and publicly available regulations.

2) Evaluate and Compare LLMs

Assessing the initial performance of cutting-edge language models, specifically ChatGPT-4 by OpenAI and Gemini by Google, to determine their suitability for fine-tuning. Both LLMs were evaluated by posing the same set of questions derived from the compiled dataset. Their responses were then analysed to gauge performance. The evaluation focuses on key metrics such as accuracy, precision, recall, and F1 score to provide a comprehensive assessment of each model's capabilities. This step establishes a baseline performance for each model and identifies areas needing improvement.

3) Set Selection Criteria

Establishing a robust selection framework to choose the best-performing model based on predefined evaluation metrics and practical considerations. A selection criteria matrix was created, incorporating evaluation metrics (accuracy, precision, recall, F1 score) and additional factors such as model accessibility, availability to users, cost implications, and adaptability to fine-tuning. The matrix guided the decision-making process, ensuring a balanced evaluation of each model's technical performance and practical usability.

4) Fine-tune the Selected LLMs

Enhancing the performance of the selected model (either CustomGPT or Gemma) to better handle questions specific to engineering law. The fine-tuning process involved training the selected model on the specialized dataset developed in Step 1. This step aimed to improve the model's ability to provide accurate and relevant answers tailored to the specifics of engineering law. Steps 1 and 4 are interconnected, as the data compilation directly influences the fine-tuning effectiveness.

3.6.2) Models Phase 2 Steps: Evaluation and Testing of Fine-Tuned

5) Evaluation of Fine-tuned Models by Researcher

Conducting a detailed evaluation of the fine-tuned models, CustomGPT and Gemma, to ensure they meet the desired performance standards. This step helped verify enhancements post fine-tuning. The same evaluation metrics (accuracy, precision, recall, F1 score) were used to maintain consistency and comparability with initial evaluations.

6) Select and Refine Best Model

Choosing the best-performing fine-tuned model and further refined it to enhance performance and usability. The selected model (CustomGPT) was refined by adding more data, adjusting parameters, and incorporating additional insights. This iterative process ensured the model is continually improved. Tacit, explicit, and hybrid knowledge were used to guide the refinement process, ensuring comprehensive coverage of relevant information.

7) Expert Evaluation:

Validating the refined model (CustomGPT) through a detailed assessment by domain experts. Experts from relevant the field tested the model and provided feedback based on their specialized knowledge and experience. The expert feedback helped to identify gaps or areas for improvement that were not apparent during the initial evaluations.

8) Incorporate Expert Knowledge:

Enhancing the model based on the feedback received from experts. The insights and recommendations from experts were integrated into the model, refining its responses and improving overall accuracy and relevance. Tacit knowledge from expert feedback was crucial in this step, ensuring the model benefits from deep domain expertise. Steps 5 and 8 are interconnected as when experts provide their feedback, they build upon the researchers' findings, offering deeper insights and domain-specific knowledge that can further refine the model. Therefore, the initial evaluation by researchers was critical for preparing the model for expert scrutiny, making these steps inherently interconnected in the iterative process of model improvement.

3.6.3) Phase 3 Steps: Results

9) Test with Recruited Users

Evaluating the fine-tuned model (CustomGPT) in real-world scenarios with actual users. Recruited users test the model by asking it questions and providing feedback on its performance. This step assessed the model's usability and practical applicability. The user feedback highlighted the model's strengths and areas for improvement, providing valuable insights for further refinement.

10) Analyse Impact on Decision Making:

Analysing the impact of the model on decision-making processes. This involved comparing decision outcomes, decision speed, decision confidence, and other relevant metrics to determine how the model influences user behaviour and compliance efforts. Step 10 is interconnected with Step 6, ensuring that the findings from user testing inform further model refinements.

These methodological steps are elaborated upon throughout this chapter, detailing each phase of the research and the specific activities undertaken to develop and validate the custom GPT model.

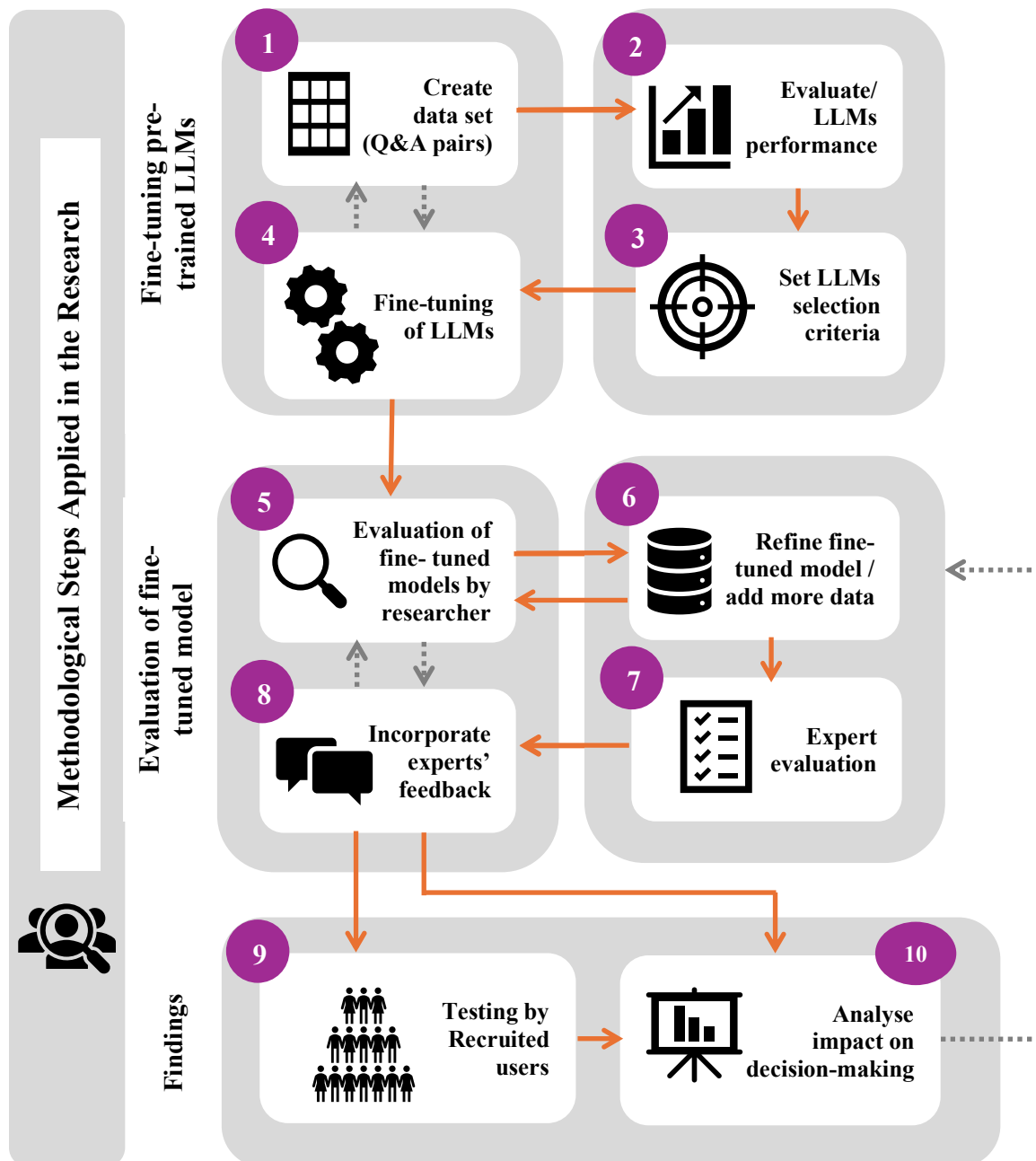


Figure (3.2): Methodological steps applied in the research (Source: Student's work)

3.7) Research Approach for Phase 1 Steps: Fine-Tuning Pre-trained Models (LLMs)

The methodology steps 1, 2, 3 & 4 of this research served as a blueprint for assessing the capabilities of advanced LLMs – specifically ChatGPT-4 and Gemini. This section details a comprehensive framework for evaluating the capabilities of advanced LLMs, specifically ChatGPT-4 and Gemini, within the context of Bahrain's engineering law, which regulates the engineering profession. This evaluation was based on a robust dataset created from 45 articles related to the law, further enhanced by human-AI collaboration.

The methodology of this phase followed a series of carefully structured steps: constructing the dataset, systematically interrogating the LLMs, and analytically grading their responses. Each phase aimed to measure the performance of the models using precise metrics and to assess their practical applicability within the legal framework of engineering. A significant component of the analysis involved rigorous quantitative assessment, ensuring that the evaluation metrics provided an objective and comprehensive understanding of the models' capabilities.

The following sub sections provide a detailed explanation of this systematic approach, emphasizing the thoroughness and scope of the comparative analysis, with a particular focus on the quantitative methods employed.

3.7.1) Step 1: Dataset

Bahraini building code serves as a comprehensive guidebook for building permit regulations within the Kingdom of Bahrain. Comprising 17 chapters,

this document addresses various aspects related to building construction, safety, and compliance. Each chapter focuses on specific requirements enforced by 17 authorities and directorates. Notably, Chapter (14) is related to the regulation of engineering professions (Benayat, 2025).

The dataset used for this part is a legal document that regulates the practice of engineering professions in the Kingdom of Bahrain. It is a PDF file that can be accessed online, and it contains 45 articles that address various aspects of practicing the engineering professions in the Kingdom. The dataset covers the regulations and requirements enforced by the CRPEP. The CRPEP is the official body in the Kingdom of Bahrain that grants licenses to all Engineering Offices and Engineers in the country, as per engineering law (CRPEP, 2014).

The initial phase commenced with the extraction of legal text from the designated engineering law document. This raw text went through a meticulous cleaning process, employing established Natural Language Processing (NLP) techniques such as tokenization, stemming/lemmatization, and stop word removal to eliminate noise and standardize the text for optimal LLM comprehension. The dataset was analysed according to five characteristics: volume, variety, velocity, veracity, and value. The analysis indicated that the dataset was small, but it contained both structured and unstructured data that required further processing and annotation. While the dataset was static, it might require periodic updates to reflect the changes in the law or its interpretation. Moreover, the dataset was reliable and accurate, though it posed some challenges in terms of ambiguity or contradiction in different contexts and scenarios. The dataset was valuable for the research, as it provided relevant information that could be used to generate realistic and meaningful questions and answers. The research methodology involved using

natural language processing techniques to extract information from the document and generate questions and answers based on the content and purpose of the document. The research evaluation measured the performance and accuracy of the ML model using various metrics and benchmarks.

The 5Vs—Volume, Velocity, Variety, Veracity, and Value—are crucial characteristics of big data. These 5Vs guide the design, training, and deployment of Q&A models, ensuring they perform effectively in real-world scenarios (Abdelkader & Ibrahim, 2023; AR Paiva & T Tasiden, 2013). Table (3.1) gives a summary of dataset description:

Table (3.2): Characteristics of the dataset for Q&A generation

Dataset	Size	Format / at	Content	The 5 Vs of Big Data					
				Volume	Velocity	Variety	Veracity	Value	
Bahrain Engineering Practice Law 2014	Small	PDF / online	45 articles on regulating the practice of engineering professions	Small, may need augmentation or advanced NLP	Static, may need updates	Structured and unstructured, need processing and annotation	Reliable and accurate, may have ambiguity or contradiction	Valuable, provide relevant info for Q&A generation	

a) Dataset Expansion through AI-Humann Collaboration

As the available dataset was relatively small, and the records of questions frequently asked by CRPEP's clients were not available to the researcher at the time of conducting this study. The study embarked on a collaborative approach, utilizing both AI capabilities and human expertise to curate a dataset of 235 diverse questions. This dataset was meticulously constructed from 45 articles of law, distributed among 11 clauses, with each article contributing five questions of distinct types. These question types were carefully chosen to encapsulate various analytical dimensions: factual, open-

ended discussion, comparison, critical thinking, and hypothetical scenarios, ensuring comprehensive coverage of potential query spectrums. From literature review, these types of questions were identified as frequently asked questions related to regulatory domains. The approach applied in this step was elaborated by Al Nama et al. (2024), who leveraged human expertise and AI for engineering regulatory data expansion in their case study with ChatGPT (Al Nama et al., 2024).

Both LLMs, ChatGPT-4 and Geminin, were instructed to generate a set of Q&A required for ML by feeding each LLM with the 45 Articles separately. They were then instructed to generate a set of 5 Q&A at least that include factual, hypothetical, theoretical, open ended, and comparative question types.

Figure (3.3) and (3.4) depict examples of instruction chats with Chatgpt-4 and Gemini respectively to generate a set of Q&A related to Article (2) of the engineering law:

Refer to the below law article to generate a set of 5 questions and answers that include factual, hypothetical, theoretical, open ended, and comparative question types. The set is to be used as a part of dataset required for machine learning and testing. Arrange the questions and answers in a tabular format, the table should have 5 columns for serial number, reference article, question type, question, and answer: Article 2

It is prohibited to practice any engineering profession without a licence obtained from the council in accordance with the provisions of this law, its implementing regulations, and decisions issued in accordance with its provisions. A licenced engineer may not practice any engineering profession in any field, branch, or category other than the field, branch, or category in which he/she is licenced to practice.

Here is a table with five questions and answers based on Article 2 of Bahrain's Law No. (51) of 2014, organized by question type.

Serial Number	Reference Article	Question Type	Question	Answer
1	Article 2	Factual	What is required to legally practice any engineering profession according to Article 2 of Bahrain's Law No. (51) of 2014?	To legally practice any engineering profession, one must obtain a licence from the council in accordance with the provisions of the law, its implementing regulations, and decisions issued in accordance with its provisions.
2	Article 2	Hypothetical	What would happen if an engineer practices in a field or category other than the one they are licensed for?	If an engineer practices in a field or category other than the one they are licensed for, they would be in violation of Article 2 of Bahrain's Law No. (51) of 2014, as it prohibits practicing any engineering profession outside the specific field, branch, or category for which the engineer is licensed.
3	Article 2	Theoretical	Why is it important to restrict engineers to practice only within their licensed field, branch, or category?	It is important to restrict engineers to practice only within their licensed field, branch, or category to ensure they have the necessary expertise and qualifications for the specific area of engineering they are engaged in. This helps maintain professional standards, safety, and the integrity of the engineering profession.
4	Article 2	Open-ended	How might the licensing requirements in Article 2 impact the quality and safety of engineering projects in Bahrain?	The licensing requirements in Article 2 likely enhance the quality and safety of engineering projects in Bahrain by ensuring that only qualified and licensed engineers, who have been vetted and approved by the council, are allowed to practice. This reduces the risk of unqualified individuals undertaking engineering tasks, which could lead to substandard work and potential safety hazards.
5	Article 2	Comparative	How does the licensing requirement for engineers in Bahrain's Law No. (51) of 2014 compare to the licensing requirements in another country?	The licensing requirement for engineers in Bahrain's Law No. (51) of 2014 is similar to many other countries that also mandate licensure to ensure competency and regulate the profession. For example, in the United States, engineers must be licensed by the state board, and they are also restricted to practice within their specific area of expertise. This common approach underscores the global importance of licensure in maintaining professional standards and protecting public welfare.

Figure (3.3): Example of an instruction chat with ChatGPT-4 to generate a set of Q&A related to article (2) - (Source: Student's work)

E Refer to the below law article to generate a set of 5 questions and answers that include factual, hypothetical, theoretical, open ended, and comparative question types. The set is to be used as a part of dataset required for machine learning and testing. Arrange the questions and answers in a tabular format, the table should have 5 columns for serial number, reference article, question type, question, and answer: Article 2

It is prohibited to practice any engineering profession without a licence obtained from the council in accordance with the provisions of this law, its implementing regulations, and decisions issued in accordance with its provisions. A licenced engineer may not practice any engineering profession in any field, branch, or category other than the field, branch, or category in which he/she is licenced to practice.

Law No. 51 Q&A Dataset (Sample)

Serial Number	Reference Article	Question Type	Question	Answer
1	Article 2	Factual	Is a license mandatory to practice engineering in Bahrain?	Yes, according to Article 2, a license obtained from the relevant council is mandatory.
2	Article 2	Hypothetical	Can a licensed civil engineer legally consult on a project involving software development?	No, based on Article 2, licensed engineers can only practice within their specific field, branch, or category of licensure.
3	Article 2	Theoretical	Could there be exceptions for engineers to work outside their licensed field in emergencies?	This is a theoretical question. While Article 2 doesn't explicitly mention exceptions, specific circumstances might be considered under implementing regulations or through legal interpretations. However, definitive answers would require further legal analysis.
4	Article 2	Open-Ended	What are the potential consequences for an engineer caught practicing without a license in Bahrain?	This is an open-ended question. Potential consequences could include fines, project stoppages, legal action, and reputational damage.
5	Article 2	Comparative	How does Article 2 compare to licensing requirements for professional engineers in the United States?	This is a comparative question. Research on engineering licensing in the US is needed for a detailed comparison. Points of comparison could include licensing procedures, limitations on practice outside licensed fields, or enforcement mechanisms.

Figure (3.4): Example of an instruction chat with Gemini to generate a set of Q&A related to article (2) – (Source: Student's work)

b) Human Expert Review and Refinement

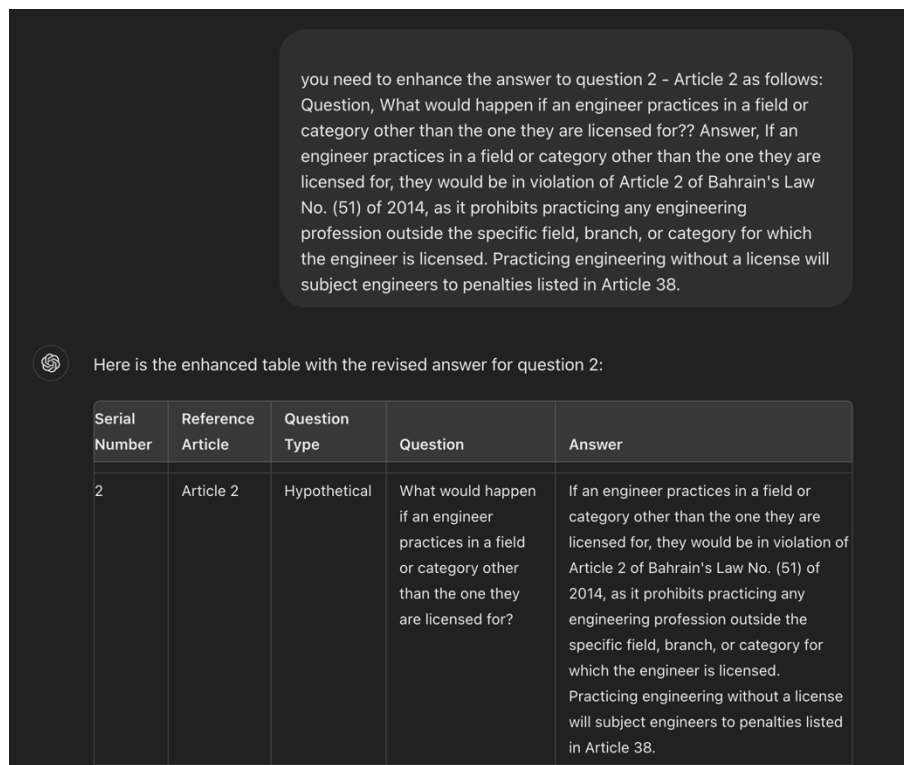
To ensure both thematic relevance to the engineering law document and factual accuracy within the engineering legal domain, the human researcher, acting as the subject matter expert, meticulously reviewed and refined the LLM's initial attempts at generating Q&A related to engineering law. This iterative process of human-AI collaboration involved several key steps after generating preliminary set of Q&A pairs as described in the above step No. (2.1.1):

1. **Human Evaluation:** The researcher meticulously evaluated the LLM's output. This evaluation focused on two key aspects:
 - **Thematic Relevance:** Did the Q&A accurately reflect the content and key concepts within the engineering law document?
 - **Factual Accuracy:** Were the answers sounded and consistent with established principles of engineering law?
2. **Iterative Feedback:** Based on the evaluation, the researcher provided targeted feedback to the LLM through guidance mechanisms. This involved:
 - Highlighting irrelevant or inaccurate Q&A.
 - Suggesting alternative phrasing or providing additional details for improvement.
 - **Offering sample Q&A pairs that demonstrate the desired characteristics for the final dataset.**
3. **Refined Output:** Through this iterative feedback loop, the LLM refined its Q&A generation based on the researcher's expertise. Each iteration potentially led to a more accurate and comprehensive set of Qs&As related to engineering law.

This collaborative approach leveraged the LLMs, ChatGPT-4 and Gemini 1.5, ability to generate a vast initial set of questions, while the researcher's domain-specific knowledge ensured the relevance, accuracy, and legal soundness of the final Q&A dataset.

This revision expanded the explanation to include the key steps involved in the human-AI collaborative process. It also emphasized the researcher's role as the legal domain expert guiding the LLM towards producing a robust Q&A dataset specifically focused on engineering law.

Figure (3.4) and (3.5) depict examples of instruction chats with Chatgpt-4 and Gemini respectively to generate enhanced answers related to article (2).



you need to enhance the answer to question 2 - Article 2 as follows:
 Question, What would happen if an engineer practices in a field or category other than the one they are licensed for?? Answer, If an engineer practices in a field or category other than the one they are licensed for, they would be in violation of Article 2 of Bahrain's Law No. (51) of 2014, as it prohibits practicing any engineering profession outside the specific field, branch, or category for which the engineer is licensed. Practicing engineering without a license will subject engineers to penalties listed in Article 38.

Here is the enhanced table with the revised answer for question 2:

Serial Number	Reference Article	Question Type	Question	Answer
2	Article 2	Hypothetical	What would happen if an engineer practices in a field or category other than the one they are licensed for?	If an engineer practices in a field or category other than the one they are licensed for, they would be in violation of Article 2 of Bahrain's Law No. (51) of 2014, as it prohibits practicing any engineering profession outside the specific field, branch, or category for which the engineer is licensed. Practicing engineering without a license will subject engineers to penalties listed in Article 38.

Figure (3.5): Example of instruction chat with ChatGPT-4 to generate an enhanced answer related to article (2) – (Source: Student's Work)

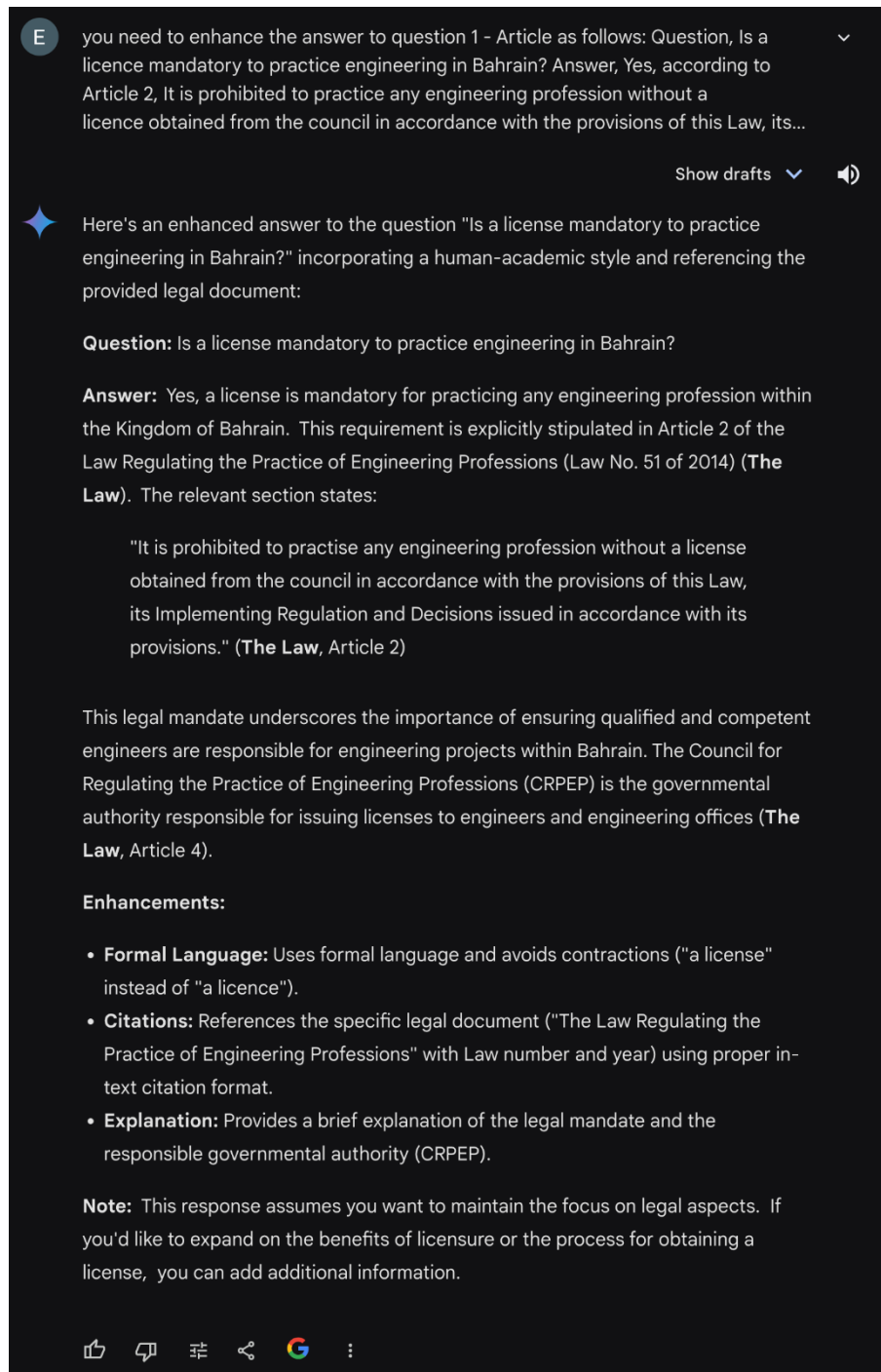


Figure (3.6): Example of instruction chat Gemini to generate an enhanced answer related to article (2) – (Source: Student's work)

c) Prompt Design Optimization

Throughout the dataset expansion process, prompt engineering played a critical role in guiding the LLMs to generate contextually accurate and legally sound content. Initial prompts were designed to be clear, specific, and aligned with the structure of engineering law provisions. These were tested iteratively, with refinements based on the quality of generated outputs and the researcher's expert feedback. Adjustments included modifying the level of detail, framing questions around specific articles, and using exemplar Q&A pairs to reinforce the expected format and tone. This aligns with recent findings that effective prompt engineering significantly improves LLM performance in domain-specific applications (Marvin et al., 2024; White et al., 2023).

3.7.2) Step (2): Evaluate the Performance of LLMs

This step delves into the detailed evaluation of the performance of ChatGPT-4 and Gemini, using a systematic approach to understand their capabilities and limitations. The evaluation combines qualitative insights and quantitative analysis to provide a comprehensive assessment of the models.

a) LLMs Interrogation and Response Grading

Both LLMs, ChatGPT-4 and Gemini, were subjected to the dataset questions. The responses were meticulously evaluated and graded on a scale where '0' denoted an incorrect answer, '0.5' indicated a partially correct response, and '1' signified a fully correct answer. An integral part of this assessment also involved noting whether the models could supplement their answers with additional information or reference their responses to preceding queries, which provided insight into their contextual awareness and knowledge integration capabilities.

Figure (3.6) depicts partial of the evaluation sheet. The approach applied in this step was elaborated by Al Nama & Mahmud (2024), who analysed the performance of ChatGPT-4 and Gemini by Questions Category in Engineering Regulations. The full evaluation is included in Appendix (A) (Al Nama & Mahmud, 2024)

Article No	Caluse Title	Question Type	Article Details	Question	Gemini Answer	Gemini - Answer Grade	Comments re. Gemini Answer	GPT-4 Answer	GPT-4 Answer Grade	Comments re. GPT-4 Answer
9	Licenses	Factual	Article 9 Engineers working in government ministries, authorities and public entities shall be licensed to practice one of the engineering professions, after ensuring that they	What is the requirement for engineers working in government ministries, authorities and public entities to practice engineering profession?	They must be licensed for the specific engineering profession they practice.	0.5	partially correct answer	Engineers must be licensed to practice one of the engineering professions, after ensuring that they satisfy the engineering qualification prescribed in clause 3 of Article 7 of this Law.	1	correct answer
9	Licenses	Critical Thinking	satisfy the engineering qualification prescribed in clause 3 of Article 7 of this Law.	Why do you think the Law requires the license request to be submitted by the employing entity?	Streamlines the process and ensures responsibility lies with the government agency to verify qualifications.	0.5	partially correct answer	Possible answer: The Law requires the license request to be submitted by the employing entity to ensure that the	1	correct answer
9	Licenses	Hypothetical	The license request shall be submitted by the employing entity; such license is free of charge and shall remain valid for the period of employment without the need to be renewed.	What would happen if an engineer working in government ministries, authorities and public entities does not satisfy the engineering qualification prescribed in clause 3 of Article 7 of this Law?	The employer cannot submit a license request for them, effectively preventing them from practicing engineering within the government entity.	1	correct answer	The engineer would not be eligible to obtain a license to practice engineering profession in Bahrain, unless he/she fulfills other requirements or conditions specified in the Law or its Implementing Regulation.	1	correct answer
9	Licenses	Open-ended Discussion	The disciplinary action against the aforesaid employees shall be in accordance with the laws and regulations applicable to their jobs as determined by their employer.	What are some examples of "disciplinary action" that could be taken against engineers working in government ministries, authorities and public entities?	Disciplinary action will follow the existing regulations of the employing entity, but could include warnings, reprimands, demotions, and even dismissal.	1	correct answer	This is a subjective question that may have different opinions and perspectives. Some possible examples are: - Verbal or written warning, reprimand,	1	correct answer, with additional information
9	Licenses	Comparison		How does the licensing process for engineers working in government ministries, authorities and public entities differ from other countries or regions?	Some countries may require government engineers to follow the same licensing process as private sector engineers, while others may have separate provisions with varying degrees of complexity.	1	correct answer	This is a subjective question that may have different opinions and perspectives. Some possible points of comparison are: - Other countries or regions might have different criteria or standards for	1	correct answer, with additional information
10	Licenses	Factual	Article 10 To obtain a license for establishing an engineering office in a form of a sole proprietorship, the following conditions	What specific experience criteria must a sole proprietor fulfill to meet the "practiced professional engineering" requirement?	The required experience criteria should be specified in the Implementing Regulation. It could involve a minimum number of years of experience, types of projects worked on, or passing	0.5	partially correct answer	A sole proprietor must have practiced professional engineering for a period as determined by the Implementing Regulation, either consecutively	0.5	partially correct answer
10	Licenses	Critical Thinking		Does requiring the owner to fulfill all conditions of Article 7 potentially	There's potential for redundancy if Article 7 requirements are extensive.	0	incorrect / irrelevant	This depends on the conditions listed in Article 7. If those	0	incorrect / irrelevant

Figure (3.6): Evaluation of ChatGPT-4 and Geminin answers (Source: Student's work)

b) Evaluation Metrics and Analysis

Accuracy, precision, recall, and F1 score are commonly used evaluation metrics in ML. Accuracy measures the overall correctness of a classifier's predictions. Precision measures the proportion of true positive predictions out of all positive predictions, while recall measures the proportion of true positive predictions out of all actual positive instances. F1 score is the harmonic mean of precision and recall, providing a balanced measure of a classifier's performance. These metrics are important for assessing the effectiveness of binary, multi-class, and multi-labelled classifiers. Accuracy and F1 score are widely used, it is important to consider other metrics for a comprehensive evaluation (Jude Chukwura Obi, 2023; Riyanto et al., 2023; Silhavy & Silhavy, 2023).

1. **Accuracy:** Measures the proportion of correct answers out of all questions. Accuracy is a straightforward metric that provides an overall indication of how well the models perform in providing correct answers. It is commonly used in NLP model evaluations and provides a clear, easy-to-understand measure of performance (D. Manning et al., 2009) .

- **Formula:**

$$Accuracy = \frac{\text{total number of questions}}{\text{number of correct answers}}$$

2. **Precision:** Indicates the proportion of true positive answers out of all answers given by the model. Precision is crucial when the cost of false positives is high. In the context of legal documents, providing incorrect

information can lead to significant misunderstandings, making precision a vital metric (Jordan et al., 2006).

- **Formula:**

$$Precision = \frac{true\ positive + false\ positive *}{true\ positive}$$

**Assuming "Partially Correct Answers" are considered False Positives*

3. **Recall:** Measures the proportion of true positive answers out of all actual relevant answers. Recall is important for ensuring that the model captures as many relevant answers as possible. In legal contexts, missing relevant information can be detrimental, so high recall is desirable (Jordan et al., 2006).

- **Formula:**

$$Recall = \frac{true\ positive + false\ negative *}{true\ positive}$$

**Assuming "Incorrect Answers" are considered False Negatives*

4. **F1 Score:** Provides a harmonic mean of precision and recall, offering a single metric for balancing both. The F1 score is a balanced metric that considers both precision and recall, making it ideal for evaluating models where both false positives and false negatives are equally important. It provides a comprehensive view of the model's performance in handling complex queries (Scikit-Learn, 2024).

- **Formula:**

$$F1\ Score = 2 \times \frac{precision \times recall}{precision + recall}$$

To quantify the proposed models' performance, standard evaluation metrics which are precision, recall, and F1 score were employed. Precision measured the accuracy of the provided answers, recall assessed the models' ability to retrieve all relevant instances, and the F1 score, being the harmonic mean of precision and recall, offered a balanced view of the models' accuracy in both retrieving relevant information and omitting irrelevant details. The inclusion of recall was crucial, as it evaluated the completeness of the models' answers, ensuring they were not only correct when they are providing an answer but also that they provided an answer whenever it was appropriate to do so. Other studies used similar approach to evaluate the performance of LLMs, including a study by (Almasre, 2024) which discusses the evaluation metrics used to assess the performance of the custom GPT model, including accuracy, precision, recall, and F1 score.

3.7.3) Step (3): Establishing Selection Criteria for Optimal LLMs in Fine-Tuning

In this step, the methodology focused on defining the criteria for selecting the most suitable LLM for fine-tuning. This involved a comprehensive evaluation to determine which model best met the specific needs and stringent requirements of engineering practice compliance within the Kingdom of Bahrain. The goal was to ensure that the chosen model was not only high-performing but also aligned with the regulatory and practical standards required in this domain.

a) Objective Comparison for Model Suitability

The heart of this part of methodology lied in establishing an empirical comparison between ChatGPT-4 and Gemini. This comparison was not merely a performance showdown but a structured analysis to determine which

model was best suited for developing a Q&A model. The ultimate aim was to discern the most apt AI model that aligned with the stringent requirements of engineering practice compliance within the Kingdom of Bahrain.

b) Choosing the Right LLM for Fine-tuning

The methodology began by evaluating two prominent LLMs, ChatGPT-4 and Gemini. While both models demonstrated impressive performance across various question types, achieving accuracy scores exceeding 0.9 as further discussed in Chapter 4 of this research. As the research progressed, the landscape of large LLMs underwent a significant shift. Initially, the focus was on comparing general-purpose models like ChatGPT-4 and Gemini for their suitability (Google, 2024b; OpenAI, 2024).

The recent launch of ChatGPT-4o by OpenAI introduced the concept of custom GPTs (OpenAI, 2024). These models can be trained on domain-specific datasets, offering the potential for increased accuracy and effectiveness in a particular field like engineering law. This development sparked an investigation into whether a custom GPT trained on engineering law data could be a more appropriate solution. A similar approach was taken by Almasre (2024) in her study on the development and evaluation of a custom GPT for assessing students' typography designs. Almasre's research demonstrated that custom GPTs could provide consistent and objective evaluations, highlighting the potential benefits of domain-specific customization (Almasre, 2024).

Coinciding with this shift in focus, Google released Gemma, an open-source, lightweight LLM designed specifically for developers (Google, 2024c). Gemma's key strengths lied in its efficient computational requirements and its open-source nature. This allowed for customization

through fine-tuning on the engineering law dataset, potentially leading to the creation of a robust Q&A system specifically tailored to this domain.

Therefore, the research focus had shifted from general-purpose LLMs to custom GPTs and Gemma. This shift reflected the growing importance of domain-specific customization in building effective Q&A systems and aligned the research with the latest advancements in LLM technology.

Table (3.3) below illustrates the criteria set to compare between ChatGPT-4, Custom GPT, Gemini, and Gemma:

Table (3.3): Comparison matrix between Custom GPT and Gemma (Source: Student's work)

Criteria	By OpnAI (OpenAI, 2024)		By Google (Google, 2024b, 2024c)		Recommendation
	ChatGPT-4	Custom GPT	Gemini (Google)	Gemma (Google)	
Cost Implications	Paid tiers available (OpenAI Pricing)	Paid tiers available (OpenAI Pricing)	Free (through Google Colab) (Gemini Info)	May have paid tiers (Gemma Info)	Gemini is ideal for cost-conscious projects.
API Availability (Bahrain)	Available globally (OpenAI API)	Available globally (OpenAI API)	Limited public access (Gemini Info)	Limited public access (Gemma Info)	ChatGPT-4 and Custom GPT offer more established APIs, Gemini and Gemma's future accessibility looks promising.
User Friendly	User-friendly interfaces available (OpenAI User Guide)	Customizable for specific needs (OpenAI User Guide)	Requires technical knowledge (Gemini Info)	Requires technical knowledge (Gemma Info)	ChatGPT-4 and Custom GPT are more accessible for non-technical users.
Case Studies	Extensive case studies available (OpenAI)	Extensive case studies available (OpenAI Case Studies)	Limited case studies (Gemini Info)	Limited case studies (Gemma Info)	ChatGPT-4 and Custom GPT have more documented use cases*.

Criteria	By OpnAI (OpenAI, 2024)		By Google (Google, 2024b, 2024c)		Recommendation
	ChatGPT-4	Custom GPT	Gemini (Google)	Gemma (Google)	
	Case Studies)				
Continuous Development	Actively developed (OpenAI Research)	Actively developed (OpenAI Research)	Actively developed (Gemini Info)	Actively developed (Gemma Info)	Both are continuously improving, backed by their respective organizations.
Performance Results	Accuracy (93.2%), Precision (93.9%), Recall (99.1%), F1-Score (96.5%) (refer to chapter 4)	Customizable for high performance (OpenAI Research)	Accuracy (85.1%), Precision (88.9%), Recall (95.2%), F1-Score (91.9%) (refer to chapter 4)	Competitive accuracy and precision (Gemma Info)	While ChatGPT-4 has a slight edge, Gemini and Gemma offer strong performance. Custom GPT can be tailored to meet specific needs.
Additional Information	Provides detailed responses (OpenAI Research)	Customizable for detailed responses (OpenAI Research)	Detailed and contextual responses (Gemini Info)	Detailed and contextual responses (Gemma Info)	Gemini and Gemma are valuable for regulatory documents requiring comprehensive explanations.
Comparison Task	Strong in factual accuracy, critical thinking, and open-ended discussions (OpenAI Research)	Customizable for specific tasks (OpenAI Research)	Strong in factual accuracy and critical thinking (Gemini Info)	Strong in factual accuracy and critical thinking (Gemma Info)	ChatGPT-4 excels in comparisons, but Gemini and Gemma can be improved with further training on specific regulatory documents. Custom GPT can also be optimized for specific regulatory requirements.

Criteria	By OpnAI (OpenAI, 2024)		By Google (Google, 2024b, 2024c)		Recommendation
	ChatGPT-4	Custom GPT	Gemini (Google)	Gemma (Google)	
*	Recent Case Studies 1. OpenAI: ○ ChatGPT-4: Utilized in customer support to automate responses, reducing response time by 60% (OpenAI Research). ○ Custom GPT: Implemented in a legal firm to draft documents, improving accuracy and consistency by 70% (OpenAI Research). 2. Google: ○ Gemini: Applied in healthcare to analyse patient data, enhancing diagnosis accuracy by 15% (Google Research). ○ Gemma: Used in financial services to detect fraud, increasing detection rates by 20% (Google Research).				

3.7.4) Step (4): Fine-Tuning LLMs

Building upon the selection criteria findings discussed in the above section, this step supported the suitability assessment of both CustomGPT and Gemma. In this step, both CustomGPT and Gemma were fine-tuned, and their performance then tested post-fine-tuning. The objective was to create Q&A models specialized in addressing regulatory inquiries related to engineering law based on the most suitable LLM. This involved the detailed process of tailoring the chosen LLMs to meet the specific needs of the regulatory framework, ensuring precise and contextually appropriate responses. The following subsections detail the steps of fine-tuning both models.

a) Custom GPT

According to (OpenAI, 2023b) the launch of custom Generative Pre-trained Transformers (GPTs) by OpenAI on November 6, 2023, represents a notable advancement in artificial intelligence. These custom GPTs enable

users to create specialized versions of ChatGPT tailored for specific purposes, thereby enhancing their ability to generate precise and contextually appropriate responses across various domains.

Custom GPT models are fine-tuned versions of pre-trained language models designed to meet the unique needs of specific domains. This customization process involves leveraging domain-specific data during fine-tuning to enhance the model's understanding and generation of relevant content. Based on the information available on (OpenAI, 2023b) the following steps can be taken to refine these models:

- **Customization Process:** Users can initiate the creation of a custom GPT by engaging in a dialogue, providing specific instructions, and incorporating additional domain-relevant knowledge. This process is designed to be user-friendly and does not require coding skills, making it accessible to a broad audience.
- **Enhancing Accuracy:** The integration of domain-specific data during fine-tuning ensures that the custom GPT can accurately interpret and respond to queries within that domain. This is particularly valuable in fields such as law, medicine, and regulatory compliance, where precision and context are critical.
- **Utilizing Built-in Capabilities:** Custom GPTs can be endowed with specific functionalities, such as web searching, image generation, or data analysis, tailored to meet the unique requirements of the domain. Users define these capabilities to enhance the model's effectiveness in performing specialized tasks.

b) Steps to Create a Custom GPT for “Engineering Law”

One of the key objectives of this research is to create a model that can assist with information and queries related to engineering law and its implementing regulations. This involved understanding the law's structure, key articles, and regulations governing the practice of engineering professions in Bahrain. Based on the information available on (OpenAI, 2023b) the following steps were taken to refine the model:

- 1) Document Collection:** The initial step involved gathering all relevant documents available on the CRPEP’s website, in addition to generating a dataset of Q&A pairs. The data collected included the engineering law and its implementing regulations, CRPEP's forms and circulars, as well as prime minister edicts and relevant decisions.
- 2) Data Cleaning:** The collected data was cleaned and organized to ensure it was accurate and consistent. This involved removing duplicates, correcting errors, and filling in any missing values. The cleaned data was then structured in a format that the machine could easily read. This step was essential for creating a custom GPT model specialized in addressing queries related to engineering law as by ensuring the data is clean and well-organized, the model's ability to understand and respond to complex legal questions in the engineering field is expected to be improved significantly.
- 3) Create Repository:** A centralized repository was created for all collected documents to ensure ease of access and reference during the model training process.
- 4) Content Analysis:** This phase involved analysing the documents to extract key provisions, regulations, and procedures. Essential sections

such as licensing requirements, fee structures, disciplinary actions, and obligations were identified. The documents were then indexed for quick reference, ensuring that the GPT could retrieve and reference specific sections accurately.

5) Instruction Design: Specific instructions were developed to guide the GPT's responses, ensuring clarity and precision. These instructions included:

- **Reference to Engineering Law and Related Documents:** This ensured that responses were grounded in the relevant legal and regulatory framework, fostering reliability and trustworthiness.
- **Indicate Document and Section/Article:** Enhancing the credibility of responses by allowing users to verify the information independently, thus promoting transparency and trust.
- **Provide Background Information or Context:** Offering context alongside facts to help users understand the broader implications and applications of the information.
- **List Multiple Answers Separately:** Organizing responses clearly to aid in navigation and understanding.
- **Request Users to Specify the Level of Detail:** Tailoring responses to the user's needs, increasing efficiency and user satisfaction.
- **Highlight Urgent Questions and Answers:** Prioritizing critical issues to ensure timely information.
- **Include Relevant Keywords:** Improving clarity and relevance by aligning the response closely with the user's intent.

6) Model Training: The training phase utilized the indexed documents to ensure the GPT comprehended the structure, key terms, and specific

- requirements laid out in the engineering law and its regulations. The training process was customized to focus on legal language and context-specific responses, enhancing the model's capability to interpret and apply legal provisions accurately.
- 7) **Testing and Validation:** The GPT was tested with sample queries to verify its ability to accurately reference the engineering law, regulations, and forms. Responses were validated for correctness, relevance, and comprehensiveness. A feedback loop was implemented to make necessary adjustments based on errors or inaccuracies observed during testing.
 - 8) **Deployment:** The GPT was deployed with access to the indexed documents, ensuring it could retrieve and reference specific sections accurately. A user-friendly interface was developed to allow professionals to interact with the GPT efficiently. Figure (3.7) depicts the interface for the developed CustomGPT.
 - 9) **Continuous Update and Maintenance:** Regular updates are suggested to the GPT with any amendments, new regulations, or additional relevant documents to ensure it remained current and accurate. The GPT's performance along with experts and users feedback were considered to identify areas for improvement and ensure ongoing accuracy and relevance.

This methodology ensured the creation of a robust and reliable GPT tailored to address regulatory inquiries related to engineering law, providing accurate and contextually appropriate responses to support engineering practice compliance within the Kingdom of Bahrain. The approach of customizing a GPT model for specific domains are currently growing, and the

steps of tailoring these GPTs can vary based on the objectives of creating the model, in a recent study “Development and Evaluation of a Custom GPT for the Assessment of Students’ Designs in a Typography Course” by (Almasre, 2024) details the creation of a custom GPT model named “Typography Evaluator.” The customization process involved defining the model's objective, setting up conversation starters, adding extra knowledge through grading rubrics and evaluation metrics, and selecting required capabilities like image processing and data analysis. The model was designed to provide consistent and objective evaluations of students' typography designs, ensuring fairness and efficiency. The custom GPT was tested for reliability and quality of feedback, demonstrating its potential to enhance educational assessments by offering detailed and structured feedback based on predefined criteria.

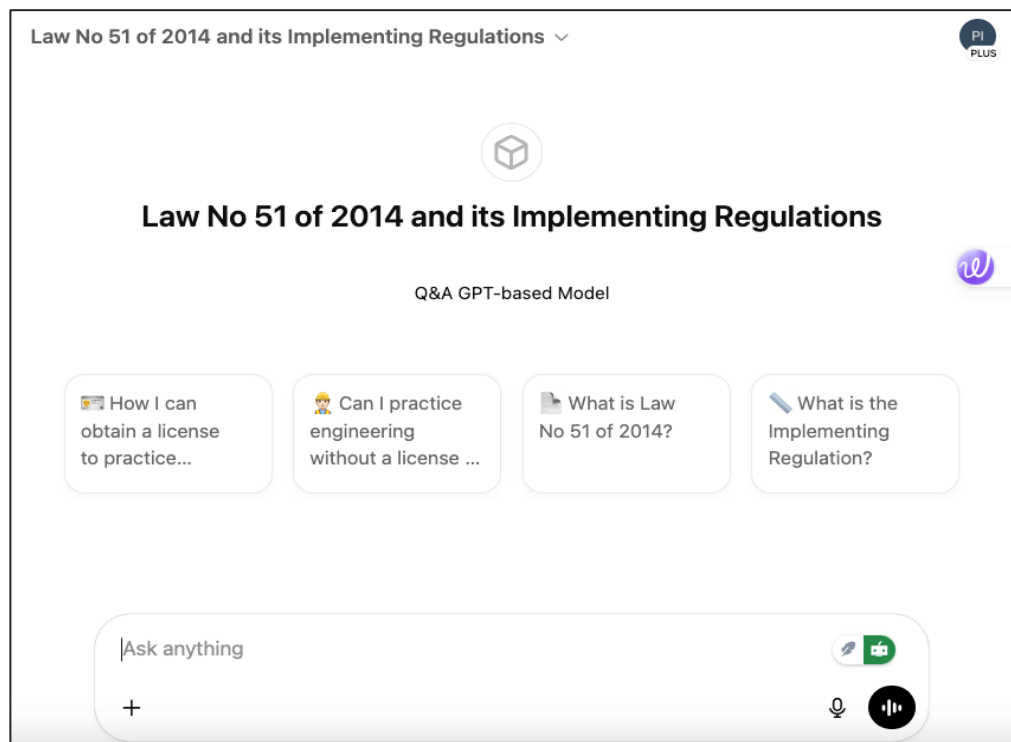


Figure (3.8): The interface for the developed CustomGPT (Source: Student's work)

c) Gemma

This section of the methodology details the approach of fine-tuning the Gemma 2B model, an LLM, using Google Colab, a cloud-based platform. This method involved data extraction, cleansing, and preparation, followed by fine-tuning using LoRA Keras and enhancing retrieval capabilities with multimodal RAG.

▪ Rationale for Tool and Technique Selection

- **Gemma 2B and Colab Synergy:** Within the Gemma family of LLMs, Gemma 2B emerges as the optimal choice for regulatory Q&A applications. While Gemma 7B might boast superior accuracy in certain domains, its resource-intensive nature can present challenges regarding training and deployment. In contrast, Gemma 2B's lightweight architecture facilitates faster training and deployment cycles, a crucial factor for research methodologies with limited computational resources or tight time constraints (Google, 2024c).
- **Colaboratory (Colab):** Colab provides a convenient web-based environment for writing and executing Python code. Notably, it offers free access to GPUs, making it particularly advantageous for tasks related to ML, data science, and research endeavours. Colab empowers researchers to effortlessly create interactive notebooks, perform data analysis, visualize information, and seamlessly collaborate (Google, 2024a). This user-friendly environment simplifies the fine-tuning process, especially for those without extensive technical expertise in LLM manipulation.

- **LoRA Keras:** LoRA was chosen for its efficiency in fine-tuning large models with limited computational resources. By focusing on low-rank updates, LoRA reduces the number of trainable parameters, making the fine-tuning process more resource-efficient while maintaining performance (Hu et al., 2021). This is crucial for handling LLMs like Gemma 2B within the computational limits of a typical Colab environment.
- **RAG (Retrieval-Augmented Generation):** RAG enhances the model's ability to answer complex queries by integrating relevant external documents into the response generation process. This is particularly advantageous for applications that involve detailed and nuanced legal documents (acorn.io, 2024).

▪ **Data Preparation**

- **PDF extraction:** The text from engineering law was extracted using the “PyMuPDF” library. This library is well-suited for extracting text from PDF files, converting them into a format that can be processed for ML tasks (PyMuPDF, 2024).
- **Data Cleansing and Preparation:** The extracted text was cleaned to remove any non-relevant content, such as headers, footers, and page numbers. This step ensures that the text data is accurate and suitable for generating Q&A pairs.
- **Generation of Q&A Pairs:** Q&A pairs were generated based on the cleansed text and stored in an XLSX file. The pairs were manually reviewed to ensure accuracy and relevance to the legal content. The dataset was then formatted using the Hugging Face

“datasets” library to prepare it for fine-tuning (HuggingFace, 2024).

- **Fine-tuning with Gemma 2b with LoRA Keras**

- **Model loading:** The Gemma 2b model and its tokenizer were loaded using the `transformers` library. This involved downloading the pre-trained model and preparing it for further training.
- **LoRA integration:** The LoRA (Low-Rank Adaptation) technique was applied to the model using Keras. As LoRA helps in fine-tuning large models efficiently by focusing on low-rank updates, significantly reducing the number of trainable parameters. This approach was particularly beneficial for training LLMs like Gemma 2b within the computational limits of a typical Colab environment (Hu et al., 2021).
- **Running on Colab:** The entire process of fine-tuning was conducted on Google Colab, leveraging its GPU runtime. Colab provided the computational resources necessary for training large models, making it an ideal platform for this task.

- **Application of RAG for Enhanced Retrieval Capabilities**

After fine-tuning, RAG was employed to enhance the model's retrieval capabilities. RAG optimized the model's performance by incorporating relevant external documents into the response generation process. This integration was crucial for improving the model's practical usability, particularly for legal documents

that often involve detailed and nuanced queries (acorn.io, 2024; Raj et al., 2024).

▪ **Evaluation of the Fine-tuned Model**

- The fine-tuned model was evaluated using a separate set of Q&A pairs to ensure its accuracy and reliability. Evaluation metrics such as F1 score, and exact match were used to assess the model's performance.

3.7.5) Step (5): Evaluation of Fine-Tuned LLMs

The evaluation aimed to compare the performance of the two fine-tuned models CustomGPT and Gemma 2B fine-tuned using LoRA Keras with LangChain in responding to queries related to engineering law. The law comprises 45 articles, and the evaluation was focused on 90 questions generated from these articles. In recent research by Almasre (2024) regarding the development and evaluation of a custom GPT for assessing students' typography designs. Almasre's study used a detailed rubric to evaluate students' designs, and human evaluators were trained to ensure consistency in their assessments beside evaluating the custom GPT model by using performance metrics such as accuracy, precision, recall, and F1 score to provide a comprehensive analysis of its capabilities (Almasre, 2024).

a) Question Types and Allocation

The generated Q&A pairs include five types of questions:

- 1. Factual:** Questions that seek specific information directly from the text.

2. **Hypothetical:** Questions that pose scenarios and ask for responses based on the text.
3. **Theoretical:** Questions that involve conceptual understanding and interpretations.
4. **Open-Ended:** Questions that require comprehensive and detailed answers.
5. **Comparative:** Questions that ask for comparisons between different elements or articles.

Given that factual questions are the most frequently asked and form the core purpose of the engineering law to provide factual answers, the evaluation emphasized this type of questions.

b) Evaluation Design

For each of the 45 articles of engineering law, two questions were allocated:

6. **One Factual Question:** to test the models' ability to provide precise and accurate information.
7. **One Question from the Other Four Types:** to evaluate the models' versatility in handling different types of queries.

In total, 90 questions were used for evaluation, ensuring a balanced assessment across all articles and question types.

c) Evaluation Procedure

The evaluation process involved selecting 90 questions to ensure coverage of all articles and types within engineering law. Researcher with expertise in the field then assessed the responses generated by each model. The

performance metrics (accuracy, precision, recall, and F1 score) were calculated based on the evaluations provided by the human expert. By using these evaluation metrics, the comparative analysis provided insights into the strengths and weaknesses of each model in handling the specific requirements of legal document queries, ensuring a robust assessment of their capabilities.

3.7.6) Step (6): Refine Fine-tuned Model and Add More Data

This step served as a crucial bridge between Step 5, where the fine-tuned model was evaluated by the researcher, and Step 7, where the model underwent expert evaluation. This step is interconnected with Step 5, and also indirectly linked with Step 8 through the incorporation of experts' feedback, as Steps 5 and 8 were interdependent.

After evaluating the fine-tuned models in Step 5 and determining that CustomGPT was the best option, the researcher pursued further refinement of the model. This involved adjusting parameters and enhancing the model by incorporating additional data and knowledge available online. The aim was to create a comprehensive and robust Q&A model capable of addressing a wide range of regulatory inquiries related to engineering law.

The refinement process included the following activities:

- **Parameter Adjustment:** Fine-tuning the model's parameters to improve its accuracy, efficiency, and relevance. This involved iterative testing and tweaking to achieve optimal performance.
- **Data Enrichment:** Feeding the model with additional relevant data to broaden its knowledge base. This included incorporating:
 - Implementing regulations

- Licensing forms
- Fee structures
- Updated regulations and amendments
- Any other pertinent documents or information available online
- **Comprehensive Knowledge Integration:** Ensuring that the model integrated this new data seamlessly with the existing knowledge base, thereby enhancing its ability to provide accurate and contextually appropriate responses.

By performing these refinements, the model was better equipped to handle a wider variety of queries, offering users a complete and enriched experience. This step ensured that when the model proceeded to expert evaluation in Step 7, it was already optimized to a high standard, increasing the likelihood of positive feedback and successful validation.

Furthermore, this step was indirectly connected with Step 8, as the enhancements and refinements made here were also benefit from and incorporate the ongoing expert feedback loop. The continuous integration of expert insights ensures that the model remained dynamic and continuously improved in accuracy and relevance, ultimately providing a reliable tool for users. The refinement of CustomGPT involved leveraging domain-specific knowledge and user-centric feedback to enhance the model's performance in addressing regulatory queries. Similar to approaches detailed in the travel chatbot use case, iterative fine-tuning was employed to align the model's outputs with real-world expectations. By incorporating retrieval-augmented techniques and iterative Q&A testing, the refinement process ensured precision in responses while maintaining contextual relevance (Meyer et al., 2024).

3.7.7) Step (7): Evaluation and Continuous Improvement of the Fine-Tuned Model through Independent Expert

This step in the research methodology involved conducting expert interviews. This step was crucial for validating the model, removing biases in evaluation, and establishing a feedback loop to ensure the model remains flexible and up to date with ongoing expert input.

The primary purposes of the expert interviews are threefold:

1. **Validation of the Model:** Experts provided critical insights into the accuracy and relevance of the model's responses, helping to validate the model's performance against real-world standards and expectations.
2. **Bias Removal:** By incorporating diverse expert opinions, the evaluation process became more robust, mitigating potential biases that might arise from a singular perspective.
3. **Continuous Improvement Loop:** Establishing a mechanism for ongoing feedback ensured that the model could be continuously updated and improved based on expert advice, even after its initial launch.

Building on techniques described in domain-specific refinement studies, the fine-tuned CustomGPT model underwent rigorous evaluation by independent experts. Experts assessed the model's responses to ensure contextual accuracy, relevance, and domain alignment. Feedback loops, as highlighted in the travel chatbot study, allowed iterative improvements by incorporating expert insights to reduce biases and address deficiencies (Meyer et al., 2024). Similarly, the Retrieval-Augmented Fine-Tuning (RAFT) study emphasized the integration of retrieval-augmented methods to refine outputs

and enhance the adaptability of models in specialized domains (T. Zhang et al., 2024).

a) Participant Selection

In the context of the CRPEP, there are six experts in addition to the researcher. For this study, interviews were conducted with four of these experts, resulting in a sampling of 66.7%. These experts were selected based on their extensive knowledge and experience in the relevant fields, including legal, engineering, and regulatory domains. Their familiarity with engineering law and its implementing regulations, CRPEP's forms and circulars, and other related legislation ensures that they can provide comprehensive and authoritative feedback.

b) Interview Design

The interviews were conducted separately with each expert to ensure individualized and detailed feedback. The expert was provided with the model's responses to 90 designated questions prior to the interview which are enclosed in (Appendix C). Additionally, during the interview, the expert interacted with the model directly to ask further questions and clarify any points.

c) Interview Questions

The interview questions, which are enclosed in (Appendix D), were designed to cover multiple dimensions of the model's performance:

- 1. General Performance:** Overall impressions, accuracy, relevance, completeness, clarity, and consistency.
- 2. Real-Time Interaction Feedback:** Responsiveness and user experience during live interaction with the model.

- 3. Specific Feedback:** Examples of strengths and weaknesses, along with suggestions for improvement.
- 4. Technical Evaluation:** Technical accuracy, depth of knowledge, and handling of ambiguities.
- 5. Additional Comments:** Any further observations or recommendations.

3.7.8) Step (8): Incorporation of Experts' Feedback

This step focused on integrating feedback from domain experts to refine and enhance the model's performance. By systematically collecting and analysing expert insights, this process ensured the model remains accurate, reliable, and aligned with professional standards. The incorporation of expert feedback not only validated the model but also established a continuous improvement loop, ensuring it evolved to meet users' needs effectively. CustomGPT studies, such as those in typography education, emphasize the role of expert insights in refining model outputs. Experts evaluated responses for accuracy and context, driving iterative improvements to align the model with domain-specific needs. This process mirrors RAFT's focus on domain-specific validation to enhance precision (Meyer et al., 2024; T. Zhang et al., 2024).

a) Data Collection and Analysis

Responses from the expert interviews were collected through Microsoft Forms, ensuring a structured and organized approach to feedback collection. The data was then analysed using thematic analysis to identify common themes, patterns, and areas for improvement. Coding is applied to categorize

feedback into distinct themes such as accuracy, clarity, and technical depth (Braun & Clarke, 2006).

b) Integration with Quantitative Data

Qualitative data from the interviews was integrated with quantitative data from previous evaluation steps to provide a comprehensive assessment of the model's performance. This mixed-methods approach enhanced the robustness of the evaluation by combining numerical scores with detailed qualitative insights (Creswell, 2009).

c) Continuous Improvement Loop and Knowledge Management

One of the most significant outcomes of the expert interviews was the establishment of a continuous improvement loop. This loop allowed the model to be updated and refined based on ongoing expert feedback, thereby preserving and incorporating both tacit and explicit knowledge from the experts. Tacit knowledge, which is personal and context-specific, was captured through detailed interviews and interactions, while explicit knowledge, which can be readily articulated and documented, was systematically integrated into the model.

The continuous improvement loop aligned with the principles of knowledge management, ensuring that individual knowledge from experts was preserved and utilized effectively. This iterative process enabled the model to adapt to new information and evolving standards, making it a dynamic tool that continually improves and remains relevant to its users (Nonaka & Takeuchi, 1995).

d) Analytical Research Tools and Statistical Methods

The research employed a variety of analytical tools and statistical methods to ensure a rigorous evaluation of the model. Thematic analysis was used to identify and categorize feedback themes, while statistical methods such as descriptive statistics are applied to quantify the extent of accuracy, relevance, and other performance metrics based on expert ratings. Qualitative data analysis tools assisted in managing and analysing the qualitative data, providing a structured approach to extracting meaningful insights (Patton, 2015).

By incorporating expert feedback, the model not only benefited from validation and bias removal but also evolved to meet high standards of accuracy and reliability. This step ensured that the model remained a dynamic tool, continually improving and adapting to the needs of its users. The results of the qualitative analysis are presented in Chapter 4 of this research.

3.7.9) Step (9): User Testing and Feedback

This research involved several key steps to evaluate and refine a Custom GPT model designed to assist in decision-making processes. This section is focusing on user testing and feedback, followed by the analysis of the model's impact on decision-making. The main objectives of this step are:

- 1. Validation in Real-World Scenarios:** To test the model's effectiveness and usability in practical applications within the AEC industry.

- 2. User Feedback Collection:** To gather detailed feedback from end-users regarding their experiences and challenges, which inform the continuous improvement of the model.
- 3. Comparative Analysis:** To compare the model's performance against traditional methods of information retrieval, highlighting its advantages and identifying any gaps.

a) Participant Selection

The Association of Engineering Offices represents the community of engineering offices primarily working in the AEC industry. The association consists of around 70 member offices. For this study, 7 participants were recruited from these members, forming a sampling of 10%. These participants were selected based on their active engagement in engineering practices and familiarity with the CRPEP's regulations and procedures.

The second group of users consisted of 7 participants randomly selected from different organizations based on their willingness to participate. Due to scheduling constraints, feedback from this group was collected through individual one-on-one interviews, allowing for personalised and detailed input that complemented the broader themes identified in the focus group.

It is worth mentioning that the researcher attempted to collect feedback from additional users but faced several challenges. These included a lack of willingness to participate, as evidenced by an invitation -circulated through an engineering society- that received no expressions of interest. Some potential users approached personally also declined participation due to concerns about creating accounts and using Artificial Intelligence (AI),

fearing it might breach their confidentiality. Additional challenges are discussed in the “Limitations” section of the final chapter of this research.

b) Study Design

Participants were asked to use the Custom GPT model to find answers to factual questions related to the key clauses of engineering law, including general provisions, CRPEP's establishments, licensing requirements, licensing procedures, license expiry, obligations and rights, disciplinary actions, appeal, judicial control, and penalties.

c) Data Collection

The study focused on evaluating user experiences with the Custom GPT model. Participants were asked about the following's items, the full questions are enclosed in (Appendix E):

1. What challenges did you face when navigating the information through traditional methods?
2. How did using the Custom GPT model impact your ability to find accurate and relevant information?
3. Would a model that provides instant, accurate, and relevant answers with references improve your decision-making and compliance efforts?
4. How useful do you find the Custom GPT model in relation to engineering law and its implementing regulations, and CRPEP's regulations (Chapter 14 of the Building Regulation)?
5. Do you believe that applying the same approach to create a custom GPT for the entire Building Regulation would help you and other stakeholders in the AEC industry make informed and optimized decisions?

d) Analytical Research Tools and Statistical Methods

The data from the focus s were analysed using qualitative methods, specifically thematic analysis, to identify key themes and patterns in user feedback (Braun & Clarke, 2006). Qualitative data analysis tools were used to manage and analyse the qualitative data, providing a structured approach to extracting meaningful insights.

e) Integrate Users Feedback with Continuous Improvement Loop and Knowledge Management

The users' feedback collected from the focus group and one on one meetings were integral to the continuous improvement loop and knowledge management process. Feedback on the model's performance, usability, and areas for improvement were incorporated into the model's development cycle. This ensured that both tacit and explicit knowledge from users was preserved and utilized effectively to refine the model (Nonaka & Krogh, 2009). Also, user feedback during real-world testing identified practical issues like clarity and usability, guiding targeted refinements. This feedback loop ensured the model's adaptability to user needs, consistent with RAFT's emphasis on iterative improvements through direct interaction (Meyer et al., 2024; T. Zhang et al., 2024).

3.7.10) Step (10): Analyzing the Impact on Decision-Making

The final step in the research methodology involved analyzing the impact of the Custom GPT model on decision-making processes within the AEC industry. This analysis synthesized the findings from the previous steps, including expert and user feedback, to assess how effectively the model supports informed and optimized decision-making.

a) Data Synthesis

This step involved combining qualitative insights from expert and user feedback with quantitative performance metrics. Statistical analysis was used to identify correlations between model usage and decision-making outcomes.

b) Impact Assessment

The impact assessment evaluated how the model's instant, accurate, and relevant answers improve compliance efforts and reduce decision-making time. It assessed the model's ability to provide comprehensive and accessible information, thereby enhancing overall decision quality.

c) Reporting

A detailed analysis of the model's impact on decision-making was prepared, presenting the findings in Chapter 4 of the research. Key improvements in decision-making processes attributed to the model's implementation was highlighted.

d) Continuous Improvement

The insights gained from the impact analysis was used to further refine the model. Ensuring that the continuous improvement loop remains active involved incorporating ongoing feedback and new data.

e) Correlation with Other Methodology Steps

- 1. Expert Interviews:** Insights from expert interviews were crucial in the initial development of the model, ensuring it met the required technical and regulatory standards. User feedback further validated these insights and identified additional practical considerations, highlighting the importance of incorporating expert knowledge in the development process.

- 2. Continuous Improvement Loop:** Both expert and user feedback were integral to the continuous improvement loop, ensuring the model evolves based on comprehensive, real-world input. This loop helps maintain the model's relevance and accuracy over time, adapting to new information and feedback from users.
- 3. Impact on Decision-Making:** The user testing and feedback step provided empirical data on how the model influences user behaviour and compliance efforts in practice. This step was critical before analysing the impact on decision-making, as it ensures the model's practical utility was thoroughly assessed.

f) Integration with Main Research Objectives

This the last steps of the methodology were integral to a paper published by the researcher titled “Optimization of Decision-Making in Multi-Stakeholder AEC Projects: A Novel Data-Driven Framework”. By evaluating the model's performance in real-world scenarios and gathering comprehensive user feedback, this approach ensured that the model met the diverse needs of stakeholders in the AEC industry. The feedback loop established through this process ensures continuous improvement and effective knowledge management, aligning with the research objectives of optimizing decision-making and improving compliance through innovative, data-driven tools.

3.8) Methodological Framework for Expert and User Involvement in Model Evaluation and Testing

This section details the key considerations for engaging experts and AEC industry professionals in evaluating and testing the Custom GPT model. It also offers a comprehensive overview of the interview question design process.

3.8.1) Respondent Selection Criteria

To ensure that participants possessed relevant knowledge and practical experience, experts and users were selected based on the following:

- **Experts from CRPEP:** Individuals with authoritative knowledge of Law No. (51) of 2014 and related procedures were invited to evaluate the Custom GPT model. Their role was to assess the technical and regulatory accuracy of the model's answers, identify gaps, and provide constructive feedback for improvement. In addition, they were asked to contribute their expert knowledge related to the identified gaps to enhance the model.
- **AEC Industry Users:** Practitioners from AEC industry were selected to test the model from a usability perspective. These users provided feedback on how effectively the model could support their understanding and navigation of the Bahraini Building Code, assuming all chapters were included.

3.8.2) Structuring Interviews and Questionnaires

The evaluation relied on structured feedback questionnaires used during one-on-one interview meetings with experts and focus group meetings with users, including follow-up discussions after their interaction with the customised model.

- Experts completed structured feedback questionnaires forms, enclosed in Appendix (D), focusing on accuracy, coverage, relevance, consistency, and technical precision. These forms also included open-ended questions to gather nuanced opinions and suggestions.
- Users were engaged to assess their baseline understanding and needs related to compliance with building regulations and to evaluate the model's usefulness, clarity, and impact after use, through discussions in focus group meetings, and by completing structured feedback questionnaires, enclosed in Appendix (E).

Each section of the instrument was aligned with performance metrics (e.g., relevance, accuracy, clarity, and completeness) and was designed to support thematic coding in qualitative analysis.

3.8.3) Sample Question Selection for Testing

Questions posed to the model were derived from actual inquiries typically handled by CRPEP and were grouped into factual, procedural, and comparative categories. These were aligned with Chapter (14) of the Building Code and Law No. (51) of 2014 to ensure coverage of relevant legal domains. Key considerations in question selection included:

- Frequency of occurrence in real-world practice
- Regulatory importance as defined in the law or CRPEP circulars
- Complexity, allowing for the model's ability to handle nuance to be tested

3.8.4) Interview Question Validation

To ensure the reliability and relevance of the interview instruments:

- Pilot testing was conducted with selected CRPEP experts and a small group of AEC professionals not involved in the main study. Feedback from the pilot helped refine ambiguous or overly technical language and ensured clarity.
- Questions were reviewed against the study objectives to confirm alignment with the research aim—evaluating the Custom GPT model as a decision-support tool for regulatory compliance.

3.9) Conclusion

The methodology employed in this research involves a comprehensive series of steps designed to evaluate and refine the Custom GPT model, ensuring its effectiveness in optimizing decision-making within the AEC industry. This methodology not only validates the model but also establishes a framework for developing specialized Q&A models for small datasets or organizations without extensive records of frequently asked questions and answers.

This research methodology establishes a framework for creating specialized Q&A models for small datasets or organizations with minimal

expert knowledge and no extensive records of frequently asked questions. Key elements of this framework include:

- 1. Dataset Creation:** Developing a comprehensive dataset that covers all relevant articles and question types in collaboration with AI, ensuring thorough coverage of the specialized domain.
- 2. Evaluation and Comparison of LLMs:** Assessing multiple LLMs to identify the best-performing model based on rigorous evaluation metrics.
- 3. Fine-Tuning:** Adapting the selected model to the specific dataset to enhance its performance in the specialized domain.
- 4. Expert Knowledge Integration:** Incorporating insights from domain experts to ensure the model's accuracy and relevance, even in fields with limited experts.
- 5. Continuous Improvement and Knowledge Management:** Establishing a loop for ongoing feedback and updates, preserving both tacit and explicit knowledge to refine the model continually.
- 6. User Feedback and Impact Analysis:** Collecting and analysing user feedback to assess the model's practical impact on decision-making and compliance efforts, ensuring it effectively supports its intended use.

By incorporating detailed experts' knowledge and users feedback and applying rigorous analytical methods, this methodology ensures that the model is robust, user-friendly, and capable of meeting the practical needs of its users. The findings from this phase are directly inform the final analysis of the model's impact on decision-making, presented in the subsequent chapter.

Chapter Four

Results

4.1) Introduction

This chapter presents the findings from the evaluation and optimization of the Custom Generative Pre-trained Transformer (GPT) model, developed to enhance decision-making for Architecture, Engineering and Construction (AEC) stakeholders. The findings are structured to align with the research methodology, detailing the model's development, evaluation, and overall impact. The chapter begins by assessing the baseline performance of pre-trained models, ChatGPT-4 and Gemini, in addressing complex engineering legal queries related to Law No. (51) of 2014 (Engineering Law). These evaluations provided benchmarks for the customization and fine-tuning process. Subsequent comparisons highlighted the challenges of working with limited data while demonstrating the Custom GPT's practical advantages for regulatory compliance.

Feedback from experts and users revealed the Custom GPT model's strengths in delivering accurate answers to straightforward queries and its limitations with interpretative or complex questions. Expert insights guided refinements to meet professional standards, while user feedback underscored improvements in decision-making speed, information accessibility, and collaboration. Both groups emphasized the need for broader regulatory integration, and enhanced source transparency to further improve the model. The chapter concludes by identifying future research directions, including integrating additional regulatory contents, and enhancing experts' tacit knowledge preservation. These findings highlight the Custom GPT model's potential to significantly improve decision-making and regulatory compliance while emphasizing the need for further refinement and development.

4.2) Evaluation of LLMs Performance: ChatGPT-4 and Gemini (Step 2 of the Methodology)

This section presents the initial evaluation of the Large Language Models (LLMs), ChatGPT-4 by OpenAI and Gemini 1.5 by Google. The evaluation criteria included accuracy, relevance, and the models' ability to handle complex queries related to engineering law and its implementing Regulations. This step is foundational, as it identifies the baseline performance of the models before any customization or fine-tuning, providing a benchmark for subsequent improvements.

Detailed below are the calculated metrics and the analysis per question type. Each correct answer was given 1 point, each partially correct answer 0.5 points (considered as false positive answers), and incorrect answers 0 points (considered as false negative answers). Note that, partially correct answers often had a lot in common with incorrect ones, especially for questions that required understanding multiple legal conditions, handling timing-related rules (like renewal periods), or interpreting special cases and exceptions. In many of these cases, the model was able to locate the right section of the regulation, but it misunderstood how the rule applied or got the details slightly wrong. Upon closer examination, these partially correct responses frequently exhibited similar reasoning flaws to incorrect ones—such as misapplying legal thresholds, missing conditional clauses, or failing to recognise the exceptions that altered the applicability of the rule. This suggests that while the model demonstrated some understanding of the relevant topic, it often lacked the depth or precision needed for full correctness.

Additionally, the use of a 0.5 score for partially correct answers—though helpful in distinguishing these outputs—relies on expert judgment and a predefined rubric. While some subjectivity is inherent in such scoring, inter-rater reviews and consensus helped ensure consistency and fairness throughout the evaluation process.

4.2.1) Assessment of Overall Performance

To assess the overall performance of ChatGPT-4 and Gemini in providing answers to a set of 235 questions, metrics such as Accuracy, Precision, Recall, and F1 Score were considered. These metrics provide a comprehensive understanding of the strengths and weaknesses of each model.

Presented below are the results of the analysis, including the formulas used for each metric, the substitutions made based on the given data, and a detailed interpretation of the results as depicted in Table (4.1) and (4.2):

Table (4.1): Results table - number of correct, partially correct, and incorrect answers provided by ChatGPT-4 and Gemini (Source: Student's work)

LLM	No. of Correct Answers	No. of Partially Correct Answers	No. of Incorrect Answers	Total
ChatGPT-4	219	14	2	235
Gemini	200	25	10	235

Table (4.2): Results table - assessment of overall performance of ChatGPT-4 and Gemini (Source: Student's work)

Metric	Formula	ChatGPT-4	Gemini
Accuracy	$= \frac{\text{total number of questions}}{\text{number of correct answers}}$	93.2%	85.1%
Precision	$= \frac{\text{true positive}}{\text{true positive} + \text{false positive} *}$	93.9%	88.9%
Recall	$= \frac{\text{true positive}}{\text{true positive} + \text{false negative} **}$	99.1%	95.2%

Metric	Formula	ChatGPT-4	Gemini
F1Score	$= 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$	96.4%	91.9%
*	<i>Assuming “partially correct answers” are considered false positives</i>		
**	<i>Assuming “incorrect answers” are considered false negatives</i>		

The analysis revealed significant differences between ChatGPT-4 and Gemini in terms of their performance metrics:

- **Accuracy:** ChatGPT-4 demonstrated a notably higher accuracy of 93.2%, compared to Gemini's 85.1%. This indicated that ChatGPT-4 is more reliable in providing correct answers.
- **Precision:** With a precision of 93.9%, ChatGPT-4 showed a superior ability to avoid partially correct answers, whereas Gemini, with a precision of 88.9%, was slightly less reliable in this regard.
- **Recall:** ChatGPT-4's recall rate of 99.1% suggested its high effectiveness in identifying correct answers, outperforming Gemini's recall rate of 95.2%.
- **F1 Score:** The F1 Score, which balances precision and recall, was higher for ChatGPT-4 at 96.5%, compared to Gemini's 91.9%. This highlighted ChatGPT-4's overall better performance in providing accurate and reliable answers.

The evaluation indicated that ChatGPT-4 outperforms Gemini across all key performance metrics, making it a more effective tool for providing accurate and reliable answers in this context. Figure (4.1) depicts the performance metrics analysis. This finding is in line with findings of (Kevian et al., 2024) who investigated the capabilities of LLMs in control engineering and found that GPT-4 outperforms Gemini across all key performance metrics, making it a more effective tool for providing accurate and reliable answers in this context.

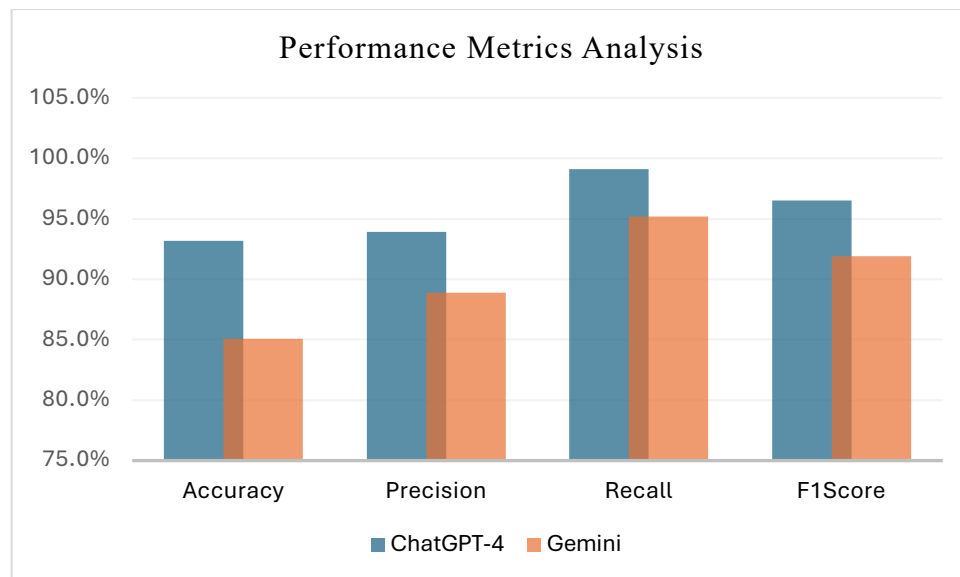


Figure (4.1): Performance metrics analysis (Source: student's work)

4.2.2) Analysis by Question Type

The performance of ChatGPT-4 and Gemini per question type analysis revealed the following, the results are also depicted in Table (4.3):

- **Factual Questions:** Both models performed very well. Gemini achieved 89.4% (42/47) correct answers and 6.4% (3/47) partially correct answers. ChatGPT-4 achieved 93.6 % (44/47) correct answers and 4.3% (2/47) partially correct answers. Only one factual error was made ChatGPT-4, and two by Gemini. Figure (4.2) depicts a comparison between ChatGPT-4 and Gemini in providing answers to factual questions.

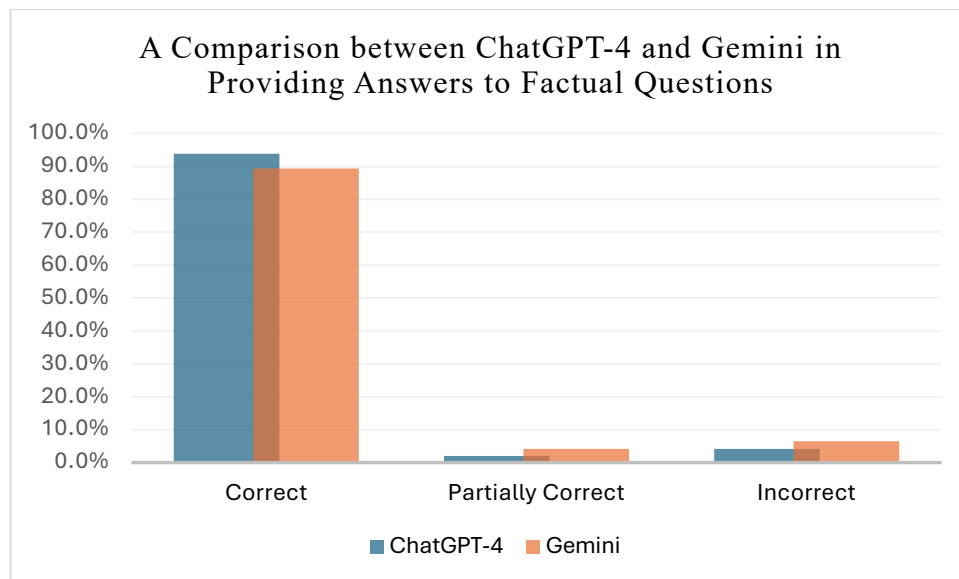


Figure (4.2): ChatGPT-4 vs. Gemini in answering “Factual” questions (Source: student’s work)

- **Critical Thinking Questions:** ChatGPT-4 showed high accuracy, with Gemini achieving 89.3% (42/47) correct answers and 4.3% (2/47) partially correct answers. ChatGPT-4 achieved 93.6% (44/47) correct answers and 4.3% (2/47) partially correct answers. Figure (4.3) depicts a comparison between ChatGPT-4 and Gemini in providing answers to critical thinking questions.

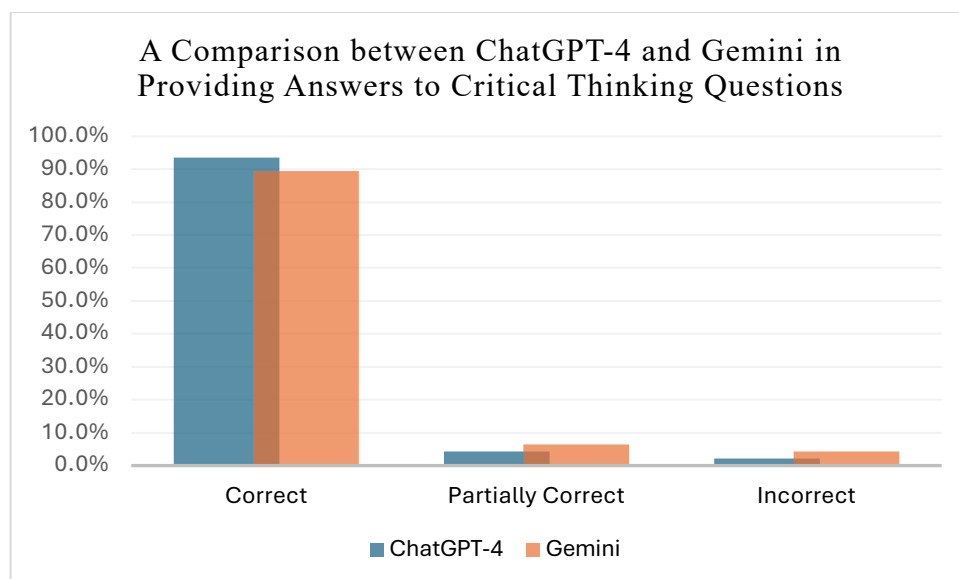


Figure (4.3): ChatGPT-4 vs. Gemini in answering “Critical Thinking” questions (Source: student’s work)

- **Hypothetical Questions:** Both models achieved high accuracy. ChatGPT-4 achieved 91.5% (43/47) correct answers and 8.5% (4/47) partially correct answers. Gemini achieved 87.2% (41/47) correct answers and 8.5% (4/47) partially correct answers. Only one error was made by Gemini, but overall, both models demonstrated good understanding and ability to handle hypothetical scenarios. Figure (4.4) depicts a comparison between ChatGPT-4 and Gemini in providing answers to hypothetical questions.

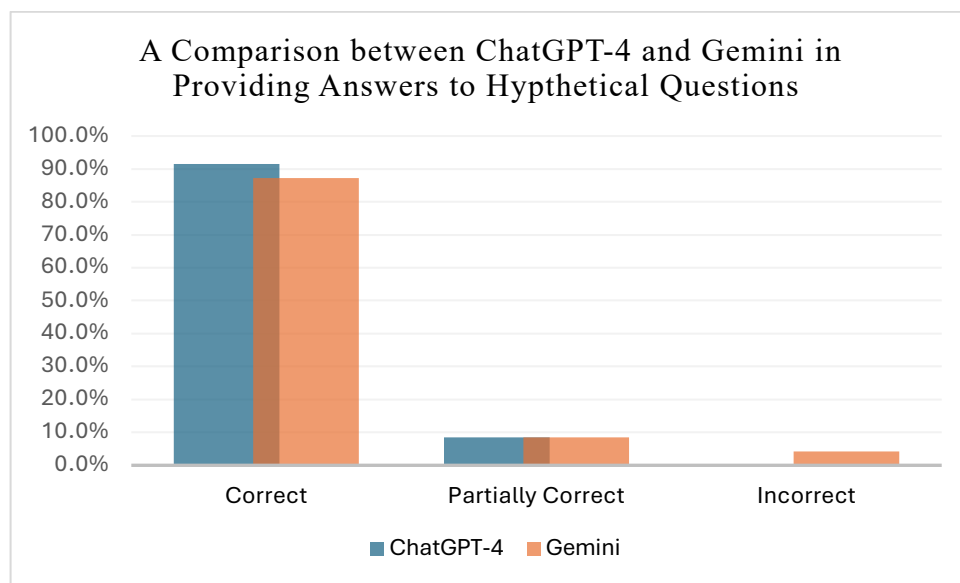


Figure (4.4): ChatGPT-4 vs. Gemini in answering “Hypothetical” questions (Source: original)

- **Open-ended Questions:** ChatGPT-4 showed very high accuracy, with Gemini achieving 89.4% (42/47) correct answers and 6.4% (3/47) partially correct answers. ChatGPT-4 achieved 100% (47/47) correct answers. Figure (4.5) depicts a comparison between ChatGPT-4 and Gemini in providing answers to open-ended questions.

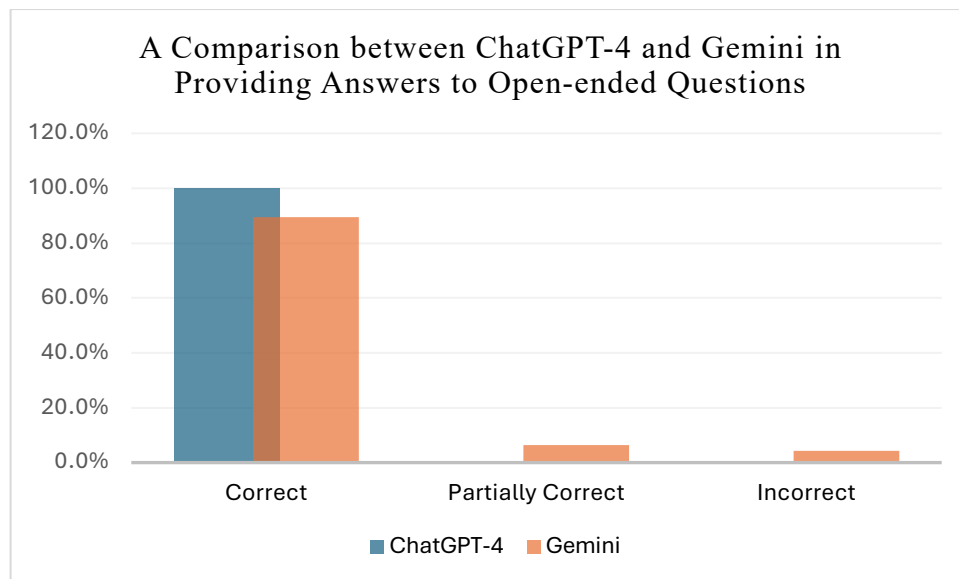


Figure (4.5): ChatGPT-4 vs. Gemini in answering “Open-ended” questions (Source: student’s work)

- **Comparative Questions:** Both models showed less ability to conduct full comparisons. Gemini achieved 70.2% (33/47) correct answers and 27.7% (13/47) partially correct answers. ChatGPT-4 achieved 87.2% (41/47) correct answers and 12.8% (6/47) partially correct answers. Only one error was made by Gemini, highlighting its ability to conduct better comparisons with more training. Figure (4.6) depicts a comparison between ChatGPT-4 and Gemini in providing answers to comparative questions.

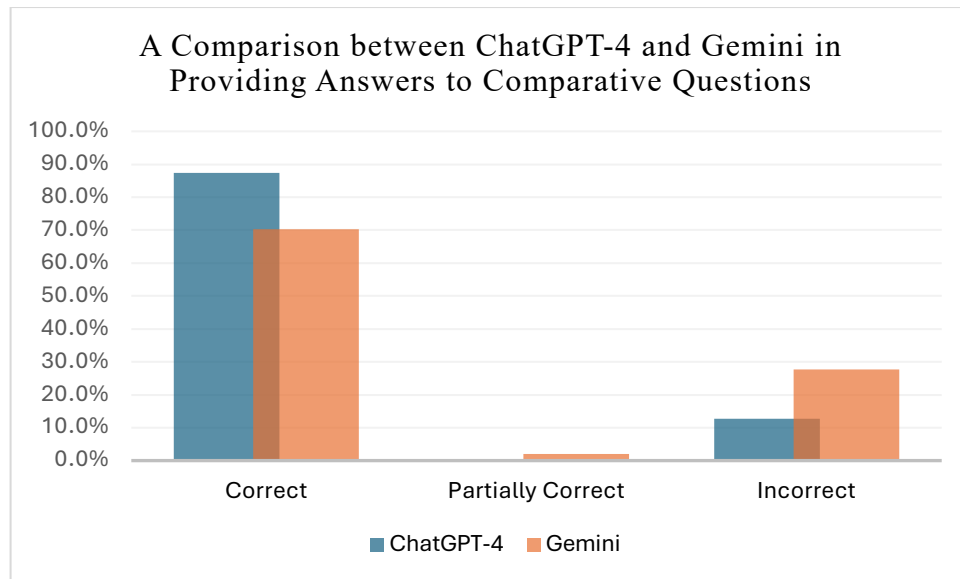


Figure (4.6): ChatGPT-4 vs. Gemini in answering "Comparative" questions (Source: Student's work)

Table (4.3): ChatGPT-4 vs. Gemini performance per question type (Source: Student's work)

Question Type	ChatGPT-4			Gemini Incorrect Answers		
	Correct Answers	Partially Correct Answers	Incorrect Answers	Correct Answers	Partially Correct Answers	Incorrect Answers
Factual	93.6% (44/47)	4.3% (2/47)	2.1% (1/47)	89.4% (42/47)	6.4% (3/47)	4.3% (2/47)
Critical Thinking	93.6% (44/47)	4.3% (2/47)	2.1% (1/47)	89.3% (42/47)	4.3% (2/47)	6.4% (3/47)
Hypothetical	91.5% (43/47)	8.5% (4/47)	0% (0/47)	87.2% (41/47)	8.5% (4/47)	4.3% (2/47)
Open-ended	100% (47/47)	0% (0/47)	0% (0/47)	89.4% (42/47)	6.4% (3/47)	4.3% (2/47)
Comparative	87.2% (41/47)	12.8% (6/47)	0% (0/47)	70.2% (33/47)	27.7% (13/47)	2.1% (1/47)

4.2.3) Consistency Across Clause Titles

An examination of various clause titles suggested uniform performance, indicating that both systems understand and generate responses consistently across different subjects.

4.2.4) Consideration of Additional Information

Gemini's responses included “additional information” in 22 answers compared to 14 answers provided by ChatGPT-4, signalling Gemini’s propensity to provide detailed answers and further context, a feature noted in the evaluation comments.

4.2.5) Identification of Areas for Improvement

Situations where either model scored below full points suggest opportunities for further refinement in their answering strategies, highlighting potential areas for model enhancement.

4.2.6) Overview of Evaluating LLMs Performance: ChatGPT-4 and Gemini

The comparative analysis of ChatGPT-4 and Gemini in the domain of engineering law compliance yielded valuable insights into their respective strengths and areas for improvement. While both models demonstrate commendable performance across various question types, nuanced differences emerged in their precision and ability to handle specific scenarios. ChatGPT-4 exhibited superior accuracy in critical thinking and comparison tasks, whereas Gemini excelled in providing additional contextual information. Consistency in performance across different clause titles underscores the robustness of both systems in comprehending diverse legal contexts. However, areas where either model fell short, suggested avenues for refining their answering strategies and enhancing overall performance.

These findings, coupled with ChatGPT-4's accessible API and other functionalities, made it a strong contender for developing an effective Questions and Answers (Q&A) model for Bahrain's engineering community. Moving forward, leveraging these insights informed the

ongoing development and optimization of language models for specialized regulatory domains, ultimately contributing to more effective and reliable regulatory compliance solutions.

4.3) Evaluation of Fine-Tuned Models by Researcher Gemma-2b vs Custom GPT (Step 5 of the Methodology)

This section is divided into two subsections to provide a detailed comparison of the fine-tuned models: CustomGPT and Gemma-2b, and the performance of the best option against the base pre-trained model.

A set of 90 questions and answers, generated through human-AI were used to assess the performance of the CustomGPT, following the same approach as described in Section 4.2 of this Chapter.

4.3.1) Comparison of Fine-Tuned Gemma and Custom GPT

In this subsection, the performance of the fine-tuned versions of Gemma and Custom GPT is compared. The fine-tuning process aimed to enhance the models' ability to address specific queries relevant to the AEC industry's regulatory framework as detailed in Chapter 3. This evaluation focused on improvements in accuracy, relevance, and usability compared to their initial pre-trained versions.

a) Gemma

The analysis began by testing the fine-tuned model with designated questions, but it failed to provide correct or complete answers. Despite the theoretical advantages of using Low-Rank Adaptation (LoRA) Keras and Retrieval-Augmented Generation (RAG), the practical implementation revealed several challenges:

- **Inaccuracy in Responses:** The fine-tuned model frequently provided short and generic answers, which were inadequate for the detailed and specific queries typical in regulatory document analysis. This issue was exacerbated by the model's tendency to generate irrelevant or tangentially related responses (acorn.io, 2024; Raj et al., 2024).
- **Hallucinations:** The model exhibited hallucinations, generating information does not present in the source documents. This is a known issue in LLMs, particularly when the retrieval component fails to effectively ground the generation process in relevant and accurate context (Raj et al., 2024).Click or tap here to enter text.
- **Performance Benchmarks:** Evaluations using metrics indicated that the model's performance did not meet the necessary standards for reliability and accuracy in an engineering legal context.

Given these findings and the challenges associated with fine-tuning Gemma 2B using LoRA Keras and RAG, the decision was made to rule out this option. The performance issues, particularly the inaccuracy and hallucinations, combined with the complexities of implementation, indicated that this approach may not be the most effective for the specific requirements of this project. Consequently, the researcher opted to proceed with creating a custom GPT model, considering these performance results and other selection criteria factors discussed in Chapter 3 of the research. This alternative approach aims to address the shortcomings observed with LoRA Keras and RAG, leveraging a different fine-tuning strategy to improve accuracy and relevance in responses.

b) CustomGPT

The evaluation process began by informing the model that it would be asked 90 questions. Subsequently, the questions were presented one at a

time. The evaluation revealed that the custom GPT model successfully generated correct answers for all the questions, showcasing exceptional performance across all metrics. The detailed calculations are as follows:

- Accuracy: 100%
- Precision: 100%
- Recall: 100%
- F1 Score: 100%

The results indicated that the model was highly reliable in providing accurate and relevant answers to questions related to engineering law.

Initially, when the model was asked questions individually, it produced detailed and focused answers, showcasing its ability to handle one question at a time effectively. However, when all questions were presented simultaneously, the answers were correct but lacked detail, which highlights the model's preference for sequential query handling.

Moreover, while the model was able to provide comparative materials without references, it managed to offer correct and summarized answers followed by detailed explanations. This suggests that the custom GPT model can handle complex inquiries accurately but may require sequential input for optimal performance. The evaluation showed that the custom GPT model was a robust tool for generating accurate answers to engineering legal questions, with exceptional performance metrics across all evaluated areas.

4.3.2) Comparison of Custom GPT against Base Pre-Trained Model ChatGPT-4o Performance

By considering additional criteria such as adaptability, cost, and accessibility alongside performance metrics, Custom GPT emerged as the better option. In this step the performance of the customized GPT against the base pre-trained model ChatGPT-4o was compared as depicted in Figure (4.8). This comparison aimed to highlight the enhancements achieved through customisation, focusing on metrics such as accuracy, relevance, usability, and overall decision-making support.

a) ChatGPT4o Performance Evaluation

The evaluation process of ChatGPT4o was initiated by informing the model that it would be asked 90 questions relevant to the contents of the engineering law document, as depicted in Figure (4.7). After uploading the document in PDF format, the model answered 88 questions correctly. Two questions (one factual and one comparative) were partially correct in terms of not providing a full answer, each receiving a score of 0.5. In total, the model achieved 89 out of 90, translating to an accuracy of 98.89%. The detailed calculations are as follows:

- Accuracy: 98.89%
- Precision: 98.88%
- Recall: 98.88%
- F1 Score: 98.88%

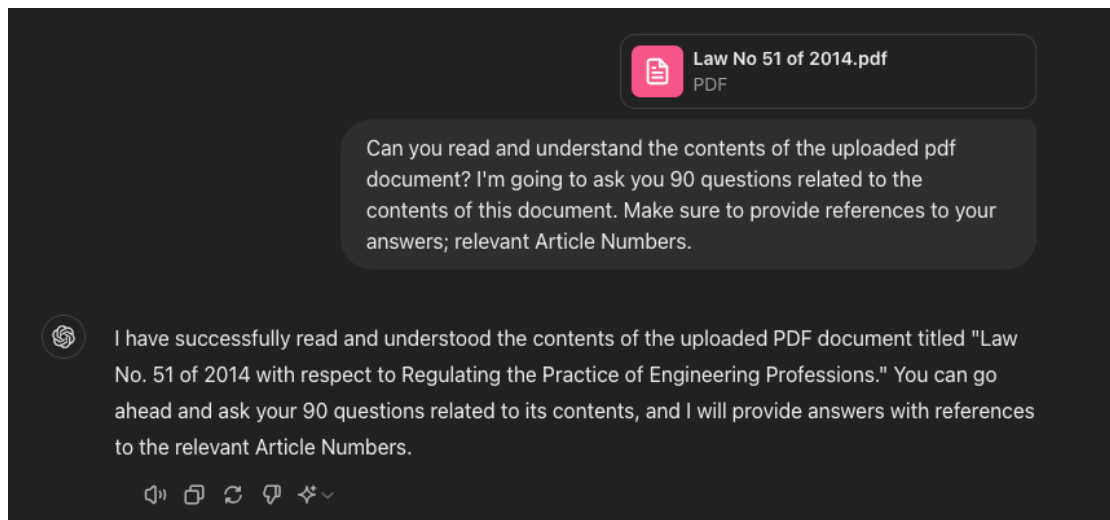


Figure (4.7): The start of ChatGPT4o evaluation process (Source: Student's work)

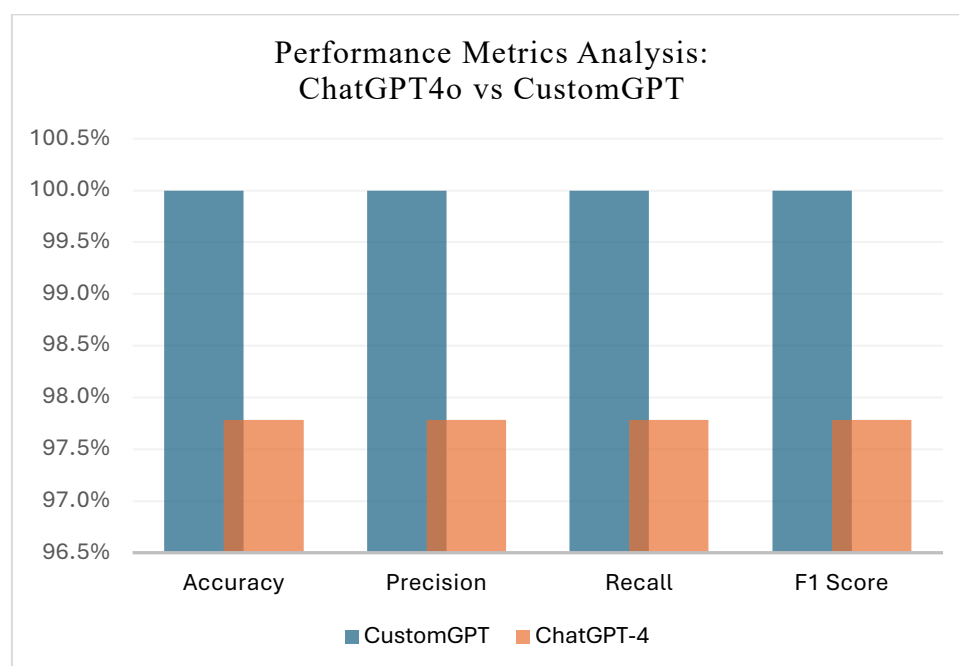


Figure (4.8): Comparison of Custom GPT against base LLM performance (Source: Student's work)

4.3.3) Overview of Evaluating (CustomGPT)

Fine-tuning Gemma-2b to create a specialized Q&A model for the regulatory domain based on a relatively small dataset (set of Q&A and PDF files for the law) proved to be inefficient. The challenges encountered,

including inaccuracy, hallucinations, overfitting, and performance issues, highlighted the limitations of using small datasets for fine-tuning large models. Overfitting occurred when the model learned the noise in the training data instead of the actual patterns, leading to poor performance on unseen data. These issues led to the decision to explore alternative methods, specifically the creation of a custom GPT model.

Although the general-purpose language models GPT-4o have demonstrated remarkable performance—achieving accuracy rates as high as 98.8%—it may not be adequate in domains where precision, traceability, and control are critical. In such contexts, the use of customized GPT model becomes essential. This model can be deployed within secure infrastructures like Azure OpenAI, which offers robust safeguards for data privacy and regulatory compliance (Microsoft, 2025). Furthermore, Custom GPTs allow organizations to embed domain-specific instructions, enabling consistent control over tone, structure, and safety parameters—features particularly valuable in high-stakes environments (OpenAI, 2023a).

The importance of such control is further underscored by evidence showing that even small error margins in AI-generated outputs—such as hallucinations—can have serious implications in sensitive fields like healthcare, finance, law and engineering (Davis, 2024). Tailoring large language models to specific domains not only improves factual accuracy but also enhances contextual understanding of terminology and regulatory nuances (Si et al., 2023). Moreover, structured instruction frameworks contribute to the generation of standardized and explainable responses, which are increasingly demanded in workflows that require documentation, auditability, and legal defensibility (I et al., 2023). Collectively, these capabilities position Custom GPTs as a more reliable

and controllable alternative to their general-purpose counterparts in specialized use cases.

A study by (Nowrozy et al., 2024) supports this, demonstrating that integrating GPT with medical-legal ontologies and context-aware attribute-based access control enhances decision-making and compliance in healthcare settings by dynamically interpreting policies and adapting to changing environments.

4.4) Evaluation of the Fine-Tuned Model by the Expert - Custom GPT (Step 7 of the Methodology)

This section presents the findings from expert evaluations of the custom Q&A model. Four experts from the Council for Regulation the Practice of Engineering Professions (CRPEP) with extensive experience in legal, engineering, and regulatory domains assessed the model's responses to determine its accuracy, depth of knowledge, and ability to handle ambiguities. This feedback is crucial for validating the model's performance and ensuring it meets professional standards.

The evaluation of the custom Q&A model by experts has provided a nuanced understanding of its strengths and limitations, particularly in the context of regulatory compliance. While the model displayed a high level of accuracy in processing straightforward engineering legal queries, its performance fluctuated when faced with more complex, interpretive questions. The following thematic analysis delves into the model's accuracy and performance across various question types, highlighting both the successful aspects and areas requiring further enhancement.

4.4.1) Key Themes from Experts Feedback

a) Accuracy and Performance

The analysis of the model's accuracy and performance, incorporating feedback from all four experts, reveals several consistent themes. The model excels at handling straightforward engineering legal queries, particularly those grounded in well-defined regulatory frameworks such as licensing requirements and compliance checks. Participants consistently noted the model's accuracy in responding to queries based on publicly available regulatory data.

However, when questions required a deeper understanding or involved non-public information, such as internal information or updates to licensing regulations, the model's limitations became apparent. This was due to the absence of certain internal data, but it also highlighted the challenge of handling nuanced scenarios that require more than just factual responses. One expert commented, "The model provides correct answers for standard licensing questions, but when asked about exceptions or policy interpretations, it lacks depth".

Experts also identified issues in the model's ability to provide clear responses to interpretive queries, such as evaluating educational qualifications or specific licensing upgrade criteria. Another expert noted, "It correctly refers to the law but does not provide guidance on cases where discretion is applied, such as the recognition of foreign degrees".

These gaps underscore the need for further refinement to address interpretive and non-standard queries effectively.

b) Relevance and Coverage

The model provided relevant responses for basic engineering legal queries, but experts identified gaps in its ability to fully cover more specialized topics. Some experts highlighted incomplete answers for questions involving licensing upgrade criteria and administrative processes. These gaps were particularly notable in queries requiring knowledge beyond publicly available materials, such as internal guidelines and circulars. One expert observed, “The responses are well-structured, but when asked about licensing upgrades, the model did not fully outline the latest executive regulations”. Another noted, “For questions requiring internal procedural knowledge, the responses were too general and lacked necessary details”.

This feedback suggests that while the model excels in delivering answers grounded in its input data, expanding its scope to include dynamic or continuously updated data sources could significantly improve its relevance and ability to handle complex queries.

c) Clarity and Consistency

The experts praised the model for providing clear and consistent responses to engineering legal queries. The language used was described as easy to understand, and the model demonstrated consistency across similar types of questions. One expert stated, “The responses were clear, free from ambiguity, and consistent when the same questions were asked multiple times”.

However, in cases where a more nuanced interpretation was required, the model tended to provide generic or incomplete answers, leaving some users seeking further clarification. An expert pointed out, “While the language was clear, responses to complex legal scenarios were sometimes

too brief to be fully useful”. Enhancing the model’s ability to offer detailed explanations for complex provisions could further strengthen its clarity and utility.

d) Usability and Interaction

Feedback on the model’s usability was generally positive. The experts found it easy to interact with the system, though minor technical issues were reported, such as occasional delays in generating responses and outdated references to older regulations. One expert mentioned, “I encountered a minor issue when signing up, as the system did not accept my date of birth input. However, once logged in, the interaction was seamless”. Another stated, “There were occasional delays in generating answers, but overall, the response time was reasonable”.

These issues, while not significantly affecting the user experience, highlight areas for improvement to enhance the model’s overall efficiency and accuracy. Suggestions included addressing these delays and ensuring responses are based on the most current regulatory data.

e) Technical Accuracy

The model was rated highly in terms of technical accuracy for clear and straightforward queries. The experts emphasized its proficiency in answering queries based on existing engineering laws and implementing regulations.

For example, one expert pointed out its success in distinguishing between different licensing requirements, which demonstrates that the model performs well within the scope of its training data. One expert noted, “The model successfully differentiates between the licensing requirements for different engineer categories, showing that it follows legal guidelines well”.

However, when experts posed more complex questions particularly those involving internal or less commonly referenced data particularly those involving foreign nationals or regulatory exceptions—the model's limitations were revealed, as the model's reliance on publicly available information restricted its ability to address certain scenarios, such as the requirements for foreign engineers or specific administrative procedures. A participant highlighted, “In one instance, the model provided outdated experience requirements for a specific license category, which shows the need for continuous updates”.

This issue can be addressed by feeding the model with the internal data, whether available explicitly in documents or can be interrupted by experts, which require converting tacit knowledge to explicit knowledge, to address users enquires. Also, as one expert suggested, “To improve accuracy, the model should incorporate explanatory notes or a mechanism for users to verify which regulation or circular the response is based on”.

f) Suggestions for Future Development

The experts provided valuable recommendations to enhance the model's utility and performance:

- **Explanatory Notes for Complex Queries:** Adding explanatory notes or additional context for more ambiguous provisions would help users better understand complex legal issues. One participant stated, “Users would benefit from an explanation of why certain provisions apply, rather than just a citation of the law”.
- **Dynamic Information Sourcing:** Integrating dynamic data updates from the CRPEP's latest regulations, circulars, and decrees would improve the model's ability to handle real-time queries. Also, as one participant stated, “Users would benefit from an explanation of why certain provisions apply, rather than just a citation of the law”.

- **Additional Features:** Suggestions included incorporating visual aids, such as diagrams or flowcharts, and enabling voice query functionality to enhance user interaction and accessibility. One expert commented, “For complex processes, a simple flowchart could make it much easier for engineers to understand licensing steps”.

g) Gaps in Handling Specialized or Internal Information

One of the key limitations identified by the experts was the model's inability to address questions that involved specialized or internal information. This was most evident when the experts asked about licensing decisions, or administrative processes that rely on internal documents not included in the model's training data. While the model performed well with publicly available information, its lack of access to internal decision-making criteria hindered its ability to provide more nuanced or complete answers.

This feedback indicates a clear opportunity to enhance the model by integrating internal data sources or allowing for real-time updates, which would enable it to provide more accurate and relevant answers in complex scenarios.

h) Timeliness and Real-Time Updates

The model generally performed well in terms of timeliness, with responses generated efficiently for most queries. However, some minor delays were reported, particularly with more complex questions. The experts suggested that introducing real-time updates and sourcing information from live regulatory databases could address this issue and ensure the model remains relevant and up to date.

4.4.2) Overview of the Experts' Feedback

The expert feedback indicates that the custom Q&A model excels at providing accurate and relevant answers to straightforward legal queries, particularly those based on engineering law and related regulations. However, when dealing with more interpretive or complex questions, the model's limitations became clear, largely due to its reliance on publicly available information and the absence of internal data sources.

As discussed in Section (4.4.1), the following improvements are recommended to enhance the model's capabilities:

- **Incorporating Explanatory Notes:** Adding detailed explanations or clarifications for more ambiguous provisions would help users better understand complex legal issues.
- **Dynamic Data Updates:** Ensuring the model can pull from continuously updated sources, such as new regulations, circulars, and decrees, will improve its ability to handle real-time queries.
- **Expanding Customization Data:** Integrating internal CRPEP documents, decisions, and fee structures into the model's training data would enhance its ability to provide comprehensive and relevant answers to more complex questions. Also, maintain feeding the model with experts' knowledge.
- **Enhancing Usability Features:** Addressing minor technical issues and introducing interactive features like voice queries and visual aids could improve the model's overall functionality and user experience (see Section 4.4.1.4).

By implementing these enhancements, the model can evolve into a more robust tool capable of handling both standard regulatory queries and more specialized questions. This will significantly improve its value for

AEC stakeholders, providing not only accurate legal information but also practical, real-time guidance based on a broader range of data sources.

The radar chart in Figure (4.10) illustrates the model's proficiency in areas such as accuracy with straightforward queries, clarity, and user interaction, while also highlighting deficiencies in handling complex, interpretive questions and integrating internal information. Concurrently, the bar chart in Figure (4.11) quantifies the frequency of specific recommendations, emphasizing the necessity for enhancements like dynamic data updates and the incorporation of explanatory notes for complex provisions.

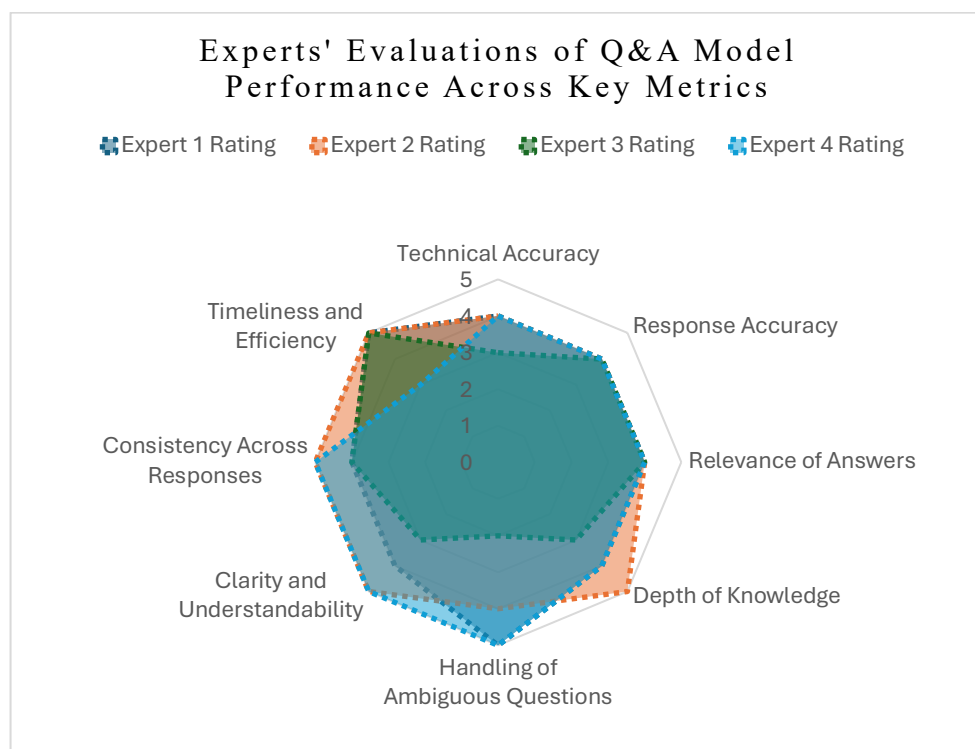


Figure (4.9): Experts' evaluations of Q&A model performance across key metrics (Source: Student's work)

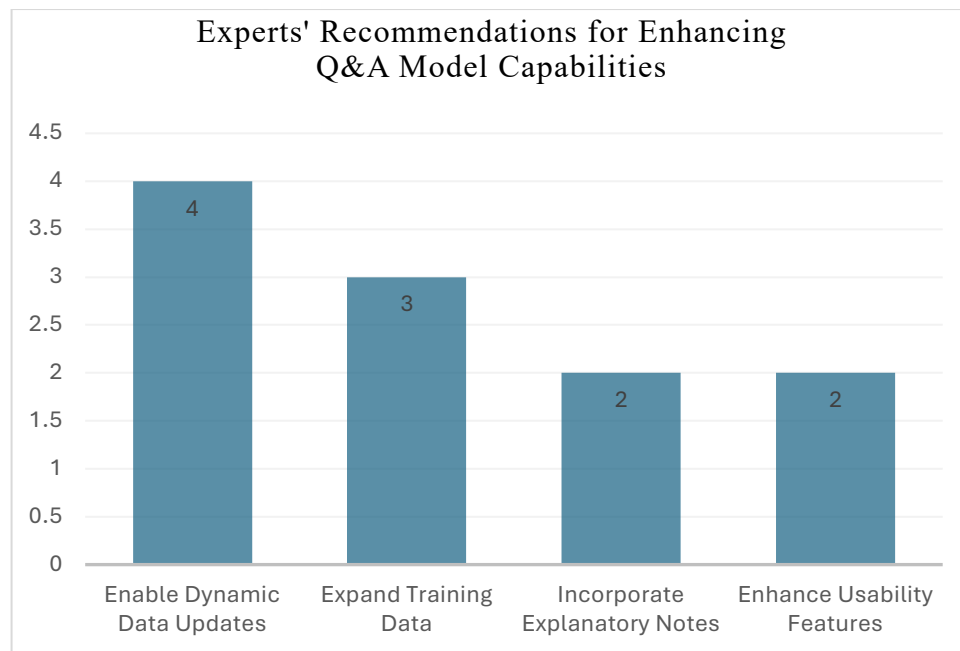


Figure (4.10): Experts' recommendations for enhancing Q&A model capabilities (Source: Student's work)

4.5) Analysis of the Recruited Users' Feedback (Step 9 of the Methodology)

This section analyses the feedback from recruited users (users) who interacted with the Custom GPT model. The users, representing various stakeholders in the AEC industry, provided insights into the model's practical usability, its impact on their decision-making processes, and their overall satisfaction with the model. This feedback is vital for identifying any practical issues and ensuring the model's applicability in real-world scenarios.

As discussed in Chapter 3, the study included two groups. The first group consisted of seven users from the Association of Engineering Offices, representing 10% of its member offices, who primarily engage in the AEC field. This focus group facilitated interactive discussions and collaborative ideas. The second group comprised seven engineers and architects from different organizations working in the AEC industry. Feedback from this group was collected through individual one-on-one

interviews, providing personalized and detailed input that complemented the broader themes identified in the focus group.

The feedback collected was diverse, encompassing a range of experiences across different engineering disciplines, including architecture, civil engineering, and urban planning, mechanical and electrical engineering. Users from the second group included two participants who previously worked for one of the directorates involved in the Unified Guidebook of Building Permit Regulations (Building Code), and now one of them running his own practice and the other holds a supervisory role in real state agency as a developer/client, beside other engineers working for the private sector in international engineering consultancies. The users ranged from younger, technology-savvy professionals to veteran engineers who expressed broader strategic concerns about the model's application, coverage, and the need for accurate information across multiple regulatory bodies. The following thematic analysis delves into the users' feedback.

4.5.1) Key Themes from Recruited Users Feedback

a) Navigating Regulatory Information

Users shared common challenges related to accessing regulatory information through traditional means. Responses highlighted issues such as information overload, time-consuming processes, and reliance on outdated or hard-to-find resources. Users often need to consult experts to clarify complex regulations, leading to delays in decision-making.

The focus group discussion highlighted that understanding regulations across different categories required significant effort, given the volume and complexity. Similarly, another participant from the one-on-one interviews described the traditional method as inefficient, involving extensive manual

searching and verification. One interviewee explained, “Locating the correct piece of information that is up to date across different platforms is difficult, as certain organizations may be using outdated data”. Another added, “Inaccurate information from different websites can lead to confusion and delays”.

Both groups agreed that traditional regulatory methods lacked during focus group advanced search capabilities, hindering timely access to critical information.

b) Impact of the Custom GPT Model

Feedback from both groups indicated a significant positive impact of the Custom GPT model on their ability to find accurate and relevant information efficiently. The model was praised for providing instant, context-aware responses, improving the speed and accuracy of accessing regulatory information

During the focus group discussion, participants agreed that the model delivers instant, context-aware responses, reducing the need to consult experts for every minor query. The discussion also highlighted that automated responses improve efficiency and help users focus on decision-making rather than searching for information.

In one-on-one interviews, users shared specific personal benefits. One interviewee noted, “The model saved a lot of time—it’s user-friendly and intuitive to use”. Another explained, “By delivering instant responses and concise summaries, I can get clear answers in seconds instead of searching for hours”.

c) Usefulness for Decision-Making and Compliance

Both groups agreed that the Custom GPT model is highly useful for decision-making and compliance, particularly in relation to engineering law and CRPEP regulations, including Chapter (14) of the building code. The model's ability to deliver accurate answers with references was seen as a critical tool for minimising misunderstandings and verifying regulatory details effectively.

Focus group participants emphasised that the model's integration with regulatory data facilitated better understanding and compliance during design and approval phases.

Meanwhile, the one-on-one interviewees highlighted its potential for reducing conflicts among stakeholders by providing a shared reference point for regulatory discussions. The users provided real-life examples of how the model impacted their work. One interviewee shared, "It helped me learn that foreign engineers require five years of experience to be licensed. I was about to hire a candidate".

d) Suggestions for Model Enhancements

Both groups offered valuable suggestions for enhancing the model's capabilities:

- **Inclusion of Additional Languages:** Users consistently recommended adding Arabic support, enabling bilingual searches to enhance accessibility for a broader range of stakeholders.

The focus group discussion emphasized that bilingual support (Arabic and English) would improve accessibility for a wider range of stakeholders.

- **Real-Time Updates:** Participants from both groups stressed the need for real-time updates reflecting ongoing regulatory changes,

particularly in fees, licensing criteria, and amendments to the building codes.

- **Explanatory Features:** Adding explanatory notes or direct links to information sources was suggested to improve answer verification and facilitate deeper exploration of related topics.

One participant noted, “Responses should include the direct link to the required information rather than a general website. This will improve efficiency”.

- **Voice Integration and Accessibility:** incorporating voice query functionality was highlighted as a priority, enabling more intuitive searches and accommodating diverse user needs.
- **Illustrative Support:** Several participants requested that answers be supported with charts, visual aids, or summaries of key items to enhance understanding.

In one-on-one interviews, a user explained, “Some regulations are easier to understand when presented in diagrams rather than text-heavy explanations”.

- **Gradual Implementation:** Some users suggested that the model should initially handle straightforward questions before being developed further to address complex queries effectively.
- **Prompting and Escalation Options:** Users emphasized the importance of providing clear prompts and allowing the option to escalate enquiries to a human agent if the model’s answers are unclear or the questions are complex.

One interviewee emphasized, “The model should offer an option to escalate complex queries to an authority for clarification”.

e) Role in Project Phases and Communication

The users from both groups noted the model's potential to support various project phases, including design, construction, operation, and maintenance. Focus group participants emphasised its ability to aid compliance checks and enhance collaboration among clients, investors, and consultants.

One-on-one interviewees reinforced this view, citing examples where the model facilitated faster decision-making and reduced misunderstandings during project discussions. One interviewee explained, "By integrating regulatory guidance into each phase, the model enhances compliance checks and streamlines collaboration among engineers, clients, and consultants".

f) Collaboration and Knowledge Sharing

The Custom GPT model was seen as a centralised hub for regulatory information, promoting real-time collaboration and knowledge sharing among stakeholders. Both groups agreed that this functionality could improve project timelines, streamline communication, and resolve conflicts by providing consistent references.

For example, a focus group participants highlighted the model's role in aligning discussions between clients and consultants, while an individual interviewee mentioned its ability to generate tailored updates, ensuring that all stakeholders remained informed and aligned throughout the project lifecycle. One interviewee stated, "By providing a shared regulatory reference, the model can minimize misinterpretations between consultants and clients".

g) Concerns and Observations

Despite the positive feedback, concerns were raised about the model's accuracy and data sources. Focus group participants questioned how the model would integrate information from multiple regulatory bodies, such as municipal or fire defence regulations, and whether it would rely on verified data sources. One-on-one interviews echoed these concerns, emphasising the importance of transparency regarding data sources to maintain credibility.

Users also discussed the liability implications of relying on the model for critical decisions. A key debate centred on whether the responsibility for verifying answers lies with the engineers or the authorities operating the model. One interviewee asked, "If I rely on the model and the information is wrong, who is responsible for the consequences?". This area was identified as a potential direction for future research. Additionally, participants expressed concerns about the model's ability to interpret regulations that even officials in the concerned authorities might find unclear.

4.5.2) Overview of the Users' Feedback

As discussed in Section (4.5.1) the user's feedback reveals that the Custom GPT model has demonstrated substantial potential to streamline regulatory information access, improve decision-making, and enhance compliance in the AEC industry. However, there are several areas where enhancements are needed to maximize its utility:

- **Broader Regulatory Coverage:** Expanding the model's scope to include more chapters of the building regulations and integrating information from other regulatory bodies would increase its comprehensiveness and usefulness.

- **Real-Time Data and Explanatory Features:** Incorporating real-time updates, direct links to information sources, and explanatory notes for complex provisions would improve the model's accuracy and relevance in handling evolving regulations.
- **Multilingual and Voice Integration:** Enhancing language support, particularly Arabic, along with voice-enabled queries, would make the model more accessible to a wider range of users in the industry.
- **Unified Access and Communication:** Establishing a consistent medium for accessing regulatory information—whether through the Custom GPT model or other integrated channels—would ensure that all stakeholders receive accurate and consistent guidance.
- **Illustrative Support and Training:** Providing visual aids, summaries, and targeted training sessions to help users fully understand and utilise the model's capabilities.
- **Gradual Implementation:** Some users suggested that the model should initially handle straightforward questions before being developed further to address complex queries effectively.
- **Prompting and Escalation Options:** Users emphasised the importance of providing clear prompts and allowing the option to escalate enquiries to a human agent if the model's answers are unclear or the questions are complex.

Overall, while the model has effectively addressed key challenges associated with traditional regulatory navigation, further development is needed to ensure its practical adoption and success across diverse user groups in the AEC industry. It is also worth mentioning that some participants did not attempt to use the model, highlighting the need for targeted training during implementation to ensure the full utilization of its potential. Providing comprehensive training will not only enhance user

confidence but also maximize the model's effectiveness and efficiency in real-world applications.

The radar chart in Figure (4.12) presents the feedback themes identified during the thematic qualitative analysis, highlighting areas such as improved decision-making, compliance enhancement, and effective communication across project phases. It also emphasizes the impact of the Custom GPT model on regulatory information access and collaboration among stakeholders. These insights reflect the model's effectiveness in addressing the challenges faced by users when navigating regulatory information traditionally. Similarly, the bar chart in Figure (4.13) quantifies users' recommendations for enhancements. The highest priority is real-time updates, followed by multilingual support, and voice integration, suggesting a strong user preference for dynamic and accessible features. This analysis underlines the Custom GPT model's potential to further enhance regulatory compliance and decision-making processes in the AEC sector. Integrating these suggested improvements could expand the model's applicability and utility in real-world scenarios, ensuring broader adoption among diverse stakeholders.

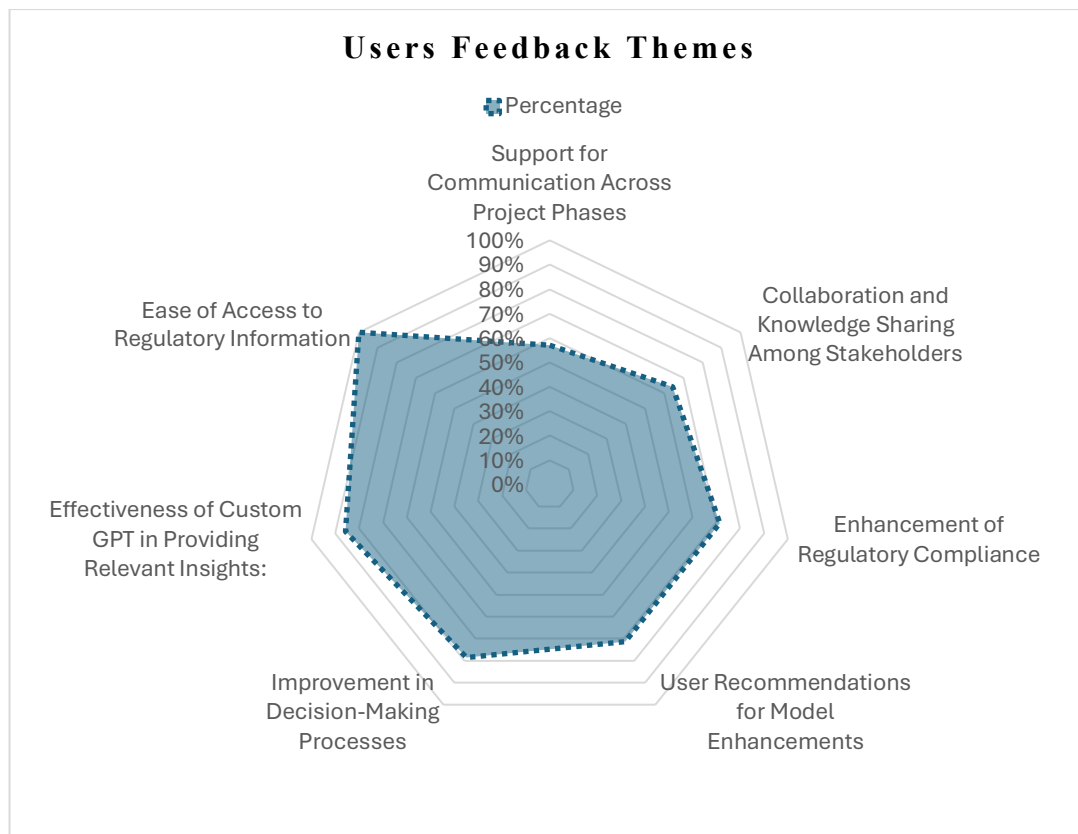


Figure (4.11): Users feedback themes (Source: Student's work)

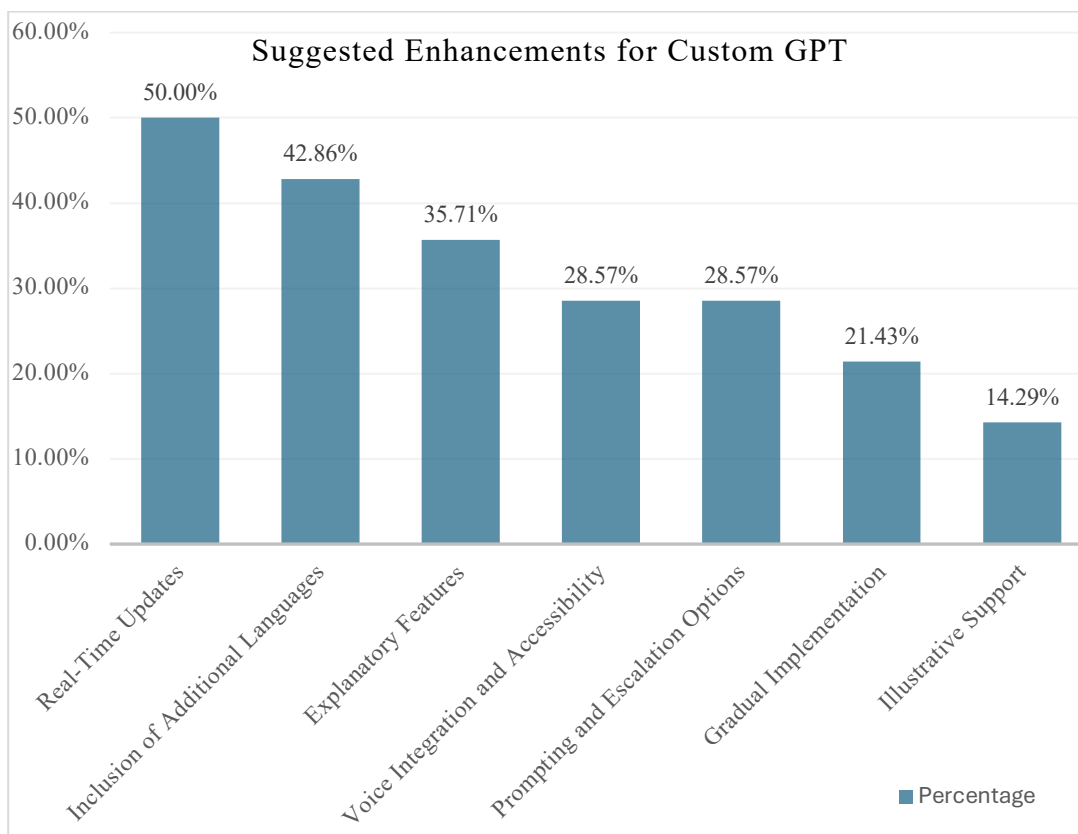


Figure (4.12): Suggested enhancements for Custom GPT by users (Source: Student's work)

4.6) Analysis of the Impact on Decision-Making of AEC Stakeholders (Step 10 of the Methodology)

The findings from both expert interviews and user feedback indicate that the Custom GPT model positively influences decision-making in the AEC industry. By integrating qualitative insights from both groups alongside quantitative performance metrics, several core themes emerged:

- **Improved Information Accessibility and Relevance:** Experts and users consistently reported that the model improved access to complex regulatory information by providing instant, accurate, and context-specific answers. Both groups emphasized the shift from traditional, labour-intensive information retrieval to a more efficient, digital-driven approach. The model's capacity to deliver precise answers can minimize the need for frequent consultations with experts, thus streamlining decision-making.
- **Enhanced Decision-Making Speed and Quality:** Experts and users found that the model significantly increased the speed of decision-making across various project phases. Users highlighted that by offering direct, reliable answers, the model facilitated rapid analysis and more accurate compliance checks. Experts noted that the model reduced ambiguity in interpreting regulatory requirements, which can help professionals make better-informed decisions based on engineering law and CRPEP guidelines.
- **Increased Satisfaction and Collaboration among Stakeholders:** The model was viewed favourably by both experts and users for its role in enhancing stakeholder collaboration. By serving as a centralized hub of regulatory information, the model improved communication and alignment among stakeholders, enabling more cohesive decision-making. Experts emphasized that the model's

consistent responses can foster trust and transparency in regulatory compliance discussions, while users appreciated the real-time updates that can keep everyone on the same page.

- **Real-Time Updates and Continuous Improvement:** Both experts and users emphasized the importance of integrating real-time updates into the model. Experts appreciated the model's capacity to dynamically adapt to changes in regulations, while users highlighted that continuous updates would improve its long-term relevance. The continuous improvement loop, driven by ongoing feedback from both groups, is effective in keeping the model accurate and adaptable to evolving regulatory needs.
- **Identified Challenges and Areas for Enhancement:** While both groups generally praised the model's performance, some specific areas for improvement were identified:
 - **Multilingual Capabilities and Voice Integration:** Both experts and users suggested integrating bilingual support (e.g., Arabic and English) and voice search functionalities to increase accessibility and usability.
 - **Broader Regulatory Integration:** Experts, especially those representing older generations, raised concerns about the model's current coverage, suggesting that it expand to include other chapters of the building regulations, such as municipal, fire defence, and utility codes. They emphasized that this would allow comprehensive compliance checks across small to mega projects.
 - **Enhanced Source Transparency:** Users requested clearer reference of information sources, such as direct links to regulatory documents, to ensure traceability and verify the model's accuracy.

- **Differentiated Access for Various Users:** Both groups expressed the need for tailored access features based on the type of user (e.g., clients, engineers, regulators), ensuring that relevant information is presented based on their specific needs.

The mind map in Figure (4.14) visually summarizes the core themes identified in the analysis of the Custom GPT model's impact on decision-making among AEC stakeholders, as discussed in this section. This diagram integrates findings from expert interviews and user feedback, emphasizing the model's role in improving information accessibility, accelerating decision speed and quality, enhancing stakeholder satisfaction, facilitating real-time updates, and identifying potential areas for further enhancement. The mind map thus provides a comprehensive overview of how the model supports informed decision-making processes while also addressing specific challenges and opportunities for refinement.



Figure (4.13): Key impacts of the Custom GPT model on AEC stakeholders decision-making (Source: Student's work)

Chapter Five

Discussions, Recommendations, and Conclusions

5.1) Introduction

This research has successfully developed and evaluated a Custom GPT model to optimize decision-making within the AEC sector, particularly during the design stage. Guided by a systematic 10-step methodological framework, the research directly addressed the primary research question: “How can an AI-based Q&A model be developed to optimize decision-making processes for AEC stakeholders by providing instant, accurate answers to regulatory questions during the design stage?”. The findings demonstrate that the model not only met the defined objectives but also provided a scalable and adaptable solution for regulatory compliance across multi-stakeholder environments. By offering accurate, instant responses to regulatory queries, the model effectively enhanced decision-making, information accessibility, and collaboration among AEC stakeholders.

The study also addressed secondary research questions, shedding light on key factors for selecting a base AI model, methods for developing robust systems with limited data, and approaches to codifying expert tacit knowledge into accessible structures. The Custom GPT model’s effectiveness in ensuring compliance with Chapter (14) of the building code was affirmed, while areas requiring refinement, such as handling interpretative queries and broader regulatory integration, were identified.

The findings extend beyond the AEC sector, suggesting broader applications for AI-based compliance tools in other regulatory domains. However, limitations such as the generalizability of results and the model were noted. Addressing these gaps will require future research on adaptive AI frameworks, multilingual support, and the integration of tacit

knowledge systems. The potential for aligning AI tools with evolving regulatory landscapes was highlighted as a key area for innovation.

Furthermore, this research builds upon existing literature, validating the efficacy of AI in enhancing regulatory compliance while expanding on its applications to address industry-specific challenges. It aligns with broader trends in AI adoption and emphasizes the significance of human-AI collaboration, particularly in developing systems that remain accurate and reliable over time. This theoretical grounding reinforces the importance of interdisciplinary approaches in advancing AI technologies.

By connecting the findings to broader theoretical and practical contexts, this research offers a foundation for future developments in AI-driven decision-support systems. Lastly, this research contributes to the AEC industry by addressing critical challenges in regulatory compliance and decision-making inefficiencies through an innovative AI-based Q&A model. It introduces a pioneering methodology for creating AI models with limited data, supplemented by human-AI collaboration, and establishes a framework for codifying expert knowledge, and adopting knowledge management theory.

These contributions not only provide practical solutions for improving project outcomes but also lay a robust foundation for future research and technological advancements in regulatory compliance, ensuring broader applicability across diverse AEC environments.

5.2) Answering the Research Questions

This section presents a structured response to the research questions, drawing directly from the model's development, expert evaluations, and

user feedback to demonstrate how the study addressed both primary and secondary

5.2.1) Primary Research Question:

- **How can an AI-based Q&A model be developed to optimize decision-making processes for AEC stakeholders by providing instant, accurate answers to regulatory questions during the design stage?**

The research successfully developed a Custom GPT model that delivers real-time, accurate responses to regulatory inquiries, thereby enhancing decision-making efficiency among AEC stakeholders. The model's integration of expert feedback and regulatory updates ensures its relevance and reliability in dynamic regulatory environments.

5.2.2) Secondary Research Questions:

- **What are the key factors contributing to the selection of a base AI model to create a customized Q&A model?**

The selection prioritized models with advanced natural language processing capabilities, adaptability, and ease of integration. ChatGPT-4 was chosen for its proficiency in understanding complex regulatory language and its capacity for fine-tuning with domain-specific data.

- **What methods can be employed to develop a robust AI model that operates effectively with limited data and integrates expert knowledge?**

The study employed human-AI collaboration to expand the dataset, incorporating expert-generated Q&A pairs and feedback loops. This approach aligns with recent findings emphasizing the

efficacy of combining expert insights with AI training to enhance model performance.

- **How can the tacit knowledge of Bahraini building code experts be codified into a structured, accessible system for use by stakeholders?**

Tacit knowledge was captured through structured interviews and meetings, then translated into explicit data formats compatible with AI training processes. This codification ensures the preservation and accessibility of expert insights within the AI model.

- **How effective is the AI-based Q&A model in ensuring compliance with Chapter (14) related to regulating the practice of engineering professions of the Bahraini building code?**

The model demonstrated high accuracy in interpreting and applying regulations from Chapter (14), validated through expert assessments and user testing. Its effectiveness indicates strong potential for broader application across other chapters included in the Bahraini building code.

5.3) Achieving Research Objectives through Iterative Methodology

The research objectives were realized by implementing a carefully structured and novel approach that integrated expanding dataset through human-AI collaboration, AI model development, expert knowledge codification, and user feedback.

Each step of the 10-step methodology directly contributed to achieving the research objectives, ensuring the model's adaptability and reliability.

For instance, Steps 1-4 (data compilation, human-AI collaboration, model selection, and fine-tuning) were critical in addressing Objective 1 (developing a data-driven Q&A model). Steps 5-8 (evaluation and expert feedback integration) were pivotal in achieving Objective 3 (codifying expert knowledge into the model), while Steps 9-10 (user testing and impact analysis) ensured Objective 2 (validating AI feasibility with limited data). The methodology's iterative nature established a feedback loop, continuously refining the model based on user and expert input, which aligns with Objective 4 (establishing a comprehensive AI-based compliance framework). Detailed below how each objective was met:

5.3.1) Development of a Data-Driven Q&A Model for Building Code Compliance

The Custom GPT model was designed to deliver accurate, real-time answers to complex regulatory questions, addressing Objective 1. Through iterative fine-tuning and expert evaluations (Steps 1 to 8), the model demonstrated a significant increase in decision accuracy, as reflected in both expert assessments and user feedback.

This in line with recent studies in similar domains, for example the study by (Ioannidis et al., 2023) explores the transformative potential of LLMs in legal and regulatory compliance. They developed advanced AI solutions, including a horizon scanning tool, an obligations generation tool, and an LLM-based expert system using OpenAI's GPT-4.

5.3.2) Feasibility of AI Models with Limited Data

Despite initial limitations in data availability, the study employed human-AI collaboration to expand the dataset through the creation of Q&A pairs (Step 1). The model's performance showed an increase in accuracy, demonstrating that AI tools can function effectively within constrained

data environments. Steps 6 and 10 established a continuous feedback loop that enabled ongoing model refinement, achieving Objective 2.

This approach aligns with (Z. Wang et al., 2023), who investigated the integration of expert knowledge into LLMs. They facilitated a dialogue between an AI investor and a financial expert using ChatGPT, incorporating feedback from human financial experts. The experimental results validate the effectiveness of this method in improving the model's ability to handle financial queries.

5.3.3) Codification of Expert Knowledge

The research successfully captured tacit knowledge from experts and transformed it into explicit, structured content (Step 8). This codified knowledge, validated through expert assessments, was integrated into the model, enabling it to provide context-specific guidance and enhance user confidence.

Many recent research explored the potential of incorporating experts feedback for finetuning LLMs, for instance, (Z. Wang et al., 2023) explored integrating expert knowledge into LLMs to create better-customized models for optimized decision-making. They initiated a dialogue between an AI investor and a financial expert using ChatGPT, incorporating feedback from human financial experts. This process resulted in a robust instruction tuning dataset of 103k multi-turn chats.

5.3.4) Establishing a Framework for AI-Based Regulatory Compliance

The iterative methodology not only developed a viable AI model but also laid the groundwork for a comprehensive framework adaptable to various chapters of the building code. This approach supports broader integration, demonstrating scalability and practicality for regulatory

compliance efforts across different AEC projects. The study's methodological steps, particularly Steps 6 and 10, were integral to establishing a feedback-driven improvement loop. This loop ensures that expert knowledge is regularly updated, user feedback is continuously integrated, and the model adapts to evolving regulatory requirements.

Figure (5.2) illustrates the iterative methodology centred on Compiled Data, showcasing feedback loops and adaptive mechanisms that refine the model. It highlights the interaction between expert inputs, user feedback, and regulatory updates to create a robust compliance framework.

Figure (5.1) provides an in-depth illustration of the Compiled Data structure, depicting the foundational elements that drive the model's adaptability and accuracy in navigating regulatory requirements across diverse AEC projects. This compiled data serves as the central input that feeds into the iterative methodology. Figure (5.1) details the breakdown of the Compiled Data, which forms the foundation of the model's knowledge base. This compiled data is categorized into three main types: Explicit Data, Tacit Data, and Hybrid Data.

- **Explicit Data:** includes structured regulatory and organizational information that is readily accessible and documented. This encompasses regulatory data such as building code chapters, relevant laws, implementing regulations, and amendments, as well as organizational data like guidelines, internal regulations, and other open-source resources. These sources provide a foundation of formal, documented knowledge critical for compliance.
- **Tacit Data:** involves knowledge that is inherently experiential and derived from individuals involved in the model's development and usage. This includes inputs from model developers, industry experts,

users, and evaluators, capturing insights and practical know-how that contribute to the model's effectiveness. Tacit Data is often less formalized but plays a vital role in refining the model's contextual understanding.

- **Hybrid Data:** represents a blend of tacit and explicit knowledge sources, enhancing the model's adaptability and continuous learning. This category is composed of contributions from Human-AI Collaboration, where human interactions with the AI help refine the model; Archived Knowledge, including resources such as handwritten documents, scripts, recorded voices, videos, photos, and diagrams that preserve tacit insights in an accessible format; and Interactive Training Data, which consists of real-time interactions like live demonstrations, question-answer sessions, and simulations that provide dynamic learning opportunities for the AI.

Together, these categories of Compiled Data provide a comprehensive foundation that enhances the model's relevance, accuracy, and flexibility in meeting the complex and evolving requirements within the AEC sector.

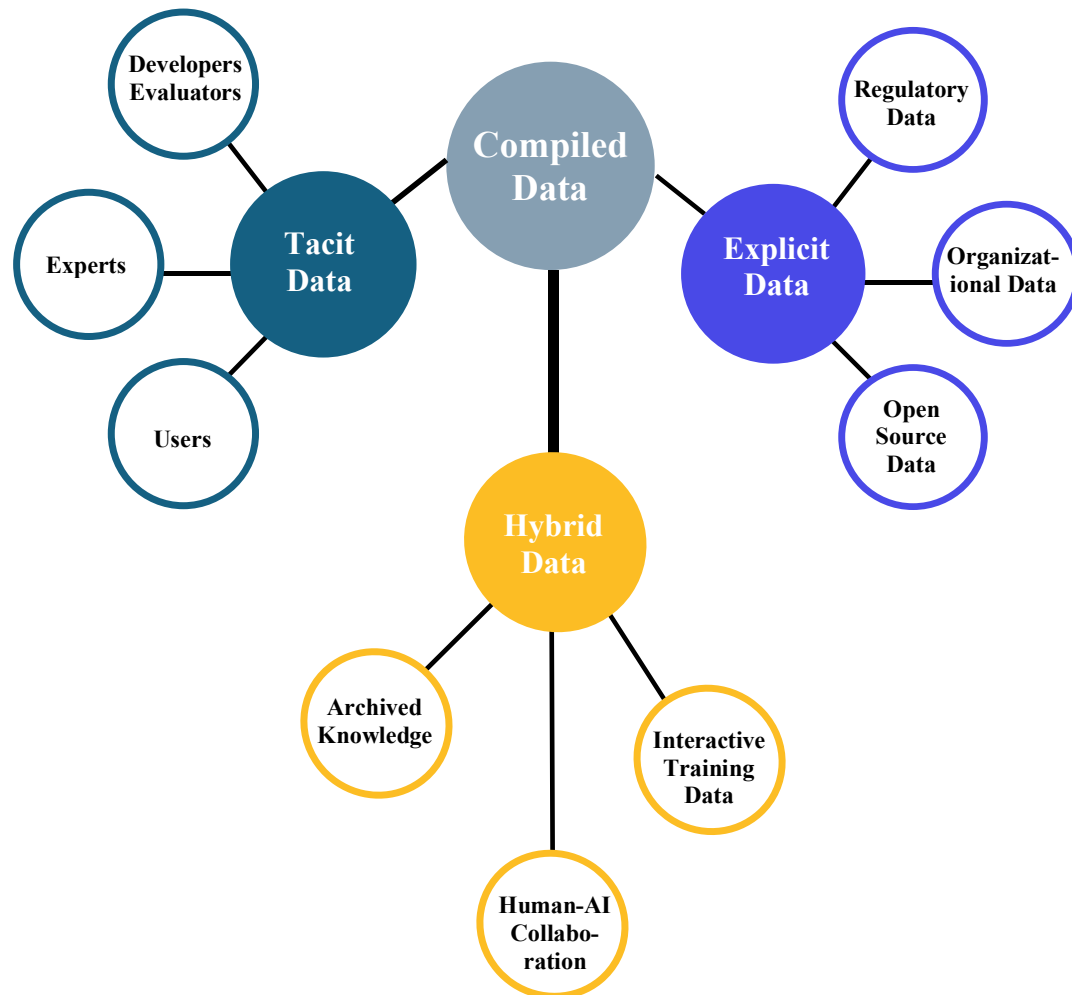


Figure (5.1): Key components of compiled data for model adaptability (Source: Student's work)

Figure (5.2) presents the comprehensive framework that integrates this data within an adaptive, feedback and data-driven model. Through data expansion, feedback loops and adaptive mechanisms, the framework continuously refines the model by incorporating expert insights, user feedback, and regulatory updates. The framework in Figure (5.2) also demonstrates how the same methodological steps can be applied to encompass additional chapters of the building code. By implementing iterative steps that involve data compilation including data generation through human-AI collaboration, evaluation and expert input, the model's knowledge base remains continuously updated and aligned with evolving regulatory standards. Together, these figures underscore the dynamic interplay between data compilation, iterative refinement, and adaptive mechanisms, fostering a resilient and evolving compliance framework that is responsive to the unique and shifting demands of the AEC industry.

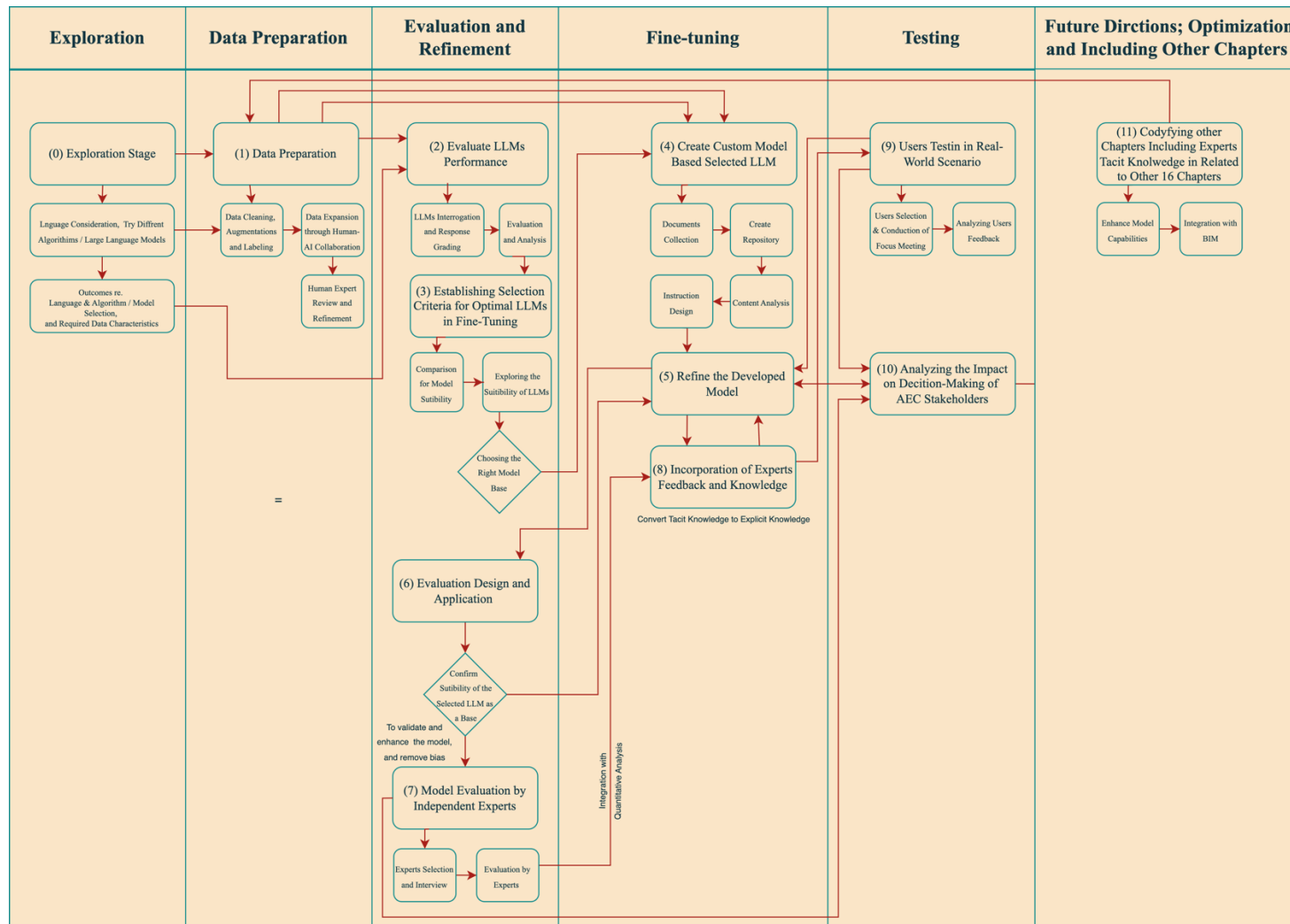


Figure (5.2): Integrated framework for data-driven compliance model with iterative feedback mechanism (Source: Student's work)

5.4) Capturing Tacit Knowledge and Enabling Knowledge Continuity

This research presents a scalable and adaptable framework as a significant contribution to the application of AI in regulatory contexts, supporting both current needs and future developments. Its flexibility extends beyond technology, enabling the model to respond to evolving regulatory content while capturing the practical insights and experiential knowledge of professionals in the field. A key challenge in regulatory domains is the tacit knowledge held by experts—often undocumented but critical for interpreting and applying building regulations. The proposed framework provides a structured approach to preserve and share this expertise by integrating expert-reviewed inputs and real-world case data. Over time, the model can serve as a living repository of regulatory intelligence, helping to bridge the gap between formal codes and professional practice, and supporting shared learning across the AEC industry.

This work also addresses the issue of limited data availability, a common constraint in domain-specific AI applications. As detailed in Chapter 3, this was overcome through human–AI collaboration, where expert input and iterative refinement were used to enrich and validate the dataset. Importantly, the customised model developed in this research is designed to serve a diverse range of AEC stakeholders—including clients, consultants, architects, engineers, project managers, contractors, suppliers, and regulatory authorities concerned with building regulations. By making regulatory knowledge more accessible and actionable across these groups, the framework supports improved compliance, coordination, and decision-making.

By combining regulatory automation with knowledge preservation, this research demonstrates how AI frameworks like CustomGPTs can support both technical innovation and institutional learning, making regulatory knowledge more accessible, sustainable, and future-ready.

5.5) Strategic Implementation, Adaptation, and Governance Considerations

This section presents the practical and strategic considerations for implementing the proposed framework. It discusses its alignment with national priorities discussed in Chapter 1, the skills and training required for adoption, the potential for broader application across regulatory domains, and the ethical and legal implications of its use.

5.5.1) Strategic Alignment and Implementation

The developed AI model aligns with Bahrain's Vision 2030 and the government program for 2023-2026 as discussed in Chapter 1, which emphasizes sustainable development and technological innovation. By streamlining regulatory compliance processes, the model contributes to efficient project delivery and governance. For effective implementation, it is recommended that authorities integrate the model into existing regulatory workflows, ensuring that decision-makers have access to real-time compliance support.

5.5.2) Empowering Adoption through Training for all Stakeholders

Successful adoption of the AI model necessitates comprehensive training programs for all stakeholder groups. Short-term training should focus on

familiarizing users with the model's functionalities, while long-term programs should aim to develop AI literacy and adaptability. For regulatory authorities, organizational-level integration of the model can enhance staff performance by enabling faster, more consistent responses to stakeholder inquiries and improving the accuracy of regulatory interpretation. Structured training for internal staff—including engineers, planners, and inspectors—will ensure the model is effectively embedded into daily workflows. Additionally, incorporating AI-related competencies into educational curricula for future professionals will ensure sustained engagement and proficiency across the AEC industry.

5.5.3) Adaptability of the Framework

While this research focused on developing a model tailored specifically to Chapter (14) of Bahrain's Building Code, the underlying framework was intentionally designed to be flexible and adaptable. By adjusting the input data and fine-tuning processes, the model can be extended to cover additional chapters of the code or applied in entirely different regulatory domains and jurisdictions even in regions. This adaptability not only enhances its relevance within the local context but also supports broader application across various regulatory environments.

Importantly, the framework was developed using the technologies available up to June 2024, including large language models and tools such as OpenAI's CustomGPT. However, it is well-positioned to integrate emerging AI technologies as they become available. New releases from leading developers like OpenAI and Google, as well as recent advancements from Chinese providers such as DeepSeek and Manus, can be accommodated with minimal structural changes. This future-oriented design ensures that the

model can evolve alongside the rapid progress in AI, maintaining its effectiveness and usability over time.

5.5.4) Ethical and Legal Implications

The deployment of AI in regulatory compliance, especially within the AEC industry, raises critical ethical and legal concerns. While AI-powered Q&A systems—such as customized GPTs—can significantly enhance efficiency and decision-making processes, their use must be guided by principles of transparency, accountability, and accuracy to uphold stakeholder trust and public interest (European Union Artificial Intelligence Act, 2024; OECD, 2019).

Professionals such as architects and engineers are governed by statutory obligations and professional codes of ethics. As such, they must not place full reliance on AI-generated outputs. Instead, they are expected to exercise professional judgment and due diligence by verifying AI responses against authoritative regulatory sources. This aligns with the broader view that AI should serve as a decision-support tool, not as a substitute for expert knowledge or legal interpretation (OpenAI, 2025). Ensuring responsible AI use requires organizations to implement safeguards such as disclaimers, audit trails, and human oversight to mitigate risks of misinformation, liability, or regulatory breaches. These measures help maintain compliance and reinforce ethical accountability in professional practice.

5.5.5) Trustworthy and Sustainable AI Use

Beyond legal and ethical obligations, the responsible use of AI in regulatory contexts must consider two additional dimensions: trustworthiness and sustainability. According to the OECD Principles on Artificial

Intelligence (2019), AI systems must be trustworthy meaning they should be transparent, accountable, fair, and robust throughout their lifecycle. This is especially critical in regulated domains like AEC, where public safety and compliance are paramount (OECD, 2019).

On the sustainability front, while large AI models do require significant computational resources, CustomGPTs offer a more targeted and efficient alternative. By fine-tuning AI to specific domains—such as regulatory compliance—they reduce unnecessary processing and improve relevance, thereby lowering overall resource demands. Furthermore, when hosted on optimized cloud infrastructure, these systems can balance computational efficiency with performance, supporting environmentally responsible innovation (Strubell et al., 2019).

5.6) Research Challenges and Limitations

Despite the promising potential of developing AI based Q&A model to optimize decision-making processes and ensure regulatory compliance in the AEC industry, there are several limitations to the study that must be acknowledged. These limitations pertain to the generalizability of the findings, data availability and quality, ML model limitations, stakeholder acceptance and usability, ethical considerations, availability of expert opinions, technological evolution, resource constraints, and interdisciplinary integration. These limitations are further detailed below:

a) Generalizability

The proposed research is based on data available up to June 2024 and focuses on regulatory compliance in the context of the AEC industry.

While the AI based Q&A model is developed using data from specific regulatory environments, its generalizability to other regions or jurisdictions may be limited. Differences in building regulations, cultural contexts, and data availability across various regions could impact the model's applicability and effectiveness. For instance, building regulations in one country may significantly differ from those in another, limiting the model's direct applicability.

b) Statistics on Building Violations Availability

it is important to note the lack of publicly announced statistics and relevant published studies on building violations. However, the researcher conducting this study, is aware of the existence of violations through being involved in the field and through newspaper reports and governmental directions. Additionally, steps taken to reduce violations through AI, as discussed in Chapter (1) of this research, further highlight the issue. This acknowledgment underscores the necessity of this research and its alignment with ongoing efforts to improve regulatory practices and enhance compliance in the construction sector.

c) Data Availability and Quality

The research relies heavily on the availability and quality of data related to building regulations. Variations in data structure, format, and content can pose challenges. The accuracy, completeness, and consistency of the data used are critical for the model's performance. Any biases or gaps in the dataset could potentially introduce limitations to the findings. Additionally, data availability for training and testing the AI model is a significant factor, and limited access to high-quality data could affect the model's development.

d) Machine Learning (ML) Model Limitations

The effectiveness of the ML model in accurately identifying compliance issues depends on several factors, including the quality of the training dataset and the chosen model architecture. The model's performance may be influenced by the limitations of the selected ML algorithms, potential overfitting, underfitting, or biases within the training data. These factors could impact the model's ability to generalize accurately to new and similar data, thus affecting its reliability and applicability in real-world scenarios.

e) Availability of Expert Opinions

The qualitative analysis of experts' opinions relies on the availability and willingness of experts to participate. Limited participation or biased expert opinions may impact the comprehensiveness and validity of the qualitative data analysis. The research outcomes depend significantly on the quality and breadth of input from industry experts.

f) Stakeholder Acceptance and Usability

The research involves testing the developed model through experiments with stakeholders. The success of implementing the model in practice hinges on stakeholder acceptance and usability. Factors such as the training provided to stakeholders, their familiarity with the technology, and their willingness to use AI based tool. Any limitations in stakeholder acceptance or usability may impact the practicality and real-world application of the proposed model. It is worth mentioning that an invitation to participate in the testing was distributed through the formal engineers' society in the Kingdom, but no one expressed interest in participating. Additionally, the researcher faced challenges

in approaching potential users represented in his occupation due to a conflict of interest in seeking support directly from architects and engineers in the industry.

g) Ethical Considerations

The research aims to ensure ethical conduct by considering ethical implications and obtaining expert opinions through qualitative data collection methods. However, potential limitations related to privacy, data security, and confidentiality of stakeholders' information must be addressed to mitigate ethical concerns. Ensuring the ethical handling of data and maintaining the confidentiality of sensitive information are paramount to the study's integrity and acceptance.

h) Technological Evolution

Given the rapid pace of advancements in AI and ML technologies, there is a risk that the tools and models used in this research may become outdated. The study is based on the available technology as of June 2024, and subsequent developments could introduce new capabilities or methods that were not considered, potentially impacting the relevance of the findings. Frequent updates to regulatory clauses and evolving interpretations demand regular updates to maintain model accuracy, which may require human oversight or automated data integration. The proposed framework is designed with adaptability, offering scalability for new regulatory chapters and flexibility to evolve with technological advancements, mitigating long-term risks.

i) Resource Constraints

The development and implementation of the AI based Q&A model require significant computational resources, time, and expertise. Limited access to these resources could constrain the scope of the research, affect the depth and breadth of the model's development and testing phases.

j) Interdisciplinary Integration

The research involves integrating knowledge from multiple disciplines, including AI, construction management, and regulatory compliance. Ensuring effective communication and collaboration across these disciplines can be challenging, and any gaps in interdisciplinary integration could impact the comprehensiveness and applicability of the research findings.

Recognizing these limitations is essential for understanding the scope and potential impact of the research. While the study aims to provide a robust solution for ensuring regulatory compliance and optimizing decision-making processes in the AEC industry, addressing these limitations in future research can enhance the model's applicability, reliability, and overall effectiveness.

5.7) Recommendations and Future Directions**a) Regulatory Adoption**

Given the framework's adaptability and potential to enhance regulatory processes, it is recommended that authorities responsible for specific chapters of the Bahraini Building Code consider adopting the proposed AI-based model. Its implementation could improve access to regulatory

information, support consistent compliance, and promote effective knowledge sharing among AEC stakeholders.

b) Broader Application

Beyond individual chapters, the framework could be extended to cover additional areas of the Building Code—such as fire safety, utilities, or municipal planning regulations. This would allow for a more comprehensive, cross-disciplinary decision-support tool. Future research should investigate how expanding the scope affects model accuracy, usability, and value across complex, multi-phase AEC projects.

c) Tacit Knowledge Preservation

Future studies should address the challenges of preserving expert tacit knowledge in evolving regulatory environments. Research could focus on creating a more dynamic, AI-based repository of expert insights that can be continually updated and verified for accuracy.

d) Adaptive AI Frameworks

The methodology's iterative framework should be tested across different domains, such as environmental compliance, safety protocols, or public infrastructure projects. This would determine its versatility and potential to be applied beyond the AEC sector, contributing to broader AI-based compliance research.

e) Human-AI Collaboration

While this research highlighted the benefits of human-AI collaboration, further exploration is needed to understand how user interactions can be optimized to maintain model accuracy over time. Studies could

investigate how human intervention in dataset expansion and feedback integration can be structured to ensure sustained model performance.

f) Integration with BIM Systems

Developing seamless integration with Building Information Modelling (BIM) can further enhance regulatory compliance checks within existing digital workflows.

g) Liability for Model Errors

Addressing liability in cases where the model provides incorrect or misleading information is a crucial area for future research. For instance, if the model provides a flawed answer during the design phase, leading to project delays or financial losses, who bears the responsibility? Future studies should explore the establishment of clear accountability frameworks—whether it is the engineer's duty to verify answers or the responsibility of the authority managing the model to ensure data accuracy. Legal and ethical considerations surrounding AI liability in decision-making must be further investigated.

h) Enhancing Multilingual Capabilities

Expanding the Custom GPT model to support multilingual interfaces (e.g., Arabic alongside English) will broaden its applicability across diverse user groups in the AEC sector. Future research should focus on assessing the effectiveness of multilingual models in improving accessibility and compliance accuracy in varied regulatory contexts.

i) Illustrative Support and User Prompts

Exploring the use of illustrative tools such as charts and visuals to enhance model explanations, as well as refining the model's prompting

system, could make interactions more intuitive. Research could also assess the feasibility of adding a “refer to human agent” feature when the model encounters complex queries or ambiguous responses.

j) Gradual Implementation of Model Capabilities

Future studies could examine phased approaches to deploying AI models, starting with straightforward regulatory questions and gradually incorporating more complex scenarios as the system is trained and validated. This method could mitigate risks associated with premature reliance on AI in sensitive areas.

k) Ensuring Consistency Across Authorities

Research should also explore strategies to ensure alignment and consistency in regulations across authorities before integrating them into the model. This would reduce conflicts caused by discrepancies in regulatory interpretations and ensure the model provides unified answers.

In summary, The Custom GPT model demonstrated notable effectiveness in improving decision-making and regulatory compliance within the AEC sector. The study’s contributions—technological, methodological, and theoretical—highlight the potential for AI to revolutionize regulatory compliance efforts by enhancing decision accuracy, improving stakeholder collaboration, and creating a more adaptable framework for future use. Ongoing research is needed to refine the framework, address identified gaps, and expand the model’s capabilities, especially in terms of multilingual support and broader regulatory integration. The roadmap provided by this study emphasizes the importance of adaptability, human-AI synergy, and

sustainable model evolution as essential components for optimizing decision-making in multi-stakeholder AEC projects through a novel, data-driven framework.

References

- Abdelkader, B., & Ibrahim, B. (2023). Contribution to decision-making in the big data industry based on the multiparametric similarity measure for Pythagorean fuzzy sets. *Journal of Logic and Computation*, Volume 33, Issue 3, April 2023, Pages 517–535, 33(3), 517–535. <https://academic.oup.com/logcom/article-abstract/33/3/517/6647540?redirectedFrom=fulltext>
- Abioye, S. O., Oyedele, L. O., Akanbi, L., Ajayi, A., Davila Delgado, J. M., Bilal, M., Akinade, O. O., & Ahmed, A. (2021). Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. In *Journal of Building Engineering* (Vol. 44). Elsevier Ltd. <https://doi.org/10.1016/j.jobe.2021.103299>
- Abrar Shukri. (2022, February 13). إخالة 89 من أصل 238 مخالفة بناء بمدينة حمد للنياية. 11999.
- acorn.io. (2024, May 13). *Fine-Tuning LLMs: Top 6 Methods, Challenges and Best Practices*. Open Source Software for Building AI Assistants: Learning Center. <https://www.acorn.io/resources/learning-center/fine-tuning-llm>
- Adeloye, A. (2023). Applications of Artificial Intelligence (AI) in the construction industry: A review of Observational Studies. *Applied Sciences Research Periodicals*. <https://www.researchgate.net/publication/372764029>
- Adu, J., Owusu, M. F., Martin-Yeboah, E., Pino Gavidia, L. A., & Gyamfi, S. (2022). A Discussion of Some Controversies in Mixed Methods Research for Emerging Researchers. *Methodological Innovations*, 15(3), 321–330. <https://doi.org/10.1177/20597991221123398>
- Ahmed, I., Kajol, M., Hasan, U., Datta, P. P., Roy, A., & Rokonuzzaman, R. (2023). *ChatGPT vs. Bard: A Comparative Study*. https://www.researchgate.net/publication/371799069_ChatGPT_vs_Bard_A_Comparative_Study
- Ahmed, J. O., Salih, A. M., Ahmed, J. O., Hiwa, D. S., Salih, A. M., Salih, R. Q., Hassan, H. A., Mahmood, Y. M., Mohammed, S. H., & Abdalla, B. A. (2025). Assessment of Chat-GPT, Gemini, and Perplexity in Principle of Research Publication: A Comparative Study. *Barw Medical Journal* |, 3(1), 2–6. <https://doi.org/10.58742/bmj.v2i4.140>
- AI at Meta. (2024). *Meta*. <https://ai.meta.com/>
- Aktharsha, S. U. (2011). *A Study on Knowledge Management and its Impact on Organizational Performance in Public Sector with Special Reference to Bharat Heavy Electricals Limited* [Bharathidasan Institute of Management]. <https://shodhganga.inflibnet.ac.in/handle/10603/9632>
- Al Nama, E., & Alalawi, A. (2023). The Adoption of Automated Building Code Compliance Checking Systems in the Architecture, Engineering, and Construction Industry. *IEEE International Conference on Cyber Management and Engineering (CyMaEn)*. <https://doi.org/10.1109/CyMaEn57228.2023.10050930>
- Al Nama, E., & Mahmud, M. (2024). Building the Blueprint for AI-powered compliance Checking: Analyzing the Performance of ChatGPT-4 & Gemini by Question Category

- in Engineering Regulations. *The International Conference on Computing in Civil and Building Engineering (ICCCBE) - 2024*.
- Al Nama, E., Mahmud, M., & Al Madhoob, H. (2024, December). Leveraging Human Expertise and AI for Engineering Regulatory Data Expansion: A Case Study with ChatGPT. *International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETIS 2024)*. <https://icetis24.asu.edu.bh/>
- Almasre, M. (2024). Development and Evaluation of a Custom GPT for the Assessment of Students' Designs in a Typography Course. *Education Sciences*, 14(2). <https://doi.org/10.3390/educsci14020148>
- Alsajri, A., Salman, H., & Steiti, A. (2024). Generative Models in Natural Language Processing: A Comparative Study of ChatGPT and Gemini. *Babylonian Journal of Artificial Intelligence*.
- AMI Environmental. (2025). *Compliance and Safety: Environmental Consultants' Roles in Building Projects*.
- Ansari, R., Banihashemi, S. A., Taherkhani, R., & Moradi, S. (2022). Decision Support System for Analyzing Key Performance Indicators in Construction Projects Management. *International Journal of Engineering, Transactions A: Basics*, 35(5), 865–874. <https://doi.org/10.5829/ije.2022.35.05b.03>
- Anthropic. (2024). *Cluadi*. <https://www.anthropic.com/claude>
- API platform, & OpenAI. (2023). *Transforming Work and Creativity with AI*. <https://openai.com/product>
- AR Paiva, & T Tasiden. (2013). (12) *United States Patent*. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/<https://patentimages.storage.googleapis.com/53/6a/7f/940e0ac09c2cd1/US8412651.pdf>
- Autodesk. (2023). *Design and Build with BIM - Building Information Modeling*.
- Autodesk. (2025, August 15). *Building Information Modeling*. <https://www.autodesk.com/solutions/aec/bim>
- Azhar, S. (2011). Building Information Modeling (BIM): Trends, Benefits, Risks, and Challenges for the AEC Industry. *Journal of Leadership and Management in Engineering*, 11, 241–252. [https://doi.org/10.1061/\(ASCE\)LM.1943-5630.0000127](https://doi.org/10.1061/(ASCE)LM.1943-5630.0000127) You can use this format in Mendeley to ensure it's correctly cited. If you need a nymore help with citations or anything else, feel free to ask!
- Azzouz, A., & Papadonikolaki, E. (2020). Boundary-Spanning for Managing Digital Innovation in the AEC Sector. *Architectural Engineering and Design Management*. <https://doi.org/10.1080/17452007.2020.1716693>
- Baskerville, R., & Dulipovici, A. (2006a). The Theoretical Foundations of Knowledge Management. *Knowledge Management Research and Practice*, 4(2), 83–105. <https://doi.org/10.1057/palgrave.kmrp.8500090>
- Baskerville, R., & Dulipovici, A. (2006b). The Theoretical Foundations of Knowledge Management. *Knowledge Management Research and Practice*, 4(2), 83–105. <https://doi.org/10.1057/palgrave.kmrp.8500090>
- Benayat. (2025). *Benayat*. <https://www.benayat.bh/building-permits/>
- Benchmark International. (2024). *2024 Global Architecture, Engineering & Construction (AEC) Industry Report*. <https://www.benchmarkintl.com/insights/2024-global-architecture-engineering-construction-aec-industry-report/>

- Bhatia, R. (2023). History in the Revolution of QSAR: A Review. *PharmaTutor*. <https://www.pharmatutor.org/articles/history-revolution-of-qsar-quantitative-structural-activity-relationship>
- Bhattacharya, S., & Momaya, K. (2021). Actionable Strategy Framework for Digital Transformation in AECO Industry. *Engineering, Construction and Architectural Management*. <https://www.emerald.com/insight/content/doi/10.1108/ECAM-07-2020-0587/full/html>
- BNA. (2021, January 17). مخالفات البناء السكنية تنتهك الخصوصية وتؤثر على الواجهات الحضرية. *11607*.
- Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Bukowski, D. (2023, October 25). *Context is Everything: How to Approach Fine-Tuning a Grounded LLM Application*. Medium. <https://medium.com/@bukowski.daniel/context-is-everything-how-to-approach-fine-tuning-a-grounded-llm-application-36418c8dc6ad>
- Cano, E. L., Moguerza, J. M., Ermolieva, T., & Yermoliev, Y. (2017). A Strategic Decision Support System Framework for Energy-Efficient Technology Investments. *TOP*, 2. <https://doi.org/10.1007/s11750-016-0429-9>
- Cash, P., Isaksson, O., Maier, A., & Summers, J. (2022). Sampling in Design Research: Eight Key Considerations. *Design Studies*, 78. <https://doi.org/10.1016/j.destud.2021.101077>
- Chandler, C., Foltz, P. W., & Elvevåg, B. (2022). Improving the Applicability of AI for Psychiatric Applications through Human-in-the-loop Methodologies. *Schizophrenia Bulletin*, 48(5), 949–957. <https://doi.org/10.1093/schbul/sbac038>
- Chen, R., Liang, J., Gao, S., Wan, F., & Quan, X. (2024). Self-Evolution Fine-Tuning for Policy Optimization. *Findings of the Association for Computational Linguistics: EMNLP 2024*. <http://arxiv.org/abs/2406.10813>
- Chen, X., Chang-Richards, A., Ling, F. Y. Y., Yiu, T. W., Pelosi, A., & Yang, N. (2025). Digital Technologies in the AEC Sector: a Comparative Study of Digital Competence among Industry Practitioners. *International Journal of Construction Management*. <https://doi.org/10.1080/15623599.2024.2304453>
- Chen, Y., Ren, Z., Hu, B., & Zheng, H. (2023). Investigation of the Critical Factors Influencing Multi-Stakeholders' Participation in Design Optimization of EPC Projects. *Buildings*, 13(7). <https://doi.org/10.3390/buildings13071654>
- Chu, Y., Cao, H., Diao, Y., & Lin, H. (2023). Refined SBERT: Representing Sentence BERT in Manifold Space. *Neurocomputing*, 555. <https://doi.org/10.1016/J.NEUCOM.2023.126453>
- Compliance Chain. (2024, January 26). *Compliance in Construction (Best Way in 2023)*. <https://compliancechain.com/compliance-in-construction-best-way-in-2023/>
- Creswell, J. W. (2009). *Research Design- Qualitative, Quantitative, and Mixed Methods Approaches*. [chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.ucg.ac.me/skladiste/blog_609332/objava_105202/fajlovi/Creswell.pdf](https://efaidnbmnnnibpcajpcglclefindmkaj/https://www.ucg.ac.me/skladiste/blog_609332/objava_105202/fajlovi/Creswell.pdf)
- Cretu, C. (2023, October 23). *How Does ChatGPT Actually Work? An ML Engineer Explains*. Scalable Path. <https://www.scalablepath.com/machine-learning/chatgpt-architecture-explained>

- CRPEP. (2014a). *Law No. (51) of 2014 with respect to Regulating the Practice of Engineering Professions*. <https://www.crpep.bh/>
- CRPEP. (2014b). *The Council for Regulating the Practice of Engineering Professions*. <http://www.crpep.bh/>
- D. Manning, C., Raghavan, P., & Schütze, H. S. (2009). *Introduction to Information Retrieval*.
- Debs, L., & Hubbard, B. (2023). Gathering and Disseminating Lessons Learned in Construction Companies to Support Knowledge Management. *Construction Economics and Building*.
- Deng, Y., Lu, P., Yin, F., Hu, Z., Shen, S., Zou, J., Chang, K.-W., & Wang, W. (2024). Enhancing Large Vision Language Models with Self-Training on Image Comprehension. *ArXiv*. <http://arxiv.org/abs/2405.19716>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. <https://doi.org/10.18653/v1/N19-1423>
- Edwards, B. (2023, April 5). *ChatGPT vs Google Bard: Which is better? We Put Them to the Test*. <https://arstechnica.com/information-technology/2023/04/clash-of-the-ai-titans-chatgpt-vs-bard-in-a-showdown-of-wits-and-wisdom/>
- Eilouti, B. (2018). Scenario-based design: New applications in metamorphic architecture. *Frontiers of Architectural Research*, 7(4), 530–543. <https://doi.org/10.1016/j.foar.2018.07.003>
- Environmental Protection Agency. (2025). *EPA*. <https://www.epa.gov/>
- European Commission. (2021, June 4). *Standard Contractual Clauses for Data Transfers between EU and non-EU Countries*. <https://www.mendeley.com/reference-manager/library/all-references/b525a984-2ac3-3cc7-be5f-c753c20a3838>
- European Commission. (2022, September 28). *Regulatory Framework Proposal on Artificial Intelligence*.
- European Union Artificial Intelligence Act. (2024). *European Union AI Act*. <https://artificialintelligenceact.eu/ai-act-explorer/>
- Fadoul, A., Tizani, W., & Arturo Osorio-Sandoval, C. (2020). A Knowledge-Based Model for Constructability Assessment of Buildings Design Using BIM. *The 18th International Conference on Computing in Civil and Building Engineering*, 147–159. https://link.springer.com/chapter/10.1007/978-3-030-51295-8_13
- Fernandez, T., & Ackerson, D. (2023, July 12). *Word Embeddings: Giving Your ChatBot Context For Better Answers*. Semaphore. <https://semaphoreci.com/blog/word-embeddings#:~:text=Word%20embeddings%20can%20be%20visualized,meanings%20for%20any%20input%20text.&text=The%20actual%20mathematics%20behind%20all%20this%20is%20beyond%20me>
- Finn, G. A. (1989). AEC Experience with Expert Systems in Construction. *Computing in Civil Engineering*. <https://www.semanticscholar.org/paper/AEC-Experience-with-Expert-Systems-in-Construction-Finn/ed75c20b897fa1e0b923194dfb819e0d8e41f4a0>

- Flannelly, L. T., Flannelly, K. J., & Jankowski, K. R. B. (2014). Independent, Dependent, and Other Variables in Healthcare and Chaplaincy Research. *Journal of Health Care Chaplaincy*, 20(4), 161–170. <https://doi.org/10.1080/08854726.2014.959374>
- Google. (2024a, April). *Colaboratory (Colab)*. <https://colab.research.google.com/>
- Google. (2024b, April). *Gemini Official Blog*. <https://blog.google/products/gemini/>
- Google. (2024c, April). *Gemma Open Models*. <https://ai.google.dev/gemma>
- Government of Bahrain. (2025). *Government Plan (2023-2026)*. <https://pmo.gov.bh/Media/Pdf-Section/GovPlan2023-2026.pdf>
- Grljevic, O., & Zita, B. (2011). Knowledge Management in Expert Systems Development. *Proceedings of the Management, Knowledge and Learning International Conference 2011*. <https://www.researchgate.net/publication/239807932>
- Gunel, B., Du, J., Conneau, A., & Stoyanov, V. (2021a). Supervised Contrastive Learning for Pre-Trained LANGUAGE Model Fine-Tuning. *The Thirty-Fourth Annual Conference on Neural Information Processing System*. <https://arxiv.org/abs/2011.01403>
- Gunel, B., Du, J., Conneau, A., & Stoyanov, V. (2021b, April 2). *Supervised Contrastive Learning for Pre-trained Language Model Fine-tuning*. <http://arxiv.org/abs/2011.01403>
- Guo, D., Onstein, E., & Rosa, A. D. La. (2021). A Semantic Approach for Automated Rule Compliance Checking in Construction Industry. *IEEE Access*, 9, 129648–129660. <https://doi.org/10.1109/ACCESS.2021.3108226>
- Guo, H., Lin, J.-R., & Yu, Y. (2023). Intelligent and Computer Technologies' Application in Construction. *Buildings*, 13(3), 641. <https://doi.org/10.3390/buildings13030641>
- Gupta, H., Sawant, S. A., Mishra, S., Nakamura, M., Mitra, A., Mashetty, S., & Baral, C. (2023). Instruction Tuned Models are Quick Learners. *ArXiv*. <http://arxiv.org/abs/2306.05539>
- Haitao, Z., Ning, G., Yong, J., Shutao, X., & Zhao, C.-Z. (2014). Automatic Question-Answering System. *Journal of Information Technology*. <https://doi.org/10.1080/01446190801965368>
- Halevi Hochwald, I., Green, G., Sela, Y., Radomyslsky, Z., Nissanholtz-Gannot, R., & Hochwald, O. (2023). Converting Qualitative Data in to Quantitative Values Using a Matched Mixed-Methods Design: A New Methodological Approach. *Journal of Advanced Nursing*, 79(11), 4398–4410. <https://doi.org/10.1111/jan.15649>
- Health and Safety Executive. (2025). *Construction*. <https://www.hse.gov.uk/construction/index.htm>
- Hemmer, P., Schellhammer, S., Vössing, M., Jakubik, J., & Satzger, G. (2022). Forming Effective Human-AI Teams: Building Machine Learning Models that Complement the Capabilities of Multiple Experts. *Thirty-First International Joint Conference on Artificial Intelligence*, 2478–2484. <https://github.com/ptrckhmmr/human-ai-teams>
- Houlsby, N., Giurui, A., Jastrzebski, S., Morrone, B., de Laroussilhe, Q., Gesmundo, A., Attariyan, M., & Gelly, S. (2019). Parameter-Efficient Transfer Learning for NLP. *Proceedings of the 36th International Conference on Machine Learning (ICML)*. <http://arxiv.org/abs/1902.00751>
- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). *LoRA: Low-Rank Adaptation of Large Language Models*. <http://arxiv.org/abs/2106.09685>

- HuggingFace. (2024). *Transformers Documentation*.
<https://huggingface.co/docs/transformers/en/index>
- IBM Watson Health. (2024). *IBM Watson Health*.
<https://www.ibm.com/industries/healthcare>
- Indurthi, S. R., Zhou, W., Chollampatt, S., Agrawal, R., Song, K., Zhao, L., & Zhu, C. (2024). Improving Multilingual Instruction Finetuning via Linguistically Natural and Diverse Datasets. *ArXiv*. <http://arxiv.org/abs/2407.01853>
- Information & eGovernment Authority. (2024). *Artificial Intelligence - National Portal of Bahrain*. Artificial Intelligence - National Portal of Bahrain
- Ioannidis, J., Harper, J., Quah, M. S., & Hunter, D. (2023). Gracenote.ai: Legal Generative AI for Regulatory Compliance. *Proceedings of the Third International Workshop on Artificial Intelligence and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2023)*. <https://ssrn.com/abstract=4494272>
- J. Kim, S. Chung, S. Moon, & S. Chi. (2022). Feasibility Study of a BERT-based Question Answering Chatbot for Information Retrieval from Construction Specifications. *2022 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*. <https://ieeexplore.ieee.org/document/9989625/authors#authors>
- Jacobi, O. (2025). *Exploring Architectures and Capabilities of Foundational LLMs*. Aporia. <https://www.aporia.com/learn/exploring-architectures-and-capabilities-of-foundational-llms/>
- Jasim, N. A., Aljumaily, H. S. M., Varouqa, I. F., & Al-Zwainy, M. S. (2021). Building Information Modeling and Building Knowledge Modeling in Project Management. *Computer Assisted Mechanics and Engineering Sciences*, 28(1), 3–16. <https://doi.org/10.24423/cames.302>
- Jessica Taylor. (2019, June 20). *What Are Question-Answering Systems?* <https://lucidworks.com/post/what-are-question-answering-systems/>
- Jordan, M., Kleinberg, J., & Schölkopf, B. (2006). *Pattern Recognition and Machine Learning*. [chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf](https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf)
- Jude Chukwura Obi. (2023). A Comparative Study of Several Classification Metrics and Their Performances on Data. *World Journal of Advanced Engineering Technology and Sciences*, 8(1), 308–314. <https://doi.org/10.30574/wjaets.2023.8.1.0054>
- Juell, P., & Paulson, P. (2003). Using Reinforcement Learning for Similarity Assessment in Case-Based Systems. *IEEE Intelligent Systems*.
- K2 Integrity. (2023). *Construction Safety: How Robust Compliance Programs Can Ensure Worker Safety*. <https://www.k2integrity.com/en/knowledge/expert-insights/2023/construction-safety-how-robust-compliance-programs-can-ensure-worker-safety/>
- Kadlčík, M., & Štefánik, M. (2024). *Self-training Language Models for Arithmetic Reasoning*. <http://arxiv.org/abs/2407.08400>
- Kamers, R. (2018). *Using the Scenario Planning Method in a Systematic Way to Develop a New Business Model and Reduce Future Uncertainty within the Construction Industry*. University of Twente.

- Keary, T. (2024). Bard vs. ChatGPT: Which is Better & What's the Difference? In *Techopedia*. <https://www.techopedia.com/bard-vs-chatgpt-which-is-better>
- Kevian, D., Syed, U., Guo, X., Havens, A., Dullerud, G., Seiler, P., Qin, L., & Hu, B. (2024). Capabilities of Large Language Models in Control Engineering: A Benchmark Study on GPT-4, Claude 3 Opus, and Gemini 1.0 Ultra. *ArXiv Preprint, arXiv:2404.03647*. <http://arxiv.org/abs/2404.03647>
- Knight, W., & Goode, L. (2023, November 13). *Bard (Chatbot)*. Wikipedia. https://en.wikipedia.org/wiki/Bard_%28chatbot%29
- Korkmaz, K. A., & Bahidrah, S. (2018). Implementation of Knowledge Management in Construction Projects. *Advancements in Civil Engineering & Technology*, 2(3). <https://doi.org/10.31031/acet.2018.02.000539>
- Koroteev MV. (2021). *BERT: A Review of Applications in Natural Language Processing and Understanding*. <https://www.semanticscholar.org/paper/BERT%3A-A-Review-of-Applications-in-Natural-Language-Koroteev/879eaab2275a364549809560b42f0fef357ebbce>
- Kovacevic, M., Nie, J.-Y., & Davidson, C. (2008). Providing Answers to Questions from Automatically Collected Web Pages for Intelligent Decision Making in the Construction Sector. *Journal of Computing in Civil Engineering*, 22(1), 3–13. [https://doi.org/10.1061/\(asce\)0887-3801\(2008\)22:1\(3\)](https://doi.org/10.1061/(asce)0887-3801(2008)22:1(3))
- Lau, J. (2024, October 26). *Bard vs. ChatGPT: What's the Difference? [2024]*. <https://zapier.com/blog/chatgpt-vs-bard/>
- Lavikka, R., Kallio, J., Casey, T., & Airaksinen, M. (2018). Digital Disruption of the AEC Industry: Technology-Orientated Scenarios for Possible Future Development Paths. *Construction Management and Economics*, 36(11), 635–650. <https://doi.org/10.1080/01446193.2018.1476729>
- Leitão, D., Saleiro, P., Figueiredo, M. A. T., & Bizarro, P. (2022). Human-AI Collaboration in Decision-Making: Beyond Learning to Defer. *ArXiv*. <https://arxiv.org/abs/2206.13202>
- Lester, B., Al-Rfou, R., & Constant, N. (2021). The Power of Scale for Parameter-Efficient Prompt Tuning. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. <http://arxiv.org/abs/2104.08691>
- Li, X., Liu, S., & Sun, Y. (2022). A GP-Based Hierarchical Objectives Decision-Making Method for Building Energy Efficiency Optimization. *Buildings*, 12(1). <https://doi.org/10.3390/buildings12010052>
- Lin, T.-H., Huang, Y.-H., & Putrantu, A. (2022). Intelligent Question and Answer System for Building Information Modeling and Artificial Intelligence of Things Based on the Bidirectional Encoder Representations from Transformers model. *Automation in Construction*, 142.
- Little, B. (2025). *The 7 Knowledge Management Trends Shaping 2025*. <https://bloomfire.com/blog/knowledge-management-trends/>
- Liu, J., Liu, J., Wang, Q., Wang, J., Cai, X., Zhao, D., Wang, R. L., & Yan, R. (2023). Retrieval-based Knowledge Transfer: An Effective Approach for Extreme Large Language Model Compression. *ArXiv*. <http://arxiv.org/abs/2310.15594>
- López Espejel, J., Ettifouri, E. H., Yahaya Alassan, M. S., Chouham, E. M., & Dahhane, W. (2023). GPT-3.5, GPT-4, or BARD? Evaluating LLMs Reasoning Ability in Zero-

- Shot Setting and Performance Boosting through Prompts. *Natural Language Processing Journal*, 5, 100032. <https://doi.org/10.1016/j.nlp.2023.100032>
- M. Toman. (2007). Utilization Of Knowledge Management And Information Technologies Theory In Water Resources Management. *Water Resources Management*, 103. <https://www.witpress.com/elibrary/wit-transactions-on-ecology-and-the-environment/103/17577>
- Ma, X., Wang, Z., Ng, P., Nallapati, R., & Xiang, B. (2019). Universal Text Representation from BERT: An Empirical Study. *ArXiv*. <http://arxiv.org/abs/1910.07973>
- Maali, O., Shalwani, A., Lines, B., Hurtado, K., & Sullivan, K. (2022, March 7). Are Different Innovations More Challenging to Implement? A Comparison of Different Types of Changes in the AEC. *Construction Research Congress*. <https://ascelibrary.org/doi/10.1061/9780784483978.094>
- Malsane, S., Matthews, J., Lockley, S., Love, P. E. D., & Greenwood, D. (2015). Development of an Object Model for Automated Compliance Checking. *Automation in Construction*, 49, Part A, 51–58.
- Marc Robert, J., Moulet, L., Lizarralde, G., Davidson, C., Nie, J., & Sylva, L. Da. (2006). Finding Out: A System for Providing Rapid and Reliable Answers to Questions in the Construction Sector. *Construction Innovation*, 6(4), 250–261. <https://doi.org/10.1108/14714170610713926>
- Marvin, G., Hellen, N., Jjingo, D., & Nakatumba-Nabende, J. (2024). *Prompt Engineering in Large Language Models* (pp. 387–402). https://doi.org/10.1007/978-981-99-7962-2_30
- Mateus, R. J. G., Pinto, F. S., Fauth, J., Azenha, M., Granja, J., Veludo, R., Muniz, B., Reis, J., & Marques, P. (2024). Improving Regulations for Automated Design Checking Through Decision Analysis Good Practices: A Conceptual Application to the Construction Sector. *Flexible Automation and Intelligent Manufacturing: Establishing Bridges for More Sustainable Manufacturing Systems (FAIM 2023)*, 160–169. https://doi.org/10.1007/978-3-031-38241-3_19
- Mathot, M., Coenders, J., & Rolvink, A. (2016, September 30). Spatial Structures in the 21st Century. *Proceedings of the IASS Annual Symposium 2016*.
- Maureira, C., Pinto, H., Yepes, V., & Garcia, J. (2021). Towards an AEC-AI Industry Optimization Algorithmic Knowledge Mapping: An Adaptive Methodology for Macroscopic Conceptual Analysis. *IEEE Access*, 9, 110842–110879. <https://doi.org/10.1109/ACCESS.2021.3102215>
- Meiru Che, B. S.; M. E. (2011). *Scenario-Based Architectural Design Decisions Documentation and Evolution*. The University of Texas at Austin.
- Meyer, S., Singh, S., Tam, B., Ton, C., & Ren, A. (2024). *A Comparison of LLM Finetuning Methods & Evaluation Metrics with Travel Chatbot Use Case*. <http://arxiv.org/abs/2408.03562>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *ArXiv*. <http://arxiv.org/abs/1301.3781>
- Ministry of Finance and National Economy, B. (2023). *Bahrain Economic Report*. https://www.mofne.gov.bh/media/td2n1wia/beq_v6_en.pdf
- Ministry of Finance and National Economy, B. (2024a). *Bahrain Economic Quarterly Report I Q2 2024*. <https://www.mofne.gov.bh/media/1flbokdj/beq2-2024-en.pdf>

- Ministry of Finance and National Economy, B. (2024b). *Bahrain Economic Quarterly Report I Q3 2024*. https://www.mofne.gov.bh/media/0q3cevt/beq_en.pdf
- Ministry of Works, B. (2010). *Project Management System – Procedures*. Project Management System – Procedures
- Mohammed, B. A. A. (2024). *Risk and Stakeholder Management in Mega Projects Beyond the Realms of Theory Chief Roads Special Projects Section Ministry of Works and Housing Kingdom of Bahrain*. Risk and Stakeholder Management in Mega Projects Beyond the Realms of Theory
- Moodley, K., Smith, N., & Preece, C. N. (2008). Stakeholder Matrix for Ethical Relationships in the Construction Industry. *Construction Management and Economics*, 625–632. <https://www.tandfonline.com/doi/full/10.1080/01446190801965368?scroll=top&needAccess=true>
- Moon, S., Lee, G., & Chi, S. (2022). Automated System for Construction Specification Review Using Natural Language Processing. *Advanced Engineering Informatics*, 51. <https://doi.org/10.1016/j.aei.2021.101495>
- Munawar, H. S., Ullah, F., Qayyum, S., & Shahzad, D. (2022). Big Data in Construction: Current Applications and Future Opportunities. *Big Data and Cognitive Computing*, 6(1). <https://doi.org/10.3390/bdcc6010018>
- Mundra, N., Doddapaneni, S., Dabre, R., Kunchukuttan, A., Puduppully, R., & Khapra, M. M. (2023). *A Comprehensive Analysis of Adapter Efficiency*. <http://arxiv.org/abs/2305.07491>
- Musa, M. (2023). *Expert System For Maintenance and Rehabilitation of All Types Of Pavements (ExFMRAP)*. <https://doi.org/10.22541/au.168339610.00329839/v1>
- Muthumanickam, N. K., Brown, N., Duarte, J. P., & Simpson, T. W. (2023). Multidisciplinary Design Optimization in Architecture, Engineering, and Construction: A Detailed Review and Call for Collaboration. In *Structural and Multidisciplinary Optimization* (Vol. 66, Issue 11). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1007/s00158-023-03673-y>
- Nabavi, A., Ramaji, I., Sadeghi, N., & Anderson, A. (2023). Leveraging Natural Language Processing for Automated Information Inquiry from Building Information Models. *Journal of Information Technology in Construction*, 28, 266–285. <https://doi.org/10.36680/J.ITCON.2023.013>
- Nair, S., Yang, E., Lawrie, D., Duh, K., McNamee, P., Murray, K., Mayfield, J., & Oard, D. W. (2022). Transfer Learning Approaches for Building Cross-Language Dense Retrieval Models. *ArXiv*. <http://arxiv.org/abs/2201.08471>
- National Fire Protection Association. (2025). *NFPA*. <https://www.nfpa.org/>
- Nematzadeh, H., Enayatifar, R., Mahmud, M., & Akbari, E. (2019). Frequency Based Feature Selection Method Using whale Algorithm. *Genomics*, 111(6), 1946–1955. <https://doi.org/10.1016/j.ygeno.2019.01.006>
- Nguyen, Z., Annunziata, A., Luong, V., Dinh, S., Le, Q., Ha, A. H., Le, C., Phan, H. A., Raghavan, S., & Nguyen, C. (2024). Enhancing Q&A with Domain-Specific Fine-Tuning and Iterative Reasoning: A Comparative Study. *ArXiv*. <http://arxiv.org/abs/2404.11792>

- Nonaka, I., & Krogh, G. von. (2009). Tacit Knowledge and Knowledge Conversion: Controversy and Advancement in Organizational Knowledge Creation Theory. *JSTOR*, 20, 635–652.
- Nonaka, I., & Takeuchi, H. (1995). *The Knowledge - Creating Company*.
- Nowrozy, R., Ahmed, K., & Wang, H. (2024). GPT, Ontology, and CAABAC: A Tripartite Personalized Access Control Model Anchored by Compliance, Context and Attribute. *ArXiv Preprint, arXiv:2403.08264*. <http://arxiv.org/abs/2403.08264>
- Nwadike, A., & Wilkinson, S. (2022). Challenges Facing Building Code Compliance in New Zealand. *International Journal of Construction Management*. <https://www.tandfonline.com/doi/abs/10.1080/15623599.2020.1801336>
- Occupational Safety and Health Administration. (2025). *OSHA*. <https://www.osha.gov/aboutosha>
- OECD. (2019). *OECD AI Principles overview*. <https://oecd.ai/en/ai-principles>
- Olawumi, T. O., Chan, D. W. M., Wong, J. K. W., & Chan, A. P. C. (2018). Barriers to the integration of BIM and sustainability practices in construction projects: A Delphi survey of international experts. *Journal of Building Engineering*, 20, 60–71. <https://doi.org/10.1016/j.jobbe.2018.06.017>
- Onatayo, D., Onososen, A., Oyediran, A. O., Oyediran, H., Arowoia, V., & Onatayo, E. (2024). Generative AI Applications in Architecture, Engineering, and Construction: Trends, Implications for Practice, Education & Imperatives for Upskilling—A Review. *Architecture*, 4(4), 877–902. <https://doi.org/10.3390/architecture4040046>
- OpenAI. (2023a). *Embeddings*. <https://platform.openai.com/docs/guides/embeddings/what-are-embeddings>
- OpenAI. (2023b, November 6). *Introducing GPTs*. <https://openai.com/index/introducing-gpts/>
- OpenAI. (2024, April). *OpenAI*. <https://openai.com/>
- OpenAI. (2025, January 29). *Open AI Usage policies*. <https://openai.com/policies/usage-policies/>
- Pang, W., Zhou, C., Zhou, X.-H., & Wang, X. (2024). Phased Instruction Fine-Tuning for Large Language Models. *Findings of the Association for Computational Linguistics: ACL 2024*. <http://arxiv.org/abs/2406.04371>
- Panicker, A. D., & Venkitakrishnan, S. (2012). Question Classification Using Machine Learning Approaches. *International Journal of Computer Applications*, 48(13), 975–888.
- Parliamentary Affairs Editor. (2021, December 18). تخفيف عقوبات مخالفات البناء.. غرامة 500 دينار فقط بدلاً من 10 آلاف. 11942. <https://www.alayam.com/alayam/local/938772/News.html>
- Patton, M. Quinn. (2015). *Qualitative Research & Evaluation Methods* (4th ed.).
- Pauli, J., Hoffmann, M., & Bergmann, R. (2023). Similarity-Based Retrieval in Process-Oriented Case-Based Reasoning Using Graph Neural Networks and Transfer Learning. *The International FLAIRS Conference Proceedings*.
- Pfeiffer, J., Kamath, A., Rücklé, A., Cho, K., & Gurevych, I. (2020). AdapterFusion: Non-Destructive Task Composition for Transfer Learning. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*. <http://arxiv.org/abs/2005.00247>
- Plummer, M. L. (2021). *Open Architecture*. Ryerson University.

- Pokhariyal, G. P. (2019). Importance of Moderating and Intervening Variables on the Relationship between Independent and Dependent Variables. *International Journal of Statistics and Applied Mathematics*, 4(5). www.mathsjournal.com
- Pott, R. W. M., & Nortjé, S. (2021). The Use of Student Question-Posing in Reactor Design to Encourage an Open-Ended Approach to Learning. *European Journal of Engineering Education*, 46(6), 874–888. <https://doi.org/10.1080/03043797.2021.1923662>
- Power, D. J. (2002). *Decision Support Systems: Concepts and Resources for Managers*. Quorum Books. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/<https://scholarworks.uni.edu/cgi/viewcontent.cgi?article=1066&context=facbook>
- Prabatha, T., Hewage, K., & Sadiq, R. (2022). An Energy Performance Contract Optimization Approach to Meet the Competing Stakeholder Expectations under Uncertainty: A Canadian Case Study. *Sustainability (Switzerland)*, 14(7). <https://doi.org/10.3390/su14074334>
- Premakumar, V. N., Vaiana, M., Pop, F., Rosenblatt, J., Schwerz De Lucena, D., Ziman, K., & Graziano, M. S. A. (2024). Unexpected Benefits of Self-Modeling in Neural Systems. *ArXiv*.
- Pressbooks. (2025). *Units of Analysis and Units of Observation*. <https://pressbooks.pub/scientificinquiryinsocialwork/chapter/7-3-unit-of-analysis-and-unit-of-observation/>
- Primožič, L., & Kutnar, A. (2024). Key Decision Factors of Professional Stakeholders (Architects, Engineers, Constructors) When Deciding for Sustainable Construction. *Frontiers in Built Environment*, 10. <https://doi.org/10.3389/fbuil.2024.1420163>
- PyMuPDF. (2024). *PyMuPDF Documentation*. <https://pymupdf.readthedocs.io/en/latest/>
- Q&A Software. (2023). In *Wikipedia*.
- Qihuang, Z., Ding, L., Liu, J., Du, B., Zhong, Q., Ding R, L., & Tao R, D. (2023). Can ChatGPT Understand Too? A Comparative Study on ChatGPT and Fine-tuned BERT. *ArXiv*. <https://doi.org/10.48550/arXiv.2302.10198>
- Rafsanjani, H. N., & Nabizadeh, A. H. (2023). Towards Digital Architecture, Engineering, and Construction (AEC) Industry through Virtual Design and Construction (VDC) and Digital Twin. *Energy and Built Environment*, 4(2), 169–178. <https://doi.org/10.1016/j.enbenv.2021.10.004>
- Raj, M. J., VM, K., Warriar, H., & Gupta, Y. (2024). Fine Tuning LLM for Enterprise: Practical Guidelines and Recommendations. *ArXiv*. <https://doi.org/10.48550/arXiv.2404.10779>
- Rane, N., Choudhary, S., & Rane, J. (2024). Gemini versus ChatGPT: Applications, Performance, Architecture, Capabilities, and Implementation. *Journal of Applied Artificial Intelligence*, 5(1), 69–93. <https://doi.org/10.48185/jaai.v5i1.1052>
- Rau, D., Wang, S., Déjean, H., & Clinchant, S. (2024). Context Embeddings for Efficient Answer Generation in RAG. In *IEEE International Conference on Program Comprehension* (Vols. 2022-March). IEEE Computer Society. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>
- Ren, R., & Zhang, J. (2021). A New Framework to Address BIM Interoperability in the AEC Domain from Technical and Process Dimensions. *Advances in Civil Engineering*, 2021. <https://doi.org/10.1155/2021/8824613>

- Riyanto, S., Imas, S. S., Djatna, T., & Tika Dewi Atikah. (2023). Comparative Analysis using Various Performance Metrics in Imbalanced Data for Multi-Class Text Classification. *International Journal of Advanced Computer Science and Applications*; West Yorkshire, 14(6). <https://www.proquest.com/openview/0aee35afdb3b2075ac884a6ec288dbbc/1?pq-origsite=gscholar&cbl=5444811>
- Robert, J.-M., Moulet, L., Lizarralde, G., Davidson, C., Nie, J.-Y., & Sylva, L. Da. (2006). Finding Out: A System for Providing Rapid and Reliable Answers to Questions in the Construction Sector. *Construction Innovation*. <https://doi.org/10.1108/14714170610713926>
- Rubin, R. A. (2010). Review of Legal Aspects of Architecture, Engineering, and the Construction Process. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*.
- Russell, S. E., Wrobel, A. L., Dean, O. M., Berk, M., Dodd, S., Ng, C. H., Malhi, G. S., Cotton, S. M., Sarris, J., & Turner, A. (2022). Mixed Methods Thematic Analysis of a Randomised Controlled Trial of Adjunctive Mitochondrial Agents for Bipolar Depression. *Clinical Psychopharmacology and Neuroscience*, 20, 300–310. <https://doi.org/10.9758/cpn.2022.20.2.300>
- Sadek, K., El-Bastawissi, I., Raslan, R., & Sayary, S. (2020). IMPACT OF BIM ON BUILDING DESIGN QUALITY. *BAU Journal - Creative Sustainable Development*, 1(1). <https://doi.org/10.54729/2789-8334.1000>
- Salmanova, S. S. S. (2024). The Role of Expert Systems Within Artificial Intelligence. *Proceedings of Azerbaijan High Technical Educational Institution*, 335–342.
- Salvalai, G., Sesana, M. M., Dell'Oro, P., & Brutti, D. (2023). Open Innovation for the Construction Sector: Concept Overview and Test Bed Development to Boost Energy-Efficient Solutions. *Energies*, 16(14). <https://doi.org/10.3390/en16145522>
- Sara Najeeb. (2022, January 21). 1000 مخالفات بقاء في 3 سنوات وتعليق العقوبات قلل المخالفات 11976.
- Sari, R., & Pekerikli, M. K. (2020). An Investigation of Comparison and Evaluation of Official BIM Documents Released in the USA, UK and Turkey. *Journal of Construction Engineering, Management & Innovation*, 3(1), 67–84. <https://doi.org/10.31462/jcemi.2020.01067084>
- Schröder, C., Heyer, G., & Ai, S. (2024). *Self-Training for Sample-Efficient Active Learning for Text Classification with Pre-Trained Language Models*.
- Scikit-Learn. (2024). *F1_score*. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html
- Shao, Z., Gong, Y., Shen, Y., Huang, M., Duan, N., & Chen, W. (2023). Enhancing Retrieval-Augmented Large Language Models with Iterative Retrieval-Generation Synergy. *Findings of the Association for Computational Linguistics: EMNLP 2023*. <https://doi.org/10.18653/v1/2023.findings-emnlp.620>
- Shi, Y., Zi, X., Shi, Z., Zhang, H., Wu, Q., & Xu, M. (2024). Enhancing Retrieval and Managing Retrieval: A Four-Module Synergy for Improved Quality and Efficiency in RAG Systems. *ArXiv*. <http://arxiv.org/abs/2407.10670>
- Silhavy, R., & Silhavy, P. (2023). Artificial Intelligence Application in Networks and Systems. *Proceedings of 12th Computer Science On-Line Conference 2023, Volume 3*. <https://link.springer.com/book/10.1007/978-3-031-35314-7>

- Sleiman, H. A., & Robert, S. (n.d.). *A Survey on Reasoning on Building Information Models Based on IFC*.
- Stam, L., Ostuzzi, F., & Heylighen, A. (2022). Open design: an actual topic in architectural education. *International Journal of Technology and Design Education*, 32(1), 667–693. <https://doi.org/10.1007/s10798-020-09607-9>
- Strobbe, T., Pauwels, P., Verstraeten, R., De Meyer, R., & Campenhout, V. (2012). Optimization in Compliance Checking Using Heuristics: Flemish Energy Performance Regulations (EPR). *EWork and EBusiness in Architecture, Engineering and Construction*.
- Strobelt, H., Kinley, J., Krueger, R., Beyer, J., Pfister, H., & Rush, A. M. (2022). GenNI: Human-AI Collaboration for Data-Backed Text Generation. *IEEE Transactions on Visualization and Computer Graphics*, 28(1), 1106–1116. <https://doi.org/10.1109/TVCG.2021.3114845>
- Strubell, E., Ganesh, A., & McCallum, A. (2019). *Energy and Policy Considerations for Deep Learning in NLP*. <http://arxiv.org/abs/1906.02243>
- Sujan, S. (2020). *A holistic and Systemic Model of Collaboration in the AEC Industry* [University of Liverpool, School of Engineering]. <https://doi.org/10.13140/RG.2.2.33419.72485>
- Sujan, S. F., Jones, S. W., Kiviniemi, A., Wheatcroft, J. M., & Mwiya, B. (2020). Holistically Assessing Collaborative Culture in the AEC Industry. *Journal of Information Technology in Construction*, 25, 272–286. <https://doi.org/10.36680/j.itcon.2020.016>
- Sujan, S. F., Kiviniemi, A., Jones, S. W., Wheathcroft, J. M., & Hjelseth, E. (2019). Common biases in client involved decision-making in the AEC industry. *Frontiers of Engineering Management*, 6(2), 221–238. <https://doi.org/10.1007/s42524-019-0026-3>
- Tamkeen. (2025). *Tamkeen*. <https://www.tamkeen.bh/en/>
- Terenchuk, S., Pasko, R., Panko, O., & Zaprivoda, V. (2021). Models, Methods, and Means of Reproduction of Expert Knowledge in Intelligent Support System Building Technical Expertise. *Scientific Journal of Astana IT University*, 6, 76–87. <https://doi.org/10.37943/aitu.2021.43.51.007>
- The Daily Tribune. (2024, November 19). *Building Permit Approval Made Simple*. Building permit approval made simple
- Tubaro, P. (2022). *Learners in the loop: hidden human skills in machine intelligence*. 163, 110–129. <https://doi.org/10.3280/SL2022-163006i>
- Turing. (2023). *A Guide on Word Embeddings in NLP*. <https://www.turing.com/kb/guide-on-word-embeddings-in-nlp>
- Tzagkarakis, S. I., & Kritas, D. (2023). Mixed Research Methods in Political Science and Governance: Approaches and Applications. *Quality and Quantity*, 57, 39–53. <https://doi.org/10.1007/s11135-022-01384-y>
- Umar, T. (2022). Challenges of BIM Implementation in GCC Construction Industry. *Engineering, Construction and Architectural Management*, 29(3), 1139–1168. <https://doi.org/10.1108/ecam-11-2019-0608>
- Varavenko, V. E., Lyapustina, N. A., & Kovalev, D. V. (2022). Mitigation Tools for Risks Assigned to the Contractor in the EPC Contract: the Experience of the International

- Federation of Consulting Engineers and Russian Civil Legislation. *Pravo i Politika*. <https://doi.org/10.17803/1994-1471.2022.4.55-65>
- Villaschi, F. S., Carvalho, J. P., & Bragança, L. (2022). BIM-Based Method for the Verification of Building Code Compliance. *Applied System Innovation*, 5(4). <https://doi.org/10.3390/asi5040064>
- Wang, H., Meng, X., & McGetrick, P. J. (2018). Involving Knowledge of Construction and Facilities Management in Design through the BIM Approach. *Proceedings of the Creative Construction Conference (2018)*, 568–575. <https://doi.org/10.3311/cc2018-075>
- Wang, N., Issa, R. R. A., & Anumba, C. J. (2021). Query Answering System for Building Information Modeling Using BERT NN Algorithm and NLG. *Computing in Civil Engineering 2021 - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2021*, 425–432. <https://doi.org/10.1061/9780784483893.053>
- Wang, N., Issa, R. R. A., & Anumba, C. J. (2022a). NLP-Based Query-Answering System for Information Extraction from Building Information Models. *Journal of Computing in Civil Engineering*, 36(3). [https://doi.org/https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0001019](https://doi.org/https://doi.org/10.1061/(ASCE)CP.1943-5487.0001019)
- Wang, N., Issa, R. R. A., & Anumba, C. J. (2022b). Query Answering System for Building Information Modeling Using BERT NN Algorithm and NLG. *Computing in Civil Engineering 2021 - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2021*, 425–432. <https://doi.org/10.1061/9780784483893.053>
- Wang, Z., He, B., Yang, Y., Shen, C., & Peña-Mora, F. (2020). Building a next-generation AI Platform for AEC: A Review and Research Challenges. *37th CIB W78 Information Technology for Construction Conference*, 27–45. <https://doi.org/10.46421/2706>
- Wang, Z., Wang, J., Wu, J., & Zhang, X. (2023). An Effective Data Creation Pipeline to Generate High-quality Financial Instruction Data for Large Language Model. *ArXiv Preprint, arXiv:2308.01415*. <http://arxiv.org/abs/2308.01415>
- Wang, Z., Wang, Z., Le, L., Zheng, H. S., Mishra, S., Perot, V., Zhang, Y., Mattapalli, A., Taly, A., Shang, J., Lee, C.-Y., & Pfister, T. (2024). *Speculative RAG: Enhancing Retrieval Augmented Generation through Drafting*. <http://arxiv.org/abs/2407.08223>
- Wayessa, G. O., & Nygren, A. (2023). Mixed Methods Research in Global Development Studies: State-Sponsored Resettlement Schemes in Ethiopia. *European Journal of Development Research*, 6, 1440–1464. <https://doi.org/10.1057/s41287-023-00588-8>
- Węglarz, A. (2018). The Concept of Expert System Supporting the Increase of Energy Efficiency in Buildings. In *Design Solutions for nZEB Retrofit Buildings* (pp. 115–140). IGI Global. <https://doi.org/10.4018/978-1-5225-4105-9.ch005>
- Węglarz, A., & Gilewski, P. G. (2017). Application of Expert Systems in the Construction Sector. *MATEC Web of Conferences*. <https://doi.org/10.1051/mateconf/20171100001>
- Wenwu, Q. (2018). *Automatic Questioning and Answering Method and System*. <https://typeset.io/papers/automatic-questioning-and-answering-method-and-system-2ym6e8hcjf>

- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*. <http://arxiv.org/abs/2302.11382>
- Wilson, S. (2023, October 23). *Embeddings: What They are and Why They Matter*. <https://simonwillison.net/2023/Oct/23/embeddings/>
- Wolf, F., & Fink, G. A. (2024). Self-Training for Handwritten Word Recognition and Retrieval. *International Journal on Document Analysis and Recognition*. <https://doi.org/10.1007/s10032-024-00484-9>
- Wu, J., Asce, S. M., Akanbi, T., Zhang, J., & Asce, A. M. (2022). Constructing Invariant Signatures for AEC Objects to Support BIM-based Analysis: Automation through Object Classification. *Computing in Civil Engineering*. https://www.researchgate.net/publication/361671307_Constructing_Invariant_Signatures_for_AEC_Objects_to_Support_BIM-Based_Analysis_Automation_through_Object_Classification
- Xu, S., Li, W., Tang, L. C., Yang, L., & Tang, Q. (2018). Artificial Intelligence Assisted Professional Work in BIM: A Machine Reasoning Extension. *Creative Construction Conference (2108)*, 16–23. <https://doi.org/10.3311/ccc2018-003>
- Yasser, A., & Abdulkarim, A. (n.d.). *Effect of Redefining the Public Sector's Role on the Stakeholder's Satisfaction: Benayat System as a Case Study*.
- Yeo, W. J., Ferdinan, T., Kazienko, P., Satapathy, R., & Cambria, E. (2024). Self-training Large Language Models through Knowledge Detection. *Findings of the Association for Computational Linguistics: EMNLP 2024*. <http://arxiv.org/abs/2406.11275>
- Ying, W., Wang, D., Hu, X., Zhou, Y., Aggarwal, C. C., & Fu, Y. (2024). Unsupervised Generative Feature Transformation via Graph Contrastive Pre-training and Multi-objective Fine-tuning. *ArXiv*. <http://arxiv.org/abs/2405.16879>
- Yun-xi, X. (2008). Construction Project Stakeholder Collaboration Group Decision Making Based on Entropy Theory. *Chinese Journal of Management Science*. <https://www.semanticscholar.org/paper/Construction-Project-Stakeholder-Collaboration-on-Yun-xi/ce40e23d9bca9b545f556a428b62667fbf6eb414>
- Yuxia, L., & Ruonan, R. (2021). Automatic Question-Answering System Based on Semi-Supervised Learning and Text-to-SQL Model. *Journal of Artificial Intelligence Research*. <https://doi.org/10.1613/jair.1.12345>
- Zerhoudi, S., & Granitzer, M. (2024). PersonaRAG: Enhancing Retrieval-Augmented Generation Systems with User-Centric Agents. *ArXiv*. <http://arxiv.org/abs/2407.09394>
- Zhang, B., Liu, Z., Cherry, C., & Firat, O. (2024). When Scaling Meets LLM Finetuning: The Effect of Data, Model and Finetuning Method. *ArXiv*. <http://arxiv.org/abs/2402.17193>
- Zhang, L., Jijo, K., Setty, S., Chung, E., Javid, F., Vidra, N., & Clifford, T. (2024). Enhancing Large Language Model Performance to Answer Questions and Extract Information More Accurately. *ArXiv*. <http://arxiv.org/abs/2402.01722>
- Zhang, T., Patil, S. G., Jain, N., Shen, S., Zaharia, M., Stoica, I., & Gonzalez, J. E. (2024). *RAFT: Adapting Language Model to Domain Specific RAG*. <http://arxiv.org/abs/2403.10131>
- Zhang, Z., Broyd, T., & Ma, L. (2023a). Rule Capture of Automated Compliance Checking of Building Requirements: A Review. *Proceedings of the Institution of Civil*

- Engineers: Smart Infrastructure and Construction*, 176(4), 224–238.
<https://doi.org/10.1680/jsmic.23.00005>
- Zhang, Z., Broyd, T., & Ma, L. (2023b). Rule capture of automated compliance checking of building requirements: a review. *Proceedings of the Institution of Civil Engineers: Smart Infrastructure and Construction*, 176(4), 224–238.
<https://doi.org/10.1680/jsmic.23.00005>
- Zhang, Z., Ma, L., & Broyd, T. (2022). Towards Fully-Automated Code Compliance Checking Of Building Regulations: Challenges For Rule Interpretation And Representation. *European Conference on Computing in Construction*.
- Zheng, Z., Chen, K.-Y., Cao, X.-Y., Lu, X.-Z., & Lin, J.-R. (2023). *LLM-FuncMapper: Function Identification for Interpreting Complex Clauses in Building Codes via LLM*.
- Zheng, Z., Lu, X.-Z., Chen, K.-Y., Zhou, Y.-C., & Lin, J.-R. (2022). *Pretrained Domain-Specific Language Model for General Information Retrieval Tasks in the AEC Domain*. https://www.researchgate.net/publication/359130279_Pretrained_Domain-Specific_Language_Model_for_General_Information_Retrieval_Tasks_in_the_AEC_Domain
- Zhilyaev, D., Binnekamp, R., & Wolfert, A. M. R. (2022). Best Fit for Common Purpose: A Multi-Stakeholder Design Optimization Methodology for Construction Management. *Buildings*, 12(5). <https://doi.org/10.3390/buildings12050527>
- Zhong, B., He, W., Huang, Z., Love, P., Tang, J., & Luo, H. (2020). A Building Regulation Question Answering System: A deep Learning Methodology. *Advanced Engineering Informatics*, 46.
<https://www.sciencedirect.com/science/article/abs/pii/S1474034620301658?via%3Dihub>
- Zhou, H., Liu, T., Ma, Q., Yuan, J., Liu, P., You, Y., & Yang, H. (2023). LoBaSS: Gauging Learnability in Supervised Fine-tuning Data. *ArXiv*. <http://arxiv.org/abs/2310.13008>
- Zhu, B., Niu, Y., Lee, S., Hur, M., & Zhang, H. (2023). *Debiased Fine-Tuning for Vision-language Models by Prompt Regularization*. <http://arxiv.org/abs/2301.12429>

Appendices

Appendix (A):

LLMs Interrogation and Response Grading,

ChatGPT-4 and Gemini

Due to its length, Appendix (A) is stored in a cloud drive and can be accessed at:

[\[https://bit.ly/3F9pmJ3\]](https://bit.ly/3F9pmJ3).

Appendix (B):

Samples of Questions Generated through Human-AI Collaboration and Answers Used as Training Data

Due to its length, Appendix (B) is stored in a cloud drive and can be accessed at:

[<https://bit.ly/3F9pmJ3>].

Appendix (C):

Samples of Questions Generated through Human-AI Collaboration and Answers Provided by CustomGPT, for Expert Feedback

Due to its length, Appendix (C) is stored in a cloud drive and can be accessed at:

[<https://bit.ly/4i91bcd>].

Appendix (D):

Interview Questions for Experts

Expert Feed-back on Q&A Model Performance

Dear Expert,

Thank you for participating in this research study to evaluate the performance of a custom Q&A model. The purpose of this research is to assess and enhance the model's ability to provide accurate, relevant, and clear responses to questions related to Law No. 51 of 2014 concerning the Regulation of the Practice of Engineering Profession, its Implementing Regulations, CRPEP's Forms and Circulars, and other related legislation.

Before completing this form, please review the model's responses to the 90 designated questions and interact with the model. Your feedback collected through this form will help identify strengths and areas for improvement in the model.

The main objective of this evaluation is to incorporate your expert knowledge into the model. Insights on technical accuracy, depth of knowledge, and handling of ambiguities are particularly valuable, along with any suggestions for improvement. Please note that your responses will be kept confidential and used solely for the purpose of this research. If you have any questions or require further assistance, please feel free to contact me.

Instructions:

1. Sign the participant consent form (attached) before proceeding with the evaluation.
2. Review the model's responses to the 90 designated questions (to be provided after the introductory session).
3. Sign up for ChatGPT (ChatGPT | OpenAI).
4. Engage with the model using the provided link to ask additional questions or seek clarifications.
5. Complete this expert feedback form, offering your insights on various aspects of the model's performance.
6. **Please note:** The model is for trial purposes only, exclusively for participants in this research study. It is not an official tool released by the concerned authority, and the link to the model will be disabled after the study is completed. Therefore, **please do not share this link with anyone else.**

Thank you for your valuable contribution to this research.
Entesar,

* Required

Expert Information

1. Please provide information about your name (optional), your role and designation (mandatory) *

General Performance

2. What is your overall impression of the model's performance? *

3. How well do you think the model understands and answers the questions? *

4. Rate the overall accuracy of the model's responses. *



5. Provide examples of responses that were exceptionally accurate or inaccurate. *

6. Rate the relevance of the model's answers to the questions asked. *



7. Did the model provide any off-topic or irrelevant information? *

☐ Yes

☐ No

8. Did the model's answers cover all aspects of the questions? *

☐ Yes

☐ No

9. Were there any important details or aspects missing from the responses? *

10. How clear and understandable were the model's responses? *



11. Were there any instances where the language used was confusing or ambiguous? *

12. Rate the consistency of the model's responses across different questions and sessions. *



13. Did you notice any contradictions or discrepancies in the answers provided? *

Real-Time Interaction Feedback

General

14. How timely and efficient were the model's responses during the interaction? *



15. Did the model respond promptly to your questions? *

☐ Yes

☐ No

16. How was your experience interacting with the model? *

17. Did you encounter any usability issues or difficulties? *

Specific Feedback

18. Can you provide examples of responses where the model performed particularly well? *

19. What aspects of these responses made them stand out? *

20. Can you provide examples of responses where the model underperformed? *

21. What specific issues did you notice in these responses? *

22. What specific improvements would you suggest to enhance the model's performance? *

23. Are there any additional features or capabilities that you believe would benefit the model? *

Technical Evaluation

24. How technically accurate were the model's responses, especially in complex or specialized topics? *



25. What specific technical knowledge or details do you think should be added to the model? *

26. Were there any technical errors or misunderstandings in the answers? *

27. How well did the model demonstrate depth of knowledge in the subject matter? *



28. Were the responses sufficiently detailed and informative? *

29. How effectively did the model handle ambiguous or unclear questions? *



Additional Comments

30. In your opinion, how effective is the AI-based Q&A model in ensuring compliance with Chapter 14 of the building code? *

31. In your opinion, how can the tacit knowledge of experts in building code in general and chapter 14 in particular be codified into a structured, accessible system for use by stakeholders? *

32. Do you have any additional comments or observations about the model's performance? *

33. Is there anything else you would like to share regarding the evaluation? *

34. Did the model seek clarification or make reasonable assumptions? *

This content is neither created nor endorsed by Microsoft. The data you submit will be sent to the form owner.

Appendix (E):

Meeting Questions for Use

Focus Group Meeting with Users (2)

Dear Participant,

Thank you for participating in this research study. Your involvement will help us assess the adaptability and usefulness of a custom Q&A model specifically designed to address questions related to the Architecture, Engineering, and Construction (AEC) industry. The primary goal of this study is to gather feedback from end-users like you on how effectively the model supports practical applications within this field.

Before the focus group meeting, you are invited to use the model through the link provided below. The model is designed to answer questions related to **Law No. 51 of 2014** with respect to **Regulating the Practice of Engineering Professions** and its **Implementing Regulations**, which form the basis for **Chapter 14** of the Building Code (Unified Guidebook of Building Permit Regulations).

You can explore the model to answer questions such as:

- How can I obtain a license to practice engineering in Bahrain?
- Can I practice engineering without a license in Bahrain?
- What is Law No. 51 of 2014?
- What is the Implementing Regulation?
- What does CRPEP stand for?

Following your interaction with the model, you will join a focus group discussion to share your insights and experiences. The date, time, and venue for this discussion will be confirmed later.

Instructions:

1. **Read and sign the participant consent form (attached)** before using the model and attending the focus group meeting. Please bring the signed form with you to the meeting.
2. **Sign up for ChatGPT** (ChatGPT | OpenAI).
3. **Engage with the model** using the provided link to ask questions or seek clarifications related to **Law No. 51 of 2014 and its Implementing Regulations only**.
4. **Complete this user feedback form**.
5. **Please note:** This model is for trial purposes only, exclusively for participants in this focus group meeting. It is not an official tool released by the concerned authority, and the link to the model will be disabled after the study is completed. Therefore, **please do not share this link with anyone else**.

Should you have any questions or need further clarification before our session, please do not hesitate to contact me. Your participation is highly valued, and I look forward to our collaborative efforts in advancing this research.

Thank you once again for your significant contribution.

Entesar,

* Required

1. Please provide information about your engineering discipline (e.g. architect, civil engineer, etc.), designation, total years of experience, and organization name (optional). *

2. What challenges did you face when navigating the information through traditional methods?

*

3. How did using the custom GPT model impact your ability to find accurate and relevant information? *

4. Would a model that provides instant, accurate, and relevant answers with references improve your decision-making and compliance efforts? *

5. How useful do you find the Custom GPT model in relation to Law No. 51 of 2014 and its implementing regulations and CRPEP's regulations (Chapter 14 of the Building Code ((Unified Guidebook of Building Permit Regulations)))? *

6. Do you believe that applying the same approach to create a custom GPT for the entire Building Regulation (Unified Guidebook of Building Permit Regulations) would help you and other stakeholders in the AEC industry make informed and optimized decisions? *

7. Are there any specific features or capabilities you believe would enhance the model's performance? *

8. How do you foresee the Custom GPT model assisting you during different phases of a project, from design to construction and completion, and finally to operation and maintenance? *

9. After using the custom GPT model during this meeting, can you provide examples of how it helped you make decisions more efficiently? *

10. In what ways do you think the custom GPT model could improve collaboration and communication among different stakeholders in a project? *

11. Do you have any additional comments or observations about the model's performance? *

This content is neither created nor endorsed by Microsoft. The data you submit will be sent to the form owner.

الملخص

يمثل الامتثال التنظيمي في قطاع الهندسة المعمارية والإنشائية والبنية التحتية تحديًا مستمرًا لأصحاب المصلحة، حيث تؤدي التعقيدات القانونية والاختلافات في التفسيرات إلى تأخير المشاريع وزيادة التكاليف. تسعى هذه الدراسة إلى معالجة هذه المشكلات من خلال تطوير نموذج مخصص قائم على الذكاء الاصطناعي وقادر على توفير إجابات دقيقة وفورية للأسئلة التنظيمية، مما يساعد في تحسين عملية اتخاذ القرار وضمان الامتثال للوائح. تم تطبيق الدراسة في مملكة البحرين، حيث استخدم الفصل (14) من الدليل الموحد لاشتراطات البناء والتشريعات الأخرى المتعلقة به كمصدر أساسي للبيانات. قامت الدراسة بمقارنة عدة نماذج للذكاء الاصطناعي (LLMs) بهدف تحديد النموذج الأكثر كفاءة في معالجة مجموعة بيانات صغيرة متعلقة بالقوانين واللوائح التنظيمية. وأظهرت النتائج أن الـ (CustomGPT) هو الخيار الأمثل، حيث تفوق في التكيف مع البيانات المحدودة وتقديم إجابات دقيقة ومتسقة مقارنةً بالنماذج الأخرى، حيث حقق نسبة أداء بلغت 100% في جميع مؤشرات التقييم، في حين سجل النموذج الأساسي ChatGPT-4.0 دقة بلغت 98.8%.

وتعتمد منهجية البحث على إطار عمل من عشر خطوات تم تصميمه خلال هذه الدراسة، يجمع بين الذكاء الاصطناعي والتدخل البشري والخبرات المتخصصة لضمان دقة النموذج وملاءمته للسياق التنظيمي. ويستند البحث إلى نظرية إدارة المعرفة (KM)، والتي تتيح تحويل المعرفة الضمنية لدى الخبراء إلى معلومات واضحة ومنظمة، مما يعزز إمكانية الوصول إليها ويوفر مرجعًا موثوقًا للمتخصصين في المجال، وشملت مقابلات الدراسة 66.7% من الخبراء في الجهة المسؤولة عن تطبيق الفصل (14).

وقد أظهرت نتائج الدراسة أن النموذج يحسن دقة القرارات، ويسهل الوصول إلى المعلومات التنظيمية، ويعزز التعاون بين مختلف الأطراف المعنية. كما يتميز بإمكانية التوسع ليشمل مجالات تنظيمية أخرى، ما يجعله أداة قوية لدعم التحول الرقمي في إدارة الامتثال. ورغم الفعالية العالية للنموذج، إلا أن الدراسة تُقر ببعض التحديات، مثل الحاجة إلى تحسين توافر البيانات، والتكيف مع الاختلافات القانونية، والتعامل مع الاستفسارات المعقدة التي تتطلب تفسيرات أعمق.

وتوصي الدراسة بتطوير إمكانيات دعم اللغات المتعددة، واعتماد أطر ذكاء اصطناعي أكثر تكيفاً، ودمج النموذج مع أنظمة نمذجة معلومات البناء (BIM) لتعزيز تطبيقاته العملية وجعله أكثر توافقاً مع احتياجات القطاع. وتقدم هذه الدراسة مساهمة مهمة في مجال استخدام الذكاء الاصطناعي لضمان الامتثال التنظيمي، حيث توفر إطاراً قابلاً للتطوير يمكن توظيفه في تحسين عمليات اتخاذ القرار، وتعزيز التعاون بين أصحاب المصلحة، ودعم الابتكار في المشاريع الهندسية متعددة الأطراف.



جامعة البحرين

كلية إدارة الأعمال

قسم الإدارة والتسويق

تحسين عملية اتخاذ القرار في مشاريع العمارة والهندسة والإنشاء متعددة أصحاب المصلحة: إطار منهجي مبتكر قائم على البيانات

أطروحة مقدمة استكمالاً لمتطلبات الحصول على درجة الدكتوراه في تحليل الأعمال

مقدم من

انتصار جاسم مهدي النامة
19990435

الإشراف

د. مقصود محمود
المشرف

د. هدى المدحوب
المشرف المشارك

مملكة البحرين

مايو 2025