

Univerza v Ljubljani  
Fakulteta *za gradbeništvo*  
*in geodezijo*



**FAITH SHARON TANGARA**

INTEGRATING BIM MODELS AND AI MODELS FOR COST  
ESTIMATION

MASTER THESIS

SECOND CYCLE MASTER STUDY PROGRAMME BUILDING  
INFORMATION MODELLING – BIM A+

Ljubljana, 2023

Univerza v Ljubljani  
Fakulteta *za gradbeništvo*  
*in geodezijo*



**FAITH SHARON TANGARA**

INTEGRATING BIM MODELS AND AI MODELS FOR COST  
ESTIMATION

INTEGRACIJA MODELOV BIM IN MODELOV AI ZA OCENO  
STROŠKOV



European Master in  
Building Information Modelling

Master thesis No.:

Supervisor:  
Assist. Prof. Tomo Cerovšek, Ph.D.

Ljubljana, 2023



## **ERRATA**

<b>Page</b>	<b>Line</b>	<b>Error</b>	<b>Correction</b>
-------------	-------------	--------------	-------------------

*»This page is intentionally blank«*

## **BIBLIOGRAFSKO – DOKUMENTACIJSKA STRAN IN IZVLEČEK**

<b>UDK:</b>	<b>004.946:69-047.74(043.3)</b>
<b>Avtor:</b>	<b>Tangara Faith Sharon</b>
<b>Mentor:</b>	<b>doc. dr. Tomo Cerovšek</b>
<b>Somentor:</b>	<b>Integracija modelov BIM in modelov AI za oceno stroškov</b>
<b>Tip dokumenta:</b>	<b>Magistrsko delo</b>
<b>Obseg in oprema:</b>	<b>74 str., 34 sl., 6 pregl.,</b>
<b>Ključne besede:</b>	<b>Ocena stroškov, popisi, umetna inteligenca, veliki jezikovni modeli, BIM</b>

### **Izvelek:**

Ocena stroškov ostaja ključni del projektnega vodenja, ki usmerja načrtovanje in sprejemanje odločitev skozi vse faze projektov. Tradicionalne metode so omejene zaradi kompleksnosti gradbenih elementov, zahtevane natančnosti in kratkih rokov. Kljub napredku informacijskega modeliranja zgradb (BIM), se pojavlja vse večja potreba po natančnejših, učinkovitejših in avtomatiziranih metodah izdelave ocen stroškov. Zato se magistrska naloga osredotoča na boljše razumevanje izdelave popisov in stroškovnikov na osnovi BIM ter potencial avtomatizacije.

Cilj študije je prikazati potencial umetne inteligence za izboljšanje natančnosti, učinkovitosti in racionalizacijo procesov pri izdelavi popisov in oceni stroškov, kar bi lahko služilo kot ogrodje sistema za podporo odločanju popisovalcem in drugim deležnikom projektov. Pri raziskovanju potenciala umetne inteligence se osredotočamo zlasti na napredno analitiko in velike jezikovne modele LLM (angl. Large Language Models), kot je ChatGPT, in na druge tehnologije v povezavi z BIM. Z integracijo LLM, BIM in AI bi lahko izdelavo popisov in oceno stroškov bistveno izboljšali tako v smislu kakovosti in učinkovitosti, medtem bi znatno zmanjšali obseg »ročnega dela«. Da bi lahko bolje izkoristili potencial umetne inteligence in BIM, bi bilo treba izboljšati interakcijo med LLM in modeli BIM ter uvesti napredno analitiko zgodovinskih podatkov kot učnih nizov, s katerimi bi bilo mogoče ogrodje aplicirati na razne vrste in velikosti projektov.

*»This page is intentionally blank«*

## **BIBLIOGRAPHIC– DOKUMENTALISTIC INFORMATION AND ABSTRACT**

**UDC:** 004.946:69-047.74(043.3)

**Author:** Tangara Faith Sharon

**Supervisor:** Assist. Prof. Tomo Cerovšek, Ph.D.

**Title:** Integrating BIM Models and AI Models for Cost Estimation

**Document type:** Master thesis

**Scope and tools:** 74 p., 34 fig., 6 tab.,

**Keywords:** Cost estimation, quantity take off, artificial intelligence, large language models, BIM

### **Abstract:**

Cost estimation remains a critical part of project management, guiding the budgeting and decision making throughout project phases. Traditional methods are limited due to complexity of building elements, required accuracy, and tight schedules. Despite recent advancements in Building Information Modelling (BIM) there is a growing demand for more accurate, efficient, and automated methods of cost estimation. Therefore, the thesis focus is given to understanding BIM-based Quantity Take-Off (QTO) and automation potential.

The goal of the research is to demonstrate the potential of AI to enhance accuracy, efficiency, and streamline cost estimation processes that would support especially cost estimators, and also other project stakeholders. This research explores the potential of AI, specifically Large Language Models (LLMs) like ChatGPT, and other technologies in conjunction with BIM models. By integrating LLM, BIM, and AI cost estimation could be improved in terms of quality and efficiency, while manual work could be reduced significantly. To fully harness the potential of AI and BIM the interaction between LLMs and BIM models should be enhanced, as well as advanced analytics of historical data that would serve as learning sets should be put in place to allow for framework's applicability on diverse project types and scales.

*»This page is intentionally blank«*



## ACKNOWLEDGEMENTS

I'm profoundly grateful to GOD for granting me the opportunity to pursue this master program and to the BIM A+ Consortium for providing the platform for my academic and career growth. Prof. Tomo Cerovsek's exceptional guidance and expertise has been instrumental during my research period. My heartfelt appreciation to my dear course mates for their camaraderie and support, which have enriched my education and made this journey truly memorable.

My gratitude extends to my beloved family, especially my dad and siblings, whose enduring love and sacrifices have underpinned my determination. I am also appreciative of my best friend, whose unwavering friendship and support have been a constant source of motivation throughout this academic endeavor.

I extend my sincere thanks to my friends, who were both physically present and connected remotely, for their encouragement and unwavering belief in my capabilities. Finally, I wish to acknowledge my employer, manager, and fellow colleagues for their understanding and flexibility, which allowed me to balance my academic pursuits with my professional commitments.

Your unwavering support, encouragement and prayers didn't go unnoticed.

Thank you.

*»This page is intentionally blank«*

## TABLE OF CONTENTS

<b>ERRATA.....</b>	<b>i</b>
<b>BIBLIOGRAFSKO – DOKUMENTACIJSKA STRAN IN IZVLEČEK.....</b>	<b>iii</b>
<b>BIBLIOGRAPHIC– DOKUMENTALISTIC INFORMATION AND ABSTRACT .....</b>	<b>v</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>vii</b>
<b>TABLE OF CONTENTS .....</b>	<b>ix</b>
<b>INDEX OF FIGURES .....</b>	<b>xi</b>
<b>LIST OF TABLES.....</b>	<b>xii</b>
<b>1. INTRODUCTION.....</b>	<b>1</b>
<b>1.1. Background and motivation.....</b>	<b>1</b>
<b>1.2. Research objectives.....</b>	<b>3</b>
<b>1.3 Value Proposition.....</b>	<b>4</b>
<b>1.4 Structure of the thesis.....</b>	<b>5</b>
<b>2. LITERATURE REVIEW .....</b>	<b>6</b>
<b>2.1 BIM based quantity take off and estimation in different project phases.....</b>	<b>6</b>
<b>2.2 AI techniques and applications.....</b>	<b>10</b>
<b>2.3 Interface between BIM and AI .....</b>	<b>13</b>
<b>2.4 Machine learning applications.....</b>	<b>16</b>
<b>2.5 Examples AI to quantity surveying .....</b>	<b>21</b>
<b>3. METHODOLOGY .....</b>	<b>25</b>
<b>3.1 Scope Definition .....</b>	<b>25</b>
<b>3.2 Methodological framework.....</b>	<b>26</b>
<b>4. PROTOTYPE SYSTEM DEVELOPMENT .....</b>	<b>29</b>
<b>4.1 Requirements analysis .....</b>	<b>29</b>
<b>4.1.1 Functional Requirements:.....</b>	<b>29</b>
<b>4.1.2 Non-functional Requirements:.....</b>	<b>30</b>

---

<b>4.1.3. Assumptions and Dependencies</b> .....	31
<b>4.2 Use case diagrams</b> .....	32
<b>4.2.1 Relationship Establishment</b> .....	35
<b>4.3 Functional model</b> .....	37
<b>4.4 System architecture</b> .....	38
<b>4.4.1 Proposed Framework for Ai-Driven Cost Estimation</b> .....	46
<b>4.4.2 Integrating AI with BIM Models</b> .....	46
<b>4.4.3 IfcCostItem</b> .....	51
<b>4.4.4 Integrating AI with Cost Data</b> .....	53
<b>4.4.5. General database architecture for AI application</b> .....	55
<b>4.5 Interface prototype</b> .....	57
<b>5. CASE STUDY</b> .....	<b>60</b>
<b>5.1 Project description</b> .....	60
<b>5.2 Standard methods for BIM based.</b> .....	61
<b>5.3 Demonstration of non-standard methods</b> .....	62
<b>6 DISCUSSION AND CONCLUSIONS</b> .....	<b>65</b>
<b>6.1 Assessment of the Results</b> .....	65
<b>6.2 Accuracy of the Semantic Parsing</b> .....	67
<b>6.2 Limitations of the study</b> .....	69
<b>6.3 Future work</b> .....	70
<b>7 REFERENCES</b> .....	<b>72</b>

## INDEX OF FIGURES

<i>Figure 1:Structure of the thesis framework</i> .....	6
<i>Figure 2:Sequence of quantity takeoff throughout a project's life span (Алиабаб et al., 2017) ...</i>	7
<i>Figure 3:Procedure for identifying and choosing research studies. ....</i>	9
<i>Figure 4:Overview of artificial intelligence and its subfields (Abioye et al., 2021) .....</i>	11
<i>Figure 5:Market Basket Analysis ARM Services Process Flow (Oracle E59126-02,2014.).....</i>	19
<i>Figure 6:QTO based on classification system. ....</i>	22
<i>Figure 7:UNIFORMAT II is the elemental classification (Charette &amp; Marshall, 1999) .....</i>	23
<i>Figure 8: Methodological framework for AI-driven cost estimation.....</i>	26
<i>Figure 9:General use case by cost estimator .....</i>	32
<i>Figure 10:Use case diagram for automated estimation process .....</i>	34
<i>Figure 11:Interaction sequence diagram for cost estimation.....</i>	35
<i>Figure 12:AS-IS Cost estimation process of BIM models. ....</i>	39
<i>Figure 13:5D framework for BIM models .....</i>	40
<i>Figure 14:Classification framework for cost estimation (BIM A+ 5) .....</i>	41
<i>Figure 15:Cost elements of an external wall.....</i>	45
<i>Figure 16:Cost breakdown of individual walls of a model.....</i>	45
<i>Figure 17:Process of extraction of keywords (Mengtian Yin, 2023).....</i>	49
<i>Figure 18:Semantic mapping of building (Pauwels P, Van Deursen D., 2011).....</i>	50
<i>Figure 19:IFC Cost Composition (Building Smart Industry Foundation Classes 4.0.2.1) .....</i>	52
<i>Figure 20:Cost elements in a model and its prices.....</i>	52
<i>Figure 21:Architecture of the AI application .....</i>	55
<i>Figure 22: Retrieval-augmented generation of responses.....</i>	57
<i>Figure 23:Setting up sample query as parameter for the prototype system .....</i>	58
<i>Figure 24:User prompt matches trained data .....</i>	59
<i>Figure 25:3D model of apartment building for case study.....</i>	60
<i>Figure 26:Open BIM-oriented QTO procedure (Choi et al., 2015).....</i>	61
<i>Figure 27:Information retrieval of floor finish.....</i>	62
<i>Figure 28:Information retrieval of staircase and cost database.....</i>	63
<i>Figure 29: Section of cost databse.....</i>	63

<i>Figure 30:General information retrieved from ChatGPT</i> .....	64
<i>Figure 31:Information retrieved from the model</i> .....	65
<i>Figure 32:Semantic grouping of elements</i> .....	66
<i>Figure 33:Data preprocessing from the model (Zheng &amp; Fischer, 2022)</i> .....	67
<i>Figure 34:Cost description annotated and tagged for mapping.</i> .....	67

## LIST OF TABLES

Table 1:Implementation of AI subdomains in the construction sector (Abioye et al., 2021).....	13
Table 2:AI implementation in quantity surveying. ....	17
Table 3:Classification breakdown of a model.....	43
Table 4:Modules of Langchain .....	54
Table 5:Semantic parsing evaluation of user prompts vs information retrieval .....	68
Table 6: SWOT analysis of proposed prototype .....	69

## **1. INTRODUCTION**

### **1.1. Background and motivation**

Cost estimation is a crucial aspect of construction project management as it provides a basis for budgeting and decision-making processes (Sayed et al., 2020). However, cost estimation is a complex and time-consuming task, which necessitates the need to develop accurate and reliable cost estimation methods. Traditional cost estimation methods, such as the floor area method or cube method have limitations in terms of accuracy and reliability, especially in the early stages of design when information is limited. In recent years, the use of Building Information Modelling (BIM) has become increasingly popular in the construction industry, with potential benefits for cost estimation (Wang et al., 2019).

The process of integrating cost and scheduling data into the 5D model can be intricate, especially when dealing with large, complex projects that involve various construction techniques. Even with BIM, the process of cost estimation in construction projects remains a complex task due to the need to merge information from various sources of both the model cost elements and the market prices. However, this complexity opens opportunities for artificial intelligence (AI) to play a significant role in simplifying these challenges.

AI-enabled approximation tools enhance the precision of cost forecasts by examining extensive datasets and discerning patterns and correlations among diverse project variables. Through machine learning algorithms, these tools can infer from historical data, delivering accurate cost estimations for projects tailored to their unique project parameters.

Cost consultants, entrusted with the provision of accurate cost estimates and aiding clients in making informed decisions regarding construction projects, can significantly benefit from AI-based tools. These instruments facilitate the generation of cost predictions rooted in historical data analysis and project parameters. By enabling cost consultants to identify potential cost reductions and providing insights into project feasibility, AI tools help clients to make data-informed investment decisions.

Moreover, the amalgamation of AI-based cost approximation and consultancy tools with BIM can significantly amplify the precision and efficiency of cost estimation. BIM presents a 3D model of the project, enabling the automatic extraction of data pertinent to project parameters and material quantities. Through the integration of AI-based cost estimation tools with BIM, project teams can automate the cost estimation process and enhance the accuracy of estimates.

Therefore, the motivation behind this research is to explore the potential of AI for estimation in the construction industry. The research will investigate how AI models can be integrated with BIM models for the accuracy and efficiency of cost estimation in the early stages of design, particularly for schematic designs, and how it can improve cost consultancy services.



## 1.2. Research objectives

The purpose of this study is to investigate the effectiveness of using AI for cost estimation of BIM models and to explore how this approach can reduce the time spent in generating cost estimates at the preliminary phases of a project.

The specific objectives of the research are:

1. **To investigate the current practice of BIM-based quantity takeoff and cost estimation:** understand the traditional and BIM-based methods used for quantity takeoff and cost estimation in construction. Identify any existing limitations or areas that might benefit from a new approach.
2. **To study the principles and applications of different AI techniques:** explore the methodology and effectiveness of AI with a focus on potential relevance to the construction sector.
3. **To conceptualize a framework for integrating AI into construction quantity takeoff and cost estimation:** develop a new model that combines AI technique with the prevalent practices of quantity takeoff and cost estimation.
4. **To validate the proposed framework using real-world construction data:** gather construction project data to apply the newly designed framework. Test the model's efficiency and reliability, compare the output with traditional methods, and evaluate its accuracy.
5. **To measure the impact of the AI solution on construction quantity takeoff and cost estimation:** assess the potential benefits of applying the new market basket analysis solution

to quantity takeoff and cost estimation. Quantify the benefits and improvements over traditional methods in terms of time, and cost.

### **1.3 Value Proposition**

The contemporary process of delivering construction projects entails a vast volume of data, which poses significant challenges when relying on manual analysis using conventional rule-based techniques. This approach is not only laborious but also time-consuming. Quantity surveyors face the demanding task of performing quantity take-offs, measurements, rate analysis, generating bills of quantities and cost estimation. Additionally, they must handle progress valuations and respective progress payments and variations. Implementing AI-driven solutions, already successfully employed in various industries, offers the potential to enhance productivity, optimize resource allocation, mitigate risks, and ultimately achieve better outcomes for projects (Doe, 2021).

Undoubtedly, integrating these tools and methods presents unique challenges for project-driven and project-organized businesses like quantity surveying, in contrast to more repetitive and somewhat predictable processes found in industries such as manufacturing. However, the contemporary landscape offers a multitude of open-source tools, technological advancements, and the widespread adoption of Building Information Modeling (BIM). These factors significantly lower the barriers to entry for leveraging data already collected and applying machine learning algorithms and analytics to unearth valuable insights that augment decision-making processes.

This study showcases the practical application of modern machine learning algorithms and methods in quantity take-off (QTO) and cost estimation, effectively bridging the gap between academia/research and practical implementation. Furthermore, this research discusses use cases, offering examples of how quantity surveyors can seamlessly integrate AI to achieve prompt results.

## **1.4 Structure of the thesis**

This dissertation is structured in six chapters: introduction, review of the state of art, methodological framework, system development, case study and conclusions. On the review of the state of the art was researched about the BIM-based estimation methods, showing its concepts. Existing AI techniques with the respective applications in the industry are also explored as well as the integration of AI with BIM. The third chapter highlights the methodological framework of cost estimation the proposed workflow is introduced. Besides, the workflow was demonstrated to be applied to a hypothetical project. On the case study was applied an application of the proposed work, gathering information from a BIM model, interaction with user prompts, and using it on a future project under development. The results and considerations regarding the of the workflow were also presented. On the last chapter were made final comments about the proposed dissertation work, what were the perceived benefits of its application and suggestion for future developments.

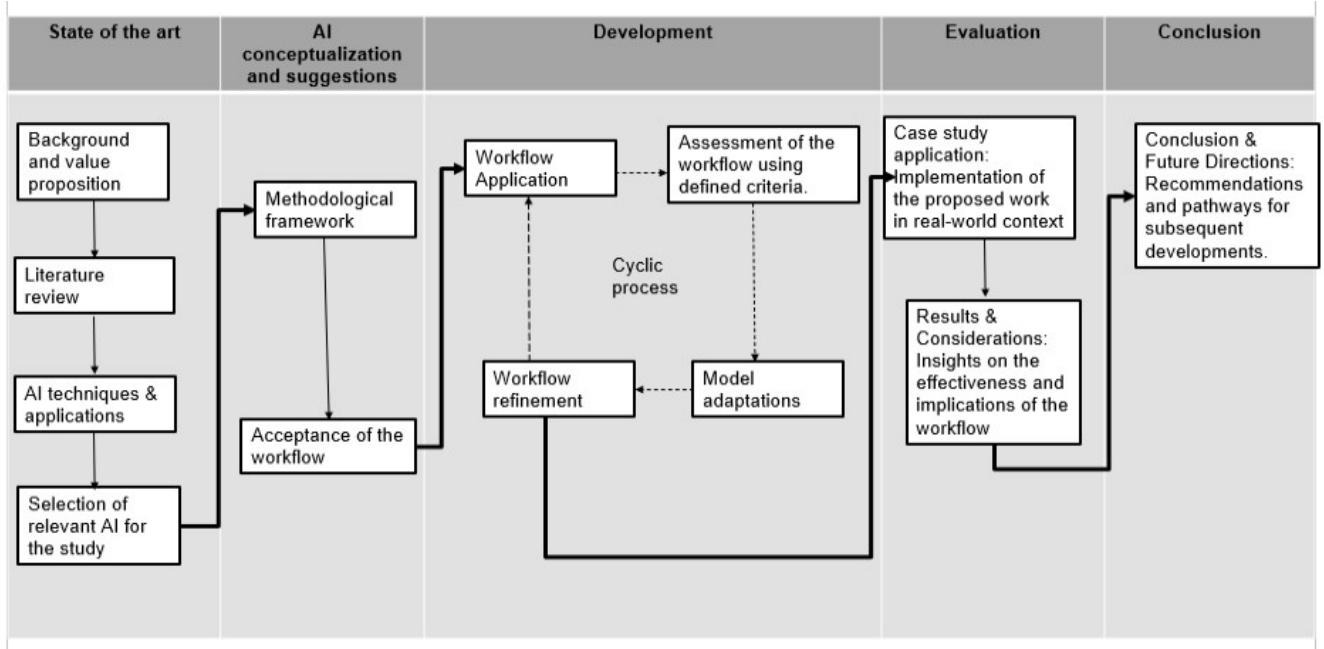


Figure 1: Structure of the thesis framework

## 2. LITERATURE REVIEW

### 2.1 BIM based quantity take off and estimation in different project phases

The process of quantity take-off (QTO) is an intricate and indispensable component that plays a pivotal role throughout the entire life cycle of a construction project. Its significance is derived from the multifaceted functions it fulfills at various stages of the project, from preliminary estimation, tendering, construction planning, and cost control (Wijayakumar & Suranga Jayasena, 2013). When utilizing BIM for quantity takeoff, the process becomes more efficient and accurate. BIM models contain comprehensive and detailed metadata about the project's elements, such as spatial relationships, measurements, shapes, and orientations of these elements, providing a visual understanding of the project's physical characteristics. Additionally, BIM models may incorporate

basic properties such as materials and even textures. This rich data allows for automated and precise extraction of quantities, eliminating manual calculations and potential errors (Tang et al., 2022).

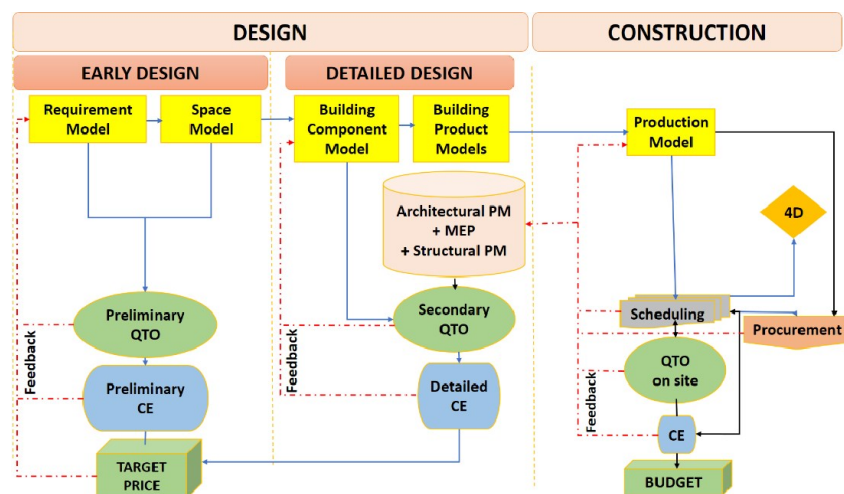


Figure 2: Sequence of quantity takeoff throughout a project's life span (Алиабаб et al., 2017)

In the preliminary stages of a construction project, QTO serves as the bedrock for establishing a reliable and realistic cost estimate. By meticulously quantifying and itemizing the materials, resources, and labor required for the project, it provides a solid foundation upon which the initial cost estimate is built. This estimation serves as a fundamental reference point for project stakeholders, enabling them to establish feasible budgets. BIM-based QTO at this stage leverages the detailed information embedded in the model to generate accurate quantities. By integrating the automated quantity extraction process with the BIM model, estimators can quickly and reliably obtain the quantities of materials, components, and resources required. This integration not only saves time but also improves the accuracy of the estimates, as any changes made to the BIM model are automatically reflected in the quantities (Tang et al., 2022).

During the tendering stage, QTO serves as a determinant of both project costs and the duration of construction activities. Contractors and bidders heavily rely on accurate and detailed BIM-based QTOs to generate more comprehensive and competitive bids. The detailed and accurate quantities obtained from the BIM models facilitate precise cost calculations, leading to more accurate pricing. This, in turn, enhances the contractor's ability to provide competitive proposals to potential clients (Tang et al., 2022).

As the project progresses to the construction stage, quantity take-off continues to be of utmost importance. The BIM model serves as a dynamic tool to review quantities and associated information. This ensures that everyone is working from the same accurate data. Furthermore, BIM-based QTO facilitates efficient construction planning and resource allocation. The detailed quantities extracted from the BIM model allow for optimized scheduling and sequencing of activities. Contractors can allocate resources based on the accurate quantities, minimizing waste and improving productivity. Additionally, any changes or revisions made to the BIM model automatically updates the quantities, enabling real-time adjustments to the construction plan. This enables effective cost control and helps maintain the economic balance of the contractor's finances throughout the project's execution (Olsen & Taylor, 2017).

In essence, the accuracy of the QTO is paramount for maintaining financial equilibrium. It provides a comprehensive analysis of productivity and the diverse types of costs associated with a specific project. By diligently quantifying the required materials, resources, and labor, it enables contractors to make informed decisions, manage costs, optimize resource allocation, and ensure the successful completion of the project within the allocated budget.

This comprehensive overview of BIM-based QTO and estimation in different project phases underscores the importance of BIM in enhancing the accuracy and efficiency of the construction

process. However, with the increasing adoption of BIM-based QTO, artificial intelligence (AI) integration is becoming increasingly necessary to facilitate efficient QTO processes and cost analysis for better estimation and feasibility studies too. Therefore, there is a need for developing and applying new methods and tools that can not only improve efficiency of quantification but also contribute to cost analysis for completeness, and consistency.

In pursuit of the literature review objective, a general framework presented in Figure 2 is proposed to critically assess the synthesis of BIM and AI. The initial stage of this research involves a comprehensive literature search using the Google Scholar database to retrieve relevant scientific articles. After selecting suitable literature that satisfies the selection criteria, it is gathered and stored in a database, which serves as the basis for further exploration. Connected Papers platform is then employed to generate informative conceptualizations, facilitating systematic mapping and exploration of the structural and dynamic aspects of the curated literature dataset. Finally, information analysis is conducted to facilitate knowledge discovery, leading to insights into the potential of AI in developing new cost estimation processes for BIM models.



Figure 3: Procedure for identifying and choosing research studies.

## 2.2 AI techniques and applications

Artificial Intelligence (AI) involves the creation and development of machines and computer programs that exhibit intelligence. While it shares similarities with the endeavor of utilizing computers to comprehend human intelligence, AI is not strictly limited to techniques that are observable in biological systems (IBM, 2013). AI can be classified into three types: Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGI), and Artificial Super Intelligence (ASI). ANI, also known as weak AI, is a form of AI where machines demonstrate intelligence in a specific domain such as chess playing, sales prediction, movie suggestions, language translation, and weather forecasts. AGI, or strong AI, aims to create machines that operate at the same level as humans. It involves creating machines capable of solving a range of complex problems in different domains, controlling themselves autonomously, with their own thoughts, strengths, and weaknesses. This remains a major goal of AI, but it has proven to be challenging and elusive. ASI is concerned with building machines that surpass human capabilities across multiple domains (Abioye et al., 2021).

To understand the current state of AI in the construction industry, it is important to identify the main sub-tasks of AI. These include machine learning, natural language processing, knowledge-based programming, optimization, automated scheduling.





Figure 4: Overview of artificial intelligence and its subfields (Abioye et al., 2021)

Machine Learning (ML), a subfield of AI, involves the creation of algorithms that allow computers to learn from and make decisions or predictions based on data. In the construction industry, ML can be used to predict project outcomes based on historical data. (Egwim et al., 2021), discusses how a supervised learning model could analyze past construction projects to predict potential delays, enabling proactive mitigation strategies. Computer Vision, another multidisciplinary field of AI, aims to simulate the human visual system, enabling machines to interpret and understand visual data (Abioye et al., 2021). Computer vision can be used to analyze images or video from construction sites, identifying potential safety hazards or tracking the progress of ongoing work. This can facilitate decision-making, improve safety standards, and ensure timely project completion. Automated Planning and Scheduling involves creating systems that can plan actions to achieve a goal (Ghallab et al., 2004). This could involve scheduling tasks for a project,

considering factors such as resource availability and potential risks. This can lead to more efficient use of resources and improved project timelines (Abioye et al., 2021).

Robotics, an interdisciplinary field involving the design and operation of robots, is increasingly being incorporated into construction. Robots, equipped with sensors and actuators, can perform specialized tasks such as bricklaying or material transportation, increasing efficiency and reducing the risk of human injury. Natural Language Processing (NLP) involves creating computational models that mimic human linguistic capabilities. NLP could be used to analyze project documents or communications, extracting key information and insights (Abioye et al., 2021).

Lastly, Optimization in AI involves making decisions that provide the best outcomes given a set of constraints (Bini, 2018). This could involve optimizing resource allocation or project schedules to minimize costs and maximize efficiency. This is often achieved through the use of evolutionary algorithms, a family of metaheuristic algorithms created with the advent of AI.

Table 1 provides an overview of how various subfields of Artificial Intelligence (AI) are applicable to different areas of construction. It serves as a quick reference matrix to understand the intersection of AI technology and construction tasks, showcasing where each AI subfield can be leveraged to enhance efficiency and effectiveness in the industry.

<i>Construction Application Areas</i>	<i>Machine Learning</i>	<i>Computer Vision</i>	<i>Automated Planning &amp; Scheduling</i>	<i>Robotics</i>	<i>Knowledge-based Systems</i>	<i>Natural Language Processing</i>	<i>Optimization</i>
<i>Health and Safety</i>	✓	✓			✓	✓	
<i>Scheduling</i>	✓		✓				✓

<i>Cost Estimation</i>	✓					✓
<i>Legal (Contracts Conflict Management)</i>	✓	✓		✓	✓ &	
<i>Supply chain &amp; Logistics</i>	✓	✓		✓	✓	✓
<i>Site Monitoring &amp; Performance Evaluation</i>	✓	✓		✓	✓	✓
<i>Material Management</i>	✓			✓	✓	✓
<i>Offsite Assembly</i>	✓		✓			✓
<i>Plant and Equipment Management</i>	✓			✓	✓	✓
<i>Project Planning</i>	✓			✓	✓	✓
<i>Knowledge Management</i>	✓				✓	✓
<i>Design Risk Management</i>	✓				✓	✓
<i>Temporary Structures</i>	✓				✓	✓
<i>Bids/Tenders</i>	✓				✓	✓
<i>Energy Management</i>	✓				✓	✓
<i>Sustainability</i>	✓				✓	✓

Table 1: Implementation of AI subdomains in the construction sector (Abioye et al., 2021).

### 2.3 Interface between BIM and AI

The widespread adoption BIM in the AEC industry has motivated previous studies to investigate relationship between BIM and AI and the potential benefits of BIM in addressing AEC challenges (Pan & Zhang, 2023). Given the need for greater digitalization in the construction industry, it is

essential to reveal the meaningful relationship between BIM and AI to guide smart construction. In retrospect, these studies have predominantly depended on bibliometric analyses, which have proven to be inadequate in comprehensively examining the practical applications and potential advancements of BIM in qualitative terms. While bibliometric analyses can reveal insightful data on co-authorship trends within a field of research, they cannot provide a holistic understanding of the practicality and implementation of BIM and AI in the construction industry (Pan & Zhang, 2023).

To achieve a more thorough analysis, some studies have attempted to examine the contribution of BIM to specific topics such as: BIM in supporting green building development (Lu et al., 2017), reviewing the combination of BIM and geographical information systems (GIS) in a sustainable built environment (Wang et al., 2019), and presenting a comprehensive review of the current situation of BIM and internet of things (IoT) and device integration (Pan & Zhang, 2023). Although these reviews have primarily revealed the practical applications of BIM for specific purposes or processes, they have not placed sufficient emphasis on elevating BIM development to a higher level of intelligence.

Several studies have sought to combine BIM applications with specific subfields of AI, including Natural Language Processing (NLP). (Abdirad & Mathur, 2021) conducted research on construction project work and highlighted the significant dependence of construction practice on direct verbal and symbolic communication for knowledge and information sharing. This reliance can be attributed to the longstanding historical significance and inherent effectiveness of speech as a primary and instinctive mode of communication. The integration of voice interaction with BIM systems aims to establish a more natural and authentic representation of construction practices (Egwim et al., 2021). Additionally, other studies have explored the integration of AI and BIM with

various Industry 4.0 tools, such as Internet of Things (IoT) (Yao et al., 2017), smart cities, blockchain, augmented reality and quantum computing (Egwim et al., 2021).

Compared to conventional analytical approaches, AI is superior in tackling complex problems in the real world in a more automated and reliable way. It is helpful in facilitating a smarter decision-making process under less dependency on human experience and knowledge, aiming to maximize the chance of success in realizing the project goal. As earlier noted, the adoption of various AI technologies would bring sustainable benefits in terms of automation, improved efficiency, and quality project delivery to the AEC industry. This increasing usefulness of AI has resulted in a surge of research on the implementation of AI in the AEC industry since the beginning of the 21st century (Pan & Zhang, 2023). However, research regarding the integration of AI with BIM-based quantity takeoff is still in its initial stages, but there is potential for significant development in the upcoming years.

There are various possible reasons behind AI not reaching its full potential in the construction domain. This could be due to the fragmented nature of the industry, the rapid advancement in tech advancement vs slow adoption, high cost of software to use, large historical data, etc. (Abioye et al., 2021). Such foregoing limitations inspire this study to focus on the integration of BIM and AI models into QTO processes and cost estimation.

Since BIM and AI offer unique benefits in their respective domains, it is logical to focus on seamlessly integrating the rapidly growing array of AI methods into BIM-related projects. This integration aims to leverage the strengths of both BIM and AI, enabling easy access to available information and the automatic generation of deep insights for complex projects with minimal

manual work (Zhao, n.d.). As research has shown, the connection between BIM and various AI methods is a burgeoning area of study that promises to significantly transform the AEC industry.

## **2.4 Machine learning applications**

It is a common occurrence that the AEC industry produces vast amounts of data every day for each project. This data comes from sources like smart devices, IoT sensors, BIM, etc. and offers a chance for the industry and its customers to analyze and benefit from insights derived from past construction data with the help of AI and Machine learning (ML). AI can be defined as the field of research that focuses on creating "intelligent agents," which refer to devices capable of "perceiving and interpreting their environment and taking actions that optimize their chances of achieving a particular goal" (Bini, 2018). ML on the other hand is considered as a subfield of AI. It is a type of AI that gains knowledge from experience and enhances its ability to perform tasks over time (Bini, 2018). There are many ML algorithms such as Decision Trees, Genetic Algorithm, Logistic Regression, Neural Networks, Linear Regression, Nearest-Neighbor Mapping, Random Forests, and K-Means Clustering for implementing ML models. Selection of an appropriate ML algorithm is influenced by several factors including its usability, precision, training duration, and other considerations. With respect to construction, (Goh & Chua, 2013) suggested neural network analysis as an ML algorithm for construction safety management, (Chou & Lin, 2012) used four ML algorithms to present on predicting disputes in public-private partnership projects and (Zou & Ergan, 2019) suggested three ML algorithms to predict the impact of construction projects on urban quality of life.

A limited number of studies have explored the integration of AI and ML in construction QTO and analysis of cost historical data. Table 2 shows several studies related to quantity surveying and AI implementation.

*Table 2: AI implementation in quantity surveying.*

<b>Topic</b>	<b>QS related subject</b>	<b>AI Algorithm</b>	<b>Source</b>
Towards Automated Construction Quantity Take-Off: An Integrated Approach to Information Extraction from Work Descriptions	Quantity takeoff	Natural Language Processing (NLP)	(Tang et al., 2022)
Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning.	Cost estimation	Regression Analysis	(Kim et al., 2004)
Research on construction cost control based upon BP neural network and theory of constraint	Cost control	Backpropagation	(Han & Lu, 2010)
Web-based conceptual cost estimates for construction projects using Evolutionary Fuzzy Neural Inference Model	Cost estimation	Evolutionary Fuzzy Neural Inference Model	(Cheng et al., 2009)
Non-parametric bill-of-quantities Estimation of concrete road bridges' Superstructure: an artificial neural Networks approach	Cost estimation	Feedforward Neural Network	(Marinelli et al., n.d.)

As at the time of this study, there is no specific literature that has explored the use of ML algorithm and models for QTO and cost estimation. This inspires the motive for this research.

Machine learning techniques, such as Market Basket Analysis (MBA), have found innovative applications in various fields, including project management. MBA, a method rooted in the

principles of unsupervised learning, is traditionally utilized in the retail industry to discern patterns of products that are frequently purchased together. This is achieved using algorithms like the Apriori algorithm, which is designed to uncover hidden associations within large datasets (Silvers, 2011). When applied to cost estimation, MBA can analyze patterns of resource usage and cost databases across a multitude of projects during their respective construction phases. This analysis can provide insights into the quantities and combinations of resources that are likely to be required for new projects, thereby enhancing the accuracy of cost estimation (Yan et al., 2022).

MBA employs association rules in its data mining processes. Association rule mining (ARM) refer to a set of if/then statements used to discover relationships between variables in large data sets, relational database, or other information repository (Agrawal et al., 1993). Association rules are used to discover relationships between items that are frequently co-occurring in a dataset.



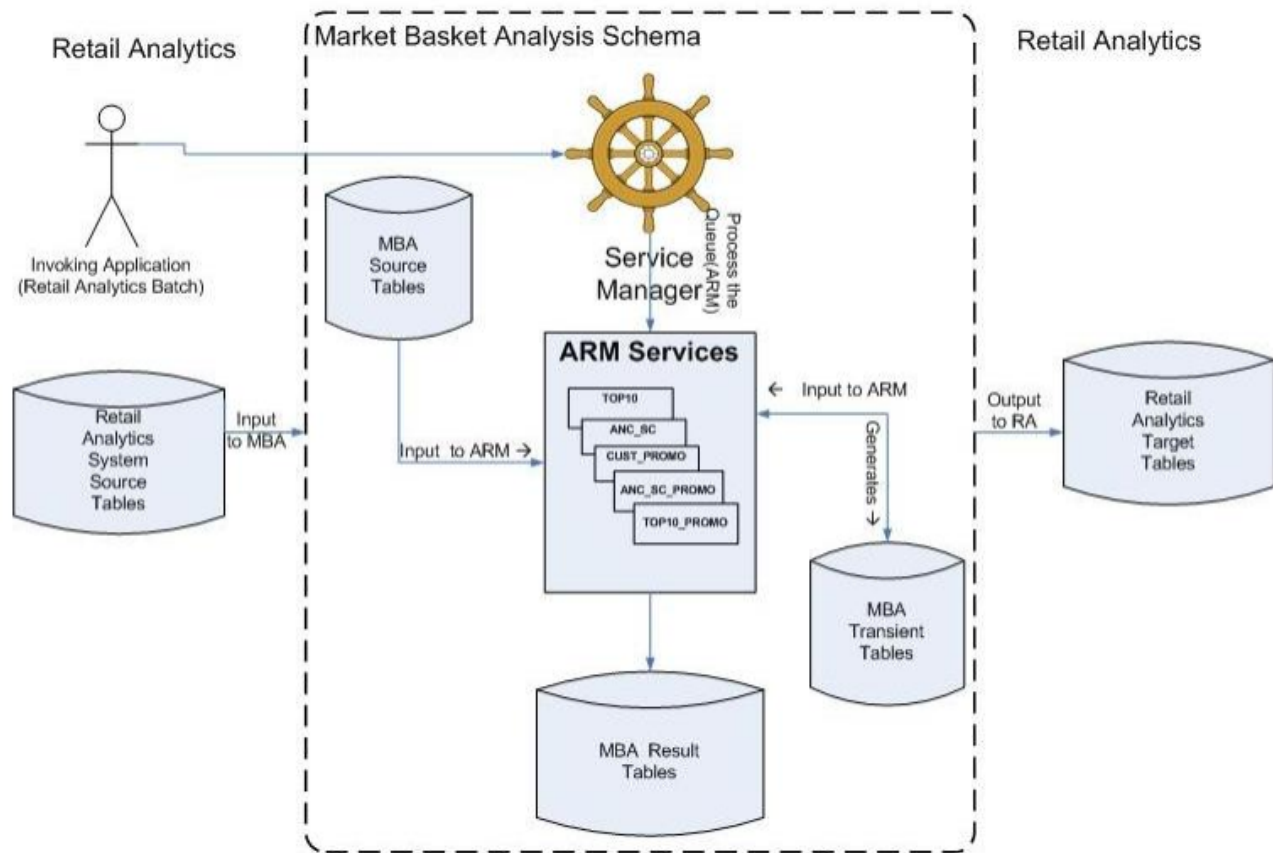


Figure 5: Market Basket Analysis ARM Services Process Flow (Oracle E59126-02, 2014.)

In the field of data mining, the fidelity of the input data forms the bedrock for the validity of the associations derived from Association Rule Mining (ARM) algorithms. This underscores the quintessential role of data accuracy and granularity in engendering association rules that are not merely reliable but also pragmatic in their application (Yan et al., 2022). The importance of precise data cannot be overstated, in construction. Such information forms the bedrock upon which the ARM algorithm functions to explain associations that could be instrumental in providing pivotal insights into the trajectory of the project.

Detailed data about material types and quantities can yield associations that shed light on trends in material usage, connecting specific materials to distinctive stages of construction or specific BIM elements of the model.

To enable such analyses, the user defines the entities within their BIM authoring tool for each layer and the composite as a whole. For instance, a concrete layer can be represented as an `IfcSlab`, `IfcWall`, or `IfcColumn`, while insulation can be denoted as an `IfcCovering`. A well-structured database within the BIM authoring tool is essential to support these definitions. It is crucial that the definition of entities follows a global function rather than an object-specific one.

During the export process, the BIM authoring tool appends the relevant entities and their corresponding Property Sets (`Psets`) based on the IFC standard to the respective layers and composites. This ensures that the necessary information is retained for subsequent analyses.

In the background, a mapping system is employed to transfer the properties from the `IfcBuildingElement` to the corresponding `IfcBuildingElementPart`. This ensures that the relevant attributes and characteristics are accurately propagated to the individual components of the building elements, enabling comprehensive analysis and evaluation.

By incorporating precise data, including comprehensive material quantities and their associated properties, the ARM algorithm can effectively uncover meaningful associations. These associations, in turn, link specific materials to distinct BIM elements.

The Levels of Detail (LOD) play a significant role in determining the precision and richness of data available for Association Rule Mining (ARM). While LOD 100, provides elementary, symbolic information about building elements, often proving too generic for detailed ARM

exploration, LOD 400 delivers comprehensive information on element assemblies, complete with fabrication, assembly, and detailing data (Meyer.C, 2019).

This advanced level of detail allows ARM to probe deeper, uncovering intricate associations such as the link between specific material specifications and distinctive systems like flooring, walling, or HVAC systems etc. Through the analysis of precise cost and quantity data, ARM can surface associations between different materials, their quantities, and the overarching project cost. Furthermore, ARM can inform scheduling decisions. Precise data detailing the chronology of different construction stages can manifest associations between project milestones and their durations. The recognition of such associations can lead to the formulation of more precise project timelines, thereby enhancing the efficiency of resource allocation and scheduling.

The associations deduced from this data can significantly enhance the decision-making process across planning, budgeting, and scheduling phases, thereby reiterating the emphasis on data integrity in BIM models for ARM.

## **2.5 Examples AI to quantity surveying**

Traditionally, the extraction of quantities from designs and drawings has been a manual process, often laborious and susceptible to human error. A salient application of AI in quantity surveying is the automation of QTO from BIM models (Tang et al., 2022). BIM facilitates the rapid and precise calculation of model material quantities, enabling project teams to utilize the output for material ordering and cost estimation.

A fundamental prerequisite for an automated QTO process is the implementation of a system that adheres to a structured organization of information. This type of organization, known as a Work

Breakdown Structure (WBS), is integral to the extensive scope of a project life cycle. It is essential for the WBS to be universally adopted by all stakeholders involved in the project to circumvent potential conflicts, errors, and omissions (Egwim et al., 2021). The WBS is usually built on a series of increasingly detailed levels across different domains, each identifiable by a specific code. The most effective way to apply this organization to a BIM model is to input the codes in ID form. To ensure data flow consistency, it's crucial that the various systems and documentations used throughout the AEC life cycle follow the same organizational structure (Zou & Ergon, 2019).

Outline Level	Code	Name	Description	Unit Cost	Quantity Type	Quantity Unit	Quantity Formula
1.1	01	01_Structural Works					
1.1.1	01.01	Beams					
1.1.1.1	01.01.01	01_Formwork					
1.1.1.1.0	01.01.01.01	ST-Precast_Rectangular_	ST-Precast_Rectangular_f	\$ 73.77	Area	m'	[Length]' * (b + 2' h)
1.1.1.1.0	01.01.01.02	ST-Concrete_Rectangular	ST-Concrete_Rectangular_f	\$ 2.85	Area	m'	[Length]' * (b + 2' h)
1.1.1.1.0	01.01.01.03	ST-Concrete_Rectangular	ST-Concrete_Rectangular_f	\$ 2.85	Area	m'	[Length]' * (b + 2' h)
1.1.1.1.0	01.01.01.04	ST-Concrete_Rectangular	ST-Concrete_Rectangular_f	\$ 2.85	Area	m'	[Length]' * (b + 2' h)
1.1.1.1.0	01.01.01.05	ST-Concrete_Rectangular	ST-Concrete_Rectangular_f	\$ 2.85	Area	m'	[Length]' * (b + 2' h)
1.1.1.2	01.01.02	02_Reinforcement					
1.1.1.2.0	01.01.02.01	ST-Precast_Rectangular_	ST-Precast_Rectangular_f	\$ 2.48	Mass	kg	[Volume]' [Reinforcement]
1.1.1.2.0	01.01.02.02	ST-Concrete_Rectangular	ST-Concrete_Rectangular_f	\$ 3.00	Mass	kg	[Volume]' [Reinforcement]
1.1.1.2.0	01.01.02.03	ST-Concrete_Rectangular	ST-Concrete_Rectangular_f	\$ 3.00	Mass	kg	[Volume]' [Reinforcement]
1.1.1.2.0	01.01.02.04	ST-Concrete_Rectangular	ST-Concrete_Rectangular_f	\$ 3.00	Mass	kg	[Volume]' [Reinforcement]
1.1.1.2.0	01.01.02.05	ST-Concrete_Rectangular	ST-Concrete_Rectangular_f	\$ 3.00	Mass	kg	[Volume]' [Reinforcement]
1.1.1.3	01.01.03	03_Concrete pouring					
1.1.1.3.0	01.01.03.01	ST-Precast_Rectangular_	ST-Precast_Rectangular_f	\$ 340.72	Volume	m'	[Volume]
1.1.1.3.0	01.01.03.02	ST-Concrete_Rectangular	ST-Concrete_Rectangular_f	\$ 258.82	Volume	m'	[Volume]
1.1.1.3.0	01.01.03.03	ST-Concrete_Rectangular	ST-Concrete_Rectangular_f	\$ 258.82	Volume	m'	[Volume]
1.1.1.3.0	01.01.03.04	ST-Concrete_Rectangular	ST-Concrete_Rectangular_f	\$ 258.82	Volume	m'	[Volume]
1.1.1.3.0	01.01.03.05	ST-Concrete_Rectangular	ST-Concrete_Rectangular_f	\$ 18.37	Volume	m'	[Volume]
1.1.1.4	01.01.04	04_Concrete finishing					
1.1.1.4.0	01.01.04.01	ST-Precast_Rectangular_	ST-Precast_Rectangular_f	\$ 13.13	Volume	m'	[Volume]
1.1.1.4.0	01.01.04.02	ST-Concrete_Rectangular	ST-Concrete_Rectangular_f	\$ 15.05	Volume	m'	[Volume]
1.1.1.4.0	01.01.04.03	ST-Concrete_Rectangular	ST-Concrete_Rectangular_f	\$ 15.05	Volume	m'	[Volume]
1.1.1.4.0	01.01.04.04	ST-Concrete_Rectangular	ST-Concrete_Rectangular_f	\$ 15.05	Volume	m'	[Volume]
1.1.1.4.0	01.01.04.05	ST-Concrete_Rectangular	ST-Concrete_Rectangular_f	\$ 15.05	Volume	m'	[Volume]
1.1.1.5	01.01.05	05_Steel beams					
1.1.1.5.0	01.01.05.01	ST-UB_Universal_Beam -	ST-UB_Universal_Beam - S	\$ 4.15	Mass	kg	[Volume]' [Specific Weigh
1.1.1.5.0	01.01.05.02	ST-UB_Universal_Beam -	ST-UB_Universal_Beam - S	\$ 5.31	Mass	kg	[Volume]' [Specific Weigh
1.1.1.5.0	01.01.05.03	ST-HSS_Round_Structura	ST-HSS_Round_Structura	\$ 6.24	Mass	kg	[Volume]' [Specific Weigh
1.1.2	01.02	Roofs					

Figure 6: QTO based on classification system.

Classification systems, which serve as the foundational framework, offer a structured approach to organizing and categorizing information. This facilitates the efficient extraction of quantities from BIM models. There are different types of these classification systems; MasterFormat to OmniClass, Uniformat, and UniClass. For the purpose of this study, Uniformat II mainly serves as the basis for elemental cost analysis. The necessity for an elemental classification like Uniformat II is particularly evident during the economic evaluation of building alternatives at the design stage. While MasterFormat 95, a classification based on products and materials, may seem

a logical choice for preparing detailed cost estimates, it often proves time-consuming, costly, and inappropriate at the early design stages. It is at these early stages that economic analysis is most crucial in determining the economically efficient choices among building alternatives. Therefore, cost estimates based on an elemental classification like Unifomat II provide the necessary cost information for analysts to evaluate building alternatives in a cost-effective manner.

UNIFORMAT General Services Administration (GSA)	CANADIAN INSTITUTE OF QUANTITY SURVEYORS (CIQS)	THE ROYAL INSTITUTION OF CHARTERED SURVEYORS (RICS-UK)	CONSTRUCTION ECONOMICS EUROPEAN COMMITTEE (CEEC)
01 FOUNDATIONS	A1 SUBSTRUCTURE	1.0 SUBSTRUCTURE	(1) SUBSTRUCTURE
011 Standard foundations	A11 Foundations	2.0 SUPERSTRUCTURE	SUPERSTRUCTURE
012 Special foundations	A12 Basement excavation	2.1 Frame	(2) Frame
02 SUBSTRUCTURE	A2 STRUCTURE	2.2 Upper floors	(3) External walls
021 Slab on grade	A21 Lowest floor construction	2.3 Roof	(4) Internal walls
022 Basement excavation	A22 Upper floor construction	2.4 Stairs	(5) Floors
023 Basement walls	A23 Roof construction	2.5 External walls	(6) Roofs
03 SUPERSTRUCTURE	A3 EXTERIOR ENCLOSURE	2.6 Windows and exterior doors	(7) Stairs
031 Floor construction	A31 Walls below grade	2.7 Interior walls & interior partitions	(8) Windows & external doors
032 Roof construction	A32 Walls above grade	2.8 Interior doors	(9) Internal doors
033 Stair construction	A33 Windows & entrances	3.0 INTERNAL FINISHES	FINISHES
04 EXTERIOR CLOSURE	A34 Roof covering	3.1 Wall finishes	(10) Internal wall finishes
041 Exterior walls	A35 Projections	3.2 Floor finishes	(11) External wall finishes
042 Exterior doors & windows	B1 PARTITIONS & DOORS	3.3 Ceiling finishes	(12) Floor finishes
05 ROOFING	B11 Partitions	4.0 FITTINGS AND FURNITURE	(13) Ceiling finishes
06 INTERIOR CONSTRUCTION	B12 Doors	4.1 Fittings and furnishings	(14) EQUIPMENT AND FURNISHINGS SERVICES
061 Partitions	B2 FINISHES	5.0 SERVICES	(15) Plumbing
062 Interior finishes	B21 Floor finishes	5.1 Sanitary appliances	(16) Heating
063 Specialties	B22 Ceiling finishes	5.2 Services equipment	(17) Ventilating & air- conditioning
07 CONVEYING SYSTEMS	B23 Wall finishes	5.3 Disposal installations	(18) Internal drainage
08 MECHANICAL	B3 FITTINGS & EQUIPMENT	5.4 Water installations	(19) Electrics
081 Plumbing	B31 Fittings & equipment	5.5 Heat source	(20) Communication
082 HVAC	B32 Equipment	5.6 Space heating & air treatment	(21) Lifts, escalators, etc.
083 Fire Protection	B33 Conveying systems	5.7 Ventilation systems	(22) Protective installations
084 Special mechanical systems	C1 MECHANICAL	5.8 Electrical installation	(23) Miscellaneous services inst.
09 ELECTRICAL	C11 Plumbing & drainage	5.9 Gas installation	EXTERNAL SITE WORKS
091 Distribution	C12 Fire protection	5.10 Life & conveyor installation	(24) Site preparation
092 Lighting & power	C13 HVAC	5.11 Protective installations	(25) Site enclosure
093 Special electrical systems	C14 Controls	5.12 Communication installations	(26) Site fittings
10 GENERAL CONDITIONS & PROFIT	C2 ELECTRICAL	5.13 Special installations	(27) Site services
11 EQUIPMENT	C21 Services & distribution	5.14 Builders work in connection with services	(28) Site Buildings
111 Fixed & moveable equipment	C22 Lighting, devices & heating	5.15 Builders profit & attendance on services	(29) Hard and soft landscaping
112 Furnishings	C23 Systems & ancillaries	6.0 EXTERNAL WORKS	(30) PRELIMINARIES
113 Special construction	D1 SITE WORK	6.1 Site works	
12 SITE WORK	D11 Site development	6.2 Drainage	
121 Site preparation	D12 Mechanical site services	6.3 External services	
122 Site improvements	D13 Electrical site services	6.4 Minor building work	
123 Site utilities	D2 ANCILLARY WORK		
124 Off-Site work	D21 Demolition		
	D22 Alterations		

Figure 7: UNIFORMAT II is the elemental classification (Charette & Marshall, 1999)

## Predictive Cost Estimation

Another application of AI in quantity surveying is predictive cost estimation through machine learning (ML). Neural networks, a subset of ML, are computational models designed to identify patterns in large datasets (Goh & Chua, 2013). (Xi'an Dianzi & Keji Daxue Xuebao, 2020) utilized RSMMeans as a synthetic data source due to the lack of comprehensive data from real construction projects. Their aim was to demonstrate the feasibility and utility of intelligent ANN models during the value engineering phase for preliminary construction estimation. After training the neural

networks, they identified relationships between project specifications and material usage. Consequently, when the model encountered new project specifications, it was able to accurately predict the required material quantities.

### **Large Language Models (LLMs)**

Research for this type of AI is still very vast and informs the focus of this paper since it offers significant potential for reshaping quantity surveying practices. LLMs are a type of foundation model tailored for text and text-like data, with training grounded on vast datasets that can extend to petabytes (Saka et al., 2023). These models are underpinned by three core components: data, the transformer architecture, and training. In quantity surveying, LLMs can analyze project specifications, extract relevant data, and cross-reference with historical project information to derive precise information required for estimation. Their proficiency in sifting through extensive textual data ensures heightened accuracy in retrieving information, thereby enhancing preliminary cost estimation.

### **3. METHODOLOGY**

#### **3.1 Scope Definition**

The scope of this research is to explore the use of Artificial Intelligence (AI) in cost estimation for building projects and how it can improve the process of cost estimation in schematic designs. The research focuses on how AI models can be applied to cost estimation models to improve their accuracy and reduce errors and examine their effectiveness in providing cost elements data for estimation of construction projects.

Although the research project aims to provide apt cost estimates, the output is limited to data fed into the framework. The fully automated process of retrieving cost information for this research relies on using Large Language Models (LLMs) such as ChatGPT, whose knowledge is based on information available up until September 2021. ChatGPT has no access to real-time data. As at the time of this research, OpenAI had developed a beta mode of gpt-4 that had internet access for real time search the web for resources, cross-check information, reference specific data, and provide up-to-the-minute information.

However, cost data also relies on other external factors such as economic conditions, regulatory costs, geographical location etc. This justifies the use of existing past cost databases and repositories like RSMeans from which LangChain will breakdown into vectors that are fed to ChatGPT and given as output based on specific user queries.

### 3.2 Methodological framework

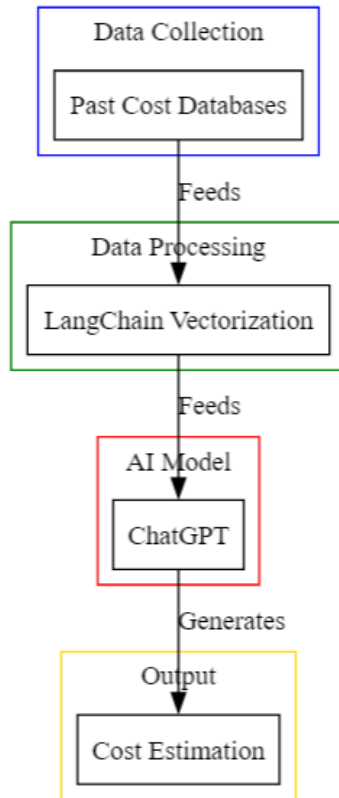


Figure 8: Methodological framework for AI-driven cost estimation.

The methodological framework for this research is designed to explore the application of LLMs in cost estimation for building projects. The framework is structured to ensure a systematic and comprehensive approach to the research, and it is divided into several key stages, each of which contributes to the overall objective of improving the accuracy and reducing errors in cost estimation models.

1. **Data Collection:** The first stage of the framework involves the collection of data from past cost databases. These databases provide historical cost data that is essential for training the AI model. The quality and relevance of the data collected at this stage



significantly impacts the accuracy of the cost estimation model. Therefore, the reference cost database used is from RSMeans (up to 2018) a comprehensive construction cost database that provides detailed data on materials, labor, and other project-related expenses to ensure the data's reliability and applicability.

2. **Data Cleaning and Preprocessing:** Once the data is collected, it undergoes cleaning and preprocessing. This stage involves removing any irrelevant, incomplete, or erroneous data and transforming the data into a format suitable for the AI model. This step is crucial to ensure the integrity of the data fed into the model, which directly influences the model's performance.
3. **Model Training and Validation:** The cleaned and preprocessed data is then used to train the AI model. During this stage, the model learns to predict cost estimates based on the input data. The model's performance is validated using a subset of the data not used during training. This validation process helps assess the model's accuracy and its ability to generalize to new, unseen data.
4. **AI Model Application:** The trained and validated AI model, in this case, chatbot, is then applied to generate responses to prompts related to providing estimation. The model uses the patterns and relationships it learned during training to predict the costs of construction projects.
5. **Post-processing:** The output from the AI model is then post-processed. This stage involves transforming the model's output into a format that is understandable and usable for the end-users. It may also involve additional checks and balances to ensure the output's accuracy and reliability.

6. **Output Generation:** The final stage of the framework is the generation of the cost estimation. This is the product of the research and represents the application of AI in improving the process of cost estimation in schematic designs and cost consultancy.

This methodological framework provides a systematic approach to the research, ensuring that each stage of the process is carefully considered and executed. It highlights the importance of data quality, model training and validation, and output processing in the application of AI in cost estimation. The framework also underscores the potential of AI in improving traditional cost estimation methods, paving the way for more accurate and reliable cost predictions in the field of building projects.

## 4. PROTOTYPE SYSTEM DEVELOPMENT

This section provides a detailed specification for the development of a system aimed at automating the process of cost estimation for BIM models. The system is intended to interact with existing BIM model to extract relevant metadata of cost elements, interact with cost databases and provide relevant response that will help with cost estimation. The system key features include integrated cost data, user prompts and data extraction from BIM models. With an intuitive user interface, users can easily load BIM models, input queries, and view results. The system monitors model changes and user input for up-to-date estimates, verifies information accuracy, checks element dependencies, and requires approval for impactful actions.

### 4.1 Requirements analysis

The system will function as a standalone application that seamlessly integrates with BIM authoring tools to extract key cost elements from BIM models, including quantities, materials, and specifications, and utilize algorithms to calculate cost estimates. The prototype system will offer a user-friendly interface for convenient interaction with the BIM model, enabling input of project-specific parameters and providing visual representations of cost estimation results.

#### 4.1.1 Functional Requirements:

##### 1. BIM Data Extraction:

- The system should have the capability to extract relevant cost element data from BIM models, encompassing quantities, materials, and specifications.
- It should support multiple BIM file formats including both **proprietary and open formats**, ensuring compatibility.

##### 2. Cost Estimation Calculation:

The system should have the capacity to use AI algorithms like ARM or market basket analysis to determine material relationships and the cost elements for accurate cost estimation based on metadata of the BIM models. For instance, when examining concrete, these algorithms would come up with correlations between concrete mix ratios, aggregate selections, and reinforcement materials. By analyzing these patterns, the system could effectively outline cost factors associated with such raw materials, labor, and maintenance, providing a comprehensive elements to be considered for estimation.

### 3. User Interface:

- The system's user interface should be intuitive, user-friendly, and visually appealing.
- Users should be able to interact with the BIM model, input project-specific parameters, and view responses efficiently.
- Clear visualizations and intuitive controls should enhance navigation and data input processes.

### 5. Reporting and Export:

- The system should generate comprehensive cost estimation reports that summarize calculated costs, including breakdowns based on labor, materials, and other cost components.
- Reports should be exportable in widely used formats, such as PDF and Excel.

#### **4.1.2 Non-functional Requirements:**

### 6. Performance:

- *The system should perform cost estimation calculations efficiently, delivering timely results even for complex BIM models.*

*- Low latency: It should be capable of handling large BIM models without significant performance degradation.*

#### 7. Scalability:

*- The system should be scalable to accommodate future enhancements and updates in BIM modeling software and cost estimation methodologies.*

#### 8. Security:

*- Appropriate security measures should be implemented to ensure the confidentiality and integrity of BIM model data and cost estimation results.*

*- User authentication and access control mechanisms should be employed to authorize system access.*

### **4.1.3. Assumptions and Dependencies**

In the development and implementation of the automated cost estimation system, some assumptions and dependencies are considered in its successful operation and accuracy of results.

#### **Assumptions:**

1. **Automatic Quantity Generation and Mapping to Cost Items:** QTOs are automatically generated from the BIM models and these quantities are directly mapped to cost items in the cost estimation process. This assumption is critical as it forms the basis for the system's ability to automate the cost estimation process, thereby reducing manual effort and potential for errors.
2. **Consistent Pricing Data:** The system also assumes that the pricing data used for cost estimation is up-to-date and consistent. This means the system presumes that the cost data it uses for estimation accurately reflects the current market prices for materials, labor, and

other cost items. This assumption is essential to ensure that the cost estimates generated by the system are realistic and reliable.

### Dependencies:

The BIM Model Cost Estimation Automation System is dependent on several external factors for its operation.

1. **BIM authoring tool and model format:** The system is heavily reliant on the BIM authoring tool and accurate exportation of model formats to provide accurate and detailed BIM models.

## 4.2 Use case diagrams

The system development is articulated using a Unified Modeling Language (UML) Use Case diagram, which provides a graphical representation of the system's functionality as perceived by the system's end-users.

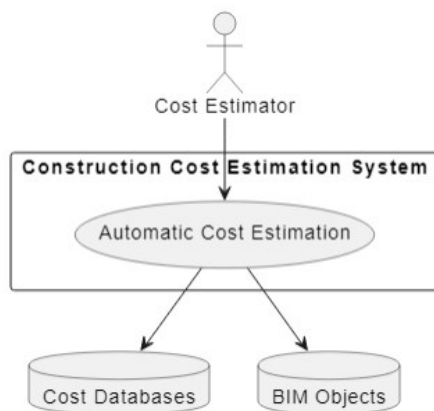


Figure 9: General use case by cost estimator

Following the detailed functional and non-functional requirements of the system, a detailed use case diagram below provides a comprehensive overview of the system's functionality and its interaction with external entities.

The diagram identifies five key actors: Users, BIM Modeling Software, Database/Storage, Report Recipients, and Large Language Models.

Users are the primary operators who provide project-specific parameters and revise BIM models. The BIM Modeling Software provides the BIM models, which are integral to the system's operations. The Database/Storage serves as the repository for the system's data and cost estimation results. Report Recipients are the end receivers of the cost estimation reports generated by the system. Large Language Models represents advanced AI models that the system interacts with for complex data processing and analysis.

The system's operations are represented as eight use cases. These include the provision of project-specific parameters, extraction of data from BIM models, storage of data and cost estimation results, return of cost estimation results, sending of cost estimation reports, revision of BIM models, updating of cost estimation results, and interaction with Large Language Models.

The interactions between the actors and the use cases are depicted by arrows, demonstrating the flow of information or actions between them. Notably, the system's interaction with Large Language Models expands its capabilities, enabling it to leverage advanced AI techniques for more accurate and efficient cost estimation.

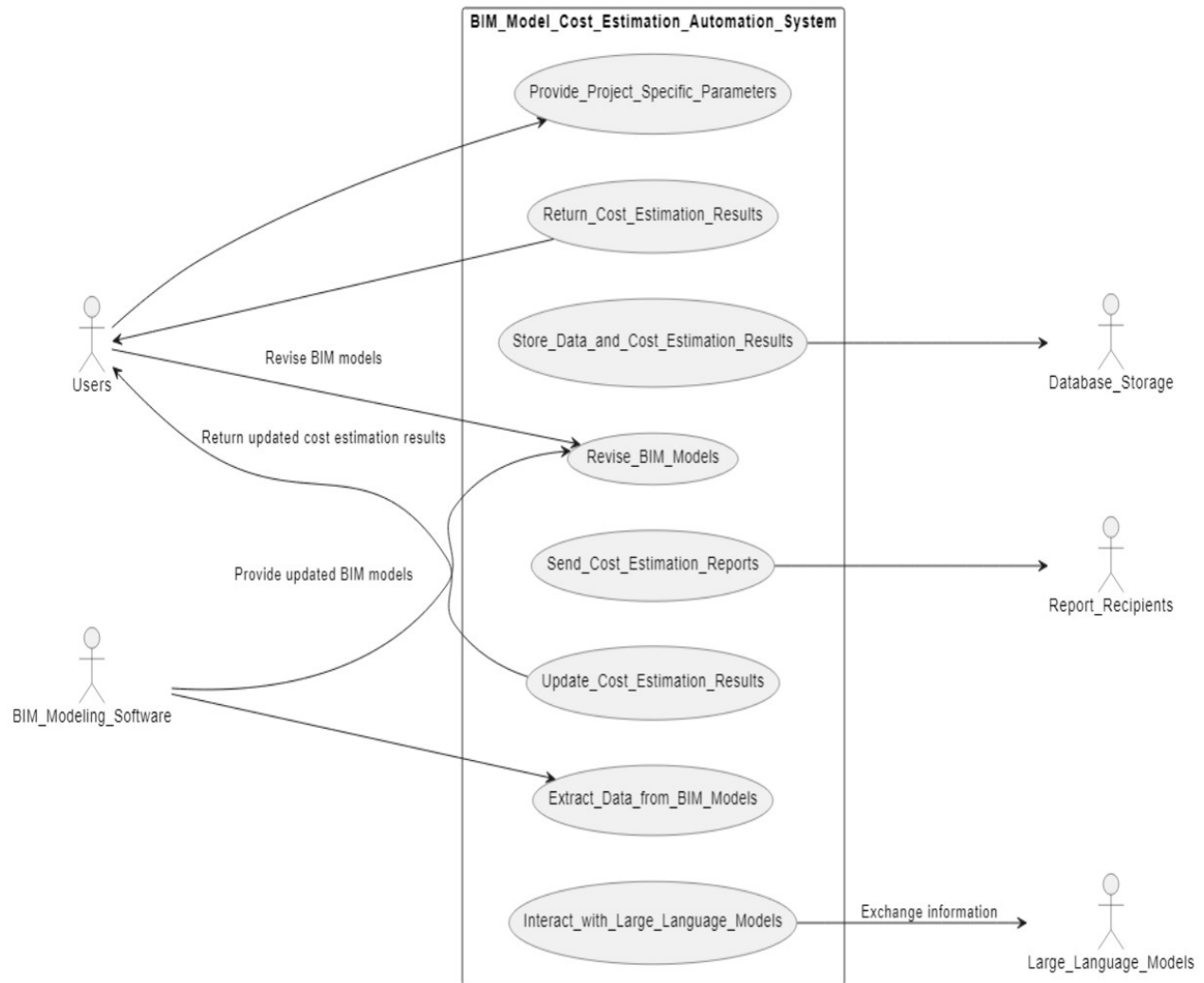


Figure 10: Use case diagram for automated estimation process



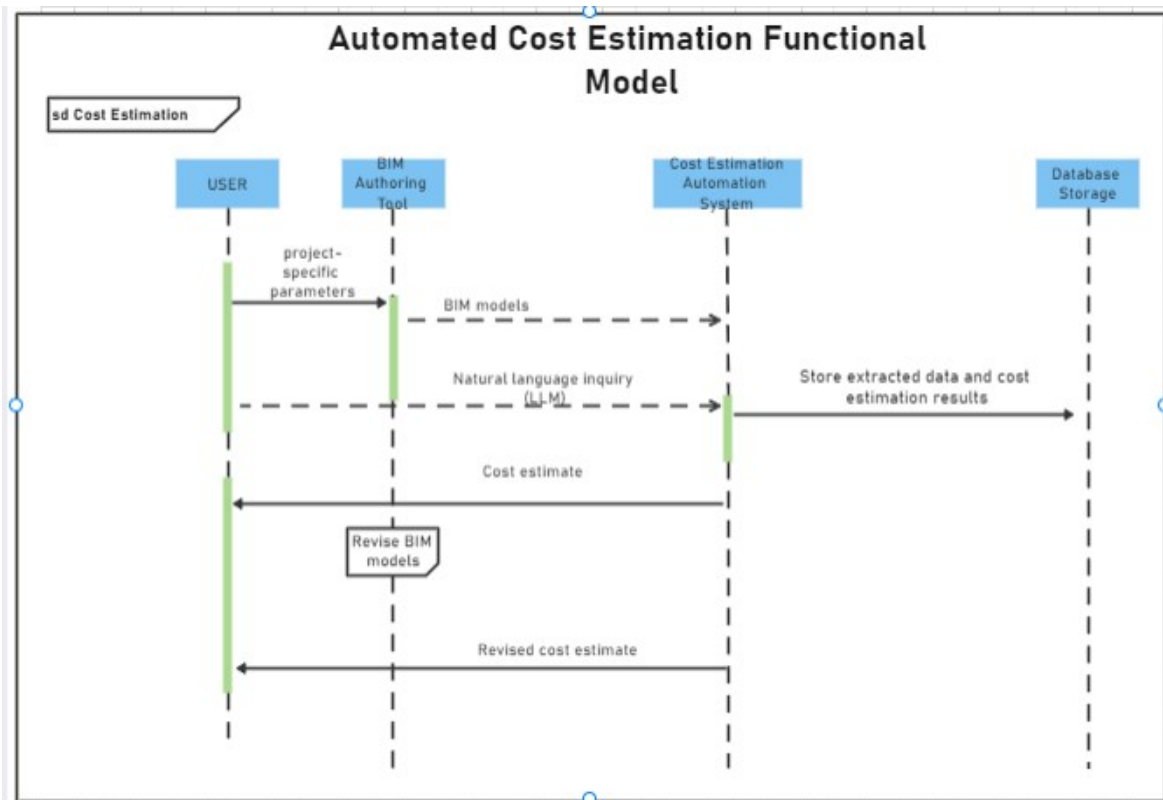


Figure 11: Interaction sequence diagram for cost estimation.

#### 4.2.1 Relationship Establishment

The relationships between the actors and the use cases in the UML class diagram are established as follows:

1. **Cost Estimation Automation System and Users:** The system interacts with users, who provide project-specific parameters and revise BIM models. This interaction is represented by the methods *Provide\_Project\_Specific\_Parameters()* and *Revise\_BIM\_Models()* in the Users class.

2. **Cost Estimation Automation System and BIM Modeling Software:** The system extracts data from BIM models provided by the BIM Modeling Software. This is represented by the method *Provide\_BIM\_Models()* in the BIM Modeling Software class.
3. **Cost Estimation Automation System and Database/Storage:** The system stores data and cost estimation results in the Database/Storage. This is represented by the method *Store\_Data()* in the Database/Storage class.
4. **BIM Model Cost Estimation Automation System and Report Recipients:** The system sends cost estimation reports to the Report Recipients. This is represented by the method *Receive\_Reports()* in the Report Recipients class.
5. **Cost Estimation Automation System and Large Language Models:** The system interacts with Large Language Models for advanced data extraction, analysis and cost estimation. This is represented by the method *Process\_Data()* in the Large Language Models class.

These relationships establish the flow of information and actions between the system and the external entities, providing a clear understanding of how the system operates and interacts with its environment.

### 4.3 Functional model

The functional model of the automation system is an amalgamation of various machine learning (ML) applications. These applications are integrated into the system, serving to automate and optimize the multifaceted stages of the cost estimation process. The objective is to enhance the precision of cost estimation, reduce the margin of error, and improve the overall efficiency of the process.

1. **Market Basket Analysis (MBA):** This ML application is an advanced analytical tool that is employed to discern patterns in the selection of construction materials and services. By scrutinizing the data from past projects, the system can predict the probability of certain items being used concurrently. This predictive capability aids in optimizing procurement strategies and minimizing costs, thereby contributing to the overall financial efficiency of the project.
2. **Machine Vision Application:** This application leverages the power of machine vision to analyze BIM models and enumerate the quantity of specific elements, such as doors, plumbing items such as WCs, or furniture pieces. This automated counting mechanism significantly curtails the time required for quantity take-off and enhances the accuracy of the cost estimation, thereby contributing to the system's efficiency.
3. **Predictive Modeling:** This ML application utilizes historical data and project-specific parameters to predict the cost of a project. The predictive model takes into account a plethora of factors such as project size, geographical location, complexity, and materials to generate an accurate cost estimate. This predictive capability allows for more precise budgeting and financial planning.

4. **Data Mining:** This involves the rigorous analysis of large datasets to identify patterns and trends that can inform the cost estimation process. For instance, data mining can reveal trends in material costs, labor rates, or project timelines, which can be used to refine the cost estimates.
5. **Large Language Models:** LLMs are also trained on large datasets, enabling them to understand and generate human-like text. These AI uses embeddings to discover patterns of relationships between words, phrases, or other types of data. For example in some past databases structural element will be referred to as 'beam' in some instances and in other cases as 'girder' and the associated costs remain the same. LLMs can identify this relationship and provide potential correlated expenses.

These ML applications are integrated into the system in a modular manner, allowing for their individual or combined use, depending on the specific requirements of each project.

#### 4.4 System architecture

The 'AI-IS' cost estimation process of BIM models relies on a systematic development of the design, from a basic conceptual model to a detailed and comprehensive model. Each stage in the process demands precise information and specifications, with the quality of the cost estimate being dependent on the precision and accuracy of the information input into the BIM model (Eastman, Teicholz, Sacks, & Liston, 2011).

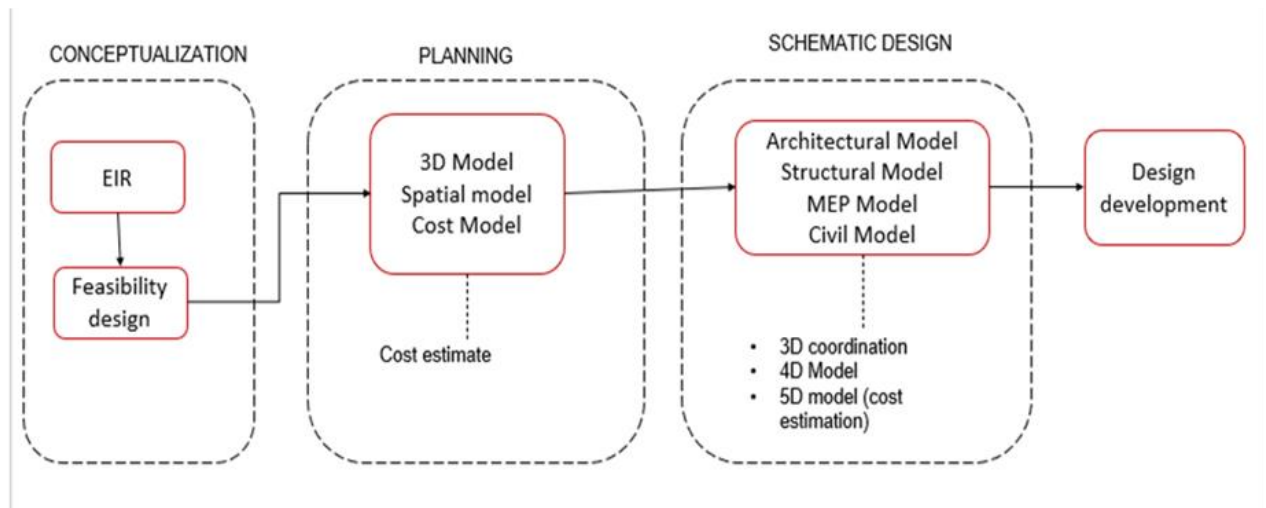


Figure 12: AS-IS Cost estimation process of BIM models.

Figure 12 shows a conceptual 5D workflow that yields cost estimates. The process starts from a federated BIM model that doesn't have any cost information or any reference to the cost information. However, the model has been carefully exported with base quantities which are automatically calculated with respect to rules of measurement. The quantity properties are assigned cost information linked to the model elements. The cost information is basically an outside source that is implemented and through the series of defined rules linked to the model elements manually. Cost database highly depends on various factors such as geographical location, market condition, regulatory requirements, etc.

Mapping cost information to all the model elements results in a fully completed 5D BIM model with integrated cost data. From this cost model, a detailed bill of quantities (BOQ) is derived which is exported in desirable formats.

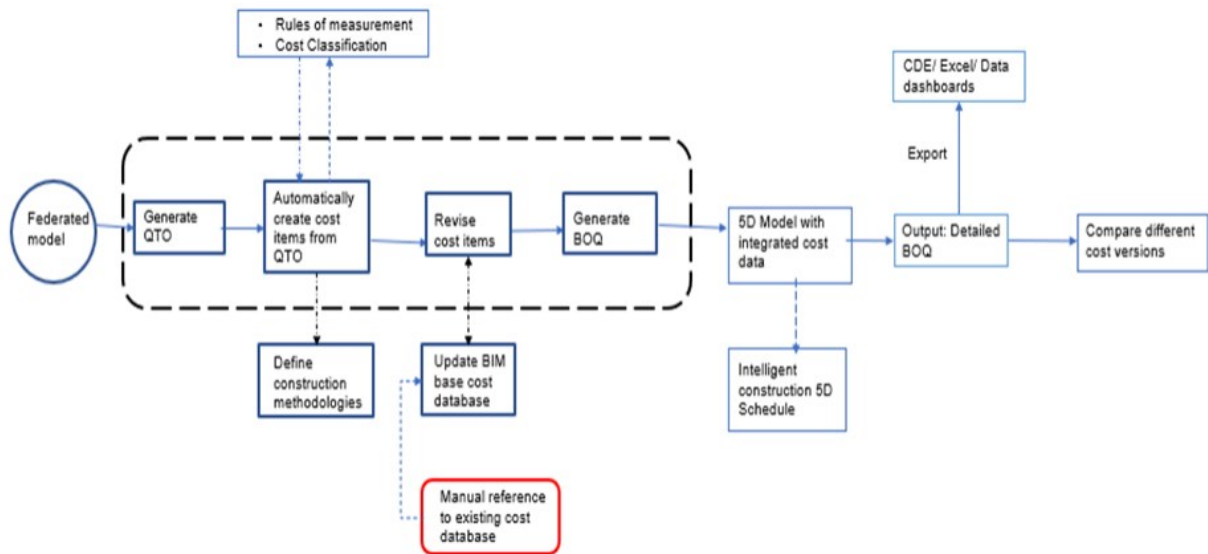


Figure 13:5D framework for BIM models

Cost classification plays a crucial role in achieving structured cost estimates. To establish a systematic approach, data sets are classified based on the ontology, entities, classes, objects, taxonomy, and attributes associated with the substructure and superstructure components.

In cost classification, ontology involves the study of the components and their relationships within the substructure and superstructure (Abanda et al., 2017). This understanding helps in identifying the relevant entities such as foundation systems, basement walls, floor slabs, columns, beams, walls, and roof elements.

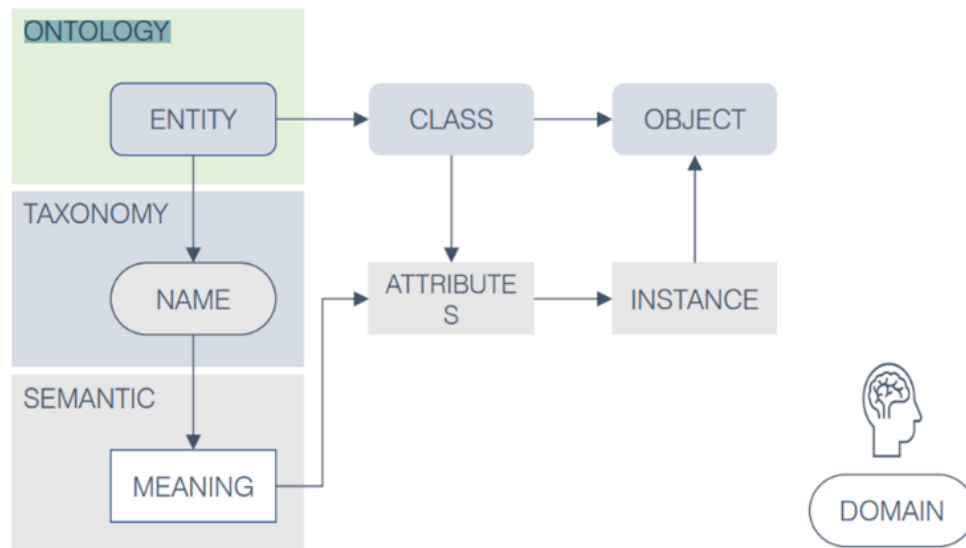


Figure 14: Classification framework for cost estimation (BIM A+ 5)

Classes are employed to categorize these entities, grouping them according to their nature and function. For instance, entities are classified into classes such as "foundation systems," "basement walls," "floor slabs," "columns," "beams," "walls," and "roof elements." This classification enables a structured approach to cost estimation by organizing the entities into distinct categories.

Objects are specific instances within a class. These objects represent individual occurrences of the entities. Examples of objects could include a particular foundation system or a specific column, each with its own unique cost implications.

Taxonomy is employed to create a hierarchical classification system that organizes the entities and classes. By establishing levels such as "Substructure" and "Superstructure," and subcategories within them, the taxonomy provides a structured framework for cost classification. This hierarchy aids in organizing and categorizing the components based on their relationships and characteristics.


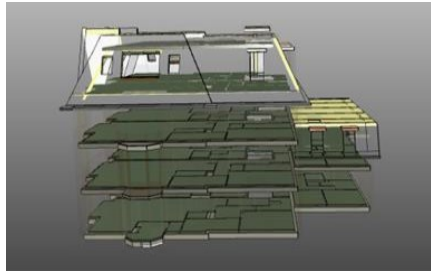
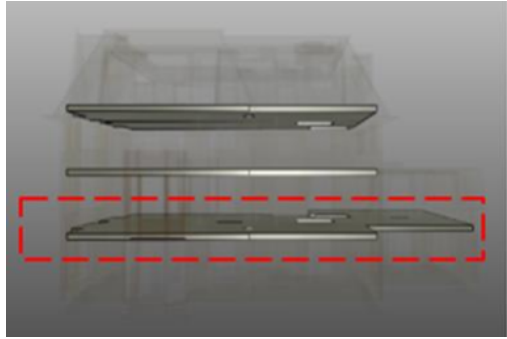
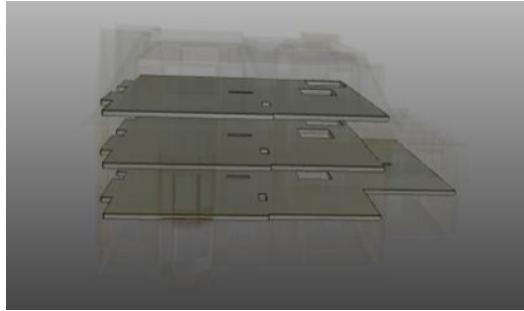
Attributes represent the characteristics or properties of the entities. These attributes, such as dimensions, material, and load-bearing capacity, influence the cost estimation process. By

considering these attributes, a more accurate and detailed cost estimate can be achieved, taking into account the specific characteristics of each component.

Undeniably, the cost estimation process employed for BIM models demonstrates a commendable level of comprehensiveness and efficacy. It embodies a systematic approach that aligns with the complexity of construction projects, providing information into potential cost implications from the conceptualization stage to the final design. However, this well-established process, when scrutinized from the vantage point of the rapidly evolving field of Artificial Intelligence (AI), exhibits certain limitations.



Table 3: Classification breakdown of a model

<p>Building level</p>		<p>Drawings Layers Model Views</p> <p>Click to Filter &lt;Filter is Empty&gt;</p> <p>Name</p> <ul style="list-style-type: none"> <li>Drawing             <ul style="list-style-type: none"> <li>IfcBeam</li> <li>IfcBuildingElementPart</li> <li>IfcBuildingElementProxy</li> <li>IfcColumns</li> <li>IfcCovering</li> <li>IfcDistributionElement</li> <li>IfcDoor</li> <li>IfcFlowSegment</li> <li>IfcMember</li> <li>IfcRailing</li> <li>IfcSlab</li> <li>IfcSpace</li> <li>IfcStair</li> <li>IfcWall</li> <li>IfcWallStandardCase</li> <li>IfcWindow</li> </ul> </li> </ul>
<p>Component level</p>		<p>IfcSlab – Different types of slab materials</p> <ul style="list-style-type: none"> <li>IfcSlab             <ul style="list-style-type: none"> <li>01 plaatmateriaal - multiplex 18</li> <li>02 beton gewapend - prefab</li> <li>04 vloertegels 200x200 10</li> <li>04 Zink</li> <li>99 Isolatie - Glaswol 190mm 190</li> <li>99 Isolatie - Glaswol 190mm 206</li> <li>IFC schoonloopmat 8</li> <li>IFC_betonvloer_ghw_345</li> <li>IFC_betonvloer_ghw_300mm 300</li> <li>IFC_breedplaat_betonvloer_ghw_200mm 200</li> <li>IFC_breedplaat_betonvloer_ghw_220mm 220</li> <li>IFC_breedplaat_betonvloer_ghw_290mm 290</li> <li>IFC_breedplaat_schil_60mm 60</li> <li>IFC_schilplaat_schillement 177</li> <li>IFC_schilplaat_schillement_191 191</li> <li>IFC_schilplaat_191_omverrek_schillement 100</li> <li>IFC_schilplaat_191_omverrek_schillement 131</li> <li>IFC_schilplaat_191_omverrek_schillement 155</li> <li>IFC_schilplaat_schillement_8_2=4,00 206</li> <li>IFC_schilplaat 90</li> <li>IFC_schilplaat_omverrek 90</li> </ul> </li> </ul>
<p>Element level</p>		<p>Concrete slab on different levels</p> <ul style="list-style-type: none"> <li>IFC_betonvloer_pretfab_200mm 200</li> <li>IFC_breedplaat_betonvloer_ghw_220mm 220</li> <li>IFC_breedplaat_betonvloer_ghw_290mm 290</li> </ul> <p>vloerveld          vloerveld          vloerveld</p>
<p>Material level</p>		<p>290mm thick concrete slab</p> <ul style="list-style-type: none"> <li>IfcSlab             <ul style="list-style-type: none"> <li>01 plaatmateriaal - multiplex 18</li> <li>02 Beton gewapend - prefab</li> <li>04 vloertegels 200x200 10</li> <li>04 Zink</li> <li>99 Isolatie - Glaswol 190mm 190</li> <li>99 Isolatie - Glaswol 190mm 206</li> <li>IFC schoonloopmat 8</li> <li>IFC_betonvloer_ghw_345</li> <li>IFC_betonvloer_ghw_300mm 300</li> <li>IFC_betonvloer_pretfab_200mm 200</li> <li>IFC_breedplaat_betonvloer_ghw_200mm 200</li> <li>IFC_breedplaat_betonvloer_ghw_220mm 220</li> <li>IFC_breedplaat_betonvloer_ghw_290mm 290</li> <li>IFC_breedplaat_schil_60mm 60</li> </ul> </li> </ul>

While BIM has significantly transformed the construction industry, the infusion of AI promises an even more profound potential impact. AI, with its vast capabilities for learning, prediction, and automation, presents novel methods for addressing challenges inherent in the traditional BIM-based cost estimation process. In this context, it becomes imperative to reflect upon these limitations and identify areas within the BIM cost estimation process where AI can not only supplement but also elevate the existing procedures. This necessitates a nuanced understanding of the interplay between BIM and AI, recognizing the strengths of the former while being cognizant of the transformative potential of the latter.

Despite the robustness of the BIM process, it's important to acknowledge that it operates within certain parameters that might constrain its overall effectiveness in a progressively digitized and data-driven construction industry. Therefore, a comprehensive critique of the process, under the lens of AI advancements, allows for a more holistic understanding of the future of cost estimation in construction. This not only reveals the areas where BIM's manual and semi-automated processes might be streamlined but also unearths potential avenues for predictive and dynamic cost analysis, fostering a more proactive approach to cost management in construction projects.

**Manual Input Requirement:** The current process is dependent on the precision and completeness of the input data (Eastman et al., 2011). Although BIM software does aid in automating cost estimation, it still requires considerable manual input. This is an area where LLMs can be trained on database information and significantly enhance efficiency by automatically learning from past projects' data to provide more precise initial cost estimates.

**Limited Predictive Capability:** While BIM offers real-time cost estimation, its ability to predict future costs based on changing market conditions is limited. MBA which employs predictive

analytics can analyze historical cost data to forecast future cost implications, thus providing a more dynamic and proactive approach to cost estimation.

Code	Description	Quantity	Unit	Rate	Sub-Total	Factor	Total	
	IfcWallStandardCase				185,062		185,062	
	Basic Wall:External wall 1Basic Wall:External wall 1:	33	m2	454.00	14,982		14,982	
A:Code	B:Description	C:Quantity	D:Unit	E:Rate	F:Factor	G:Total	H:Include	I:
	Material Costs	1	m2	100.00		100.00	100.00	
	Labor Costs	1	hr	20.00		20.00	20.00	
	Equipment Costs	1	hr	200.00		200.00	200.00	
	Waste/Overage Costs		\$			10.00	10.00	
	Shipping/Delivery Costs	1	hr	50.00		50.00	50.00	
	Contingency Costs		\$			74.00	74.00	

Figure 15: Cost elements of an external wall

	IfcWallStandardCase				185,062		185,062	
A:Code	B:Description	C:Quantity	D:Unit	E:Rate	F:Subtotal	G:Factor	H:Total	I:Use
	Basic Wall:External wall 1Basic Wall:External wall 1:351460	33	m2	454.00	14,982		14,982	
	Basic Wall:External wall 1Basic Wall:External wall 1:351461	60	m2	454.00	27,240		27,240	
	Basic Wall:External wall 1Basic Wall:External wall 1:351462	38	m2	454.00	17,252		17,252	
	Basic Wall:External wall 1Basic Wall:External wall 1:351463	65	m2	454.00	29,510		29,510	
	Basic Wall:Interior wall 1Basic Wall:Interior wall 1:351637	19	m2	1,434.00	27,246		27,246	
	Basic Wall:Interior wall 1Basic Wall:Interior wall 1:351738	15	m2	1,434.00	21,510		21,510	
	Basic Wall:Interior wall 1Basic Wall:Interior wall 1:352692	12	m2	1,434.00	17,208		17,208	
	Basic Wall:Interior wall 1Basic Wall:Interior wall 1:352771	8	m2	1,434.00	11,472		11,472	
	Basic Wall:Interior wall 1Basic Wall:Interior wall 1:352988	7	m2	1,434.00	10,038		10,038	
	Basic Wall:Interior wall 1Basic Wall:Interior wall 1:353116	6	m2	1,434.00	8,604		8,604	

Figure 16: Cost breakdown of individual walls of a model

**Inadequate Learning from Past Projects:** The current BIM process may not adequately incorporate insights from completed projects into the cost estimation of new ones. In contrast, AI, can 'learn' from past project data to refine and improve cost estimates for future projects (Egwin et al., 2021)

#### **4.4.1 Proposed Framework for Ai-Driven Cost Estimation**

The reliability of cost data in a 5D BIM model is largely dependent on the accuracy and currency of the sources of this data. BIM models metadata, vendor quotes, historical data, and industry standards serve as the primary sources of this data (Tang et al., 2019). Vendor quotes, the most direct reflection of the current market, must be solicited with meticulous specifications and scrutinized to ensure that they reflect the true cost of the resources under consideration.

Historical data also serve as a reliable basis for cost estimation, but it needs to be indexed to the current market conditions using appropriate escalation indices to counter the effects of inflation and other economic changes (Zhang, 2015). Industry databases, such as RSMeans data, provide benchmark values for many construction elements and should be used judiciously, with adjustments for local conditions and specific project circumstances.

Regularly reviewing and updating cost data is necessary to maintain its relevance in an ever-evolving market. This process should be institutionalized and rigorously followed to ensure that the cost data used in the BIM model is the most accurate reflection of the current market (Olatunji, Sher & Gu, 2010).

#### **4.4.2 Integrating AI with BIM Models**

Schemas like IFC outline the mechanism through which distinct BIM software applications interface with each other. Nonetheless, as emphasized in the International Framework for Dictionaries (IFD) white paper (BuildingSMART, 2008), the presence of a controlled vocabulary encapsulating construction terminology is paramount for facilitating data exchange. By incorporating multi-language terminology, the IFD facilitates the mapping of concepts to IFC

entities and attributes (Saka et al., 2023), thereby distinguishing specific linguistic instances from concepts. Therefore, by employing an IFD library and the semantic web, terminology or its linguistic equivalent in a specific language can be readily mapped to an entity within a data schema akin to the IFC (Lin et al., 2013).

The initial stage involves defining a structured framework for specification documents, focusing on the ontology and semantics. The documents present key cost elements, manufacturing details, typically defined by code, description, unit of measurement, and price information. Each item undergoes a price analysis, which considers factors such as material costs, labor expenses, transportation, and equipment rentals to determine the final price.

The framework defines four main cost entities: construction work item, product item, labor item, and equipment item. Once these entities are identified, the next step is to establish a common set of labels to categorize unstructured data found in technical specifications pertaining to these four different cost entities.

To define these labels, the process begins by analyzing item descriptions within the specification documents. Cost items are categorized according to their class to obtain a diverse sample of descriptions. Parameters affecting the price are extracted from these descriptions. The set of attributes is then discussed with industry practitioners through interviews to validate the labels.

The labels are further validated through an iterative process involving additional cost item descriptions. These labels can be grouped into four types of attributes:

1. Attributes related to item classification (ID code, product class, typology).

2. Attributes related to item characteristics (material, finishing, shape, size, function, physical performance, sub-components, standards, certificates, supply mode, yield).
3. Attributes related to miscellaneous information (including, excluded, notes, application).
4. Attributes related to unit price (unit of measurement, price).

The outcome of this phase is the creation of a common environment where all data related to generating a new cost item are stored, typically in a relational database.

Once the labels are defined, Natural Language Processing (NLP) techniques are evaluated for structuring information. Classification techniques are particularly explored as they are deemed most suitable for achieving the research objectives. These classification techniques can be categorized as unsupervised (machine learning without human supervision) or supervised (machine learning under human supervision).

### **Extraction And Association of Keywords**

Keywords signify user intent and ideally, should correlate with entities or attributes within a data schema such as IFC. The process of extracting and mapping keywords lays the groundwork for comprehending user intentions and retrieving BIM data. Figure 16 suggests a process to effectively manage keyword extraction and mapping (Mengtian Yin, 2023).

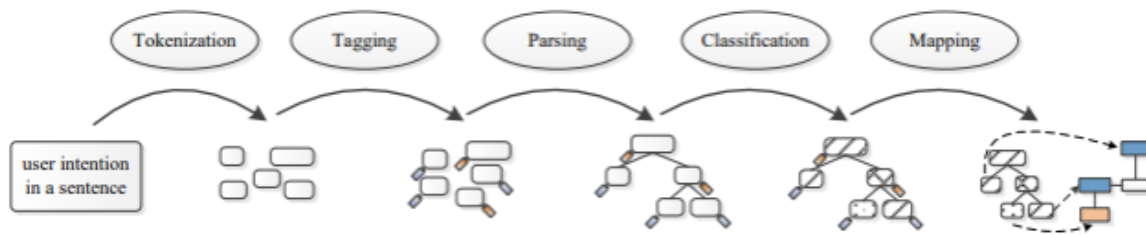


Figure 17: Process of extraction of keywords (Mengtian Yin, 2023).

- (1) Tokenization: This step involves segmenting user input into individual words. Various Natural Language Processing (NLP) algorithms are deployed for text processing in different languages.
- (2) Tagging: In this stage, each word within a sentence is labeled according to its grammatical category, i.e., noun, verb, adjective, and so on.
- (3) Parsing: Leveraging the tagging based on parts of speech, the syntactic structure of the sentence is analyzed, thereby determining the relationships amongst the different segments of the sentence.
- (4) Classification: Each word, based on its part of speech, is categorized into different collections. This aids in deciding whether a word should correlate with an entity or its attribute, with such words being designated as "keywords."
- (5) Mapping: Each keyword is associated with an entity or its attribute, following the relationship structures between the concepts in IFD and entities within a data schema.

The IFD Library functions as an accessible resource that defines and explains concepts and terms, assigning them unique identification numbers. To ensure a clear comprehension of information exchanged using IFC, it's vital to rely on concepts defined within IFD. By utilizing the correct GUID, entities can be accurately understood, along with their synonyms, plural forms, and multilingual names. A prior IFD prototype (Pan & Zhang, 2023) has expanded in two

ways to facilitate keyword mapping, based on an analysis of quantity takeoff standards. The process begins with synonym expansion, linking terms with the same meaning, such as "pillar" and "column," to a standard term, "column" in this case. Subsequently, form standardization aligns various term forms, including singular/plural, different spellings, and abbreviations, with a standard term. Using IFD's mapping mechanics, extracted data keywords can be matched to IFC entities or attributes. This entails substituting keywords with their standard forms to account for variations, followed by a synonym check (Pauwels P, Van Deursen D.,2011).

Finally, with the existing relationships between concepts and IFC entities/attributes, the keywords can be successfully mapped to the corresponding IFC entities and attributes, facilitating data retrieval. This process, underpinned by the principles of semantic mapping, on Uniform Resource Identifiers (URIs) to obtain globally unique identifiers, thereby enabling the standard exchange of information. For instance, the terms "quantity," "beams," and "storey" are systematically mapped to "IfcProperty," "IfcBeam," and "IfcBuildingStorey," respectively.

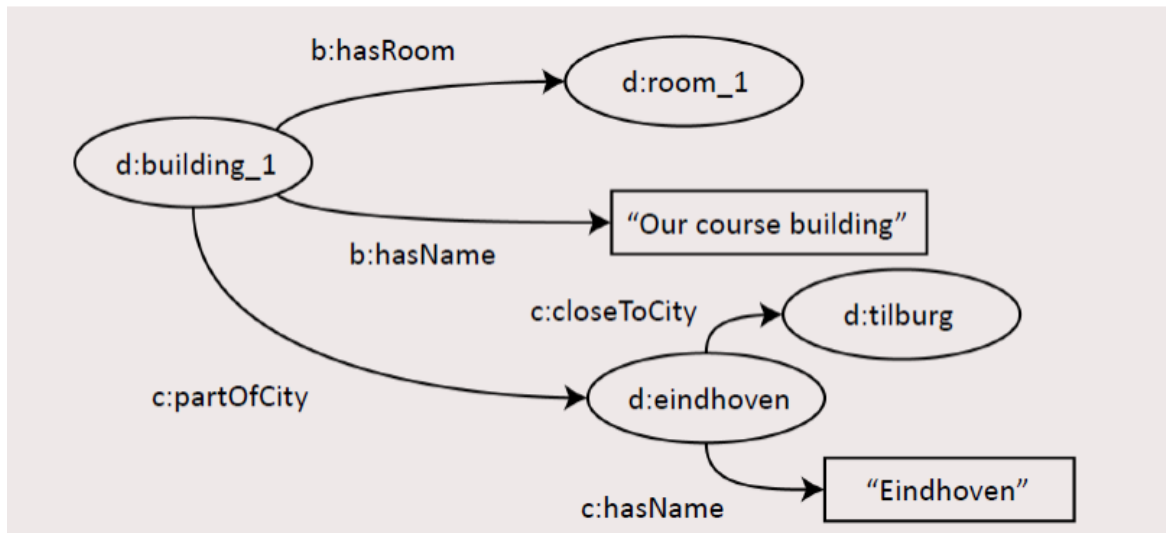


Figure 18: Semantic mapping of building (Pauwels P, Van Deursen D., 2011)



### 4.4.3 IfcCostItem

"IfcCostItem" represents an entity categorized under `IfcControl` and functions as a representation of cost-related elements, including asset costs, service costs, work execution costs, lifecycle costs, cost estimates, and budgets (BuildingSMART, 2022a). Within the IFC schema, each element, including `IfcCostItem`, possesses a set of attributes, with "Name" and "Description" being the primary ones. "Name" is used to group instances within a hierarchical framework, while "Description" provides textual information about the item within a cost schedule (BuildingSMART, 2022b).

`IfcCostItem` includes class-specific attributes such as "PredefinedType," "CostQuantities," and "CostValue." It can establish links to one or more "IfcCostValue" instances, representing cost elements like unit costs, total costs, or unit costs associated with quantities. These quantities can be individual or sourced from building elements. Additionally, an "IfcCostValue" can represent an original value, or a value derived from formulas. `IfcCostItem` also enables the creation of hierarchical cost assemblies through the "IfcRelNests" relationship, facilitating a structured arrangement of cost items (BuildingSMART, 2022b).

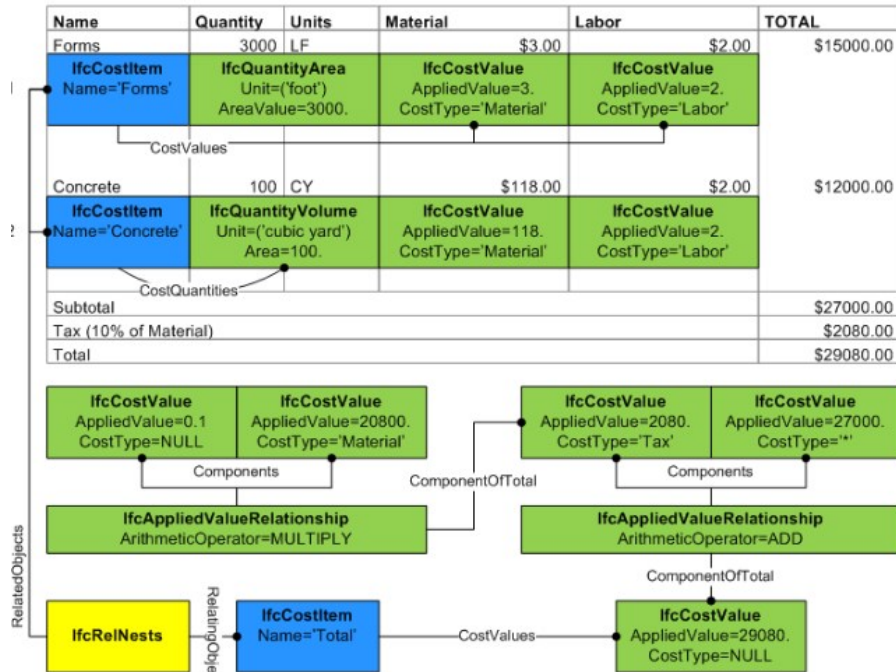


Figure 19: IFC Cost Composition (Building Smart Industry Foundation Classes 4.0.2.1)

To provide clear descriptions, the cost item is broken down into attributes. This breakdown is essential to identify the fundamental characteristics of the cost items, because the IFC schema does not make this clear. The cost items stored in the cost database will then be utilized for the purpose of reference to generate cost estimates.

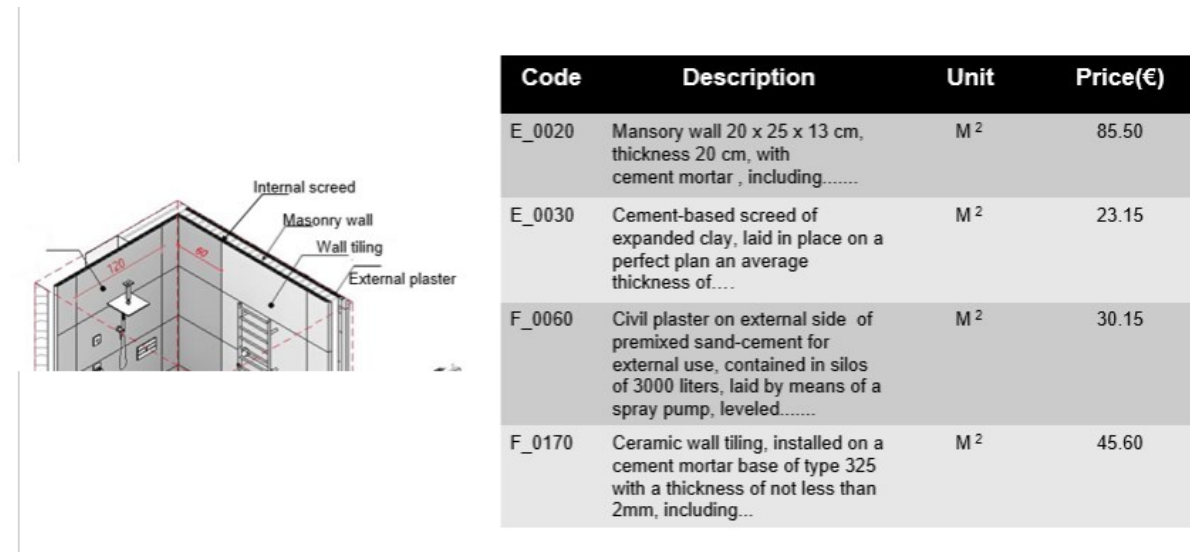


Figure 20: Cost elements in a model and its prices

#### **4.4.4 Integrating AI with Cost Data**

As previously stated, historical cost data plays a critical role in preparing cost estimates. To interact iteratively with large cost databases, LangChain, an AI framework is employed. The concept behind LangChain as an open-source framework is that provides a versatile platform for combining large language models (LLMs), such as GPT3.5 or GPT-4, with external sources of computation and data. The LangChain framework is designed around two fundamental principles: data-awareness and agentic capabilities. Its data-aware nature allows seamless connectivity between language models and diverse sources of data, enabling comprehensive analysis of cost-related information. At the same time, the agentic capabilities of LangChain empower language models to interact intelligently with their environment, facilitating efficient extraction and utilization of cost data.

LangChain incorporates a set of core modules that are regarded as the building blocks of any application powered by large language models (LLMs), representing the core abstractions. These modules are designed with standard, extendable interfaces, and the framework also offers external integrations and pre-built implementations for convenient use. The modules include:

Table 4: Modules of Langchain

<b>Models</b>	This module supports different model types and integrations, enabling the integration of various language models into the LangChain framework.
<b>Prompts</b>	The prompts module focuses on the management, optimization, and serialization of prompts, facilitating the effective utilization of language models within the framework.
<b>Memory</b>	The memory module allows the persistence of state between consecutive calls made to a chain or agent, ensuring continuity and context-awareness during interactions.
<b>Indexes</b>	Language models become more powerful when combined with application-specific data. The indexes module provides interfaces and integrations for loading, querying, and updating external data, thus enhancing the capabilities of language models.
<b>Chains</b>	Chains in LangChain are structured sequences of calls, either to a language model or a different utility, enabling the development of complex workflows and facilitating interaction between different components.
<b>Agents</b>	An agent represents a chain in which a language model, given a high-level directive and a set of tools, repeatedly makes decisions, executes actions, and observes the outcomes until the directive is successfully completed.
<b>Callbacks</b>	Callbacks in LangChain enable the logging and streaming of intermediate steps within a chain. This functionality eases the observation, debugging, and evaluation of the internal processes of an application.

#### 4.4.5. General database architecture for AI application

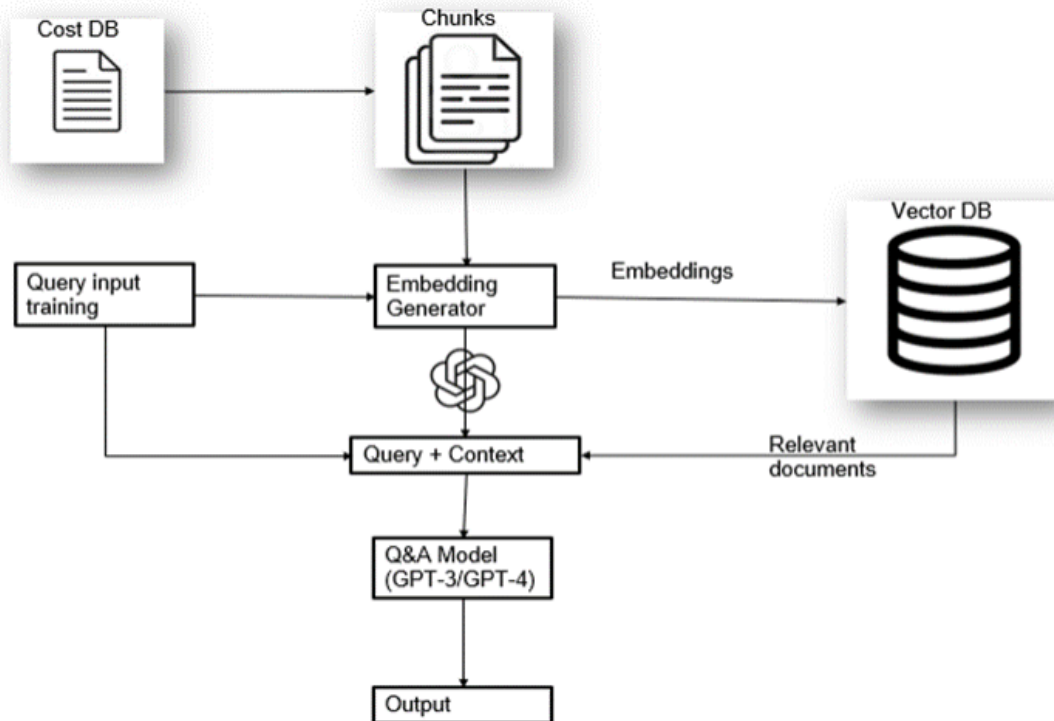


Figure 21: Architecture of the AI application

##### 4.4.5.1 Data Ingestion

###### (a) Load data sources to text

The first step in any data processing task is to load the data. In this case, it means reading text data from various sources. These sources could be anything from text files, web pages, databases, APIs, or even images and audio files that have been transcribed into text. The data needs to be converted (or "loaded") into a format that is usable for the subsequent steps, which often means converting it into plain text. It may also involve cleaning or pre-processing the data to remove any extraneous

information or to standardize the format. Historical cost data is fed into the system for further reference.

### **(b) Chunk text**

After the text has been loaded, the next step is to break it down into smaller pieces, or "chunks". This is necessary because most language models, including those based on machine learning, have a limit to the amount of text they can process at once due to memory constraints. For example, models based on the transformer architecture, such as BERT or GPT, have a maximum sequence length. The chunking process can be as simple as splitting the text every 'n' words or characters, or it could involve more complex processes like parsing the text into sentences or paragraphs.

### **(c) Embed text**

Once the text has been chunked, each chunk is then converted into a numerical representation, or "embedding". This is done using a language model that has been trained to understand the meaning and context of words and phrases. The resulting embeddings capture the semantic meaning of the text in a form that can be used for various LLMs tasks.

### **(d) Load embeddings to Vector Database**

After the embeddings have been created, they are then stored in a database, often referred to as a vector database or vector store. This is a specialized type of database that is designed to handle high-dimensional vector data and to support efficient operations like nearest-neighbor search, which allows us to find the vectors that are most similar to a given query vector.

Storing the embeddings in a vector database allows for faster and more efficient retrieval of similar chunks of text when performing tasks like semantic search or document classification. This can be crucial in applications where the amount of data is large and performance is a key concern.

#### 4.4.5.2. Retrieval & Generation

Large Language Models (LLMs) have been trained on a vast amount of text data. They generate text by predicting the next word in a sequence, given the previous words.

One common method for doing this is through a process known as "retrieval-augmented generation". This is a two-step process that involves both retrieval of relevant information and generation of a response based on that information.

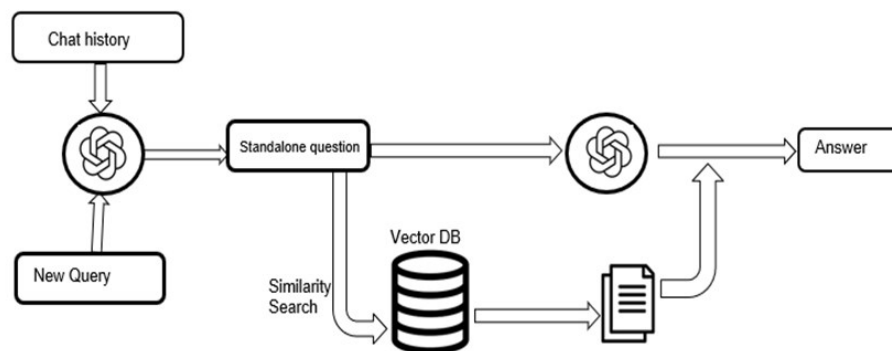


Figure 22: Retrieval-augmented generation of responses

## 4.5 Interface prototype

With the goal of querying IFC-based BIM data. For example, many application programming interfaces (APIs) are available to ease the use of programming languages to access the model repository. These APIs parse and manipulate the data in IFC instance data (e.g. Xbim. Essentials). A step forward was taken by the development of an open query language,

(a) **Receive user question:** The process begins when the user inputs a question or request. This question serves as the initial prompt for the LLM.



Key	Value	Private	Allow blank	Optional	Queryst.	Long
message	what are the cost elemen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
uuid	123	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Add parameter

Figure 23: Setting up sample query as parameter for the prototype system

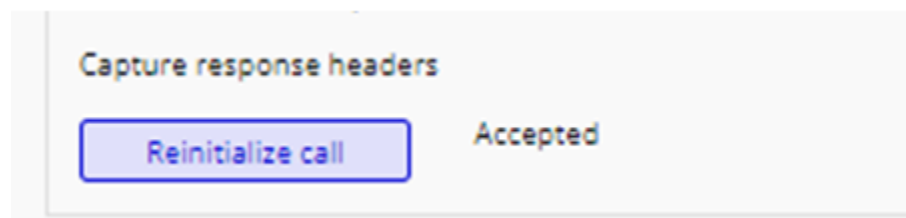
(b) **Lookup documents in the index relevant to the question (similarity search):** The system then searches an index of documents to find those that are most relevant to the user's question. This is typically done using a similarity search, which identifies documents that are semantically similar to the user's question.

(c) **Construct a PromptValue from the question and relevant documents:** Once the relevant documents have been identified, they are combined with the user's question to construct a new prompt. This prompt, referred to as a PromptValue, contains both the user's original question and the information from the relevant documents.

(d) **Pass the PromptValue to LLM:** The PromptValue is then passed to the LLM. The LLM uses this PromptValue as its new input, and generates a response based on this information.

(e) **Get back the result and return to the user:** The LLM's response is then returned to the user. This response should be more informed and accurate, as it is based on the additional information provided by the relevant documents.





*Figure 24: User prompt matches trained data*

This process allows LLMs to provide more accurate and informed responses to user questions, even when those questions pertain to information the LLM was not originally trained on. However, it's important to note that the effectiveness of this process depends on the quality and relevance of the documents in the index, as well as the ability of the similarity search to accurately identify the most relevant documents.

## 5. CASE STUDY

### 5.1 Project description

The BIM model used to demonstrate the interaction of LLMs, the BIM models themselves, and external cost databases represents a proposed seven-story student apartment complex. The complex has a plinth area of 2150 square meters, with the ground floor elevated 1.00 meter above ground level and a first-floor height of 3.50 meters. From the second to the fifth floors, a consistent floor height of 3.05 meters is maintained, while the sixth floor has a height of 2.60 meters. The seventh floor features a ceiling height of 2.45 meters. The structural composition consists of reinforced concrete with lightweight concrete masonry walls, each with finishes on either side, and reinforced concrete floor slabs. Each room is equipped with functional floor finishes bounding it. The model is designed with comprehensive specifications and metadata, facilitating subsequent information exchange.

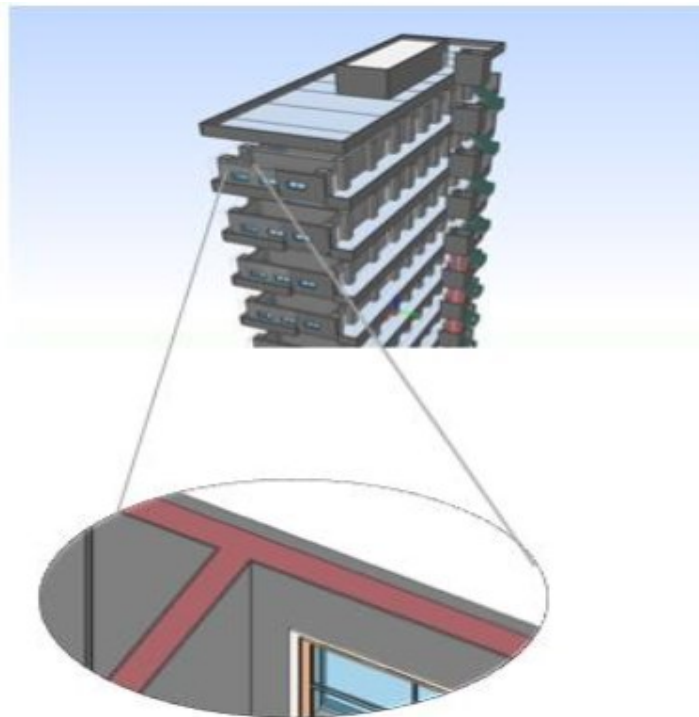


Figure 25:3D model of apartment building for case study

## 5.2 Standard methods for BIM based.

The ideal process for estimating a model consists of a systematic series of four crucial steps: BIM modeling, physical quality verification, property verification, and finally, the actual process of quantity take-off. In the initial phase of BIM modeling, the application of BIM authoring tools that are compatible with the Industry Foundation Classes (IFC) format comes into play (Choi et al., 2015). This step is crucial as it sets the basis for the subsequent steps in the QTO process.

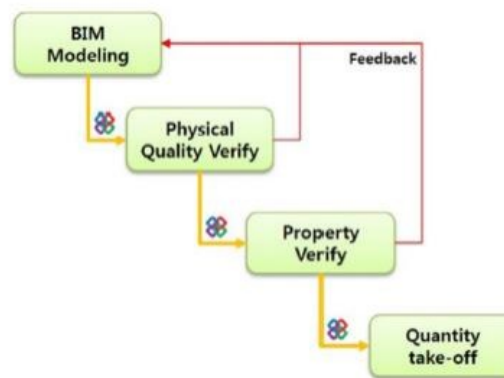


Figure 26: Open BIM-oriented QTO procedure (Choi et al., 2015)

Following the BIM modeling phase, the physical quality of the resulting BIM model is scrutinized. This is executed by deploying commercial software like the Solibri Model Checker (SMC), an industry-accepted platform for this task. This phase ensures the credibility of quantity information ascertained from the BIM model, a critical component for generating reliable estimates.

The third stage involves property verification, which is an extensive process of extracting structural elements from the BIM model. This step also includes a meticulous assessment of the construction code, a process that forms the backbone for the eventual expansion of the estimation work.

Finally, the BIM model is subjected to a thorough check of physical data for the computation of quantities. This forms the bedrock for the schematic estimation of the frame and ensures the accuracy of the estimation. This holistic process, when applied appropriately, offers a robust and accurate method for model estimation, ensuring a BIM model's quality directly influences the result's precision.

### 5.3 Demonstration of non-standard methods

The figure below shows the interaction between the user and the system with the model begin the source of data. In this case, the user wanted to know the different floor finish for different spaces in the model and their respective thickness

Provide a table of the floor finish and their respective thickness measurement

Based on the provided PDF content, a table of the floor finish and their respective thickness measurements can be created as follows:

Room/Space	Floor Finish	Thickness (mm)
Lounge	Cera.tiles	200
Dining	Cera.tiles	200
Guest room	Cera.tiles	200
Kitchen yard	P.V	5
Bath	Cera.tiles	200
WC	P.V	200
Master Bedroom	Cera.tiles	200
Wardrobe	P.V	200

Figure 27: Information retrieval of floor finish

In a different case, the user prompts the system to provide information for the staircase flight construction.

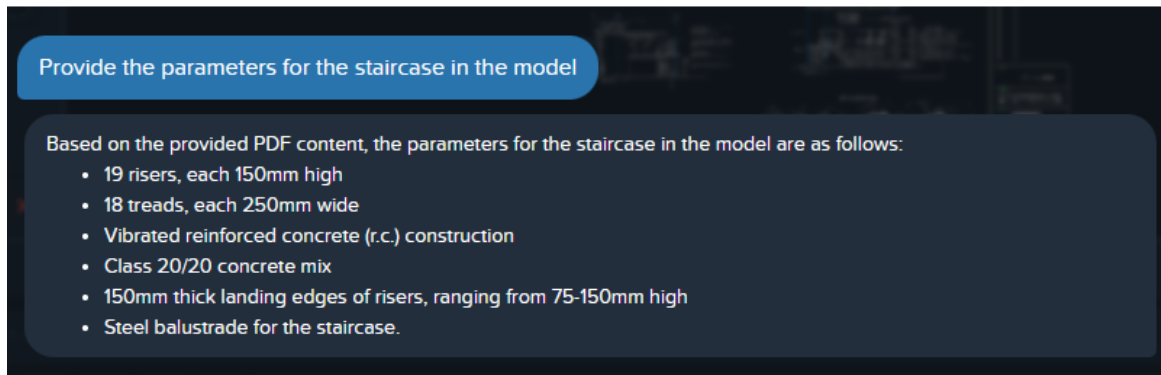


Figure 28: Information retrieval of staircase and cost database

While a BIM model cannot entirely have the above metadata input in a model, in this case, the system works in conjunction with trained cost database to include this information. This additional information is carefully considered as an estimate and not the exact material until proven otherwise.

ITEM	DESCRIPTION	UNIT
D	<b>CONCRETE WORKS</b> (The estimator will use his discretion to weight the rates for concrete location)	
a	<b>Mass Concrete</b>	
1	Plain concrete mix 1:4:8 blinding	CM
2	Mass concrete Class 15/20(1:3:6) in foundations	CM
3	100 mm thick surface bed Class 15/20(1:3:6)	SM
b	<b>Vibrated Reinforced Concrete Class 20/20mm (1:2:4)</b>	
1	Foundations / Ground beams / stairs / Column & bases	CM
2	150mm thick ground / suspended floor slab	SM
c	<b>Class 25/20mm (1:1 ½ :3 )</b>	
1	Foundations / Ground beams / stairs /Column & bases	CM

Figure 29: Section of cost database

The figure below shows a general query prompt by the user to determine the cost elements that are to be factored to achieve the floor finish in the model.

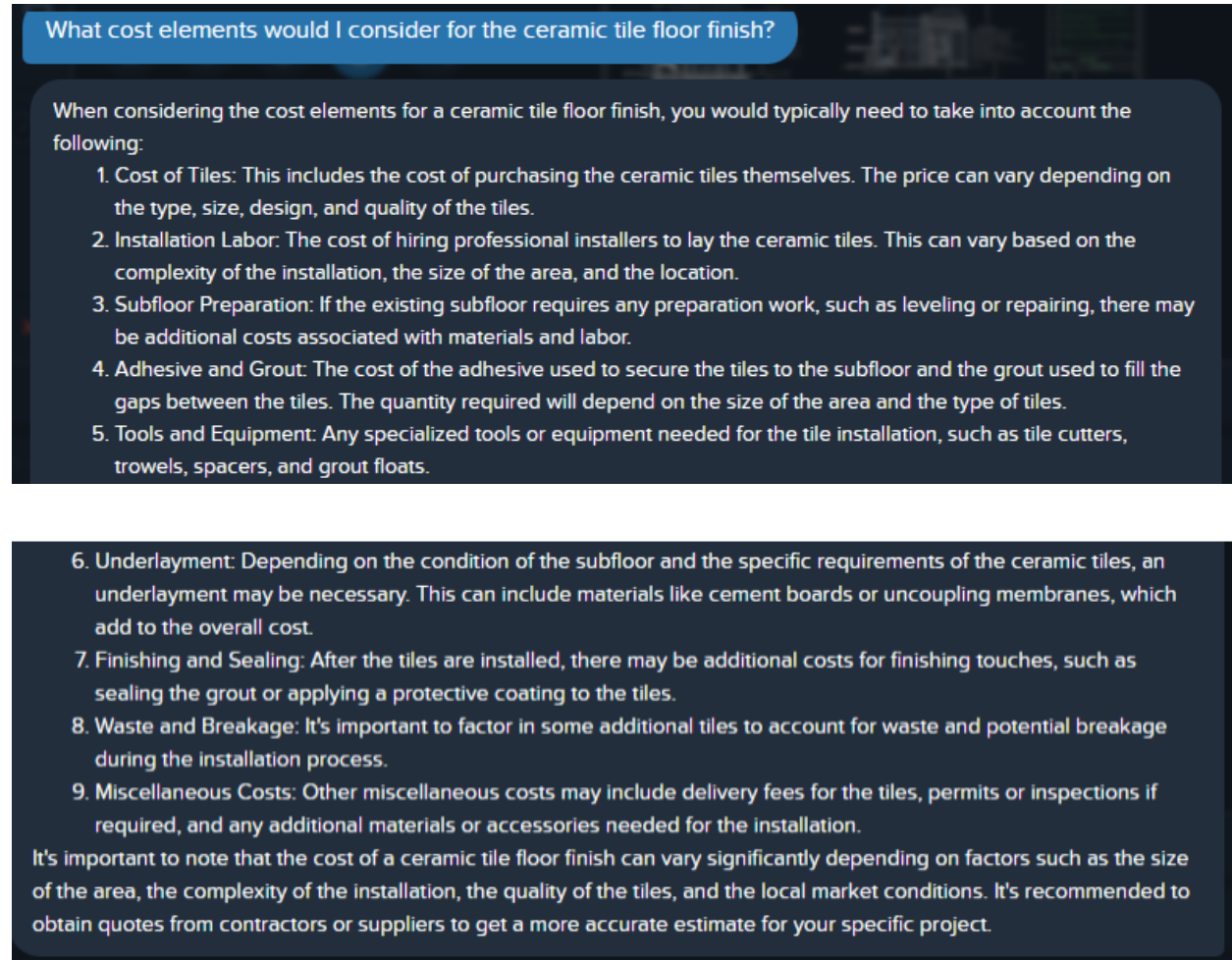


Figure 30: General information retrieved from ChatGPT

As seen, this is based on a general ChatGPT knowledge base to inform the user on cost items to consider when preparing a detailed cost estimate. The instant and prompt measurement and responses helps to retrieve information from models quickly and easily navigate the model facilitating the estimation process.

## 6 DISCUSSION AND CONCLUSIONS

### 6.1 Assessment of the Results

As seen previously on the demonstration of the prototype, the performance of the prototype is as expected. However, this is entirely true if the model metadata is input correctly for accurate information retrieval. User prompts, using the appropriate model metadata resulted to accurate or expected responses for estimation quantities.



Figure 31: Information retrieved from the model.

The system is primarily designed to generate responses from the model database and to pull out specific details from the model. For cost estimation purposes, the system is also linked to ChatGPT which has existing data for general query prompts such as *'cost elements for the floor finish'*. This includes common model features such as floors, walls, windows, and count elements such as doors. For ChatGPT model its temperature parameter was set to 0 to minimize the randomness of text generation.

To ensure accuracy of semantic search element terms like "pillar" and "column" are grouped under one standard term, "column." Similarly, "beams" and "girders," often used interchangeably, are standardized to avoid confusion. The library also accounts for different word forms, such as singular/plural or varying spellings. In data extraction, these standardized terms get matched to the right IFC categories.

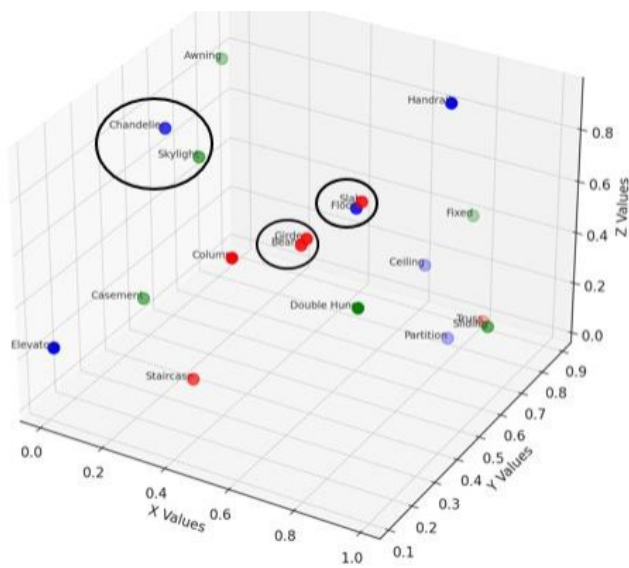


Figure 32: Semantic grouping of elements

The system's efficiency was examined by enriching the existing dataset with parameters and their values for NL requests. Both selection and projection parameters were added, as well as the value obtained from the user's request and its related predicted value in the database record (Zheng & Fischer, n.d.). Using low code tool Bubble, OpenAI API code and Pandas library, the system was tested for estimation. Data preprocessing involved extracting information from the



model and cleaning up using Python and stored in BSON format for ease of prompts and querying.

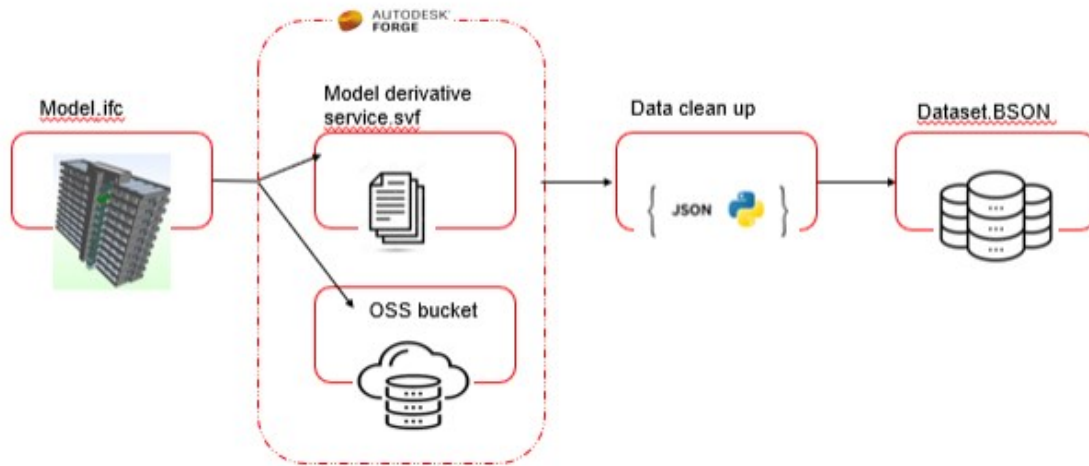


Figure 33: Data preprocessing from the model (Zheng & Fischer, 2022)

## 6.2 Accuracy of the Semantic Parsing

Code	Description	Unit	Price(€)
E_0020	Masonry wall 200 mm x 250 mm x 130 mm, thickness 20 cm, with cement mortar, including the burden of shoulder formation, turn, edges, pilasters, internal work surfaces.	M <sup>2</sup>	85.50
E_0030	Cement-based screed of expanded clay, laid in place on a perfect plan configured according to pre-established slopes, for an average thickness of 10 cm, including every other burden and teaching to give the finished work as a rule of art	M <sup>2</sup>	23.15

Annotations in the image: 'Cost item' points to the 'Code' column; 'Quantity Set' points to the 'Unit' column; 'Property set' points to the 'Description' column.

Figure 34: Cost description annotated and tagged for mapping.

In the research conducted, a prototype was designed to interpret user queries in natural language and efficiently access cloud-based BIM databases. Through this interface, users can retrieve specific building-related information without necessitating an in-depth knowledge of BIM.

To test the efficiency of semantic parsing of annotated texts, DOCCANO, is used for text classification, sequence labeling, and sequence to sequence tasks. The annotated texts is based on the IfcCostItem schema as shown in figure 34 and further broken down to specificity such as finishes, physical properties and geometry.

Table 5 shows successful parsing of annotated text from the user prompts and the retrieved value from semantic search of the database.

User prompt	Annotated Value	Cost Item	Retrieved Value
<b>What is the floor finish of the master bedroom?</b>	Floor	Floor finish (Property set)	Ceramic tiles
<b>Provide the thickness of the plaster of internal wall</b>	Thickness	Plaster thickness (Quantity set)	25mm

Table 5: Semantic parsing evaluation of user prompts vs information retrieval

A significant contribution of this research lies in the integration of the LLM model into the BIM cost elements and cost database framework. Historically, constructing such a system demanded considerable technical expertise, particularly in natural language processing and semantic analysis. However, by incorporating this model, many of the conventional complexities associated with syntactic analysis, semantic interpretation, and model training were effectively circumvented. Moreover, the AI model obviated the need for predefined response templates and served as a comprehensive external knowledge base. One of the more salient features of this approach was the capacity for iterative refinement through prompt adjustments.

The flexibility and adaptability of this natural language-based development not only streamlined the prototype creation but also enabled testing using different BIM models. This represents a marked departure from previous methodologies that were predominantly reliant on specialized technical experts for estimation methods. From the above-presented results, the AI estimation model prototype achieved its anticipated objectives, with successful result visualizations.

## 6.2 Limitations of the study

Testing unveiled certain limitations in the chosen development methodology. A SWOT Analysis, detailing the strengths, weaknesses, opportunities, and threats of the prototype for this use case, is provided in the table below.

Table 6: SWOT analysis of proposed prototype

<p style="text-align: center;"><b>Strengths:</b></p> <p><b>Time Efficiency:</b> AI-tools for estimation significantly reduce the time required for manual calculations and data analysis. With automation, AI models can process vast amounts of model data quickly, providing estimations in a fraction of the time it would take for manual calculations.</p> <p><b>Scalability:</b> The AI systems is highly scalable and can handle large datasets without compromising performance. As the volume of model data increases, AI algorithms can handle the complexity and still deliver accurate estimations consistently. This scalability is crucial for complex models that have a vast array of variables.</p>	<p style="text-align: center;"><b>Weaknesses:</b></p> <p><b>Data Quality and Availability:</b> Integrating AI with cost data relies heavily on the availability of high-quality, well structured, and relevant data. If the data used for training the AI model is incomplete or inaccurate, the responses result to flawed estimations.</p> <p><b>Cost data availability:</b> Ensuring data quality and accessibility can be a challenge, especially in cases where historical cost data is limited or not readily available.</p>
<p style="text-align: center;"><b>Opportunities:</b></p> <p><b>Enhanced Decision-Making:</b> Accurate estimations provided by AI can support decision-making processes by providing insights into project feasibility, and resource allocation.</p>	<p style="text-align: center;"><b>Threats:</b></p> <p><b>Reliance on Historical Data:</b> AI models for cost estimation heavily depend on historical cost data. However, if the industry undergoes significant changes, such as technological advancements or market disruptions, historical data may not accurately represent current cost dynamics. It is essential to regularly evaluate the relevance of historical data and supplement it with real-time data to ensure accurate cost estimations.</p>

	<b>Ethical Considerations:</b> Integrating AI with cost data raises ethical concerns, especially regarding data privacy. Cost data often includes sensitive information and especially with regards to tendering, and proper measures must be in place to protect the privacy and confidentiality of individuals and organizations involved. Ethical considerations should be addressed throughout the development and implementation of AI systems for estimation.
--	---

### 6.3 Future work

Future development of the system would primarily revolve around the integration of BIM authoring tools into the proposed framework. This envisioned progression entails adopting an iterative development approach to refine the existing prototype, marking a departure from the conventional method where the prototype autonomously loads the federated BIM model. Instead, in its more advanced form, the system would empower users to directly query the model within the BIM authoring tool, thereby fostering a more integrated and immersive user experience.

The incorporation of Linked Data Models (LLMs) into BIM authoring tools would establish a synergistic ecosystem of data interchange. Notably, a consequential outcome of this integration lies in the system's potential to proficiently generate comprehensive cost estimation reports in diverse formats, thus significantly augmenting the overall utility and adaptability of the system for enhanced construction project management practices.

In addressing the pivotal questions raised for consideration:

- How do we collect and analyze data from tens of thousands of previous projects, including planned and actual events, in an industry that traditionally lacks reliable and structured data records? It is acknowledged that many previous projects did not possess the digital maturity required for extensive data preservation. However, our present strategy seeks to mitigate

this challenge by meticulously organizing project data well in advance, employing comprehensive tagging of all information.

- What data must be collected on-site, either daily or weekly, to empower an AI tool for real-time progress analysis and future event prediction? These questions underscore the critical importance of data collection and its role in enhancing construction practices. Therefore, the proposed system is not only committed to advancing construction project management through BIM integration but also aims to efficiently address the data collection needs necessary for AI-driven insights. This holistic approach marks a significant step towards more informed and efficient construction project management.

## 7 REFERENCES

- Abanda, F. H., Kamsu-Foguem, B., & Tah, J. H. M. (2017). *BIM-New rules of measurement ontology for construction cost estimation*. <https://doi.org/10.1016/j.jestch.2017.01.007>
- Abdirad, H., & Mathur, P. (2021). Artificial intelligence for BIM content management and delivery: Case study of association rule mining for construction detailing. *Advanced Engineering Informatics*, 50, 101414. <https://doi.org/10.1016/J.AEI.2021.101414>
- Abioye, S. O., Oyedele, L. O., Akanbi, L., Ajayi, A., Manuel, J., Delgado, D., Bilal, M., Akinade, O. O., & Ahmed, A. (2021). Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. *Journal of Building Engineering*, 44, 2352–7102. <https://doi.org/10.1016/j.jobe.2021.103299>
- Agrawal, R., Imieliński, T., & Swami, A. (1993). *Mining association rules between sets of items in large databases*. 207–216. <https://doi.org/10.1145/170035.170072>
- Bini, S. A. (2018). Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *The Journal of Arthroplasty*, 33(8), 2358–2361. <https://doi.org/10.1016/J.ARTH.2018.02.067>
- Building Smart Industry Foundation Classes 4.0.2.1*. (n.d.).
- Charette, R. P., & Marshall, H. E. (1999). *UNIFORMAT II elemental classification for building specifications, cost estimating, and cost analysis*. <https://doi.org/10.6028/NIST.IR.6389>
- Cheng, M. Y., Tsai, H. C., & Hsieh, W. S. (2009). Web-based conceptual cost estimates for construction projects using Evolutionary Fuzzy Neural Inference Model. *Automation in Construction*, 18(2), 164–172. <https://doi.org/10.1016/J.AUTCON.2008.07.001>
- Choi, J., Kim, H., & Kim, I. (2015). Open BIM-based quantity take-off system for schematic estimation of building frame in early design stage. *Journal of Computational Design and Engineering*, 2(1), 16–25. <https://doi.org/10.1016/j.jcde.2014.11.002>
- Chou, J.-S., & Lin, C. (2012). Predicting Disputes in Public-Private Partnership Projects: Classification and Ensemble Models. *Journal of Computing in Civil Engineering*, 27(1), 51–60. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000197](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000197)

- Egwim, C. N., Alaka, H., Toriola-Coker, L. O., Balogun, H., & Sunmola, F. (2021). Applied artificial intelligence for predicting construction projects delay. *Machine Learning with Applications*, 6, 100166. <https://doi.org/10.1016/J.MLWA.2021.100166>
- Ghallab, M., Nau, D., & Traverso, P. (2004). *Automated Planning: theory and practice*. <https://books.google.com/books?hl=en&lr=&id=uYnpze57MSgC&oi=fnd&pg=PP1&ots=XnQZQ5KKWd&sig=xsyJS7r1LR-7i0Ve62Uw4GRyJnk>
- Goh, Y. M., & Chua, D. (2013). Neural network analysis of construction safety management systems: a case study in Singapore. *Http://Dx.Doi.Org/10.1080/01446193.2013.797095*, 31(5), 460–470. <https://doi.org/10.1080/01446193.2013.797095>
- Han, X., & Lu, M. (2010). Research on construction cost control based upon BP neural network and theory of constraint. *2010 International Conference on Management and Service Science, MASS 2010*. <https://doi.org/10.1109/ICMSS.2010.5576622>
- Kim, G. H., An, S. H., & Kang, K. I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and Environment*, 39(10), 1235–1242. <https://doi.org/10.1016/J.BUILDENV.2004.02.013>
- Lin, J.-R., Hu, Z.-Z., & Zhang, J. (2013). *BIM ORIENTED INTELLIGENT DATA MINING AND REPRESENTATION*.
- Lu, Y., Wu, Z., Chang, R., & Li, Y. (2017). Building Information Modeling (BIM) for green buildings: A critical review and future directions. *Automation in Construction*, 83, 134–148. <https://doi.org/10.1016/J.AUTCON.2017.08.024>
- Marinelli, M., Dimitriou, L., Fragakakis, N., & Lambropoulos, S. (n.d.). *Association of Researchers in Construction Management*. 853–862.
- Market Basket Analysis Overview*. (n.d.). Retrieved May 4, 2023, from [https://docs.oracle.com/cd/E76441\\_01/orase/pdf/141/html/implementation\\_guide/Output/market-basket-analysis.htm](https://docs.oracle.com/cd/E76441_01/orase/pdf/141/html/implementation_guide/Output/market-basket-analysis.htm)
- Mengtian Yin, L. T. C. W. S. X. X. L. H. Y. (2023). An ontology-aided, natural language-based approach for multi-constraint BIM model querying. *Research Gate*, 6–9.

- Meyer, C. (n.d.). *BIM-based LCA of Buildings: The Influence of LOD*. Retrieved May 21, 2023, from <https://agnhb.tugraz.at>
- Olsen, D., & Taylor, J. M. (2017). Quantity Take-Off Using Building Information Modeling (BIM), and Its Limiting Factors. *Procedia Engineering*, 196, 1098–1105. <https://doi.org/10.1016/J.PROENG.2017.08.067>
- Pan, Y., & Zhang, L. (2023). *Integrating BIM and AI for Smart Construction Management: Current Status and Future Directions*. 30, 1081–1110. <https://doi.org/10.1007/s11831-022-09830-8>
- Pauwels, P., Van Deursen, D., De Meyer, R., Van De Walle, R., Van Campenhout, J., De Roo, J., & Verstraeten, R. (2011). A semantic rule checking environment for building performance checking. *Automation in Construction*, 20(5), 506–518.
- Saka, A. B., Oyedele, L. O., Akanbi, L. A., Ganiyu, S. A., Chan, D. W. M., & Bello, S. A. (2023). Conversational artificial intelligence in the AEC industry: A review of present status, challenges and opportunities. In *Advanced Engineering Informatics* (Vol. 55). Elsevier Ltd. <https://doi.org/10.1016/j.aei.2022.101869>
- Sayed, M., Abdel-Hamid, M., & El-Dash, K. (2020). Improving cost estimation in construction projects. *Https://Doi.Org/10.1080/15623599.2020.1853657*, 23(1), 135–143. <https://doi.org/10.1080/15623599.2020.1853657>
- Silvers, F. (2011). Data Warehouse Designs : Achieving ROI with Market Basket Analysis and Time Variance. *Data Warehouse Designs*. <https://doi.org/10.1201/B11692>
- Tang, S., Liu, H., Almatared, M., Abudayyeh, O., Lei, Z., & Fong, A. (2022). Towards Automated Construction Quantity Take-Off: An Integrated Approach to Information Extraction from Work Descriptions. *Buildings* 2022, Vol. 12, Page 354, 12(3), 354. <https://doi.org/10.3390/BUILDINGS12030354>
- Tang, S., Shelden, D. R., Eastman, C. M., Pishdad-Bozorgi, P., & Gao, X. (2019). A review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends. *Automation in Construction*, 101, 127–139. <https://doi.org/10.1016/J.AUTCON.2019.01.020>



- Wang, H., Pan, Y., & Luo, X. (2019). Integration of BIM and GIS in sustainable built environment: A review and bibliometric analysis. *Automation in Construction*, 103, 41–52. <https://doi.org/10.1016/J.AUTCON.2019.03.005>
- What is Artificial Intelligence (AI)? | IBM.* (n.d.). Retrieved July 5, 2023, from <https://www.ibm.com/topics/artificial-intelligence>
- Wijayakumar, M., & Suranga Jayasena, H. (2013). Automation of bim quantity take-off to suit qs's requirements. *The Second World Construction Symposium*. <http://dl.lib.uom.lk/handle/123/18649>
- Xi'an Dianzi, & Keji Daxue Xuebao. (2020). COMPARATIVE ANALYSIS OF CONSTRUCTION COST ESTIMATION USING ARTIFICIAL NEURAL NETWORKS. *Journal of Xidian University*, 14(7). <https://doi.org/10.37896/JXU14.7/146>
- Yan, H., Ma, M., Wu, Y., Fan, H., & Dong, C. (2022). Overview and analysis of the text mining applications in the construction industry. In *Heliyon* (Vol. 8, Issue 12). Elsevier Ltd. <https://doi.org/10.1016/j.heliyon.2022.e12088>
- Yao, X., Zhou, J., Zhang, J., & Boer, C. R. (2017). From Intelligent Manufacturing to Smart Manufacturing for Industry 4.0 Driven by Next Generation Artificial Intelligence and Further on. *Proceedings - 2017 5th International Conference on Enterprise Systems: Industrial Digitalization by Enterprise Systems, ES 2017*, 311–318. <https://doi.org/10.1109/ES.2017.58>
- Zhao, H. (n.d.). *Automation of Quantity Takeoff and Material Optimization for Residential Construction Manufacturing*.
- Zheng, J., & Fischer, M. (n.d.). *BIM-GPT: a Prompt-Based Virtual Assistant Framework for BIM Information Retrieval*.
- Zou, Z., & Ergan, S. (2019). *Leveraging Data Driven Approaches to Quantify the Impact of Construction Projects on Urban Quality of Life*. <https://arxiv.org/abs/1901.09084v3>
- Алшабаб, Ш., Халил, & Петроченко. (2017). *BIM-Based Quantity Takeoff*.