Bridging AI Readiness and Application:

Prototyping a Strategy-Aligned Language Model for Quality Insights at Skanska

A Comprehensive Study of Organizational AI Maturity, Applied NLP Development, and Scalable Implementation in Construction Quality Management

Master's Thesis in Complex Adaptive Systems, and Quality and Operations Management

LISA LÖVGREN
OLIVIA TURUNEN

DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025
www.chalmers.se

Master's thesis 2025

Bridging AI Readiness and Application: Prototyping a Strategy-Aligned Language Model for Quality Insights at Skanska

A Comprehensive Study of Organizational AI Maturity, Applied NLP Development, and Scalable Implementation in Construction Quality Management

LISA LÖVGREN
OLIVIA TURUNEN

Department of Civil and Environmental Engineering
Chalmers University of Technology
Gothenburg, Sweden 2025

Bridging AI Readiness and Application: Prototyping a Strategy-Aligned Language Model for Quality Insights at Skanska
A Comprehensive Study of Organizational AI Maturity, Applied NLP Development, and Scalable Implementation in Construction Quality Management
LISA LÖVGREN, OLIVIA TURUNEN

Supervisor: Rasmus Rempling, Department of Civil and Environmental Engineering
Supervisor: Peter Samuelsson, Skanska AB Examiner: Rasmnus Rempling, Department of Civil and Environmental Engineering

Master's Thesis 2025
Department of Civil and Environmental Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Image generated with ChatGPT 4o, April 2025, prompted: "Construction and AI" merged with resulting category clusters of data subset.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2025

Bridging AI Readiness and Application: Prototyping a Strategy-Aligned Language Model for Quality Insights at Skanska
A Comprehensive Study of Organizational AI Maturity, Applied NLP Development, and Scalable Implementation in Construction Quality Management
LISA LÖVGREN, OLIVIA TURUNEN
Department of Civil Engineering, Chalmers University of Technology

Abstract

The construction industry is under increasing pressure to improve efficiency, reduce costs, and enhance sustainability. While other sectors have advanced in AI adoption, construction remains comparatively behind. This thesis explores how artificial intelligence (AI) can support decision-making in construction, with a focus on Quality Management at Skanska Sweden AB.

First, organizational AI Readiness was assessed through interviews and workshops using established organizational frameworks. This reveals both strategic interest and practical challenges in applying AI. Second, an operational use case was explored by developing an AI prototype that processes historical quality deviation texts. The prototype was developed with the purpose of creating value for the Quality Department by providing insight. Using natural language processing (NLP), the prototype explored a weakly supervised classification approach combining unsupervised clustering, pseudo-labelling via zero-shot learning, and a fine-tuned transformer classifier (XLM-R and SBERT). Two promising category types, incident type and affected building component, were identified and co-developed with domain experts to structure the data.

The results show that while AI readiness is moderate, initiatives often remain siloed due to limited infrastructure, resources, and unclear ownership. Skanska shows a growing awareness and curiosity around AI and there is potential to learn from international practices within the company. However, although large volumes of data available, barriers remain particularly in terms of the availability of structured and labelled data. There is also a need for further AI-specific expertise, and it remains challenging to integrate new tools into established workflows. The prototype demonstrates practical value by visualizing patterns in text data, enabling the Quality Department to adopt a more data-driven and preventive approach. While weak supervision proved challenging due to limited label quality and model sensitivity, the final classifier achieved approximately 67% accuracy through fine-tuning with a manually labelled dataset, accounting of 6‰. Despite this, the approach successfully enabled structured insights into issue frequency, duration, and distribution across projects. The prototype also serves as a scalable proof of concept, illustrating how tailored AI solutions can accelerate digital transformation in construction.

Keywords: Artificial Intelligence (AI), Natural Language Processing (NLP), Language Model Prototype, Text Classification, AI Readiness, Quality Management, Change Management, Construction Industry, Digital Transformation.

Acknowledgements

We would like to express our sincere gratitude to everyone who contributed to the completion of this thesis.

First and foremost, we would like to thank our academic supervisor at Chalmers University of Technology, Rasmus Rempling, for their guidance and encouragement throughout the research process, and company supervisor, Peter Samuelsson, for their engagement and reliability. We are also grateful to Skanska for providing the opportunity to conduct this thesis within the organization and for sharing valuable insights, supporting data access, and engaging in fruitful discussions. We also want to express our gratitude towards Brosamverkan for enabling our international exchange. Finally, we want to thank all participants in the workshops and interviews for their time and expertise, which significantly enriched the outcomes of this work.

Lisa Lövgren, Olivia Turunen, Gothenburg, May, 2025

List of Acronyms

Below is the list of acronyms used throughout this thesis listed in alphabetical order:

List of Abbreviations

AEC	Architecture, Engineering and Construction
AI	Artificial Intelligence
BERT	Bidirectional Encoder Representations from Transformers
BIM	Building Information Modeling
DL	Deep Learning
GDPR	General Data Protection Regulation
GPT	Generative Pre-trained Transformer
IoT	Internet of Things
KMeans	K-Means Clustering
KPI	Key Performance Indicator
LLM	Large Language Model
LOTClass	Learning with Out-of-the-box Classifier for Text Classification
LSTM	Long Short-Term Memory
MEGClass	Mixed Expert Guided Classification
ML	Machine Learning
NLI	Natural Language Inference
NLP	Natural Language Processing
NN	Neural Network
PCA	Principal Component Analysis
POS	Part-of-speech
RNN	Recurrent Neural Network
RQ	Research Question
SBERT	Sentence-BERT (Bidirectional Encoder Representations from Transformers for Sentence Embeddings)
SQL	Structured Query Language
STS	Semantic Textual Similarity
t-SNE	t-distributed Stochastic Neighbour Embedding
X-Class	Explainable Classifier for Weakly Supervised Text Classification
XLM-R	Cross-lingual Language Model - RoBERTa
ZSL	Zero-Shot Learning

Contents

List of Figures

List of Tables

1
Introduction

As one of the world's largest and most traditional sectors, the construction industry is facing a pivotal moment in its digital evolution. It is increasingly challenged by global conditions such as rising labor and material costs, supply chain disruptions, and growing sustainability demands. While these external pressures cannot be fully controlled, the industry can respond by transforming its internal processes and accelerating artificial intelligence (AI) adoption. While artificial intelligence has already transformed industries such as finance and manufacturing, construction continues to lag behind, held back by fragmented processes and limited standardization. This thesis explores how AI can become a catalyst for change in construction, not by replacing human expertise, but by supporting it. Focusing on a case within Skanska Sverige AB, the study investigates both organizational readiness and practical application through the development of an AI prototype for quality and deviation management. The goal is to demonstrate how AI can unlock insights, enhance preventive work, and lay the foundation for scalable, long-term value creation with AI, which will be obtained by implementing an AI prototype for natural language processig (NLP) of large amount of texts.

1.1 Background

The background section introduces the broader characteristics and challenges of the construction industry, outlines the transformative potential of AI, and presents the case company and the organizational setting in which this research is conducted.

1.1.1 Construction Industry

The construction industry is characterized by complexity, fragmentation, and a high degree of project uniqueness. Standardization is difficult to achieve, as even buildings based on identical designs need to be adapted to varying geological and climat conditions. Each project involves new owners, contractors, subcontractors, and suppliers working together. This setup makes it challenging to implement repeatable processes, which limits the potential for automation and the integration of intelligent technologies. [4]

In addition, construction projects follow non-linear and unstructured workflows, where tasks are interdependent and often executed in parallel by subcontractors with varying digital maturity. The physical environment is constantly changing, making coordination and real-time information sharing critical and difficult. External factors such as noise, dust, and instability further complicate data collection and reduce the reliability of technical systems. These characteristics create uncertainty, inefficiencies, and barriers to innovation. [4]

Some of the challenges currently facing the construction industry are linked to broader societal issues and global conditions, for example, rising labour and material costs caused by worldwide shortages. While these external factors cannot be fully controlled, the industry can respond by transforming its internal processes and operations. By adopting new technologies, the construction sector can increase efficiency, improve resource management, and become more adaptable in a rapidly changing world. A shift in mindset to remain competitive and sustainable is also essential. [5]

1.1.2 Artificial Intelligence

In recent years, AI has shifted from a theoretical concept to a practical tool reshaping industries. Rather than referring to a single technology, AI includes a collection of methods, such as machine learning, computer vision, and natural language processing, that allow systems to interpret data, recognize patterns, and support decision-making. These capabilities are now used to optimize supply chains, predict equipment failures, and automate complex tasks across sectors.

Although construction has historically been one of the least digitized industries, this is starting to change. With a global value exceeding USD 12 trillion (at the time corresponding to SEK 126 trillion), the construction sector is under increasing pressure to modernize in response to labour shortages, rising costs, and sustainability goals [6]. AI offers a way forward: from enhancing safety through image recognition on job sites, to optimizing schedules using predictive algorithms, to integrating real-time data with Building Information Modeling (BIM) [2].

Investment trends reflect this growing interest, between 2020 and 2022, over USD 50 billion (at the time corresponding to SEK 550 billion) was invested globally in AEC (architecture, engineering, and construction) technologies, with a significant portion directed toward late-stage ventures [6]. Although still in early stages of adoption compared to other industries, AI holds considerable promise in addressing long-standing challenges in construction. From improving schedule reliability and reducing safety risks to enhancing resource planning and data integration, AI-based solutions offer a pathway toward greater efficiency and control, if the industry can overcome barriers related to culture, fragmentation, and digital maturity.

1.1.3 Case Company

Skanska AB is one of the world's leading construction and project development companies, founded in Sweden in 1887. With over 135 years of experience, the company has grown into a global player operating in selected markets in the Nordics, Europe, and the United States. Skanska Sweden employs approximately 6 700 people and had an operating income of SEK 1,2 billion in 2024 [7]. Their focus is on building a better society and being a leader in sustainable solutions, quality, safety, and ethics [8].

In Sweden, Skanska operates through several business units under Skanska Sweden AB. The core construction and civil engineering operations consist of four main divisions: Skanska Hus (Building Construction), Skanska Väg och Anläggning (Civil Infrastructure), Skanska Industrial Solutions, and Skanska Rental. Construction and civil engineering activities account for approximately 85% of Skanska Sweden's total revenue [9].

Skanska Swedens's operations are structured around three overarching business streams: Construction, Residential Development, and Commercial Property Development. The Construction business stream includes both Civil construction, i.e. roads, tunnels and bridges etc, and Building construction (=Hus). This study focuses exclusively on Skanska Hus, where operations are divided into regional units. The study takes its starting point in the region of Skanska Hus Göteborg, broadening to focusing on national quality function within Skanska Hus's building operations in Sweden, specifically within the unit.

Furthermore, interviews have also been conducted with representatives from Skanska's U.S. organization in New York City, and Skanska UK, to explore successful applications of artificial intelligence in other parts of the world. This provided a broader perspective on how AI readiness and implementation strategies may differ across geographical and organizational contexts.

1.2 Purpose

The overall purpose of this study is to explore how AI can enhance strategic decision-making and operational efficiency within the construction industry, focusing specifically on Skanska Hus, Skanska Sverige. The study takes its starting point in the recognition that, while many industries have already made significant progress in adopting AI technologies, the construction sector remains comparatively behind, largely due to its project-based nature, fragmented workflows, and varying levels of digital maturity. In this context, the study seeks to understand how the construction industry can begin to close this gap by identifying concrete opportunities, organizational prerequisites, and value-generating use cases for AI. Through an exploratory, inductive approach, the research is divided into three interconnected tracks, each addressing a distinct aspect of AI adoption and aligned with the study's three research questions.

The first focus is to analyse organizational AI readiness on a high level, assessing the current state and maturity level within Skanska from an external perspective. By applying established frameworks and theory on digital and AI transformation, learning from other industries, this part of the study identifies internal conditions, barriers, and enablers for AI implementation.

The second focus is narrowing in on the area of quality management, identified during the first exploratory phase, as an area with both business value and technical feasibility. Through creating an AI-prototype, this part of the study analyses how AI-generated insights can be integrated into decision-making and knowledge-sharing in a construction context. It also reflects on broader organizational implications for scaling AI solutions beyond isolated use cases.

The third focus concerns the practical implementation of AI and the technical performance and potential of the prototype itself. Here, an AI prototype is developed using NLP and ML to structure and analyse text-based quality data. Through iterative testing, evaluation, and visualization of the results produced from the prototype, the goal is to create an efficient and accurate model.

In conclusion, the study aims to contribute to a more nuanced understanding of how the construction industry, in this case Skanska AB, can begin to adopt, implement, and benefit from AI, specifically using text-processing to interpret quality data.

1.3 Research Questions

With background to above Chapter 1, the report will address three research questions (RQs), which will be accompanied by relating theory, implementation, results and analysis. The RQs are the following:

- **RQ 1:** *What is Skanska's current state of AI readiness, and what organisational strengths and challenges exist in terms of AI adoption?*

- **RQ 2:** *How can an AI-prototype create value within the department of Quality Management?*

- **RQ 3:** *How can a language model prototype be developed and used to efficiently extract and visualize insights from quality-related construction texts?*

1.4 Limitations

This study is subject to several limitations that should be acknowledged. First, the findings are primarily based on data and observations from a single Swedish construction company, which may limit the generalizability of the conclusions to the broader industry. Although the company is an important actor on the Swedish market, its internal processes, digital maturity, and AI-related initiatives may not fully reflect those of smaller firms or companies operating in other regions or contexts.

Second, the assessment is conducted from the perspective of an external party and does not include full access to internal strategic documentation, proprietary datasets, or project-level decision-making. This means that some conclusions, particularly those related to organizational readiness, internal barriers, or technology adoption trajectories are based on secondary sources or publicly available information rather than first-hand implementation data.

Finally, while the analysis attempts to incorporate a broad view of AI technologies, it does not encompass all emerging fields or niche use cases. The emphasis is placed on applications that are currently most relevant to the construction sector, based on industry reports and peer-reviewed literature.

2

Frame of Reference

This Section outlines the theoretical foundation for the study, starting by covering the construction industry's unique characteristics. To understand how AI can be successfully adopted, the Section further explores theoretical frameworks for managing organizational change and assessing AI readiness, along with strategic approaches to AI adoption. Finally, it also introduces key AI components relevant to the prototype, including data criteria, language models, and pre-trained architectures. Together, these perspectives offer a comprehensive reference for understanding how construction firms can approach and implement AI effectively.

2.1 Construction

The construction industry differs from many other industries due to its fragmented and project-based nature, where buildings are assembled through sequential, yet disconnected, processes. This discontinuity has contributed to the industry's slow industrial and digital development. While technologies such as Internet of Things (IoT), big data, cloud computing, and AI have driven digitalization across many industries, particularly manufacturing, the construction sector remains behind. In recent years, however, these technologies have begun to be applied in construction. Despite this progress, their adoption remains limited to isolated areas, highlighting the need for a more integrated and systemic implementation [4].

The following Section outlines key characteristics of the construction industry to provide a foundation for understanding the potential barriers to the application of AI.

2.1.1 Construction Industry Characteristics

Standardized construction plans are rare, as buildings based on identical designs still differ due to varying geological and climatic conditions. As a result, each construction project is unique, making standardization difficult. For example, generating detailed and reusable bills of materials for future projects is highly complex. This uniqueness also comes from the dynamic composition of project teams, where owners, contractors, subcontractors, and suppliers vary from project to project. The construction industry operates through project unique production systems. The development of automated and integrated intelligent systems requires modular and repeatable components and processes, an approach that remains difficult to implement in this highly fragmented sector.

Furthermore, construction projects follow non-linear processes and are typically organized in an unstructured way. Rather than forming a sequential chain, tasks are interlinked through shared resources and parallel activities. A significant part of the work is carried out by subcontractors, whose varying levels of digital maturity and project involvement affect the reliability of the information provided, both during execution and in the aftermarket. These fragmented information flows contribute to miscommunication among project participants and hinder effective documentation for future use.

Coordination between participants is crucial, as overlapping workspaces, task sequences, and movement paths often create conflicts. This becomes even more challenging because both the location and the environment change throughout the project. In addition, construction equipment, materials, and labor need to be continuously relocated as the work progresses.

Furthermore, construction projects are highly complex and uncertain. Although construction plans are often detailed, they are frequently modified during execution to adapt to dynamic environments, which can result in delays, rework, quality deficiencies, and claims. To manage this uncertainty, project managers tend to incorporate large margins and risk buffers into the planning. While this helps prevent issues, it can also lead to inefficient use of resources.

Unlike a clean and controlled environment like in other industries, construction sites are often harsh and, as mentioned, unpredictable. Factors like noise, dust, mud, and even risks of geological disasters, pose significant challenges for data collection, network communication, and the reliability of intelligent systems. Moreover, workers exposed to such environments may lack real-time information, limiting their ability to respond to dangerous situations. This insecurity also reduces their willingness to engage with technical or automated equipment. [4].

Furthermore, the construction sector faces persistent challenges related to organizational inertia and data fragmentation, both of which limit the adoption and scalability of digital technologies, including AI. Organizational inertia in construction refers to the sector's resistance to change, driven by old practices, fragmented project structures, and legacy systems that slow the uptake of new technologies. At the same time, construction data is often siloed across different platforms and stored in non-standardized formats, restricting opportunities for cross-project integration and advanced analytics [10]. While construction companies generate large volumes of data, this data is often fragmented and lacks standardized formats, which hinders cross-project learning, collaboration, and data-driven decision-making. Low digital maturity in construction firms compounds this challenge, as proactive leadership and clear data governance structures are often missing [11]. These combined insights underscore a dual challenge for construction firms seeking to leverage AI: they must overcome cultural and structural barriers while simultaneously ensuring that project data is consistent, accessible, and actionable [10, 11].

2.1.2 Problem Area and Applications

AI has emerged as a cornerstone technology in the Fourth Industrial Revolution, transforming the way industries operate through data-driven decision-making, automation, and predictive capabilities [2]. Despite these advancements, the construction industry, valued at over USD 12 trillion globally (at the time corresponding to SEK 126 trillion) [6], has traditionally lagged behind in terms of digitization and AI adoption. Characterized by fragmented stakeholders, manual workflows, and analogue tools, the construction sector has historically demonstrated resistance to technological change [2]. A McKinsey research highlights that, as of 2018, construction ranked among the least digitized sectors when compared with twelve other industries [1]. However, growing demands for infrastructure, increasing labour shortages, and regulatory pressures for transparency have begun to catalyse a digital transformation across Architecture, Engineering, and Construction (AEC) industries.

This transformation is being accelerated by AI technologies that promise to address many of the sector's most persistent challenges, such as cost overruns, safety incidents, and inefficiencies in planning and resource allocation. For instance, AI-powered image recognition can identify unsafe worker behavior from site footage, while machine learning models can optimize scheduling by evaluating millions of potential timelines. Enhanced analytics platforms are being used to monitor sensor data in real time, improving both predictive maintenance and operational decision-making [1]. McKinsey estimates that from 2020 to 2022, global investment in AEC technology reached USD 50 billion (at the time corresponding to SEK 550 billion), an 85% increase compared to the previous three years, with 1,229 deals closed during that period [6]. This trend is illustrated in Figure 2.1, which shows both the sharp increase in funding and the growing number of deals in AEC technology over the past few years. These investments underscores a growing recognition that AI an opportunity for the future of construction. However, the construction industry shows both low current AI adoption and low future investment compared to other industries, positioning it closer to the “Falling behind” in Figure 2.2. This suggests a slow pace of digital transformation, which may limit its ability to capitalize on AI.

While enthusiasm around AI in construction is growing, the industry's structural and operational characteristics pose unique challenges to large-scale AI adoption. Construction firms are often small, project based entities with limited digital infrastructure and constrained IT budgets, typically spending only 1-2% of their revenue on IT, compared to 3-5% in other sectors [6].

AI applications in construction can be categorized across multiple domains: project planning and scheduling, site monitoring, resource and waste optimization, health and safety analytics, contract management, and supply chain logistics [2]. For example, AI enables more accurate cost estimation and scheduling by leveraging large

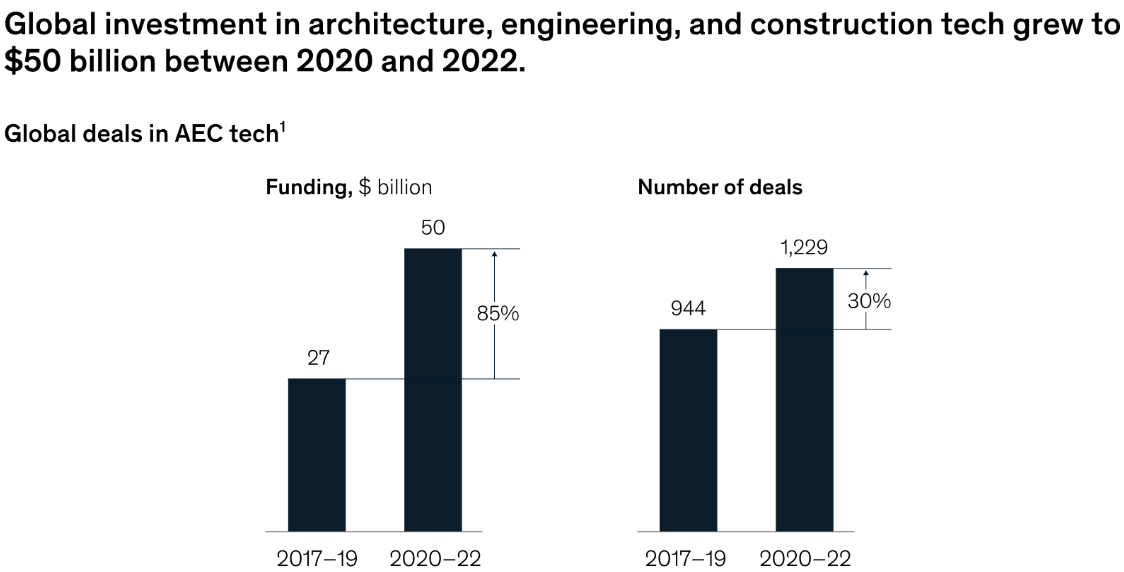


Figure 2.1: Tech investments in Construction[1].

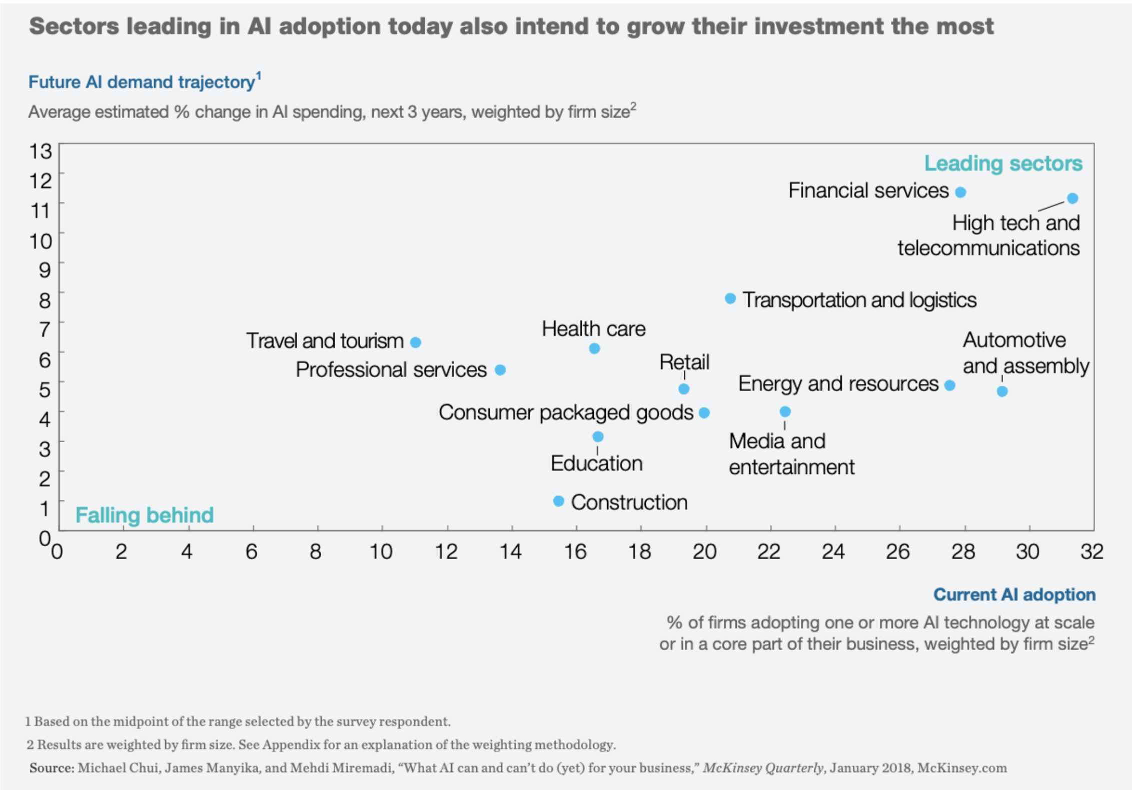


Figure 2.2: AI Adoption in Construction vs Other Industries [1].

datasets from previous projects and external factors such as weather or site conditions [2].

AI is also being used to optimize resource allocation and minimize material waste. By analyzing historical and real-time data, intelligent algorithms can forecast material demand, suggest optimal delivery timing, and reduce storage costs, contributing to both sustainability and profitability. Furthermore, AI can analyze data from sensors, drones, and connected machines to provide construction site analytics. These tools support real-time monitoring of productivity, safety risks, and performance bottlenecks, enabling more responsive site management [2].

Additionally, AI is used to read and analyze complex construction contracts. These AI tools can point out risky clauses, unclear parts, or mistakes, which helps make better purchasing decisions and lowers the risk of legal problems. AI-powered audit systems are also being implemented to ensure financial accuracy by cross-referencing billing data, flagging anomalies, and aligning invoices with real-world progress. This strengthens financial governance and supports more transparent reporting structures [2].

Despite these opportunities, several barriers remain, seen in Figure 2.3. Cultural resistance to change, high initial deployment costs, and a shortage of AI talent continue to hinder adoption. These challenges are illustrated in the lower left quadrant of the figure, which highlights issues such as ethics, governance, and limited internet connectivity. In addition, concerns around data ownership, transparency, and cybersecurity (shown in the top right quadrant) remain unresolved. Nevertheless, the trajectory is clear, the construction industry stands at a pivotal moment where AI can redefine its operational norms. Companies that act early and strategically to incorporate AI, will gain competitive advantages in cost efficiency, project reliability, and overall value delivery [2].

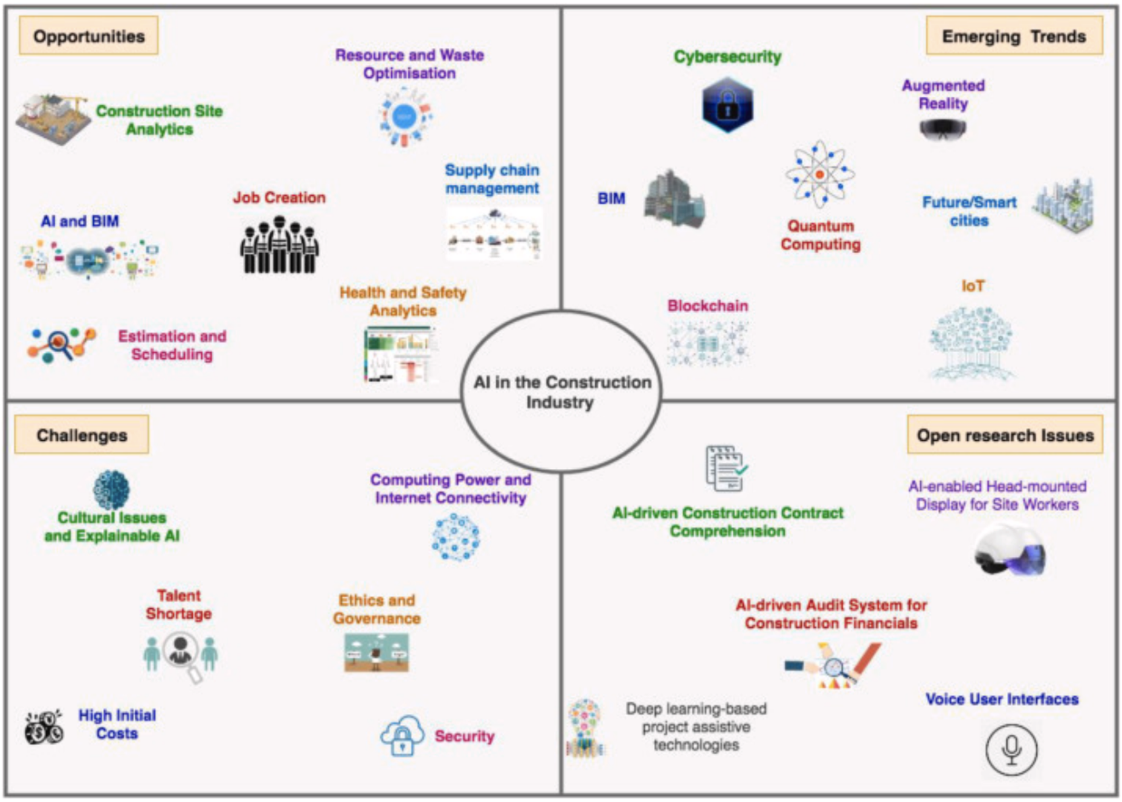


Figure 2.3: Overview of AI Use in the Construction Industry [2].

2.2 Business Frameworks

This chapter establishes the theoretical foundation for understanding how businesses adopt and integrate new technologies. It begins with a broad overview of change management, then narrows down to established frameworks for technology adoption, ultimately focusing on AI readiness and AI adoption strategies. Together, these perspectives provide a basis for analysing how businesses can integrate AI effectively, ensuring alignment between technology, organizational structures, and industry conditions. The frame of reference presented will later be used in the analysis to assess construction companies like the case company, Skanska Sweden, can structure their approach to AI use.

2.2.1 Change Management

Change management focuses on how organizations and individuals adapt to organizational transitions and change. One of the earliest models, proposed by Lewin in 1947 [12], includes the stages of unfreezing, moving, and refreezing. Since then, several other models have been introduced in the literature. While these models vary, they all emphasize the need for a structured approach to managing change and recommend appointing a change agent to lead the process. However, applying traditional change models directly to the construction industry can be challenging due to its unique characteristics, the industry's specific nature requires adaptations to these frameworks for them to be effective [12].

Successfully adopting AI within complex, project-oriented organizations in the construction industry requires not only technical capability but also a structured approach to change management. Kotter's Eight-Stage Model for Leading Change offers an approach for this transition. His model emphasizes the importance of establishing a sense of urgency, creating a guiding coalition, and developing a vision for change, all of which are critical for aligning AI initiatives with organizational priorities and engaging employees in the transformation process. Importantly, Kotter also highlights the need for short-term wins and continuous reinforcement, both of which can help ensure that AI is seen as a credible, valuable tool rather than an abstract or disruptive innovation. This model underscores the idea that even well-designed AI tools will struggle to gain traction if the organizational environment is not ready to support them [13].

Technology Acceptance Model
A widely recognized framework for understanding how individuals adopt and use new technologies or services is the Technology Acceptance Model, first introduced by Davis in 1989 [14]. It is based on the fact that the user's decision to accept and use a technology is primarily influenced by Perceived Usefulness and perceived ease of use. Perceived Usefulness is about how an individual believes that using a particular technology will improve their performance in the specific task or work. Perceived Ease of Use is about how an individual believes that using a particular system will be free from effort [14].

AI Readiness
The AI Readiness Framework, developed by Holmström in 2022, evaluates an organization's ability to implement and use AI in a way that adds value to the organization. It is structured around four key dimensions: technologies, activities, boundaries, and goals. The framework also helps organizations to address key bottlenecks in AI adoption. This framework also serves as the basis for subsequent research on AI readiness, which is further elaborated in the next Section 2.2.2 [15].

2.2.2 AI Readiness Framework

Building on Holmström's study, Tehrani et al. studied 52 multinational corporations on their orgaizational AI readiness, they conducted 52 semi-structured in-depth interviews with decision makers across different organizations. Ultimately, the findings identify eight key dimensions that influence an organization's ability to successfully implement AI. The study primarily focuses on the following, but is not strictly limited to them: natural language processing, computer vision, image recognition, and deep learning. The research is focusing on a model for organizational readines, and one for AI adoption strategies [3].

Organizational readiness refers to how ready the organization is in terms of cognitive, emotional, and behavioural preparedness toward a change, before the a activity is started. Organizational readiness should be acquired before starting a change initiative, and therefore consists of a pre-assesment of organizational capabilities, to help identify what is needed, and to control the risk factors with a change initiative. AI readiness refers to how prepared an organization is to adopt and use AI. Many organizations expect AI to increase their productivity, however many struggle with adopting AI due to lack of important infrastructure and organizational readiness. Many managers are also unsure of wheter their organization is ready to implement AI, and if so how to do it. Due to AI's complex nature, it's readiness can not only rely on traditional readiness theories [3].

The *AI-Readiness Framework* is an organizational framework that can be categorized in eight cathegories, as shown in Figure 2.4: (1) Environmental Readiness, (2) Technological Readiness, (3) Informational Readiness, (4) Infrastructural Readiness, (5) Data Readiness, (6) Participants' Readiness, (7) Customers' Readiness, and (8) Process Readiness.

Figure 2.4: Schematic model of AIR [3].

(1) Environmental Readiness
This dimension refers to the organizational, technical, competitive, cultural, and regulatory environment within which an organization operates. This includes the macro-economic environment, organizational culture, and leadership. The regulatory environment affects AI implementation, as some countries have more supportive or restrictive regulations in areas such as personalized data usage, cloud computing, and AI-driven automation. These regulations determine whether companies can freely adopt AI or must navigate legal constraints. An open organizational culture fosters collaboration, tech-friendliness, and knowledge-sharing, making AI adoption easier. Leaders who convince employees of AI's benefits and address their concerns increase willingness to adopt AI. Ensuring alignment among employees and management further simplifies implementation. In summary, companies need a supportive regulatory environment, an open culture, and strong leadership to successfully implement AI.

(2) Technological Readiness
This dimension focuses on an organization's level of technological maturity, which includes the availability of necessary resources, a strong track record of using advanced technologies, and sufficient IT support, as successful adoption often relies on a well-established technological foundation. Emphasis is placed on the organization's historical experience and culture of working with technology, where AI solutions such as chatbots or virtual assistants are not viewed as standalone fixes, but rather as advanced tools that build upon an already mature digital environment. Moreover, the ability to process and manage data is essential. Organizations that are actively adopting AI typically already have these technical competencies. Lastly, sufficient and reliable IT support is crucial to avoid technical bottlenecks that could otherwise hinder the smooth implementation and operation of AI technologies.

(3) Informational Readiness
Distinct from Data Readiness, this dimension refers to "*people's meaningful understanding of a specific issue.*" In this context, it concerns the decision maker's knowledge of the relevant AI use case, how AI is applied within the industry, the specific problem at hand, and how AI can be used to address that problem. The focus lies specifically on the decision maker, rather than the broader organization or team. To make informed and strategic decisions, decision makers must have a deeper knowledge of AI's practical applications and potential. Furthermore, the decision to implement AI should be strategically significant, with the potential to substantially impact operations and increase organizational profitability. Since AI adoption is cost-intensive, it is essential that the decision maker has a clear understanding of the problem to be solved, as well as an awareness of the AI solutions available on the market in order to identify the most suitable option.

(4) Infrastructural Readiness
The availability and suitability of foundational resources are crucial for successful AI implementation. This dimension includes three main categories: human resources, financial resources, and IT resources. Financial resources are considered one of the

most critical and challenging components, as AI implementation is costly. It is not only about having access to sufficient funds, but also about ensuring that it is continuous and flexible to enable quick adaptation to changing market conditions. AI requires ongoing investment for updates and maintenance, given the rapid pace of technological advancement. In terms of human resources, organizations must ensure access to both internal and external talent with the necessary skills to support AI initiatives. This readiness can be developed by training existing employees and by recruiting experts who already have relevant knowledge. A key part of infrastructural readiness is also the ability to bridge the HR gap, not only by hiring individuals with strong technical skills, but also by seeking those with cross-industry domain knowledge who can contextualize AI applications within the organization's specific field. IT resources include the organization's technical infrastructure, such as computers, networks, and programming environments. Important capabilities here involve storage capacity, computing power, scalability, and security, all of which are necessary for AI to function effectively. The absence of the right IT infrastructure can significantly hinder a company's ability to adopt and benefit from AI technologies.

(5) Data Readiness
This dimension refers to the availability of large amouts, high-quality, and relevant data required to feed and support AI. It is important to distinguish data from information, as previously explained: data can be seen as raw, often meaningless symbols, while information is the result of organizing and interpreting data to create meaning, something that humans use to solve problems or make decisions. In the context of AI, data serves as the input that enables algorithms to function and learn, whereas information is primarily used and created by humans. The volume and quality of data are crucial for the performance of AI models.

(6) Participants' Readiness
The psychological and behavioural preparedness of individuals within and around the organization to adopt and work with AI. For employees, this readiness includes three key aspects: acceptance, trust, and knowledge and skills. In many organizations, staff are not yet sufficiently familiar with AI, making knowledge and training critical factors for successful adoption. A common barrier is the lack of trust, as some employees fear that AI may replace their jobs. This can create resistance and act as a bottleneck in the implementation process. Therefore, participants must not only be capable, but also willing to embrace change and see AI as a supportive tool rather than a threat. Managerial readiness is equally important. For AI to create value, it must be aligned with the organization's strategic goals, both in the short and long term. Managers play a crucial role in shaping attitudes toward AI and setting the direction for its integration. Finally, readiness among external stakeholders and partners is also essential, although it is more difficult to influence. Partners must be willing to work with AI-based systems and adapt their processes accordingly. Traditional mindsets and reluctance to change among partners can slow down or even block AI adoption, highlighting the need for alignment beyond the organization's boundaries.

(7) Customers' Readiness
This dimension refers to how well organizations are prepared to address customers' needs, privacy concerns, and acceptance of AI technologies. It is essential that companies have clear plans in place for managing potential issues related to AI use, in order to minimize risks, particularly in industries handling customer transactional data. In such cases, organizations must communicate transparently with customers to build trust and avoid misunderstandings. While customers may not require information about AI used internally, their acceptance becomes more critical when AI is used in customer-facing touchpoints or involves the use of customer data.

(8) Process' Readiness
The last dimension includes three key components: operational integration, feedback mechanisms, and integrated communication. To fully leverage the value of AI, operational integration must be in place, meaning that different teams and functions within the organization collaborate effectively. This cross-functional integration is essential for spreading the benefits of AI across departments and ensuring consistent outcomes. A strong feedback mechanism is also critical. Continuous feedback allows AI systems to improve their performance and better adapt to the specific needs of teams or the organization as a whole. Since these needs may evolve over time, a constant feedback loop helps maintain the relevance and effectiveness of AI solutions. Finally, integrated communication plays a vital role in preventing system failures caused by miscommunication. Clear and consistent communication across the organization supports smoother implementation and operation of AI technologies.

2.2.3 AI Adoption Strategies

To successfully adopt AI, organizations must define a clear strategy, as the absence of one is a major barrier to implementation. McKinsey research underscores that lacking a defined AI strategy is among the most significant challenges faced by managers [16]. Ultimately, the value of AI lies not in the technology itself, but in how effectively it is integrated into organizational processes.

Building on this, Tehrani et al. [3] identified AI adoption strategies that organizations can apply individually, or in combination, depending on their context and readiness profile, shown in Figure 2.5. Each strategy aligns with specific AI readiness dimensions, meaning that the strength or weakness of certain factors can shape which approach is most appropriate to use.
In addition to internal capabilities, strategy selection also depends on external factors: (1) whether the organization's main goal is cost efficiency or differentiation, and (2) the perceived level of risk in AI adoption. Cost-driven firms might prioritize partnerships, while differentiation-focused firms may prefer crawling or guinea pig approaches to test new innovations. In this study, four of the five strategies will be presented.

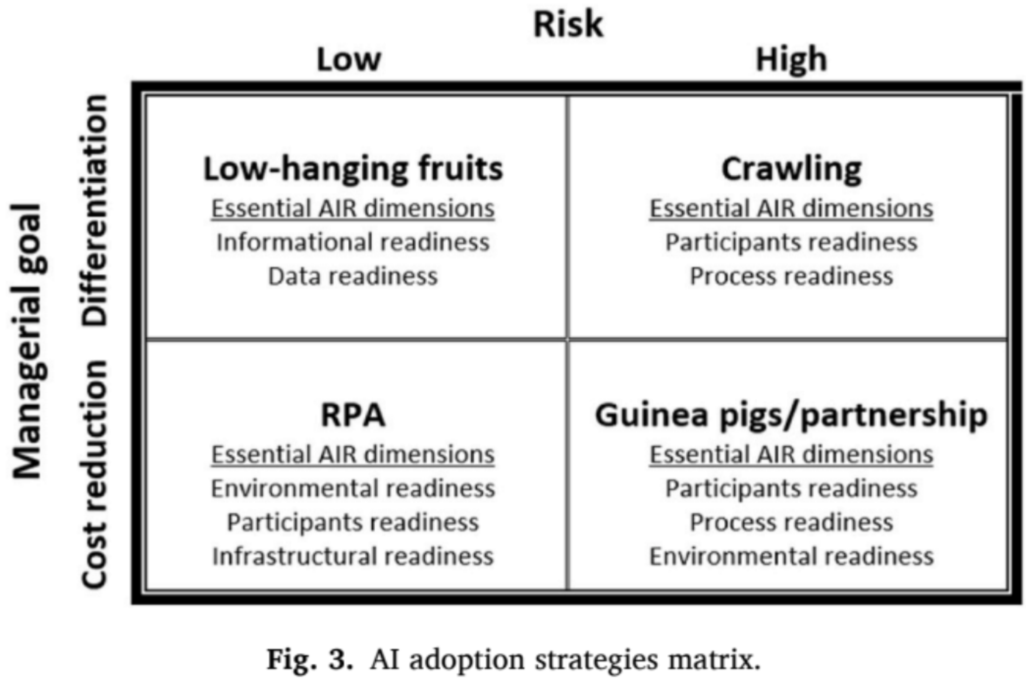


Figure 2.5: Adoption Strategies [3].

Adoption Strategies

- The Low-Hanging Fruit Strategy: This is a practical starting point for firms with strong data and informational readiness but a reluctance to take on significant risk. It involves identifying straightforward use cases, such as reporting or marketing automation, that can be implemented quickly and deliver early wins, helping build momentum for further adoption.
- The Crawling Strategy: This strategy focuses on gradual, iterative AI adoption. Organizations begin with smaller-scale pilots and expand based on lessons learned. This approach is best suited to organizations with limited financial flexibility but strong willingness to experiment and learn, requiring strong process and participant readiness.
- The Guinea Pig Approach: This approach is good for larger firms to learn indirectly by observing or partnering with smaller, more agile actors who experiment with AI. This enables risk transfer while still gaining insights. The approach is especially relevant when internal readiness is moderate, but there is willingness to innovate.
- The Partnership Strategy: Lastly, this strategy focuses on engaging external AI consultants or technology partners to compensate for limited internal capabilities. These partnerships provide access to both infrastructure and expertise, while supporting shared capability development. Environmental readiness and trust in external consultant are essential enablers.

2.3 Prototype

The following section covers relevant theory connected to the implementation of the prototype with the purpose of explaining, motivate and demonstrate the usages of different methods and models. As the foundation of the prototype is AI and machine learning, the framework naturally consists of AI in general, and text-processing models in specific. Hence, first described is overall theory followed by an introduction of models and algorithms used.

2.3.1 Artificial Intelligence

Artificial Intelligence (AI) is a broad concept, spanning multiple fields such as machine learning (ML), neural nets (NN) and deep learning (DL), see Fig. 2.6. The umbrella term refers to the section of computer science where machines and computers are built to be able to perform complex tasks, mimicking the human intelligence and ability to reason, predict and analyse. Not a system alone, AI is rather implemented in machines or systems unlocking their potential and sense of intelligence [17].

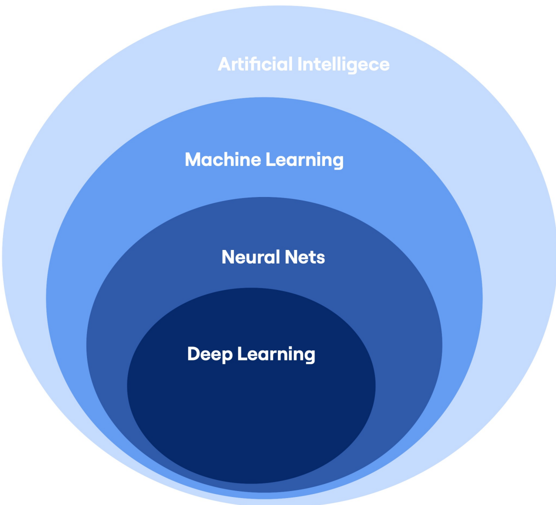


Figure 2.6: Overview of models implemented.

Machine Learning: ML is a subset of AI that enables systems to automatically learn from data and improve their performance without being explicitly programmed. Instead of following hardcoded rules, ML algorithms identify patterns and relationships in data to make predictions or decisions. ML models can be trained under various learning paradigms depending on the availability and quality of labeled data. While supervised and unsupervised learning remain foundational approaches, recent developments in weakly supervised learning have introduced scalable solutions for real-world data constraints:

Supervised Learning: Supervised learning involves training a model using fully labelled data, where both the input and corresponding output (or ground truth) are known. This approach enables models to learn direct mappings from inputs to outputs, and is widely used in classification and regression tasks. However, supervised learning relies heavily on high-quality, annotated data which is a resource that is often costly and labour-intensive to produce at scale. In this paradigm, common techniques include decision trees, support vector machines, and regression models such as linear or logistic regression.

Unsupervised Learning: In contrast, unsupervised learning is performed without labelled data. The model is presented only with input features and must uncover patterns or structures autonomously. Common applications include clustering, where data points are grouped based on similarity, e.g., using K-Means Clustering (KMeans), and association analysis, where relationships between variables are discovered without human supervision. This makes unsupervised learning well-suited for exploratory analysis, anomaly detection, and image segmentation, among other tasks.

Semi-supervised/Weakly supervised Learning: Weakly supervised learning refers to a spectrum of learning strategies that leverage noisy, incomplete, or imprecise labelling to train predictive models. This paradigm addresses the challenge of limited or imperfect supervision which is a common scenario in real-world applications where manual labelling is expensive or impractical. Three primary forms of weak supervision are identified:

- Incomplete Supervision: Training on datasets where only a portion of the data is labelled. Techniques like semi-supervised learning, transfer learning, and active learning fall under this category.
- Inexact Supervision: Using coarse-grained labels, such as class-level annotations instead of instance-level, which may lack detailed specificity but still provide learning signals.
- Inaccurate Supervision: Occurs when labels contain errors or noise, often due to human mistakes or automatic labeling heuristics. Models must then learn to tolerate or correct for mislabeled examples [18].

Zero-shot and Few-shot Learning: Zero-shot learning (ZSL) is a supervised machine learning paradigm in which a model is expected to correctly classify data from previously unseen classes, that is, categories that were not represented in the labelled training data. Unlike traditional supervised learning, which relies on annotated examples for each class, zero-shot learning requires the model to generalize to entirely new concepts without direct exposure during training. This is particularly important in real-world scenarios where obtaining labelled examples is impractical, expensive, or impossible. Examples include rare disease detection, emerging topics in text classification, or recognizing obscure object categories in vision tasks. For instance, while humans can distinguish tens of thousands of object categories, it is infeasible to provide labelled data for every possible class a model might encounter. To achieve generalization without labelled examples, ZSL methods typically rely on auxiliary information, such as textual descriptions of the target classes and semantic embeddings or class attributes [19].

Transfer Learning: Transfer learning or domain adaptation, has the purpose of using a trained model for a new task, instead of training it from scratch. Common in ZSL, transfer learning is often used in methods that focuses on semantic embeddings. An example would be to use Bidirectional Encoder Representations from Transformers (BERT), which is pre-trained on language data, to convert newly seen words into vector embeddings. With transfer learning it is also possible to by recognizing one type of text or image, simultaneously identify other unseen ones[**?**].

Neural Nets and Deep Learning: NNs are a type of ML algorithm inspired by the human brain's structure, composed of layers of interconnected nodes (neurons). These networks are particularly effective at learning complex, non-linear relationships in data.

DL is a specialized form of neural networks with many hidden layers. It excels in handling unstructured data such as text, audio, and images. DL is the foundation of many recent breakthroughs in AI, including speech recognition, image generation, and language models like BERT and Generative Pre-trained Transformer (GPT) [20].

Embeddings as a Foundation for Language Understanding
Embeddings are a method of representing objects, such as words, sentences, images, or audio, as vectors in a continuous numerical space. These representations are designed so that semantically similar inputs are located close together in this space. In the context of NLP, embeddings allow machine learning models to capture contextual and semantic relationships between words and texts.

Unlike manual feature engineering, embeddings are learned directly from data using neural networks or other algorithms. This allows the model to identify complex patterns in language that are not easily defined by rules or discrete categories. Embeddings enable tasks such as text classification, clustering, and semantic search by converting raw text into a format that machine learning models can process effectively [21].

2.3.2 Language Models

LMs are probabilistic models that assign likelihoods to sequences of words. Rather than assessing grammatical correctness, they measure how "natural" a word sequence is based on patterns learned from real-world text. This capability enables a wide range of natural language processing tasks, such as part-of-speech tagging, lemmatization, summarization, translation, and question answering. There are two main categories of language models:

Neural Network-based Models

NN use word embeddings to represent words as vectors and capture semantic meaning. Early models like Word2Vec improved representation, but struggled with deeper context. RNNs introduced the ability to handle sequential data by maintaining memory over time, allowing them to process inputs like text and speech. However, RNNs are limited by their strictly sequential nature, which slows down training on longer sequences and makes parallelization difficult. Transformers addressed these limitations by enabling parallel processing and introducing attention mechanisms that weigh the importance of each word relative to others in a sentence. This architecture powers state-of-the-art models like BERT and GPT, which are pre-trained on large-scale text corpora and can perform a wide range of tasks by leveraging deep contextual understanding and large parameter capacities [22].

Natural Language Processing
NLP refers to the field of artificial intelligence focused on enabling machines to understand, interpret, and generate human language. Since its origins in the 1950s, NLP has evolved into a suite of algorithms and tools that support tasks such as part-of-speech tagging (POS), sentiment analysis, named entity recognition, parsing, and machine translation. NLP powers a wide range of applications including speech recognition, text classification, customer service chatbots, and content recommendation systems.

At its core, NLP applies structured linguistic rules and statistical techniques to textual data. While effective in specific tasks, traditional NLP models often require domain-specific tuning and can struggle with language ambiguity, contextual subtleties, and low-resource languages.

Large Language Models
Large language models, such as OpenAI's GPT and Google's BERT, represent a major advancement in language understanding and generation. Built on transformer architecture and trained on massive text corpora using deep learning, LLMs can generate coherent, contextually relevant text, answer questions, summarize content and more, often with little or no task-specific fine-tuning.

LLMs differ from traditional NLP systems in scale and generality. They are not limited to specific rule-based tasks but can perform a wide range of language tasks through learned contextual understanding. Core technologies behind LLMs include self-attention mechanisms, DNN, and massive parallel training, which allow them to adapt flexibly to diverse domains [23].

2.3.3 Models in Selection

The AI prototype is built on algorithms and pre-trained models which will be covered in this section. An overview can be seen in 2.1.

MEGClass: MEGClass (Mutually Enhancing Granularities for Text Classifica-

Component	Type	Function
MEGClass	Weakly Supervised Classifier	Combines clustering, embeddings, and pseudo-labeling
BERT	Monolingual LLM	Contextual language understanding in English
XLM-R	Multilingual LLM	Cross-lingual language understanding
Sentence-BERT	Sentence Embedding Model	Generate semantically meaningful sentence vectors
MiniLM	Lightweight Multilingual Model	Efficient multilingual sentence encoding
KB Swedish SBERT	Swedish Sentence Embedding Model	Embed Swedish sentences for similarity tasks
spaCy	NLP Processing Toolkit	Tokenization, lemmatization, POS-tagging
KMeans Clustering	Clustering Algorithm	Group similar sentence embeddings

Table 2.1: Overview of models, tools, and algorithms used in the NLP classification pipeline.

tion), is a state-of-the-art method for extremely weakly supervised text classification, requiring only the surface names of target classes to operate without any labeled data. The model was designed to overcome limitations in prior approaches that typically treat word-, sentence-, and document-level information independently, which can lead to incorrect pseudo-labels when topic cues are ambiguous or inconsistent across granularity levels.

MEGClass introduces a multi-granular approach, where words, sentences, and document representations are allowed to mutually enhance each other. This results in a more robust and context-aware estimation of class labels, even in challenging real-world texts. Where core innovations of MEGClass include:

- Class-Oriented Sentence Representations: The model computes class-indicative sentence embeddings by aligning sentences with class name representations, emphasizing discriminative terms.
- Class Distribution Estimation: Rather than assigning each document a single label, MEGClass estimates a class probability distribution, allowing it to gauge classification confidence and reduce mislabelling.
- Contextualized Document Representations: Through a multi-head self-attention network, the model creates enriched document vectors that capture hierarchical context from sentence-level signals.
- Iterative Feedback Mechanism: Confidently classified documents are used to refine class representations iteratively. This improves class alignment across documents and reduces error propagation.
- Pseudo-Labelling and Classifier Fine-Tuning: A subset of the most confidently classified documents is used as pseudo-labelled data to fine-tune a downstream classifier, making the model applicable to unseen examples.

MEGClass has demonstrated superior performance across several benchmark datasets, particularly in scenarios with long documents and fine-grained classes. Its effectiveness lies in its ability to leverage minimal supervision while still generating high-quality pseudo-training sets, outperforming earlier methods like Learning with Out-of-the-box Classifier for Text Classification (LOTClass) and Explainable Classifier for Weakly Supervised Text Classification (X-Class) [24].

K-Means Clustering: KMeans clustering is one of the most widely used unsupervised learning algorithms in machine learning. It operates on unlabeled data, partitioning it into k distinct, non-overlapping clusters, where each data point is assigned exclusively to the cluster with the nearest centroid. This method assumes no prior knowledge about the data labels and aims to uncover inherent structure based on similarity. The core idea behind KMeans is to minimize the intra-cluster variance, or more precisely, the sum of squared Euclidean distances between each data point and its assigned cluster centroid [25].

BERT and Sentence-BERT: One of the most significant advancements in (NLP) is the introduction of BERT, developed by Devlin et al. (2018). BERT is a transformer-based model pre-trained on large text corpora, designed to understand the context of words in a bidirectional manner. It has achieved great results in a wide range of NLP tasks, such as question answering, sentence classification, and semantic textual similarity (STS).

However, BERT is inherently designed as a cross-encoder, meaning that for sentence pair tasks, both sentences are jointly input into the model. While this joint attention mechanism increases performance for tasks requiring fine-grained comparisons, it introduces a significant computational bottleneck. For example, comparing 10,000 sentences to each other using BERT requires approximately 50 million inference computations, which can take over 65 hours on a high-performance GPU. This makes BERT unsuitable for large-scale tasks such as clustering, semantic search, or retrieval [26].

To address this limitation, Sentence-BERT (SBERT) was proposed by Reimers and Gurevych (2019). SBERT modifies the BERT architecture by applying a siamese or triplet network structure, allowing it to generate fixed-size sentence embeddings. Instead of comparing sentences within the model during inference, SBERT maps each sentence to a vector space such that semantically similar sentences are close together. These embeddings can then be efficiently compared using standard similarity metrics such as cosine similarity.

The key innovation lies in the training objective. SBERT is fine-tuned on Natural Language Inference (NLI) datasets, which teach the model to distinguish between similar, contradictory, and neutral sentence pairs. During inference, the model independently encodes each sentence, which enables tasks like clustering, semantic search, and zero-shot classification to be performed orders of magnitude faster than with BERT, while still maintaining high accuracy. Moreover, SBERT introduces a pooling strategy (typically mean pooling) over the output token embeddings to produce the final sentence vector. This differs from BERT's default use of the [CLS] token, which has been shown to produce suboptimal results for sentence-level representations [27].

Two SBERT models are used in the prototype, to generate vector embeddings of Swedish texts. `KB Swedish sentence-BERT` is a bilingual Swedish-English sentence embedding model developed by the National Library of Sweden (KB-Lab). It uses

KB-Bert, a Swedish BERT model as base-encoder, and all-mpnet-base-v2 as a English teacher model [28]. The other model `paraphrase-multilingual-MiniLM-L12-v2` model is a lightweight and multilingual SBERT model published by the Sentence-Transformers team. It is trained on paraphrase data from multiple languages, including Swedish, and provides sentence embeddings of 384 dimensions. Due to its small size and high speed, it is especially suited for resource-efficient applications [29].

XLM-R: Cross-lingual Language Model - RoBERTa (XLM-R) is a multilingual transformer model developed by META to improve cross-lingual natural language understanding. The model is trained using self-supervised learning techniques and addresses the challenge of transferring knowledge between languages without requiring additional task-specific data in the target language. It builds upon earlier models like XLM and multilingual BERT but overcomes key limitations by incorporating a substantially larger and more diverse training dataset, over two terabytes of filtered data, covering a broader range of languages, including many low-resource ones that previously lacked large-scale labelled or unlabelled corpora[30].

SpaCy: `SpaCy` is an open-source Python library, providing advanced and efficient NLP on text, and offering features such as tokenization and POSTagging. Unlike BERT, `SpaCy` is older and though operating at higher speed, less capable of capturing contextual information [31].

2.3.4 Data Criteria

In AI development, especially in natural language processing and weakly supervised learning, the quality of input data is critical. Several criteria are used to evaluate whether data is suitable for training in machine learning contexts. These criteria ensure that models are able to learn effectively, avoid systematic errors, and generalize to new inputs. The following criteria were used to evaluate the data in terms of its structure.

Completeness: Data completeness refers to if a dataset includes all the information necessary to address its intended purpose. It evaluates if the data sufficiently covers the full scope of a given question, if there are any gaps, or biases, that could distort results. Incomplete datasets can cause inaccurate analyses, including misreported metrics and biased decisions, which can undermine an organization's confidence in data-driven insights [32].

Data Volume: Data volume refers to the amount of information available to train AI models, and it is especially critical in natural language processing and weakly supervised learning. Large datasets enable models to detect complex patterns and structures. AI training requires large volumes of both structured and unstructured data to meet increasing model complexity. For weakly supervised approaches, unlabelled data at scale is essential to compensate for the lack of explicit annotations. Without sufficient volume, models risk underperformance and poor adaptability to real-world tasks [33].

Relevance Relevance refers to the degree to which data contributes directly to the specific objective or task the model is designed to solve. Relevant data provides meaningful context that enhances the model's ability to identify important patterns, make accurate predictions, and avoid distraction from noise or unrelated variables. When irrelevant information is included, it can dilute learning signals and reduce overall model performance [34].

Consistency: Consistency ensures that data maintains a uniform format, structure, and meaning across sources and time. For AI, this is essential to enable accurate parsing and interpretation. Inconsistent data, such as varying labels or formats, can confuse models, lead to errors, and reduce the reliability of outputs [35].

Noise: Data noise refers to irrelevant or incorrect data that can confuse AI models and reduce their accuracy. It includes errors, outliers, and unnecessary information that do not help the model's task. Effective data validation techniques are needed to detect and remove noise, ensuring cleaner data and better AI performance [36].

Bias: Bias in AI occurs when models produce unfair results due to prejudices in training data or design choices. Biased data, such as underrepresentation of certain groups, can cause inaccuracies and unreliable predictions. To reduce bias, it's essential to use diverse data, transparent algorithms, and ongoing evaluation to promote fairness and equity [37].

Representativeness: Representativeness in datasets for AI models refers to how comprehensively the data captures the diversity and variability present in real-world scenarios. A representative dataset should reflect the full spectrum of possible inputs, covering varied contexts and distributions, to ensure robust and reliable generalization. Without representativeness, models risk developing biases or overfitting, resulting in poor performance when faced with new or diverse inputs. Hence, carefully ensuring representativeness is crucial to enhancing the accuracy, fairness, and effectiveness of AI systems [38].

3

Methods

This section outlines the research process, design, and methods used in the study to ensure transparency, repeatability, and a clear understanding. The chosen method is designed to systematically address the research questions and support the development of a solution driven prototype.

In addition, this section presents the overarching research framework, the data collection strategies employed (including interviews, workshops, and document analysis), and the techniques used for data interpretation. A structured approach was adopted to ensure that the findings are robust, relevant, and reproducible.

Finally, the study includes a discussion of research quality considerations, including validity, reliability, and ethical aspects such as data privacy, informed consent, and responsible data handling. These aspects were carefully managed to ensure the integrity and credibility of the research.

3.1 Research Process

The process began by identifying and defining the overall research focus. This led to the formulation of RQ 1: *"What is Skanska's current state of AI readiness, and what organisational strengths and challenges exist in terms of AI adoption?"*. Through an open-ended investigation, a specific area within the company was identified where AI could create tangible value, forming the basis of RQ 2: *"How can an AI-prototype create value within the department of Quality Management?"*. Building on this, RQ 3: *"How can a language model prototype be developed and used to efficiently extract and visualize insights from quality-related construction texts?"* was formed accordingly, and is discussed in detail in Section 4.

These activities were carried out in parallel: organizational insights were qualitatively collected and analysed using the AI Readiness Framework, while the prototype was iteratively developed to meet domain-specific needs. Given the complexity of AI adoption in the construction sector, an exploratory approach was essential to uncover relevant challenges, opportunities, and industry-specific conditions. The timeline and phases of the study are presented in Figure 3.1.

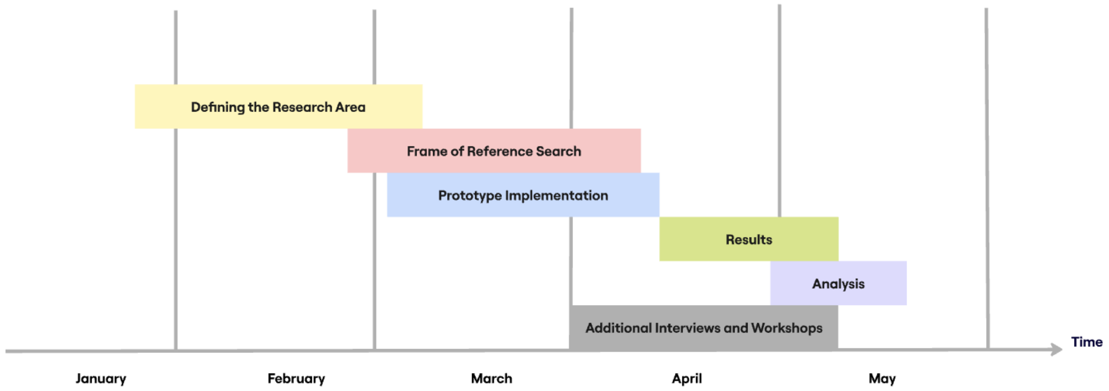


Figure 3.1: Gantt chart illustrating the research process timeline.

3.2 Research Design

The study initially adopts an inductive reasoning approach, where broader conclusions are derived from specific observations rather than tested against a predefined hypothesis. This method is particularly valuable in exploratory research, as it allows for the development of new insights and conceptual frameworks grounded in empirical data.

At the same time, the study incorporates existing theoretical models, like the AI Readiness Framework, to guide the interpretation of qualitative findings from interviews and workshops. Incorporating established theory adds a deductive dimension to the study, as parts of the data collection are guided by predefined theoretical constructs.

Overall, the research follows an abductive reasoning process, characterized by iterative movement between empirical data and theoretical insights. This approach allows for ongoing refinement of both theoretical understanding and practical implementation, which makes it particularly suitable for studying complex issues like AI adoption in the construction industry [39].

3.3 Qualitative Methods Used

The study began with an exploratory phase that included an initial stakeholder workshop, followed by several semi-structured qualitative interviews. These qualitative insights helped to refine the research focus and informed both the structured development of the AI prototype and the analysis of AI readiness. The frame of reference was developed alongside the interviews, prototype, and workshops. It had two main goals: to support the study's methodology and to provide a theoretical foundation for the analysis and discussion.

Finally, additional workshops and structured interviews were conducted to gather feedback and revisit insights from the initial session, and to triangulate findings.

3.3.1 Workshops

Two separate qualitative workshops were conducted, to brainstorm ideas, gather feedback on the process, as well as collecting information on the AI in the organization.

1. Idea Generating Workshop

The initial workshop was inspired by the AIM-workshop methodology, first introduced by Shiba, Shoji in 1987, which is an approach to identify and analyze complicated or complex problems in a collaborative and structured way. The process typically begins with clarifying shared aspirations or challenges, followed by creative brainstorming to generate potential solutions. This methodology is particularly valuable in exploratory projects where diverse perspectives are essential for defining needs, generating solutions, and building stakeholder commitment [40].

The workshop was conducted with five stakeholders from Skanska Hus Gothenburg, working in different stages of the construction process (*byggprocessen*). This workshop followed an exploratory focus group approach, where the participants engaged in guided discussions to explore attitudes, ideas, and expectations related to AI in construction. Focus groups are useful for capturing diverse perspectives and encouraging interaction, allowing participants to reflect, react, and build on each other's insights [39]. This dynamic helped generate practical input for both the prototype development and the AI readiness analysis.

The purpose was to gain a deeper understanding of stakeholder workflows, challenges, and needs of the employees, to guide the final identification of an AI use case. The first part of the workshop was structured around open-ended questions to identify inefficiencies, time-consuming tasks, and key pain points in the participants' daily work. The second part of the workshop focused on data usage, examining the types of data used, stakeholders' data literacy, and their perspectives on data-driven decision-making. Finally, participants were given space to freely generate ideas and reflect on the overall purpose of the study. To ensure unbiased insights, no predefined concepts or solutions about AI were introduced to the participats, allowing them to express their views without any bias towards AI use. Finally, the concept of AI was introduced to the participants, to further collect ideas about use cases. The reason AI was mentioned later in the workshop was not to make the participants think about solutions, but instead about problems to solve. Insights from the first workshop were further used to guide the interviews, guide the prototype development, and shape the frame of reference.

2. Feedback Workshop

A second workshop was conducted with the purpose of gathering feedback on the prototype, ensuring it aligned with stakeholder needs and addressed previously identified challenges. Participants were introduced to the prototype findings and encouraged to reflect on its usability, functionality, and integration into their existing workflows. The perceived value of the prototype was also assessed, with discussions centered around how the prototype could contribute to value creation within the organization and support ongoing quality management efforts.

As part of the workshop preparation, participants received a pre-workshop task which involved reviewing a file containing 10–20 clusters of deviations. These clusters had been grouped based on titles and descriptions. Participants were asked to reflect on potential patterns across the clusters, name each category, and identify the most cost-driving deviations based on their professional experience. They were also encouraged to think freely about alternative ways to categorize deviations, for instance, in terms of risk reduction, cost savings, or recurrence prevention. In addition, participants were asked to consider which descriptive elements (e.g., actions, damage types, locations) were most relevant for analysis. This preparatory activity served to engage participants ahead of the workshop and ensured that the feedback session was rooted in both practical experience and reflective input, ultimately enhancing the relevance and depth of the discussions.

The workshop insights were then prioritized to identify the most impactful modifications, forming the basis for final prototype adjustments. This iterative approach strengthened stakeholder engagement and ensured that the solution was validated before further development. The final prototype was built to contribute value by addressing these identified priorities.

3.3.2 Interviews

The study applied a qualitative interview approach, where the initial interviews in the explorative phase were positioned between semi-structured and unstructured formats. An initial interview guide was used to loosely cover two main areas: (1) the organization's perceived value of AI and readiness, and (2) the feasibility of developing a prototype in their context. However, the guide served more as a flexible support than a strict script. Interviewees were encouraged to speak freely, and the interviewer followed up on relevant points as they emerged. This conversational and open style allowed participants to highlight what they viewed as most important.

Following the first workshop, two participants were interviewed, and they recommended additional interviewees through snowball sampling, which is a method in which initial participants assist in identifying and recruiting additional relevant interview subjects [39]. This approach made it possible to identify additional relevant stakeholders within the Skanska organization, beyond the Skanska Hus Gothenburg department. This helped shape both the prototype and the analysis of AI readiness. In total, 13 individuals in the Swedish organization were interviewed using the this phase. Respondents and their roles can be found in Appendix A. Some of them were interviewed more than once to gather further information. The interviews were conducted primarily in person; if the interviewee was not available onsite, the interview was held digitally via Microsoft Teams. Interviews were manually transcribed to capture immediate insights and nuances. In some cases, transcription was supported by AI software, in which case permission was obtained from the interviewee.

The first phase of interviews was thus more exploratory, defining the scope of the study, highlighting contextual nuances, and identifying patterns related to AI readiness. These interviews also informed the design and development of the AI prototype, ensuring it addressed stakeholder needs and priorities.

In addition, a second round of six semi-structured interviews were conducted to validate and triangulate the initial interview and workshop findings. Triangulation involves using multiple data sources, methods, or researchers to increase the trustworthiness and credibility of research findings [39]. These interviews consisted of two professionals from the US organization at the company's New York office, one individual from the UK office, and three additional interviews with Swedish employees, two of whom had also participated in the initial round of interviews. While the study primarily focuses on a Swedish context, the international perspective exceeds the scope but contributes to triangulating findings. The purpose of the international interviews was to understand how AI is applied in practice within an international context at Skanska, and to compare our own AI model against their workflows.

These additional interviews followed a structured interview guide, which can be found in Appendix B. The approach was grounded in the eight AI Readiness dimensions proposed by Tehrani et al. [3]. This strengthened the empirical foundation for answering Research Question 1 (RQ1) and enabled a critical assessment of the prototype's alignment with strategic objectives.

3.3.3 Litterature Search

To construct the frame of reference, the scientific database Scopus was used to identify peer-reviewed and high-quality literature relevant to the research topic. Systematic and well-documented literature searches are essential for ensuring transparency, reliability, and replicability in academic research [39]. Therefore, a structured approach was applied when selecting keywords, which were derived from key concepts in the research questions and refined iteratively during the process. Common Boolean operators (e.g., AND, OR) were used to combine terms and narrow down the results.

To assess the relevance and quality of the sources, search results were filtered based on citation count and publication outlet, with a focus on highly cited articles published in reputable journals. In areas related to artificial intelligence, which is a fast evolving field, particular emphasis was placed on recently published articles to ensure that the theoretical foundation reflects the latest developments and current state of the field with constant updates in the area.

Overall, litterature search covered the areas of the construcion indsutry, relevant businesss frameworks. The frame of reference also covers relevant aspects of AI, including specific models, to later reinforce their applicability in the study's context. This analysis of existing research, also known as secondary studies, refers to examining and interpreting data or findings originally collected by others [39].

3.3.4 Qualitative Data Analysis

The interview data were analyzed thematically, using the AI Readiness framework as a guiding structure. Thematic coding is a qualitative analysis method used to identify and organize patterns or themes within interview data [39]. Each transcript was reviewed, and segments were divided into themes according to the eight dimensions of AI readiness. Within each of these eight initial themes, approximately five key codes were identified, each capturing an essential aspect of the theme.

Following this first categorization, a second layer of analysis was performed by connecting these codes to five overarching second-layer themes: (1) Organizational Readiness Gaps and Ownership Challenges, (2) Employee Trust and Safety, (3) Technical and Data Foundation Gaps, and (4) Early Signs of Adoption and Competitive Opportunity. Finally, these second-layer themes lay the basis for the discussion chapter, and captures frequently mentioned insights from the interviewees. Altogether, this analysis forms the basis for answering RQ1 and supports the formulation of managerial implications for Skanska.

3.4 Defining the Research Area

Following a series of workshops, interviews, and an initial exploration of both the construction industry and Skanska Hus's operational practices, potential areas of interest for AI application were identified. These areas were evaluated based on technical feasibility, data availability, and potential business value.

3.4.1 Quality Management in the Construction Process as the Chosen Area

The area, *Quality Management in the Construction Process*, was selected following a series of workshops and careful evaluation (partly in accordance with Section 2.3.4). It was identified as a domain characterized by a continuous and centralized inflow of data, rich in metadata and largely untapped textual content. This combination offered strong technical feasibility NLP and high business value through the potential to extract deeper insights and conclusions from the data.

Employees across all stages of the construction process interact with Skanska's ACC system, where they log deviations (*avvikelser*). A deviation refers to any activity or outcome that fails to meet specified requirements, thereby affecting work quality,

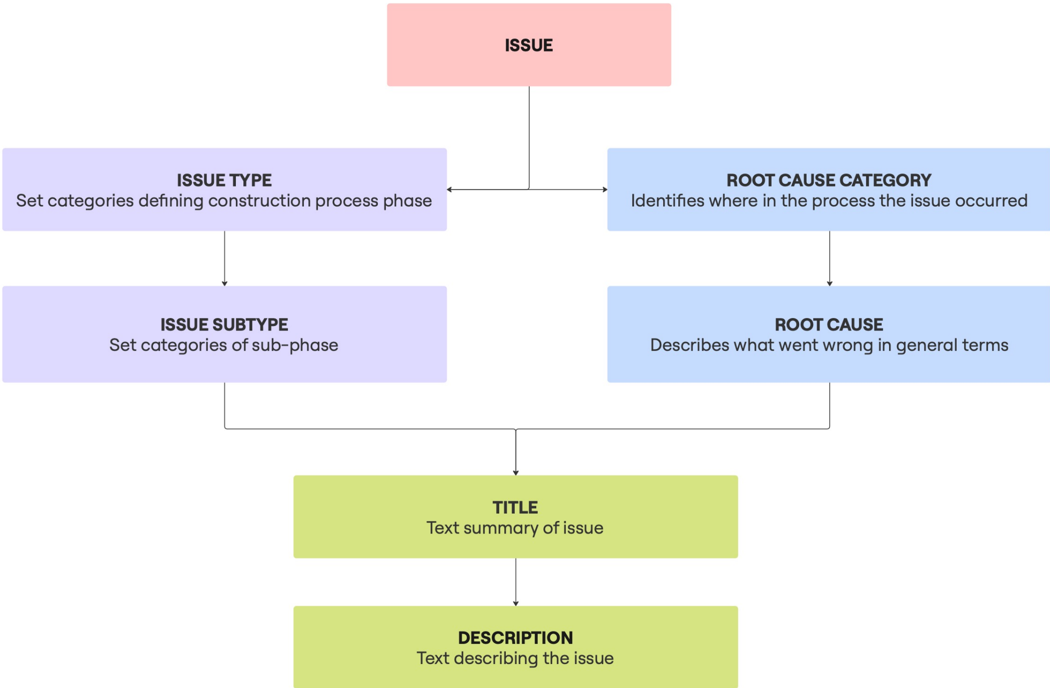


Figure 3.2: Overview of the currently reporting structure, representing each reported issue and focusing on the construction process phases, e.g. inspection and production.

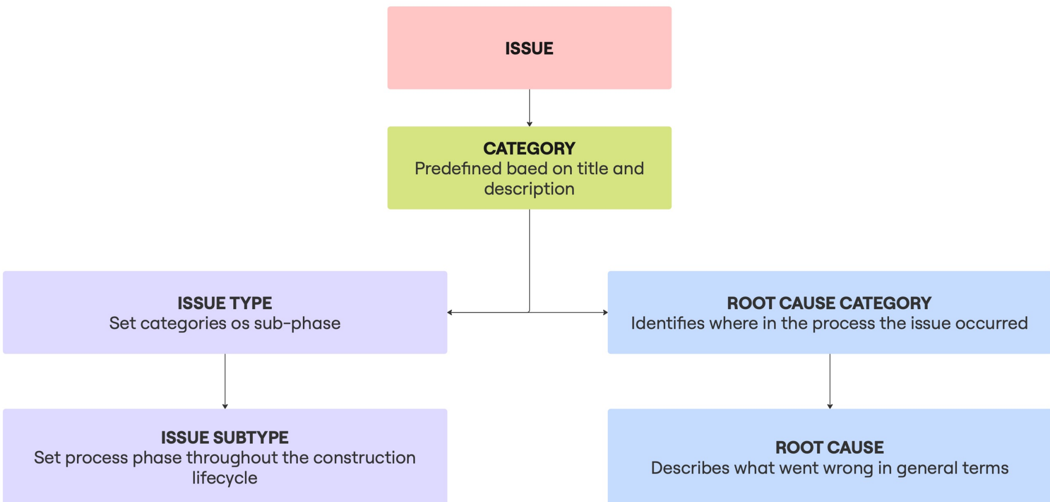


Figure 3.3: Overview of the desired reporting structure, representing each reported issue and focusing on the category and nature of the issue, e.g. construction part and action.

the final product, or the surrounding environment. Managing these deviations is essential for organizational learning, continuous improvement, and compliance with both contractual obligations and external standards. Notably, deviations are most commonly reported during the production and aftermarket phases, rather than during inspection or product development. Historically, such issues were documented in the BIM360 system, whose historical data remains accessible. Today, deviations are registered according to a construction process hierarchy (see Figure 3.2). There is a recognized need for improved oversight and timely handling of deviations to prevent costly consequences in later project phases.

The ACC/BIM360 platforms collect and store quality-related deviations and issues using predefined categories and root cause classifications tailored to the different stages of the building process. This system design helps prevent users from improvising their own taxonomies and reinforces the understanding that documentation serves a broader analytical purpose. Users are also prompted to provide a title and a description summarizing the issue. As of May 2024, Skanska introduced a new set of predefined categories within ACC, enabling better tracking, sorting, and analysis of deviations. This update enhances consistency in documentation and supports systematic analysis to detect patterns, facilitate learning, and improve quality performance over time.

Given the volume and structure of the existing data, there is a significant opportunity to improve data utilization in order to support the Quality Department's proactive quality management. This study therefore aimed to demonstrate how AI, specifically through NLP, can extract insights and enhance the categorization of issues. By restructuring issue reporting to emphasize what construction part was affected or what type of incident occurred, the organization can foster more proactive strategies and reduce both the time and cost associated with recurring deviations. This revised reporting logic is illustrated in Figure 3.3, and showcased in Section 4.

3.4.2 Other Areas of Interest

During the exploratory phase multiple areas of possible interest were introduced. These, however, were not deemed fit to use as focus scope according to their strategic and integrated business complexity, as well as the data criteria found in Section 2.3.4. The following section presents considered areas.

Risk assessment in the Cost Estimation of Construction Projects
The second area of interest focuses on cost calculations, including an added risk factor, which must be made for each project request. This risk assessment is often based on underlying data, supplier prices, but primarily on experience-based knowledge and team discussions. Since each project is unique, there is no exact method for this process, it relies heavily on experience based knowledge. Cost estimates vary from person to person and across different regions. Ultimately, many projects end up costing significantly less than initially estimated, meaning that costs are often overestimated due to higher-than-necessary risk additions. While there is a large

amount of available data, there is no structured approach to conducting an analysis. There is a need for a tool to support this process, but it remains a complex area, as it involves many individuals and their experience-based assessments. Therefore, this area was not considered further.

The Process of Reviewing and Verifying Accuracy in Project Planning
The third area of interest relates to the validation process throughout the construction value chain. From initial architectural designs, through technical planning and engineering, to the execution phase on site. A recurring issue identified is the lack of reliable, complete, and up-to-date project information being transferred between actors. This leads to time-consuming double-checking activities, such as manual cross-verification of drawings, bills of quantities, and technical specifications, to ensure no critical information has been omitted or misinterpreted. These inefficiencies stem largely from communication breakdowns and fragmented project documentation practices, which are common challenges in construction project management. Although addressing this area would offer significant value, it was deemed too complex for the current study due to the heterogeneous nature of the data (including both text-based and image-based documents) and the difficulty of standardizing such unstructured information streams.

Samläsning
The third area, Samläsning, is loosely linked to the previous area, it refers to the process of harmonizing terminologies and ensuring consistent communication across project participants. Differences in how design elements, construction materials, or processes are labelled and described create misunderstandings that complicate project execution and quality assurance efforts. This semantic misalignment often leads to errors, rework, and delays. Although this area is highly relevant and ties into broader issues of digitalization and it was ultimately excluded from this study. The decision was based on the complexity of the communication challenges involved and the broad organizational changes required to address them systematically.

3.5 Ensuring High-Quality Research

To maintain transparency and reflect on potential limitations, the following methodological considerations were acknowledged.

- Positive bias in participants: Most interviewees are highly open to change, innovation, and AI, which may skew results toward optimistic views and underrepresent resistance or concerns.
- Limited stakeholder diversity: Due to time and access constraints, some relevant roles and departments were not included, potentially narrowing the scope of insights.
- Triangulation: Qualitative data triangulation was done by combining insights from semi-structured interviews, workshops, expert perspectives from international colleagues, and a literature review.
- Researcher bias: Our academic and professional backgrounds may have in-

fluenced how we framed questions and interpreted data, possibly reinforcing preconceptions.

- Context-specific findings: The research is based on a specific organizational context, which may limit the generalizability of the results to other companies or industries.
- Sampling bias through snowball method: Snowball sampling may have limited the diversity of perspectives, as participants often recommended like-minded colleagues, potentially biasing the sample toward individuals with similar views or experiences.
- Retrospective framework application: The AI Readiness Framework was applied after the initial interviews were conducted, meaning the interviews were not originally designed with the framework in mind. This may have limited the depth or precision with which some categories were addressed. To address this potential issue, additional interviews were later conducted.

3.5.1 Reliability

Reliability refers to the consistency and dependability of research process and findings, indicating the extent to which the results can be replicated under similar conditions [39]. In this study, several measures have been taken to ensure reliability across both qualitative and quantitative components.

For the quantitative elements, consistent themes were used when categorizing interview data, allowing for meaningful comparison across scenarios and interviewees. The use of the AI Readiness Framework's predefined categories supported a structured and repeatable coding process during analysis, helping to reduce subjectivity and enhance consistency in interpretation.

In the context of the literature review, a systematic and replicable approach to data collection was followed. This included clearly defined queries, as well as documentation of search terms and databases used, to ensure transparency and enable the process to be retraced or replicated by other researchers.

For the qualitative interviews, reliability was supported by using a semi-structured interview guide, ensuring that core themes were consistently explored across all participants while allowing flexibility for contextual depth. Detailed notes were taken during all interviews, and in some cases, interviews were recorded to ensure that no important details were missed. In those instances, participants were asked for their consent prior to recording. All interviews were analyzed using a consistent coding structure based on the eight categories of the AI Readiness Framework, ensuring coherence and comparability throughout the analytical process. Triangulation through researcher collaboration and peer feedback further contributed to the dependability of the findings.

3.5.2 Replicability

Replicability refers to the extent to which a study can be repeated with similar results, given the same procedures and context [39]. While full replication is rarely possible in qualitative research due to the interpretive nature of the data and the context-specific responses, steps were taken in this study to enhance procedural transparency and allow for methodological replication.

To support replicability, interview guides were developed. These guides were used consistently across all interviews, ensuring that the same two core themes were addressed with each participant: organizational value an AI-readiness, and protoype feasibility.

Furthermore, the data collection process including how participants were selected, how consent was obtained, and how interviews were conducted and recorded has been clearly documented. All interviews followed the same format, and detailed notes were taken to support the analysis. When interviews were recorded, participants were informed and asked for consent beforehand.

The analysis followed a transparent process using predefined eight themes from the AI Readiness Framework, which enhances the possibility for other researchers to follow the same steps and compare results in a similar contexts. Although the findings themselves are context-dependent, the approach taken in this study is replicable and can serve as a basis for further research in similar organizational settings.

3.5.3 Validity

Validity in qualitative research concerns the accuracy and trustworthiness of the findings and whether the research truly reflects the phenomenon being studied. In this study, several steps were taken to strengthen both internal and construct validity. Internal validity refers to the credibility and truthfulness of the findings, while construct validity concerns whether the concepts are accurately represented and measured through the research design [39].

To ensure internal validity, all participants were interviewed using interview guides. This consistency in the interviews helped ensure that comparable data was collected across all interviews while still allowing for flexibility and depth in participants' responses. Clarifying questions and reflective listening were used to confirm interpretations during the interviews.

Construct validity was strengthened during the analysis phase, where interview insights were systematically mapped using the eight categories of the AI Readiness Framework. Although the framework was not used during the interviews themselves, it provided a structure for analysing and organizing the data. This approach enabled interpretation of diverse qualitative inputs and supported alignment between empirical insights and theoretical constructs.

Additionally, validity was enhanced through triangulation of data sources (e.g., interviews from two countries and various organizational roles), as well as peer feedback and regular discussions with supervisors, which helped challenge assumptions and validate emerging interpretations. Together, these measures contribute to a valid and transparent research process.

3.5.4 Ethical Aspects

General Data Protection Regulation (GDPR) compliance was ensured throughout the study. All participants provided their consent for participation, including permission to record some interviews and to use AI-based transcription tools. In addition, explicit consent was obtained to access and analyze company data and related informational documents. This careful approach aligns with ethical guidelines and ensures that participants' rights and data privacy were fully respected.

3.5.5 Use of AI Tools During the Thesis

AI tools were used throughout the thesis to support text processing, transcription, and coding. These tools enhanced efficiency and quality but did not replace human judgment.

Language models such as ChatGPT were used to improve clarity and academic tone in written sections. Suggestions were always reviewed and revised by the authors to ensure accuracy and alignment with project goals.

AI-assisted transcription was partially used for workshop and interview material. All transcripts were manually checked and corrected to ensure contextual accuracy.

Prompt-based code generation tools were used to check syntax, refining function structures and assist in troubleshooting. All generated code was verified and adapted by the authors.

4

Implementation of AI Prototype

The following chapter presents the process for implementing the AI prototype, including creating a pipeline for operation and function, and choosing sufficient models. In accordance with RQ2: *"How can an AI-prototype create value within the department of Quality Management?"* and RQ3: *"How can a language model prototype be developed and used to efficiently extract and visualize insights from quality-related construction texts?"*, the prototype has been tuned and selected to align with company needs and optimize its performance simultaneously.

The needs of the Quality Department, as outlined in Section 1.1.3, served as the primary motivation for enabling restructuring of the hierarchy of quality issue reporting, an area often standardized within the construction industry (see Figure 3.2). The prototype was developed with a clear focus on efficiency and accuracy, targeting the classification and prioritization of quality issue types (e.g., insufficient painting, cracks in concrete, etc.), see Figure 3.3. Currently, such issues are described freely within predefined input fields labelled `TITLE` and `DESCRIPTION` which lack standardization and structure. Given the nature of the data and the volume of reports, the problem was well-suited for a machine learning-based approach. This context motivated the application of NLP techniques and ML algorithms in the development of the prototype, a focus on weak supervision and reducing general human intervention.

4.1 Prototype

Figure 4.1: Overview of the 5 stages of implementation, including the 3 models, including, raw swedish construction data, clustering wia SBERT + KMeans, category definition through workshop, zero-shot pseudo-labels and transformer-based XLM-R classifier.

The implementation consisted of five stages, involving the development and evaluation of three distinct models, see Figure 4.1. In the first stage, the primary objective was to collect, extract, and properly format the dataset for downstream processing. The second stage introduced a clustering-based model using KMeans to explore and benchmark potential semantic groupings within the text data. In the third stage, a workshop was conducted to validate and refine candidate category labels, providing a human-guided baseline for classification as well as manually label a subset of the data. The fourth stage involved the development of a weakly supervised model based on pseudo-labelling, which leveraged semantic similarity to assign labels to unlabelled text. Finally, the fifth stage focused on semi-supervised learning and generalization, where a transformer-based classifier (XLM-R and KBLab's sentence-BERT) was fine-tuned using both human-labelled and pseudo-labelled examples.

Across all stages, each of the three models demonstrated the capability to segment the dataset into meaningful groups. Notably, the two latter models also produced groups based on the pre-defined category labels.

The following sections provide a detailed presentation and discussion of each implementation stage.

4.1.1 Data and Data Collection

The dataset was around 120 000 data-points, collected from March 2019 until March 2025, and stored in Structured Query Language (SQL) and extracted from the database available in Snowflake. The dataset consists of reported quality issues and corresponding features chosen in order to maximize valuable information and consider the trade-off between interpretability and possible insights.

The data points were filtered on the unit `5000 hus`, and with standardization flags `STANDARDS_ISSUE` is not null, meaning they follow a standard structure of reporting, with pre-defined categories for `ISSUE_TYPE` and `ISSUE_SUBTYPE`, which has been mandatory since 2024. The 19 features chosen can be seen in Table 4.1 below:

Data Features	
SK_ISSUE	STANDARDS_ISSUE
PROJECT_NUMBER	STANDARDS_ROOT_CAUSE
PROJECT_NAME	OPENED_DATE
CREATED_DATE	CLOSED_DATE
TITLE	START_DATE
DESCRIPTION	END_DATE
ISSUE_TYPE	LATITUDE
ISSUE_SUBTYPE	LONGITUDE
ROOT_CAUSE	REGION_NAME
ROOT_CAUSE_CATEGORY	

Table 4.1: Overview of key data features available in the dataset and used in the prototype

The current reporting hierarchy focused on `ISSUE_TYPE` with possible values were stages of the building process, such as `INSPECTION` and `PRODUCTION`, and `ROOT_CAUSE`, with values including `QUALITY` and `PRODUCTION`. Each issue was categorized in order to find correct department, meanwhile the aspired hierarchy of the data, has focus on cause and incident rather than process, in order to easily conclude key areas for improvement and gain valuable insights within quality design and management.

4.1.2 Initial Clustering Model

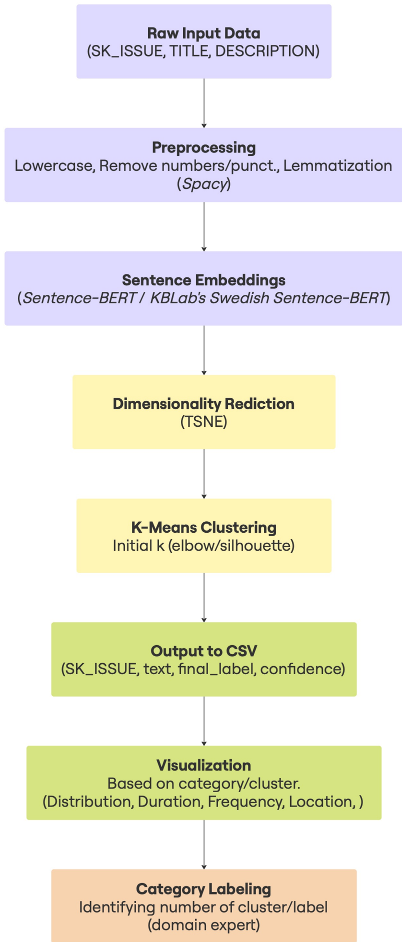


Figure 4.2: Overview of the initial clustering method, using KMeans clustering.

In the early stages of the project, an unsupervised clustering model was developed to explore the structure of the dataset and identify potentially meaningful groupings. Overview model architecture is seen in Figure 4.2. This process served both as a way to benchmark automatic categorization and as an exploratory method for identifying suitable category titles for later classification.

Each data entry consisted of a short title and a longer description, `TITLE` and `DESCRIPTION`. These were combined into one single string and converted into a semantic vector space, using a Swedish version of Sentence-BERT (`KBLab/sentence-bert-swedish-cased`). This model produced a dense 768-dimensional sentence embeddings that are optimized for semantic similarity tasks. All texts were first cleaned and formatted before being passed through the model, and re-occurring standard words such as "BYGG" were removed as they automatically were clustered into one group.

To enable clustering and visualize patterns in the dataset, the embeddings were further processed using t-distributed Stochastic Neighbour Embedding (t-SNE), which allowed us to reduce the high-dimensional vectors to two dimensions. While principal component analysis (PCA) was considered, it was not explicitly used in the final implementation. Instead, t-SNE was preferred due to its ability to reveal cluster structure more clearly during manual inspection.

Clustering was performed using the KMeans algorithm, which partitions data into a predefined number of clusters k. A range of values for k was explored, generally between 10 and 20, using both the elbow method and silhouette scores as guides. However, these metrics did not provide a clear consensus, so the final choice of k was ultimately based on a combination of metric outputs and qualitative judgment through visual inspection of the t-SNE plots and the text-components of the clusters. A value of 15 clusters was selected, aligning quite well with the expected diversity in the data.

The resulting clusters were used in two main ways. First, they provided a useful benchmark for later evaluation of classification performance (entire dataset). Second, they served as the basis for the evaluation of possible categories (consisting of 5% of the entire dataset), which would be used in the following models to train a weakly supervised model inspired by MEGClass.

4.1.3 Workshop-Based Label Design and Manual Annotation

To derive meaningful and operationally relevant category labels from the clustered data, a structured workshop was conducted involving key stakeholders from the quality departement, operations department and aftermarket, roles with relevent deep expertise in construction documentation. Participants included experienced field workers, operational managers, and senior project managers. The goal was to review and interpret the semantic clusters generated from a 5% sample of the dataset and collaboratively assign category names that reflected both construction terminology and practical categorization needs.

During the workshop, participants analysed representative text examples from each cluster and discussed recurring patterns, terms, and reporting styles. A key insight that emerged was the value of structuring the categories along two complementary dimensions: one representing the *building parts* involved (e.g., *Window*, *Balcony*, *Exterior ground*) and the other representing the *Incident* (e.g., *Construction defect*, *To be adjusted*, *Mark*, *Other*), as seen in Table 4.2:. This dual perspective was seen as highly actionable in operations, as it captures both the object of concern and the nature of the deviation, enabling more precise sorting, triaging, and response.

Following the categorization, manual labelling was performed to support classifier fine-tuning. Around 10 representative examples per category were manually annotated based on workshop guidelines. This included interpreting ambiguous texts and applying expert judgment on borderline cases, guided by the operational experience shared during the workshop. These manually labelled examples served as a high-quality reference set for supervised learning.

To strengthen the training process, additional manual labelling was carried out, as described in Table 4.3. Given the dataset's class imbalance, oversampling was applied to ensure equal representation across labels, resulting in labelled datasets accounting for 6 ‰of the total dataset. The final labelled dataset combined manually labelled data and pseudo-labelled examples with various composition, enabling a weakly supervised training approach with reduced human effort while maintaining semantic precision.

This collaborative and structured labelling process ensured that the classification system was not only data-driven but also aligned with real-world practices and terminology, increasing both model interpretability and potential for operational adoption.

Labels: Building Element	**Label: Common Issues**
Joint	Construction defect
Painting	Damaged product
Installations	To be adjusted
Wall	Mark
Window	Other
Balcony	
Roof	
Floor	
Door	
Exterior ground	
Interior fittings	
Other	

Table 4.2: Overview of chosen labels through workshop; building elements and common issue categories

4.1.4 Category Classification through Pseudo-Labelling and Classifier Implementation

To enable scalable and consistent classification of Swedish construction-related issues, we developed a two-stage weakly supervised classification pipeline inspired by the MEGClass framework (see Figure 4.3). The approach combines semantic embeddings, similarity-based pseudo-labelling, and optional semi-supervised fine-tuning via a transformer classifier. The full pipeline consists of a core method

Manually Labeled Data: Distribution Across Categories	
Construction Part	**Number of Issues (Manual Labels)**
Installation	58
Målning	49
Fönster	45
Inredning	39
Vägg	36
Fog	35
Golv	27
Balkong	25
Dörr	24
Utvändig mark	21
Tak	21
Övrigt	6
Incident Type	**Number of Issues (Manual Labels)**
Byggfel	114
Märke	29
Justering	10
Övrigt	2
Justeras	2
Skadad produkt	2

Table 4.3: Distribution of manually labelled quality issues across construction parts and incident types.

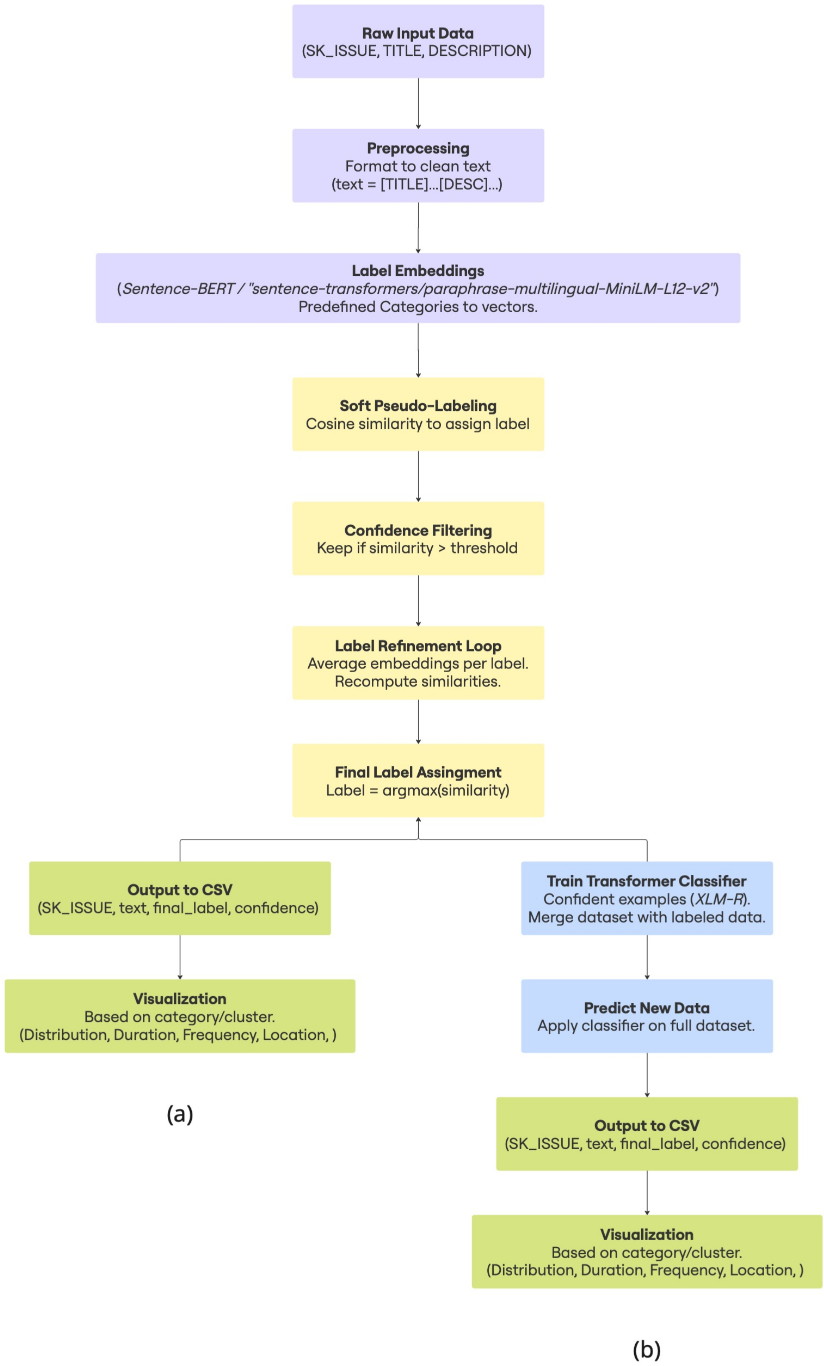


Figure 4.3: Overview of the weakly supervised MEGClass-inspired model, utilizing pseudolabels. The (a) approach is only categorization, meanwhile model (b) is a semi-supervised classifier based on transformers.

(embedding-based pseudo-labelling with refinement) and an extended version including a trainable classifier for downstream prediction. These methods are later compared to each other as well as to the traditional clustering using KMeans for benchmarking and validation (see Figure 4.2).

Initially, the raw input dataset consisted of unstructured text records with three columns: a unique identifier (`SK_ISSUE`), a short title (`TITLE`), and a longer problem description (`DESCRIPTION`). These fields were merged into a structured format with simple tags to preserve their semantic role: [`TITLE`] ... [`DESC`] .... This format was designed to help the model distinguish between possible headline-like and descriptive content, which is important for capturing subtle contextual differences.

Next, a multilingual Sentence-BERT model (`sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2`) was used to embed both the documents and a manually curated set of predefined category titles. These embeddings were then used to assign initial soft pseudo-labels to each document by computing the cosine similarity between a document vector and each label vector. The label with the highest similarity was selected, and the corresponding similarity score was treated as a confidence score. To ensure label quality and reduce noise, only examples with a confidence above a predefined thresholds (0.4 and 0.8 was used) were kept for further steps, while others were temporarily excluded.

In the next stage, label refinement was performed. For each label, the embeddings was averaged of all documents that were confidently assigned to that category. This resulted in a set of refined label vectors that better represented the semantic content of each group. The cosine similarity was recalculated between all document embeddings and the refined label embeddings, allowing the final pseudo-labels to be assigned based on the updated similarity scores.

In the base version of the model (a), these refined pseudo-labels were directly used to classify the entire dataset. In the extended (b)-version, a transformer model was additionally fine-tuned (`xlm-roberta-base` and `KB-Lab-sentenceBERT`) on the, by the model, confidently pseudo-labelled subset merged with the human-labelled examples to improve generalization with the highest confidence score of 1. The size of the pseudo-labelled dataset used for fine-tuned varied in the trials, and it ranged from 1 percent to 0 percent of the total dataset size. A custom weight function was created increasing the weights, hence the impact, of the manually labelled dataset. The fine-tuned classifier was then used to reclassify the entire dataset. This version allowed for learning a more nuanced decision boundary and potentially improved performance on unseen or ambiguous cases. In order to compute and train the model, the high performance cluster Alvis via OnDemand portal was utilized. The Nvidia GPU's T4, V100 and A40 was used and modules Python 3.9.5 and CUDA 11.8.0 loaded.

The outputs from both versions included the final label, the raw input text, the corresponding confidence score, and the ID. These results were saved and used for downstream visualization and statistical analysis. Finally, both approaches—pseudo-label-only and classifier-enhanced—were compared against each other as well as to a third baseline approach using KMeans clustering over SBERT embeddings.

4.1.5 Visualization of Results

The final classification outputs from all methods, pseudo-label only, classifier-enhanced, and KMeans clusters, were stored as structured .csv files, including columns for the issue ID (`SK_ISSUE`), the combined input text, the assigned final label (or cluster), and a confidence score where applicable. Based on the ID the data was merged with the rest of features as seen in Table 4.1 for a complete dataset again.

For interpretability and downstream analysis, and also for presenting for the Quality and Deviations team, a set of visualizations focused on category-level insights was developed, specifically focused on the issues with `ISSUE_SUBTYPE: AVVIKELSE` (deviation). These included:

- Cluster Visualizations: Using t-SNE applied to embeddings, 2D scatter plots with colors representing assigned categories or clusters was generated, helping evaluate how well-separated the semantic spaces were.
- Category Distributions: Number of examples per category/cluster with/without region, allowing quick spotting of imbalances or dominant classes.
- Temporal Duration and Frequency: When timestamp metadata was available, occurrences per category over time were plotted.
- Median Duration: The median duration of open cases.
- Issue Type Distribution: Per cluster showing in what part of construction process issues are reported.
- Deviation Distribution: What main categories and types of deviations were reported.
- Issue Duration: The median of duration of ongoing issues, before being resolved.
- Geographic Distribution: If location information was included in the data, the location of the reported issue was shown on a map.

These visualizations not only supported qualitative evaluation of classification quality, but also laid the groundwork for potential downstream use cases, such as anomaly detection, trend analysis, or operational planning within the construction sector.

4.2 Limitations in Prototype Implementation

While the prototype demonstrates the feasibility of applying ML and NLP to construction-related quality issue classification, several limitations affect its generalizability, scalability, and evaluation. These limitations stem from both technical constraints and organizational conditions encountered during the development process. The following section outlines key challenges related to data, model selection, labelling practices, and infrastructure capacity.

Language Barrier: One limitation of the prototype is its reliance on Swedish-language data, which restricts generalizability. While the use of multilingual models such as Sentence-BERT and XLM-R supports cross-lingual embeddings, the actual texts and labels remain Swedish-specific. This makes the system tailored to the national context and may require significant adaptation for use in other languages or regions.

Company Specific: The dataset originates from a specific actor within the industry, Skanska, and may reflect internal terminology, naming conventions, and organizational processes. This raises concerns regarding external validity, as similar models trained on this data might not perform as well when applied to data from other construction companies with different documentation practices.

Label Provenance and Relevance: Another important limitation concerns label provenance, understanding who assigned each label, how it was generated, and whether it was reviewed or validated. In this project, category definitions were developed through clustering and refined in collaboration with domain experts. However, once labels were applied more broadly through pseudo-labelling or manual annotation, the origin of individual labels became less clear. Without traceability, it becomes difficult to assess the reliability of the labels used to train and evaluate the model. For example, it may be unclear whether a misclassification is due to model error or a questionable label. This uncertainty makes it harder to refine the system systematically. Moreover, labelling strategy and provenance management often depend on the company's internal workflows and resources. In a larger organization, labels may come from multiple sources—different teams, automated systems, or even outsourced annotation—which introduces strategic challenges. Ensuring consistent and transparent labelling practices is not just a technical issue but also an organizational one, particularly when scaling AI solutions across departments or business units.

Outdated Models: In terms of model architecture, although well-performing and resource-efficient models were selected, the project used components (e.g., SBERT MiniLM, early MEGClass logic) that have been available for several years. While these models are not obsolete, newer alternatives with higher performance and flexibility are emerging. Only during recent years NLP and LLM have evolved significantly, making a model from a few years back possibly outdated. The decision to use these models was driven by practical constraints, including computational limitations. Limited access to GPUs and memory capacity required optimization of batch sizes, simplified training loops, and reduced data usage during fine-tuning stages. Also, the original MEGClass model was based on older versions of Python packages and utilizing methods and available tools high-performing at that time. With today's standards, even though inspired by a high-achieving model, the method could by using other more recent inspiration model possibly perform better.

4.3 Ensuring High-Quality Data

The quality of the dataset forms the foundation of any machine learning system, particularly in natural language processing tasks where semantic nuances and domain-specific language can significantly impact model performance. In this project, ensuring high-quality data was not only a prerequisite for reliable classification, but also a critical factor in achieving generalizable and ethically sound results. This section outlines the core dimensions through which data quality was assessed and safeguarded.

The section begins with an assessment of reliability, addressing potential biases and imbalances in both the dataset and models. It then outlines the replicability of the pipeline, including model configurations and data accessibility. The validity of the system's predictions is evaluated based on alignment with domain expectations and real-world usage. Lastly, ethical considerations are discussed, focusing on data privacy, transparency, and responsible deployment of AI in industrial contexts.

4.3.1 Reliability

The reliability of the classification system depends heavily on the quality and diversity of the underlying data. One challenge is the presence of bias in both the data source and the model. Since the dataset is based on internal documentation, it may reflect organizational biases such as how certain issues are documented or prioritized. For example, it has been speculated that among employees, choosing `Deviation` as a `ISSUE_SUBTYPE` might to some extent be avoided purposely or subconsciously.

Similarly, pre-trained transformer models can carry inherited bias from their large-scale pretraining data. The models may also themselves exhibit reliability issues when applied to short or ambiguous texts, particularly when they lack context or contain domain-specific terminology not covered by the pre-training corpus.

Finally, another issue is the imbalance in data volume across categories. While the overall dataset is large, some clusters and classes are slightly under-represented. This affects the model's ability to generalize and increases the risk of overfitting to dominant categories. Although, the manyally labelled data is balanced to prevent this.

4.3.2 Replicability

To ensure replicability, it is crucial to provide detailed documentation of the model versions, training parameters, and dataset preprocessing steps. In this project, model selection included openly available versions of Sentence-BERT and XLM-R, with training conducted using fixed random seeds and HuggingFace's `Trainer()` interface.

However, not all parameters (e.g., clustering hyperparameters, similarity thresholds, or label refinement steps) were exhaustively tuned or logged based on performance of models and need in result. Additionally, the data used is proprietary, which poses a challenge for external replication. Making a synthetic or anonymized sample available could partially mitigate this issue and increase transparency.

4.3.3 Validity

Evaluating validity in a text classification system involves a standard metric, such as confusion matrices for a quantitative perspective, to measure predictive accuracy. These showcases class-accuracy to ensure fair evaluation across imbalanced categories. However, it is hard to define accuracy among the result as there are extremely many data points and whether or not it is actually truthful is upon a domain expert to decide.

Furthermore, because the texts are short and domain-specific, construct validity is equally important. The labels must represent semantically coherent and operationally useful groupings. To address this, workshop participants reviewed cluster outputs and manually adjusted category definitions to better reflect the real-world use cases.

Temporal validity also matters in this setting. Some texts describe issues relevant only during certain phases of a project or a defined way of reporting. A model trained on older data may fail to recognize newer terminology or patterns, especially as the organization evolves.

Lastly, the data used was derived from actual company documentation, ensuring a high degree of real-world relevance. However, the lack of full context in many short texts could still limit interpretability.

4.3.4 Ethical Aspects

The prototype was developed with awareness of core ethical considerations related to AI deployment. One key concern was compliance with the GDPR. Formal access to the dataset was granted by the case company, which owns and manages the data. The dataset was screened to avoid the inclusion of personal data, and no fields containing solely names or other directly identifying information were used during model training. However, as with any free-text input, there remains a minor risk that names or identifying references could appear sporadically in fields such as `DESCRIPTION`. This limitation highlights the importance of ongoing data governance and anonymization procedures when working with operational text data.

Second, transparency in training data remains an open challenge. While models like XLM-R are openly available, their training corpora include data from web sources that may not be fully transparent or vetted. This might raise concerns about hidden biases or ethical issues embedded in the foundation models.

The increasing use of AI in decision support raises questions of responsibility and interpretability. The system outputs category predictions, but these should be used with a bit of caution and human judgment, as the actual classification and what it is based on might be hard to understand. This is especially important in quality- or safety-related documentation in construction projects.

Lastly, the classifier's decisions may unintentionally reflect systemic issues, such as under-reporting of certain categories, or biases introduced during the label workshop phase. These risks may reinforce the need for careful deployment, continuous evaluation, and human-in-the-loop supervision.

5 Results

This section presents the empirical results related to the three research questions. RQ1: "*What is Skanska's current state of AI readiness, and what organisational strengths and challenges exist in terms of AI adoption?*" is assessed based on thematic findings from interviews and organizational insights, identifying both strengths and barriers to implementation. RQ2: *"How can an AI-prototype create value within the department of Quality Management?"* explores how the prototype delivers strategic and operational value, including its outputs and data-driven insights that enhance the department's work. Lastly, RQ3: *"How can a language model prototype be developed and used to efficiently extract and visualize insights from quality-related construction texts?"* is explored through evaluating the prototype's performance and its capacity to extract structured insights and enable data-driven analysis. The findings are presented through quantitative metrics, visualizations, and interpretive observations.

5.1 Skanska's AI-readiness

This Section presents the summarised results of RQ1 from the interviews and workshops, structured around the eight themes of AI Readiness described in the frame-of-reference chapter 2.2.2. The summaries highlight the most important insights from each interview, connected to its specific theme. The full 12-page results, including detailed quotes and nuanced findings, are available in Appendix C.

Environmental Readiness
In the Swedish context, the general attitude toward AI was positive but passive, with operational staff and managers showing curiosity, but also uncertainty about how AI applies to their roles. While tools like Copilot and Sidekick exist, many employees are unaware of them or unsure how to use them, and there is no formal structure to encourage their adoption. Younger employees are perceived to be more open to experimenting with digital tools, while older staff tend to be more hesitant. The IT department has taken steps to raise awareness through training and examples, but adoption remains fragmented and driven by individual interest rather than organizational support.

Internationally, the US and UK respondents described a more proactive and engaged environment, where AI is seen as a strategic necessity and embedded into planning and operational discussions. In the US, managers actively participate in testing and refining AI of tools, actively working towards a culture of learning and adaptation. In the UK, there is strong strategic interest in AI, though implementation varies across business units, and interest has not always translated into action.

Technological readiness
Technological readiness within Skanska Sweden is characterised by fragmentation and legacy systems that are not designed with AI in mind. Digital tools are mostly used for documentation and compliance, lacking the structure and integration needed to support AI applications. Several respondents noted that while there are many digital systems in use, they are not set up for AI, making AI implementation difficult without significant changes to the systems in place. Despite these limitations, there are early signs of experimentation and interest in integrating AI into existing systems, such as using Copilot to generate proposal texts for different purposes. However, these efforts are isolated and not yet part of a larger strategy. Employees expressed frustration with manual data handling and system silos, highlighting the need for better coordination and standardization within the organization.

In contrast, the US and UK interviews described more advanced technological systems. In the US, teams have built centralized data infrastructure and in-house software platforms specifically designed to support AI tools. In the UK, while usage of AI-enabled tools is still low, there is recognition of the need to build workflows around these tools to realize their potential.

Informational Readiness
Informational readiness at Skanska Sweden is characterized by an awareness of AI at the leadership level but a lack of clear guidance and understanding at the operational level. While managers and decision-makers generally see the potential of AI, many employees feel unsure about how AI could specifically support their work. Respondents pointed to the need for more practical examples, use cases, structured communication to turn interest into action, and internal champions to guide the change.

There are some positive signs, such as internal efforts to share examples and train staff, but these are not yet widespread or systematic. Some employees described a gap between high-level enthusiasm and the concrete guidance needed to implement AI in day-to-day tasks, to avoid AI becoming a buzzword unless someone explains how it applies in your role. AI needs to be tied to operational pain points, and employees need something that tells them what you can do with AI, in this exact context. Furthermore, respondents highlight that the presence of more specialized data and analytics roles contributes to a higher level of practical literacy around AI among managers.

In contrast, interviews with US and UK representatives highlighted more developed informational practices. For instance, in the US, managers and employees see AI as a support tool, integrated into operations and tied to practical outcomes. In the UK, forums and knowledge-sharing sessions are helping to connect leadership ambitions with local applications of AI.

Infrastructural Readiness

Swedish interviewees consistently pointed to a lack of formal roles, designated budgets, and clear ownership of AI-related work. While there is general enthusiasm for digitalization, competing priorities and resource constraints limit the capacity to develop and sustain AI initiatives. Some departments rely on temporary efforts and informal champions, rather than structured support. The IT and digital roles are mainly focused on maintaining existing systems rather than supporting new AI initiatives. Respondents described a need for clear mandates, handover routines, and governance structures to ensure that promising pilots do not stall. Several suggested that small, cross-functional groups could be a way to ensure continuity and build on early efforts.

In the US and UK, interviewees described more mature support systems that link data, technology, and business goals. For example, in the US, strategy teams are integrated with technology experts.

Data Readiness

Skanska Sweden's data landscape is consisting of information from projects and documentation, but this data is often unstructured, inconsistently entered, and scattered across disconnected systems. Several interviewees highlighted that much of the data was initially gathered for compliance and reporting, not for learning or analytics, making it difficult to use it for AI applications. Despite these challenges, there is a growing awareness of the gap and the potential value of transforming data practices. Participants described efforts to retrofit data processes, and emphasized the importance of improving metadata and labeling, particularly if AI is to be applied at scale.

The contrast with US and UK operations was clear, both have invested in proactive data governance and structured data environments. In the US, years of data cleaning and alignment have enabled real-time predictive models in the area of safety.

Participants' Readiness

Participants' readiness in Skanska Sweden is generally characterized by curiosity and cautious openness towards AI. Most employees see AI as a tool to improve workflows and make their jobs easier, especially when it's positioned as a support mechanism rather than a replacement for expertise. There's a clear interest in AI tools that save time, reduce repetitive work, and help avoid risks, but only if the purpose and benefits of AI are transparent and directly relevant to daily tasks.

A recurring theme is that many employees still lack a basic understanding of AI, and they need clear onboarding and training to feel confident using new tools. Trust and transparency are essential, and workers want to know what AI is doing and that it's working in their favour, not behind the scenes in ways they don't understand.

Furthermore, some employees noted that they hesitate to report deviations because they fear being blamed or misinterpreted, which leads to less reliable data available.

From an international perspective, examples from the US and UK reinforce that trust grows when employees see real, role-relevant value. Peer learning and sharing of positive experiences help build momentum.

Customers' Readiness
Currently, clients do not specifically ask for AI solutions, although some customers include expectations for the use of digital tools. While there is clear internal enthusiasm for AI, it is not yet matched by external customer pressure. Internally driven efforts and operational needs have historically been the driving force, and Skanska does typically not innovate for branding or show. Rather, it responds to explicit client demands. Still, they acknowledged that if clients begin to request AI-enabled services, it could significantly accelerate broader adoption within the organization.

In the US and UK, however, customer readiness was described as more established, even if often indirect. In the US, AI initiatives are integrated into project planning, focusing on delivering customer value in areas like safety and compliance. When AI helps achieve KPIs or regulatory standards, it is easier to justify and secure investment in AI capabilities.

Process' Readiness
Interviewees described an environment where workflows are often reactive, shaped by project-specific demands rather than standardized processes across the organization. This variability limits the integration of AI tools, which depend on consistent processes and shared understanding to generate reliable results.

AI initiatives are often unsuccessful because existing workflows are not designed to support data-driven decisions or automated analysis. Furthermore, the absence of defined feedback loops and structured processes prevents these efforts from evolving beyond temporary solutions. Some employees observed that while AI solutions targeting administrative bottlenecks, such as time-consuming document handling or reporting can be helpful, there is a lack of clear guidelines or internal champions to turn these into scalable solutions.

In contrast, international examples showed how embedding AI tools into operational KPIs and project delivery processes can foster consistent use and continuous improvement. This highlights the importance of aligning processes and workflows with AI capabilities.

5.2 The Value to the Quality Department

This section presents results in from workshops and interviews with employees from the Quality department, as well as the illustrating data findings for quality insights, focusing on covering RQ2.

5.2.1 Insights from the Quality Department

The results of interviews and workshops with the quality management team show that they consistently express a desire to gain better insight into what is actually happening within construction processes, to see both the deviation reportings and their root causes in a more efficient and systematic way. Currently, while they find some value in their Power BI dashboards, these tools are not fully developed for in-depth or proactive analysis. As a result, much of their time is spent manually navigating unstructured information, limiting their ability to focus on strategic quality improvements.

A senior manager representative from Skanska Sweden's Quality Department, emphasized that current deviation management processes are heavily manual and fragmented. Deviations have historically been reported in free text, leading to inconsistent terminology and varied categorization practices across projects. This lack of structure requires significant manual effort to identify patterns and conduct analyses. The respondent described the existing approach as reactive and time-consuming, noting that quality managers often spend extensive time reviewing individual cases without effective methods to gain an overview of them.

The AI-prototype developed during the thesis directly addressed these challenges, and the respondent particularly valued the prototype's ability to cluster deviations based on issue type rather than process phase, offering a more meaningful structure aligned with how quality teams actually work. The respondent highlighted that this approach is valuable because it was based on a real and immediate operational need, rather than a theoretical or abstract AI application.

The prototype's visualizations, showing frequency, distribution, and average duration of deviation types, were seen as highly valuable in supporting data-driven prioritization of cases. By highlighting recurring issues, the tool could shift quality work from reactive problem-solving towards a more preventative and strategic focus. The respondent underscored that this would be well-received, as it directly reduces the time spent on manual, repetitive tasks and enables earlier intervention in projects. One example of a previous root cause investigation involved a recurring deviation in a specific area of the buildings basements. This deviation was identified manually through observations. In this case, the AI-driven classifications could have highlighted the issue and provided insights into what was actually happening, doing so in a much more efficient way that the quality team would value. Furthermore, the respondent emphasized that structured, visual data makes it easier to communicate insights and build a shared understanding of quality challenges across the organization.

5.2.2 Illustrating Text-Based Data Results for Quality Insights

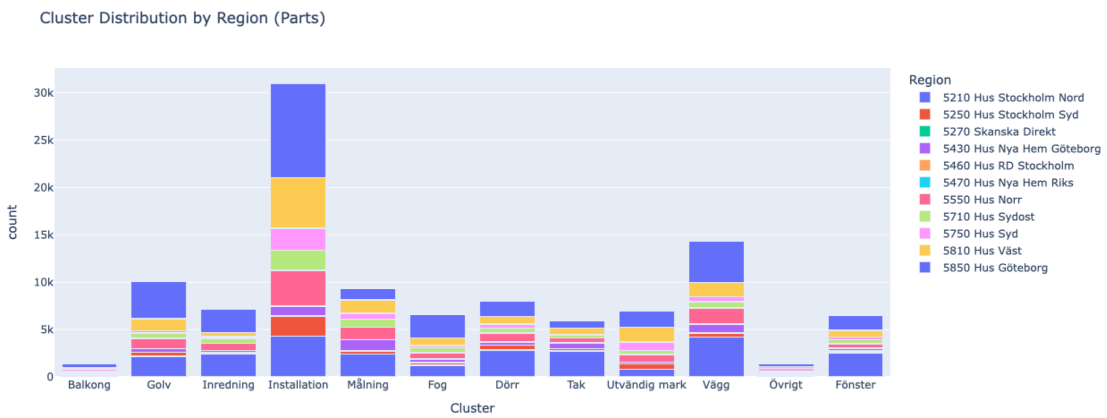
This subsection presents key results from the text classification model applied to quality-related metadata. The results are outputted from the classifier model fine-tuned only on the manually labelled data, as its accuracy exceeded that of the merged pseudo and manually labelled fine-tuning (see Section 5.3.1).

The analysis focuses on unstructured text fields, aiming to surface patterns and insights relevant to proactive quality assurance and risk assessment. The visualizations and findings are intended to support the department in identifying recurring issues, understanding thematic trends, and enhancing data-driven decision-making processes. An overview of categorized issues is seen in Table 5.1.

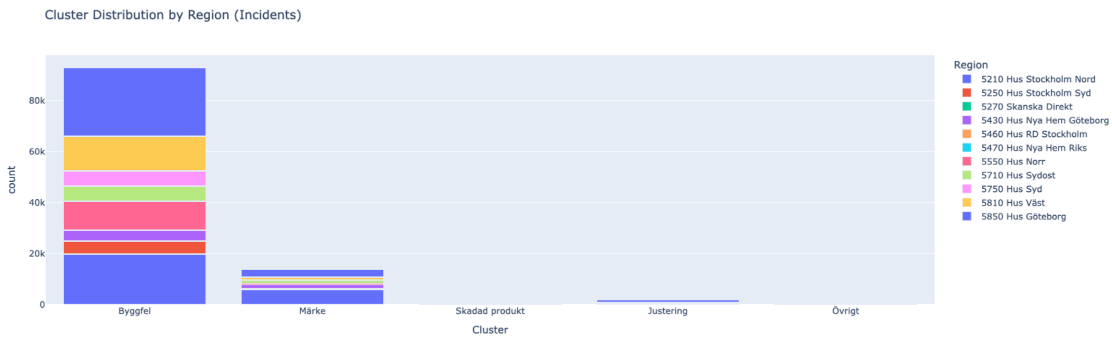
Categorization Output – Distribution Across Categories	
Construction Part	**Number of Issues**
Installation	30,987
Vägg	14,347
Golv	10,085
Målning	9,339
Dörr	7,975
Inredning	7,120
Utvändig mark	6,958
Fog	6,629
Fönster	6,516
Tak	5,956
Övrigt	1,419
Balkong	1,403
Incident Type	**Number of Issues**
Byggfel	92,864
Märke	13,911
Justering	1,843
Skadad produkt	104
Övrigt	12

Table 5.1: Distribution of AI-classified quality issues by construction part and incident type.

The distribution of categories across regions is presented in Figure 5.1. Figure 5.1a shows that the majority of reported issues related to construction parts are linked to installation. As expected, the larger regions account for a higher number of total reports. Figure 5.1b illustrates that most issues are associated with construction defaults across regions, and that the reporting is somewhat stable across regions.
Each category has an internal issue type distribution, which is showcased in Figure 5.2. In the Figures 5.2a and 5.2b a consistent trend is presented, where issues are



(a) The distribution of issues related to construction parts, per region, with a majority in *Installation*

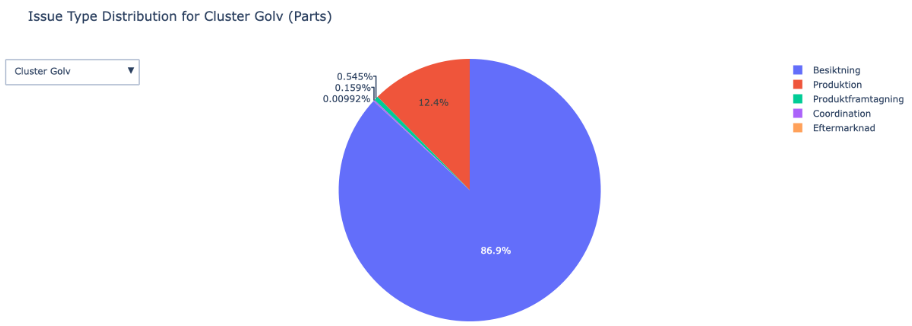


(b) The distribution of issues related to incidents, per region, with a majority in *Byggfel*.

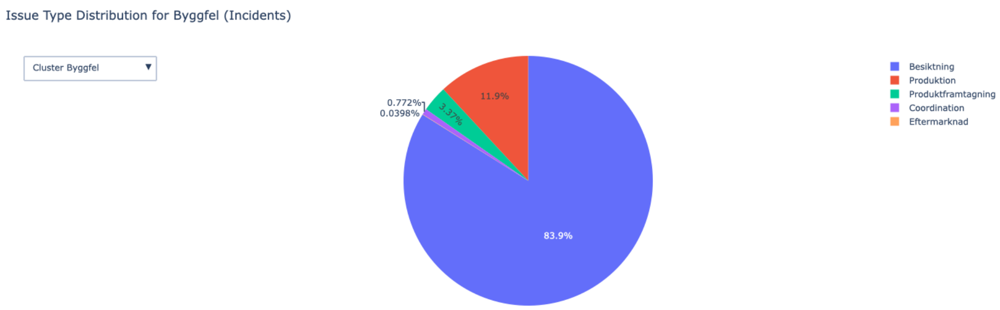
Figure 5.1: Indices to class: [0-Bream 1-Perch 2-Pike 3-Roach 4-Silverbream 5-Smelt 6-Whitefish].Confusion matrices showing the accuracies for different classification methods.

most commonly reported during inspection, followed by production, echoed in both category types for *Golv* and *Byggfel*.

Besiktning often accounts for the highest proportion within each cluster, suggesting that most quality deviations are identified during formal inspections. *Produktion* is typically the second most common, indicating that many issues are recorded as originating during the construction phase. In contrast, other categories such as *Produktmottagning*, *Coordination*, and *Eftermarknad* appear infrequently across all clusters.



(a) Representative issue type distribution for *Golv*, during the stages of the construction process.



(b) Representative issue type distribution for *Byggfel*, during the stages of the construction process.

Figure 5.2: Distribution of stage of construction process per category.

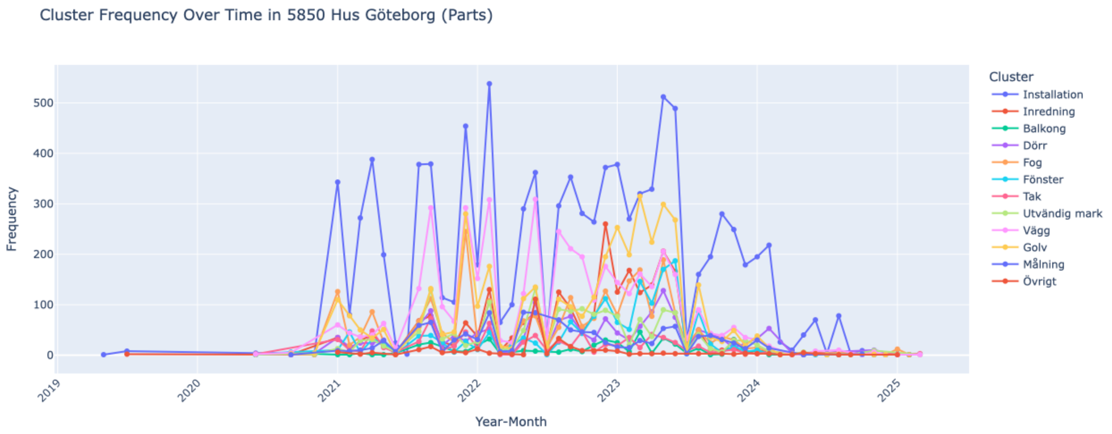
When focusing on deviations in Figure 5.3, *Avvikelser*, the result somewhat follows the overall trend in general cluster distribution as seen in Figure 5.1, with *Installation* and *Byggfel* as the most reported deviation. However, both *Installation* and *Utvändig mark* is clearly overrepresented for the construction parts categories, along with *Byggfel* and *Märke* for the categories targeting incidents.

(a) Distributions of reported deviations per construction part. *Installation* and *Utvändig mark* make up for more than 70 percent of the issues reported.

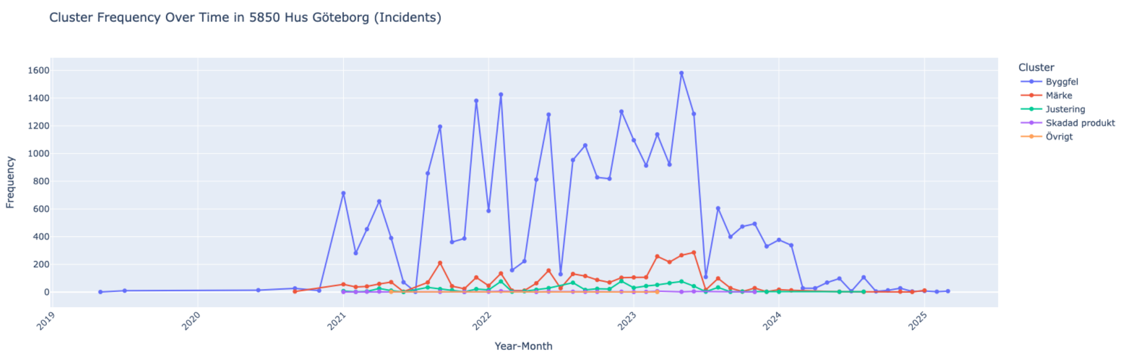
(b) Distributions of reported deviations per incident. *Byggfel* and *Märke* are the only present categories, where the former make up for more than 90 percent of the issues reported.

Figure 5.3: Distribution of the categories reported as a deviation per construction parts and incidents.

Investigated and analysed was also the frequency and time period the issues were reported. Figure 5.4 is showing peaks in waves, throughout all categories. The dataset consist of data from March 2019 until March 2025, and shown in the graphs is that during the early years few issues were reported in BIM360/ACC. Furthermore, from mid-2023 and onward, the quantity has decreased especially since mid-2024. Up until early 2024, the issue type categorization was not standardized, meaning some data-points can have been lost, as they were filtered out.



(a) The frequency of reported issues related to construction parts over time, in Hus Göteborg.

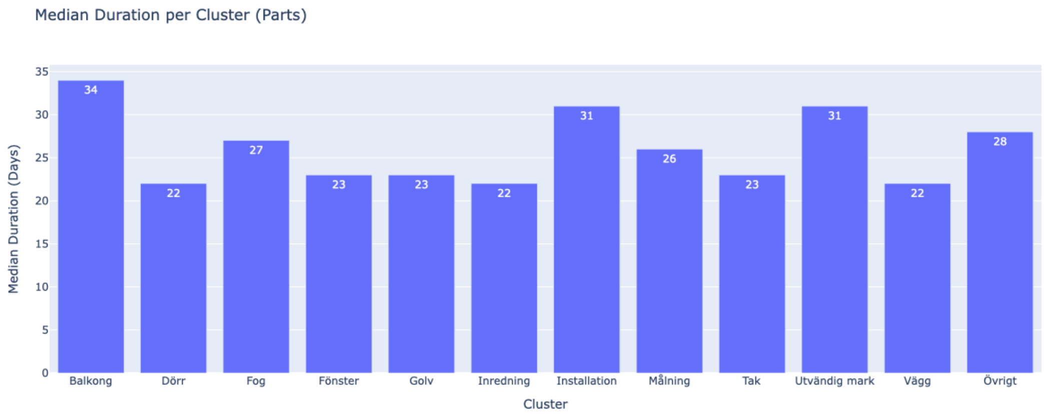


(b) The frequency of reported issues related to incidents over time, in Hus Göteborg.

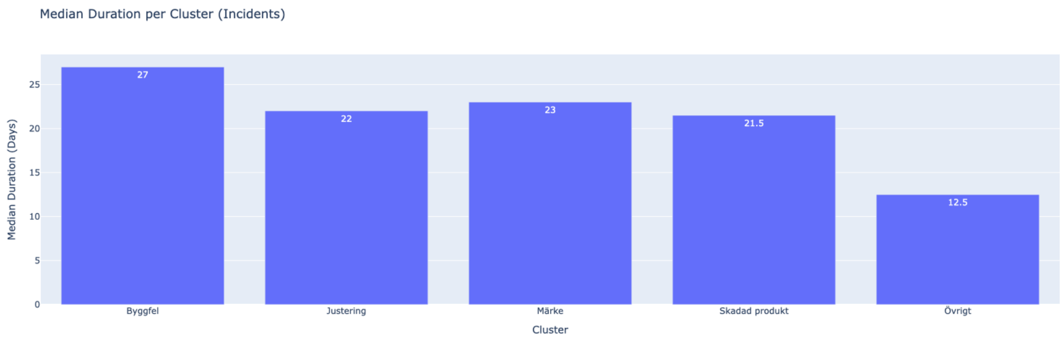
Figure 5.4: Frequency over time of reported issues in Hus Göteborg.

The median duration over all open cases per categories is seen in Figure 5.5, with possible view of median duration on region level. For the construction parts (Figure 5.5a) the time period varies between 22 to 34 days, with the longest related to balconies. For incident cases, the median duration is between 12.5 to 27 days (Figure 5.5b.

For local and geographic discoveries, Figure 5.6, maps the locations (longitude and latitude) of the projects related to the reported issues. The larger cities have a higher concentration of projects. For local trends or discrepancies, zooming in is possible.

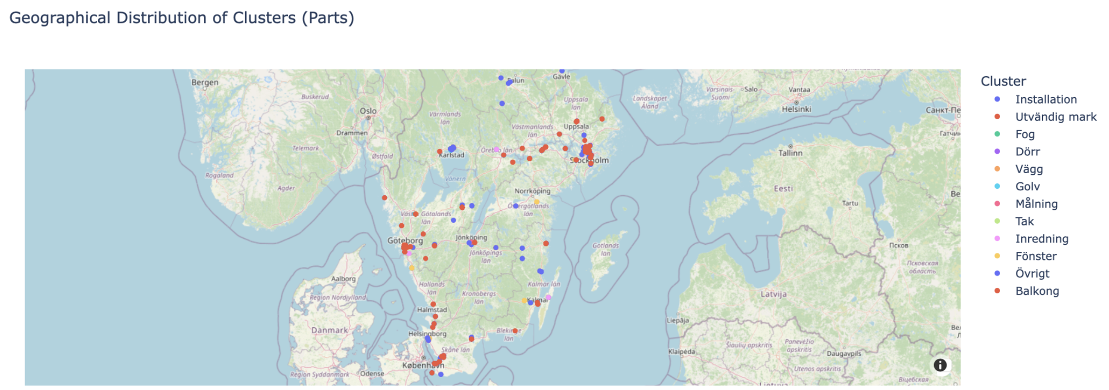


(a) The median duration over all regions per issues related to construction parts, with a values between 22 to 34 days.

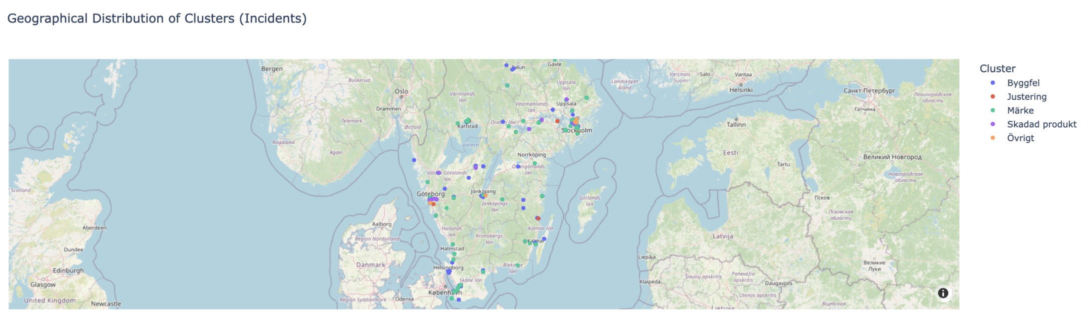


(b) The median duration over all regions per issues related to incidents, with a values between 12.5 to 27 days.

Figure 5.5: The median duration of open issues over all regions.



(a) The geographical location of construction parts issues, centered around Stockholm, Gothenburg and Malmö



(b) The geographical location of incidents issues, centered around Stockholm, Gothenburg and Malmö

Figure 5.6: The geographical locations of projects with reported issues, per category.

5.3 Classification Performance and Output Examples

This section presents the model's classification performance, including accuracy metrics and illustrative examples of the input sentences and their predicted labels. It focuses on evaluating the model's effectiveness and providing explanations to better understand the basis of its predictions aligning with RQ3.

5.3.1 Classification Performance via Confusion Matrices

To assess the classification accuracy and understand where the model succeeds and fails, confusion matrices were generated for both construction part and incident type classifications (Figure 5.7). A confusion matrix visualizes the relationship between true labels and predicted labels, revealing how often labels are correctly predicted and where misclassifications occur. The matrices were based on a manually validated subset of predictions. Since a large pre-labeled test set was not available, a targeted evaluation strategy was adopted: 25 predictions were sampled from each category, then manually verified against the true label by reviewing the underlying text.

This approach provides an insightful, albeit small-scale, diagnostic of the model's strengths and weaknesses in real-world predictions. It also enables identification of frequent misclassifications and category-specific challenges that may not be evident from aggregate metrics alone.

The matrices in Figure 5.7a and 5.7b, shows the results outputted by the classifier model, solely fine-tuned on the manually and artificially balanced labelled dataset.

The matrix in Figure 5.7c, shows the categorization results for construction parts, outputted through pseudo-labelling, when the lower limit of confident score needed to be assigned a category, was set to 0.8 (out of 1.0). Here, no incident categories fulfilled the criteria.

The matrices in Figure 5.7d and 5.7e, shows the categorization results for construction parts and incidents respectively, outputted through pseudo-labelling, when the lower limit of confident score needed to be assigned a category, was set to 0.4 (out of 1.0).

Construction Part Classification, manual labelling
The confusion matrix, Figure 5.7a, shows a mean accuracy of 67%, with strong classification performance for several categories, including *Installation*, *Fog*, *Inredning*, and *Balkong*, which are mostly predicted correctly. Some confusion is observed between *Målning* and *Vägg*, likely due to overlapping language in the input descriptions. The class *Övrigt* and *Utvändig mark* also receives misclassifications from multiple other categories, suggesting ambiguous phrasing or unclear label boundaries in the data. Overall, the model performs well on dominant and well-defined

classes but struggles with semantically similar or less distinct ones. Further refinement of label definitions and additional training examples may help reduce these confusions.

Incident Type Classification, manual labelling
The confusion matrix, Figure 5.7a, for incident type classification reveals strong model performance, mean value of 65 %, for core classes such as *Justering*, *Märke*, and *Skadad produkt*, where most examples are correctly predicted. However, *Byggfel* is frequently misclassified, particularly as *Övrigt*, indicating potential overlaps in textual patterns or insufficient separation between classes, and due to the fact of major overrepresentation. Additionally, *Övrigt* receives misclassifications from various other labels, which may reflect the presence of vague or multi-faceted issue descriptions.

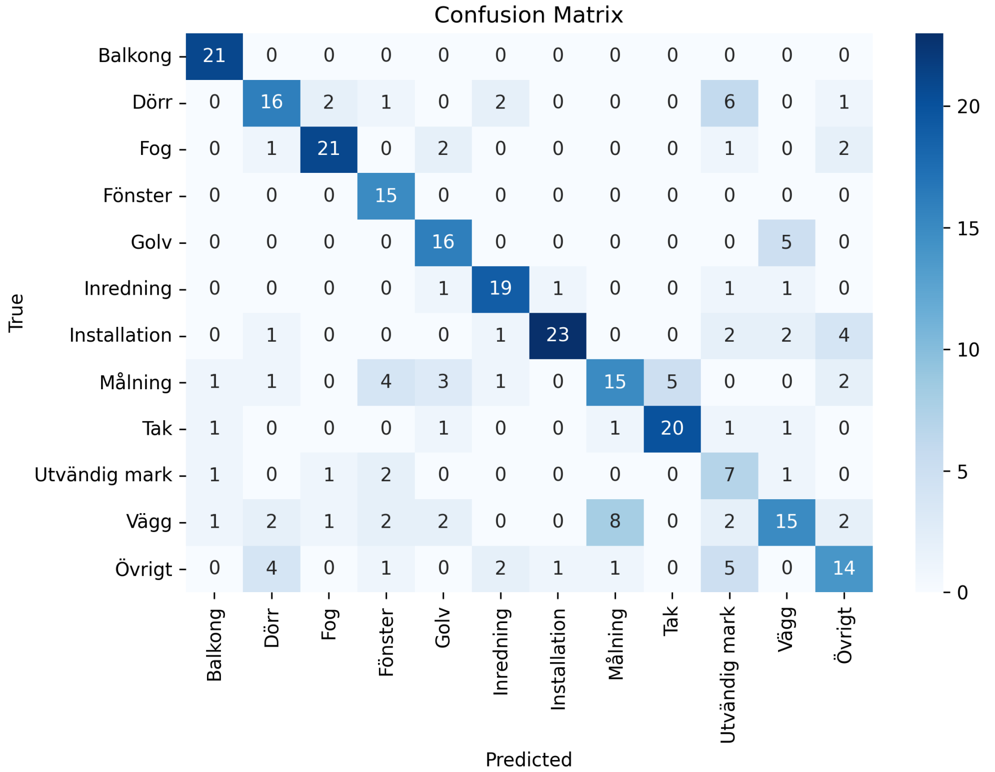
Overall, while the classifier demonstrates generally reliable accuracy, with a mean of 67% and 65%, especially on well-defined and sufficiently represented categories, the observed confusion in edge cases highlights the need for clearer label definitions, as well as enhanced training data quality and balance. These insights inform future improvements in model robustness and labelling strategy.

Construction Part Classification, pseudo-labelling 0.8
In the second matrix, Figure 5.7c, the model shows excellent performance for well-defined categories such as *Fog*, *Målning*, and *Balkong*, which are almost perfectly classified. For some categories, the model was not confident enough, and they were not predicted at all, like *Vägg*, *Utvändig mark*, and *Fönster*. This suggests the model performs best when class definitions are clear and training samples are representative.

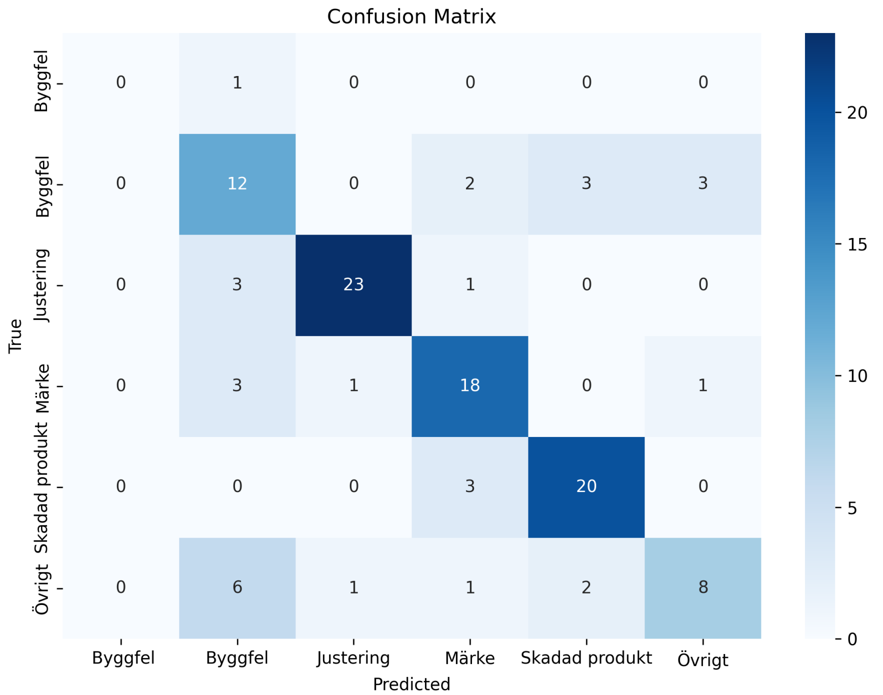
Construction Part Classification, pseudo-labelling 0.4
The third matrix, Figure 5.7d, reflects a more balanced but broader model, where several classes, including *Fönster*, *Installation*, and *Inredning*, exhibit notable confusion. Misclassifications are distributed across many classes, especially for semantically overlapping labels like *Målning*, *Vägg*, and *Tak*. The *Övrigt* class also receives a high number of misclassified instances, indicating that some descriptions may be too ambiguous for confident categorization.

While the model handles larger and frequent categories with moderate success, the broader distribution of errors suggests a trade-off between flexibility and specificity. Additional labelled data or label refinement could improve differentiation among visually or functionally similar construction parts.

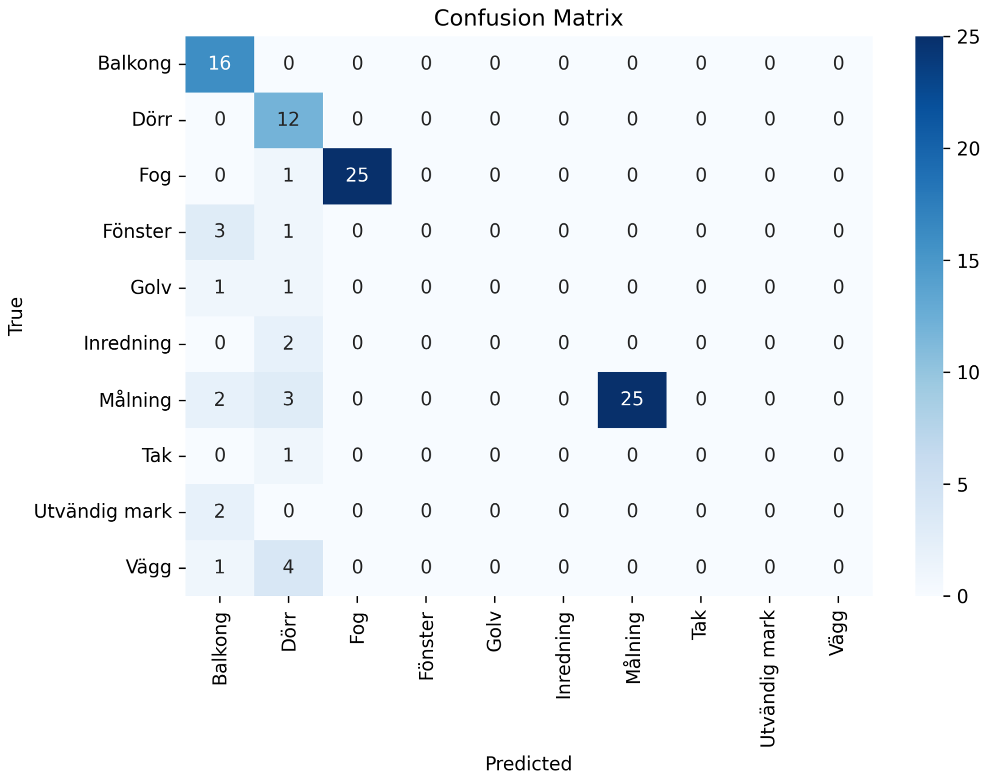
Incident Type Classification, pseudo-labelling 0.4
In the last matrix, Figure 5.7e, the classification of incident types shows significant overlap, particularly for the class *Byggfel*, which is frequently misclassified as *Märke*, *Justering*, and *Övrigt*. The matrix also reveals that *Övrigt* often functions as a fallback class, receiving instances from across the label set.



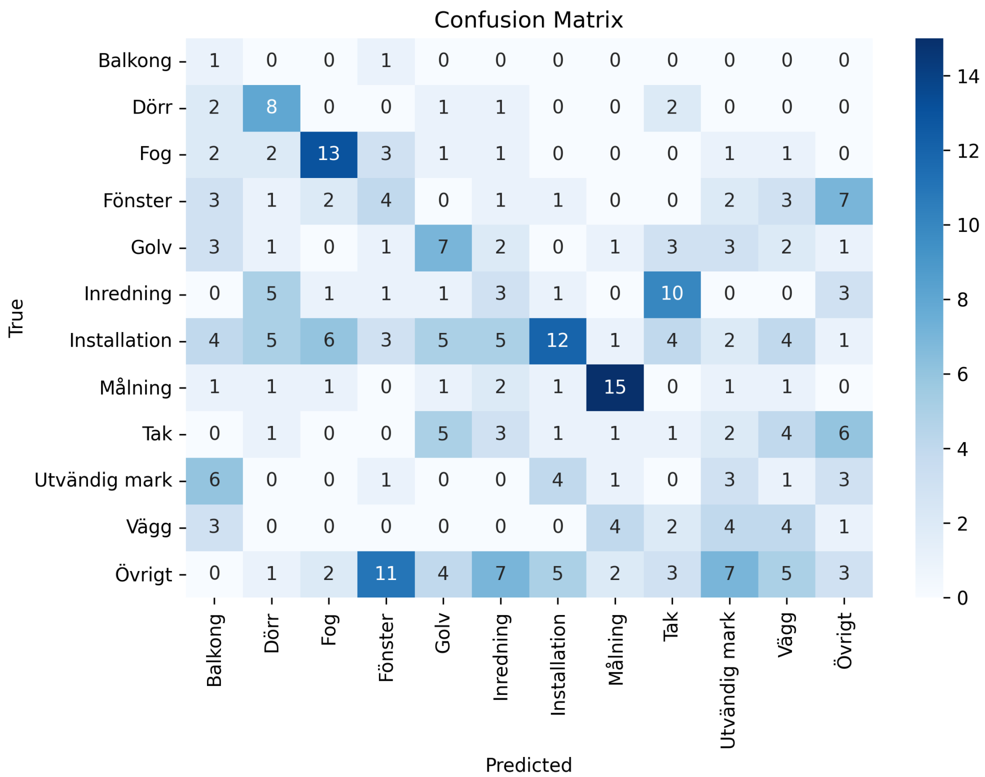
(a) Confusion matrix over the predicted and true labels per category for construction parts. The results are from model fine-tuned on manual labels only.



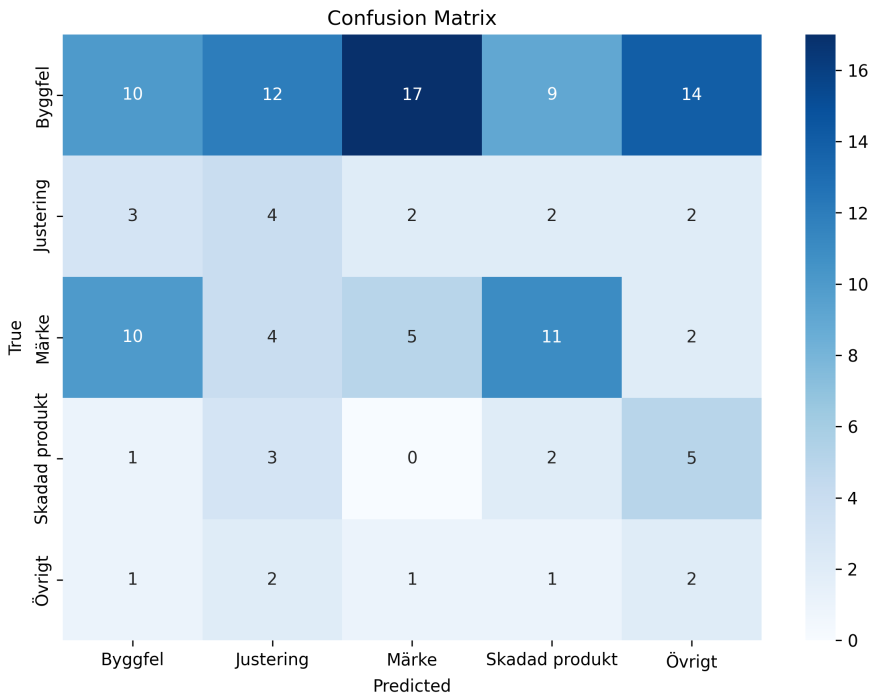
(b) Confusion matrix over the predicted and true labels per category for incidents. The results are from model fine-tuned on manual labels only.



(c) Confusion matrix over the predicted and true labels per category for construction parts. The results are from the predictions made through solely pseudo-labelling, with lower limit of confidence score 0.8.



(d) Confusion matrix over the predicted and true labels per category for construction parts. The results are from the predictions made through solely pseudo-labelling, with lower limit of confidence score 0.4.



(e) Confusion matrix over the predicted and true labels per category for incidents. The results are from the predictions made through solely pseudo-labelling, with lower limit of confidence score 0.4.

Figure 5.7: Confusion matrices showcasing the accuracy of the category predictions, based on a sample of 25 data-points per predicted category. A darker colour presents a higher accuracy.

These results highlight the need for clearer labelling guidelines and more distinct training examples. Without this, the model struggles to distinguish between operationally similar or vaguely defined incident types.

Finally, the model fine-tuned on manual labels, without pseudo-labels performed better, and to merge the two sets would introduce an exponential inaccuracy.

5.3.2 Model Interpretability and Semantic Visualization

Figure 5.8 illustrates the two-dimensional representation of sentence embeddings using t-SNE for around 120,000 data points. The visualizations display how the classifier and clustering algorithms have assigned category labels based on semantic similarity in the embedding space.

In subfigure 5.8a , construction part labels predicted by the fine-tuned transformer classifier are mapped in the embedding space. Groupings are somewhat visible for several categories such as *Installation*, *Vägg*, and *Målning*, suggesting that the model has learned to separate semantically distinct construction parts. However, overlap is generally visible between all classes, which may reflect linguistic similarity in issue descriptions.

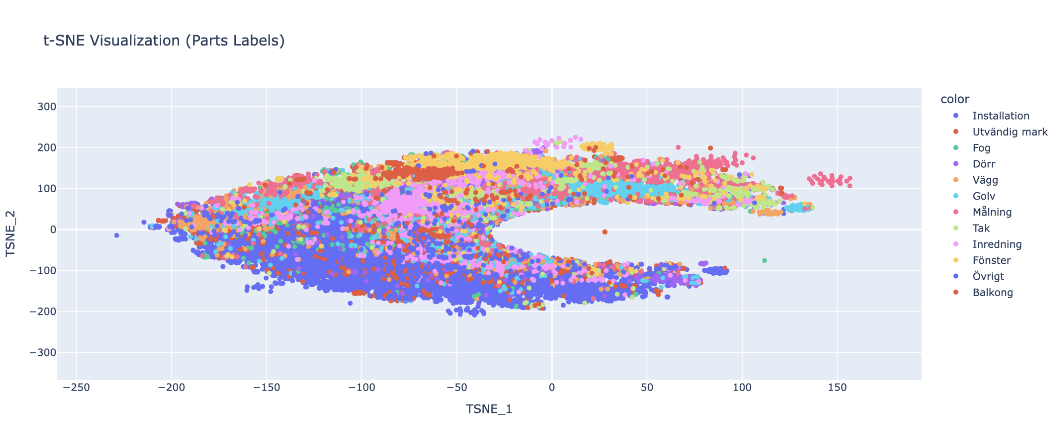
Subfigure 5.8b shows the corresponding mapping for incident type classification. Here, the embeddings appear more mixed, with significant overlap, and general overpresentation by *Byggfel*. This supports earlier findings from the confusion matrix, indicating that these categories are harder to distinguish, likely due to overlapping semantics and major class imbalance.

Subfigure 5.8c presents the result of unsupervised k-means clustering applied directly to the sentence embeddings. The clusters form distinct regions in the embedding space, indicating that the model captures latent structure in the data. However, it is important to note that these clusters are formed purely based on embedding similarity, without regard to actual category definitions. As a result, similar or spatially cohesive clusters in this view do not necessarily correspond to valid or interpretable construction or incident categories. This reinforces the need for domain knowledge and label supervision when translating unsupervised cluster assignments into meaningful classes.

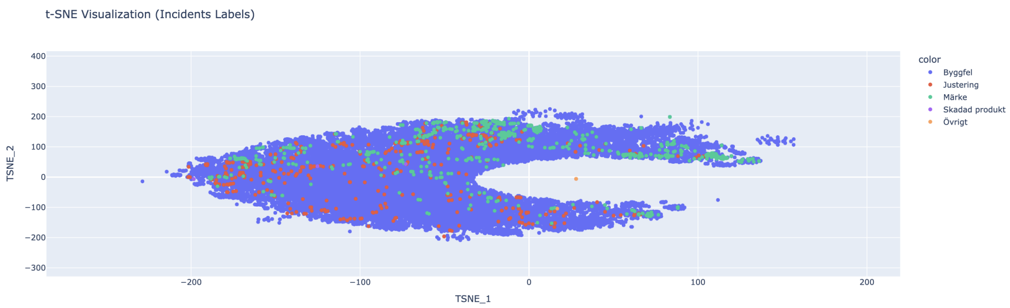
Overall, the t-SNE visualizations provide evidence that the methods capture some structure in the sentence embeddings, although their interpretations and alignment with real-world categories require additional careful evaluation.

5.3.3 Example of Word Importance for Classification

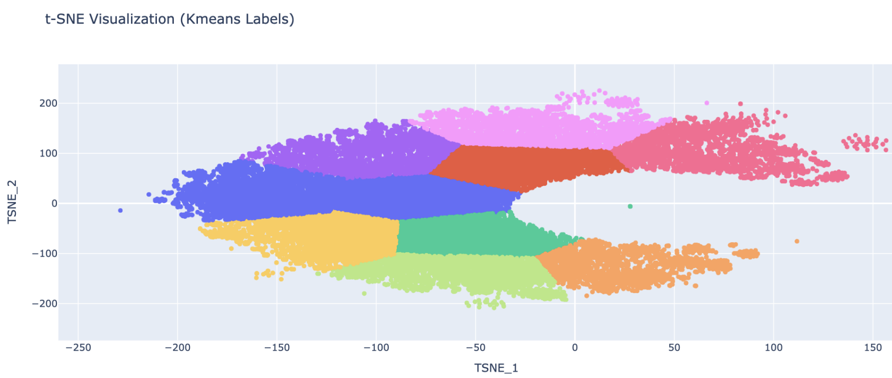
To gain a deeper understanding of the model's decision-making process, attention heatmaps and word relevance plots were analyzed for selected classification examples in Figure 5.9 and 5.10. These visualizations are based on the final attention



(a) Mapped construction part category assignment made by the fine-tuned transformer classifier, based on sentence embeddings.



(b) Mapped incident category assignment made by the fine-tuned transformer classifier, based on sentence embeddings.



(c) Mapped cluster assignment through k-means clustering, based on sentence embeddings.

Figure 5.8: 2D-visualization of the categorized sentence embeddings using t-SNE for the entrie dataset of over 100 000 data-points.

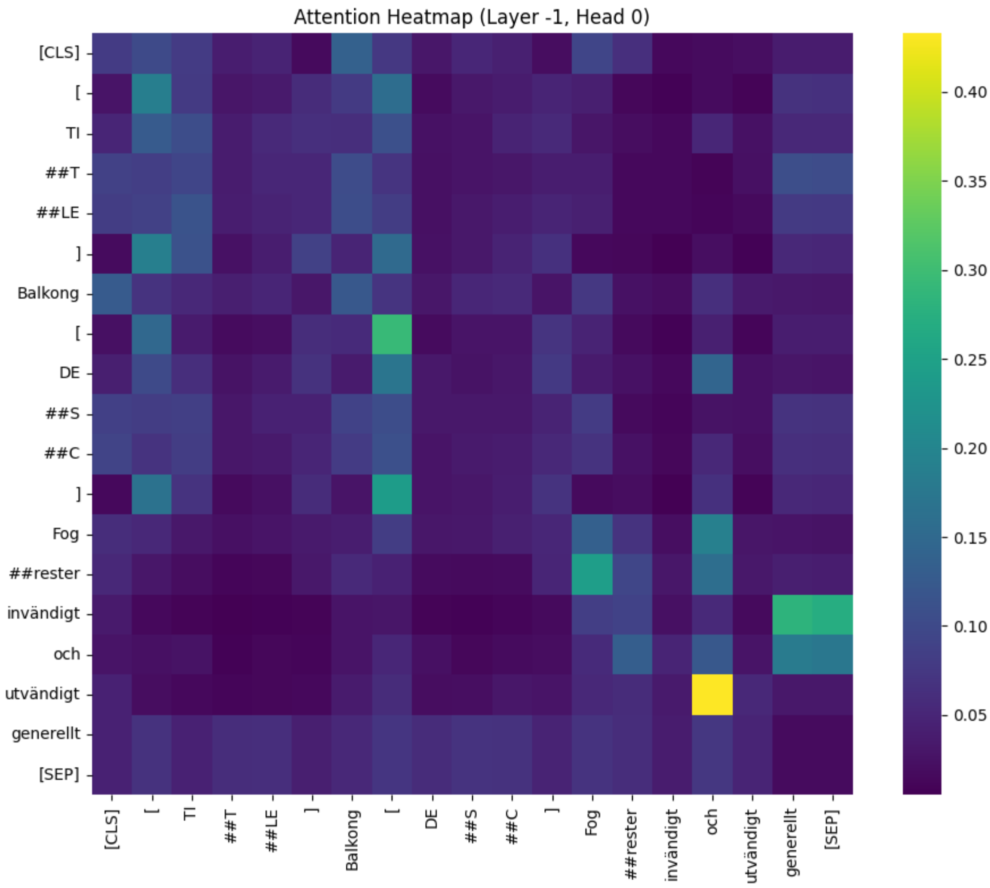
layer of the fine-tuned transformer model and highlight how different input tokens contribute to the predicted label.

In transformer models, attention mechanisms are used to determine which words in a sentence should influence the representation of each word. Self-attention allows the model to assign dynamic weights between words depending on their contextual importance. The final attention layer, in particular, is often interpreted as revealing which tokens the model focuses on most directly before outputting a classification. This provides an interpretable view of which parts of the input text were most influential for the model's decision.

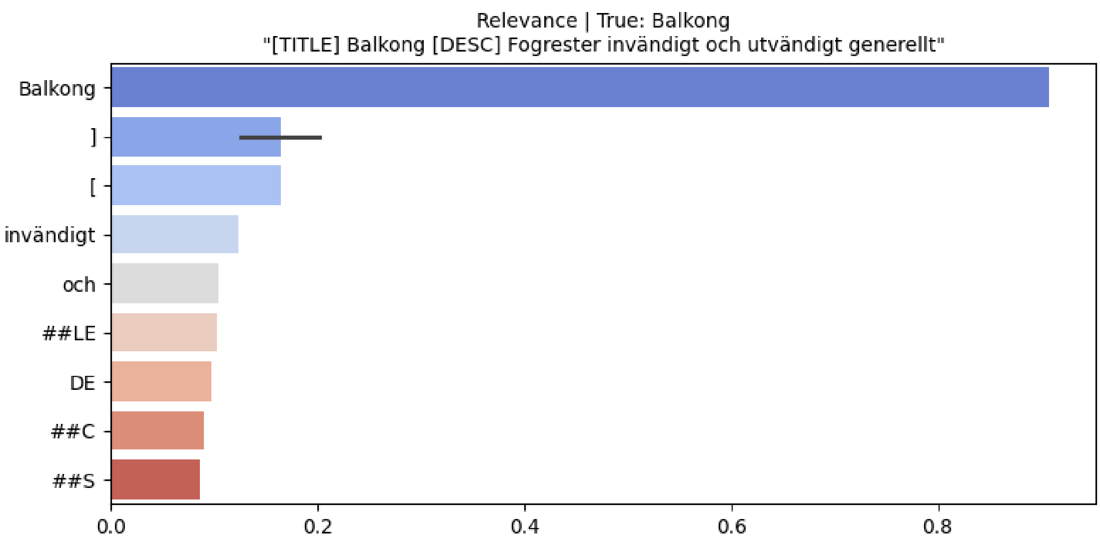
In the first example (Figure 5.9), where the true label is *Balkong*, the attention heatmap in Figure 5.9a, shows an unconcentrated focus on the word "Balkong,", while the accompanying word relevance plot in Figure 5.9b, is assigning the highest score by a clear margin to "Balkong." These relevance scores are derived from attribution methods based on back propagation and reflect how much each token contributed to the final prediction.

In contrast, the second example (Figure 5.10) corresponds to the class *Övrigt*. The input text is a question, "Vilken tjocklek yttervägg ska det vara? Vad ska det vara för tak och vilken tjocklek?", and the attention (Figure 5.10a)is focused on the question-word. Words such as "vilken," "vad," and "tjocklek" each contribute to the decision, all representing question-words (Figure 5.10b) . This reflects a common trend observed in the dataset: many texts labeled as *Övrigt* are phrased as general questions or vague specifications, lacking any strong semantic connection to a specific construction part or issue.

Overall, the attention and relevance visualizations provide insight into how the model processes and distinguishes between well-defined and ambiguous inputs, offering transparency into the internal mechanisms of transformer-based classification.

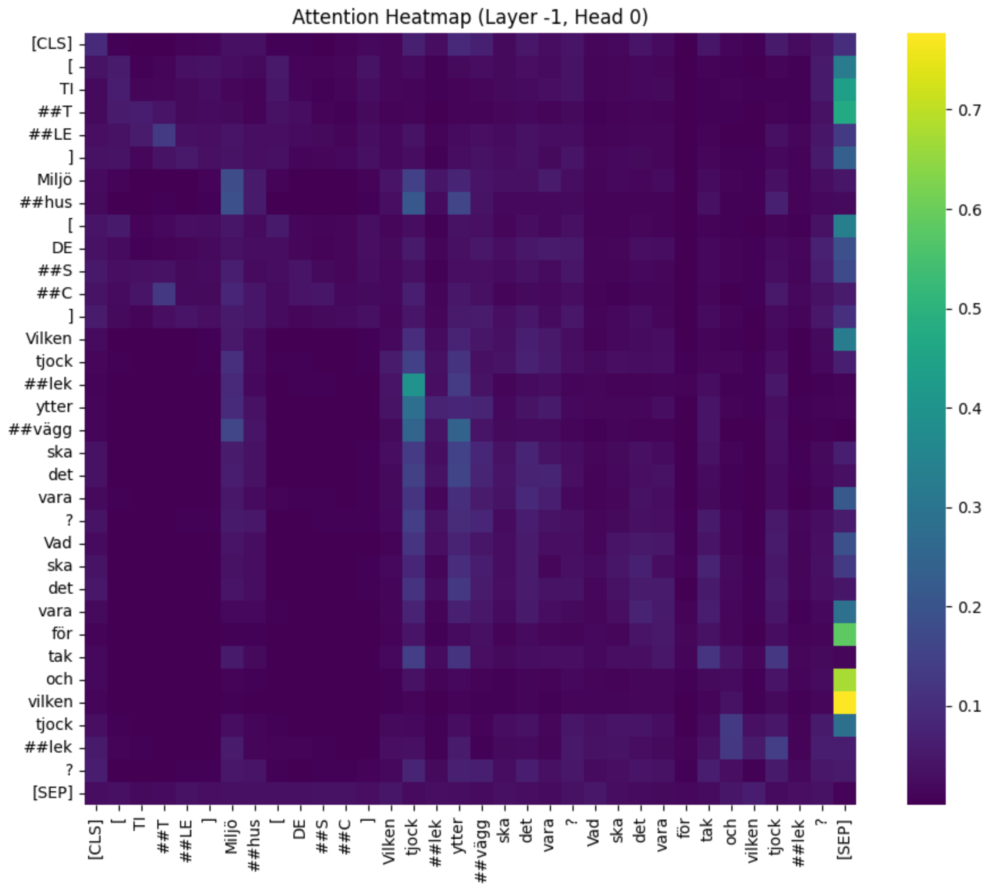


(a) Attention heatmap presenting the last layer before ouput, showcasing how the words pay attention each-other in the sentence: "[TITLE] Balkong [DESC] Fogrester invändigt och utvändigt generellt".

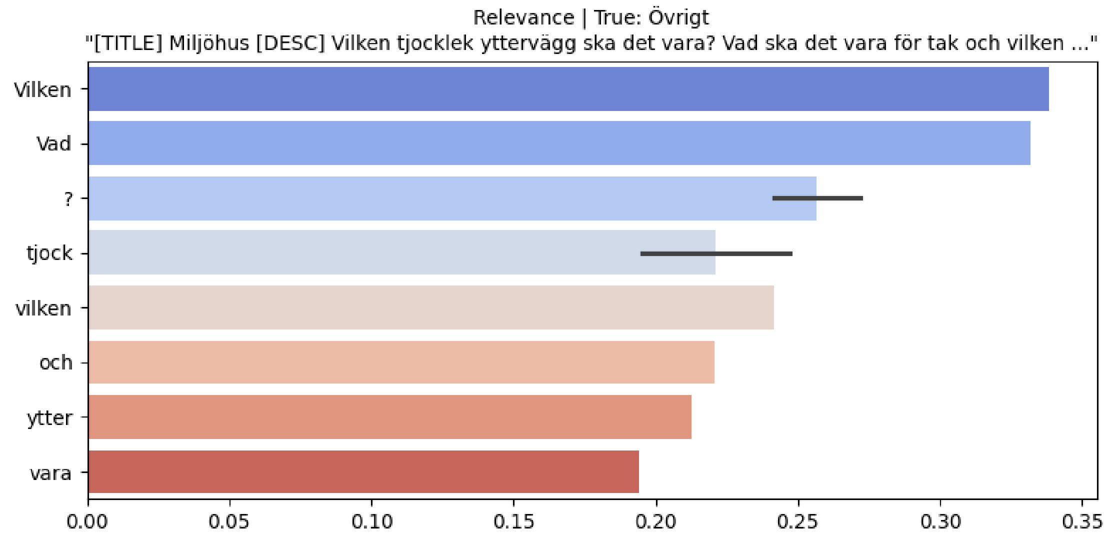


(b) In the sentence: "[TITLE] Balkong [DESC] Fogrester invändigt och utvändigt generellt", which words affect the models categorization decision positively or negatibvely.

Figure 5.9: Figure showcasing the words in the sentence "[TITLE] Balkong [DESC] Fogrester invändigt och utvändigt generellt" and their affect on model and categorization.



(a) Attention heatmap presenting the last layer before ouput, showcasing how the words pay attention each-other in the sentence: "[TITLE] Miljöhus [DESC] Vilken tjoclek yttervägg ska det vara? Vad ska det vara för tak och vilken tjocklek?".



(b) In the sentence: "[TITLE] Miljöhus [DESC] Vilken tjoclek yttervägg ska det vara? Vad ska det vara för tak och vilken tjocklek?", which words affect the models categorization decision positively or negatively.

Figure 5.10: Figure showcasing the words in the sentence "[TITLE] Miljöhus [DESC] Vilken tjoclek yttervägg ska det vara? Vad ska det vara för tak och vilken tjocklek?" and their affect on model and categorization.

6
Discussion

This section reflects on the findings in relation to the research questions and the broader context of AI implementation in the construction industry. The discussion around RQ1: *"What is Skanska's current state of AI readiness, and what organisational strengths and challenges exist in terms of AI adoption?"* highlights how Skanska's AI readiness varies across the organization. For RQ2: *"How can an AI-prototype create value within the department of Quality Management?"*, the focus shifts to the operational and strategic value of the AI prototype, emphasizing its role as a decision-support tool that transforms unstructured data into actionable insights for the quality department. RQ3: *"How can a language model prototype be developed and used to efficiently extract and visualize insights from quality-related construction texts?"* is discussed in terms of the model selection and training strategy along with performance and operational robustness, along with drawbacks. Finally, the section addresses limitations, future opportunities, and how this work contributes to the ongoing digital transformation within construction.

6.1 Discussion on AI Readiness

This section presents the second-level key themes that emerged from the eight original AI readiness themes identified through the interview results. Each theme is discussed in detail and, in some cases, supported by direct quotes. The analysis is followed by actionable recommendations and managerial implications for Skanska Sweden, inspired by change management theory outlined in Section 2.2.1 and the AI adoption strategies discussed in Section 2.2.3, all addressing RQ1.

6.1.1 Organizational Readiness Gaps and Ownership Challenges

Lack of Clear Roles, Ownership, and Follow-ups

A recurring challenge is the absence of defined roles and mandates to drive AI initiatives beyond initial experiments. Interviewees noted that idea generation isn't the problem, "We don't lack ideas. We lack budget, roles, and time to develop those". Rather, the problem lies is the lack in ownership and long-term planning. Promising AI pilots often stall once the pilot phase ends, "when you're done with your project, who owns it?". One manager described how "we try something once, but then it's no one's job to take the next step", commenting that there are "no assigned roles, no budget, no mandate" to continue beyond the experiment. Even though the IT department has offered some support, there is “no clear ownership or follow-up beyond". This gap means AI efforts remain ad-hoc and disconnected, with no accountable person or team to integrate successful AI pilots into daily operations.

Managerial implication: Establish a formal AI governance and ownership structure to ensure follow-through. This could involve assigning dedicated AI product owners or champions for each pilot, with clear mandates and resources to develop pilots into deployed solutions. By securing budget for these roles, Skanska can prevent prototypes from being abandoned and instead turn them into scalable tools. In practice, this might mean creating an internal AI group or center of excellence that owns the roadmap for all AI projects, coordinates between IT and business units, and maintains continuity. Such formal ownership will provide the mechanisms currently missing and ensure that AI innovations are not temporary, but part of a sustained strategy. It also relieves overworked IT teams by spreading responsibilities to dedicated AI leads.

Curiosity Not Turning Into Action
Within Skanska Sweden there is genuine curiosity and enthusiasm about AI, but much of this interest remains abstract, it doesn’t translate into pilot projects or process changes. Multiple interviews highlighted that employees and managers often remain in “wait and see mode" unless it’s made clear how an AI application supports their immediate goals. In other words, people hesitate to act without a tangible link to their own work outcomes. As one respondent stated, “Everyone’s excited, but that doesn’t mean implementation happens.” A clear illustration of this execution gap is that out of 200 hours of AI support available through Skanska’s Microsoft partnership, only one hour was utilized in six months, which is a sign that high-level interest has not led to action. Contributing to this stagnation is a traditionally risk-averse culture, where employees are reluctant to spend time experimenting with AI without clear incentives or leadership direction, so initial curiosity often fizzles out before any concrete pilot is attempted. Another interviewee shared a strategic dilemma, “Should we wait for the software providers to create the AI tools we need, or should we start building and testing these capabilities ourselves?” This highlights the importance of having a strategic direction to ensure that curiosity becomes tangible action.

Managerial implication: Linking AI initiatives to practical real problems, and giving teams licenses to try new tools. Leadership should sponsor a series of small, focused AI pilot projects explicitly tied to daily pain points in construction projects or business processes. It’s critical that management not only encourages these trials, but also allocates time and resources for employees to participate. Furthermore, hands-on exposure is key, and interviewees noted that enthusiasm increases after seeing AI in action during demos or prototypes. Therefore, Skanska should create more opportunities for employees to see and feel AI’s benefits through demonstrations or pilots on real projects. By showcasing benefits and quick wins, the organization can build momentum. In parallel, establishing clear incentives and recognition for teams that pioneer AI solutions will encourage others to move from curiosity to practice. Skanska should also decide whether to rely on external software providers for AI development or to invest in building and customizing AI tools internally.

Confusion and Need For Better Link To Operational Relevance
Another readiness gap lies in conceptual confusion, where many in the organization conflate AI with general “digitalization” and struggle to see how it specifically applies to their operations. Some employees see AI as just another IT tool or a fancy search function rather than a distinct capability for decision support. This confusion of definitions means AI sometimes remains as a buzzword. Without a clear story of practical value, as one interviewee pointed out, “AI remains an abstract concept, saying that ‘AI can help you’ isn’t enough.” Managers might be “positive but don’t really know how to drive it, unless someone explains how it applies in your role”. The workshop discussions highlighted that people need to understand AI in the context of real construction challenges. Awareness and readiness improve when AI is tied to actual pain points on projects. In the current state, this translation to operational relevance is weak, and many employees simply aren’t hearing relatable stories of AI solving problems like the ones they face, which leads to either overestimations of what AI is, or underestimations.

Managerial implication: Demystify AI and ground it in operational reality. Skanska should focus on targeted communication and education that frames AI in the context of work scenarios, clearly communicating its practical value and relevance. A key approach is to highlight role-specific use cases and proof of concepts where early AI pilot successes demonstrate how AI can solve real operational challenges. These concrete examples help demystify AI, showing that it is neither a magic solution nor an abstract concept. Several interviewees suggested the need for workshops or onboarding sessions to translate AI into the language of construction tasks, essentially to “start small, solve one real problem, and then expand",in line with the "Crawling strategy" [3]. Furthermore, creating a group of internal AI champions within different business units would help build understanding on the ground and serve as translators between technical AI teams and other staff, ensuring that each new AI initiative comes with clear explanations of how it works and why it matters for that team’s objectives.

6.1.2 Employee Trust and Safety

Concerns about Blame and Control
Skanska Sweden’s employees have expressed hesitation in documenting issues and deviations due to a fear of blame. Interviews revealed that some workers may be reluctant to register construction project deviations because they worry about being blamed for mistakes. This lack of psychological safety could lead to under-reporting of problems and inconsistent data entry, as individuals remain uncertain about how the data they provide might be used later on. Such under-reporting of issues not only compromises the accuracy of project data but also directly affects the effectiveness of AI tools that rely on complete, high-quality data to generate insights and predictive analytics. In practice, employees need to feel safe and in control when using digital reporting tools. Moreover, there is a risk that if AI systems are

seen as judgmental or too focused on monitoring, employees may become even more reluctant to report data, fearing that their inputs will be used against them rather than to support improvements. Building an environment where reporting data is framed as learning is critical to gain employee encouragement for AI initiatives and to ensure that the data feeding these AI systems remains reliable and comprehensive.

Managerial implication: Establish a no-blame reporting and documenting culture, where leadership can openly encourage staff to log deviations as learning opportunities, and assure them that data from these reports will be used for improvement. This will improve data quality, empowering employees to use AI-driven tools without fear and ensuring that AI systems have access to accurate and comprehensive data. This approach also aligns with the “low-hanging fruits” adoption strategy [3]. By selecting applications that quickly and easily demonstrate AI’s benefits and in this case build confidence and overcome resistance, Skanska can build trust and momentum among employees, helping them see AI as a helpful tool rather than a threat.

Fear of Replacement
Another prevalent concern is the fear that artificial intelligence might replace human roles. Some employees are skeptical about AI replacing their professional experience or judgment. While there is curiosity about AI, workers are far more receptive when they see it as a supportive assistant rather than a threat to their jobs. This perspective resonates with the Technology Acceptance Model (TAM) [14], which highlights that employees’ acceptance of AI depends largely on how useful and easy to use they perceive it to be. In fact, trust and acceptance grow when AI is clearly positioned as an assistant, not a substitute for human expertise. Employees want to retain a sense of control and feel that their knowledge remains valued. The interviews highlighted that AI tools must be transparent and designed to enhance employees’ roles, reinforcing each person’s sense of relevance in the organization. If AI is introduced as a collaborator that handles tedious tasks or provides decision support, staff are more likely to embrace it rather than fear it.

Managerial implication: Clearly communicate that AI solutions at Skanska are meant to augment employees’ work, not to replace their expertise. Involve employees in AI pilot projects and training so they maintain control and see first hand how these tools support them, thereby reducing anxiety about job displacement.

6.1.3 Technical and Data Foundation Gaps

Legacy Systems, Data Silos, and Inconsistent Data Entry
Skanska Sweden’s technical landscape is consisting of legacy systems and siloed data, consisting of inconsistent data. Many functions still rely on manual data input, leading to fragmentation and data quality issues, an issue often voiced as “frustration with the manual and inconsistent ways in which data is entered”. Information is often in separate local Excel files or isolated systems, preventing holistic analysis across projects and reflecting a lack of interoperability between tools. This fragmentation not only hinders AI’s ability to detect patterns and learn from other project’s

data, but also highlights the separation from Skanska Sweden's international departmants that have centralized their data in better ways. One US interviewee noted, "Now we have one system with consistent data flow, which is essential if you want to train reliable models".

Managerial implication: Invest in integrating core systems and establishing company-wide data standards and governance to eliminate silos and ensure data is consistently captured in AI-suitable forms.

Poor Data Alignment for AI
There is a noticeable gap between the Skanska's existing data structures and the requirements of AI initiatives. Much of Skanska's historical project data has been collected strictly for compliance and reporting purposes rather than structured for advanced analytics, severely limiting its value for AI use. In practice, key datasets often lack the detail, format, or accessibility required for algorithms to extract useful insights. Additionally, there is a large amount of knowledge and data that has not been formally registered, often instead residing in the expertise and experience of specific individuals. This creates a risk of critical information becoming siloed or lost. The goal should be to ensure that all important data is accessible and available to everyone, moving away from reliance on individual memory or experience. Here, AI can play a vital role in capturing and systematizing this knowledge, transforming it into shared resources that support better decision-making.

Managerial implication: Embed AI considerations into data collection practices to ensure that critical data is consistently captured and accessible. Also, create ways to document and share the important knowledge currently held by individual employees. This ensures that all valuable data and insights become available for everyone and can support better decisions with the help of AI.

Limited AI Configuration in Current Tools
Even where digital tools exist, Skanska has yet to configure or leverage them for AI-driven use. Interviewees noted that platforms like Power BI are used largely for reporting and lack any advanced or AI-powered analytics, as one respondee said, "We've used Power BI, but not in a way that supports AI". Similarly, current AI assistants (such as Microsoft's Copilot) are not fully utilized, meaning their potential to improve decision-making or automate tasks remains underutilized. This indicates a gap not in technology availability but in its utilization and configuration for AI purposes.

Managerial implication: Provide targeted training and support to help employees fully leverage AI features in existing software, in the construction specific context. This will help demonstrate the practical value of AI and encourage wider adoption across the organization.

6.1.4 Early Signs of Adoption and Competitive Opportunity

Customer Interest and Competitive Signals
Despite no demand for AI from most Swedish clients today, there are early indications that external expectations are beginning to shift. Interviewees noted that some clients have started to request demonstrations of modern digital tools in project deliveries, small but significant signals that AI could soon become a differentiator in winning projects. This shift suggests that competitive pressure may become a powerful motivator for Skanska to adopt AI more systematically. One interviewee summarized this dynamic: "If customers don't ask us to use AI, then what pressure is there to change?", highligting that while internal champions can push pilots forward, true momentum will likely require aligning AI initiatives with client expectations and competitive pressure. Moreover, if competitors begin adopting AI at scale, this will further differentiate those firms and accelerate the need for Skanska to respond proactively. To stay ahead, Skanska must be ready to integrate AI in project delivery as a strategic tool for innovation and market relevance.

Managerial implication: Proactively showcase digital and AI capabilities in client interactions, for example through pilot project demonstrations or tailored proposals that highlight how AI can improve project outcomes. In parallel, Skanska should monitor competitors and client requests to ensure AI efforts respond to the market's demand. This approach aligns with both the "partnership" and "guinea pig" adoption strategies [3], which encourage collaborating with external technology providers to fill capability gaps and accelerate AI adoption, and leverage learnings from early adopters and agile partners in the industry. By actively seeking partnerships and observing their competitors, Skanska can reduce internal risk and position itself for a more robust, strategic integration of AI.

Recognition of data potential
Even in areas where technical data readiness is lacking, there is a growing internal appreciation for the strategic potential of data as a differentiator in the marketplace. Interviews revealed a sense that "we have a lot of data... but it's not structured in the way AI needs", is a sign that employees recognize the current underutilized value of data for future competitive advantage. However, this awareness remains under-leveraged because data is still fragmented and not treated as a strategic asset.

Managerial implication: Define how Skanska's rich data can be used not only for internal operational efficiency, but also for delivering superior client value. Consider data as a competitive asset and elevate data governance and analytics capabilities to a strategic priority.

6.2 Discussion on Value Generating Results by AI-Prototype

The following section presents a discussion of topics related to the results generated by the developed prototype and their value, including modelling strategy, practical applicability, and future potential. The discussion focuses on how the prototype can generate value for the quality department, support decision-making, and contribute to the broader digitalization of the construction industry, covering RQ2.

6.2.1 Organizational Value and Strategic Implications for the Quality Department

The study demonstrates that the AI-prototype developed to support the Quality Department and their deviation management, holds clear and immediate value. A prominent contribution lies in addressing a persistent operational pain point through a proactive approach. Through eliminating manual, time-consuming classification of unutilized text-based data, the findings simultaneously illustrate how AI can be applied within the construction industry context.

A key insight from an interview with a senior quality manager is that the prototype's main value lies not in automation itself, but in making the potential of AI tangible and relevant within the specific context of construction. By transforming raw and unstructured text-based data into structured, categorized data, the prototype serves as both a functional tool and an educational example. It lowers the entry barrier to AI by showing, rather than telling, how AI can enhance existing quality processes. This hands-on exposure is particularly valuable in an industry where AI is often perceived as abstract or overly complex.

The value of this prototype also extends beyond the Quality Department. By offering a tangible example of AI's applicability, it creates a blueprint for other departments and use cases. The prototype acts as a "low-hanging fruit", mentioned by Thelrani et al. [3] as an AI adoption strategy. It is a manageable, clearly scoped use case that can serve as a starting point for broader AI integration. The approach of starting with a small, high-impact problem, starting by focusing on structuring data, and visualizing actionable insights, provides an example case for introducing AI elsewhere in the organization.

In this sense, the prototype functions not only as a classification tool but also as an analytical insight, turning unstructured reporting data into operational knowledge that supports strategic quality improvement. This is particularly valuable for the Quality Department, where such knowledge can support both reactive and proactive risk management. The prototype could also help institutionalize critical construction process knowledge, reducing the bottlenecks caused by over-reliance on a few key individuals' expertise in managing reporting systems and addressing issues.

However, the study also highlights critical limitations and realities of AI in construction. One such insight is that AI is not a “magic solution”, and could still be dependent on human guidance and input. Reliance on human expertise is not ideal in the long term. The goal is rather to move away from human dependence as data structures and processes mature.

In summary, the AI-prototype demonstrates value by simultaneously solving an operational problem and illustrating how AI can be introduced in a complex, low-digitized industry. However, its broader impact depends on addressing data quality and human oversight. This case thus serves as both a proof of concept and a learning exercise for Skanska’s journey towards scalable, meaningful AI integration.

6.2.2 Operational Value for the Quality Department

The AI-prototype enables automatic classification of free-text quality reports, which can reduce manual sorting effort, increase standardization, and provide immediate insights into recurring issue types.

Further, one of the most immediate and impactful outcomes of the project is the structured visibility into the underlying data itself. By organizing free-text reports into well-defined categories and visualizing trends across regions, issue types, and construction parts, the prototype provides the quality department with a new layer of insight, which can help prioritize quality interventions and identify process bottlenecks. This structured view supports existing workflows and decision-making processes by making patterns and deviations more transparent, though quality engineers and operational staff remain responsible for interpreting results, prioritizing interventions, and implementing corrective measures.

This operational impact where AI tools streamline data analysis and reduce manual work, directly reinforces the two main factors in the Technology Acceptance Model: perceived usefulness and perceived ease of use [14]. These real, practical benefits can encourage staff to view AI as an enabler of their daily work, not a burden, thus strengthening willingness to integrate these tools into everyday practice.

The trends identified in the results, such as the uneven duration of issue resolution between categories or the prevalence of inspection-phase reporting, can be used to implement proactive targeted improvement efforts. For instance, longer handling times in categories like *Balkong* or *Utvändig mark* may suggest coordination challenges or delays in resolution workflows. Similarly, the high frequency of vague or question-based inputs being assigned to *Övrigt*, implies further effort in assigning category or directions on how to formulate the cases might be needed, in order to achieve information-dense issue reports. The data also shows that most issues are categorized under Besiktning (Inspection) or Produktion (Production), underscoring the importance of quality assurance during early project phases. These insights could be used to take preventive measures and staff training, enhancing overall process maturity.

6.2.3 Future Opportunities within Quality

Beyond the immediate classification results, the developed AI prototype also presents opportunities for strategic integration within reporting workflows, based on the categorized data and established categories. The model could serve as a real-time support tool that automatically suggests a relevant category at the point of issue reporting. This would represent a shift from a reactive to a proactive approach and illustrate how the prototype can evolve from a backend analytics tool into a user-facing assistant, directly contributing to operational efficiency and data-driven quality management. By embedding classification functionality directly in the reporting interface, users could receive intelligent suggestions for categorization, improving both consistency and completeness of incoming data. This not only supports the reporting user, but also benefits the receiving side (e.g., quality engineers or project managers), who can prioritize, triage, or act more quickly based on structured inputs.

Moreover, besides enhance categorization, further potential within the Quality Department could be to utilizing other text sources in ACC/BIM360. During the early stages of exploring the available data in Snowflake, a "Comment"-section attached to the issues was found in ACC, where internal analyses, explanations and discussions . These text-based inputs were not a part of the company database, and information possibly immensely valuable for the proactive work and transparency, were lost. After highlighting the situation about the comments, and voicing their potential if being added to the database, the Quality department immediately issued a request handling this. With the comments available, categorization of root-causes and measures taken, can after similar procedures, be made available. The minimal human-in-the-loop setup and current pipeline also make the solution scalable across projects and adaptable to new categories with limited retraining.

In summary, while not without limitations, the developed AI prototype demonstrates the feasibility and value of applying natural language processing in construction quality management. It lays the foundation for broader digital integration and more intelligent handling of one of the industry's most underutilized data sources: free-text field reports.

6.3 Prototype Performance and Designing Approach

The following section presents a discussion of topics related to the developed prototype, including its performance, modeling approach, practical applicability, and potential future extensions.

6.3.1 Model Selection and Training Strategy

A central goal in this thesis was to explore effective yet practical methods for text classification in the construction domain. Given the applied context, the aim was not only to achieve accurate results but also to identify a training strategy that was relatively simple, efficient, and reproducible under real-world constraints.

Early in the project, various options were considered, including weakly supervised methods such as pseudo-labelling and exponential moving average bias correction. Intentionally the amount of human-in-the-loop work during initial prototyping was minimized, with the aspiration of being entirely unsupervised. However, these techniques introduced risks of propagating noise as seen in Section 5.3.1. To maintain model transparency and reduce training complexity, these alternatives were excluded in favour of a more straightforward supervised learning approach.

The choice to use a fine-tuned transformer model (e.g., XLM-R or KB-BERT) was motivated by the availability of general pre-trained language models capable of adapting to domain-specific tasks with limited labelled data. In our case, although only a small fraction of the full dataset was actually manually labelled, through oversampling, 58 category-examples across around 120,000 text records, the model performed well. Hence, the model had to generalize from relatively few examples per class, making transfer learning especially valuable.

Overall, the modelling strategy prioritizes simplicity and operational applicability over algorithmic complexity. The results demonstrate that even without elaborate pipelines or manual feedback loops, it is possible to achieve robust categorization in a domain as unstructured and variable as construction text reporting—provided that the model is fine-tuned and evaluated.

6.3.2 Model Performance and Operational Robustness

An essential question in this thesis is whether the trained classification model is robust enough to be applied in real-world operations. The evaluation results in the form of confusion matrices, indicate that the model performs reliably on dominant and well-defined categories such as *Installation*, *Vägg*, and *Målning*. These categories show consistent predictions, even across somewhat varied sentence structures, suggesting that the model has learned generalizable linguistic features associated with these parts. In general, the mean accuracy is between 67% and 65%, but internally between 92% and 28%. Only 6‰was used for fine-tuning with manually labelled data, which shows further promising potential. Visualization tools such as t-SNE mappings, attention heatmaps, and word relevance scores provide transparency into the model's decisions. The ability to trace predictions back to input terms adds a layer of explainability that is crucial in operational contexts, where trust and usability are key.

However, performance is weaker for ambiguous classes, such as *Utvändig mark* and *Övrigt* or certain overlapping categories. This reflects the lack of strong semantic anchors in some texts. In addition, the model struggles more when faced with vague descriptions or inputs containing multiple themes. These are quite common in free-text field reporting and pose a general challenge for classification tasks in unstructured domains.

Domain-specific pre-training on Swedish construction data and improvements in label design could further enhance performance. Equally important is a strong focus on the labelling process itself. Well-separated and clearly defined categories, aligning semantic meaning with operational utility, supported by consistent annotation guidelines, play a critical role in reducing subjective bias and ensuring continuity across labelling efforts. Such efforts not only improve data quality but also directly impact model robustness and reliability in deployment scenarios. In this project, even with expert input, certain ambiguities and overlapping concepts made manual annotation challenging, which might have affected manual labelling and validation, as these have been strongly connected to individual experience and a smaller group of people in the loop.

From a robustness perspective, the model shows sufficient consistency and semantic sensitivity to be considered deployable, particularly as a decision-support tool rather than a fully automated system. In practice, its integration would likely enhance accuracy and standardization in issue categorization, especially when used to support rather than replace human judgment.

6.3.3 Language and Domain Adaptation through Transfer Learning

The model developed in this thesis leverages transfer learning by fine-tuning a pre-trained transformer language model on a domain-specific dataset from the construction industry. This approach significantly reduces the amount of labeled data required to achieve acceptable performance, as the model already possesses a general understanding of language structure and semantics. However, transferring knowledge to a niche application domain such as Swedish construction reports still presents challenges.

Firstly, most pre-trained models are developed on general-domain corpora (e.g., news, Wikipedia, web data) and often in English. Even though multilingual models like XLM-R or Swedish-specific models such as KB-BERT were used, they may not fully capture the nuances of technical vocabulary, abbreviations, or reporting styles found in real-world construction data, confirmed during the workshop for manual labelling. Combined with the free-text nature of the input, this results in a possible mismatch between model assumptions and domain characteristics.

Secondly, language-specific issues in Swedish, such as compound words, inflection, and syntactic flexibility, can reduce model confidence and consistency in prediction. Domain-specific expressions or project-specific jargon further complicate this. While the model showed good performance in identifying well-defined categories like *Instal-*

lation or *Målning*, more abstract or inconsistent classes like *Utvändig mark* proved harder to separate.

These observations reinforce the value of continued domain adaptation. Further improvements could involve targeted pre-training on Swedish construction data (e.g., reports, standards, manuals) and expanding the labelled dataset with feedback from domain experts. Overall, transfer learning remains a powerful tool for adapting general language models to specialized industrial applications, but it requires careful handling of linguistic and contextual nuances.

Despite these challenges, the model demonstrates how categorization can be meaningfully supported by AI. With structured data, a defined label taxonomy, and a feedback mechanism to refine predictions over time, such a system could support both field users and data receivers by improving reporting accuracy and operational efficiency.

7
Conclusion

In response to RQ1, the study finds that Skanska Sweden shows a growing awareness and curiosity around AI, which is supported by large volumes of historical data available. Additionally, there is potential to learn from international practices within the company. However, although large volumes of data are available, there is a lack of structured, complete, and well-labeled data suitable for AI use. There is also further need for AI-specific expertise, and it remains challenging to integrate new tools into established workflows. Managerial implications include the risk that initiatives remain siloed due to limited infrastructure, highlighting the importance of AI training, cross-functional collaboration, and clarifying ownership structures.

Regarding RQ2, the prototype demonstrated how AI can create tangible value by structuring and visualizing previously unstructured quality issue text data. The use of expert-driven categories, for building parts and for incident type, enabled a meaningful categorization aligned with real-world practices. These insights can support proactive quality control, trend identification, and communication across teams. The analysis suggests that the Quality Department sees value in the prototype's ability to deliver structured, actionable insights from untapped text data, that directly addresses their initial goal of better understanding what happens in the construction process. This reinforces the prototype's potential as both a proof of concept and a foundation for broader AI adoption within Skanska Sweden.

For RQ3, a practical and scalable prototype was successfully implemented, capable of extracting and visualizing insights from free-text quality reports. While the goal was to focus on weakly supervised learning thorough pseudo-labelling, this approach proved challenging due to data noise, class imbalance, and label ambiguity. As a result, a method combining unsupervised clustering, and domain expert manual annotation to fine-tune a transformer model was adopted. Using only a set of 6‰labelled data the accuracy of the model categorization was measured to 67%. The work demonstrated the possibilities of reducing manual effort, visualization for interpretability and sufficient performance.

Overall, this thesis illustrates how an AI prototype can contribute to improved quality insights. By aligning organizational AI readiness with tailored technical application, construction companies can further move toward data-driven decision-making. Future work could focus on refining label granularity, leveraging domain-specific pre-training, and enhancing usability to further support adoption at scale.

Bibliography

[1] J. L. Blanco, S. Fuchs, and M. J. Ribeirinho, "Artificial intelligence: Construction technology's next frontier." https://www.mckinsey.com/~/media/McKinsey/Industries/Capital%20Projects%20and%20Infrastructure/Our%20Insights/Artificial%20intelligence%20Construction%20technologys%20next%20frontier/Artificial%20intelligence-Construction-technologys-next-frontier.pdf, 2018. McKinsey Company, Accessed: 2025-03-28.

[2] S. O. Abioye, L. Oyedele, M. Akanbi, J. M. Delgado, A. Ajayi, O. Akinade, M. Bilal, E. Bamidele, H. Owolabi, and K. Munir, "Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges," *Journal of Building Engineering*, vol. 44, p. 103299, 2021.

[3] A. Tehrani, "From ai to digital transformation: The ai readiness framework," *International Journal of Engineering Applied Sciences and Technology*, vol. 6, no. 3, pp. 50–54, 2021. Accessed: 2025-03-31.

[4] Z. You and L. Feng, "Integration of industry 4.0 related technologies in construction industry: A framework of cyber-physical system," *IEEE Access*, vol. PP, pp. 1–1, 07 2020.

[5] "The rise of artificial intelligence in construction," *Construction Today*, 2024.

[6] J. L. Blanco, D. Rockhill, A. Sanghvi, and A. Torres, "From start-up to scale-up: Accelerating growth in construction technology." https://www.mckinsey.com/industries/private-capital/our-insights/from-start-up-to-scale-up-accelerating-growth-in-construction-technology#/, 2023. McKinsey Company, Accessed: 2025-03-28.

[7] Skanska Group, "Year-end report 2024." https://group.skanska.com/490770/siteassets/investors/reports-publications/interim-reports/2024/q4-2024/skanska-q4-2024-sv.pdf, 2024. Accessed: 2025-05-29.

[8] S. Group, "About us." https://group.skanska.com/about-us, 2024. Accessed: 2025-04-05.

[9] S. Sverige, "Vår verksamhet." https://www.skanska.se/om-skanska/vaar-verksamhet, 2024. Accessed: 2025-04-05.

[10] P. M. Bosch-Sijtsema, P. Gluch, and K. Johansson, "Data-driven construction: Impacts of data and digitalization on construction projects and companies," *Construction Management and Economics*, vol. 38, no. 11, pp. 1075–1091, 2020.

[11] H. Li, W. Lu, F. Xue, J. Xu, K. Chen, and X. Chen, "Digital transformation in the architecture, engineering, and construction industry: A review and future directions," *Automation in Construction*, vol. 135, p. 104131, 2022. Accessed: 2025-05-28.

[12] W. S. Alaloul, M. S. Liew, N. A. W. A. Zawawi, I. B. Kennedy, and B. S. Mohammed, "Change management practices for adopting new technologies in the design and construction industry," *Engineering, Construction and Architectural Management*, 2020. Accessed: 2025-03-31.

[13] J. P. Kotter, *Leading Change*. Harvard Business Press, 1996.

[14] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989. Accessed: 2025-03-31.

[15] J. Holmström, "From ai to digital transformation: The ai readiness framework," *Business Horizons*, vol. 65, no. 3, pp. 329–339, 2022. Accessed: 2025-03-31.

[16] M. Chui, J. Manyika, M. Miremadi, N. Henke, R. Chung, P. Nel, and S. Malhotra, "Notes from the ai frontier: Applications and value of deep learning," 2018.

[17] G. Cloud, "What is machine learning (ml)?," 2024. Accessed: 2025-02-17.

[18] J. Kanjilal, "An introduction to weakly supervised learning," 2021. DigitalOcean, Accessed: 2025-04-28.

[19] D. Bergmann, "What is zero-shot learning?," 2023. IBM, Accessed: 2025-04-28.

[20] J. Holdsworth and M. Scapicchio, "What is deep learning?," 2023. IBM, Accessed: 2025-04-28.

[21] J. Barnard, "What is embedding?," 2023. IBM, Accessed: 2025-05-28.

[22] M. Kapronczay, "A beginner's guide to language models," 2025. Built In, Accessed: 2025-04-28.

[23] S. Vaniukov, "Nlp vs llm: A comprehensive guide to understanding key differences," 2024. Medium, Accessed: 2025-04-28.

[24] P. Kargupta, T. Komarlu, S. Yoon, X. Wang, and J. Han, "Megclass: Extremely weakly supervised text classification via mutually-enhancing text granularities," *arXiv preprint arXiv:2304.01969*, 2023. Accessed: 2025-04-28.

[25] E. Kavlakoglu and V. Winland, "What is k-means clustering?," 2024. IBM, Accessed: 2025-04-28.

[26] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018. Accessed: 2025-04-28.

[27] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," *arXiv preprint arXiv:1908.10084*, 2019. Published at EMNLP 2019, Accessed: 2025-04-28.

[28] F. Rekathati, "The kblab blog: Introducing a swedish sentence transformer," 2021. Accessed: 2025-04-28.

[29] Dataloop, "Paraphrase multilingual minilm l12 v2," 2024. Accessed: 2025-04-28.

[30] M. AI, "Xlm-r: State-of-the-art cross-lingual understanding through self-supervision," 2019. Accessed: 2025-04-28.

[31] B. Admin, "The relevance of bert and spacy in the age of emerging language models," 2024. Beige Bananas, Accessed: 2025-04-28.

[32] M. C. Data, "What is data completeness?," 2023. Accessed: 2025-06-09.

[33] D. Henschen, "Ai's appetite for data is changing data requirements," 2024. Accessed: 2025-06-09.

[34] BytePlus, “What is relevance information in ai? a comprehensive guide,” 2024. Accessed: 2025-06-09.

[35] AIMultiple, “Data quality for ai: Thefivedimensions of data health,” 2024. Accessed: 2025-06-09.

[36] P. AI, “Goodbye noise, hello signal: Data validation methods that work,” 2024. Accessed: 2025-06-09.

[37] IBM, “What is ai bias?,” 2023. Accessed: 2025-06-09.

[38] L. K. H. Clemmensen and S. K. Kjærsgaard, “Representativity in machine learning and data science,” *arXiv preprint arXiv:2203.05872*, 2022.

[39] E. Bell, A. Bryman, and B. Harley, *Business Research Methods*. Oxford University Press, 5th ed., 2018. Accessed: 2025-03-31.

[40] S. Alänge, “The affinity-interrelationship method (aim): A problem solving tool for analysing qualitative data,” tech. rep., Chalmers University of Technology, 2009.

A

List of Interviewees

This Section presents the interviewees and workshop participants who contributed their insights to the study, along with their respective roles and organizational affiliations within Skanska or as external experts.

Interview and Workshop Participants	
Role	**Country / Organization**
Operations Manager	Skanska Sweden
Digital Leader	Skanska Sweden
Digital Leader	Skanska Sweden
Project Design Manager	Skanska Sweden
Production Manager	Skanska Sweden
Development Leader	Skanska Sweden
Business Developer	Skanska Sweden
Quality Manager	Skanska Sweden
Quality Manager	Skanska Sweden
Quality and Aftermarket Manager	Skanska Sweden
Regional Manager	Skanska Sweden
Estimation Manager	Skanska Sweden
Estimation Engineer	Skanska Sweden
Digitalisation Manager	Skanska Sweden
AI Researcher	External Expert
Director of Business & Strategy	Skanska USA
Chief EHS Officer	Skanska USA
AI Capability Manager	Skanska UK & Group

Table A.1: Participants' roles and their organizational affiliations within the Skanska group and external experts.

B
Additional Interview Guide

This interview guide was used in structured follow-up interviews with representatives from Skanska Sweden, US and UK, to assess organizational AI readiness. The guide is based on the AI Readiness Framework by Tehrani et al. [3] and is structured the around eight key categories: Environmental, Technological, Informational, Infrastructural, Data, Participants', Customers', and Process' Readiness.

The main questions below were used to explore each category, and follow-up questions were asked as needed, depending on the interviewee's role and responses.

1. **Environmental Readiness:** How would you describe the general attitude toward AI within your department and within Skanska overall, both at leadership and operational levels?
2. **Technological Readiness:** To what extent are your current tools, systems, or platforms prepared to support the implementation of AI?
3. **Informational Readiness:** Do managers and decision-makers in your organization have a good understanding of how AI could be used in practice?
4. **Infrastructural Readiness:** Do you feel that your department has the necessary human, financial, and technical resources to support AI initiatives?
5. **Data Readiness:** How would you assess the quality, availability, and structure of the data in your domain for potential AI use?
6. **Participants' Readiness:** How do you think employees and external stakeholders such as subcontractors would respond to working with AI tools?
7. **Customers' Readiness:** In your experience, how open are clients or project stakeholders to adopting AI-based tools or solutions in projects?
8. **Process Readiness:** Are there any current workflows or processes where you think AI could be integrated to improve efficiency or decision-making?

C

Interview Results

This Section presents the full summaries of the respondents' answers, overarching themes, and selected quotations, from the six structured additional interviews.

Environmental Readiness

Among Swedish respondents, a cautiously optimistic view toward AI emerged. While curiosity and openness were frequently mentioned, interviewees highlighted significant variability in awareness, coupled with uncertainty about how AI applies in practice. Respondent 1 described the general mood as curious but fragmented. She noted that tools like Copilot and Sidekick are technically available, but many employees remain unaware of them or unsure how to apply them. "Skanska has a tool called Sidekick and also Copilot... but I don't know how many are aware of it or actually use it in their daily work." She emphasized that much of the communication around AI tools is delivered digitally, for example via Teams forums, but added: "It varies a lot. The information has been shared digitally, but I don't know how many read or engage with it." She also described a much of the communication around AI tools is delivered digitally, for example via Teams forums. Respondent 2 also identified a sense of openness tempered by inexperience. "Curiosity and interest, I would say. But also some hesitation, because it's something we haven't really worked with before." He suggested that younger employees tend to be more digitally inclined, while older staff might show greater skepticism. However, he credited Skanska's IT department for offering basic training and examples to raise awareness: "The IT department helps us by pushing out some trainings and examples of what we can do."

Respondent 3 provided further context, emphasizing that while leadership now discusses AI more actively, many employees remain in a "wait and see" mode. He noted that AI is still misunderstood by many, often associated with robotics or job replacement, rather than data-supported decision-making. "People are interested, but unless it's clear how it supports their own goals, they wait and see." He also remarked that Skanska Sweden's culture is typically risk-averse and not oriented toward experimentation for its own sake, making leadership endorsement and clear incentives crucial.

Workshop participants also reinforced these themes. They spoke of an open mindset but cited limited time and capacity to explore new tools without clear incentives or guidance. Across these Swedish voices, a common thread emerged: there is space for AI to grow within the organization, but it requires better communication, more tailored examples, and leadership that connects innovation to operational reality. International comparisons highlighted a more embedded perspective on AI in the US and UK contexts. Respondent 5 (US) described AI as being deeply integrated into strategic planning at Skanska USA Building. “We have a team focused on emerging technology and data. It’s a part of our broader business planning function.” She emphasized that AI is viewed not as futuristic, but as a practical tool to improve performance and reduce risk on active projects. Respondent 4 (US) presented AI as the natural progression of Skanska USA Civil’s long-term digital journey. “We digitized our management systems decades ago. Now, AI is just the next step , it’s part of the evolution, not a revolution.” His view positioned AI as an extension of existing safety and operational frameworks, highlighting a performance-driven culture where tools are expected to deliver measurable results.

Respondent 6 (UK) offered a group-level view of AI. She explained that strategic enthusiasm around AI is widespread across global business units, particularly among leaders in IT, business development, and innovation. However, she acknowledged that this interest has not always translated into action: “Everyone’s excited, but that doesn’t mean implementation happens.” For example, she cited that Skanska had only used one hour of the 200 hours of AI support available through Microsoft. Her reflections illustrate a gap between enthusiasm and execution, a challenge also present in the Swedish context.

Technological Readiness
Technological readiness within Skanska Sweden is characterized by fragmentation, legacy systems, and a general lack of integration between tools and AI capabilities. Across interviews, respondents consistently described digital infrastructure that is functional for documentation and compliance, but not built with AI in mind. Respondent 1 noted that "we need more structured information if we want AI to work," adding that historical data was collected for reporting purposes, not to be parsed by algorithms. She also commented that while tools like Power BI are in use, they are not configured to support machine interpretation or learning: "We’ve used Power BI, but not in a way that supports AI. It’s not really about advanced analysis."

Respondent 3 also shared this limitation, stating that current systems "aren’t built for machine learning yet, they’re built for documentation, not prediction." He explained that many tools were purchased with compliance in mind, and thus lack the interoperability and structured forRespondent 3 needed to train AI models. Respondent 2 offered a similar view, explaining that although Skanska Sweden has an "incredible number of tools," most are not equipped for AI enhancements: "Translating that into improving them with AI... we’re probably not quite there yet." However, he saw potential in leveraging existing systems more effectively, noting that Copilot, for instance, allowed him to paste text and receive well-formulated proposal drafts , a task that previously required more manual effort.

Several respondents noted that these early trials with tools like Copilot are promising, suggesting a willingness to explore how AI might enhance existing systems. While these efforts are not yet structured at scale, they provide useful internal examples of potential integration. In the Swedish context, a common theme was the difficulty of retrofitting existing systems. As Respondent 2 put it, "Had we known earlier, we would have structured everything differently. Now, we're trying to reverse-engineer a system that was never designed for this." This sentiment was also shared during workshops, where employees described frustration with the manual and inconsistent ways in which data is entered, especially within deviation and quality reporting. Despite these limitations, interviewees also acknowledged that the organization's broad digital footprint , with many tools already in place , could be a future advantage. With the right coordinaition, these systems could form the basis for AI-enhanced processes.

In contrast, international respondents described a more mature technological landscape. Respondent 5 (US) highlighted that Skanska USA had proactively built a data infrastructure , a centralized warehouse that supports various AI tools. "We built a data infrastructure , connected our systems, created a warehouse. It's all internal." She shared that internal chatbots, trained on proprietary operational risk data, are already in use: "You can ask the bot what the top five risks are for a given crane operation, and it will give you references to similar projects and best practices." Respondent 4 (US) described the Civil division's internally developed software platform, built over 30 years, which now supports AI-driven analytics. "We couldn't find a system that worked for us, so we built one. It now supports over 11,000 companies and 85,000 work crews." His team has integrated elements of machine learning and language models into construction workflows, providing a level of customization and control that the Swedish context currently lacks.

From a group-level perspective, Respondent 6 (UK) acknowledged that Skanska has licensed several AI-enabled tools (e.g., Copilot, Autodesk), but actual usage remains low. "Within our Microsoft agreement, we have 200 hours of AI support. After six months, we had used one." She emphasized that technical capability is not enough , business units must also build workflows around the tools, which has not yet occurred at scale. Nevertheless, the Swedish IT department has taken steps to support AI exploration by sharing internal examples and encouraging small-scale experimentation, as noted by Respondent 2 and others. These efforts, though limited, demonstrate an emerging technical support function that could grow with clearer direction.

Informational Readiness

Informational readiness within Skanska Sweden is marked by a high-level curiosity about AI, but limited clarity on what AI means in practice. Several respondents noted that while interest in AI is visible among decision makers, there is a gap between strategic ambition and operational understanding from leadership. Respondent 2 explained, "There is interest and willingness. But the understanding , that's something else, I think." He emphasized the need for internal champions who can translate AI into day-to-day language and action: "We need people with digital leadership who can help guide the rest of us."

This was shared by Respondent 1, who pointed out that while department managers may support AI conceptually, there is a lack of practical examples or guidance. She said, "Our department manager is interested, and we've had some discussions. But we need examples , what does AI really mean in our context?" She argued that without use-case-driven communication, AI remains an abstract concept: "You'd need some sort of workshop or onboarding. Just saying 'AI can help you' isn't enough." She also reflected that "managers are positive but don't really know how to drive it. It becomes a buzzword unless someone explains how it applies in your role."

Respondent 3 also highlighted the confusion around AI terminology, noting that many leaders confuse it with broader digitalization or advanced search functions. He stated, "We need to demystify AI. Start small, solve one real problem, and then expand." In his view, storytelling, internal vocabulary, and simple use cases are essential to bridge the understanding gap and move from interest to application. Despite these gaps, interviewees agreed that awareness has increased, particularly as digital discussions become more common across departments. Managers are more exposed to AI-related topics, and internal forums have helped introduce the idea of AI to wider groups.

Respondent 2 also noted that when AI is tied to operational pain points, informational readiness improves significantly. He pointed to the deviation classification prototype as a strong example: "It was good because it was based on something real. That's what makes it useful." He emphasized that "we don't need more presentations. We need something that tells us: this is what you can do with AI, in this exact context."

In contrast to Sweden, the US and UK interviewees described more developed informational infrastructures. Respondent 5 (US) explained that in her business unit, AI is no longer perceived as experimental: "It's not a pilot thing anymore. It's operational, at least for certain use cases." Her team includes data scientists, analysts, and data managers who help shape a shared understanding of how AI can support day-to-day planning and risk management. The presence of these embedded roles contributes to a higher level of practical literacy around AI among managers. Respondent 4 (US) also stressed that understanding of AI must be grounded in real-world outcomes. Drawing on his academic experience, he said, "People don't get AI until they see it solving something in their environment. We use real data, real outcomes." He underscored that alignment with performance culture and safety management reinforces AI's value, especially when communicated through ISO standards and the Deming cycle, both of which offer familiar frameworks to operational teams.

Respondent 6 (UK) provided a group-level view, describing initiatives such as the Malmö AI meetup, where 76 use cases were collected to showcase how AI is already being applied across Skanska. While she acknowledged the challenge of converting awareness into action, "We're presenting ideas, but few are moving into operational phase" , she also emphasized that visibility and peer learning were improving: "We're showing them what others are doing and asking if they want to join." She leads both technical and business AI forums that promote internal sharing of use cases and cross-unit learning, helping to create visibility and normalize experimentation.

Infrastructural Readiness
Infrastructural readiness emerged as a critical constraint in Skanska Sweden's ability to move beyond AI experimentation. Interviewees consistently noted that while digitalization is an active ambition, the formal structures, roles, and resources necessary to support AI are lacking or underdeveloped. Respondent 1 explained that while AI is a frequent topic of conversation, there is no designated ownership or responsibility: "We talk about AI, but there's no one assigned to make it happen. It's not a formal task." She noted that even when theoretical resources might exist, they are not activated in practice, due to competing priorities and limited time.

Respondent 2 shared this concern, stating: "We don't lack ideas. We lack budget, roles, and time to develop those ideas." He's essentially cautioning that without clear roles, dedicated funding, or long-term plans, the AI work risks becoming temporary and not integrated into the broader organization, "What happens when you're done with you thesis? Who owns it?" He stressed the need for structured handover, ownership, and internal resourcing, warning that promising prototypes may be abandoned without continuity mechanisms. He observed that IT and digital roles are currently stretched managing existing tools, with little capacity left to support new AI initiatives: "They're also trying to manage all the tools we already have." He added that although Skanska Sweden has a wide array of tools available, the bridge between tool usage and AI enhancement is still missing.

Respondent 3 similarly identified a lack of follow-up structures for AI experimentation: "We try something once, but then it's no one's job to take the next step. That's where it breaks." He stated clearly that his department does not currently have the resources to work meaningfully with AI. "We've only just started to touch on these questions. There are no assigned roles, no budget, no mandate." He explained that they are still in the exploratory phase and that leadership interest has not yet translated into structural support. AI work, according to Respondent 3, would require not only IT tools but also skilled data engineers and support to clean, process, and understand the data involved.

While IT competence exists in some places, there is no broader plan to activate those capabilities. AI efforts have so far been experimental and disconnected: "There's no continuous structure or next step. We experiment, but then it's over." Despite these structural gaps, respondents pointed to a few enabling elements. Respondent 2 noted that the IT department had started to support training and example sharing, suggesting a latent support structure that could be scaled. Respondent 3 also mentioned that the sustainability and quality departments have taken steps to coordinate AI efforts, although informally. These early signs of alignment point to potential nodes of infrastructure that, with investment, could support more sustainable development.

In the US and UK, interviewees described more mature infrastructures. Respondent 5 (US) explained that her strategy team works closely with business development and technology functions, creating a structure that connects AI development to strategic planning and project delivery. “The people in our strategy function sit next to data and emerging tech. That helps connect strategy with delivery.” This cross-functional setup allows for experimentation and scaling without reliance on external mandates.

Respondent 4 (US) highlighted a long-term unification effort in Skanska USA, which consolidated over 30 independent management platforms into one system: “Now we have one system with consistent data flow, which is essential if you want to train reliable models.” He argued that infrastructure should not only include IT systems, but also governance, workflow alignment, and ownership models. The internal software platform developed over decades is now capable of supporting embedded AI functionality across thousands of subcontractors. He also emphasized that internal investment, not outsourcing, had been key to long-term continuity. Respondent 6 (UK) emphasized that group-level infrastructure focuses on enablement rather than ownership. Her role involves connecting business units, avoiding tool duplication, and showcasing working examples from within the organization. While she acknowledged uneven maturity across business units, she emphasized the need to “help teams gain value from what they already have,” rather than build redundant systems.

Data Readiness
Data readiness is a recurring theme across all interviews, and it presents both a foundational opportunity and a critical barrier for AI adoption at Skanska Sweden. Interviewees highlighted a strong tradition of data collection, especially in areas such as sustainability, quality, and safety, but emphasized that data is often unstructured, inconsistently entered, and not easily usable for AI applications. Respondent 1 summarized this challenge clearly: “We have a lot of data, but it’s not structured , not in the way AI needs.” She noted that most data was originally collected for documentation and compliance, not for learning or machine interpretation: “No one thought the data would be used by an AI, so it was never set up that way.” She also observed that while tools like BIM360 and ACC house large volumes of information, the lack of standardization and labelling limits their analytical potential. Although some efforts have been made using tools like Power BI, these remain at the level of static dashboards and lack the dynamic structure needed for AI.

Respondent 2 shared these concerns, pointing out that data is often stored in local Excel sheets or siloed within individual tools, making aggregation and cross-project analysis difficult: “Some of it is stuck in Excel... but the trend is to start aggregating and centralizing it.” He described Skanska Sweden’s internal culture as responsive to data issues, stating: “We have a culture in the company where we don’t just keep quiet , we speak up and suggest improvements.” However, he emphasized that this cultural strength must be matched with technical structuring and a more coherent strategy for data governance.

Respondent 2 also highlighted the importance of testing and feedback in developing data-driven tools: “Why are we rolling out new systems without piloting them first?” Respondent 3 reinforced the issue of inconsistency in data entry. He explained that even for common deviations or observations, different terms are used across projects, making it difficult for AI to detect patterns. “Without naming conventions and tagging practices, it’s hard for AI to learn across projects.” He emphasized the importance of improving metadata, labeling, and input discipline, particularly if natural language processing (NLP) is to be applied at scale. Despite these challenges, Respondent 3 viewed the volume and potential richness of Skanska Sweden’s data as a latent asset that could support future AI use cases.

During workshops and prototype testing, employees responded positively to AI tools that could organize and cluster deviation data, even when sourced from free-text input. This suggests that better data practices could unlock significant user value if paired with intuitive tools and relevant use cases. In contrast, the US respondents described a more structured and mature data environment. Respondent 5 (US) explained that Skanska USA had spent years building out a centralized data layer: “We built out the data layer before we started talking AI. It was years of investment.” This structured foundation enables AI tools to support operations, from risk planning to subcontractor safety plans. She emphasized that the team now focuses on operationalizing data, not just storing it: “We’re helping both staff and subcontractors understand project risk and compliance with AI tools trained on internal data.”

Respondent 4 (US) noted that Skanska USA’s transition from monthly Excel reporting to real-time predictive models was only possible due to sustained efforts in data cleaning and alignment: “We can now predict optimized outcomes. That only works because we’ve spent years cleaning and aligning the data.” He underscored that useful AI models require input data that reflects real operational conditions, emphasizing structured, validated, and harmonized input as prerequisites for effective analytics. In the US context, Respondent 4 (US) raised a nuanced view on data as a strategic asset rather than just an operational resource. He emphasized that the value of data emerges not from owning it in isolation, but from how it is shared and leveraged across the ecosystem. “We’re even sharing this with our competitors... We have 11,000 companies using our software for free. Not so much in the future, but that’s actually one of the commercialization models. We just keep giving it away because then we own the data which is way more than the actual software.” Respondent 4 (US) also described how restrictive data-sharing cultures can stifle innovation, especially when companies prioritize data protection over collaboration: “You’re not really willing to share data with competitors. But the backside of that is... if you’re not sharing data, you don’t create anything new. And how do you compete?”. These reflections point to a strategic tension within Skanska’s global organization between data protection and innovation potential, and illustrate how data readiness is not only about structure and accessibility, but also about governance, ownership models, and openness to cross-organizational collaboration.

Respondent 6 (UK) offered a broader group-level view, observing that while Skanska has access to vast amounts of data, it is often not in a format conducive to AI analysis: “We have the data, but not always in a way that lets us act on it with AI.” She noted that recurring needs across business units, such as carbon tracking and safety, are emerging as shared priorities, and argued that developing standardized taxonomies and naming conventions would significantly enhance Skanska’s ability to scale AI solutions.

Participants’ Readiness
Employee readiness at Skanska Sweden reflects a generally open attitude toward AI, especially when the technology is positioned as a tool for improving workflows rather than as a disruptive force. Interviewees emphasized that while awareness and experience vary across departments and roles, most employees are curious and willing to engage with AI if its purpose and benefits are clearly communicated. Respondent 1 described employees as conditionally open: “People are curious, but they need to understand what AI does and why it’s useful.” She emphasized that scepticism may emerge if AI is perceived as replacing rather than supporting professional expertise: “If AI is summarizing data or helping with planning, I think people would use it. But if it’s about replacing your experience, that’s different.” Her reflections suggest that trust and transparency will be essential for employee acceptance.

Respondent 1 noted that employees’ reluctance to register deviations is sometimes driven by a fear of being blamed or singled out. This hesitancy is compounded by a general uncertainty about how the data they enter might be used, and by unfamiliarity with digital reporting tools. Although there is curiosity and a sense of openness to learning new methods, the psychological safety needed to openly document and share issues is not consistently present. This underscores that building a supportive environment, where deviation reporting is seen as a tool for learning and improvement rather than punishment is essential to achieving full buy, in for digital tools and, ultimately, for AI-enabled quality improvements. Respondent 2 noted that workers value tools that make their jobs easier: “We want better tools. AI should make everything faster, easier, more productive.” He used ACC as an example of a system that, while not branded as AI, is widely appreciated because of its practical utility: “ACC isn’t called an AI tool, but it makes our projects much easier.” He also stressed the importance of verifiability: “If it helps us focus on real deviations instead of scrolling through hundreds of cases, people will use it.” According to Respondent 2, adoption depends less on enthusiasm for AI and more on whether the tools feel grounded in the users’ real context.

Respondent 3 reinforced the importance of positioning AI as a support mechanism rather than a replacement. He explained that field workers, engineers, and project teams are more likely to accept AI if it clearly reduces workload or helps them avoid risks. “It’s about control. People want to know the AI isn’t doing things behind the scenes that they don’t understand.” He noted that acceptance is likely to grow if tools are transparent, simple to use, and focused on tangible outcomes. Across these Swedish interviews, a recurring theme was the need to frame AI as a value-adding enhancement to daily routines. Tools must be intuitive, relevant, and aligned with the realities of construction work to gain widespread trust. Interviewees noted that enthusiasm often increases after demonstrations or pilot usage, indicating that seeing AI in action is a key enabler of readiness. International perspectives also reinforced this approach. Respondent 5 (US) reported strong employee openness, particularly when AI tools directly supported operational tasks. “Subcontractors now use our chatbot to create their safety work plans. It saves time and points them to proven solutions.” She highlighted that the tool’s success stemmed from its contextual relevance , drawing on internal data and delivering usable outputs for on-site decisions.

Respondent 4 (US) described a similar trend, where field teams adopted AI tools readily when they saw direct benefits. “It’s not about forcing adoption. It’s about designing systems that workers want to use , because they help, not because they’re new.” He emphasized that user trust is maintained by allowing human oversight and grounding AI outputs in real-world data. Respondent 4 (US) added that their tools preserve decision-making autonomy while reducing administrative burden, which has encouraged broader acceptance. Respondent 6 (UK) does not work directly with frontline employees but emphasized the importance of internal champions , individuals who share positive experiences and model usage. “Confidence spreads when people see their peers using these tools successfully,” she explained, adding that cultural change begins with showing, not telling.

Customers’ Readiness

Customer readiness emerged as a secondary but influential factor in the organizational landscape for AI adoption at Skanska Sweden. Several interviewees noted that client expectations currently do not represent a primary driving force for AI implementation, but acknowledged that such external pressure could play a more significant role in the future. Respondent 3 Dunmark offered the most explicit reflections on this theme. He described Skanska as a company that rarely innovates for the sake of positioning or branding, stating: “Skanska isn’t a company that likes to take risks or be at the forefront , not for employer branding or PR. We do what customers ask for, and we do it well.” In his view, innovation efforts , including those involving AI , are primarily reactive to client demands. While internal motivation exists among employees, Respondent 3 expressed skepticism about its ability to generate meaningful change without corresponding expectations from clients. He emphasized that if client interest remains limited, competitive pressure may become the primary external driver for advancing AI initiatives: “If customers don’t ask us to use AI, then what pressure is there to change?” He also acknowledged that some clients already request demonstrations of modern technologies in tenders and saw this as a potential entry point for broader adoption of AI at Skanska Sweden.

Respondent 2 Burman shared this view, noting that clients currently do not explicitly demand AI-supported services, but request that Skanska Sweden uses the newest digital technologies in their deliveries. He explained that AI initiatives in his context are predominantly driven from within the organization, based on perceived internal needs rather than external requirements. While this internal motivation has value, the absence of client interest may limit the business case for more robust investment and scaling. Respondent 1 reinforced the notion that current AI-related momentum is largely internal. However, she implied that if external expectations were to emerge, particularly from clients , they could significantly influence how AI is prioritized across teams.

From the international perspective, the connection between AI and customer value was more pronounced, although often indirectly expressed. Respondent 5 (US) described how AI tools developed within Skanska USA are designed to enhance project delivery and risk reduction. While she did not elaborate on specific client reactions, she noted that these AI-enabled solutions are integrated into business planning and are likely seen by clients as value-adding features that support decision-making and safety outcomes. Respondent 4 (US) similarly tied AI to operational KPIs that are often mandated or closely monitored by clients. He explained that AI tools, such as predictive safety models, are well-aligned with customer compliance and performance expectations, particularly in regulated environments. Although he did not describe direct feedback from clients, his remarks suggest that embedding AI in service delivery can enhance perceived client value and drive adoption. Respondent 6 (UK) offered a group-level perspective, noting that she does not work directly with clients but has seen variation in client expectations across Skanska's markets. In some regions, customers have begun to request proof of digital maturity in tenders, including references to how AI or advanced analytics are used. She emphasized that while this is not yet a global norm, there are pockets of external demand that could be amplified through strategic storytelling and demonstration of successful use cases.

Process' Readiness

This section captures insights related to process readiness, including how AI initiatives are integrated into operational workflows, whether feedback mechanisms exist, and how internal communication affects adoption. While some content touches on innovation, the focus is on organizational processes and systemic structures that enable or hinder AI continuity. Process readiness at Skanska Sweden was widely described as open-minded but constrained by day-to-day operational pressures. Across interviews, respondents emphasized that employees and managers are generally positive toward innovation, including AI, but often lack the time, structure, and continuity mechanisms to carry ideas forward.

Respondent 2 described a workplace where innovation happens sporadically and under tight resource constraints: "We innovate when we can, but it's always on top of everything else." He explained that although interest is high, there is no formal ownership or internal resource to continue AI-related work once pilots or student projects end: "This should be owned by someone. Otherwise, we risk losing what you've built." He emphasized that the format of the thesis-based prototype , small, focused, collaborative , was well received and could serve as a model for future internal innovation projects. Respondent 1 shared this perspective, noting that while motivation is strong among staff, structural support for experimentation is limited: "We're motivated, but we don't have the time or structure to turn ideas into action." She suggested that internal testbeds or pilot environments , such as sandboxes, could help transform ideas into outcomes. Without such structures, however, energy around innovation tends to dissipate.

Respondent 3 described the process gaps more explicitly, highlighting how pilots often begin enthusiastically but then dissolve due to lack of clarity on next steps or long-term responsibility: "When a student builds a model that works, we should have a process for taking it forward , not just letting it die with the internship." He proposed the creation of cross-functional teams that could evaluate and continue innovation efforts beyond the pilot stage. Respondent 3 also emphasized the importance of systematic feedback loops: without mechanisms for collecting and acting on feedback, AI tools risk becoming static or irrelevant over time.

Across Swedish interviews, process readiness was framed not as a lack of will, but as a lack of ownership, follow-up mechanisms, and operational integration. Promising ideas struggle to scale due to unclear governance, siloed workflows, and insufficient structures to enable collaboration between departments. In contrast, the US and UK respondents described more embedded process structures. Respondent 5 (US) explained that innovation is embedded into project planning and execution, supported by cross-functional collaboration between strategy, data, and technology teams: "The people in our strategy function sit next to data and emerging tech. That helps connect strategy with delivery." These integrated roles ensure continuous iteration and prevent knowledge silos.

Respondent 4 (US) described a deeply rooted feedback and learning culture supported by structured internal RD. "We didn't build this system overnight. It was 30 years of learning what works and what doesn't. AI is just the latest layer." Real-time risk modeling and academic collaboration ensure that AI tools evolve with user needs, reinforcing long-term operational relevance. Respondent 6 (UK) noted that duplication of effort remains a barrier globally but pointed to her work in coordinating AI forums as a way to align communication and learning. "There's so much potential , we just need to reduce duplication and start acting like one company when it comes to AI." Her role focuses on enabling consistent internal communication and surfacing replicable practices across units.

DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY