2025 European Conference on Computing in Construction

CIB W78 Conference on IT in Construction

Porto, Portugal
July 14-17, 2025

CYBER RISK IDENTIFICATION IN CONSTRUCTION WITH LANGUAGE MODELS: THE NEXT GENERATION

Dongchi Yao, Borja García de Soto
S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi (NYUAD), Saadiyat Island, P.O. Box 129188, Abu Dhabi, United Arab Emirates

Abstract

The digitalization of the construction industry enhances efficiency through advanced technologies but increases cyber vulnerabilities. Existing research lacks comprehensive frameworks for risk identification and automation, leaving gaps in addressing cybersecurity challenges. Advances in language models offer potential, but limitations like outdated datasets and small architectures hinder their effectiveness. This study addresses these issues by collecting an up-to-date dataset to fine-tune the GPT-4o Mini model, renowned for its size and reasoning capabilities. The fine-tuned model outperforms others in identifying phase-specific cyber risks, generating a more thorough risk checklist. Its scalability suggests potential applications in broader risk management tasks, enabling industry-wide adoption.

Introduction

The construction industry is experiencing a digital revolution calling Construction 4.0, which is marked by the integration of advanced technologies such as digital twins, drones, robotics, and virtual reality. These technologies are transforming the way construction projects are designed, executed and maintained, significantly enhancing process efficiency, output quality and asset management effectiveness. However, the adoption of such high-tech solutions has increased the extent to which the industry is exposed to cyber risks. Despite the rapid advancement of the digitalization process, cybersecurity measures have not been upgraded simultaneously, which leaves the construction industry exposed to frequent cyberattacks and a major target of digital threats (Mantha and García de Soto, 2019).

This vulnerability was highlighted by cyberattacks on prominent companies like Bouygues Construction (Barbaschow, 2020), Skender Construction (Thibault, 2024), Fast Brick Robots (FBR) (Pash Chris, 2018), Marous Brothers Construction (Sawyer and Rubenstone, 2019), and Turner Construction (Stiles, 2016), etc. These incidents demonstrate the critical need for enhanced cybersecurity protocols within the industry. A significant factor contributing to these incidents, as identified in industry analyses (Mantha, García de Soto and Karri, 2021), is a widespread lack of awareness among stakeholders about the cybersecurity risks associated with different project phases. This lack of understanding leaves stakeholders ill-prepared to prevent or mitigate such risks, as ready-to-use preventive measures and actions are not clearly formulated or widely available.

Addressing this deficiency is crucial for construction companies to protect sensitive data and ensure the operational integrity of construction projects. However, as noted in the Related Works section, only a few studies on cyber risk identification exist, and most of these are limited, lacking comprehensive, industry-specific list of cyber risks (Yao and García de Soto, 2024). The methods employed in these studies also have significant weakness: they depend largely on manual execution, which introduces subjective bias and accidental oversights due to the differing levels of expertise among the implementers (Yao and García de Soto, 2024).

To address the deficiencies mentioned above and keep pace the era of generative AI, recent advancements in cyber risk identification have begun utilizing large language models (LLMs) to enhance automation in the construction industry. For example, Yao and García de Soto introduced an LLM-based approach to identify risks across project phases (Yao and García de Soto, 2024). Although the prospects are broad, has also exposed certain limitations, including reliance on outdated datasets, insufficient scale of the model architecture, and incomplete automation requiring a high level of human intervention. These challenges highlight the necessity of further development to achieve more effective models that can comprehensively identify risks and other tasks, such as risk assessment and mitigation.

Therefore, this study aims to develop a more capable language model for cyber risk identification by formulating four specific objectives: (1) collect more up-to-date construction cybersecurity text data, including academic publications, books, and guidelines; (2) fine-tune the GPT-4o Mini model using the collected dataset and our previously self-curated question-answer pairs; (3) evaluate the performance of the fine-tuned model against benchmark and other advanced models; and (4) formulate a more comprehensive list of cyber risks that might

happen in different project phases, serving as a novel industry benchmark. This study represents a significant step toward automating cybersecurity management in construction by leveraging advanced AI techniques while balancing potential costs and resource constraints.

The remainder is structured as follows: a literature review is presented first, followed by an outline of the methodology. Next, the study evaluates the model and generates the cyber risk checklist. This is followed by various in-depth discussions, and the study concludes with final remarks.

Related Works

Cybersecurity Challenges in Construction

As the construction industry gradually introduces digital technologies such as Building Information Modeling (BIM), the Internet of Things (IoT), and cloud computing, cybersecurity is facing significant challenges. These technologies have vulnerabilities that cybercriminals can exploit, including ransomware attacks, data breaches and unauthorized access to sensitive project information. Due to the fragmented structure of this industry, involving multiple stakeholders, subcontractors and suppliers, it makes the implementation of a unified security protocol complex, resulting in inconsistent protection measures throughout the supply chain. Legacy systems and diverse digital tools usually lack robust security features and are vulnerable to attacks. Cyber risks vary at different project phases such as design, procurement, construction and operation, and each stage faces unique threats. For instance, the design stage may become the target of intellectual property theft, while the construction stage may be disrupted by attacks from IoT devices. Many construction companies have limited awareness of cybersecurity at all stages and lack internal expertise, which hinders effective risk management (Mantha and García de Soto, 2019).

Cyber Risk Identification in Construction

A 2023 scoping review (Salami Pargoo and Ilbeigi, 2023) examined 45 publications related to cybersecurity in the construction sector. It includes 25 journal articles, 12 conference papers and various other sources. These studies can be classified into the following three categories: Among them, 24 focus on overall discussions, 2 are based on reviews, and 19 propose targeted solutions. Importantly, the review highlighted a significant research gap, with limited attention given to identifying cyber risks specific to the construction industry. For example, Shemov et al. (2020) explored the integration of blockchain within construction supply chains, which presents a threat-assessment framework that outlined potential cyberattacks and mitigation strategies. Similarly, Shibly and García de Soto (2020) introduced a threat modeling technique tailored to the construction sector, which was applied to a 3D concrete printing system to pinpoint system weaknesses and appropriate countermeasures. In another study, Mantha et al. (2021) proposed an initial threat model for cybersecurity in the AEC industry, using a case from the building commissioning phase to illustrate its application. Although these studies have made contributions, they rely heavily on manual processes, which are time-consuming, vulnerable to human bias, and lack responsiveness to the dynamic nature of cyber threats. This highlights the demand for cyber risk identification tools that are automated, scalable, and capable of adapting to new challenges.

Yao and García de Soto (2024) recently tried to address these limitations by using a language model to analyze a large number of online documents in order to solve cybersecurity-related questions and automatic risk identification in different project phases. Their approach integrated LLMs to simplify the identification process, which aligned with the construction industry's digital transformation and the increasing use of the use of LLMs. However, the study relied on a limited dataset (up to April 2021) and the GPT-2 architecture, and the final stage requires up to six hours of manual summary, which introduces subjectivity and limited comprehensiveness. To develop an intelligent question-answering system that can identify risks and propose effective mitigation strategies, significant improvements are still needed.

Large Language Models

LLMs can analyze extensive texts, including construction cybersecurity, and significantly improve the ability to identify cyber risks through text analysis. Language modeling calculates the probability of word sequences by decomposing sentences into conditional probabilities. And LLMs aim to maximize the likelihood of sentences from their training datasets, thereby generating coherent text aligned with the distribution of the datasets.

The advancement of LLMs driven by the transformer architecture introduced by Vaswani et al. (2017) has revolutionized NLP. The self-attention and parallel processing of transformer enabled richer word embeddings. In 2018, OpenAI's GPT (Radford *et al.*, 2018) and Google's BERT (Devlin *et al.*, 2019) leveraged transformers to improve NLP performance. Subsequent models like GPT-2 (Radford Alec *et al.*, 2019) and GPT-3 (Brown *et al.*, 2020) enhanced language generation and learning capabilities, with GPT-3 notable for its zero-shot and few-shot learning abilities.

Since GPT-3, billion-parameter LLMs have excelled in tasks such as language generation and question answering, expanding AI's role in healthcare, gaming, finance, robotics, etc. In 2023, OpenAI released GPT-4 (OpenAI *et al.*, 2023), a multimodal model handling text and images with human-level performance and enhanced reliability. Other notable LLMs include PaLM (Chowdhery *et al.*, 2022) and LaMDA (Thoppilan *et al.*, 2022), further broadening the applications and impact of LLMs across various industries.

In summary, considering the fact of deficient recognitions of cyber risks, the low level of automation, the outdated datasets for small-scale models in previous related work, and inspiration gained from language model applications in other industries, this research fine-tunes a larger

language model to detect cyber risks throughout each project phase, thus setting a new industry benchmark.

Methodology

The methodology is outlined in the flowchart shown in Figure 1, which illustrates the two-stage fine-tuning process. In Stage 1, which involves unsupervised fine-tuning, the focus is on enhancing the model's understanding of the domain-specific language associated with construction cybersecurity. This is achieved using a newly collected dataset from public sources, enabling the model to acquire domain knowledge and improve its contextual comprehension. Stage 2, which involves supervised fine-tuning, builds on this foundation by training the model to identify and adapt to diverse question-and-answer formats. This step prepares the model for downstream tasks specifically tailored to question-answering within the domain. Afterward, the model is ready to answer the relevant question to identify cyber risks throughout each construction project phase.

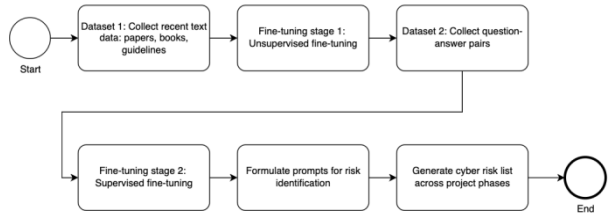

Figure 1: Methodology flowchart

Dataset

For fine-tuning, two datasets were utilized, both of which are related to either the construction industry or cybersecurity domain. The dataset for unsupervised fine-tuning was newly collected, while the question-answer pair dataset for supervised fine-tuning was the same as the one used in our previous work (Yao, 2023).

Unsupervised Fine-Tuning Dataset

Since the text data mentioned above is only till 2021, it is somewhat outdated. Therefore, this study includes more recent data. While the initial plan was to conduct a large-scale and extensive data collection, time constraints necessitated focusing on three specific categories of updated texts from 2022 to 2025. These categories include papers, books, and guidelines. This ensures that the newly trained language model stays current and relevant, avoiding reliance on outdated data. Details of the collected raw documents are provided in Table 1.

The collected documents went through several steps to prepare them for training:

- Crawling text: Tools involved included pdfplumber and python-docx, saving each document as a text string item in a list with 48 items.
- Text cleaning: The text extracted from these documents underwent a thorough and meticulous cleaning process, which included removing special characters, headers, footers, website links, newline characters, and other unnecessary elements.
- Chunking: Long text strings were split into smaller chunks, each with a maximum of 50,000 characters, ensuring no string exceeds the limit, making them suitable for the context window of GPT-4o-mini. This resulted in 133 strings. This time, the text was not split into individual sentences. Instead, the newly collected text was processed as larger chunks to allow the language model to better capture and understand context. This approach represents an improvement over our previous methodology (Yao and García de Soto, 2024). The total number of characters is 5,383,803, with 71.43% of the strings around the set maximum character limit.

Table 1: Details of the collected documents

Text source	Count	Number of characters	Total number of characters
Papers	35	1954,456	
Books	10	4193,007	6,251,097
Guidelines	3	103,634	

- Preparing the dataset for fine-tuning: The chunked list was shuffled randomly with a fixed seed for reproducibility and split into 80% training and 20% validation sets. A function converted each string into JSONL format containing system prompts, user placeholders, and assistant responses. This process ensured structured, random, and reproducible data for fine-tuning. An example dataset entry is: {"messages": [{"role": "system", "content": "You are an assistant that helps with cybersecurity risk management."}, {"role": "user", "content": " "}, {"role": "assistant", "content": string}]}

Supervised Fine-Tuning Dataset

For the supervised fine-tuning stage, we adopted the dataset from our previous work (Yao and García de Soto, 2024), which include 326 questions focused on topics in construction or cybersecurity fields, and each question accompanied by a comprehensive answer. To make the model recognize various linguistic styles, each question was rewritten as four alternative sentences, capturing both imperative and interrogative styles (Yao and García de Soto, 2024). Interrogative sentences usually end with a question mark, while imperative sentences usually end with a period. These multiple rephrasing can ensure that the model can detect subtle differences in the presentation of questions or instructions, thereby enhancing its robustness and enabling it to handle different user inquiries within both domains. This approach enables the model to recognize a query irrespective of its linguistic presentation and generate the same answer consistently, thus improving adaptability. Consequently, the dataset was expanded to 326 multiplied by five, producing a total of 1,630 question-answer pairs. Table 2 presents five example pairs related to cybersecurity and the complete dataset can be obtained from our GitHub page (Yao, 2023).

Table 2: SFT dataset examples (Yao, 2023)

Question	Answer
How are cyber threats identified?	Cyber threats may be identified through various means, including threat intelligence, security information and event management (SIEM), and user behavior analytics.
How are cyber threats commonly detected?	
Can you explain how cyber threats are identified?	
Describe the process of identifying cyber threats.	
Explain the methods for identifying cyber threats.	

Model

The chosen model is GPT-4o Mini (OpenAI, 2024), a cost-effective version of OpenAI's GPT-4o model, launched in July 2024. This model delivers advanced AI capabilities at a lower price and supports essential features like multimodal input (text and images), JSON mode, parallel function calling, and seamless API integration. With its versatile design, GPT-4o Mini is well-suited for various tasks, including customer support, data extraction, and content generation. In this study, there are three reasons for employing GPT-4o Mini to upgrade our previous language model for risk identification:

- Cheaper: At the time of this study (Jan 2025), GPT-4o Mini costs just $0.15 per million input tokens and $0.60 per million output tokens, making it more than 60% cheaper than GPT-3.5 Turbo. This cost-efficiency is crucial for construction companies, which often work with limited budgets, allowing them to adopt AI-driven risk identification solutions without excessive financial burden.
- Capable: The GPT-4o Mini achieved a score of 82% in the MMLU benchmark test, which demonstrates powerful language processing capabilities and surpasses similar compact models. Its extended context window of 128,000 tokens enables it to process complex documents and extract key information, making it highly suitable for risk identification and classification tasks. These tasks require analyzing and organizing textual data but do not involve complex reasoning processes, aligning perfectly with the model's strengths.
- Faster: The streamlined architecture and advanced features of GPT-4o Mini, such as parallel function calling, support rapid fine-tuning and deployment. This speed is essential for achieving real-time risk identification and response, enabling construction teams to act quickly and effectively in dynamic and potentially hazardous situations.

Fine-Tuning

(1) ***Unsupervised Fine-Tuning*.** We fine-tuned the gpt-4o-mini-2024-07-18 base model using supervised learning via the OpenAI API. The training process involved 2,609,454 tokens over three epochs (automatically determined by OpenAI), with a total of 318 steps, a batch size of 1, and a learning rate multiplier of 1.8, using a seed value of 374929889. The entire training process was completed in approximately 600 seconds. The fine-tuning achieved a training loss of 1.9040 and a validation loss of 2.0546, with the full validation loss recorded at 2.4408, indicating that the supervised fine-tuning is effective with relatively low loss.

(2) ***Supervised Fine-Tuning*.** After fine-tuning the model using the OpenAI API on 231,945 tokens over three epochs (1,956 steps, batch size 2, learning rate multiplier 1.8, seed 83207177), the process concluded in about 36 minutes. The final training and validation losses of 0.0002 indicate minimal deviation from target outputs, reflecting high accuracy. The 0.0156 validation loss, measured on a broader dataset, similarly denotes strong performance. This value provides a quantitative measure of model error relative to ideal predictions, highlighting the model's effectiveness in maintaining consistency across various evaluation criteria.

Model Evaluation

To demonstrate advances in cyber risk identification, we compare our fine-tuned model with baselines: the gpt-2-RL benchmark (Yao and García de Soto, 2024) and several larger GPT-4 variants (Table 4). Their contrasting architectures, scales, and training regimes create a comprehensive testbed, revealing how scaling and targeted fine-tuning boost performance.

Prompts

The prompts and reference answers are formulated based on the findings in the literature (Mantha, García de Soto and Karri, 2021), as shown in Table 3, and are consistent with the prompt formulation used in our previous work (Yao and García de Soto, 2024). The questions were designed in different styles, as outlined below, resulting in a total of 5 styles × 6 phases = 30 prompts. Each reference answer was structured as 'Cyber risks in the {key1} phase include {key2} resulting from {key3}.' (Yao and García de Soto, 2024)

- What are the cyber risks in the {key1} phase?
- Can you identify cyber risks in the {key1} phase?
- Cyber risks in the {key1} phase include
- Identify cyber risks in the {key1} phase.
- What types of cyber risks should be considered during the {key1} phase?

Table 3: Cyber risk list from Mantha et al. (2021)

Phase (Key1)	Threats (Key2)	Vulnerabilities (Key3)
Initiation	Data theft	Unsecured network transfer and cloud storage applications
Design	Proprietary information stolen	Unpatched software
Construction & Procurement	Performance degradation, physical damage	Excessive usage, fabricated chips
Commissioning	Data tampering, actuation tampering	Compromised dashboard and sensor
Operation & Maintenance	Spying, deliberate destruction	Chip insertion
Renovation & End of life	Data retrieval	Disposed sensors and equipment

Answer Generation

Each model is limited to generating responses of fewer than 50 tokens for several reasons. First, this can ensure consistent comparisons across different models, allowing for a fair evaluation of their performance. Second, by matching the length of answers in the fine-tuning dataset, the models maintain behavior similar to their fine-tuned counterparts. Third, enforcing a 50-token limit promotes conciseness, making the answers easier to understand and digest for users. This approach balances uniformity, reliability, and clarity in the model outputs.

Metrics

Generated answers are evaluated against references via: (1) BLEU, comparing n-gram overlap; higher equals closer wording. (2) ROUGE, unigram (ROUGE-1), bigram (ROUGE-2) and longest common subsequence (ROUGE-L), reports precision, recall and F1 to gauge content coverage, accuracy and balance. (3) BERTScore uses BERT embeddings to assess semantic similarity, again providing precision, recall and F1, giving a nuanced context-aware quality measure.

Results and Analysis

Table 4 shows the results of our model comparison, from which several important observations can be made.

(1) The lowest performance of gpt-2-RL. Among all the models evaluated, gpt-2-RL demonstrates the lowest performance across all metrics. This outcome is unsurprising given its relatively smaller model size and the limitations of the pretraining dataset. These constraints likely hindered its ability to generalize and perform well on downstream tasks, making it less competitive compared to the more advanced models.

(2) The effectiveness of gpt-4o-mini-stage-2. This fine-tuned model delivers the highest performance across all metrics, significantly outperforming all other models, including those in our previous work. This underscores the effectiveness of our fine-tuning methodology, which proves to be both robust and impactful. The improvements over earlier models, such as gpt-2-RL, validate the success of our approach in enhancing model performance through targeted fine-tuning strategies.

(3) Comparison with GPT-4o and GPT-4 Turbo. Expanding the model further to larger architectures like GPT-4o or GPT-4 Turbo does not yield superior results. While these larger models offer increased capacity, their performance pales in comparison to our fine-tuned gpt-4o-mini-stage-2, indicating that well-executed fine-tuning can be more critical than model size. This finding suggests that our fine-tuned model strikes a highly effective balance between model capacity and optimization, making further scaling unnecessary for this task.

In conclusion, the exceptional performance of gpt-4o-mini-stage-2 across all metrics highlights the effectiveness of our fine-tuning approach. By surpassing both smaller and larger architectures like GPT-4 Turbo, our model demonstrates its superior capability to generate answers to risk identification questions, positioning it to be an effective tool for later risk identification.

Cyber Risk Identification Application

The fine-tuned gpt-4o-mini-stage-2 model is made to identify cyber risks across project phases. The procedure consisted of submitting strategically formulated queries to the model, which then generated responses that were consolidated into a cyber risk checklist by project phases. To maintain consistency with the linguistic patterns in supervised fine-tuning, various question structures were employed in the Prompts section. To improve the model's capacity to identify varied references to project phases, key terms within the questions were intentionally diversified. The term set {key1} captures the language associated with different phases of a construction project, as presented in Table 5. This phase categorization aligns with the structure adopted in earlier threat modeling research by Mantha et al. (2021).

Table 5: Different terms of phases

Phase	Different terms
Initiation	pre-planning, concept, feasibility study, conception, early project definition, inception
Design	design development, schematic design, detailed design, planning
Construction &Procurement	construction, procurement, build, execution, implementation
Commissioning	handover, startup, completion, closeout
Operation & Maintenance	operation, maintenance, operations phase, facility management, service phase
Renovation & End of Life	renovation, end-of-life, demolition, decommissioning, retrofitting, rehabilitation, disposal

Table 4: Comparisons among various models

Model	BERTScore			BLEU	ROUGE-1			ROUGE-2			ROUGE-L		
	P	R	F		P	R	F	P	R	F	P	R	F
gpt-2-RL (benchmark)	0.836	0.852	0.844	1.1E-231	0.114	0.201	0.145	0.025	0.045	0.032	0.090	0.159	0.115
gpt-4o-mini-2024-07-18	0.856	0.884	0.869	9.8E-232	0.188	0.437	0.262	0.089	0.215	0.125	0.136	0.318	0.190
gpt-4o-mini-stage-1	0.854	0.883	0.868	9.7E-232	0.180	0.426	0.252	0.084	0.208	0.119	0.143	0.340	0.200
gpt-4o-mini-stage-2	***0.904***	***0.911***	***0.907***	***1.1E-231***	***0.304***	***0.483***	***0.366***	***0.176***	***0.285***	***0.213***	***0.294***	***0.468***	***0.355***
chatgpt-4o-latest	0.859	0.885	0.872	9.7E-232	0.182	0.436	0.256	0.085	0.213	0.120	0.137	0.329	0.192
gpt-4-turbo-2024-04-09	0.861	0.883	0.872	9.6E-232	0.179	0.447	0.255	0.081	0.213	0.117	0.135	0.341	0.193

The answers produced by the model were reviewed and compiled into a checklist of identified cyber risks. Due to page limits, only cyber risks identified for the initiation phase are shown. For the full list of cyber risks, please refer to our GitHub page (Yao, 2025).

- Weak Identity and Access Management. Inadequate access controls, such as the absence of multi-factor authentication, can allow unauthorized individuals to access sensitive project systems and data.
- Unauthorized Access to Project Documents. Without proper safeguards, critical bidding documents and project plans may be viewed, stolen, or altered by unauthorized personnel.
- Insecure Communication Channels. Using unencrypted emails or unsecured messaging platforms can lead to the interception of sensitive project information during transmission.
- Inadequate Data Encryption. Failing to encrypt data at rest and in transit makes sensitive project information vulnerable to breaches and unauthorized access.
- Phishing Attacks Targeting Key Personnel. Deceptive emails or websites can trick project managers and team members into revealing login credentials or confidential project details.
- Third-Party Vendor and Supply Chain Risks. Cyber vulnerabilities from subcontractors or technology providers with weak security measures can indirectly compromise the main project's security.
- Ransomware Attacks on Bidding Systems. Ransomware can lock access to essential project documents and systems, causing significant delays during project initiation.
- Lack of Cybersecurity Governance. The absence of formal cybersecurity policies can result in inconsistent security practices and leave the project exposed to various cyber threats.
- Insider Threats. Employees, contractors, or other insiders with access to project systems and sensitive information may intentionally or unintentionally compromise security, thereby leading to data breaches or system disruptions.
- Inadequate Incident Response Planning. Without a well-defined incident response plan, the project team may struggle to effectively address cybersecurity incidents, resulting in prolonged downtime.

Compared with the previous list that only contained 36 risk items in (Yao and García de Soto, 2024), this study identified 57 items, thereby enhancing the comprehensiveness of risk identification. The time required to review and compile the list was significantly reduced from six hours to less than 20 minutes by involving just one person in the process. This improvement in efficiency is attributed to our newly fine-tuned model, which generates more cohesive and accurate responses that are easier to interpret. Furthermore, unlike our earlier GPT-2-based model (Yao and García de Soto, 2024), the updated model can provide detailed explanations for each cyber risk item when prompted for longer descriptions. These advancements collectively represent a substantial improvement in our risk identification process.

Discussions

The Application Scenarios of the Checklist

The study proposes two principal applications for the prioritized checklist. First, it functions as a novel benchmark for project managers, which allows them to develop proactive and preventive strategies focusing on critical risks before the start of a specific phase. Second, risk analysts can employ the checklist to support project-level risk management tasks, which include identifying specific risk factors, conducting quantitative evaluations, and prioritizing critical risk factors. Moreover, the checklist is a useful tool for various stakeholders. For instance, cybersecurity teams can apply it to evaluate and upgrade existing security frameworks by revealing key vulnerabilities. Addressing these vulnerabilities supports robust, tailored security measures for construction projects. Stakeholders, such as contractors and vendors, can also utilize on the checklist to ensure their systems and communication protocols comply with necessary security standards. In addition, it can also provide information for training and helping personnel understand common cyber risks, potential impacts, and the need to adhere to security protocols. Through further fine-tuning, the checklist remains current with the ever-evolving cybersecurity landscape, keeping users informed about new threats and trends. Therefore, it remains a reliable tool to ensure timely response to critical risks at every stage of a project.

Using the Model for Other Tasks

The fine-tuned GPT-4o Mini model, demonstrating strong capabilities in cyber risk identification, also has the potential to address other cyber risk management tasks essential for construction projects, due to its upgraded model size and enhanced reasoning capabilities. The tasks include but not limited to the following.

(1) Project-Level Risk Assessment. Due to the expanded size, this model understands more details. For a certain project's stage, variables representing its current situation or a detailed description of the project's situation, maybe 100 or 200 variables such as the number of mobile ends or the frequency of information exchanges using cloud platforms can be inputted into the model. The probability of likelihoods or serious assessments about possible cybersecurity incidents like phishers' attacks and insider attacks can then be inferred by the model. Depending on how specific the inputs are about the project information, the model may need further fine-tuning.

(2) Risk Strategy Formulation. With its capability of reasoning and producing structured knowledge outputs, the model can generate a step-by-step strategy to enhance a project's cybersecurity when provided with a description of the project at a specific phase. This strategy includes aspects such as personnel configuration (e.g., dedicated IT personnel ratio), project data exchange protocols within a common data environment, and resilience enhancements, such as the number of data

backups. This function can be tailored to companies requiring different levels of strategy granularity.

(3) Cybersecurity Document Correction. For companies with their own security procedure documents already established, the corrected procedures should be entered into the model so that it might recognize any problems present in the documents and put forward corrective solutions. A simple example: the model might notice issues like inadequate access controls, outdated encryption protocols, or weak incident response plans. Once identified, the model will offer managerial actions and technological countermeasures such as clarifying security roles, establishing training programs, setting policy improvements, updating software configurations, implementing multi-factor authentication, and deploying intrusion detection systems, etc., to strengthen a construction project's cybersecurity defenses.

Industry-Wide Model Deployment

The fine-tuned GPT-4o-mini model provides a powerful industry solution for construction companies, aiming to simplify risk management. Through simple API integration, this model can be embedded in internal websites, mobile applications or existing project management software (such as Procore and Buildertrend), ensuring that stakeholders such as contractors, project managers and security officers can quickly obtain key information. Tests show that the inference time for each prompt is approximately two seconds, which is highly suitable for on-site scenarios that require rapid decision-making. The operating cost of this model is low. The cost of each million input tokens is approximately $0.15, and the cost of each million output tokens is approximately $0.60. It supports frequent real-time analysis and large-scale data processing. This economic efficiency is particularly beneficial for ongoing tasks such as incident monitoring, forecasting and scenario planning. To keep up with the constantly changing construction methods and cybersecurity standards, it is recommended to fine-tune using a dataset of approximately 100,000 tokens (about 75,000 words). This scale ensures that the model can absorb the professional vocabulary of specific industries while avoiding overfitting, and the update cost is approximately $0.30. Even if it is updated twice a month, the total cost is only $0.60. In addition, fine-tuning usually takes less than an hour, making regular updates possible. By combining public resources and proprietary documents, GPT-4o-mini can be continuously optimized to provide professional and economical risk management support for the construction industry.

Limitations and Outlooks

Due to the difficulty in obtaining such data, the datasets used lack detailed project-specific information, which limits the model's ability to assess the unique network risks of individual projects. The study did not evaluate the model's performance on a broader range of tasks, such as threat classification/detection or cybersecurity document validation, making it premature to conclude that the model can handle all risk-related tasks. Future work will focus on gathering diverse datasets to enhance the model's abilities, which are aligned with company-specific needs and may include inputs and outputs such as codes, project network diagrams, or schedules. The ultimate goal is to integrate this enhanced tool into within web/mobile platforms to offer an AI-driven cybersecurity consultant for users without specialized knowledge.

Conclusions

Within the development of AI-driven risk management, this study contributes a novel approach by fine-tuning a GPT-4o-mini model to enhance cyber risk identification across diverse project phases. This model was initially unsupervised trained based on a corpus of recent literature and guidelines and then optimized through supervised training with 1,630 pairs of questions and answers to ensure its ability to cope with various cybersecurity challenges. By systematically comparing our model's performance against other notable benchmarks, including GPT-4-based variants and our earlier gpt-2 framework, we reveal its advantages in accurately identifying and contextualizing cyber risks. In addition to the improvements in the algorithm, this study also introduces a fast and cost-effective method for generating a comprehensive cyber risk checklist. This practical output is compiled within 20 minutes by a single expert, which underscores the model's utility in delivering timely and actionable insights for project managers who seek proactive preventive measures. These findings reveal the potential of advanced large language models to improve existing models in risk management in the construction industry and bridge the gap between theoretical best practices and actual operations. Future work will enhance the accuracy of the model by integrating project-specific data, further verify its versatility through more extensive testing, and ultimately integrate it into industry-scale web and mobile applications.

Acknowledgments

This study was supported by different Centers at NYUAD. It was supported by the Center for Sand Hazards and Opportunities for Resilience, Energy, and Sustainability (SHORES), funded by Tamkeen under the NYUAD Research Institute Award CG013; the Center for Cyber Security at New York University Abu Dhabi (CCS-AD), funded by Tamkeen under the NYUAD Research Institute Award G1104; and the Center for Interacting Urban Networks (CITIES), funded by Tamkeen under the NYUAD Research Institute Award CG001.

Declaration

Language and wording in this manuscript were refined with the help of OpenAI's ChatGPT, used solely for grammar and clarity improvements.

References

Barbaschow, A. (2020) *Bouygues Construction falls victim to ransomware*, *ZDNET*. Available at: https://www.zdnet.com/article/bouygues-construction-falls-victim-to-ransomware/ (Accessed: 30 September 2023).

Brown, T.B. *et al.* (2020) 'Language Models are Few-Shot Learners'. arXiv. Available at: http://arxiv.org/abs/2005.14165 (Accessed: 2 March 2024).

Chowdhery, A. *et al.* (2022) 'PaLM: Scaling Language Modeling with Pathways'. arXiv. Available at: http://arxiv.org/abs/2204.02311 (Accessed: 2 March 2024).

Devlin, J. *et al.* (2019) 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding'. arXiv. Available at: http://arxiv.org/abs/1810.04805 (Accessed: 2 March 2024).

Mantha, B., García de Soto, B. and Karri, R. (2021) 'Cyber Security Threat Modeling in the AEC Industry: An Example for the Commissioning of the Built Environment', *Sustainable Cities and Society*, 66, p. 102682. Available at: https://doi.org/10.1016/j.scs.2020.102682.

Mantha, B.R.K. and García de Soto, B. (2019) 'Cyber Security Challenges and Vulnerability Assessment in the Construction Industry', in *Proceedings of the Creative Construction Conference 2019*. Budapest University of Technology and Economics, pp. 29–37. Available at: https://doi.org/10.3311/CCC2019-005.

Mohamed Shibly, M.U.R. and García de Soto, B. (2020) 'Threat Modeling in Construction: An Example of a 3D Concrete Printing System', in *37th International Symposium on Automation and Robotics in Construction*. Available at: https://doi.org/10.22260/ISARC2020/0087.

OpenAI *et al.* (2023) 'GPT-4 Technical Report'. arXiv. Available at: http://arxiv.org/abs/2303.08774 (Accessed: 2 March 2024).

OpenAI (2024) *GPT-4o mini: advancing cost-efficient intelligence*, *OpenAI*. Available at: https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/ (Accessed: 10 January 2025).

Pash Chris (2018) *How hackers and spies tried to steal the secrets of Australia's one-armed robot bricklayer*, *Yahoo Finance*. Available at: https://au.finance.yahoo.com/news/hackers-spies-tried-steal-secrets-103645052.html?guccounter=1 (Accessed: 30 September 2023).

Radford, A. *et al.* (2018) 'Improving Language Understanding by Generative Pre-Training', *OpenAI blog*. Available at: https://openai.com/research/language-unsupervised (Accessed: 2 March 2024).

Radford Alec *et al.* (2019) 'Language Models are Unsupervised Multitask Learners', *OpenAI blog*. Available at: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf (Accessed: 2 March 2024).

Salami Pargoo, N. and Ilbeigi, M. (2023) 'A Scoping Review for Cybersecurity in the Construction Industry', *Journal of Management in Engineering*, 39(2), p. 03122003. Available at: https://doi.org/10.1061/JMENEA.MEENG-5034.

Sawyer, T. and Rubenstone, J. (2019) *Construction Cybercrime is on the Rise*, *Engineering News-Record*. Available at: https://www.enr.com/articles/46832-construction-cybercrime-is-on-the-rise (Accessed: 23 April 2021).

Shemov, G., García de Soto, B. and Alkhzaimi, H. (2020) 'Blockchain Applied to the Construction Supply Chain: A Case Study with Threat Model', *Frontiers of Engineering Management*, 7(4), pp. 564–577. Available at: https://doi.org/10.1007/s42524-020-0129-x.

Stiles, M. (2016) *Turner Construction data breach exposes hundreds in Washington to possible fraud*, *The Business Journals*. Available at: https://www.bizjournals.com/seattle/blog/techflash/2016/04/turner-construction-data-breach-exposes-hundreds.html (Accessed: 15 July 2021).

Thibault, M. (2024) *Skender hit by ransomware attack*, *ConstructionDive*. Available at: https://www.constructiondive.com/news/skender-ransomware-attack-chicago-maine/712844/ (Accessed: 12 May 2024).

Thoppilan, R. *et al.* (2022) 'LaMDA: Language Models for Dialog Applications'. arXiv. Available at: http://arxiv.org/abs/2201.08239 (Accessed: 2 March 2024).

Vaswani, A. *et al.* (2017) 'Attention is all you need', in *Advances in Neural Information Processing Systems*.

Yao, D. (2023) *SFT Training Dataset*, *GitHub Repository*. Available at: https://shorturl.at/Wckwk (Accessed: 28 April 2023).

Yao, D. (2025) *Cyber risk checklist-EC3*, *GitHub Repository*. Available at: https://github.com/Dongchi-Yao/Project/blob/master/EC3/Cyber%20risk%20checklist-EC3.docx (Accessed: 17 January 2025).

Yao, D. and García de Soto, B. (2024) 'Enhancing cyber risk identification in the construction industry using language models', *Automation in Construction*, 165, p. 105565. Available at: https://doi.org/10.1016/j.autcon.2024.105565.