

Article

Assessing the Accuracy of ChatGPT Use for Risk Management in Construction Projects

Hande Aladağ 

Department of Civil Engineering, Yildiz Technical University, Istanbul 34349, Turkey; haladag@yildiz.edu.tr;
Tel.: +90-212-383-5258

Abstract: Artificial Intelligence (AI) is considered promising digital technology that has important opportunities for enhancing project oversight and delivering improved decision-making in the risk management domain. However, there is a limited amount of research that has evaluated AI tools' performance in risk management. Therefore, with the intention of sustaining more accurate risk-based decision-making process in the construction industry, this paper investigates the accuracy of ChatGPT in risk management for different project types. In this context, Key Performance Indicators (KPIs) related to each risk management sub-process were determined, and then a questionnaire that consisted of prompt templates was prepared for collecting data from ChatGPT. Afterwards, ChatGPT's responses were evaluated by experts with focus group sessions. The findings indicate that ChatGPT has a moderate level of performance in managing risks. It provides more accurate knowledge in risk response and risk monitoring rather than risk identification and risk analysis sub-processes. This research paves the way for future studies by demonstrating an implication of ChatGPT use for risk-based decision making. In addition, gaining insight into the precision of ChatGPT in the risk-based decision-making process will empower decision-makers to establish resilience in business operations through technology-driven risk management.

Keywords: artificial intelligence; ChatGPT; construction 5.0; construction innovation; construction projects; digitalization; risk management



Citation: Aladağ, H. Assessing the Accuracy of ChatGPT Use for Risk Management in Construction Projects. *Sustainability* **2023**, *15*, 16071. <https://doi.org/10.3390/su152216071>

Academic Editors: Chunlu Liu, Mirosław J. Skibniewski, Hui Liu and Yu Lei

Received: 29 September 2023
Revised: 10 November 2023
Accepted: 16 November 2023
Published: 17 November 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Risk management is crucial for construction projects due to the inherent uncertainties and complexities involved in the construction industry [1]. Effective risk management is essential for construction projects to minimize disruptions, control costs, ensure safety, enhance stakeholder satisfaction, and achieve project success [2]. Thus, construction project teams can improve project outcomes. Likewise, Nobanee et al. (2021) argued that sustainability and risk management are closely intertwined, and organizations must evaluate unpredictable and unknown risks for facilitating the identification of risk factors both in the present and future economic landscape [3]. In this vein, the successful execution of projects can be maintained by implicating an effective risk management process that consists of the sub-processes of risk identification, risk prioritization, risk response, and risk control [4].

Risk management provides decision-makers a systematic approach to identify, assess, and mitigate risks. It also enables decision-makers to make more informed and effective decisions. However, within the decision-making perspective, the success in risk management process generally depends on human expertise and judgment [5], and additionally requires proper data collection for overcoming the inherent complexities in construction works. At this point, it is argued that Artificial Intelligence (AI) is amongst the crucial digital technologies that have significant opportunities to solve complex problems and deliver improved decision-making in the risk management domain [6,7].

Studies in the literature show that AI has the potential to greatly enhance risk management [8,9], and intelligent risk management is necessary to help construction participants gain successful projects [10]. In this respect, AI can monitor, recognize, evaluate, and predict potential risks in terms of safety, quality, efficiency, and cost under high uncertainty [11]. By employing AI-based risk analysis, project managers can gain valuable insights to swiftly prioritize potential risks and identify proactive measures, rather than relying solely on reactive risk mitigation responses [12]. In brief, as the construction industry continues to accumulate vast amounts of data, the utilization of AI in the field of risk management is anticipated to become increasingly widespread. Consequently, AI is expected to play a significant role for project managers in risk assessment, generating decision support, automation of risk monitoring, and scenario analysis. These possible outcomes of AI use in risk management will also trigger the practical implementation of AI applications by the project managers. However, many firms are still facing difficulties to apply AI for risk analysis in construction projects because generally AI methods are found as expensive to invest in. In addition, the difficulty in the learning process of AI tools is another important obstacle that effects the implementation level of AI application for risk analysis among construction organizations [13]. Therefore, it is of significance that learning and exposing AI tool applications should be less complex since this would aid organizations to filter and handpick the appropriate AI risk analysis tools for their organization [13].

Today, there are many different tools that can be used for AI applications, such as probabilistic models, fuzzy theory, machine learning, and neural networks. ChatGPT has gained significant recognition for its capacity to promote the sharing of knowledge, assist in research pursuits, and improve the problem-solving capabilities within diverse scientific domains among the array of available AI tools [14]. In addition, findings in the literature point out that when compared to the other AI tools, ChatGPT can provide users (project managers and non-technical stakeholders) quick and improved decision-making experience in real-life applications with its ease of use and simplicity in translating technical jargon into comprehensible information. As a result of this enhanced communication and comprehension within the team, the quality of decision-making is elevated, ultimately resulting in more favorable project outcomes [15]. In this regard, this study limits its focus to ChatGPT since it is more assessable and easily usable by all project stakeholders when compared to other AI tools.

In the literature, providing certainty to the project stakeholders is seen as an important aspect of risk management in real-life implementations [16] and ChatGPT is proposed as a vital support tool with its advanced language capabilities and analytical skills for project risk management process [15]. Therefore, while implementing a formal risk management process, it is substantial to understand the accuracy of tools used within the process tangibly. Although the possible contributions of AI-driven risk management for project managers have been sufficiently pointed out in the risk management literature [17–20], there are a limited number of studies that evaluate AI tools' performance in the risk management domain in real-life implementations [21]. When the focus is set on GPT use for risk management in construction projects, the number of studies decreases. Therefore, this study aims to explore the performance of ChatGPT for risk-based decision in the construction industry. In line with this aim, the objectives of this study are as follows: (1) Determine the Key Performance Indicators (KPIs) related to each risk management sub-process for assessing the accuracy of ChatGPT. (2) Gather data from ChatGPT in consideration of the identified KPIs. (3) Analyze the accuracy of the gathered data from ChatGPT with expert evaluations.

Given the strong interest in applying AI methods in construction management to harness digital evolution [12], providing feedback on ChatGPT's accuracy in construction project risk management will advance AI applications in civil engineering. In this vein, the main contributions of the study are as follows:

- Understanding the appropriate usage areas along with the usage limitations of ChatGPT in the risk decision-making process, which provides insights for construction companies that want to be digital pioneers in the industry.
- Understanding the accuracy of ChatGPT in the risk decision-making process, which can enable decision-makers to build resilience through technology-driven risk management in business operations.
- Additionally, as one of the pioneer studies in the AI-driven risk management domain, the study indubitably contributes to the body of risk management knowledge with its findings related to ChatGPT's performance in each risk management sub-process for different construction project types.

Regarding the organization of the study, after this introduction section, Section 2 presents a literature review summary under three sub-sections, which focus on artificial intelligence technologies in general, AI use for risk management in construction projects, and ChatGPT use for risk management in construction projects, respectively. Section 3 outlines the research methodology that encompasses the determination of KPIs for evaluating the accuracy of ChatGPT in risk management processes, preparation of the questionnaire, data collection, and data analysis, respectively. Section 4 focuses on the presentation and discussion of the findings obtained from the gathered data. Finally, in Section 5, along with the theoretical and practical implications of the research, the limitations and recommendations for further studies are presented.

2. Literature Review

The following sections present summaries under three sub-sections, which focus on AI technologies in general, AI use for risk management in construction projects, and ChatGPT use for risk management in construction projects respectively.

2.1. Artificial Intelligence and the Subfields Used in Risk Management

The definition of AI states that “tasks that can be operated automatically using self-governing mechanical and electronic devices that use intelligent control” [7]. There are three types of AI conceptualizations, which are Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGI), and Artificial Super intelligence (ASI). In the field of ANI, machines demonstrate intelligence within specific domains, such as chess playing, sales prediction, movie recommendations, language translation, and weather forecasting. AGI aims to develop machines capable of solving complex problems using their own reasoning and decision-making abilities, whereas ASI focuses on constructing machines that surpass human capabilities across multiple domains [22]. As a vast umbrella term, AI includes various technologies, applications, types, and subfields. Based on a categorization provided by [22], these subcategories are: (a) machine learning, (b) computer vision, (c) natural language processing, (d) knowledge-based systems, (e) optimization, (f) robotics, (g) automated planning and scheduling (Table 1).

A literature review shows that Machine Learning (ML), Robotics, Knowledge-based Systems (KBS), Natural Language Processing (NLP), and optimization are commonly used AI subfields of risk management in the construction industry [22]. Insights into each of these subfields are presented below.

ML involves the creation and utilization of computer programs to acquire knowledge from experience or historical data, enabling the modeling, control, or prediction of various phenomena using statistical methods, without the need for explicit programming. Robotics is a multidisciplinary engineering field that encompasses the creation, production, operation, and upkeep of robots and computer systems that imitate human physical actions. KBS is a field within AI that focuses on utilizing existing knowledge to enable machine decision-making. This is achieved by creating a knowledge base, which is composed of domain expert knowledge, historical cases or experiences, and other pertinent sources of information. NLP focuses on developing computational models that replicate the linguistic abilities of humans. NLP finds applications in various domains, such as machine trans-

lation, the processing and summarization of natural language text, and the retrieval of information in multiple languages and speech recognition. Optimization is concerned with making decisions or choices that provide the best outcomes given a set of constraints [22].

Table 1. AI tools and subfield of AI, derived from [22].

Machine Learning	Computer Vision	Knowledge-Based SYSTEMS (KBS)	Optimization
<ul style="list-style-type: none"> Supervised Learning Unsupervised Learning Reinforcement Learning Deep Learning 	<ul style="list-style-type: none"> Scene Reconstruction Motion Analysis Image Restoration Recognition 	<ul style="list-style-type: none"> Expert Systems Intelligent Agents Cased-Based Reasoning Linked System 	<ul style="list-style-type: none"> Evolutionary Algorithms Genetic Algorithms Differential Evolution Particle Swarm Optimization
Robotics	Natural Language Processing (NLP)	Automated Planning and Scheduling	
<ul style="list-style-type: none"> Climbing Actuation Sensing Locomotion 	<ul style="list-style-type: none"> Text Speech 	<ul style="list-style-type: none"> Automated Planning Automated Scheduling 	

Chat Generative Pre-training Transformer, commonly known as ChatGPT, which constitutes the focus of this study, is in the field of NLP. Research in the field of NLP has primarily concentrated on the development of Large Language Representation Models (LLMs). These LLMs aim to enhance the capabilities of NLP by enabling more comprehensive understanding of human language, including tasks like translation, text classification, and conversational interactions. ChatGPT is among the numerous LLMs that have been developed within the realm of NLP research [23].

2.2. Literature Review of AI Use for Risk Management in Construction Projects

Although the potential benefits of digital technologies in construction management are reported in the extant literature, there has been a paucity of empirical research examining the AI use for managing risks [21]. An in-depth literature review was conducted to gain a comprehensive insight into the AI use studies on risk management. To perform a literature review, the Scopus scientific search engine was used due to its comprehensiveness and accuracy compared to the WoS search engine [24]. While the search was performed using the Scopus scientific search engine, (“Artificial intelligence”) OR (AI) OR (ChatGPT) AND (Risk Management) OR (Risk Management Process) AND (construction projects) were used. As a result of the literature review analysis, a total of 30 studies were found. Additionally, conference studies were excluded due to their quality issues and accessibility of research. Thus, a total of 22 articles were derived as a result of the initial analysis. Studies published before 2000 were also extruded. As a result, a total of 20 studies were listed in the database. After that, title and abstract analyses were performed to detect related studies. Thus, the new set of references were narrowed to a total of 12. According to the results of the literature review, AI use in risk management can be categorized into the following subjects:

There exists a substantial number of studies aiming to clarify AI tools and their potential use in general. For example, Chenya et al. (2022) conducted a systematic literature review to find the gaps and future research trends in intelligent risk management [10]. They found that less research was conducted to refine data source and categories, the link between digital management platforms and risk management was not considered frequently, and developing an intelligent decision support system had minor interest in general. Likewise, Khodabakhshian et al. (2023)’s study consisted of investigating and comparatively analyzing the main deterministic and probabilistic methods that can be applied during different phases of risk management in regard to scope, primary applications, advantages, disadvantages, limitations, and proven accuracy [9]. Again, the findings verify that the research areas are clustered under AI-based risk data structuralizing and pre-processing,

and AI algorithm classification for risk identification, analysis, and mitigation planning. Regona et al. (2022) identified the adoption challenges of AI, along with the opportunities offered for reducing low-level AI adoption in the construction industry [7]. To achieve this aim, their study adopted a systematic literature review approach using the PRISMA protocol. Their findings revealed that AI is particularly beneficial in the planning stage since the accuracy of event, risk, and cost forecasting plays a crucial role in construction projects' success. Their findings also revealed that AI use in construction projects provides benefits to shorten the duration of repetitive tasks and improve work processes by using big data. On the other hand, the required data acquisition and retention processes increase difficulties in the incorporation of AI on construction sites due to the fragmented nature of the industry. Another study presented an extensive survey of the applications of deep learning techniques within the construction industry in terms of preventing challenges in structural health monitoring, construction site safety, building occupancy modeling, and energy demand prediction [25]. In the work of Locatelli et al. (2021), the authors explored the potential of NLP and the combination of an NLP and BIM [26].

A literature review shows that AI is also used for predictive risk identification and risk assessment. In their work, Erfani and Cui (2022) stated that traditional expert-based approaches in the field of risk identification create difficulties in projects due to their time-consuming and expensive aspects [27]. In order to overcome these constraints, the authors introduced a data-oriented framework that utilizes historical data and artificial intelligence methods, specifically word-embedding models, to identify risks. The developed framework compares different risk factors from previous projects by taking into account the semantic meanings of words. Tirmizi and Arif (2022) created an AI-based framework that successfully tackles the issues of knowledge management and stakeholder integration in companies [28]. This framework aims to enhance the efficiency and effectiveness of reviewing contract development by pre-identifying risk factors and providing recommendations for risk mitigation strategies. Schwarz and Sánchez (2015) utilized Artificial Neuronal Networks (ANNs) together with a Monte Carlo Simulation for increasing the certainty of data by determining the inputs for the simulation process [16].

A review of the literature shows that a vast number of studies related to AI use for managing risk or uncertainty were carried out from the perspective of safety. For example, Bigham et al. (2018) proposed an AI-based platform for construction safety risks by implementing safety standards by OSHA codes for compliance [29]. Poh et al. (2018) presented a machine learning approach to develop leading indicators that classify sites in accordance with their safety risks in construction projects [30]. In their work, Koç et al. (2021) devised an extensive framework for forecasting the disability status of construction workers after accidents [31]. This framework employed four ensemble machine learning models based on tree-based algorithms: Random Forest, XGBoost, AdaBoost, and Extra Trees. Additionally, a cutting-edge optimization technique called the Genetic Algorithm (GA) was utilized for fine-tuning hyper-parameters during the prediction process.

Yaseen et al. (2020) put a focus on project delay risks. They developed a hybrid AI model with genetic algorithm optimization for predicting delay risks [32]. The achieved results of the developed hybrid model revealed a good resultant performance in terms of accuracy and classification error. The aim of Phasha (2022)'s study was to investigate the beneficial impact of using AI technologies/tools on cost overruns and risk management in construction projects [33]. The predictive models that were derived from conducting regression analysis showed that both cost overruns and risk management might decrease when AI use intensified. From a cost-risk assessment perspective, Cheng et al. (2010) developed an "estimate at completion"-based model to improve the accuracy of project cost estimations by integrating two artificial intelligence approaches, namely the Fast Messy Genetic Algorithm (fmGA) and Support Vector Machine (SVM) [34]. Afzal et al. (2021) compiled the current AI methods used for cost-risk assessment in the construction management domain and found that AI methods were limited in addressing cost overrun

issues under high uncertainty due to the limitation of the subjective risk data and complex computation [11].

Choi et al. (2021) examined AI and text-mining applications for analyzing contractors' risk in invitations to bid and contracts for Engineering Procurement and Construction (EPC) projects [8]. In this study, a Critical Risk Check module for extracting risk-involved clauses and a Term Frequency Analysis module for contractual risk extraction were developed as a digital EPC contract risk analysis tool for contractors. Likewise, Awad and Fayek (2012) put their focus on contractor prequalification [35]. Based on the need for evaluating the contractor and project-specific aspects from the surety bonding perspective, they developed a decision support system by combining fuzzy logic and expert systems.

In their study, Baryannis et al. (2019) noted that the predictive and learning capabilities of AI for supply chain risk management were still in their infancy, as little attention had been given to the development of automated solutions for decision-making [18]. Forbes et al. (2010) developed a Case-Based Reasoning (CBR)-based framework for the selection of the most suitable risk management techniques by using the similarity measuring capability of CBR [36]. In their work, Basaif et al. (2020) determined the level of awareness of Malaysian construction practitioners of using AI for risk analysis [13].

Regarding the studies in the literature related to AI use for risk management, the frequently encountered subjects are related to either performing systematic literature reviews to determine how current AI methods can be applied during different phases of risk management or generating AI-based models to prevent specific risks (i.e., cost-over and delay risks) with the use of AI tools, such as Artificial Neural Networks (ANNs), Bayesian Belief Networks (BBNs), the Genetic Algorithm (GA), and fuzzy hybrid methods. On the other hand, very few studies on evaluating ChatGPT use for risk management in construction projects were encountered.

2.3. Literature Review on ChatGPT Use for Risk Management in Construction Projects

As the number of studies with a focus on the potential benefits and challenges of AI technology use increases, studies related to how ChatGPT can be utilized to improve various aspects of project management have started to emerge [15]. Before summarizing the studies encountering ChatGPT use in risk management, it is vital to highlight that the total number of studies focusing on ChatGPT use in the project management domain is in the minority. The limited studies outside of risk management generally concern the application of ChatGPT in natural disaster prevention and reduction [37,38] and the scheduling of construction projects [23].

Regarding the studies in the literature related to ChatGPT use for risk management, Hofert (2023) investigated the extent to which ChatGPT can grasp concepts from the fields of risk, time series, extremes, and dependence [39]. Even though the study by [39] seems to share the same purpose with the current study, which is to present ChatGPT competency in quantitative risk management, the study's focus was on actuarial practice rather than construction works.

Klepo et al. (2023) conducted a study about AI use in risk management, with a focus on infrastructure projects [40]. Open-source AI (ChatGPT) was engaged for identifying the most critical risks and response strategies. In their work, risk factors effecting infrastructure projects in the water sector were identified and analyzed by project managers. Afterwards, AI was engaged to formulate adequate risk response strategies for the most critical risks. A comparison between the results (PM expert and AI strategies) constituted the findings of the study. It was indicated that the PM experts and AI strategies overlapped. This study put its focus only on the risk management process of a specific project type (EPC-based infrastructure project) and assessed the performance of ChatGPT use only in risk identification and generating a risk response strategy process.

Al-Mhdawi et al. (2023) identified the key indicators for measuring the performance of ChatGPT in managing construction risks based on ISO 31000 by conducting a focus group with nine experts in risk management [14]. Then, the authors quantified the performance

indicators of ChatGPT by measuring the validity and scalability of the identified performance indicators. As the output, Fuzzy Performance Numbers that constituted the ranking of the KPIs were obtained. As part of the methodology, ranking was conducted using a questionnaire survey and processed under a fuzzy environment. This study put its focus only on evaluating ChatGPT's performance against established ISO 31000 standards [41] and did not focus on a specific construction project type.

A literature review on the limited number of the studies dealing with ChatGPT use for risk management in construction projects shows that studies have neither assessed the performance of ChatGPT for all sub-processes of risk management nor performed a comparison of ChatGPT's performance for different type of projects to show if ChatGPT provides consistency and accuracy for different type of construction projects. Taking into account that one of the main features of a "project" is the "uniqueness", and that risk management should be tailored specifically to each project [42], the assessment of ChatGPT's accuracy for each risk management sub-process in different types of construction projects should be analyzed. Therefore, in contrast to the existing literature on ChatGPT use for risk management in construction projects, this study aims to assess the accuracy of ChatGPT use for the sub-processes of risk identification, risk analysis, risk response, and risk monitoring in different project types.

3. Research Methodology

This study has the objectives of determining the accuracy of ChatGPT use in each risk management sub-process (risk identification, risk analysis, risk response, and risk monitoring) in different project types. In this regard, this study adopted a research methodology involving two main phases. Figure 1 presents the details of the research methodology adopted within this study.

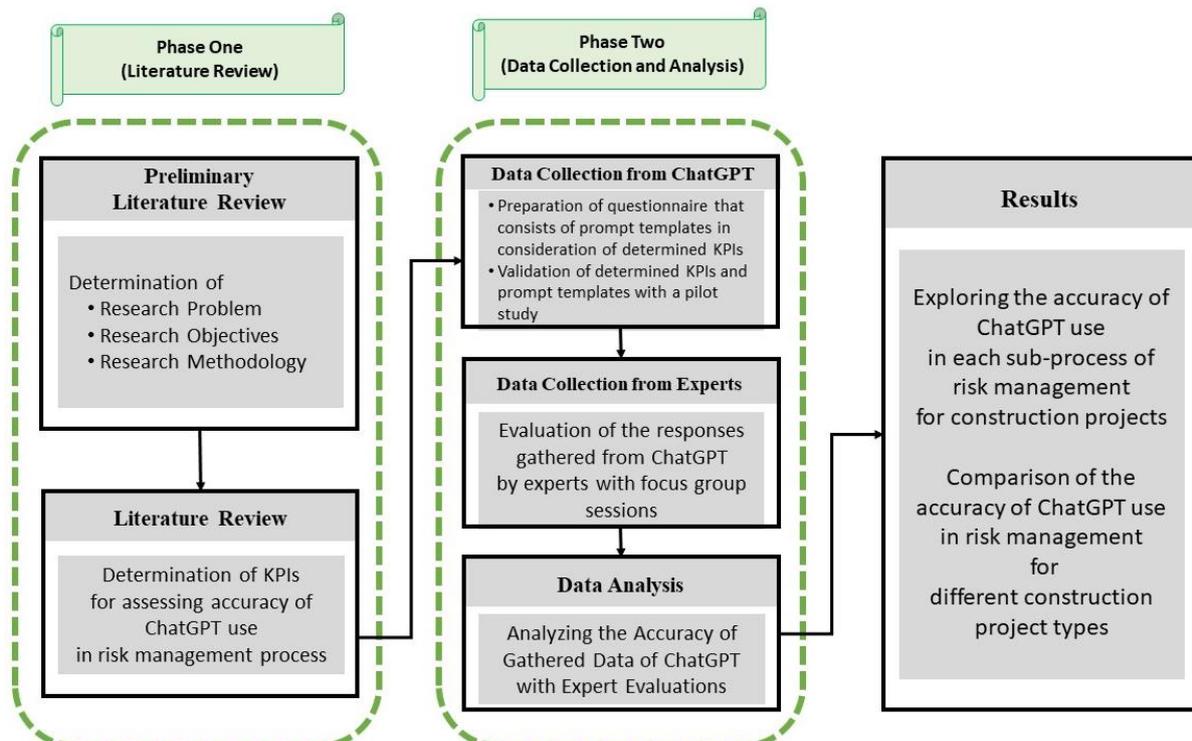


Figure 1. Research methodology.

In the first phase, a literature review was conducted to determine the Key Performance Indicators (KPIs) for assessing the accuracy of ChatGPT use in each risk management sub-process. As the second phase, based on the determined KPIs, a questionnaire that consisted of prompt templates was prepared for collecting data from ChatGPT. The responses

gathered from ChatGPT were then evaluated by experts with focus group sessions. The following sections present the details of the research methodology adopted in this study.

3.1. Determination of KPIs for Assessing the Accuracy of ChatGPT Use in Risk Management Process

In the construction management domain, KPIs are commonly used for in performance measurement because they clearly give a value to compare against the current performance. Therefore, as the first step, the KPIs related to each risk management sub-process for assessing the accuracy of ChatGPT were determined. The determination of these indicators constitute significant importance since the output gathered from ChatGPT was evaluated based on these determined KPIs.

Risk management can be defined as a structured approach that involves a logical sequence of steps (identify risks, perform risk analysis, plan and implement risk responses, and monitor risks) [16]. In existing risk management practices, project managers generate an extensive inventory of potential risks that could impact the project's goals. Once the risks are identified, qualitative and quantitative risk analyses are conducted to prioritize them based on factors such as probability, impact, and potential consequences. Afterwards, risk responses are developed to address the prioritized risks, ensuring the project remains on track and within its constraints [43]. It is also vital to continuously monitor and control risks throughout the project's life cycle to detect if any change exists in terms of new risk formation and/or alteration requirements for determined risk response strategies. Ultimately, regular risk reviews and reassessments are carried out to ensure that the project's risk profile is up to date and that necessary actions are taken accordingly [15]. Therefore, there should be KPIs relevant to the (1) risk identification, (2) risk analysis, (3) risk response, and (4) risk monitoring sub-processes, respectively.

In their study, [14] determined twelve KPIs for measuring the performance of ChatGPT in managing construction risks. Their work set the starting point for determining KPIs. Based on the literature findings related to risk management and AI use in risk management, seven additional KPIs were also obtained for assessing the accuracy of ChatGPT use in the risk management process for construction projects. These are as follows:

Capturing complexity risk interdependencies and correlating identified risks are essential for a more comprehensive risk assessment and accurate decision-making in risk management implementations [11]. However, the current risk management practices are highly criticized for ignoring the causal inferences among risk factors [9]. Therefore, KPI-2, related to the "accuracy of generating relations of identified risks", is included.

Successful and effective risk management requires a clear understanding of the risks faced by the project and business. This involves more than simply listing the identified risks and characterizing them by their probability of occurrence and impact on objectives. The large amount of risk data produced during the risk assessment process must be structured to aid their comprehension and interpretation and to allow them to be used as a basis for action. A hierarchical Risk Breakdown Structure (RBS) framework, similar to the WBS, provides a number of benefits by decomposing potential sources of risk into layers of increasing detail [44]. Considering the importance given to the RBS by PMI, KPI-3, related to RBS, was included.

Risk allocation is an important issue in risk management studies, and a huge number of studies have been dedicated to risk allocation since risk allocation promotes risk mitigation, accountability, cost efficiency, expertise utilization, and overall project success. It helps establish a clear framework for managing risks and ensures that each party assumes responsibility for the risks within their control, leading to smoother project execution and better outcomes. Additionally, appropriate risk allocation between the public and private sectors, according to their risk management capabilities, is crucial for the success of PPP projects [45–47]. Therefore, as a risk management-related study, KPI-7, related to "ability to provide proper risk allocation decisions", was included.

In the work of [15], which presents a comprehensive guide for project managers on effectively using ChatGPT within the context of PMBOK, tailored prompts for various project management tasks were presented: KPI-8: Ability to generate contingent response strategies (mitigation strategies). KPI-9: Ability to provide supportive suggestions for how to monitor risk. KPI-11: Streamlining risk reporting and communication were included after the evaluation of the prompt templates presented by [15].

Consequently, a total of 19 KPIs were determined and categorized based on the main sub-processes mentioned in [43]. Table 2 presents the explanatory information related to the determined KPIs.

Table 2. KPIs for assessing accuracy of ChatGPT use in risk management processes.

Risk Identification	KPI-1	Accuracy of risk identification	The ability of ChatGPT to effectively recognize and assess potential risks that could impact the construction project.
	KPI-2	Accuracy of generating relationships among identified risks	The extent to which ChatGPT can capture complexity–risk interdependencies and correlate identified risks in terms of their interactions.
	KPI-3	Ability to generate Risk Breakdown Structure (RBS)	The extent to which ChatGPT can break down the risks of a project, as a hierarchical outline of risk.
	KPI-4	Ability to generate new risk(s) in correspondence with new circumstances	The ability of ChatGPT to recognize and assess potential risks in terms of new circumstances and/or emerging trends within the construction industry.
Risk Analysis	KPI-5	Ability for risk assessment and prioritization	The degree to which ChatGPT consistently evaluates and prioritizes risks in accordance with the construction project’s objectives, considering factors such as probability and impact.
Risk Response	KPI-6	Ability to provide relevant risk responses	The ability of ChatGPT to propose relevant and effective risk response strategies (such as escalation, avoidance, transfer, mitigation, or acceptance) that align with the specific requirements of the construction project.
	KPI-7	Ability to provide proper risk allocation decisions	The extent to which ChatGPT can specify the accurate stakeholder that should undertake the risk based on industry trends, project-specific factors, etc.
	KPI-8	Ability to generate contingent response strategies (mitigation strategies)	Risk mitigation refers to the risk handling strategy used to eliminate or lessen the likelihood and/or consequence of a risk. In this sense, this pertains to the extent to which ChatGPT can aid project managers in generating contingent response strategies for prioritized risks (such as contingency plans, removal of high-risk elements of scope from the project, etc.), providing a proactive approach to addressing potential issues.
Risk Monitoring	KPI-9	Ability to provide supportive suggestions for how to monitor risk	The extent to which ChatGPT can support project managers in monitoring identified risks, assessing the efficacy of risk response strategies, and providing recommendations for adjustments when necessary.
	KPI-10	Flexibility to customize risk management processes	The degree to which ChatGPT outputs can be tailored and customized to meet the specific requirements and objectives of the construction project.
	KPI-11	Streamlining risk reporting and communication	The extent to which ChatGPT can produce concise and informative risk reports for stakeholders, ensuring they are consistently informed about the project’s risk profile.

Table 2. Cont.

Risk Management Tool Features	KPI-12	Consistency of responses	The degree of consistency in ChatGPT's responses when presented with similar questions or inputs.
	KPI-13	Clarity of communication	The degree of clarity with which ChatGPT can effectively communicate its responses to the user, encompassing factors such as language choice and the level of detail provided.
	KPI-14	Ability to learn and adapt to new information	The degree of ChatGPT's ability to assimilate new information and adapt its responses accordingly.
	KPI-15	Ability to handle multi-language input	The degree of ChatGPT's ability to process input in various languages, which can be valuable for international construction projects.
	KPI-16	Ease of use	The degree of user-friendliness and intuitiveness in ChatGPT, facilitating its easy utilization by project team members for risk management purposes.
	KPI-17	Compatibility with different devices and platforms	The degree of compatibility of ChatGPT with various devices and platforms, including desktop computers, mobile devices, or cloud-based platforms.
	KPI-18	Compliance with industry standards and best practices	The degree of alignment between ChatGPT's risk management processes, and industry standards, and best practices specific to the construction industry.
	KPI-19	Ability to generate data with complex scenarios	The extent to which ChatGPT can give accurate answers in the context of a complex scenario.

3.2. Preparation of Questionnaire

A questionnaire including prompt templates was prepared as a base to carry out a conversation with ChatGPT. In this step, special attention was paid to all the determined KPIs covered by the questions, and the conversation with ChatGPT was repeated considering different project types (PPP-type transportation and energy project). The selection of project types mainly depends on their complexity in terms of stakeholder relations, financial requirement, and scale size. This is because projects with intricate stakeholder relationships, such as those involving multiple organizations, regulatory bodies, or community groups, tend to have higher complexity levels since it becomes more challenging to manage their expectations, align their interests, and address potential conflicts. Additionally, construction projects with large budgets or complex funding structures tend to be more complex due to the intricacies involved in managing financial aspects effectively. The scale of a project can also contribute to its complexity. Larger projects typically involve more stakeholders, resources, and interdependencies. Additionally, larger projects may have more intricate logistical requirements, stricter regulatory compliance, and greater potential impacts, all of which add to the complexity. The PPP-type was specially chosen since PPP projects are generally characterized as risky due to the involvement of many stakeholders, the huge amount of investment, and long concession periods [48]. Likewise, infrastructure projects, such as energy and transportation projects, are, by their nature, large and complex, and often involve new technologies. They also tend to have a high degree of uncertainty due to factors such as the fluctuating price of oil and gas or tariffs. As a result, they can be subject to several risks that need to be carefully assessed.

The questionnaire was tested with a pilot study. The aim of the pilot study was to validate the determined KPIs and to obtain feedback about the accuracy of questions that will be addressed to ChatGPT. In the pilot study, two respondents provided feedback about the questionnaire (one PMP with 12 years of sector experience in international infrastructure projects and one academician who has expertise in both project management and digital technologies). The demographic information of the pilot study participants is given in Table 3.

Table 3. The demographic information of respondents participating the pilot study.

Respondents	Profession	Academic Background	Position	Experience (on Yearly Basis)
Respondent 1	Civil Engineer	Civil Engineering, Ph.D.	Project manager	12 years
Respondent 2	Civil Engineer	Construction Management, Ph.D.	Academician	10 years

The pilot study was conducted with each respondent separately in face-to-face meetings. The face-to-face meetings lasted approximately 2 h. These meetings were structured in three stages. In the first stage, the general AI use for risk management, the extent to which ChatGPT can be used for risk management processes, and the aim of the study were explained briefly. Then, as the second stage, the KPIs for assessing the accuracy of ChatGPT use in risk management process were presented, and each respondent was asked to verify the determined KPIs. In this stage, the respondents were also expected to revise and/or indicate missing KPIs, if any. In the second stage, there was consensus about the sufficiency of the determined KPIs since there were no recommendations from the experts. For the third stage of the pilot study, prompt templates that are used to carry out the conversation with ChatGPT were presented. Each respondent was asked to verify the presented questions and to add, eliminate, and/or revise the questions by considering if all of the questions were relevant with the determined KPIs and risk management sub-processes. The suggested questions were added to the prompt template. The finalized prompt templates can be found in the Supplementary Materials.

3.3. Data Collection

The data collection consisted of two main parts: (1) gathering data from ChatGPT through conversation, (2) gathering data from experts with focus group sessions to evaluate the responses of ChatGPT.

For the first part, a total of 36 questions were addressed to ChatGPT in the context of the determined 19 KPIs. The conversation took place in August 2023, and a free-access ChatGPT version was used for gathering data. A prior trial was carried out using the ChatGPT 3.5 platform. However, another free-access platform was chosen (ChatGPT Demo) due to the time-consuming process of the ChatGPT 3.5 platform in providing responses and not providing efficient answers after a certain amount of questions. ChatGPT Demo is built based on the structure of ChatGPT-4. With advanced machine learning algorithms and a flexible design, one of the most important advantages of ChatGPT Demo is that it allows users to use it for free without having to log in. The conversations were conducted as two trials using a desktop computer (one trial for a PPP-type transportation project in English, one trial for a PPP-type energy project in English). The data gathered from ChatGPT (the conversation history for each trial) can be found in the Supplementary Materials.

For the second part of the data collection, two focus group sessions were conducted with construction experts (one focus group session was conducted with five experts with expertise in PPP-type transportation projects and risk management, and the second one was conducted with five experts with expertise in PPP-type energy projects and risk management). In their study, in which deterministic and probabilistic risk management approaches in construction projects were investigated, [9] pointed out the necessity for the validation of AI-based data gathering and preprocessing tools by experts in the field and/or through case studies for the implementation of algorithms and comparison of the results. Therefore, this study put its focus on the accuracy assessment of ChatGPT for different project types with evaluations from experts in the field within focus group sessions. A focus group interview is a research method that summarizes the opinions of a group in the context of predetermined questions by the researcher(s) in order to reveal the views and attitudes of a small number of experts (5–10 participants) [49]. In focus group interviews, unlike individual interviews, participants interact with each other and influence each other with their experiences and/or perspectives [50]. The most important factors in choosing

this method are as follows: (1) it enables group members to develop different ideas in terms of its interactive nature, (2) it satisfies the researcher more in terms of validity because it is conducted face-to-face, and (3) it creates a chance to obtain a lot of data in a short time. The demographic information of the experts who participated in the focus group sessions (FGSs) is presented in Table 4.

Table 4. Demographic information of experts participating in focus group sessions.

Focus Group Session (FGS)	Expert	Position	Industrial Experience	Experience in PPP Projects	Academic Background
FGS-1 for PPP Transportation Project	Expert 1	Project Manager	12 years	7 years	Civil Engineering, M.Sc.
	Expert 2	Technical Manager	8 years	3 years	Construction Management, M.Sc.
	Expert 3	Planning and Cost Control Executive Manager	12 years	5 years	Civil Engineering, Ph.D.
	Expert 4	Project Manager	10 years	5 years	Civil Engineering, B.Sc.
	Expert 5	Deputy Director Contracts and Administrative	15 years	9 years	Civil Engineering, M.Sc.
FGS-2 for PPP Energy Project	Expert 1	Project Manager	7 years	4 years	Civil Engineering, M.Sc.
	Expert 2	Contract Manager	10 years	3 years	Construction Management, Ph.D.
	Expert 3	Project Manager	13 years	7 years	Construction Management, Ph.D.
	Expert 4	Managing Director	8 years	6 years	Construction Management, M.Sc.
	Expert 5	General Manager	16 years	12 years	Civil Engineering, M.Sc.

During the focus group sessions, the experts were expected to evaluate the accuracy of the responses that were received from ChatGPT. In this regard, the responses gathered from ChatGPT were first shared with experts. Then, the experts were asked to rate the accuracy of the ChatGPT's responses using a seven-point Likert scale as a result of the group's decision. The assignment of the agreement level to scores is as follows: (1) very low; (2) low; (3) low-moderate; (4) moderate; (5) moderate-high; (6) high; (7) very high. The questionnaire that was used to conduct the focus group sessions can be found in the Supplementary Materials.

3.4. Data Analysis

In the data analysis process, the assessment of the quantified performance of ChatGPT was evaluated by the experts for the KPIs between KPI-1 and KPI-11 under risk identification, risk analysis, risk response, and risk monitoring sub-processes. Furthermore, the assessment of ChatGPT's performance as a risk management tool was evaluated by the authors for the KPIs between KPI-12 and KPI-19. The underlying reason for this preference was that the KPIs between KPI-12 and KPI-19 are related to general features of ChatGPT (such as clarity of communication, ability to handle multi-language input, ease of use, compatibility with different devices and platforms etc.), and these can only be evaluated by the author.

In the quantitative analysis of ChatGPT accuracy for the KPIs between KPI-1 and KPI-11, the issue of how to create a group decision that all participants agreed on was a significant problem for the decision-making process, apart from the individual decisions of the experts. In the literature, it is stated that a common group decision can be formed, or the geometric mean of the values given by individual decision makers can be used if there is not consensus as a group decision [51,52]. The quantified performance of ChatGPT was evaluated on a scale of 1 to 7, where 1 represents very poor performance and 7 represents very high performance. The group decision in each FGS reflecting the performance of

ChatGPT for each question can be found in the Supplementary Materials. Through each focus group session, the experts built consensus for the performance of ChatGPT for each KPI. Therefore, adaptation of the geometric mean calculation of individual values was not required. The experts used the conversation history as an input to make a group decision related to the quantitative analysis of the accuracy of ChatGPT.

4. Results and Discussion

The results gathered from the focus group sessions regarding the evaluations of ChatGPT's performance are presented below.

4.1. Evaluation of ChatGPT's Performance in Risk Identification Sub-Process

The evaluation of ChatGPT's performance for risk identification sub-processes consisted of expert evaluation under four main KPIs, which are "accuracy of risk identification (KPI-1)", "accuracy of generating relations of identified risks (KPI-2)", "ability to generate risk breakdown structure (KPI-3)", and "continual improvement and updates (KPI-4)".

- Evaluation of ChatGPT's performance in KPI-1: The experts were additionally encouraged to identify a set of 20 risk factors specific to the construction project for their focus group session. A comparison is presented in Table 5.

Table 5. Comparison of risk factors that were identified by ChatGPT and experts.

	Risk Factors Identified by ChatGPT	Risk Factors Identified by Experts
	(√) 1. Political instability and changes in government policies.	(√) 1. Public opposition
	(√) 2. Regulatory and legal risks, including changes in legislation and regulations.	(x) 2. Termination of concessionaire by public institution.
	(√) 3. Delays in obtaining necessary permits and approvals.	(√) 3. Failure/delay in obtaining permits/approvals.
	(√) 4. Currency exchange rate fluctuations.	(√) 4. Nationalization/expropriation.
	(x) 5. Economic downturn or recession.	(√) 5. Financial risks associated with project funding (inflation, interest rate fluctuation, foreign exchange fluctuation, taxation, etc.).
	(√) 6. Construction and infrastructure risks, such as delays, cost overruns, and quality issues.	(√) 6. Market risks associated with tariff and demand changes, etc.
	(√) 7. Environmental risks, including potential impact on ecosystems and local communities.	(√) 7. Environmental concerns (environmental impact of the project, adverse site conditions, weather conditions).
	(√) 8. Public opposition and protests against the project.	(x) 8. Project objective/purpose differences among stakeholders (lack of commitment among stakeholders, etc.).
	(√) 9. Land acquisition and resettlement issues.	(x) 9. Purchasing guarantees by public institution.
	(√) 10. Financing risks, such as difficulty in securing funding or changes in interest rates.	(x) 10. Payment mechanisms
	(x) 11. Force majeure events, such as natural disasters or acts of terrorism.	(√) 11. Inadequate contract management (inadequacy of plans and specifications, inadequacy in contract enforcement).
FGS-1	(√) 12. Traffic and demand risks, including lower-than-expected usage or changes in travel patterns.	(x) 12. Integration between design and construction phases.
	(x) 13. Technology and operational risks, such as system failures or cybersecurity threats.	(√) 13. Inadequate government regulations (unstable policies, lack of regulation regarding PPP).
	(x) 14. Performance and maintenance risks, including the need for costly repairs or upgrades.	(√) 14. Completion risk in terms of delays in project timeline.
	(√) 15. Dispute resolution and contract enforcement risks.	(x) 15. Excessive design variations and construction changes.
	(√) 16. Inflation and cost escalation.	(x) 16. Wrong estimation of cost trade-offs.
	(√) 17. Revenue and cash flow risks, including lower than projected toll or fare revenues.	(√) 17. Revenue (income) risk.
	(√) 18. Market and competition risks, such as changes in market conditions or new competitors.	(√) 18. Social and political instability in host countries.
	(x) 19. Reputation and public image risks, including negative media coverage or customer dissatisfaction.	(x) 19. Improper risk-allocation decisions.
	(x) 20. Exit strategy risks, such as difficulty in selling or transferring the project at the end of the concession period.	(x) 20. Inadequate relationships with employer (public).

Table 5. Cont.

	Risk factors identified by ChatGPT	Risk factors identified by experts
FGS-2	(√) 1. Political instability and changes in government policies.	(√) 1. Completion risk in terms of delays in project timeline.
	(√) 2. Regulatory and legal uncertainties.	(√) 2. Increased costs due to unforeseen circumstances (Construction cost overrun + Operation cost overrun).
	(√) 3. Delays in obtaining necessary permits and approvals.	(√) 3. Changes in government regulations (unstable renewable energy policies; changes in tax and tariff policies).
	(√) 4. Fluctuations in energy prices and market conditions.	(√) 4. Environmental concerns (environmental impact of the project, adverse site conditions, weather conditions).
	(√) 5. Currency exchange rate fluctuations.	(x) 5. Health and safety concerns.
	(√) 6. Changes in tax and tariff policies.	(√) 6. Technological challenges.
	(√) 7. Environmental and social impact concerns.	(√) 7. Social and political instability in host countries.
	(x) 8. Volatility in fuel prices and availability.	(√) 8. Financial risks associated with project funding (inflation, interest rate fluctuation, foreign exchange fluctuation, taxation, etc.).
	(√) 9. Construction delays and cost overruns.	(√) 9. Market risks associated with energy commodity prices, and demand change, etc.
	(√) 10. Technical and operational risks.	(√) 10. Disruption in supply chain due to force majeure events.
	(√) 11. Supply chain disruptions.	(√) 11. Risk of terrorism and strikes.
	(√) 12. Natural disasters and extreme weather events.	(x) 12. Liquidity risks based on non-existence of secondary market and long payback period.
	(√) 13. Security and geopolitical risks.	(x) 13. Credit risk based on default of renewable energy projects.
	(√) 14. Labor disputes and strikes.	(√) 14. Risks based on non-existence of required infrastructure.
	(√) 15. Inadequate infrastructure and grid connection challenges.	(x) 15. Insufficient project finance supervision.
	(√) 16. Project financing and funding risks.	(x) 16. Inability of concessionaire.
	(x) 17. Counterparty risks, including non-payment or contract breaches.	(x) 17. Imperfect law and supervision system.
	(x) 18. Public opposition and community resistance.	(x) 18. Inadequate contract management (inadequacy of plans and specifications, contract enforcement).
	(√) 19. Changes in energy policies and regulations.	(√) 19. Delay in project approvals and permits.
	(x) 20. Inadequate risk management and mitigation strategies.	(x) 20. Lack of support infrastructures.

“√” Included by experts. “x” Not included by experts. Risk factors are not listed in a specific order.

For FGS-1, the general risk identification performance of ChatGPT was found to be moderate. The responses of ChatGPT covered a vast scale of generic risk factors related to political–legal, financial, and environmental issues, which are expected to occur in any PPP-type project. In addition, risk factors that might have a huge impact on project preferability, such as tariffs, demand change, and insufficiency in meeting end-user expectations for a transportation project, satisfied the focus group participants. The exit strategy risk identified by ChatGPT was approved as a proper risk factor by the experts. Likewise, focusing on the operational phase and mentioning technology and operational risks and performance and maintenance risks was determined to be appropriate. Although the majority of the risks determined by ChatGPT were also included by the experts, the absence of identifying risk factors related to stakeholder-associated risks was found to be a significant inadequacy since PPP requires a merit partnership between a public institution and the concessionaire of the project. Not emphasizing the risk factors related to the integration of design and construction phases or related issues was detected as a significant deficiency since PPP-transportation projects are generally implemented as fast-track projects. Responses related to the underlying reasons for risk identification were found to be appropriate, whereas the performance in the classification of suggested risk factors into main risk categories were comparatively not identified appropriately (some risks overlapped or fell into multiple categories).

For FGS-2, the general risk identification performance of ChatGPT was found to be moderate-high. The responses of ChatGPT covered a vast scale of generic risk factors related to political–legal, financial, and environmental issues, which are expected to occur

in any PPP-type project. Likewise, risk factors referring to fluctuations in energy prices and market conditions and changes in energy policies and regulations were expected to be the be-all and end-all factors for an energy project. Having a focus on the operational phase and mentioning technology and operational risks was determined to be appropriate. The absence of identifying risk factors related to stakeholder-associated risks was found to be a significant inadequacy since PPP requires a merit partnership among parties. Another important deficiency suggested was that “Lack of support infrastructures” was not covered by ChatGPT since the experts placed importance on this risk factor. Deeming “risk management and mitigation strategies” as an important risk factor was also found as a good addition since the risks should be undertaken by the most competent party in PPP projects where the partnership in risk management should be sustained. In addition, the risk identification performance of ChatGPT for PPP-type energy projects in different host countries was considered to be appropriate since it paid attention to the differentiations due to variations in the political, regulatory, economic, and social contexts of the two countries. In brief, the responses related to the underlying reasons for risk identification and classification were found to be appropriate. On the other hand, despite its moderate success in risk identification, ChatGPT failed to classify these risk factors. It provided a broad overview of the main risk groups that are specific to project types, but some risks overlapped or fell into multiple categories.

- Evaluation of ChatGPT’s performance in KPI-2: The general performance of ChatGPT for generating relationships among the identified risks was found to be sufficient in each focus group session. It was seen that ChatGPT considered cause-and-effect relationships or influences on each other, leading to potential correlations or dependencies. However, in some cases, ChatGPT failed to broaden the generated relationships among the identified risks with solid explanations.
- Evaluation of ChatGPT’s performance in KPI-3: ChatGPT gave a breakdown of the identified risks by using the classification that it provided previously. However, it failed to visualize this information in a visual representation as a hierarchical structure or flowchart.
- Evaluation of ChatGPT’s performance in KPI-4: The ability of ChatGPT in recognizing and assessing the potential risks in the context of new variables was found to be high. In each trial, it gave a new set of risk factors by taking into account the specific variables.

Consequently, Table 6 presents the overall evaluation of ChatGPT’s performance in risk identification sub-processes (with the number of questions in parentheses).

Table 6. Evaluation of ChatGPT’s performance in risk identification sub-processes.

	FGS-1		FGS-2	
	Group Decision Score	Overall Score	Group Decision Score	Overall Score
KPI-1	3/7	3.75 (10)	KPI-1	5/7
KPI-2	4/7		KPI-2	5/7
KPI-3	2/7		KPI-3	2/7
KPI-4	6/7		KPI-4	5/7
				4.25 (10)

The group decision score reflects the performance of ChatGPT in each KPI, and it was shown out of 7. The overall scores represent the mean value of the group decision scores for each KPI.

The findings indicate that the general risk identification performance of ChatGPT was moderate. The responses of ChatGPT covered a vast scale of generic risk factors. Additionally, it was emphasized that some expected be-all and end-all risk factors were neglected by ChatGPT, which reveals a weakness to some extent. However, it should be noted that ChatGPT presents the risk factors based on the specific context of PPP trans-

portation and energy projects for a specific host country based on industry knowledge and experience. The host country's political, economic, regulatory, and social factors contribute to the interplay among different risk factors. Conducting a thorough risk assessment considering the unique characteristics of the project will provide a more comprehensive understanding of the risk factors and their potential impacts on specific projects. Therefore, there will be always a need to provide insights based on historical data, industry trends, and project-specific factors.

4.2. Evaluation of ChatGPT's Performance in Risk Analysis Sub-Processes

The evaluation of ChatGPT's performance in risk analysis sub-processes consisted of the expert evaluations of its "ability for risk assessment and prioritization (KPI-5)". Thus, questions for understanding how ChatGPT prioritized the set of risk factors were addressed through the trials. ChatGPT's performance in risk analysis sub-processes can be emphasized as insufficient since the experts' evaluations were as follows.

For FGS-1, the experts found ChatGPT's ability for "risk assessment and prioritization" was very low. It was seen that ChatGPT assigned a high importance level for the political and regulatory risks of the specified host country, and to a certain extent, this was considered by the experts as appropriate. Nonetheless, assigning a low importance level for currency exchange rate fluctuations for the host country was not found to be an accurate response by the experts. Additionally, it was reported that there was inconsistency in determining the importance levels of risk factors. For example, the importance level was assigned as high for financing risks, such as difficulty in securing funding or changes in interest rates, whereas "currency exchange rate fluctuations", which is also a financial risk, was considered to have low importance by ChatGPT. The experts highlighted that the risk rankings provided by ChatGPT appeared to be a general indication and should be further tailored and refined. It was observed by the experts that ChatGPT did not change the ranking orders when a specific circumstance for the contractor party was defined. This situation made the experts think that the direct or indirect effects of the new specific circumstance on the other risk factors were not considered by ChatGPT. Similarly, when ChatGPT was asked to repeat its calculation by considering the probability and impact features for each risk factor for another host country, it gave the exact same risk rankings, which indicates that the accuracy of its ability in risk assessment and prioritization was not sufficient and only reflected a generic perspective. Another inconsistency in risk prioritization indicated by the experts was the logic behind how ChatGPT generated the risk importance rankings. It was emphasized that the rankings of the risks were generated based on a combination of their importance level and their potential impact on the success and viability of PPP transportation projects in the host country. However, when it was asked how the numeric values for the importance level and impact were calculated, ChatGPT argued that it used a hypothetical approach based on general industry knowledge and common risk assessment practices. ChatGPT also clarified that it did not have direct access to external sources, real-time data, any industrial statistical data, or factors, such as the frequency of past occurrences, which made the accuracy of its ability for risk assessment and prioritization questionable to the experts. Its ability to extract the risk-ranking order as a visual "probability and impact matrix" was also found to be insufficient. It generated a probability and impact matrix" by representing the impact and probability levels on a three-level evaluation scale (high/moderate/low). Therefore, the situation of not using a more sensitive scale led ChatGPT to generate a probability and impact matrix that did not show all the risk factors. In addition, it did not use a red, amber, green (RAG) rating to signify the risk scores.

For FGS-2, the experts also found ChatGPT's ability for "risk assessment and prioritization" to be low. As a result of ChatGPT's risk prioritization, "political instability and changes in government policies" was determined to be the most important risk factor, whereas "changes in energy policies and regulations" was determined to be the least important risk factor. The underlying reasons for this determination were supported by

several solid considerations. When the logic behind how the risk importance rankings were generated by ChatGPT was evaluated, it was emphasized that the rankings of the risks were generated based on a combination of factors and considerations, such as their impact on project stability, potential for financial implications, degree of predictability and controllability, and environmental and social impact concerns. However, when it was asked how the numeric values for these factors were calculated, ChatGPT argued that it used a hypothetical approach based on general industry knowledge and common risk assessment practices. It also noted that the provided rankings were a general guideline and they were not generated as an output by using a specific case study and historical data, regulatory frameworks, and industry standards. It also did not refer to any best practices in the industry or mention the exact names of any industry standards or regulatory framework related to PPP-type energy projects that can be used by professionals dealing with risk management. It only gave references to risk management standards, such as ISO 31000 and COSO Enterprise Risk Management (ERM), from the risk management perspective. In conclusion, all these inefficiencies made the accuracy of ChatGPT's ability for risk assessment and prioritization questionable to the experts. Another inconsistency in its risk prioritization indicated by the experts was the basic calculation mistakes made by ChatGPT. It put a risk factor with a low-risk probability level and a moderate impact level one step ahead of a risk factor that had moderate risk probability level and a moderate impact level. In addition, its ability to show the importance level of risk factors with a visual probability and impact matrix was found to be inadequate. According to ChatGPT's previous risk prioritization calculations, five risk factors playing a critical role in project success varied from the risk factors that were mapped out from the probability and impact matrix. This confusion arose because of the oversight of ChatGPT, i.e., it replaced a risk factor that had low probability and impact levels into the grid as having high probability and impact levels. This mistake might result in a wrong estimation for project managers who are assessing the potential impact and effort required for each task in a project. ChatGPT's accuracy in risk assessment and prioritization depending on new risks and emerging trends in the construction industry was similar since it presented quite a similar risk ranking without a broadened perspective considering the direct and indirect effects of the new situation on the other risk factors.

Consequently, Table 7 presents the overall evaluation of ChatGPT's performance on risk analysis sub-processes (with the number of questions in parentheses).

Table 7. Evaluation of ChatGPT's performance in risk analysis sub-processes.

	FGS-1		FGS-2	
	Group Decision Score	Overall Score	Group Decision Score	Overall Score
KPI-5	1/7	1 (16)	KPI-5	2/7 2 (16)

The group decision score reflects the performance of ChatGPT in each KPI and it is shown out of 7. The overall scores represent the mean value of the group decision scores for each KPI.

The evaluations show that ChatGPT was lacking in risk assessment and prioritization. In risk management implementations, decision-makers may consider additional characteristics of risk, such as urgency, proximity, dormancy, manageability, controllability, detectability, connectivity, strategic impact, and propinquity, when prioritizing individual project risks for further analysis and action. In contrast, ChatGPT generates its responses by utilizing a combination of licensed data, data generated by human trainers, and publicly available data. It has not been directly trained on specific industry reports, best practices, or regulatory frameworks, nor does it have access to proprietary databases or classified information. In parallel to [14]'s findings, caution must be exercised when relying solely on ChatGPT for risk assessment and prioritization. It is still important to perform a manual review and verification to ensure the accuracy and reliability of its risk assessments.

Although, the risk assessment and prioritization of ChatGPT might serve as a general indication and starting point for risk prioritization by the practitioners, it is crucial to gather project-specific information and involve the expertise of relevance stakeholders. But at this point, it is important to highlight that ChatGPT's unsuitability did not only arise from its lack of providing a project-specific risk assessment due to estimating the precise numerical equivalent for risk probability and impact with a hypothetical approach. In the face of a new variable that will cause a new environment or a specific need or a goal, ChatGPT failed to regenerate its assessment and prioritization with the consideration of the direct or indirect effects of the new situation on the other risk factors. Ultimately, the numerical equivalents of risk impact should be determined through a comprehensive risk assessment process that aligns with the project's needs and requirements. Therefore, ChatGPT use all by itself in risk assessment and prioritization is not recommended since it requires insights, consulting, and guidance from professionals or risk management specialists. It is recommended to refer to authoritative sources, such as industry publications, government agencies, relevant professional organizations, and consulting firms that specialize in risk assessment and management. These sources can provide the most accurate and current information for specific needs.

4.3. Evaluation of ChatGPT's Performance in Risk Response Sub-Processes

The evaluation of ChatGPT's performance in risk response sub-processes consisted of expert evaluations under three main KPIs, which are the "ability to provide a relevant risk response (KPI-6)", "ability to provide proper risk allocation decisions (KPI-7)", and "ability to generate contingent response strategies (KPI-8)".

- Evaluation of ChatGPT's performance in KPI-6: In both trials, the experts found ChatGPT's ability to "provide a relevant risk response" as very low. The experts especially emphasized that ChatGPT generally neglected more possible effects and did not suggest any proper risk mitigation measures for these circumstances. There was consensus that ChatGPT suggested more common actions that did not reflect a detailed plan. In addition, especially in trial 1, when the underlying reasons for choosing the risk response strategy was asked, it was seen that ChatGPT started to give inconsistent answers in such a way that it switched to its previous answers each time the same question was requisitioned.
- Evaluation of ChatGPT's performance in KPI-7: It was seen that in each focus group session, ChatGPT provided more accurate responses in providing proper risk allocation decisions. It also gave clear and well-defined reasons for how it provided these allocation decisions, which builds up trust for its user(s).
- Evaluation of ChatGPT's performance in KPI-8: In each focus group session, ChatGPT's performance for generating contingent response strategies for the prioritized risks in the context of a certain emerging situation was found to be very high. It provided clear and well-defined explanations of how the suggested contingent response strategies ensured project continuity.

Table 8 presents the overall evaluation of ChatGPT's performance in risk response sub-processes (with the number of questions in parentheses).

Table 8. Evaluation of ChatGPT's performance in risk response sub-processes.

	FGS-1		FGS-2	
	Group Decision Score	Overall Score	Group Decision Score	Overall Score
KPI-6	2/7	4.33 (6)	KPI-6	3/7
KPI-7	5/7		KPI-7	6/7
KPI-8	6/7		KPI-8	7/7

The group decision score reflects the performance of ChatGPT in each KPI and it was shown out of 7. The overall scores represent the mean value of group decision scores for each KPI.

4.4. Evaluation of ChatGPT's Performance in Risk Monitoring Sub-Processes

The evaluation of ChatGPT's performance in risk monitoring sub-processes consisted of expert evaluations under three main KPIs, which were its "ability to provide supportive suggestion how to monitor risk (KPI-9)", "flexibility to customize risk management processes (KPI-10)", and "streamlining risk reporting and communication (KPI-11)".

- Evaluation of ChatGPT's performance in KPI-9: In FGS-1, ChatGPT's performance in providing supportive suggestions of how to monitor risk factors was found to be reasonable and proactive. The experts emphasized that project managers can effectively monitor the identified risk factors by implementing ChatGPT's suggestions. They also agreed that the suggested risk monitoring approach may enable project managers to take timely actions to mitigate the impact and ensure the overall success of PPP-type transportation projects. On the other hand, in FGS-2, ChatGPT's performance in providing supportive suggestions on how to monitor risk factors was found to be not as successful as ChatGPT's performance of FGS-1, as the suggestions made by ChatGPT provided guidance from a narrower perspective. The experts emphasized that they should be broadened, with more inclusive and explanatory suggestions. That way, project managers can stay informed, anticipate potential challenges, and take more appropriate actions.
- Evaluation of ChatGPT's performance in KPI-10: There were no additional questions for ChatGPT for this KPI. However, for the experts, the whole conversation history revealed insight into whether the gathered responses of ChatGPT were flexible in terms of fitting the specific needs and goals of the construction project. There was consensus that, to a certain degree, ChatGPT customized its responses to fit a specific need/goal/emerging situation from a narrower perspective. However, it is believed that as more precise information/data was sustained for ChatGPT, it had the potential to customize its outputs to fit the specific needs and goals of the construction project.
- Evaluation of ChatGPT's performance in KPI-11: An incoherent situation was witnessed. In trial 1, ChatGPT created a template of an informative risk report for project managers by informing in each sub-heading what kind of information/data should be inserted into this template. Although it served as a good example for project managers by revealing what kind of information a risk report should contain, ChatGPT did not create an actual risk report in the context of its conversation history. User(s) should customize the content at the end. Then, in trial 2, ChatGPT created a template of an informative risk report for project managers in the context of its conversation history. It presented a summary on the risk assessment of the top risk factors, risk response strategies, and recommendations. It also gave a summary in the conclusion sub-heading. Another important deficiency in streamlining risk reporting and communication was the issue of generating a text-based report without visual representation.

Consequently, Table 9 presents the overall evaluation of ChatGPT's performance in risk monitoring sub-processes (with the number of questions in parentheses).

Table 9. Evaluation of ChatGPT's performance in risk monitoring sub-processes.

	FGS-1		FGS-2	
	Group Decision Score	Overall Score	Group Decision Score	Overall Score
KPI-9	6/7	4.33 (4)	KPI-9	4/7
KPI-10	3/7		KPI-10	3/7
KPI-11	4/7		KPI-11	5/7

The group decision score reflects the performance of ChatGPT in each KPI, and it is shown out of 7. The overall scores represent the mean value of group decision scores for each KPI.

The findings indicate that ChatGPT has a moderate level of suitability for providing supportive suggestions on how to monitor risks. It should be kept in mind that the suggestions made by ChatGPT provide guidance from a narrower perspective. Thus, they should be broadened with more inclusive and explanatory suggestions. When considered from this perspective, the experts' judgements are in line with [40]'s findings, which indicate that while AI can provide valuable and thought-provoking strategies to consider, the assessment of their suitability and implementation within a given context still requires human effort. The experts also agreed that the suggested risk monitoring approach may enable project managers to take timely actions to mitigate the impact and ensure the overall success of the project.

4.5. Evaluation of Tool Features of ChatGPT's Performance

The evaluation of ChatGPT's performance in Risk Management Tool Features were performed under eight main KPIs, which were "consistency of responses (KPI-12)", "Clarity of communication (KPI-13)", "ability to learn and adapt to new information (KPI-14)", "ability to handle multi-language input (KPI-15)", "ease of use (KPI-16)", "compatibility with different devices and platforms (KPI-17)", "compliance with industry standards and best practices (KPI-18)", and "ability to generate data in a complex scenario (KPI-19)" respectively. As it was indicated previously, no additional questions were addressed to ChatGPT for the KPIs between KPI-12 and KPI-19. All the ChatGPT responses were evaluated by the author if the gathered responses were in line with the expected outcomes of the related KPIs. This preference was made since the performance of ChatGPT under these KPIs was highly observable as a result of user experience, which, in this scenario, was the author. Tracking the conversation history also aids in understanding the basis of how the evaluations were performed by the author. The conversation history with ChatGPT can be found in the Supplementary Materials.

- Evaluation of ChatGPT's performance in KPI-12: It can be said that ChatGPT did not give consistent responses all the time. Even though it provided more accurate answers in trial 2 in comparison to trial 1, there were many times ChatGPT gave answers inconsistent with its previous answers. Therefore, the user tried to clarify the situation with follow-up questions within the conversation. ChatGPT can also make mistakes in basic calculations, which can lead decision makers to act with overlooked outputs if the mistakes and/or inconsistency escape their attention. With respect to this, the general performance of ChatGPT in providing consistent responses to similar questions or inputs was unfortunately found to be ineffective by the experts in each focus group session.
- Evaluation of ChatGPT's performance in KPI-13: As an overall review, ChatGPT uses clear language. It is easy to understand ChatGPT's responses, but it fails sometimes in sustaining detailed answers for specific circumstances.
- Evaluation of ChatGPT's performance in KPI-14: In each trial, it was seen that ChatGPT's performance in generating strategies based on risk allocation decisions for newly defined situations was satisfactory. However, its general performance in learning and adapting to new information seemed to be not so satisfying since it was observed by the experts that, in some cases, ChatGPT did not learn from the information given during the conversation. As a result, users tried to clarify the situation with follow-up questions within the conversation.
- Evaluation of ChatGPT's performance in KPI-15: With this aim, a set of five questions (one question from each of the five sub-processes was chosen) was asked to ChatGPT in the author's native language. The logic behind executing a conversation in a different language was to assess the ability of ChatGPT to handle multi-language input. It can be said that ChatGPT's performance in handling input in different languages was high, but its performance in sustaining precise output in multiple languages was very low. This is the reason ChatGPT provided totally different responses when the same questions were asked in English and in another language.

- Evaluation of ChatGPT's performance in KPI-16: Users can use the chosen ChatGPT platform for free without needing to log in. Users are expected to enter their questions into a chat box in text format. From this perspective, the chosen ChatGPT platform is very easy to use. Through the conversation, if there is a misspelling, it allows the users to correct it. Additionally, there is a sidebar that allows the users to access their chat logs. Users can continue to carry on their conversation whenever they want. On the other hand, users are not allowed to insert or extract visual information/data (jpeg, png, dwg, etc.). The conversation history cannot be saved as a Word document or in pdf. format for the chosen platform. One of the deficiencies of ChatGPT in risk management was found to be related to data representation. In risk management, frequently used data representation techniques include "probability and impact matrix" and "hierarchical charts". As a text-based tool, ChatGPT did not provide templates, matrix, hierarchical charts, etc. with visual presentation. Therefore, users have to convert text-based information/responses to better suit their specific needs. When it is considered that a picture paints a thousand words, ChatGPT lacks in using the power of visuals.
- Evaluation of ChatGPT's performance in KPI-17: With this aim, a set of five questions (one question from each of the five sub-processes was chosen) was asked to the same ChatGPT platform that was used in the previous trials on a mobile device. For the trial, which was carried out on the mobile device, it provided the exact same answers as that carried out on a desktop computer.
- Evaluation of ChatGPT's performance in KPI-18: Through all trials, ChatGPT always underlined that as an AI language model, its responses were generated based on a mixture of licensed data, data created by human trainers, and publicly available data. In addition, it had not been directly trained on specific industry reports, best practices, or regulatory frameworks. Through all trials, ChatGPT also did not refer to any best practices in the industry, and did not mention the exact names of any industry standards, regulatory framework related to PPP-type transportation or energy project that can be used by the professions dealing with risk management. It only referred to risk management standards, such ISO 31000 [41] and COSO ERM [53], from the risk management perspective. As a conclusion, all these inefficiencies make ChatGPT's compliance with industry standards and best practices questionable.
- Evaluation of ChatGPT's performance in KPI-19: Based on the performed trials and focus group evaluations, it can be inferred that ChatGPT generally provided general insights in terms of commonly known information and widely accepted practices. But if more precise and specific information was provided, it has the potential to give accurate answers in the context of complex scenarios. For example, in trial 1, ChatGPT was asked to perform a risk analysis using the decision tree method with a given complex scenario, as follows, and then asked to detect the riskier seller. *"The contractor first assesses the options available regarding the outsourcing of the construction materials. Here he has two options. He can either outsource from a native seller or an overseas seller. As per each option, there are two potential outcomes. While the native seller would allow the contractor to personally inspect, it is costlier. On the other hand, the overseas seller might be cheaper, but the travel expenses won't allow inspection of materials. While the Expected Monetary Value (EMV) for native sellers is 80,000 with an 80 percent chance of success. When the EMV of the overseas seller is calculated, with merely a 50 percent chance of success, the loss EMV is 15,000."* ChatGPT did a pretty good job of analyzing both situations using a decision tree. It constructed a decision tree based on the information provided and detected that the native seller option would be less risky compared to the overseas seller option.

Consequently, Table 10 presents the overall evaluation of ChatGPT tool features' performance (with the number of questions in parentheses).

Table 10. Evaluation of ChatGPT tool features' performance.

Performance of ChatGPT' Tool Features Based on Author's Experience			
KPI-12	3/7	KPI-16	6/7
KPI-13	5/7	KPI-17	7/7
KPI-14	3/7	KPI-18	2/7
KPI-15	6/7	KPI-19	4/7

4.5

The group decision score reflects the performance of ChatGPT in each KPI and it is shown out of 7. The overall scores represent the mean value of group decision scores for each KPI.

Table 11 presents ChatGPT's performance in risk management processes as a combined table. The overall scores for risk identification, risk analysis, risk response, and risk monitoring represent the mean value of the group decision scores for each KPI, whereas the overall score for ChatGPT tool features' performance was determined by the author. The group decision score reflects the performance of ChatGPT in each KPI and it is shown out of 7 (i.e., for KPI-1, the representation of "3/7" means that the group decision score for the accuracy of ChatGPT's responses for KPI-1 is 3 out of 7). In Table 11, the accuracy of ChatGPT for each risk management sub-processes is also presented as the overall score. The overall scores were calculated as the mean value of the group decision scores of each KPI under related risk management sub-processes (i.e., for risk identification sub-processes, 3.75 presents the mean value of the group decision score for KPI-1, KPI-2, KPI-3, and KPI-4). The quantified performance of ChatGPT was evaluated out of 7 (1 represents very poor performance and 7 represents very high performance of ChatGPT).

This study had the objectives of analyzing the accuracy of ChatGPT use for risk management in different construction project types. The results individually gathered from focus group sessions indicate that the overall performance of ChatGPT in each risk management sub-process was close in value. Furthermore, ChatGPT's performance for planning risk management activities, identifying a set of risks, estimating prioritization for identified risks, and suggesting response strategies for the different project types varied in accordance with the context differentiation in project types. However, the slightly better performance of ChatGPT in energy projects may have arisen due to several reasons. One possibility is that the training data used for ChatGPT may have been biased towards energy projects, resulting in a better understanding and performance in that domain compared to transportation projects. Another factor could be the possibility that ChatGPT's pre-trained models and underlying algorithms may be more suited to handling risk factors in energy projects, leading to better suggestions in that area.

The study's findings indicate that ChatGPT has a moderate level of performance in managing risk. It provided more accurate knowledge in risk response and risk monitoring compared to risk identification and risk analysis sub-processes for both project types. In line with [40]'s work, in which AI was engaged for formulating adequate risk response strategies for infrastructure projects, the study's findings reveal that AI provided some additional good and thought of worthy strategies to consider in risk response sub-processes, but human endeavor was needed in the assessment of the suitability and implementation of them within the given context. In contrast, the aspect of ChatGPT's performance that was found to be the least effective in managing construction risks was its consistency in risk assessment and prioritization [14]. This difference might arise due to the specific nuances and intricacies of PPP-type projects. It may present unique challenges that ChatGPT is not as adept at addressing when compared to the evaluation in a general context. Nonetheless, it can be concluded that the use of AI for risk management holds promise as a tool, but the overlapping of human and artificial intelligence is highly needed for efficient risk management in construction projects.

Table 11. Average values for ChatGPT's performance.

Risk Identification		Risk Analysis			Risk Response			Risk Monitoring			Risk Management Tool Features		Overall Score			
	Group Decision Score	Overall Score														
FGS-1	KPI-1	3/7	3.75 (10)	KPI-5	1/7	1 (16)	KPI-6	2/7	4.33 (6)	KPI-9	6/7	4.33 (4)	KPI-12	3/7	Based on author's experience	4.5
	KPI-2	4/7			KPI-7		5/7	KPI-10		3/7	KPI-13		5/7			
	KPI-3	2/7			KPI-8		6/7	KPI-11		4/7	KPI-14		3/7			
	KPI-4	6/7									KPI-15		6/7			
FGS-2	KPI-1	5/7	4.25 (10)	KPI-5	2/7	2 (16)	KPI-6	3/7	5.33 (6)	KPI-9	4/7	4 (4)	KPI-16	6/7		
	KPI-2	5/7			KPI-7		6/7	KPI-10		3/7	KPI-17		7/7			
	KPI-3	2/7			KPI-8		7/7	KPI-11		5/7	KPI-18		2/7			
	KPI-4	5/7									KPI-19		4/7			

5. Conclusions

This study explored the performance of ChatGPT for risk-based decision making in the construction industry. In line with this aim, first, a total of 19 KPIs related to risk management sub-processes were determined for assessing the accuracy of ChatGPT. Then, based on the determined KPIs, a questionnaire consisting of prompt templates was prepared for collecting data from ChatGPT. In the context of the prompt templates, two trials were performed with ChatGPT (one trial for a PPP-type transportation project, and one trial for a PPP-type energy project). The responses gathered from ChatGPT were then evaluated by experts with focus group sessions. In the study, two focus group sessions were conducted separately with construction experts who specialized in the specific project type and risk management. The findings of the study indicate that ChatGPT has a moderate level of performance in managing risk. It provided more accurate knowledge in risk response and risk monitoring compared to risk identification and risk analysis sub-processes within diverse project contexts. In the below sub-sections, along with the theoretical and practical implications of the research, the limitations and recommendations for further studies are emphasized.

5.1. Implications for Researchers

To the best of our knowledge, this is one of the pioneer studies contributing to today's state-of-the-art risk management domain. Based on an examination of the studies in the literature related to ChatGPT use for risk management, the novelty of this study is its focus on the evaluation of ChatGPT's performance in each risk management sub-process for different construction project types. The other studies in risk management domain have generally investigated the extent to which ChatGPT can grasp concepts only in risk identification and generating a risk response strategy process, or have assessed the performance of ChatGPT for a specific construction project type. At this point, it is believed that one of the crucial implications of this study is its focus on the evaluation of ChatGPT's performance in each risk management sub-processes for different construction project types.

5.2. Implications for Construction Industry

In PMBOK methodology, there are many processes, each with its own set of inputs, outputs, and tools. Ref. [43] presents a set of tools and techniques, such as expert judgement, data gathering, prompt lists, and meetings, that can be used in risk identification sub-processes. To address the unique risks associated with each individual project, ChatGPT can be utilized as a data-gathering tool for determining a set of potential risk factors.

For example, ChatGPT's potential set of risk factors can be used as a prompt list to form a basis when developing a project charter and risk register as well. Upon completion of the risk identification sub-process in risk management implementations, the content of the risk register includes a list of identified risks. A structured risk statement may be used to distinguish risks from their cause(s) and their effect(s). For this reason, ChatGPT might provide a discussion ground for the projects' stakeholders as a starting point. On the other hand, the expert evaluations show that ChatGPT has a high degree of ability to provide proper risk allocation decisions and to generate contingent response strategies. In risk management implementations, it is appropriate for the project team to make a response plan (often called as contingency plans or fallback plans), which will only be executed under certain predefined conditions, if it is believed that there will be sufficient warning to implement the plan [43]. It is advantageous for a project manager to have a contingency plan readily available for implementation rather than having to develop one while the risk is already causing negative impacts [54]. In line with the findings of [14]'s work that highlighting ChatGPT's strongest aspect as its ability to provide relevant risk mitigation strategies, ChatGPT can be used as an efficient tool for developing a good "Plan B" by practitioners. Additionally, from a broader perspective, AI-driven risk management contributes to sustainability by improving the speed and accuracy of risk management processes. ChatGPT can aid risk assessment, monitoring, and response, enabling organiza-

tions to quickly address emerging risks and prevent environmental and social damages. It enhances risk prediction and forecasting, facilitating better-informed decisions and effective resource allocation for sustainable outcomes. Moreover, AI-driven risk management, like ChatGPT use, can promote socio-economic sustainability by identifying and managing risks effectively, optimizing resource allocation, making data-driven decisions, adapting to market dynamics, and fostering inclusive socio-economic development. In brief, ChatGPT use for risk management in construction projects additionally aids sustainability by proactively identifying risks, facilitating communication and collaboration among stakeholders, and providing recommendations for eco-friendly construction methods and materials.

5.3. Limitations and Directions for Future Research

While this study offers insights into the potential benefits of integrating ChatGPT into risk management domain, it is essential to recognize its limitations. The limitations are two-fold. First, this study performed its analysis based on the data gathered from only one platform. Other online platforms should be used with the same set of questions to determine if the responses are consistent and if there is variation among the answers. The second limitation is the absence of training data, and hence, the main aim of this study was to assess the extent to which ChatGPT provides accurate outputs in general. The responses of ChatGPT are formulated using a combination of licensed data, data generated by human trainers, and publicly available data. It has not undergone direct training on industry reports, best practices, or regulatory frameworks, nor does it have access to proprietary databases or classified information. Thus, the rankings provided by ChatGPT can be meant to serve as a general indication and starting point for risk management implementations. However, to generate more accurate and precise outcomes, it is crucial to gather project-specific information and involve the expertise of relevance stakeholders.

As time passes, it is possible that ChatGPT's pre-trained models and underlying algorithms may be more suited to handling risk factors in construction projects, leading to better suggestions in that area. The provided results may improve as ChatGPT becomes more trained and more historical data and best practices are obtained. Therefore, inputs from real-life case studies are much needed to provide a more comprehensive understanding for the consistency and accuracy of ChatGPT. For this training, risk registers and lessons learned registers can be used as training data inputs. In line with this limitation, future studies should explore ChatGPT's capability in real-world project management implications by using specific case studies and historical data. In addition, researchers can perform studies for detecting if any differentiation occurs for the conversations taking place in different times by using the same prompt template considering the possible enhancement in ChatGPT's performance.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/su152216071/s1>, Table S1: KPIs for assessing accuracy of ChatGPT use in risk management process, Table S2: Prompt templates, Table S3: Results of FGS-1 (for PPP-type transportation project), Table S4: Results of FGS-2 (for PPP-type energy project).

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available upon request from the author.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Al-Bahar, J.; Crandall, K.C. Systematic Risk Management Approach for Construction Projects. *J. Constr. Eng. Manag.* **1990**, *116*, 533–546. [[CrossRef](#)]
2. Akintoye, A.; MacLeod, M.J. Risk Analysis and Management in Construction. *Int. Journey Proj. Manag.* **1997**, *15*, 31–38. [[CrossRef](#)]

3. Nobanee, H.; Al Hamadi, F.Y.; Abdulaziz, F.A.; Abukarsh, L.S.; Alqahtani, A.F.; AlSubaey, S.K.; Almansoori, H.A. A bibliometric analysis of sustainability and risk management. *Sustainability* **2021**, *13*, 3277. [CrossRef]
4. Wang, S.Q.; Dulaimi, M.F.; Aguria, M.Y. Risk Management Framework for Construction Projects in Developing Countries. *Constr. Manag. Econ.* **2004**, *22*, 237–252. [CrossRef]
5. Cardona, O.D. The need for rethinking the concepts of vulnerability and risk from a holistic perspective: A necessary review and criticism for effective risk management. In *Mapping Vulnerability*; Routledge: London, UK, 2013; pp. 37–51.
6. Auth, G.; Jöhnk, J.; Wiecha, D.A. A Conceptual Framework for Applying Artificial Intelligence in Project Management. In Proceedings of the IEEE 23rd Conference on Business Informatics (CBI), Bolzano, Italy, 1–3 September 2021; pp. 161–170.
7. Regona, M.; Yigitcanlar, T.; Xia, B.; Li, R.Y.M. Opportunities and adoption challenges of AI in the construction industry: A PRISMA review. *J. Open Innov. Technol. Mark. Complex.* **2022**, *8*, 45. [CrossRef]
8. Choi, S.J.; Choi, S.W.; Kim, J.H.; Lee, E. AI and Text-Mining Applications for Analyzing Contractor’s Risk in Invitation to Bid (ITB) and Contracts for Engineering Procurement and Construction (EPC) Projects. *Energies* **2021**, *14*, 4632. [CrossRef]
9. Khodabakhshian, A.; Puolitaival, T.; Kestle, L. Deterministic and Probabilistic Risk Management Approaches in Construction Projects: A Systematic Literature Review and Comparative Analysis. *Buildings* **2023**, *13*, 1312. [CrossRef]
10. Chenya, L.; Aminudin, E.; Mohd, S.; Yap, L.S. Intelligent risk management in construction projects: Systematic Literature Review. *IEEE Access* **2022**, *10*, 72936–72954. [CrossRef]
11. Afzal, F.; Yunfei, S.; Nazir, M.; Bhatti, S.M. A review of artificial intelligence-based risk assessment methods for capturing complexity-risk interdependencies: Cost overrun in construction projects. *Int. J. Manag. Proj. Bus.* **2021**, *14*, 300–328. [CrossRef]
12. Pan, Y.; Zhang, L. Roles of artificial intelligence in construction engineering and management: A critical review and future trends. *Autom. Constr.* **2021**, *122*, 103517. [CrossRef]
13. Basaif, A.A.; Alashwal, A.M.; Mohd-Rahim, F.A.; Karim, S.B.A.; Loo, S.C. Technology awareness of artificial intelligence (AI) application for risk analysis in construction projects. *Malays. Constr. Res. J.* **2020**, *9*, 182–195.
14. Al-Mhdawi, M.K.S.; Qazi, A.; Alzarrad, A.; Dacre, N.; Rahimian, F.; Buniya, M.K.; Zhang, H. Expert Evaluation of ChatGPT Performance for Risk Management Process Based on ISO 31000 Standard. 2023. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4504409 (accessed on 17 August 2023).
15. Weng, J.C. *Putting Intellectual Robots to Work: Implementing Generative AI Tools in Project Management*; NYU SPS Applied Analytics Laboratory: New York, NY, USA, 2023. Available online: <https://archive.nyu.edu/handle/2451/69531> (accessed on 17 August 2023).
16. Schwarz, I.J.; Sánchez, I.P.M. *Implementation of Artificial Intelligence into Risk Management Decision-Making Processes in Construction Projects*; Universität der Bundeswehr München, Institut für Baubetrieb: Neubiberg, Germany, 2015; pp. 357–378.
17. Aziz, S.; Dowling, M. Machine learning and AI for risk management. In *Disrupting Finance: FinTech and Strategy in the 21st Century*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 33–50.
18. Baryannis, G.; Validi, S.; Dani, S.; Antoniou, G. Supply chain risk management and artificial intelligence: State of the art and future research directions. *Int. J. Prod. Res.* **2019**, *57*, 2179–2202. [CrossRef]
19. Gevaert, C.M.; Carman, M.; Rosman, B.; Georgiadou, Y.; Soden, R. Fairness and accountability of AI in disaster risk management: Opportunities and challenges. *Patterns* **2021**, *2*, 100363. [CrossRef]
20. Khatib, E.; ZM, R.; Al-Nakeeb, A. The effect of AI on project and risk management in health care industry projects in the United Arab Emirates (UAE). *Int. J. Appl. Eng. Res.* **2021**, *6*, 1–9.
21. Rodríguez-Espíndola, O.; Chowdhury, S.; Dey, P.K.; Albores, P.; Emrouznejad, A. Analysis of the adoption of emergent technologies for risk management in the era of digital manufacturing. *Technol. Forecast. Soc. Chang.* **2022**, *178*, 121562. [CrossRef]
22. Abioye, S.O.; Oyedele, L.O.; Akanbi, L.; Ajayi, A.; Delgado, J.M.D.; Bilal, M.; Ahmed, A. Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. *J. Build. Eng.* **2021**, *44*, 103299. [CrossRef]
23. Prieto, S.A.; Mengiste, E.T.; García de Soto, B. Investigating the use of ChatGPT for the scheduling of construction projects. *Buildings* **2023**, *13*, 857. [CrossRef]
24. Franceschini, F.; Maisano, D.; Mastrogiacomo, L. Empirical analysis and classification of database errors in Scopus and Web of Science. *J. Informetr.* **2016**, *10*, 933–953. [CrossRef]
25. Akinosho, T.D.; Oyedele, L.O.; Bilal, M.; Ajayi, A.O.; Delgado, M.D.; Akinade, O.O.; Ahmed, A.A. Deep learning in the construction industry: A review of present status and future innovations. *J. Build. Eng.* **2020**, *32*, 101827. [CrossRef]
26. Locatelli, M.; Seghezzi, E.; Pellegrini, L.; Tagliabue, L.C.; Di Giuda, G.M. Exploring natural language processing in construction and integration with building information modeling: A scientometric analysis. *Buildings* **2021**, *11*, 583. [CrossRef]
27. Erfani, A.; Cui, Q. Predictive risk modeling for major transportation projects using historical data. *Autom. Constr.* **2022**, *139*, 104301. [CrossRef]
28. Tirmizi, S.A.A.; Arif, F. Conceptual Approach for the Use of Artificial Intelligence for Contractual Risk Assessment in Infrastructure Projects. *Eng. Proc.* **2022**, *22*, 12.
29. Bigham, G.F.; Adamtey, S.; Onsarigo, L.; Jha, N. Artificial Intelligence for Construction Safety: Mitigation of the Risk of Fall. In Proceedings of the SAI Intelligent Systems Conference, London, UK, 6–7 September 2018.
30. Poh, C.Q.; Ubeynarayana, C.U.; Goh, Y.M. Safety leading indicators for construction sites: A machine learning approach. *Autom. Constr.* **2018**, *93*, 375–386. [CrossRef]

31. Koc, K.; Ekmekcioğlu, Ö.; Gurgun, A.P. Integrating feature engineering, genetic algorithm and tree-based machine learning methods to predict the post-accident disability status of construction workers. *Autom. Constr.* **2021**, *131*, 103896. [[CrossRef](#)]
32. Yaseen, Z.M.; Ali, Z.H.; Salih, S.Q.; Al-Ansari, N. Prediction of risk delay in construction projects using a hybrid artificial intelligence model. *Sustainability* **2020**, *12*, 1514. [[CrossRef](#)]
33. Phasha, C. The Impact of Artificial Intelligence on Cost Overruns and Risk Management in Construction Project Management. Ph.D. Thesis, University of Johannesburg, Johannesburg, South Africa, 2022.
34. Cheng, M.Y.; Peng, H.S.; Wu, Y.W.; Chen, T.L. Estimate at completion for construction projects using evolutionary support vector machine inference model. *Autom. Constr.* **2010**, *19*, 619–629. [[CrossRef](#)]
35. Awad, A.; Fayek, A.R. A decision support system for contractor prequalification for surety bonding. *Autom. Constr.* **2012**, *21*, 89–98. [[CrossRef](#)]
36. Forbes, D.R.; Smith, S.D.; Horner, R.M.W. The selection of risk management techniques using case-based reasoning. *Civ. Eng. Environ. Syst.* **2010**, *27*, 107–121. [[CrossRef](#)]
37. Xue, Z.; Xu, C.; Xu, X. Application of ChatGPT in natural disaster prevention and reduction. *Nat. Hazards Res.* **2023**, *3*, 556–562. [[CrossRef](#)]
38. Uddin, S.J.; Albert, A.; Ovid, A.; Alsharef, A. Leveraging ChatGPT to Aid Construction Hazard Recognition and Support Safety Education and Training. *Sustainability* **2023**, *15*, 7121. [[CrossRef](#)]
39. Hofert, M. Assessing ChatGPT's Proficiency in Quantitative Risk Management. 2023. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4444104 (accessed on 17 August 2023).
40. Klepo, M.S.; Knežević, D.; Knežević, T.; Meštrović, H. Artificial Intelligence in Risk Management System on Infrastructure Projects. In *Creative Construction e-Conference 2023*; Budapest University of Technology and Economics: Budapest, Hungary, 2023; pp. 208–214.
41. ISO 31000-2018; Risk Management—Guidelines. International Organization for Standardization: Geneva, Switzerland, 2018.
42. PMBOK. *A Guide to the Project Management Body of Knowledge (PMBOK Guide)—Seventh Edition and the Standard for Project Management*; Project Management Institute, Inc.: Newtown Square, PA, USA, 2021.
43. PMBOK. *A Guide to the Project Management Body of Knowledge (PMBOK Guide)—Sixth Edition and the Standard for Project Management*; Project Management Institute, Inc.: Newtown Square, PA, USA, 2017.
44. Hillson, D. Use a risk breakdown structure (RBS) to understand your risks. In Proceedings of the Project Management Institute Annual Seminars & Symposium, San Antonio, TX, USA, 7–8 October 2002.
45. Abednego, M.P.; Ogunlana, S.O. Good project governance for proper risk allocation in public–private partnerships in Indonesia. *Int. J. Proj. Manag.* **2006**, *24*, 622–634. [[CrossRef](#)]
46. Jin, X.H.; Zhang, G. Modelling optimal risk allocation in PPP projects using artificial neural networks. *Int. J. Proj. Manag.* **2011**, *29*, 591–603. [[CrossRef](#)]
47. Ameyaw, E.E.; Chan, A.P. Risk allocation in public-private partnership water supply projects in Ghana. *Constr. Manag. Econ.* **2015**, *33*, 187–208. [[CrossRef](#)]
48. Yang, W.; Dai, D. Concession Decision Model of BOT Projects Based on a Real Options Approach. In Proceedings of the International Conference on Management Science and Engineering, Lille, France, 6–7 October 2006; pp. 307–312.
49. Chan, I.Y.S.; Leung, M.Y.; Yu, S.S.W. Managing the Stress of Hong Kong Expatriate Construction Professionals in Mainland China: Focus Group Study Exploring Individual Coping Strategies and Organizational Support. *J. Constr. Eng. Manag.* **2012**, *138*, 1150–1160. [[CrossRef](#)]
50. Wibeck, V.; Dahlgren, M.A.; Öberg, G. Learning in Focus Groups An Analytical Dimension For Enhancing Focus Group Research. *Qual. Res.* **2007**, *7*, 249–267. [[CrossRef](#)]
51. Saaty, T.L. Decision Making—The Analytic Hierarchy and Network Processes (AHP/ANP). *J. Syst. Sci. Syst. Eng.* **2004**, *13*, 1–35. [[CrossRef](#)]
52. Ishizaka, A.; Labib, A. Review of the Main Developments in the Analytic Hierarchy Process. *Expert Syst. Appl.* **2011**, *38*, 14336–14345. [[CrossRef](#)]
53. COSO ERM-2017; COSO Enterprise Risk Management-Integrating with Strategy and Performance. Committee of Sponsoring Organizations: Englewood Cliffs, NJ, USA, 2017.
54. Heimann, J.F. Contingency planning as a necessity. In Proceedings of the Project Management Institute Annual Seminars & Symposium, Houston, TX, USA, 7–16 September 2000.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.