Intelligent Agent for Construction Safety Reporting Based on Large Language Models

Le Yang[1], Botao Gu[1], Hongling Guo[1], Yuecheng Huang[1] and Dongping Fang[2]

[1]Department of Construction Management, Tsinghua University, Beijing, China
[2]School of Civil Engineering, Tsinghua University, Beijing, China
Corresponding author's E-mail: yangle23@mails.tsinghua.edu.cn

Abstract

In the construction industry, construction safety management is primarily governed by a safety management team that involves safety officers, supervisors, project manager, and owner's representatives. Safety reporting plays a vital role in safety management, serving not only to document on-site activities but also to facilitate effective communication across various management levels. With digital-technology advancements, intelligent safety reporting systems based on electronic forms have emerged. However, these systems encounter challenges such as high learning costs, delays in report submission and unstable report quality. This study presents an intelligent agent designed for safety reporting utilizing large language model (LLM) technology. The agent is deployed on cloud server, and safety personnel can connect with it through an instant messaging application to obtain real-time information. The interaction between the agent and safety personnel is similar to daily conversations, requiring minimal learning costs. Additionally, the agent seamlessly generates safety reports, offering a concise summary and subsequent work plans, thereby improving the overall quality of reports. Through its deployment on a construction site, the agent achieved an F1 score of 0.996 for hazard identification and an average score of 93.4 for generated reports. It has been recognized by construction safety personnel, proving its ability to improve the efficiency and quality of safety personnel's work and demonstrating its potential to enhance safety management capabilities.

Keywords: Construction Safety, Safety Reporting, Large Language Model, AI Agents

1. INTRODUCTION

The construction industry plays a vital role in the national economy, significantly contributes to GDP, and serves as a key driver of economic growth in numerous countries. However, it remains the most dangerous industry globally, resulting in a significant number of casualties each year. The data from the Occupational Safety and Health Administration (OSHA) highlights a concerning trend in the U.S. construction sector, with the number of fatalities remaining high since 2018 (Bureau of Labor Statistics, 2022). By 2022, the reported death reached 1,056. Similarly, in China, the construction industry has been burdened by an alarming death toll exceeding 3,000 annually since the year 2016.
The construction safety management system, undertaken by safety officers, supervisors, project managers, and owner's representatives, plays a crucial role in ensuring on-site safety and minimizing accidents and injuries. An essential component of the system is safety reporting, which documents on-site hazards and facilitates targeted communication across various management levels, enabling the team to plan subsequent work effectively. However, it is challenging for personnel to record high-quality safety reports on-site, as it requires safety personnel to spend a significant amount of time writing them after returning to the office. Additionally, the quality of the reports is influenced by the cultural proficiency of the safety personnel. Despite the development of digitization, there is still a lack of suitable tools to improve the efficiency of safety reporting. In an attempt to minimize reporting costs, some electronic form systems have been developed, providing predefined options to describe potential hazards and situations. However, the existing forms and the ever-changing nature of daily safety hazards, coupled with the complexity of on-site environments, give rise to substantial discrepancies that impose significant adjustment costs on users. Overall, the lack of digitization in

safety reporting not only burdens on-site personnel but also hampers projects' ability to timely and comprehensively gather hazard data, impeding the improvement of safety management.
The recent focus on the ChatGPT product signifies the advent of the artificial general intelligence (AGI) era. Artificial intelligence agents are now recognized as a pivotal step towards achieving AGI, as they showcase the potential for a wide array of intelligent activities. An agent refers to an artificial entity that perceives the environment, makes decisions, and takes actions. Agents possess varying degrees of autonomy, reactivity, proactiveness, and social abilities, enabling them to analyze and respond to external inputs while providing appropriate feedback. They can interact with humans or operate autonomously to accomplish tasks, thereby enhancing work efficiency. The development of agents relies on their inherent general capabilities, and the recently emerged large language models (LLMs) demonstrate remarkable proficiency in knowledge acquisition, instruction understanding, generalization, planning, and reasoning. Additionally, LLMs exhibit effective natural language interaction with humans, rendering them well-suited as foundational models for agents. Numerous studies have emerged utilizing LLMs to construct agents, effectively enhancing task efficiency and alleviating user workload.
Therefore, this study presents the development of an intelligent agent for construction safety reporting leveraging large language model technology, along with a novel approach to reporting. This study begins with a review safety reporting and provide an overview of the agent and LLM fundamentals. Then, it proposes a three-part framework design for the agent, encompassing perception, brain, and action modules. This framework effectively demonstrates the agent's capabilities in hazard identification, report generation, and interactive communication with humans through instant messaging (IM). To optimize the utilization of LLM's general capabilities for these tasks, task-specific prompts are specifically designed. Moreover, field tests were conducted in practical projects, receiving recognition from personnel, and achieving an F1 score of 0.996 for hazard identification and an average score of 93.4 for report generation. Furthermore, this study thoroughly discusses the limitations and challenges associated with this agent.
Our primary contributions are three-fold: (1) An LLM-based agent framework for construction safety scenarios is proposed, which is applicable to various safety tasks, thereby reducing user learning costs and improving efficiency; (2) An LLM prompt process for hazard identification and report generation is established, achieving high hazard identification rates and report quality; (3) The intelligent agent is deployed and tested on-site, effectively collecting and addressing observed hazards, demonstrating the system's ability to enhance work efficiency and safety management capabilities.

2. LITERATURE REVIEW

Safety reporting. Safety reporting has a longstanding history as an integral component of safety management practices (Hallowell et al. 2013, Jazayeri et al. 2017). It entails the documentation and monitoring of on-site behaviors. Safety reporting is not an isolated safety management approach but rather integrated into broader processes such as Behavior-Based Safety (BBS), Near Miss Reporting (NMR), or Safety Management Systems (SMS) (Oswald et al. 2018). In the NMR process, safety reporting encourages employees to report incidents that nearly resulted in accidents, injuries, or property damage (Jones et al. 1999). The aim is to identify and comprehend the underlying causes and contributing factors of these incidents, with the goal of preventing future accidents and enhancing overall safety. Similarly, within the Safety Observation Reporting (SOR) process, safety reporting encourages all on-site individuals to report unsafe acts or conditions, which can then be addressed by the health and safety team through appropriate corrective actions.
However, the challenges faced by safety reporting are often disregarded. On the side of on-site personnel, the intricate nature of the work environment often hinders providing timely and comprehensive descriptions of environmental factors. Additionally, limited education or lack of experience of personnel makes it challenging for them to accurately describe hazards that are less observable or occur infrequently. From the perspective of project management personnel, untimely and incomplete hazard reports can result in distorted representations that fail to accurately reflect the on-site conditions. Overall, these issues significantly contribute to the production of low-quality safety reports within certain safety management practices, thereby impeding effective management and

hindering improvements in project safety standards.

Large language model. A large language model is a deep neural network model characterized by a large parameter space. The development of the Transformer architecture, coupled with the growth of computational resources and advancements in large datasets, has facilitated the training of LLMs with billions of parameters (Thirunavukarasu et al. 2023). As a consequence, these advancements have endowed LLMs with capabilities that smaller models lack, including language comprehension, generation, and logical reasoning. This architecture significantly enhances their ability to summarize, translate, predict, and generate text that exhibits human-like characteristics.

ChatGPT is a conversation model launched by OpenAI in October 2022. It has demonstrated remarkable proficiency in interacting with humans, showcasing its expansive knowledge base, reasoning abilities, and contextual tracking across multiple conversational turns. As a result, ChatGPT has emerged as the most powerful chatbot to date. Notably, other companies and organizations have introduced their own versions of LLMs, such as Meta's LLaMA (Touvron et al. 2023) and ZhipuAI's ChatGLM (Du et al. 2021), exhibiting distinct variations in terms of scale and performance.

The integration of the construction industry with LLM has been a subject of recent research. Uddin et al. (2023) guided students in utilizing ChatGPT to aid in hazard identification on construction sites, producing results that demonstrated a significant improvement in hazard recognition levels. Prieto et al. (2023) employed ChatGPT to generate a construction schedule for a construction project, employing a logical approach to meet the specified scope requirements. You et al. (2023) introduced RoboGPT, which utilizes ChatGPT for automated sequence planning in robot-based assembly tasks within the construction domain. Zheng et al. (2023) presented BIM-GPT, a prompt-based virtual assistant (VA) framework that enables users to engage in natural language queries, summarization, and retrieval of BIM-related information.

Agent. In the field of artificial intelligence, the term "agent" refers to an artificial entity that is capable of perceiving the surrounding environment using sensors, making decisions based on the perceived information, and taking responsive actions using actuators (Wooldridge et al. 1995, Russell et al. 2010). Despite substantial efforts made since the mid-20th century to develop AI agents, their adaptability across diverse scenarios has been limited, and notable progress in terms of their inherent general capabilities has remained elusive. However, in recent times, the advent of LLMs and their demonstrated proficiency in knowledge acquisition, comprehension of instructions, reasoning, and planning has inspired researchers to explore the utilization of LLMs as foundational components for constructing agents.

In the literature review conducted by Xi et al. (2023), a conceptual framework is proposed, focusing on agents based on LLMs. This framework comprises three distinct modules: the brain, perception, and action. The brain module serves as the agent's core, responsible for storing external information, knowledge, and internal memory. Additionally, it performs crucial tasks such as natural language processing, decision-making, reasoning, and planning. The perception module functions as the framework's sensor, tasked with receiving information from the external environment. This includes conventional text-based inputs as well as broader multimodal inputs like visual, auditory, and tactile cues. Lastly, the action module serves as the agent's actuator, enabling it to generate textual outputs, take concrete actions, and utilize tools to provide feedback to the external environment.

3. DEVELOPMENT OF INTELLIGENT AGENT FOR SAFETY REPORTING

This study presents the development of an intelligent agent for safety reporting as a critical element of construction project safety management. The architecture of the agent is depicted in Figure 1. Referring to the framework of agent proposed by Xi et al. (2023), the agent is composed of three modules: perception, brain, and action. The perception module is responsible for receiving reports from users, allowing personnel to interact with the agent via an IM (instant messaging) system. Acting as a sensor, the IM can receive textual inputs, capture images, and process real-time voice inputs. Once the external information is received, it is transmitted to the brain module for further processing. The brain module first determines the type of input and proceeds with the corresponding processing, such as recording hazards observed during inspections or generating today’s report. Subsequent processing is triggered by the brain module, which leverages the inherent general capabilities of LLMs to address these issues or invokes external tools. For instance, in the case of hazards, the agent generates a corresponding webpage for the user. Regarding report generation, the agent relies on the planning and reasoning capabilities of the LLM to summarize the safety hazards observed during the day and provide future plans. The action module is capable of providing feedback to users in the form of text or a website, thereby influencing the safety management practices of the construction project.

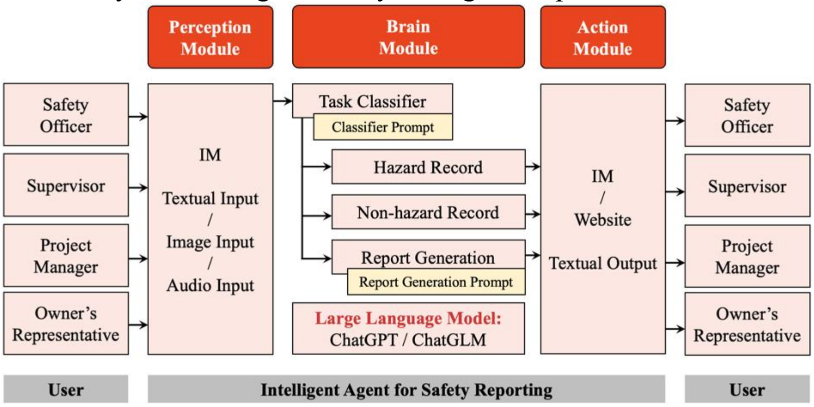

Figure 1. Technical framework of intelligent agent.

This study also presents a novel approach that utilizes an agent to facilitate safety reporting for personnel, illustrated in Figure 2. During on-site operations, when encountering a hazard, safety personnel can engage in a dialogue through IM, utilizing text, images, and voice inputs (a). Once the hazard is identified, the agent generates a structured hazard report and presents it on a web page. Personnel could edit and resolve the hazard according to the actual situation (b). These reports are stored and made readily accessible for subsequent analysis. At the end of each day, personnel can conveniently review the safety reports generated throughout the day via IM (c).

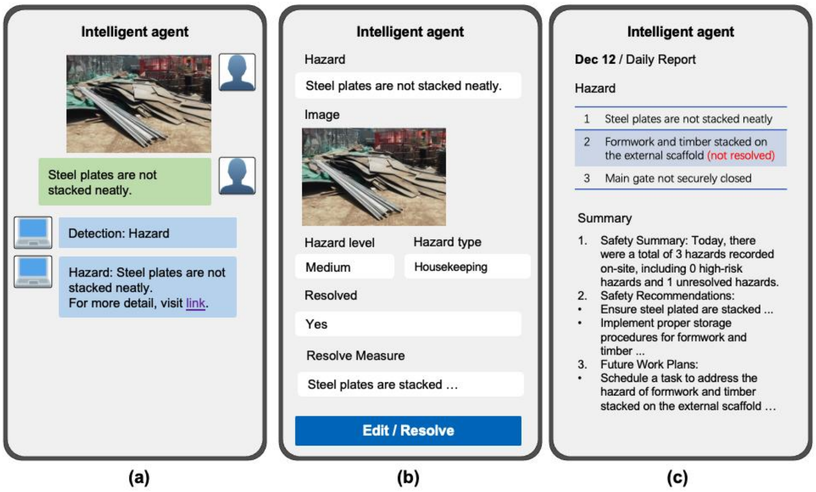

Figure 2. Safety reporting approach based on agent. (a) Personnel record hazard images and description via the agent. (b) Personnel edit and resolve the hazard. (c) The daily report page lists and summarizes the recorded hazards.

Hazard Identification. The primary task for the agent is to determine whether the user's input relates to safety hazards. To accomplish this, this study employed prompt engineering and devised a dedicated classification prompt that enables the LLM to assess the severity of the input in terms of hazards. For instance, the LLM would assign a high score to a hazard description such as "steel plates are not stacked neatly," while assigning a low score to an activity like "we conducted routine safety training today." A threshold for hazard identification was established through tests conducted, based on hazards observed on-site. Inputs with a score equal to or higher than the threshold are classified as hazards. The prompt design for hazard identification is depicted in Figure 3.

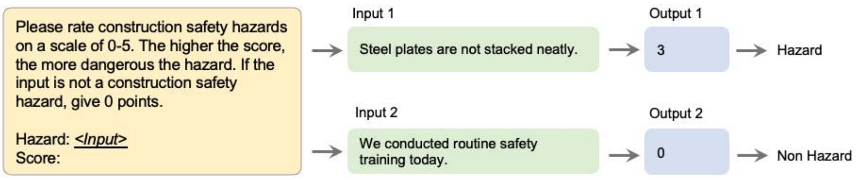

Figure 3. Prompt design of hazard identification

Report Generation. At the conclusion of each day's work, the agent compiles a collection of daily hazards and generates reports in formats such as PDF/Word. These reports encompass all identified hazards, an overall safety assessment for the day, safety suggestions for the ongoing work, and future work arrangements. Safety officers utilize these reports to plan their future activities, including addressing any unresolved hazards from previous days. The report generation process involves extracting all hazards and combining them with dedicated prompts, which are then utilized by the LLM to generate comprehensive reports. Leveraging its domain-specific knowledge and reasoning capabilities, the LLM excels in producing coherent and articulate reports. The prompt design for report generation is illustrated in Figure 4.

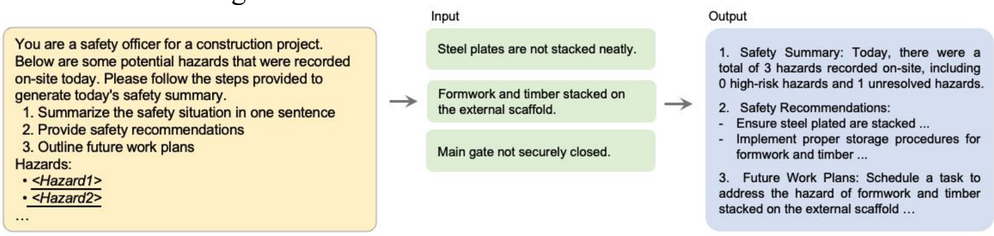

Figure 4. Prompt design of report generation.

Underlining LLM. The above implementations are constructed based on the underlying LLM and its associated technologies. The LLM serves as the core framework of the agent, offering crucial capabilities in dialogue and logical analysis. For testing purposes, this study utilized OpenAI's ChatGPT and ZhipuAI's ChatGLM as the foundational models. Simultaneously, this study introduces corresponding prompts for the two above functions. The design of prompts, known as prompt engineering, enhances the controllability and standardization of the model's responses.

4. SITE TRIAL & DISCUSSION

Site trial. The agent was deployed in a complex steel structure project undertaken by Company H in Xi'an, China, from July to September, in order to evaluate its practicality. The test participants consisted of safety officers and supervisors who were encouraged to report any hazards observed during daily inspections and generate safety reports at the conclusion of each workday. Subsequently, interviews were conducted with users. They reported that the agent effectively records hazard and facilitates corrective actions. They also provided positive feedback on the generated safety reports, indicating a favorable reception of the agent. Following the on-site testing phase, this study carried out a comprehensive data analysis utilizing the collected hazard records and safety reports.

To facilitate hazard identification, a total of 117 hazard records were extracted from field observations, classified according to company’s specifications, as listed in Table 1. Additionally, an extra 23 instances of non-hazardous information, consisting of safety training and inspections, were incorporated to assess the agent's classification performance concerning these inputs. This study evaluated the accuracy of ChatGPT (GPT-4) and ChatGLM (chatglm_turbo), achieving F1 scores of 0.996 and 0.987, respectively, with ChatGPT demonstrating superior accuracy. These results indicate that the current LLM exhibits substantial knowledge and reasoning capabilities, performing effectively in decision-making and judgment tasks.

Table 1. Hazards reported from project

Hazard Type	Count	Example
Working at heights	12	Workers standing on the guardrail of the aerial work platform.

Fire	25	Lack of fire extinguishers in the basement.
Electrical	27	Presence of bulging and damaged cable wires.
Equipment	11	Damaged anti-toppling device on the scissor lift.
Housekeeping	42	Accumulated water not cleared at the construction site.

For report generation, interviews were conducted with two personnel involved in the project, from which positive feedback on the acceptance of the safety reports was obtained. Moreover, five daily reports derived from the aforementioned hazards were selected and LLMs were employed as evaluators to assess the quality of the content generated. Zheng et al. (2024) indicates that using advanced LLMs to score generated results is a feasible method for quality assessment. This study utilized both ChatGPT and ChatGLM to generate reports based on the same set of inputs, and these reports were independently scored by both models on a scale of 0-100. The scoring results were subsequently averaged. The average scores for ChatGPT and ChatGLM were 93.4 and 92.2, respectively, indicating that the former produced reports with superior text quality.

Limitation and challenge. The agent still exhibits certain limitations. Its applicability to larger-scale projects remains uncertain, as the agent has solely undergone testing in small-scale steel structure projects, without conclusive evidence of its suitability for larger-scale endeavors. The agent may encounter challenges when confronted with more complex site environments, diverse user roles, larger data volumes, and dynamic personnel communication. Moreover, the system relies on human involvement, as the perception and action modules are depended upon personnel collaboration, thereby restricting its autonomy and proactiveness. Furthermore, there arises a necessity to introduce human evaluation. While LLM scoring is employed to assess the performance of generated reports, it possesses limitations in capturing crucial aspects such as content accuracy, plan effectiveness, and potential content omissions.

5. CONCLUSION

This study presents an intelligent agent specifically designed for safety reporting, accompanied by a novel approach for personnel to report hazards. Users can engage in communication with the agent in a manner similar to their interactions with colleagues, thereby lowering learning costs and enhancing communication efficiency. By leveraging the natural language processing and reasoning capabilities of LLMs, this agent reduces the time and effort dedicated by personnel to report writing, as well as enhances the quality of the generated reports. The deployment of this agent at a construction site yielded remarkable results, achieving an F1 score of 0.996 for hazard identification and an average score of 93.4 for the generated reports, thereby demonstrating its ability to enhance the work efficiency of personnel and the project management capabilities.

There is considerable potential for future improvements in this research. Functionally, the current handling of hazards is in its initial stage, offering opportunities for further exploration of information extraction from textual input. Broadening the support for managing a wider range of activities would be beneficial. Additionally, as the number of reports accumulates, extracting valuable safety management information from a substantial amount of text becomes a challenge. Furthermore, comprehensive on-site experiments are necessary to ensure the stability and quality of the generated text, while avoiding the production of harmful content and illusions. Lastly, future advancements may involve the agent taking the initiative to initiate interactions with humans, rather than solely relying on passive information reception from users.

6. ACKNOWLEDGMENTS

We would like to acknowledge the project was funded by China Postdoctoral Science Foundation (GZC20231233), and was supported by the National Science Foundation of China No.72401153 and Shuimu Tsinghua Scholar Program.

REFERENCES

Bureau of Labor Statistics. 2022. Fatal Occupational Injuries by Industry and Event or Exposure, All United States, 2022.

Du, Z., Qian, Y., Liu, X., Ding, M., Qiu, J., Yang, Z., & Tang, J. 2021. GLM: General language model pretraining with autoregressive blank infilling. arXiv preprint arXiv:2103.10360.

Jazayeri, E., & Dadi, G. B. 2017. Construction safety management systems and methods of safety performance measurement: A review. Journal of Safety Engineering, 6(2), 15-28.

Hallowell, M. R., Hinze, J. W., Baud, K. C., & Wehle, A. 2013. Proactive construction safety control: Measuring, monitoring, and responding to safety leading indicators. Journal of construction engineering and management, 139(10), 04013010.

Jones, S., Kirchsteiger, C., & Bjerke, W. 1999. The importance of near miss reporting to further improve safety performance. Journal of Loss Prevention in the process industries, 12(1), 59-67.

Oswald, D., Sherratt, F., & Smith, S. 2018. Problems with safety observation reporting: A construction industry case study. Safety science, 107, 35-45.

Prieto, S. A., Mengiste, E. T., & García De Soto, B. 2023. Investigating the Use of ChatGPT for the Scheduling of Construction Projects. Buildings, 13(4), 857.

Russell, S. J., & Norvig, P. 2010. Artificial intelligence a modern approach. London.

Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. 2023. Large language models in medicine. Nature medicine, 29(8), 1930-1940.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... & Lample, G. 2023. LLaMA: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971.

Uddin, S. M. J., Albert, A., Ovid, A., & Alsharef, A. 2023. Leveraging ChatGPT to Aid Construction Hazard Recognition and Support Safety Education and Training. Sustainability, 15(9), 7121.

Wooldridge, M., & Jennings, N. R. 1995. Intelligent agents: Theory and practice. The knowledge engineering review, 10(2), 115-152.

Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., ... & Gui, T. 2023. The rise and potential of large language model based agents: A survey. arXiv preprint arXiv:2309.07864.

You, H., Ye, Y., Zhou, T., Zhu, Q., & Du, J. 2023. Robot-Enabled Construction Assembly with Automated Sequence Planning Based on ChatGPT: RoboGPT. Buildings, 13(7), 1772.

Zheng, J., & Fischer, M. 2023. BIM-GPT: a Prompt-Based Virtual Assistant Framework for BIM Information Retrieval. arXiv preprint arXiv:2304.09333.

Zheng, L., Chiang, W. L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., ... & Stoica, I. 2024. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena[J]. Advances in Neural Information Processing Systems, 2024, 36.