SASBE 2025 aims to encourage the international exchange of innovative ideas between researchers from academia and industry. In addition to knowledge dissemination, the conference offers a valuable platform for professional networking, particularly benefiting university professors, graduate students, and postdoctoral researchers.

Research Article

Toward Safer Integration of Active Back-Support Exoskeletons in Construction through Domain-Specific Language Models

Okunola Akinwale[1], Abiola Akanmu[2], Houtan Jebelli[3]

Myers Lawson School of Construction, Virginia Tech, Blacksburg, VA, United States.
Myers Lawson School of Construction, Virginia Tech, Blacksburg, VA, United States.
Civil and Environmental Engineering, University of Illinois Urbana-Champaign, IL, United States.
Correspondence: abiola@vt.edu

Abstract

Active back-support exoskeletons have shown promise in reducing back-related musculoskeletal disorders in labor-intensive occupations. Yet, adoption in construction remains constrained by fragmented evidence on their biomechanical benefits, human factors risks, and practical usability. This study presents the development of a domain-adapted large language model fine-tuned with the Retrieval-Augmented Fine-Tuning framework to systematically synthesize and communicate evidence relevant to active back-support exoskeleton use in construction. The training dataset was developed from a review of twenty peer-reviewed studies spanning nine construction tasks and seven active back-support exoskeleton models, covering outcomes related to muscle activity, range of motion, perceived discomfort, exertion, usability, cognitive load, fall risk, and sociotechnical adoption factors. These studies identified both facilitators (e.g., productivity gains, posture correction, and device durability) and barriers (e.g., restricted mobility, thermal discomfort, and task–device incompatibility) shaping adoption in construction workflows. A total of 3,650 question–answer pairs were generated with distractors and Chain-of-Thought reasoning and used in a teacher–student distillation process with GPT-4o and GPT-4o-mini. The fine-tuned model achieved a validation accuracy of 88% and demonstrated stable generalization without overfitting, supported by low validation loss. In head-to-head evaluation against the baseline, the fine-tuned model achieved reliable scores in coherence, relevance, and harmlessness, a 10% improvement in response completeness (96% vs. 86%), and a 2% increase in factual accuracy (82% vs. 80%). The results demonstrate the feasibility of deploying fine-tuned large language models as interactive decision-support tools for exoskeleton adoption in construction, advancing the intersection of artificial intelligence, biomechanics, and occupational safety.

Keywords: Language Models; Construction Safety; Exoskeleton Adoption; Retrieval-Augmented Fine-Tuning; Human Factors.

Highlights

- Active back-support exoskeletons show both biomechanical benefits and adoption barriers.
- RAFT fine-tuned model improved accuracy, completeness, and safe decision support.
- Adoption requires balancing safety, usability, productivity, and long-term cost.

1 Introduction

Active back-support exoskeletons are increasingly recognized as promising interventions for reducing back-related musculoskeletal disorders in the construction industry. Compared with passive devices, active back-support exoskeletons offer more mechanical assistance (Reimeir, Calisti, Mittermeier, Ralfs, & Weidner, 2023), making them suitable for the physically demanding and high-risk tasks that contribute to elevated rates of musculoskeletal injuries in construction (Akinwale Okunola, Akanmu, & Yusuf, 2023). Despite this potential, most commercially available active back-support exoskeletons are designed for general industrial settings, particularly manual material handling, and often misalign with the dynamic and irregular demands of construction work.

Empirical evaluations of active back-support exoskeletons in construction tasks report mixed findings. Documented benefits include reduced muscle activity (Ojha, Guo, Jebelli, Martin, & Akanmu, 2024), improved range of motion (Schwartz, Desbrosses, Theurel, & Mornieux, 2023), and decreased discomfort (Akinwale Okunola, Akanmu, & Yusuf, 2023). However, unintended consequences such as restricted mobility (A. Okunola, Afolabi, Akanmu, Jebelli, & Simikins, 2024), localized pressure points (Ojha et al., 2024), uneven load distribution (Ali, Fontanari, Schmoelz, & Agrawal, 2021), increased fall risk (Akinwale Okunola, Akanmu, & Jebelli, 2024), and elevated cognitive load (Akanmu, Okunola, Jebelli, Ammar, & Afolabi, 2024) have also been observed. In parallel, studies examining adoption factors reveal divergent stakeholder priorities. While managers and corporate leaders emphasize productivity and cost-effectiveness, frontline workers and safety officers prioritize usability, comfort, and risk reduction (Crea et al., 2021). This diversity of perspectives complicates exoskeleton selection and hampers effective integration into construction workflows.

Although active back-support exoskeletons are intended to reduce musculoskeletal risks and improve productivity, inappropriate device selection may negate these benefits and amplify risks. For instance, Akanmu et al. (2024) reported increased cognitive load from bulky devices during carpentry tasks requiring frequent posture adjustments. Similarly, Zhu, Weston, Mehta, and Marras (2021) found that cognitive demands can undermine biomechanical gains and accelerate user fatigue. High procurement and maintenance costs further exacerbate the risks of poor device fit, creating financial barriers to adoption. These challenges highlight the need for decision-support models that can integrate biomechanical evidence with stakeholder priorities to generate actionable guidance for exoskeleton deployment.

Large language models (LLMs) such as GPT, BERT, and LLaMA are pretrained to generate human-like text across diverse domains (Hadi et al., 2023). When fine-tuned with domain-specific corpora, these models often outperform their general-purpose counterparts by contextualizing technical evidence for specialized applications (Susnjak et al., 2025). Recent advances in construction demonstrate this potential: Pu, Yang, Li, and Guo (2024) fine-tuned GPT-4o-mini for inspection reporting, while Koppel, Tšernikova, and Kalvet (2024) adapted GPT-4 as a virtual ergonomic risk assessment assistant. Traditional approaches such as supervised fine-tuning and retrieval-augmented generation each carry limitations—overfitting and hallucinations in supervised fine-tuning, and irrelevant retrievals in retrieval-augmented generation (Chung, Vo, Kizilkale, & Reite, 2024). The Retrieval-Augmented Fine-Tuning (RAFT) framework integrates both methods, reducing hallucinations and improving contextual grounding (Chung et al., 2024; Zhang et al., 2024). Despite the progress in artificial intelligence for construction safety, scarce studies have applied fine-tuned LLMs to the domain of exoskeleton adoption, where biomechanics, human factors, and stakeholder dynamics intersect.

This study develops and evaluates a domain-specific LLM fine-tuned with the RAFT framework to support active back-support exoskeleton adoption in construction. The model synthesizes biomechanical and human factors evidence while incorporating stakeholder perspectives, offering contextually relevant recommendations for device-specific deployment. The contribution lies in bridging artificial intelligence capabilities with construction ergonomics, advancing decision-support infrastructure that can inform safer and more effective exoskeleton integration in industry practice.

2 Background

This section reviews the current state of knowledge on back-support exoskeletons in construction, emphasizing human factors evidence, stakeholder perspectives, and the potential for artificial intelligence-driven decision support. It also identifies research gaps that motivate the development of a domain-specific large language model.

2.1 Human Factors Evidence on Benefits and Risks of Exoskeletons in Construction

Human factors risk assessments of active back-support exoskeleton across various construction tasks have demonstrated both biomechanical benefits (Lei et al., 2024; Akinwale Okunola, Akanmu, Jebelli, & Afolabi, 2025; Sposito, Fanti, Poliero, Caldwell, & Di Natali, 2024) and potential risks (Akanmu et al., 2024; Akinwale Okunola et al., 2024; Sposito et al., 2024). For instance, Akinwale Okunola, Akanmu, Jebelli, et al. (2025) evaluated an active back-support exoskeleton during framing tasks and reported reductions in muscle activity, range of motion, physical exertion, and discomfort. Similarly, Sposito et al. (2024) found that active back-support exoskeleton use during railway construction reduced perceived fatigue, while Lei et al. (2025) observed decreased muscle activity and workload during bending and lifting tasks. However, risks have also been identified. Sposito et al. (2024) reported user discomfort in the thighs and thermal discomfort during railway construction tasks. Akanmu et al. (2024) found increased cognitive load associated with active back-support exoskeleton use during carpentry framing, and Akinwale Okunola et al. (2024) highlighted elevated fall risk during framing tasks. These findings highlight the complex trade-offs between benefits and risks, emphasizing the need for nuanced understanding in the selection and adoption of active back-support exoskeleton among construction stakeholders.

2.2 Stakeholder Perspectives, Adoption Barriers, and the Potential of Artificial Intelligence for Decision Support

In addition to the task-specific risks and benefits of active back-support exoskeleton, construction stakeholders hold varying priorities regarding the facilitators and barriers to adoption, which influence appropriate exoskeleton selection. A. Okunola et al. (2024) examined these factors from the stakeholders' perspective, identifying facilitators such as productivity gains, cost-effectiveness, injury reduction, durability, low cognitive load, posture correction, and ease of maintenance. Reported barriers included incompatibility with tasks, poor performance on uneven surfaces, excessive weight, perceived fall risk, restricted mobility, thermal discomfort, and snagging hazards. Similarly, Gutierrez et al. (2024) identified facilitators, including reduced fatigue, improved task performance, increased productivity, standardization of expectations, workforce retention, and positive social perception. Barriers noted were concerns about hygiene and reuse, heat discomfort, cost sensitivity, fall risk, mobility limitations, poor fit for female workers, pressure points, durability issues, and bulkiness (Dunson-Todd, Nik-Bakht, & Hammad, 2025).

While understanding the human factors risks and benefits of active back-support exoskeleton is important, stakeholder perceptions of adoption facilitators and barriers also play a critical role (Crea et al., 2021). The ability to synthesize these factors is essential for guiding effective active back-support exoskeleton selection and supporting broader adoption in the construction industry. LLMs, which are advanced artificial intelligence pre-trained on an extensive amount of data to synthesize and generate human-like language (Hadi et al., 2023), presents an opportunity to bridge this synthesis gap. By translating complex, multi-faceted scientific insights into accessible, actionable narratives, LLMs can help stakeholders better understand the nuanced trade-offs involved in active back-support exoskeleton adoption. Beyond general application, these models are known for their ability to be fine-tuned for a specific domain purpose, providing in-depth knowledge in a particular area, which often outperforms their general-purpose counterparts (Hadi et al., 2023). For example, Pu et al. (2024) fine-tuned a pretrained GPT-4o-mini for driving AutoRepo, a system that uses construction information to generate periodic inspection reports, which could help stakeholders in understanding the status of construction activity on site. Koppel et al. (2024) fine-tuned GPT-4 to function as a virtual risk analysis assistant, specifically designed to support ergonomic risk assessments within the construction industry. These examples demonstrate the feasibility of domain-specific fine-tuning for construction safety applications. However, unlike prior efforts focused on general safety reporting or risk assessment, the domain of exoskeleton adoption presents a unique challenge that spans biomechanics, cognitive ergonomics, sociotechnical dynamics, and economic considerations. A domain-specific model for active back-support exoskeleton adoption must therefore be trained on datasets that capture not only scientific literature but also stakeholder perceptions and task-specific constraints.

2.3 Research Gap

Although these studies highlight the feasibility of adapting LLMs for targeted applications, the effectiveness of such models is largely determined by the relevance of the selected datasets. Unlike other subsets of artificial intelligence, such as machine learning, LLMs require datasets primarily in text form rather than continuous sensor data. Synthesizing peer-reviewed documents in a domain area has been used for fine-tuning LLMs to become experts in that domain (Susnjak et al., 2025). For example, Susnjak et al. (2025) fine-tuned GPT-3.5 using a curated dataset derived from seventeen peer-reviewed biomedical studies, enabling the model to more effectively synthesize and interpret medical literature. Given that LLMs can be adapted for specific domains, approaches such as Supervised Fine-Tuning, which involves training on labelled data, and RAG, which enables inference from external documents, each have limitations (Chung et al., 2024; Zhang et al., 2024). Supervised Fine-Tuning may lead to hallucinations due to overfitting, while Retrieval-Augmented Generation can introduce inaccuracies through the retrieval of irrelevant or misleading documents. To address these issues, Zhang et al. (2024) proposed RAFT, which combines the strengths of both methods. RAFT improves accuracy and reduces hallucinations by training the model to selectively use relevant retrieved content. The authors describe Supervised Fine-Tuning as similar to preparing for a closed-book exam, Retrieval-Augmented Generation as an open-book exam without preparation, and RAFT as a combination of both—studying for the exam while also having access to the book, ensuring more reliable and contextually grounded outputs. This study builds on these advancements by employing the RAFT methodology to curate, synthesize, and fine-tune a domain-specific LLM that captures the breadth of biomechanical, cognitive, and stakeholder factors influencing active back-support exoskeleton adoption in construction. This approach aims to provide stakeholders with an Artificial Intelligence-driven tool capable of generating contextually relevant, evidence-based recommendations adapted to the multifaceted realities of construction environments.

3 Methodology

Figure 1 presents the overall workflow, which is structured into four sequential stages: (1) literature review, (2) data preparation and pre-processing, (3) fine-tuning of a domain-specific large language model, and (4) model performance evaluation.

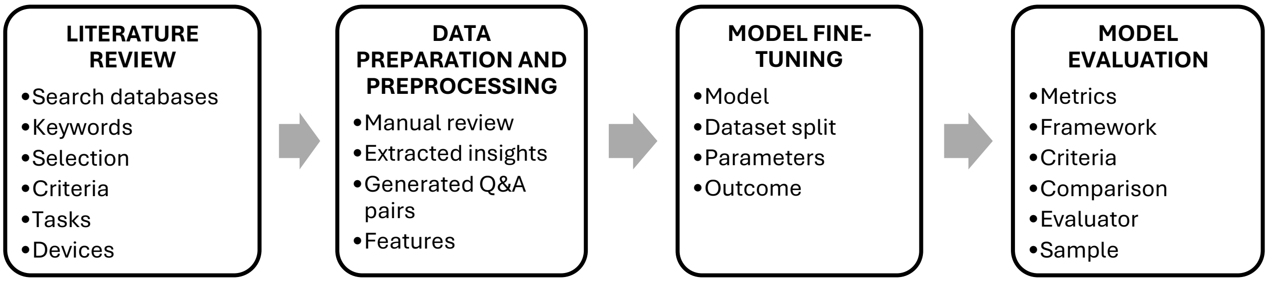


Figure 1. Overview of methodology.

3.1 Literature Review

The first stage consisted of a review of peer-reviewed studies to compile a dataset on the biomechanical benefits, human factors risks, and adoption considerations of active back-support exoskeletons in construction. Searches were performed across Google Scholar, ScienceDirect, IEEE Xplore, and the American Society of Civil Engineers Library using combinations of keywords including “active exoskeletons AND construction”, “powered exoskeletons AND construction”, and “wearable robotic devices AND construction”. A total of 20 studies, published between 2022 and 2025, were selected according to inclusion criteria that emphasized evaluations of active back-support exoskeletons across diverse construction tasks. Selection prioritized articles reporting both objective biomechanical metrics (e.g., electromyographic activity and kinematic range of motion) and subjective measures (e.g., perceived discomfort and cognitive load), in line with best practices in applied human factors research. The reviewed tasks included manual material handling, carpentry, railway maintenance, masonry, flooring, rebar tying, overhead drilling, dynamic lifting, and load carrying. Exoskeletons covered spanned commercially available systems (e.g., Cray X, Iron Hand, Apogee, StreamEXO, and soft robotic suits) as well as laboratory-developed prototypes.

3.2 Data Preparation and Preprocessing

The second stage involved transforming insights from the reviewed studies into a dataset suitable for RAFT. Each paper was manually reviewed to ensure core findings related to the biomechanical, psychological, and physiological implications of exoskeleton use are represented in the studies. For example, findings such as “active back-support exoskeletons reduce lumbar strain but increase shoulder discomfort during prolonged tasks” were reviewed. From each of the identified studies, practitioner-oriented Q&A pairs were generated (e.g., "What is the reduction in lumbar range of motion when using a back-support exoskeleton?"), with answers synthesizing empirical evidence, references, and Chain-of-Thought reasoning to approximate expert logic. A total of 3,650 Q&A pairs were produced using the RAFTDataGen algorithm, powered by the GPT-4o model on Microsoft Azure's AI cloud infrastructure. This algorithm supports the incorporation of distractors and explicit reasoning chains, ensuring dataset richness and reducing overfitting (Wu, Ding, Shen, & Tao, 2025). The resulting dataset size substantially exceeded comparable RAFT-based fine-tuning studies in adjacent domains (Sager, Cabaza, Cusack, Bass, & Dominguez, 2024; Shi, Kazda, Schmitter, & Gupta, 2025), strengthening model generalizability.

3.3 Model Fine-Tuning

The curated dataset was used to fine-tune GPT-4o-mini, a lightweight variant of GPT-4o optimized for speed and reasoning depth in applied decision-support contexts. The dataset was split into 70% training (8,030,685 tokens), 15% validation, and 15% testing subsets. Training followed OpenAI's recommended fine-tuning procedures, with a batch size of 32, a learning rate of 5×10^{-5}, and three epochs to balance convergence and generalization. RAFT fine-tuning enabled the model to internalize domain-specific reasoning patterns, such as linking electromyographic evidence of muscle fatigue with subjective reports of discomfort. This capacity is particularly critical for construction safety decision-making, where real-time sensor data are limited and perceptual insights must be inferred.

3.4 Model Evaluation

The performance of the fine-tuned model was first assessed using automated metrics, including validation accuracy and loss curves. In addition, the LLM-as-a-Judge framework was employed, which involves using independent LLMs to evaluate the outputs of the fine-tuned model (Zheng et al., 2023). This method has been shown to achieve up to 80% agreement with human raters and is valued for its scalability, consistency, and cost-effectiveness in evaluating LLM outputs (Zheng et al., 2023). Although human judgment remains the gold standard due to its nuanced contextual understanding, LLM-as-a-Judge provides an efficient alternative for large-scale evaluation. A single-answer grading protocol, as proposed by Zheng et al. (2023), was adopted, in which model responses were scored against predefined criteria. The evaluation dimensions included coherence (clarity and structure of the response), relevance (alignment with the prompt), completeness (coverage of essential points), factual accuracy (truthfulness of content), and harmlessness (absence of harmful or biased information). This framework has been widely adopted in recent fine-tuning studies (Saroufim et al., 2025). For the evaluation, 20 Q&A pairs were randomly sampled from the reserved 15% test set. These queries were presented to both the fine-tuned model and the baseline GPT-4o-mini via the Microsoft Azure Playground. Outputs from each model were then independently scored by LLaMA 4, which served as the evaluation agent. Scores across criteria were aggregated, and results were compared to determine the performance gains of the fine-tuned model relative to the baseline.

4 Results

This section presents the results of the study, encompassing both the findings from the literature review and the performance evaluation of the fine-tuned model.

4.1 Literature Review

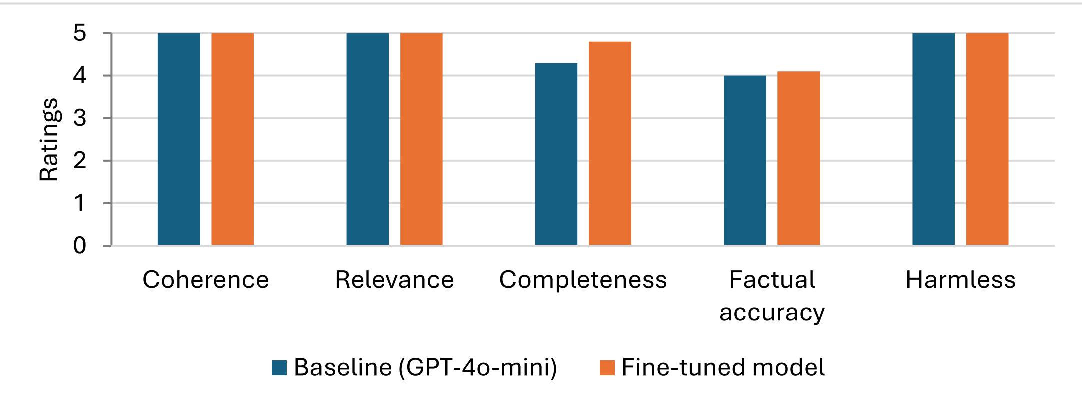
The literature review process identified twenty peer-reviewed studies, which collectively advanced understanding of the biomechanical effects of active back-support exoskeletons in construction and the sociotechnical factors shaping their adoption. Across these studies, common biomechanical metrics included reductions in muscle activity (via electromyography), improvements in range of motion, and decreases in perceived discomfort and exertion. For example, Akinwale Okunola, Akanmu, Jebelli, et al. (2025) demonstrated significant reductions in lumbar muscle activation during framing tasks, while (Sposito et al., 2024) reported decreased fatigue in railway construction when using active back-support exoskeletons. At the same time, challenges such as increased cognitive workload, thermal discomfort, and restricted mobility. Factors that may increase fall risks were also documented (Akinwale Okunola et al., 2024; Sposito et al., 2024). In addition to physiological and ergonomic evaluations, the studies highlighted adoption facilitators and barriers from multiple stakeholder perspectives. Facilitators included productivity gains, reductions in worker compensation claims, posture correction, device durability, and ease of maintenance. Barriers included task–device incompatibility, restricted mobility, excessive device weight, thermal discomfort, and hygiene concerns associated with shared use. These findings underscore the multifaceted nature of integrating active back-support exoskeletons into construction workflows and the need for decision-support tools that synthesize both quantitative biomechanical data and qualitative stakeholder insights.

4.2 Model Fine-Tuning Performance

Following dataset preparation, the RAFT-based fine-tuning of GPT-4o-mini was successfully completed on the curated dataset of 3,650 Q&A pairs. The model achieved validation accuracy of 88% by the third epoch, indicating stable learning and effective generalization without notable overfitting, as shown by the low validation loss. The RAFT framework's integration of distractor documents and Chain-of-Thought reasoning enhanced contextual discernment, enabling the model to separate relevant biomechanical insights from unrelated content. This improvement translated into more nuanced, task-specific recommendations for active back-support exoskeleton deployment in construction scenarios.

4.3 LLM-as-a-Judge Evaluation

The LLM-as-a-Judge evaluation offered a detailed comparison of the fine-tuned GPT-4o-mini and the base model across 20 randomly selected Q&A pairs. As shown in Figure 2, the fine-tuned model outperformed the base model across all metrics. Specifically, it achieved perfect scores of 5 out of 5 on coherence, relevance, and harmlessness. This indicates consistently clear, contextually appropriate, and safe outputs. On completeness, the fine-tuned model scored 4.8 (96%) compared to 4.3 (86%) for the base model, demonstrating its ability to deliver more comprehensive answers that connected



biomechanical outcomes with adoption concerns. For factual accuracy, the fine-tuned model achieved 4.1 (82%), slightly higher than the base model's 4.0 (80%). These results confirm the RAFT framework's effectiveness in embedding domain-specific expertise into a compact, deployable model.

Figure 2. Performance evaluation.

5 Discussion

The development and evaluation of the fine-tuned LLM in this study introduce an AI-driven approach to supporting the selection and adoption of active back-support exoskeletons in construction. The model's ability to synthesize biomechanical insights alongside stakeholder priorities addresses a critical gap in current exoskeleton evaluation frameworks, which often emphasize either physiological metrics or qualitative perceptions in isolation (De Looze, Bosch, Krause, Stadler, & O'Sullivan, 2016). By integrating these dimensions, the fine-tuned model provides a holistic lens through which professionals can evaluate trade-offs in active back-support exoskeleton deployment.

The literature review formed a critical empirical foundation, ensuring that the training data reflected the multifaceted impacts of active back-support exoskeletons—from muscle activity and range of motion to cognitive workload and fall risk. For example, reductions in muscle activation ranging from 12–22% during manual material handling tasks have been documented (Lazzaroni et al., 2020), while decreases in range of motion are also evident in carpentry (Akinwale Okunola, Akanmu, Jebelli, et al., 2025) and flooring tasks (Akinwale Okunola, Akanmu, & Yusuf, 2023). Equally important, the synthesis of facilitators and barriers from diverse stakeholders added a practical layer that is often underrepresented in biomechanical research. This dual perspective enabled the model to produce recommendations that are both scientifically grounded and operationally relevant. The RAFT-based fine-tuning demonstrated the value of combining teacher-student distillation with domain-specific datasets. The resulting validation accuracy of 88%, along with qualitative improvements in completeness and factual accuracy, indicates meaningful gains over general-purpose LLMs. Such enhancements are vital for decision-support applications, where comprehensiveness and factual precision underpin trust and compliance (Bommasani, 2021). The LLM-as-a-Judge evaluation further reinforced these strengths, particularly in generating coherent, relevant, and safe outputs. Research has shown that powerful LLMs like GPT-4 can align with human preferences over 80% of the time in evaluations (Zheng et al., 2023). Nonetheless, LLM-based evaluators remain prone to biases (e.g., position bias, verbosity bias, and self-enhancement bias) as well as limited reasoning consistency (Zheng et al., 2023). These findings suggest that while LLM-as-a-Judge offers a cost-effective alternative to human evaluation, careful prompt engineering and bias mitigation are essential.

From a practical standpoint, the fine-tuned model represents a pathway toward integrating artificial intelligence into construction safety and ergonomics planning. By enabling stakeholders to query the model on task-specific active back-support exoskeleton recommendations, cognitive risks, and adoption barriers, organizations can make data-driven decisions that balance productivity with worker well-being. The lightweight architecture of GPT-4o-mini further supports deployment in resource-constrained environments such as mobile or on-site decision aids. Nonetheless, limitations remain. The model's knowledge is bounded by the breadth of available literature, which, though comprehensive, may not fully capture emerging device technologies or contextual nuances such as environmental hazards and workforce diversity (e.g., gender-specific ergonomic responses). Additionally, while the RAFT framework reduced hallucinations, residual inaccuracies persist, as reflected in factual accuracy scores. These challenges point to the need for ongoing dataset expansion, iterative retraining, and user-in-the-loop mechanisms to refine reliability over time. Furthermore, given the documented biases in LLM-as-a-Judge evaluations (Zheng et al., 2023), future work should incorporate human expert validation or ensemble judging strategies to bolster evaluation robustness.

6 Conclusions

This study introduced a domain-specific LLM fine-tuned with the RAFT framework to synthesize biomechanical insights and sociotechnical factors influencing active back-support exoskeleton adoption in construction. By curating a dataset of 3,650 Chain-of-Thought Q&A pairs derived from twenty peer-reviewed studies, the model embedded empirical evidence and stakeholder perspectives. The fine-tuned GPT-4o-mini achieved 88% validation accuracy and outperformed the base model on completeness and factual accuracy, as validated through LLM-as-a-Judge evaluation. The model provides a scalable decision-support tool that assists construction professionals, ergonomists, and safety officers in evaluating task-specific suitability of active back-support exoskeletons. However, the model's insights remain constrained by the scope of available literature and may not reflect rapidly evolving technologies or diverse construction contexts. While RAFT minimized hallucinations, consistently high factual accuracy remains a challenge in domains requiring interdisciplinary reasoning.

Future research should expand the dataset to cover additional exoskeleton types (e.g., upper-limb supports and multi-joint exosuits), integrate multimodal data (e.g., sensor streams, real-time usage analytics, and environmental factors), and develop human-in-the-loop systems to combine AI recommendations with expert oversight. These steps would improve precision, adaptability, and trustworthiness. This study establishes a foundation for the intelligent integration of AI and biomechanics in construction, advancing safer, more ergonomic work environments through technology-informed decisions. Continued progress in this direction may accelerate the responsible adoption of wearable robotics across physically demanding industries, contributing to long-term occupational health and productivity gains.

Funding

This material is based upon work supported by the [Anonymized for Review] under Grant Nos. [Anonymized for Review] and [Anonymized for Review]. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the [Anonymized for Review].

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author.

Conflicts of Interest

The authors declare no conflict of interest.

References

Akanmu, A., Okunola, A., Jebelli, H., Ammar, A., & Afolabi, A. (2024). Cognitive load assessment of active back-support exoskeletons in construction: A case study on construction framing. *Advanced Engineering Informatics, 62*, 102905. https://doi.org/10.1016/j.aei.2024.102905

Ali, A., Fontanari, V., Schmoelz, W., & Agrawal, S. K. (2021). Systematic review of back-support exoskeletons and soft robotic suits. *Frontiers in bioengineering and biotechnology, 9*, 765257. https://doi.org/10.3389/fbioe.2021.765257

Bommasani, R. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*. https://doi.org/10.48550/arXiv.2108.07258

Chung, I., Vo, P., Kizilkale, A. C., & Reite, A. (2024). Efficient In-Domain Question Answering for Resource-Constrained Environments. https://doi.org/10.48550/arXiv.2409.17648.

Crea, S., Beckerle, P., De Looze, M., De Pauw, K., Grazi, L., Kermavnar, T., Masood, J., O'Sullivan, L.W., Pacifico, I., Rodriguez-Guerrero, C. & Vitiello, N. (2021). Occupational exoskeletons: A roadmap toward large-scale adoption. Methodology and challenges of bringing exoskeletons to workplaces. *Wearable Technol, 2*, e11. https://doi.org/10.1017/wtc.2021.11

De Looze, M. P., Bosch, T., Krause, F., Stadler, K. S., & O'Sullivan, L. W. (2016). Exoskeletons for industrial application and their potential effects on physical work load. *Ergonomics, 59*(5), 671–681. https://doi.org/10.1080/00140139.2015.1081988

Dunson-Todd, M., Nik-Bakht, M., & Hammad, A. (2025). Proposed Standard Test to Evaluate Back-Support Exoskeleton Efficacy for Rebar Workers: Test Design and Initial Implementation. *Journal of Computing in Civil Engineering, 39*(5), 04025071. https://doi.org/10.1061/JCCEE5.CPENG-6481

Gutierrez, N., Ojelade, A., Kim, S., Barr, A., Akanmu, A., Nussbaum, M. A., & Harris-Adamson, C. (2024). Perceived benefits, barriers, perceptions, and readiness to use exoskeletons in the construction industry: Differences by demographic characteristics. *Applied ergonomics, 116*, 104199. https://doi.org/10.1016/j.apergo.2023.104199

Hadi, M. U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Wu, J. & Mirjalili, S. (2023). Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects. *Authorea Preprints, 1*, 1–26. https://doi.org/10.36227/techrxiv.23589741.v8

Koppel, T., Tšernikova, O., & Kalvet, T. (2024). *Virtual risk assistant–a rule-based risk assessment tool based on large language models*. Paper presented at the ISPIM Innovation Symposium. https://www.proquest.com/docview/3089910193/fulltextPDF/7138002BA4B34EBEPQ/1?accountid=14826&sourcetype=Conference%20Papers%20&%20Proceedings

Lazzaroni, M., Tabasi, A., Toxiri, S., Caldwell, D. G., De Momi, E., van Dijk, W., de Looze, M.P., Kingma, I., van Dieën, J.H. & Ortiz, J. (2020). Evaluation of an acceleration-based assistive strategy to control a back-support exoskeleton for manual material handling. *Wearable Technologies, 1*, e9. https://doi.org/10.1017/wtc.2020.8

Lei, T., Liang, K., Xu, J., Li, H., Seo, J. O., & Heung, K. H. (2025). Effects on muscular activity and usability of soft active versus rigid passive back exoskeleton during symmetric lifting tasks. *Scientific Reports, 15*(1), 29839. https://doi.org/10.1038/s41598-025-14500-3

Lei, T., Seo, J., Liang, K., Xu, J., Li, H., Zhou, Y., Khan, M. & Heung, K.H. (2024). Lightweight active soft back exosuit for construction workers in lifting tasks. *Journal of Construction Engineering and Management, 150*(7), 04024073. https://doi.org/10.1061/JCEMD4.COENG-14490

Ojha, A., Guo, H., Jebelli, H., Martin, A., & Akanmu, A. (2024). Assessing the Impact of Active Back Support Exoskeletons on Muscular Activity during Construction Tasks: Insights from Physiological Sensing. In *Computing in Civil Engineering 2023* (pp. 340–347). https://doi.org/10.1061/9780784485248.041

Okunola, A., Afolabi, A., Akanmu, A., Jebelli, H., & Simikins, S. (2024). Facilitators and barriers to the adoption of active back-support exoskeletons in the construction industry. *J Safety Res, 90*, 402–415. https://doi.org/10.1016/j.jsr.2024.05.010

Okunola, A., Akanmu, A., & Jebelli, H. (2024). Fall risk assessment of active back-support exoskeleton-use for construction work using foot plantar pressure distribution. *Advanced Engineering Informatics, 62*, 102626. https://doi.org/10.1016/j.aei.2024.102626

Okunola, A., Akanmu, A., Jebelli, H., & Afolabi, A. (2025). Assessment of active back-support exoskeleton on carpentry framing tasks: Muscle activity, range of motion, discomfort, and exertion. *International Journal of Industrial Ergonomics, 107*, 103716. https://doi.org/10.1016/j.ergon.2025.103716

Okunola, A., Akanmu, A. A., & Yusuf, A. O. (2023). Comparison of active and passive back-support exoskeletons for construction work: range of motion, discomfort, usability, exertion and cognitive load assessments. *Smart and Sustainable Built Environment, 14*(3), 582–598. https://doi.org/10.1108/SASBE-06-2023-0147

Pu, H., Yang, X., Li, J., & Guo, R. (2024). AutoRepo: A general framework for multimodal LLM-based automated construction reporting. *Expert Systems with Applications, 255*, 124601. https://doi.org/10.1016/j.eswa.2024.124601

Reimeir, B., Calisti, M., Mittermeier, R., Ralfs, L., & Weidner, R. (2023). Effects of back-support exoskeletons with different functional mechanisms on trunk muscle activity and kinematics. *Wearable Technologies, 4*, e12. https://doi.org/10.1017/wtc.2023.5

Sager, N., Cabaza, T., Cusack, M., Bass, R., & Dominguez, J. (2024). Rethinking Retrieval Automated Fine-Tuning in an evolving LLM landscape. *SMU Data Science Review, 8*(2), 2. https://scholar.smu.edu/datasciencereview/vol8/iss2/2

Saroufim, M., Perlitz, Y., Choshen, L., Antiga, L., Bowyer, G., Puhrsch, C., Guessous, D., Rao, S., Chauhan, G., Kumar, A. & Kumar, J.P. (2025). Neurips 2023 llm efficiency fine-tuning competition. *arXiv preprint arXiv:2503.13507*. https://doi.org/10.48550/arXiv.2503.13507

Schwartz, M., Desbrosses, K., Theurel, J., & Mornieux, G. (2023). Biomechanical consequences of using passive and active back-support exoskeletons during different manual handling tasks. *International Journal of Environmental Research and Public Health, 20*(15), 6468. https://doi.org/10.3390/ijerph20156468

Shi, L., Kazda, M., Schmitter, C., & Gupta, H. (2025). Improving LLM-Powered EDA Assistants with RAFT. *arXiv preprint arXiv:2506.06500*. https://doi.org/10.48550/arXiv.2506.06500

Sposito, M., Fanti, V., Poliero, T., Caldwell, D. G., & Di Natali, C. (2024). Field assessment of active BSE: Trends over test days of subjective indicators and self-reported fatigue for railway construction workers. *Heliyon, 10*(12). https://doi.org/10.1016/j.heliyon.2024.e33055

Susnjak, T., Hwang, P., Reyes, N., Barczak, A. L., McIntosh, T., & Ranathunga, S. (2025). Automating research synthesis with domain-specific large language model fine-tuning. *ACM Transactions on Knowledge Discovery from Data, 19*(3), 1–39. https://doi.org/10.1145/3715964

Wu, Y., Ding, L., Shen, L., & Tao, D. (2025). Robust Knowledge Editing via Explicit Reasoning Chains for Distractor-Resilient Multi-Hop QA. *arXiv preprint arXiv:2509.01468*. doi:https://doi.org/10.48550/arXiv.2509.01468

Zhang, T., Patil, S. G., Jain, N., Shen, S., Zaharia, M., Stoica, I., & Gonzalez, J. E. (2024). *Raft: Adapting language model to domain specific rag.* Paper presented at the First Conference on Language Modeling. https://doi.org/10.48550/arXiv.2403.10131

Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E. & Zhang, H., (2023). Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in Neural Information Processing Systems, 36*, 46595–46623. https://doi.org/10.48550/arXiv.2306.05685

Zhu, Y., Weston, E. B., Mehta, R. K., & Marras, W. S. (2021). Neural and biomechanical tradeoffs associated with human-exoskeleton interactions. *Appl Ergon, 96*, 103494. https://doi.org/10.1016/j.apergo.2021.103494