

ORIGINAL RESEARCH PAPER

A RAG-Based Decision Support System for Occupational Exposure Assessment of VOCs: Design, Implementation, and Evaluation

Hadis Vahedi¹, Mehrnoush Shamsfard², Ehsan Toofaninejad³, Zahra Vatankhah Mohammadabadi², Fatemeh Dorfeshan⁴, Somayeh Farhang Dehghan^{5,1*}

¹Department of Occupational Health and Safety Engineering, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

²Natural Language Processing (NLP) Lab, Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran

³E-Learning in Medical Sciences Department, School of Medical Education and Learning Technologies, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

⁴Student Research Committee, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

⁵Environmental and Occupational Hazards Control Research Center, Research Institute for Health Sciences and Environment, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Received: 2025/12/29

Accepted: 2026/07/18

ABSTRACT

Introduction: Assessing exposure to volatile organic compounds (VOCs) in workplaces requires timely access to accurate, standardized guidance on sampling and analysis. This study aimed to develop and evaluate a decision-support chatbot, based on the Retrieval-Augmented Generation (RAG) architecture, to assist occupational health professionals in selecting appropriate sampling and analysis methods for VOCs in accordance with NIOSH guidelines.

Material and Methods: This research was of an applied-developmental type. In this study, relevant NIOSH methods for sampling and analysis of VOCs were extracted, structured, and organized into a knowledge base. User needs were identified through semi-structured interviews with occupational health professionals. Based on this data, a generative AI-based chatbot was designed and implemented using the RAG architecture. The system performance was evaluated using expert evaluation and a chatbot usability questionnaire.

Results: The developed chatbot received an overall user-friendliness score of 3.88 out of 5, indicating good user acceptance. Participants noted improved access speed for standard technical information and the high transparency of the system's responses. Expert evaluation also confirmed the scientific validity of the retrieved content and its compliance with NIOSH standards.

Conclusion: The proposed RAG-based decision-making chatbot provides an effective and innovative tool to support occupational exposure assessment to VOCs. Integrating generative AI with standard occupational health guidelines can improve the efficiency of decision-making and access to expert information in practice.

Keywords: Occupational health, Chatbot, Retrieval-Augmented Generation (RAG), NIOSH standards, Volatile organic compounds (VOCs), Decision support system (DSS)

HOW TO CITE THIS ARTICLE

Vahedi H., Shamsfard M., Toofaninejad E., Vatankhah Mohammadabadi Z., Dorfeshan F., Farhang Dehghan S. A RAG-Based Decision Support System for Occupational Exposure Assessment of VOCs: Design, Implementation, and Evaluation. *J Health Saf Work*. 2026; 16(2): 345-370.

* Corresponding Author Email: somayeh.farhang@gmail.com

Copyright © 2026 The Authors.

Published by Tehran University of Medical Sciences

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license (<https://creativecommons.org/licenses/by-nc/4.0/>). Non-commercial uses of the work are permitted, provided the original work is properly cited.

1. INTRODUCTION

Volatile organic compounds (VOCs) represent a major class of chemical hazards in occupational environments, commonly encountered across industries such as petrochemicals, printing, paint manufacturing, laboratories, and healthcare settings. Occupational exposure to VOCs can lead to a range of acute and chronic health effects, including respiratory irritation, neurological disturbances, hepatic and renal damage, and, in some cases, increased cancer risk. Accurate assessment of workplace exposure to VOCs is therefore a cornerstone of occupational health programs, requiring precise selection and execution of sampling and analytical methods. The National Institute for Occupational Safety and Health (NIOSH) provides a comprehensive and scientifically validated collection of standardized methods for VOC monitoring—widely regarded as a global reference in occupational hygiene. However, despite their scientific rigor, NIOSH methods are often challenging to apply in real-world settings due to their technical complexity, fragmented structure, and extensive documentation. Occupational health professionals frequently struggle to quickly identify the most appropriate method for a given compound and workplace context, correctly interpret procedural details, or troubleshoot implementation issues under time pressure. This gap between standardized guidance and practical application can lead to suboptimal exposure assessments, misinterpretations of risk, and ultimately, inadequate worker protection. While digital tools and AI-driven assistants have emerged in recent years to support decision-making in health-related domains, few are tailored to the highly technical, document-intensive, and safety-critical nature of occupational hygiene. There remains a clear need for intelligent, interactive systems capable of delivering structured, authoritative, and context-

aware guidance on NIOSH methods in real time. To address this gap, we developed HAVANA—a specialized generative AI-powered chatbot built on a Retrieval-Augmented Generation (RAG) architecture—designed to assist occupational health experts in selecting, interpreting, and applying NIOSH methods for VOC sampling and analysis (Fig 1).

2. MATERIAL AND METHODS

This applied developmental study followed a user-centered design and iterative development approach across four integrated phases. First, a qualitative needs assessment was conducted through semi-structured interviews with 12 occupational health professionals, purposively sampled to ensure diversity in expertise, organizational roles, and field experience. Participants included industrial hygiene practitioners, academic faculty, technical managers, public health officials, and executives from occupational hygiene service companies, all with confirmed experience in VOC exposure assessment and in implementing NIOSH methods. Interviews focused on identifying real-world challenges, information needs, and preferred interaction patterns during method selection and execution. Data collection continued until informational saturation was achieved. Transcripts were analyzed using thematic content analysis to extract key user requirements and question archetypes.

In the second phase, these insights informed the design of a structured set of user-oriented questions, categorized into ten thematic types (e.g., chemical properties, risk assessment, sampling protocols, analytical procedures). Each question type was mapped to a specific response-generation mechanism—either a Large Language Model (LLM), the RAG pipeline, or a hybrid approach—based on whether the response required general

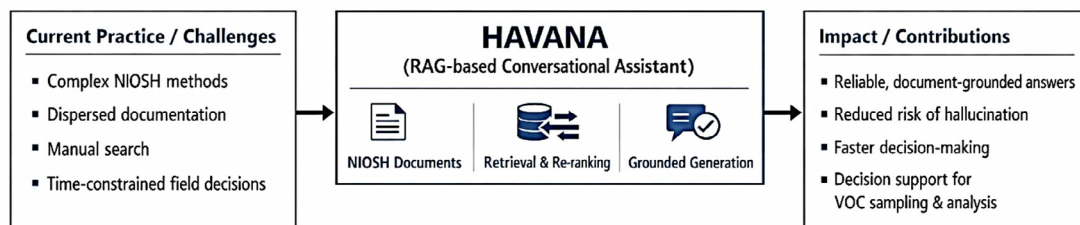


Fig. 1: Conceptual overview of current challenges in using NIOSH methods and the HAVANA RAG-based solution for reliable decision support in VOC sampling and analysis.

Table 1: Thematic Question Types and Response Mechanisms

No.	Question Type	Question Nature	Interview Frequency	Response Mechanism
1	Chemical properties	Identification, physicochemical characteristics, and applications	9	LLM
2	Exposure risk management	Exposure limits, hazard indices, and safety decisions	8	LLM & RAG
3	Sampling protocols	Method selection, adsorbent media, and environmental conditions	8	RAG
4	Technical sampling details	Step-by-step procedural instructions and field protocols	7	RAG
5	Laboratory analysis	Sample preparation and instrumental analysis (GC/MS, GC-FID)	7	RAG
6	Technical analytical details	Instrument settings, column specifications, and detectors	6	LLM
7	Toxicological aspects	Clinical symptoms and toxicity mechanisms	6	RAG
8	Calibration & QA/QC	Instrument validation and quality assurance protocols	5	RAG
9	Exposure calculations	TWA formulas, correction factors, and concentration determination	5	LLM
10	General troubleshooting	Technical support and initial problem-solving	4	LLM

Note: LLM: Large Language Model; RAG: Retrieval-Augmented Generation; GC/MS: Gas Chromatography-Mass Spectrometry; GC-FID: Gas Chromatography-Flame Ionization Detector; QA/QC: Quality Assurance/Quality Control; TWA: Time-Weighted Average.

knowledge, precise retrieval from standard documents, or interpretive synthesis.

The third phase involved the technical development of the HAVANA chatbot. The core knowledge base was constructed by extracting, structurally segmenting, and preprocessing official NIOSH method documents (PDF format). Unlike fixed-size text chunking, we implemented a logic-preserving segmentation strategy that respected document sections (e.g., sampling procedure, analytical method, QA/QC), ensuring contextual coherence. These chunks were embedded using a scientific-domain-optimized embedding model and stored in a Qdrant vector database with associated metadata (e.g., NIOSH method number, content type). The RAG pipeline included query rewriting (for English retrieval), dense vector search, and cross-encoder-based re-ranking to enhance retrieval precision. Retrieved chunks were passed as context to a state-of-the-art LLM (GPT-4o), constrained to generate responses solely from the retrieved evidence, and the resulting output was translated into clear, step-by-step Persian.

Finally, in the fourth phase, the system was evaluated using both quantitative and qualitative methods. Usability was assessed using the validated Persian version of the Chatbot Usability Questionnaire (CUQ), which was completed by the same 12 experts after interactive testing. Additionally, open-ended feedback forms captured qualitative insights on scientific accuracy, alignment with NIOSH standards, clarity, and limitations.

3. RESULTS AND DISCUSSION

The expert panel comprised 12 professionals (7 male, 5 female) with diverse backgrounds: 5 in industrial hygiene services, 3 in analytical laboratories/public health departments, 2 in academia, and 2 as HSE managers. Their average work experience was 9 years ($SD = 1.95$), ensuring the practical relevance of the input.

Ten thematic question types were identified (Table 1), with the most frequent relating to chemical properties (9 mentions), exposure risk management (8), and sampling protocols (8). The response-generation mechanism was intelligently assigned: RAG was used for questions requiring exact procedural details (e.g., sampling flow rates, analytical parameters), LLM for general or conceptual queries (e.g., VOC health effects), and hybrid RAG+LLM for interpretive tasks (e.g., risk-based decision-making).

- Technical evaluations demonstrated significant performance gains from design choices:
- Structured chunking improved retrieval accuracy from 39% (fixed-size) to 73% on a 200-question test set.

- Re-ranking with a cross-encoder increased correct responses from 62% to 92% in a 50-question benchmark.

- Optimal $k = 10$ balanced precision (37.2%) and recall (89.4%) in retrieval, avoiding noise overload while ensuring coverage.

According to Table 2, Usability results were highly favorable (mean CUQ score = 3.88/5). “Response

Table 2: Mean and Standard Deviation of Chatbot Usability Questionnaire (CUQ) Components (n = 12)

No.	Usability Component	Mean Score	Standard Deviation
1	Clarity of responses	4.09	0.83
2	Relevance of content to user needs	3.73	1.10
3	Response speed	3.82	1.08
4	Ease of interaction	3.91	0.54
5	Conversation continuity	3.64	1.21
6	Overall user experience	3.73	1.27
7	Understanding of the user's prompt	3.91	1.04
8	Appropriateness of tone and communication style	4.09	0.94
9	Responsiveness to specific and technical needs	3.91	1.04
10	Ability to learn and improve during interaction	2.27	1.49

Note: Scores are based on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). Higher scores indicate more favorable perceptions of usability.

clarity” and “communication tone” received the highest scores (4.09), while “conversation continuity” and “learning capability” received the lowest scores (2.27), reflecting limitations in multi-turn dialogue. Experts unanimously praised the system’s scientific accuracy, adherence to NIOSH standards, and time-saving potential, though some noted room for improvement in contextual memory during extended interactions.

The findings confirm that a RAG-based architecture is well-suited for high-stakes, standards-driven domains like occupational hygiene. By grounding responses in authoritative documents, HAVANA overcomes a key limitation of generic LLMs—namely, the risk of hallucination or outdated advice. The user-centered design process ensured that the system addresses real-world needs, particularly the translation of complex standards into actionable steps. The hybrid response strategy (LLM vs. RAG vs. combined) proved essential: no single mechanism could cover the full spectrum of user queries, from factual lookup to risk interpretation. The emphasis on structured document chunking and re-ranking aligns with recent research showing that the quality of data preprocessing often outweighs model size in determining RAG performance. Usability results highlight a critical trade-off: while clarity and reliability were strengths, conversational persistence remains a challenge for current LLMs in specialized contexts. Nonetheless, experts valued accuracy and traceability over conversational fluency—a preference consistent with other technical domains (e.g., clinical or regulatory AI). HAVANA thus represents a pragmatic balance between usability and reliability, positioning it as a decision support tool rather than an autonomous agent.

4. CONCLUSIONS

This study demonstrates the feasibility and value of a RAG-powered chatbot as a decision support system for occupational health professionals managing VOC exposure. HAVANA successfully bridges the gap between complex NIOSH documentation and on-the-ground decision-making by delivering precise, traceable, and user-friendly guidance. The integration of structured knowledge engineering, evidence-based retrieval, and constrained generation ensures high scientific fidelity, while the user-centered design process guarantees practical relevance. Although limitations exist—particularly in multi-turn dialogue and reliance on input document quality—the system was well received by experts and shows strong potential for real-world deployment. Future work should expand the knowledge base to include other international standards (e.g., OSHA, ISO), conduct field trials in industrial settings, and enhance contextual memory for complex interactions. Ultimately, HAVANA exemplifies how generative AI, when carefully constrained and grounded in domain-specific knowledge, can empower professionals without replacing human expertise—offering a scalable model for intelligent support in other standards-intensive fields.

5. ACKNOWLEDGMENT

This article is based on an MSc thesis conducted in the Department of Occupational Health and Safety at Work, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran (Thesis Code: 43011152). This study was supported by Shahid Beheshti University of Medical Sciences (Grant No. 43018675). The

authors acknowledge the use of the Gemini large language model (version 1.5 Flash) to support language refinement and improve the expression of certain sections of this manuscript. All outputs generated with AI assistance were critically reviewed, revised as necessary, and approved by the authors. The authors assume full responsibility for the final content of the manuscript.

6. ETHICAL CODE

The study received ethical approval from the Ethics Committee of Shahid Beheshti University of Medical Sciences (IR.SBMU.PHNS.REC.1403.047) and adhered to the Declaration of Helsinki (2013) and national biomedical research ethics guidelines. Written informed consent was obtained from all

participants before the semi-structured interviews, and they were assured of the voluntary nature of participation and their right to withdraw at any time.

Personal identifiers were removed or coded, and all interview data were securely stored with restricted access. No identifiable information appears in the reported results. NIOSH standard data were used as publicly available sources with proper citation. The authors declare no conflicts of interest and confirm the independence of data analysis and interpretation.

7. FUNDING

This study was financially supported by Shahid Beheshti University of Medical Sciences (Grant No. 43018675).

سامانه تصمیم‌یار مبتنی بر RAG برای ارزیابی مواجهه شغلی با VOCs:

طراحی، پیاده‌سازی و ارزیابی

حدیث واحدی^۱، مهرنوش شمس‌فرد^۲، احسان طوفانی نژاد^۳، زهرا وطن‌خواه محمدآبادی^۴، فاطمه درفشان^۴،

سمیه فرهنگ دهقان^{۵،*}

^۱گروه مهندسی بهداشت حرفه ای و ایمنی کار، دانشکده بهداشت و ایمنی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.

^۲آزمایشگاه NLP، دانشکده علوم و مهندسی کامپیوتر، دانشگاه شهید بهشتی، تهران، ایران

^۳گروه یادگیری الکترونیکی در علوم پزشکی، دانشکده آموزش پزشکی و فناوریهای یادگیری، دانشگاه علوم پزشکی شهید بهشتی،

تهران، ایران

^۴کمیته تحقیقات دانشجویی، دانشکده بهداشت و ایمنی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.

^۵مرکز تحقیقات کنترل عوامل زیان آور محیط و کار، پژوهشکده علوم بهداشتی و محیط زیست، دانشگاه علوم پزشکی شهید

بهشتی، تهران، ایران

تاریخ دریافت: ۱۴۰۴/۱۰/۸، تاریخ پذیرش: ۱۴۰۵/۴/۲۷

چکیده

مقدمه: ارزیابی مواجهه با ترکیبات آلی فرار (VOCs) در محیط‌های کاری نیازمند دسترسی به‌موقع به راهنمایی‌های دقیق و استاندارد برای نمونه‌برداری و آنالیز است. هدف این مطالعه، توسعه و ارزیابی یک چت‌بات تصمیم‌یار مبتنی بر معماری Retrieval Augmented Generation (RAG) به‌منظور کمک به متخصصان بهداشت حرفه‌ای در انتخاب روش‌های مناسب نمونه‌برداری و آنالیز VOCها بر اساس دستورالعمل‌های NIOSH بود.

روش کار: این پژوهش از نوع کاربردی-توسعه‌ای بود. در این مطالعه، روش‌های مرتبط NIOSH برای نمونه‌برداری و آنالیز VOCs استخراج و ساختاربندی شده و در قالب یک پایگاه دانش سازمان‌دهی شدند. نیازهای کاربران از طریق مصاحبه‌های نیمه‌ساختاریافته با متخصصان بهداشت حرفه‌ای شناسایی شد. بر اساس این داده‌ها، یک چت‌بات مبتنی بر هوش مصنوعی مولد با استفاده از معماری RAG طراحی و پیاده‌سازی شد. عملکرد سامانه با استفاده از ارزیابی خبرگان و پرسشنامه کاربرپسندی چت‌بات مورد بررسی قرار گرفت.

یافته‌ها: چت‌بات توسعه‌یافته امتیاز کلی کاربرپسندی ۴/۸۸ از ۵ را کسب کرد که نشان‌دهنده پذیرش مناسب کاربران است. شرکت‌کنندگان به بهبود سرعت دسترسی به اطلاعات فنی استاندارد و شفافیت بالای پاسخ‌های سیستم اشاره کردند. همچنین ارزیابی خبرگان صحت علمی محتوای بازایی‌شده و انطباق آن با استانداردهای NIOSH را تأیید کرد.

نتیجه‌گیری: چت‌بات تصمیم‌یار مبتنی بر RAG پیشنهادی، ابزاری مؤثر و نوآورانه برای پشتیبانی از ارزیابی مواجهه شغلی با VOCs فراهم می‌کند. ادغام هوش مصنوعی مولد با دستورالعمل‌های استاندارد بهداشت حرفه‌ای می‌تواند کارایی تصمیم‌گیری و دسترسی به اطلاعات تخصصی را در عمل بهبود بخشد.

کلمات کلیدی: بهداشت حرفه‌ای، چت‌بات، بازایی تقویت‌شده با تولید (RAG)، استاندارد NIOSH، ترکیبات آلی فرار (VOCs)، سیستم پشتیبان تصمیم‌گیری (DSS)

مقدمه

ترکیبات آلی فرار (VOCs) به‌عنوان یکی از گسترده‌ترین آلاینده‌های شیمیایی در محیط‌های کاری، در طیف وسیعی از صنایع راهبردی از جمله پتروشیمی، صنایع رنگ، چاپ، آزمایشگاه‌های تشخیص طبی و مراکز تولیدی مورد استفاده قرار می‌گیرند. مواجهه شغلی با این ترکیبات به دلیل ویژگی‌های فیزیکوشیمیایی خاص آن‌ها، می‌تواند پیامدهای بیولوژیکی متعددی را در پی داشته باشد. این پیامدها از تحریکات ساده سیستم تنفسی و اختلالات عصبی کوتاه‌مدت آغاز شده و در صورت مواجهه مزمن، به آسیب‌های پایدار کبدی، کلیوی و افزایش معنادار ریسک ابتلا به سرطان‌های شغلی منجر می‌شود (۱). بر این اساس، پایش دقیق و مستمر مواجهه شغلی با VOCs و انتخاب پروتکل‌های صحیح نمونه‌برداری و آنالیز، نه‌تنها یک الزام قانونی، بلکه از ارکان اصلی برنامه‌های صیانت از سلامت نیروی کار محسوب می‌شود (۲).

در حال حاضر، دستورالعمل‌های ارائه‌شده توسط موسسه ملی ایمنی و بهداشت شغلی ایالات متحده (NIOSH)، معتبرترین منبع مرجع برای ارزیابی این آلاینده‌ها در سطح بین‌المللی هستند (۳). این استانداردها مجموعه‌ای حجیم از روش‌های تخصصی را شامل می‌شوند که هر یک دارای جزئیات فنی بسیار دقیق از جمله نوع جاذب (Sorbent)، حجم بهینه نمونه، دبی جریان پیشنهادهی، شرایط حساس نگهداری و انتقال، تکنیک‌های پیشرفته آنالیز آزمایشگاهی و پروتکل‌های سخت‌گیرانه کنترل کیفیت (QC) می‌باشند (۴، ۵). با وجود این غنای علمی، ساختار متنی سنگین و پراکندگی اطلاعات در میان صدها فایل جداگانه، سبب شده است که بهره‌برداری عملی از آن‌ها در شرایط واقعی محیط کار—که اغلب با محدودیت زمان و منابع همراه است—با چالش‌های عملیاتی جدی مواجه گردد.

کارشناسان بهداشت حرفه‌ای در فرآیند ارزیابی، مدام با پرسش‌های کلیدی روبرو هستند: کدام روش استاندارد برای یک ترکیب خاص در حضور تداخل‌کننده‌های محیطی مناسب‌تر است؟ آیا تجهیزات موجود در آزمایشگاه با

الزامات متد همخوانی دارد؟ و پارامترهای عملیاتی چگونه باید برای یک شیفت کاری خاص تنظیم شوند؟ (۶). پاسخ به این پرسش‌ها در حال حاضر مستلزم جستجوی دستی در اسناد طولانی، مقایسه تطبیقی چندین روش مشابه و تفسیر فنی نکات ظریف است؛ فرآیندی که علاوه بر زمان‌بر بودن، ریسک خطای انسانی در تصمیم‌گیری را به‌شدت افزایش می‌دهد. مطالعات نشان می‌دهند که شباهت ظاهری روش‌های NIOSH و تفاوت‌های بسیار ظریف در محدودیت‌های نمونه‌برداری، می‌تواند منجر به انتخاب متد نامتناسب با سناریوی واقعی مواجهه گردد (۶). چنین خطاهایی نه تنها قابلیت مقایسه‌پذیری نتایج را از بین می‌برند، بلکه با برآورد نادرست (بیش‌برآورد یا کم‌برآورد) میزان مواجهه، مستقیماً بر تفسیر ریسک سلامت و اثربخشی اقدامات کنترلی و مدیریتی تأثیر منفی می‌گذارند.

با ظهور تحولات دیجیتال، توسعه ابزارهای هوشمند پتانسیل بالایی برای تسهیل این فرآیندهای پیچیده نشان داده است (۸-۱۰). با این حال، نسل‌های اولیه چت‌بات‌ها و ابزارهای دیجیتال موجود، اغلب یا به ارائه اطلاعات عمومی بسنده می‌کنند و یا به دلیل عدم اتصال به پایگاه‌های دانش تخصصی و به‌روز، خروجی‌های آن‌ها فاقد اعتبار لازم برای استناد در گزارش‌های رسمی بهداشت حرفه‌ای است. در واقع، خلأ پژوهشی بنیادین در این حوزه، نبود یک سامانه تصمیم‌یار هوشمند و تعاملی است که بتواند پل ارتباطی مستقیمی میان «نیاز لحظه‌ای کارشناس» و «پایگاه دانش عظیم متدهای NIOSH» برقرار کند، به‌گونه‌ای که پاسخ‌های ارائه شده کاملاً مستند و مبتنی بر متن صریح استاندارد باشد.

در این راستا، بهره‌گیری از معماری بازبازی تقویت‌شده با تولید (RAG) به عنوان یکی از پیشرفته‌ترین راهکارهای هوش مصنوعی مولد، می‌تواند این چالش را برطرف سازد (۱۱). برتری کلیدی RAG در این است که پاسخ‌های مدل به‌جای اتکا به حافظه درونی و احتمالی (که مستعد پدیده توهم یا Hallucination است)، بر پایه استخراج هدفمند اطلاعات از یک منبع خارجی تأیید شده (پایگاه



شکل ۱: شماتیک کلی چالش‌ها و رویکرد پیشنهادی برای پشتیبانی از ارزیابی مواجهه شغلی با ترکیبات آلی فرار (VOCs) در بهداشت حرفه‌ای.

ربات گفتگوی تخصصی مبتنی بر هوش مصنوعی در حوزه بهداشت حرفه‌ای انجام شد که بتواند فرآیند انتخاب و تفسیر روش‌های استاندارد نمونه برداری و آنالیز مواجهه شغلی با ترکیبات آلی فرار (VOCs) را تسهیل کند. رویکرد کلی مطالعه، یک مطالعه توسعه‌ای-کاربردی بود که در آن از تلفیق روش‌های کیفی، مهندسی داده و پردازش زبان طبیعی برای طراحی یک ربات گفتگوی پشتیبان تصمیم‌سازی استفاده شد (۱۳). مراحل طراحی، توسعه و ارزیابی سامانه به‌گونه‌ای مستندسازی شدند که امکان بازتولید چارچوب پژوهش برای مطالعات مشابه فراهم باشد. نمای کلی مراحل انجام پژوهش و ارتباط آن‌ها با اهداف اصلی مطالعه در شکل ۲ ارائه شده است. در مرحله اول، نیازهای اطلاعاتی و چالش‌های کاربران از طریق مصاحبه‌های کیفی شناسایی شد (۱۴). در مرحله دوم، بر اساس خروجی مرحله اول، طراحی دانش شامل سازمان دهی اسناد استاندارد NIOSH و تدوین سوالات کاربر محور به عنوان مبنای منطق گفتگو و سناریوهای پرسش پاسخ انجام گرفت (۱۵). مرحله سوم به توسعه ربات گفتگوی تخصصی مبتنی بر معماری بازایی تقویت شده با تولید اختصاص یافت (۱۱، ۱۶، ۱۷) و در نهایت، در مرحله چهارم، کاربرد پذیری ربات گفتگو توسعه یافته با استفاده از پرسشنامه استاندارد و بازخورد خبرگان ارزیابی شد (۱۸، ۱۹). این توالی مرحله‌ای موجب شد که توسعه ربات گفتگوی هاوانا صرفاً یک پیاده سازی فنی نباشد، بلکه نظام مند، کاربر محور و متکی بر استانداردهای معتبر شکل گیرد. مراحل پژوهش در ادامه ذکر شده است.

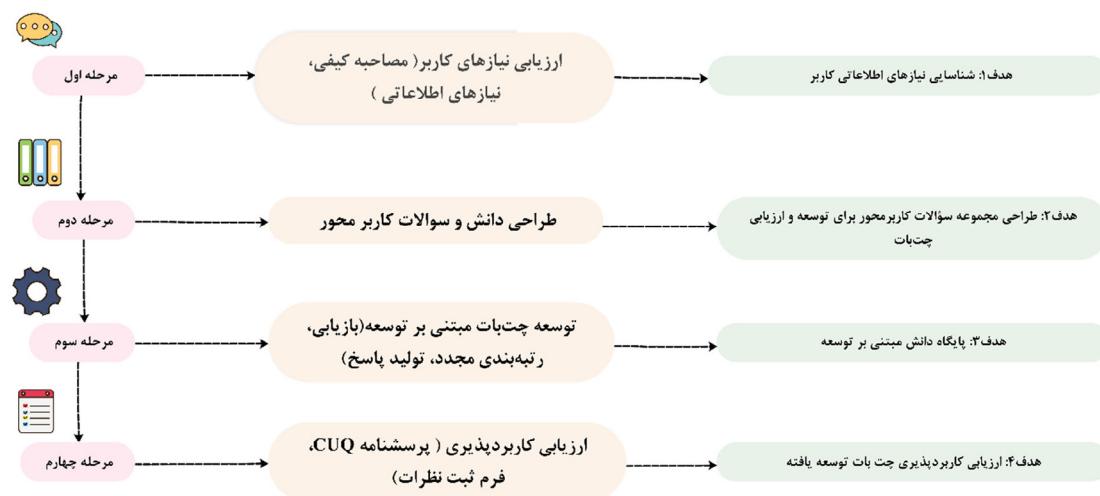
داده اسناد NIOSH تولید می‌شود (۱۲). این رویکرد، قابلیت اعتماد سامانه را برای کاربردهای حساس بهداشتی و ایمنی به‌طور معناداری افزایش می‌دهد.

بر این اساس، مطالعه حاضر با هدف طراحی، پیاده‌سازی و ارزیابی سامانه هوشمند «هاوانا» (HAVANA) بر پایه معماری RAG انجام شد تا به‌عنوان یک دستیار تصمیم‌یار متخصص، کارشناسان را در مسیر انتخاب و اجرای دقیق روش‌های نمونه‌برداری VOCs راهنمایی کند. این پژوهش برای دستیابی به هدف نهایی، چهار هدف عملیاتی و گام‌به‌گام زیر را دنبال کرده است:

۱. تحلیل چالش‌ها و الگوهای جستجوی متخصصین بهداشت حرفه‌ای در مواجهه با روش‌های استاندارد از طریق مطالعه کیفی و مصاحبه‌های نیمه‌ساختاریافته.
۲. تدوین مجموعه‌ای از پرسش‌های کلیدی و کاربرمحور برای پوشش کامل مراحل انتخاب متد، عملیات میدانی و تحلیل آزمایشگاهی.
۳. پیاده‌سازی معماری RAG شامل مراحل پیش‌پردازش، چابک‌بندی اسناد NIOSH و ایجاد پایگاه داده برداری برای تولید پاسخ‌های مستند.
۴. سنجش اثربخشی و تجربه کاربری سامانه توسعه‌یافته از دیدگاه متخصصان حوزه با استفاده از پرسشنامه استاندارد کاربردپذیری چت‌بات (CUQ) و تحلیل کیفی بازخوردها.

روش کار

این پژوهش با هدف طراحی، توسعه و ارزیابی یک



شکل ۲: هم‌راستا سازی مراحل روش کار پژوهش با اهداف اصلی مطالعه

رسیدن به اشباع داده‌ها^۱ ادامه یافت. به این معنا که پس از انجام ۱۲ مصاحبه، مضمون یا کد جدیدی از داده‌ها استخراج نشد و بنابراین این تعداد برای پوشش نیازهای اطلاعاتی کاربران کافی تشخیص داده شد (۲۰). هدف اصلی، دستیابی به دیدگاه‌های چندسطحی از اجرایی‌ها (مثل کارشناسان صنعت)، تصمیم‌گیرندگان (مدیران)، نظارتی‌ها (مسئولان فنی و معاونت‌های بهداشتی) و نظری‌ها (هیئت علمی) بود. این توزیع به صورت غیر تصادفی و هدفمند انجام شد تا از یک سو، اشباع اطلاعاتی حاصل شود و از سوی دیگر، پوشش کاملی از نیازهای کاربردی در محیط‌های واقعی (صنعتی، آموزشی، نظارتی) فراهم گردد. مدت زمان هر مصاحبه بین ۳۰ تا ۴۵ دقیقه بود. برای این منظور، یک راهنمای مصاحبه از پیش طراحی شده مورد استفاده قرار گرفت که شامل پرسش‌های باز درباره نیازهای اطلاعاتی، چالش‌های عملی در استفاده از روش‌های استاندارد NIOSH، ابهام‌های رایج در فرآیند نمونه‌برداری و آنالیز و انتظارات کاربران از یک ربات گفتگو راهنما بود (۲۱). ساختار باز سوالات امکان بیان آزادانه تجربه‌ها و دیدگاه‌های مشارکت‌کنندگان را فراهم می‌کرد. تمامی مصاحبه‌ها با رضایت آگاهانه مشارکت‌کنندگان ضبط صوتی شدند تا دقت و جامعیت داده‌ها حفظ شود.

1. Data saturation

۱. مرحله اول: شناسایی نیازهای اطلاعاتی کاربران

مرحله اول این پژوهش با رویکرد کیفی اکتشافی و با هدف شناسایی نظام‌مند نیازهای اطلاعاتی، چالش‌ها و انتظارات کاربران نهایی در زمینه نمونه‌برداری و آنالیز ترکیبات آلی فرار بر اساس روش‌های استاندارد NIOSH انجام شد. این مرحله به عنوان مبنای مفهومی طراحی ساختار تعاملی ربات گفتگو و تعیین الزامات اطلاعاتی ربات گفتگو در نظر گرفته شد.

مشارکت‌کنندگان این مرحله شامل ۱۲ متخصص بهداشت حرفه‌ای با پیشینه‌های متنوع حرفه‌ای بودند که تجربه عملی یا آموزشی آن‌ها در حوزه ارزیابی مواجهه شغلی و اجرای روش‌های نمونه‌برداری و آنالیز NIOSH تأیید شده بود. این گروه شامل کارشناسان صنعت، اعضای هیئت علمی، مدیران صنعتی، مسئولان فنی، کارشناسان معاونت‌های بهداشتی/بهداشت کار و مدیران شرکت‌های خدمات بهداشت حرفه‌ای بود. انتخاب افراد به صورت نمونه‌گیری هدفمند و معیارهای ورود به مطالعه شامل حداقل سه سال سابقه فعالیت در حوزه بهداشت حرفه‌ای یا آنالیز ترکیبات شیمیایی، آشنایی عملی با روش‌های استاندارد NIOSH و تجربه مشارکت در فرآیند نمونه‌برداری یا تحلیل VOCs بود. فرآیند نمونه‌گیری تا

پس از پایان مصاحبه‌ها، فایل‌های صوتی به صورت کامل به متن تبدیل^۱ شدند. فرآیند پیاده سازی با دقت انجام گرفت تا محتوای معنایی گفتار، تاکیدها و شرح تجربیات مشارکت‌کنندگان تا حد امکان بدون حذف یا ساده سازی ثبت شود (۲۲). متون حاصل، داده‌های خام کیفی این مرحله را تشکیل دادند.

تحلیل داده‌ها با استفاده از تحلیل محتوای کیفی با سازمان‌دهی تماتیک انجام شد (۲۳). در گام نخست، متون پیاده‌سازی شده مصاحبه‌ها وارد مرحله تحلیل اولیه با بهره‌گیری از ابزارهای مبتنی بر مدل‌های زبانی بزرگ برای تسهیل مرور داده‌ها (۲۴، ۲۵). این ابزارها به منظور تسهیل مرور داده‌ها، شناسایی الگوهای پرتکرار، استخراج گزاره‌های معنادار و پیشنهاد کدهای اولیه مورد استفاده قرار گرفتند. نقش این ابزارها صرفاً کمکی و پشتیبان بوده و فرآیند تحلیل نهایی به طور کامل تحت کنترل پژوهشگر انجام شد. در مرحله بعد، پژوهشگر با بازبینی دقیق تمامی متون و خروجی‌های پیشنهادی ابزارهای هوش مصنوعی، فرآیند کدگذاری دستی را انجام داد. کدهای مشابه در قالب خوشه‌های مفهومی گروه‌بندی شدند (۲۶) و این خوشه‌ها طی یک فرآیند تکرار شونده مقایسه، ادغام و پالایش، به مضامین اصلی تبدیل شدند. استخراج مضامین تا دستیابی به انسجام مفهومی، پوشش مناسب داده‌ها و اشباع تحلیلی ادامه یافت.

مضامین نهایی استخراج‌شده به‌عنوان نمایانگر نیازهای اطلاعاتی کلیدی کاربران نهایی، مبنای طراحی مراحل بعدی پژوهش، شامل تدوین ساختار سوالات کاربر محور، طراحی منطق تعامل ربات گفتگو و توسعه معماری ربات گفتگو قرار گرفتند. نتایج حاصل از این مرحله در بخش یافته‌ها به صورت تفکیک شده گزارش شده‌اند.

۲. مرحله دوم: طراحی مجموعه سوالات کاربر محور

مرحله دوم پژوهش با تمرکز بر طراحی نظام مند مجموعه‌ای از سوالات کاربر محور انجام شد. هدف این مرحله، تبدیل نیازهای اطلاعاتی شناسایی شده در

مرحله نخست به الگوهای پرسش‌پاسخ ساختارمند بود. الگوهایی که منطق تعامل کاربران با ربات گفتگو و چارچوب پاسخ‌دهی را مشخص می‌کنند. در این مرحله، سوالات به عنوان هسته مفهومی طراحی ربات گفتگو در نظر گرفته شدند و نقش آن‌ها تعیین‌کننده‌ی نوع اطلاعات موردنیاز، نحوه سازمان‌دهی دانش و مسیرهای گفتگویی ربات بود (۲۷). مبنای طراحی سوالات، داده‌های حاصل از مصاحبه‌های نیمه ساختاریافته با خبرگان در مرحله نخست پژوهش بود. این داده‌ها به عنوان ورودی طراحی مورد استفاده قرار گرفتند تا مجموعه سوالات، بازتاب دهنده‌ی سناریوهای واقعی استفاده از روش‌های استاندارد NIOSH در محیط‌های کاری باشد (۲۸). بر همین اساس، فرآیند طراحی به گونه‌ای انجام شد که سوالات از نظر زبانی به ادبیات رایج کاربران نزدیک بوده و از منظر مفهومی، با منطق کاربردی روش‌های استاندارد هم راستا باشند.

در فرآیند طراحی، مجموعه‌ای از معیارها برای تدوین و پالایش سوالات در نظر گرفته شد، از جمله:

- استفاده از زبان طبیعی متناسب با کاربران هدف؛
- بازنمایی موقعیت‌های واقعی تصمیم‌گیری در فرآیند نمونه‌برداری و آنالیز؛
- قابلیت پشتیبانی پاسخ‌ها بر مبنای محتوای مستند روش‌های استاندارد NIOSH.

به‌منظور سامان‌دهی منطق پاسخ‌دهی ربات، سوالات طراحی شده در قالب تیپ‌های اطلاعاتی مفهومی دسته بندی شدند. این تیپ‌ها چارچوبی انتزاعی برای تعیین نوع پاسخ مورد انتظار (مانند توضیح مفهومی، راهنمای اجرایی یا محاسباتی) فراهم کردند و به طراحی مسیرهای تعاملی ربات کمک نمودند، بدون آن‌که وارد جزئیات پیاده سازی فنی شوند (۲۹). در ادامه، بر اساس ساختار سوالات و تیپ‌های اطلاعاتی تعریف شده، الزامات دانشی ربات گفتگو مشخص گردید. به عبارت دیگر، تعیین شد که برای پاسخ‌گویی به هر گروه از سوالات، چه نوع محتوایی از نظر مفهومی مورد نیاز است (نظیر توصیف متنی، اطلاعات جدولی یا روابط محاسباتی). این گام نقش واسط میان

1. Transcription

توانایی بالایی در استخراج شباهت معنایی دقیق بین پرسش و اسناد فنی دارد. انتخاب مدل بر اساس ارزیابی اولیه عملکرد آن در متون تخصصی حوزه سلامت و بهداشت حرفه‌ای انجام گرفت (۳۴). بردارهای معنایی تولید شده، به همراه متادیتاها، در یک پایگاه داده برداری^۳ مبتنی بر کیودرنت^۴ ذخیره شدند. انتخاب این پایگاه داده به دلیل پشتیبانی از جستجوی برداری، امکان فیلترگذاری پیشرفته بر اساس متادیتا و قابلیت مقیاس پذیری انجام شد. این ساختار، بازیابی ترکیبی مبتنی بر شباهت معنایی و محدودیت‌های مفهومی (نظیر نوع روش یا بخش سند) را ممکن می‌سازد. خط لوله نهایی RAG شامل چند مرحله متوالی است. در گام نخست، پرسش کاربر تحلیل می‌شود تا مشخص گردد ماهیت آن عمومی است یا به اطلاعات فنی و استاندارد وابسته است. شکل ۳ ماتریس تصمیم‌گیری برای هدایت پرسش‌های کاربران به ماژول‌های پاسخ‌دهی (RAG)، LLM و Hybrid در معماری ربات گفت‌وگوی HAVANA.

در پرسش‌های تخصصی مرتبط با روش‌های NIOSH، به منظور افزایش دقت بازیابی (با توجه به انگلیسی بودن اسناد مرجع)، پرسش به زبان انگلیسی بازنویسی می‌شود (۳۵). سپس پرسش به بردار معنایی تبدیل شده و جستجوی اولیه در پایگاه داده برداری انجام می‌گیرد. چانک‌های بازیابی شده در مرحله بعد با استفاده از یک سازوکار رتبه‌بندی مجدد مبتنی بر Cross-Encoder پالایش می‌شوند تا مرتبط‌ترین واحدهای اطلاعات انتخاب شوند (۳۶). علاوه بر این، در صورت وجود ارجاعات درون متنی مرتبط (مانند جداول یا توضیحات تکمیلی)، چانک‌های وابسته نیز به مجموعه نهایی افزوده می‌شوند. در گام نهایی، مجموعه چانک‌های منتخب به عنوان زمینه^۵ به یک مدل زبانی بزرگ در سطح GPT-4O ارسال می‌شود تا پاسخ نهایی تولید گردد (۳۷). این مدل به دلیل توانایی بالا در درک متون تخصصی، تولید پاسخ ساختارمند و عملکرد مناسب در وظایف پرسش و پاسخ علمی انتخاب شد. تولید

طراحی مفهومی سوالات و آماده سازی فنی منابع دانشی را ایفا کرد و از پراکندگی رویکرد در مراحل بعد جلوگیری نمود. در این مرحله، تمرکز پژوهش صرفاً بر تعریف منطق دانشی و طراحی مفهومی سوالات کاربرمحور بوده است. جزئیات فنی مربوط به استخراج اسناد، پیش پردازش، ساختاردهی داده‌ها و چانک‌بندی منابع دانشی به صورت مستقل در مرحله سوم پژوهش تشریح شده‌اند (۳۰).

۳. مرحله سوم: طراحی و توسعه ربات گفتگوی هوشمند مبتنی بر معماری بازیابی تقویت شده با تولید (RAG)

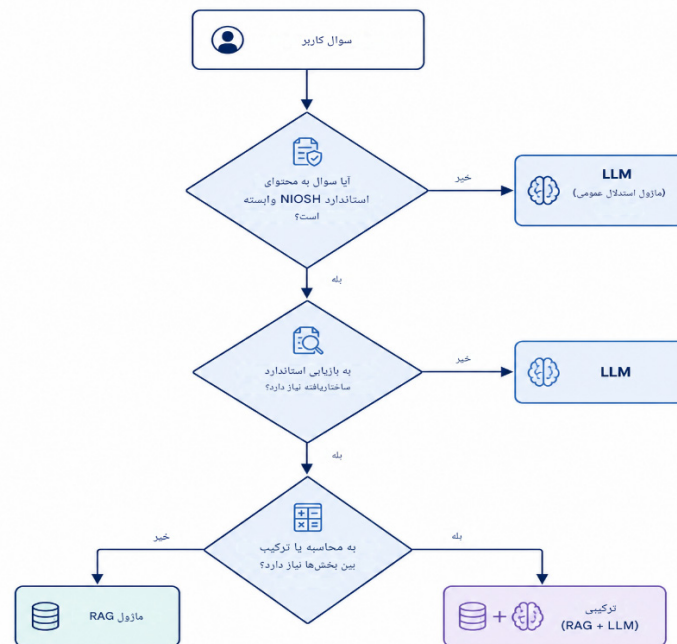
در این مرحله، ربات گفتگوی پژوهش با استفاده از معماری بازیابی تقویت شده با تولید (RAG) طراحی و توسعه یافت. هدف از به کارگیری این معماری، تضمین پاسخ دهی مستند، دقیق و منطبق با مفاد روش‌های استاندارد NIOSH و در عین حال کاهش وابستگی مدل زبانی به حافظه درونی و کاهش خطر تولید پاسخ‌های غیرمستند بود (۳۱، ۳۲).

اسناد رسمی روش‌های استاندارد NIOSH، که عموماً به صورت فایل‌های PDF و با ساختاری ثابت منتشر می‌شوند، به عنوان هسته پایگاه دانش ربات گفتگو مورد استفاده قرار گرفتند. این اسناد شامل بخش‌های استاندارد نظیر مقدمه، دامنه کاربرد، تجهیزات، روش نمونه‌برداری، روش‌های آنالیز آزمایشگاهی، جداول عملکردی و توضیحات محاسباتی هستند (۴). به منظور تسهیل بازیابی معنایی دقیق، هر سند پس از پیش پردازش به واحدهای معنایی مستقل^۱ تقسیم شد. در این فرآیند، متن اصلی، جداول و توضیحات همراه آن‌ها به گونه‌ای سازمان‌دهی شدند که هر چانک بتواند به صورت مستقل بازیابی و در تولید پاسخ مورد استفاده قرار گیرد. برای هر چانک، متادیتاهایی از جمله شماره روش NIOSH، نوع محتوا، و بخش مربوطه ذخیره شد تا امکان فیلترگذاری هدفمند فراهم شود (۳۳).

برای بازنمایی معنایی چانک‌ها، از یک مدل امبدینگ^۲ آموزش دیده بر متون علمی انگلیسی استفاده شد که

1. Chunks
2. Embedding

3. Vector database
4. Qdrant
5. Context



شکل ۳: چارچوب هدایت پرسش‌ها به ماژول‌های پاسخ‌دهی در معماری ربات گفت‌وگوی HAVANA

در این پژوهش، یک پاسخ زمانی «صحیح» تلقی شد که:

۱. بخش بازایی‌شده توسط سامانه با متن مرجع موجود در Golden Set انطباق مفهومی داشته باشد؛
۲. پاسخ تولیدشده از نظر محتوای فنی، پارامترهای کمی، و دستورالعمل‌های اجرایی با متن اصلی NIOSH سازگار باشد؛
۳. هیچ‌گونه تحریف، حذف بحرانی یا افزودن اطلاعات غیرمستند در پاسخ مشاهده نشود.

اعتبار مجموعه طلایی از طریق بازبینی و اجماع خبرگان تضمین شد تا اطمینان حاصل شود که پاسخ‌های مرجع دقیقاً بازتاب‌دهنده راهنمای رسمی NIOSH هستند. این رویکرد امکان ایجاد یک معیار ساختاریافته و قابل اتکا برای سنجش عملکرد ماژول بازایی و تولید پاسخ را فراهم کرد.

به منظور تسهیل تعامل کاربر با ربات و طرح پرسش‌های تخصصی، یک رابط گفت‌وگوی متنی برای ربات HAVANA (هاوانا) طراحی شد که نمای کلی آن به همراه صفحه‌ای از چت در شکل ۴ نشان داده شده است.

پاسخ به‌گونه‌ای محدود شده است که مدل صرفاً بر اساس اطلاعات بازایی‌شده عمل کند و از ایجاد محتوای خارج از پایگاه دانش اجتناب شود. پاسخ نهایی به زبان فارسی و با ساختاری علمی، شفاف و مرحله‌به‌مرحله ارائه می‌شود. به‌منظور ارزیابی دقت بازایی و صحت پاسخ‌های تولیدشده توسط ربات HAVANA، یک مجموعه مرجع استاندارد مبتنی بر اسناد رسمی NIOSH طراحی شد. در گام نخست، مجموعه‌ای از روش‌های نمونه‌برداری و آنالیز منتخب NIOSH که مرتبط با اهداف پژوهش بودند گردآوری شد. این اسناد بر اساس ساختار درونی خود (مانند دامنه کاربرد، تجهیزات موردنیاز، شرایط نمونه‌برداری، مراحل آنالیز، کالیبراسیون و کنترل کیفیت) به بخش‌های مفهومی منسجم تقسیم شدند.

در گام دوم، برای هر سؤال ارزیابی از پیش تعریف‌شده، خبرگان حوزه بهداشت حرفه‌ای بخش یا بخش‌های دقیق متن NIOSH که حاوی پاسخ فنی معتبر بودند را شناسایی و استخراج کردند. این قطعات استخراج شده به عنوان «پاسخ مرجع» آن سؤال در مجموعه طلایی ثبت شدند.



شکل ۴: نمای کلی رابط کاربری ربات گفت‌وگوی HAVANA و نمونه‌ای از تعامل کاربر با ربات برای دریافت راهنمایی در زمینه انتخاب و اجرای روش‌های نمونه‌برداری و آنالیز ترکیبات آلی فرار بر اساس استانداردهای NIOSH.

- وضوح و قابل فهم بودن پاسخ‌ها،
- میزان مفید بودن چت‌بات در انجام وظایف تخصصی
- کیفیت تعامل و رضایت کلی کاربر را پوشش می‌دهد. پاسخ‌ها بر اساس مقیاس لیکرت پنج‌درجه‌ای ثبت شدند.

نسخه مورد استفاده در این پژوهش، نسخه فارسی روان سنجی شده پرسشنامه CUQ بود که روایی و پایایی آن در مطالعات پیشین تایید شده است (۳۹). تکمیل پرسشنامه پس از یک دوره تعامل عملی کاربران با چت‌بات هاوانا انجام گرفت تا ارزیابی کاربردپذیری مبتنی بر تجربه واقعی استفاده از ربات گفتگو صورت گیرد.

۴.۲. ارزیابی کیفی با فرم ثبت نظرات

در کنار ارزیابی کمی، از یک فرم ثبت نظرات باز برای دریافت دیدگاه‌های تفصیلی کاربران و خبرگان حوزه بهداشت حرفه‌ای استفاده شد. این فرم با هدف تکمیل نتایج کمی پرسشنامه CUQ طراحی شد و امکان بیان آزادانه نظرات درباره:

- نقاط قوت و محدودیت‌های ربات گفتگو،
- دقت علمی پاسخ‌ها،
- میزان انطباق محتوای ارائه شده با روش‌های استاندارد NIOSH،
- و کیفیت راهنمایی چت‌بات در سناریوهای واقعی کاری را فراهم می‌کرد.

۴. مرحله چهارم: ارزیابی کاربردپذیری و بازخورد کاربران

در مرحله چهارم پژوهش، ارزیابی ربات گفتگو توسعه یافته با تمرکز بر کاربردپذیری، کیفیت تعامل و میزان رضایت کاربران از چت‌بات HAVANA انجام شد. ارزیابی کاربردپذیری با مشارکت ۱۲ نفر از متخصصان بهداشت حرفه‌ای که تجربه کار با ربات را داشتند انجام شد. هدف این مرحله، بررسی میزان سهولت استفاده، وضوح و مفید بودن پاسخ‌های ارائه شده و ارزیابی تجربه کلی کاربران در تعامل با ربات گفتگو بود. به منظور دستیابی به ارزیابی جامع و چند بعدی، از دو ابزار مکمل کمی و کیفی استفاده شد که امکان تحلیل هم زمان شاخص‌های کاربردپذیری و برداشت‌های تفسیری کاربران را فراهم می‌کرد (۲۸).

۴.۱. ارزیابی کمی کاربردپذیری با پرسشنامه CUQ

برای سنجش کمی کاربردپذیری چت‌بات، از پرسشنامه روانسنجی شده^۱ CUQ استفاده شد. این ابزار یکی از پرسشنامه‌های متداول در ارزیابی تجربه کاربری چت‌بات‌هاست و ابعاد مختلف تعامل انسان-چت‌بات را از دیدگاه کاربران مورد بررسی قرار می‌دهد (۳۸). پرسشنامه CUQ شامل مجموعه‌ای از گویه‌هاست که حوزه‌هایی مانند:

- سهولت یادگیری و استفاده از ربات گفتگو،

1. Chatbot usability questionnaire

جدول ۱: مشخصات فردی و حرفه‌ای مشارکت‌کنندگان پژوهش (n=۱۲)

ردیف	متغیر	طبقه بندی	تعداد	درصد
۱	جنسیت	زن	۵	۴۲
۲		مرد	۷	۵۸
۳	سطح تحصیلات	کارشناسی	۵	۴۲
۴		کارشناسی ارشد	۴	۳۳
۵		دکتری	۳	۲۵
۶	حوزه فعالیت	مسئول فنی - مدیر شرکت خدمات بهداشت حرفه ای	۵	۴۲
۷		آزمایشگاه آنالیز- کارشناس معاونت	۳	۲۵
۸		هیئت علمی	۲	۱۷
۹		کارشناس و مدیر HSE صنایع	۲	۱۶

متخصصان حوزه بهداشت حرفه‌ای با تنوع مناسب از نظر جنسیت، سطح تحصیلات و حوزه فعالیت حرفه‌ای بود. مشارکت‌کنندگان در حوزه‌های مختلفی از جمله مسئولان فنی سنجش آلاینده‌ها، کارشناسان آزمایشگاه‌های آنالیز، اعضای هیئت علمی دانشگاه‌ها و متخصصان HSE در صنایع فعالیت داشتند. میانگین سابقه کاری آنان ۹ سال (انحراف معیار ۱/۹۵) و دامنه سابقه کاری بین ۶ تا ۱۲ سال گزارش شد که نشان‌دهنده برخورداری پنل از تجربه حرفه‌ای مرتبط با موضوع پژوهش است.

یافته‌های مرتبط با هدف دوم: طراحی سوالات کاربرمحور
به‌منظور شناسایی نیازهای اطلاعاتی کاربران و طراحی سوالات اصلی ربات گفتگوی هاوانا، داده‌های حاصل از مصاحبه‌های نیمه‌ساختاریافته با خبرگان تحلیل و سوالات مطرح‌شده در قالب تیپ‌های محتوایی مجزا طبقه‌بندی شدند. هر تیپ بیانگر نوعی از تعامل کاربر با سیستم در فرآیند ارزیابی مواجهه شغلی با ترکیبات آلی فرار است.

مطابق نتایج ارائه شده در جدول ۲، مجموعاً ده تیپ اصلی از سوالات کاربرمحور شناسایی شد که طیف متنوعی از نیازهای اطلاعاتی در حوزه ارزیابی مواجهه شغلی با VOCs را پوشش می‌دهد. بیشترین فراوانی مربوط به سوالات مرتبط با خواص ترکیبات شیمیایی، ارزیابی و مدیریت ریسک مواجهه و انتخاب روش‌های

استفاده از این ابزار کیفی، امکان شناسایی جنبه‌هایی از تجربه کاربری را فراهم ساخت که صرفاً از طریق شاخص‌های عددی قابل استخراج نبودند و به تحلیل عمیق‌تر عملکرد ربات کمک می‌کرد. در نهایت، داده‌های حاصل از پرسشنامه CUQ و فرم ثبت نظرات کاربران و خبرگان، مبنای تحلیل نتایج کمی و کیفی در بخش یافته‌های پژوهش قرار گرفتند و یافته‌های هر یک به صورت تفکیک شده گزارش شده‌اند.

یافته‌ها

یافته‌های این پژوهش بر اساس چهار هدف اصلی مطالعه سازمان‌دهی شده‌اند. نتایج هر بخش در قالب جداول مربوطه ارائه شده و مهم‌ترین الگوهای مشاهده‌شده به‌صورت توصیفی گزارش شده‌اند.

یافته‌های مرتبط با هدف اول: ویژگی‌های پنل خبرگان
به‌منظور توصیف ترکیب پنل خبرگان شرکت‌کننده در مراحل طراحی، توسعه و ارزیابی ربات گفتگوی هاوانا، اطلاعات مربوط به مشخصات فردی، سطح تحصیلات، حوزه فعالیت حرفه‌ای و سابقه کاری جمع‌آوری و به‌صورت توصیفی تحلیل شد. این اطلاعات به‌عنوان زمینه‌ای برای درک اعتبار نظرات و داده‌های حاصل از مصاحبه‌ها و ارزیابی‌ها مورد استفاده قرار گرفت.
مطابق جدول ۱، پنل خبرگان مطالعه شامل ۱۲ نفر از

جدول ۲: تیپ‌های سوالات کاربرمحور شناسایی شده و فراوانی اشاره خبرگان

ردیف	تیپ سوالات	ماهیت سوالات طراحی شده	تعداد تکرار در مصاحبه	مکانیزم تولید پاسخ
۱	خواص و مشخصات ترکیبات شیمیایی	شناسایی، ویژگی فیزیکی و کاربرد ترکیبات VOCs	۹	LLM
۲	ارزیابی و مدیریت ریسک مواجهه	حدود مجاز تماس، شاخص‌های خطر و تصمیم‌گیری ایمنی	۸	LLM & RAG
۳	انتخاب و پروتکل‌های نمونه‌برداری	انواع روش‌ها، انتخاب لوله جذب و شرایط کاری	۸	RAG
۴	جزئیات فنی روش‌های استاندارد نمونه‌برداری	نکات اجرایی دقیق هر روش	۷	RAG
۵	فرآیندهای آنالیز آزمایشگاهی	مراحل آماده‌سازی نمونه و آنالیز در GC/MS یا GC-FID	۷	RAG
۶	جزئیات فنی روش‌های استاندارد آنالیز	پارامترهای تنظیم دستگاه، نوع ستون، دما و آشکارساز	۶	LLM
۷	جنبه‌های سم‌شناختی ترکیبات	علائم بالینی، مکانیسم سمیت و اثرات بهداشتی	۶	RAG
۸	کالیبراسیون و تضمین کیفیت	کنترل صحت دستگاه‌ها، کالیبراسیون، QA/QC نتایج	۵	RAG
۹	محاسبات مواجهه و دوزیمتری	فرمول‌های TWA، ضرایب تصحیح، محاسبات غلظت	۵	LLM
۱۰	مسائل متفرقه و حل مشکل	پرسش‌های عمومی و راهکارهای اولیه رفع اشکال	۴	LLM

LLM: : Large Language Model
RAG: Retrieval-Augmented Generation
VOCs: Volatile Organic Compound

زبانی (LLM) پاسخ داده می‌شوند. این طبقه‌بندی مبنای طراحی منطق پاسخ‌دهی در معماری چت‌بات قرار گرفت و امکان هدایت هدفمند سوالات به مناسب‌ترین ماژول پردازشی را فراهم کرد.

یافته‌های مرتبط با هدف سوم: توسعه و ارزیابی عملکرد سیستم RAG

مقایسه روش‌های چانک‌بندی اسناد - برای بررسی تأثیر روش چانک‌بندی بر عملکرد سیستم RAG، دو رویکرد «چانک‌بندی با اندازه ثابت» و «چانک‌بندی مبتنی بر ساختار اسناد NIOSH» بر روی ۲۰۰ سؤال فنی مقایسه شدند (جدول ۳). نتایج نشان داد استفاده از چانک‌بندی مبتنی بر ساختار سند موجب افزایش چشمگیر تعداد پاسخ‌های صحیح از ۷۸ به ۱۴۶ مورد شد و هم‌زمان تعداد پاسخ‌های نادرست به‌طور قابل توجهی کاهش یافت. این یافته بیانگر نقش کلیدی

نمونه‌برداری بود. تحلیل ماهیت سوالات نشان داد که نیاز اطلاعاتی کاربران از نظر وابستگی به اسناد استاندارد و سطح تفسیر موردنیاز متفاوت است. بر این اساس، سوالات در سه مسیر پاسخ‌دهی اصلی طبقه‌بندی شدند: ۱. سوالات مبتنی بر اطلاعات استاندارد: شامل پرسش‌هایی که نیازمند دسترسی دقیق به مقادیر استاندارد، مراحل اجرایی و جزئیات روش‌های رسمی NIOSH هستند. این دسته از طریق ماژول بازیابی مبتنی بر اسناد (RAG) پاسخ داده می‌شوند. ۲. سوالات ترکیبی (LLM + RAG): شامل پرسش‌هایی که علاوه بر بازیابی اطلاعات استاندارد، نیازمند تفسیر یا ترکیب مفاهیم هستند؛ مانند برخی سوالات مرتبط با ارزیابی ریسک و تصمیم‌گیری ایمنی. ۳. سوالات عمومی و مفهومی: شامل پرسش‌هایی درباره ویژگی‌ها یا کاربردهای کلی ترکیبات که وابستگی مستقیمی به مقادیر استاندارد ندارند و عمدتاً توسط مدل

جدول ۳: مقایسه عملکرد روش‌های چانک‌بندی در پاسخ‌دهی به سؤالات فنی (n=۲۰۰)

ردیف	نوع چانک‌بندی	تعداد سؤالات آزمون	پاسخ صحیح	پاسخ نادرست
۱	چانک‌بندی با اندازه ثابت	۲۰۰	۷۸	۱۲۲
۲	چانک‌بندی مبتنی بر ساختار سند	۲۰۰	۱۴۶	۵۴

جدول ۴: مقایسه عملکرد سیستم در حالت با ریرنکر و بدون ریرنکر (n=۵۰)

ردیف	وضعیت ربات	تعداد سؤالات آزمون	پاسخ صحیح
۱	بدون استفاده ریرنکر	۵۰	۳۱
۲	با استفاده از ریرنکر	۵۰	۴۶

جدول ۵: شاخص‌های عملکرد بازیابی برای مقادیر مختلف k (n=۱۰۰)

ردیف	مقدار k	اسناد مرتبط (TP)	اسناد غیرمرتبط (FP)	اسناد مرتبط بازیابی نشده (FN)	دقت	فراخوانی
۱	k=۵	۲۴۸	۲۵۲	۱۶۸	۴۹/۶٪	۵۹/۶٪
۲	k=۱۰	۳۷۲	۶۲۸	۴۴	۳۷/۲٪	۸۹/۴٪
۳	k=۱۵	۳۹۶	۱۱۰۴	۲۰	۲۶/۴٪	۹۵/۳٪

نظر دقت و فراخوانی بررسی شد. مطابق جدول ۵، عملکرد سیستم برای مقادیر مختلف پارامتر k (تعداد چانک‌های بازیابی‌شده) از نظر شاخص‌های دقت (Precision) و فراخوانی (Recall) ارزیابی شد. نتایج نشان داد با افزایش مقدار k، میزان فراخوانی افزایش یافته اما دقت کاهش می‌یابد که بیانگر ایجاد مبادله میان جامعیت و دقت بازیابی است. در میان مقادیر بررسی‌شده، مقدار k=10 بهترین توازن میان این دو شاخص را فراهم کرد و به‌عنوان مقدار بهینه در پیکربندی نهایی سیستم انتخاب شد.

یافته‌های مرتبط با هدف چهارم: ارزیابی کاربردپذیری ربات گفتگو

به منظور ارزیابی تجربه کاربری و میزان پذیرش ربات گفتگوی هاوانا، از پرسشنامه استاندارد CUQ و فرم ثبت نظرات خبرگان استفاده شد. نتایج به صورت کمی و کیفی گزارش می‌شود. ارزیابی کاربردپذیری ربات با استفاده از پرسشنامه استاندارد CUQ انجام شد که عمدتاً ابعاد سهولت استفاده، وضوح تعامل و پذیرش کاربر را پوشش می‌دهد. بنابراین، نتایج این بخش صرفاً منعکس‌کننده

حفظ انسجام ساختاری اسناد استاندارد در بهبود دقت بازیابی و تولید پاسخ است. همان‌طور که در جدول ۳ مشاهده می‌شود، استفاده از چانک‌بندی مبتنی بر ساختار سند موجب افزایش قابل توجه تعداد پاسخ‌های صحیح و کاهش پاسخ‌های نادرست شده است.

- اثر استفاده از ریرنکر بر دقت بازیابی

به‌منظور ارزیابی نقش ریرنکر در انتخاب اسناد مرتبط‌تر، عملکرد سیستم در دو حالت (با ریرنکر) و (بدون ریرنکر) بر روی ۵۰ سوال آزمون مقایسه شد (جدول ۴). استفاده از ریرنکر موجب افزایش تعداد پاسخ‌های صحیح از ۳۱ به ۴۶ مورد شد که نشان‌دهنده بهبود معنی‌دار در کیفیت انتخاب قطعات متنی بازیابی‌شده است. بر اساس نتایج جدول ۴، تعداد پاسخ‌های صحیح در حالت استفاده از ریرنکر افزایش یافته است.

- تعیین مقدار بهینه پارامتر k

برای تعیین مقدار مناسب پارامتر k (تعداد چانک‌های بازیابی‌شده)، عملکرد سیستم برای مقادیر مختلف k از

جدول ۶: میانگین و انحراف معیار مؤلفه‌های کاربردپذیری بر اساس CUQ (n=۱۲)

ردیف	مؤلفه کاربردپذیری	میانگین امتیاز	انحراف معیار
۱	وضوح پاسخ	۴/۰۹	۰/۸۳
۲	تناسب محتوای ارائه شده با نیازهای کاربر	۳/۷۳	۱/۱۰
۳	سرعت پاسخ دهی	۳/۸۲	۱/۰۸
۴	سهولت تعامل	۳/۹۱	۰/۵۴
۵	استمرار مکالمه	۳/۶۴	۱/۲۱
۶	تجربه کلی کاربر	۳/۷۳	۱/۲۷
۷	قابلیت درک زبان کاربر	۳/۹۱	۱/۰۴
۸	تناسب لحن و سبک ارتباطی	۴/۰۹	۰/۹۴
۹	پاسخ گویی به نیازهای خاص و تخصصی	۳/۹۱	۱/۰۴
۱۰	قابلیت یادگیری و بهبود در طول تعامل	۲/۲۷	۱/۴۹

مراحل پیشنهادی نمونه‌برداری و قابلیت استفاده عملی در شرایط واقعی کار در سطح قابل قبولی قرار دارند. بر اساس بازخوردها، ربات به‌ویژه در انتخاب اولیه روش‌های مناسب و تنظیم پارامترهای نمونه‌برداری عملکرد منسجم و قابل اعتمادی ارائه کرد.

بحث

هدف این پژوهش ارائه و ارزیابی یک سیستم مکالمه‌ای مبتنی بر معماری RAG برای پشتیبانی از انتخاب و تفسیر روش‌های نمونه‌برداری VOCs مطابق استانداردهای NIOSH بود. تحلیل یافته‌ها نشان می‌دهد که ترکیب طراحی کاربرمحور، پردازش ساختاریافته اسناد مقرراتی و بهینه‌سازی چندلایه بازایی می‌تواند ظرفیت مدل‌های زبانی را در حوزه‌های تخصصی و پرریسک به‌طور معناداری ارتقا دهد. در ادامه، یافته‌ها از چهار منظر نظری، فنی، کاربردی و سیاستی تحلیل می‌شوند.

بحث پیرامون هدف اول: ترکیب و اعتبار پنل خبرگان

یافته‌های مربوط به هدف نخست نشان داد که پنل خبرگان پژوهش از تنوع حرفه‌ای، تحصیلی و تجربی مناسبی برخوردار بوده است. حضور هم‌زمان متخصصان دانشگاهی، کارشناسان بهداشت حرفه‌ای و مسئولان فنی آزمایشگاه‌ها سبب شد فرآیند نیازسنجی و ارزیابی سامانه

تجربه کاربری ربات بوده و به‌تنهایی بیانگر کفایت فنی تصمیم‌ها نیست.

- نتایج کمی پرسشنامه CUQ

طبق داده‌های جدول ۶، میانگین کلی امتیاز کاربردپذیری سیستم ۳/۸۸ از ۵ بود که نشان‌دهنده پذیرش مناسب ربات از سوی کاربران تخصصی است. در میان مؤلفه‌ها، وضوح پاسخ و تناسب لحن و سبک ارتباطی بالاترین میانگین امتیاز را کسب کردند، در حالی که قابلیت یادگیری و بهبود در طول تعامل کمترین امتیاز را به خود اختصاص داد. سایر مؤلفه‌ها از جمله سهولت تعامل، سرعت پاسخ‌دهی و پاسخ‌گویی به نیازهای تخصصی در محدوده قابل‌قبول ارزیابی شدند.

- نتایج کیفی فرم ثبت نظرات

تحلیل کیفی نظرات خبرگان نشان داد که مهم‌ترین نقاط قوت ربات شامل دقت علمی پاسخ‌ها، انطباق با استانداردهای NIOSH و کاهش زمان دسترسی به اطلاعات تخصصی بوده است. در مقابل، برخی خبرگان به محدودیت تداوم مکالمه در تعاملات طولانی اشاره کردند. همچنین ارزیابی کیفی عملکرد ربات به‌عنوان یک ابزار پشتیبان تصمیم‌گیری نشان داد که پاسخ‌ها از نظر تطابق فنی با روش‌های استاندارد NIOSH، شفافیت

صرفاً بازتاب‌دهنده یک دیدگاه تخصصی خاص نباشد، بلکه طیفی از سطوح تصمیم‌گیری - از سیاست‌گذاری و تفسیر استانداردها تا اجرای میدانی نمونه‌برداری - را دربر گیرد. میانگین سابقه کاری نزدیک به ۹ سال نیز نشان می‌دهد که اغلب مشارکت‌کنندگان با چالش‌های واقعی مرتبط با انتخاب روش نمونه‌برداری، تفسیر نتایج و ارزیابی مواجهه شغلی مواجه بوده‌اند. چنین تجربه‌ای اعتبار داده‌های کیفی حاصل از مصاحبه‌ها و همچنین قضاوت‌های ارائه‌شده در مرحله ارزیابی کاربردپذیری را تقویت می‌کند (۴۰). از منظر توسعه سامانه‌های مبتنی بر هوش مصنوعی، این ترکیب چند تخصصی اهمیت مضاعفی دارد. در حوزه‌های مقررات محور مانند بهداشت حرفه‌ای، شکاف اصلی نه در کمبود اطلاعات، بلکه در ترجمه عملی استانداردهای فنی به تصمیم‌های اجرایی است. بنابراین، اگر طراحی سامانه صرفاً بر مبنای دیدگاه نظری یا صرفاً بر اساس تجربه اجرایی انجام شود، خطر سوگیری عملکردی یا کاهش قابلیت تعمیم افزایش می‌یابد. مشارکت هم‌زمان خبرگان دانشگاهی و اجرایی در این پژوهش امکان استخراج الزامات عملکردی ربات را به‌گونه‌ای فراهم کرد که هم با مبانی علمی هم‌راستا باشد و هم با محدودیت‌ها و نیازهای محیط‌های کاری واقعی سازگار بماند. این رویکرد با اصول طراحی مشارکتی در سامانه‌های سلامت دیجیتال همسو است.

Stratton و همکاران (۲۰۲۰) نشان دادند که ابزارهای تصمیم‌یار زمانی بیشترین کارایی را دارند که نیازهای متنوع ذینفعان - از کاربران اجرایی تا تصمیم‌گیرندگان سازمانی - در مراحل اولیه طراحی لحاظ شود (۴۱). همچنین Bloch و Sacks (۲۰۲۰) در توسعه سامانه‌های هوشمند پشتیبان ایمنی در صنایع ساختمانی گزارش کردند که مشارکت چندبعدی خبرگان، اعتبار و پذیرش عملی سیستم را به‌طور معناداری افزایش می‌دهد (۲۶). یافته‌های پژوهش حاضر این دیدگاه را در بستر بهداشت حرفه‌ای و استانداردهای NIOSH تأیید می‌کند و نشان می‌دهد که تنوع تخصصی پنل خبرگان می‌تواند به ارتقای هم‌زمان دقت علمی، کارآمدی اجرایی و پذیرش حرفه‌ای

سامانه منجر شود.

با این حال، نتایج ارزیابی کاربردپذیری نشان داد که مؤلفه «یادگیری سامانه و تداوم تعامل» در مقایسه با سایر ابعاد امتیاز پایین‌تری کسب کرده است. این یافته را می‌توان تا حدی ناشی از ماهیت معماری مورد استفاده دانست؛ زیرا در معماری‌های مبتنی بر بازیابی و تولید (RAG) که به‌صورت بدون حافظه طراحی می‌شوند، پاسخ‌دهی عمدتاً بر اساس بازیابی اطلاعات مرتبط از پایگاه دانش انجام می‌شود و سامانه حافظه پایداری از تعاملات قبلی کاربر نگهداری نمی‌کند. در نتیجه، هر پرسش به‌صورت مستقل پردازش می‌شود و امکان یادگیری تدریجی از الگوهای تعامل کاربران محدود می‌ماند. با این حال، این موضوع را می‌توان نه صرفاً یک محدودیت، بلکه فرصتی برای توسعه‌های آتی در نظر گرفت. ادغام معماری‌های بازیابی محور با لایه‌های حافظه محور، مانند حافظه مبتنی بر نشست کاربر، ردگیری زمینه مکالمه و مدل‌های مبتنی بر ویژگی‌های کاربر، می‌تواند در نسخه‌های آینده سامانه به بهبود تداوم تعامل، شخصی‌سازی پاسخ‌ها و یادگیری تدریجی از رفتار کاربران منجر شود. چنین رویکردی به‌ویژه در سامانه‌های پشتیبان تصمیم در حوزه بهداشت حرفه‌ای اهمیت دارد، زیرا کاربران اغلب در قالب زنجیره‌ای از پرسش‌های مرتبط درباره روش‌های نمونه‌برداری، حدود مواجهه و تفسیر نتایج با سامانه تعامل می‌کنند.

در مجموع، نتایج هدف اول بیانگر آن است که اعتبار سامانه توسعه‌یافته صرفاً حاصل بهینه‌سازی فنی معماری RAG نبوده، بلکه ریشه در طراحی مشارکتی و بازتاب‌دهی نیازهای واقعی کاربران نهایی دارد؛ عاملی که در حوزه‌های ایمنی و سلامت شغلی پیش‌نیاز توسعه ابزارهای قابل اعتماد و قابل استفاده محسوب می‌شود (۱۵).

بحث پیرامون هدف دوم: تیپ‌بندی سوالات و طراحی تعامل کاربرمحور

یافته‌های هدف دوم نشان داد که بخش عمده‌ای از سوالات کاربران بر تصمیم‌گیری‌های عملی و فنی، به ویژه انتخاب روش نمونه‌برداری و ارزیابی ریسک مواجهه،

چنین رویکردی با نتایج Jin و همکاران (۲۰۲۴) هم‌راستا است که نشان دادند در حوزه‌های تخصصی و پرریسک، سامانه‌های هوشمند باید هم‌زمان قابلیت پاسخ‌گویی عمومی و ارائه راهنمایی‌های دقیق و مستند را داشته باشند تا آمادگی عملی کاربران افزایش یابد (۴۳). یافته‌های پژوهش حاضر این دیدگاه را در زمینه بهداشت حرفه‌ای تقویت می‌کند و نشان می‌دهد که معماری‌های هیبریدی، به‌ویژه در حوزه‌هایی با استانداردهای فنی پیچیده، نسبت به مدل‌های منفرد از کارایی و انطباق‌پذیری بیشتری برخوردارند. در مجموع، نتایج هدف دوم نشان می‌دهد که تیپ‌بندی سوالات نه تنها به طراحی تعامل کاربرمحور کمک کرده، بلکه مبنای منطقی برای انتخاب و بهینه‌سازی معماری RAG فراهم آورده است. این هم‌راستایی میان الگوی نیاز کاربران و ساختار فنی سامانه، یکی از عوامل کلیدی در ارتقای دقت، قابلیت استناد و پذیرش حرفه‌ای ربات توسعه‌یافته محسوب می‌شود.

بحث پیرامون هدف سوم: عملکرد معماری RAG و بهینه‌سازی فنی

- اهمیت چانک‌بندی ساختاریافته اسناد

یافته‌های حاصل از مقایسه روش‌های چانک‌بندی نشان داد که چانک‌بندی مبتنی بر ساختار اسناد استاندارد NIOSH به‌طور معناداری عملکرد بازیابی را بهبود می‌بخشد. این نتیجه نشان می‌دهد که حفظ منطق درونی اسناد فنی، از جمله مرزبندی مفهومی بخش‌ها، جداول و مراحل اجرایی، نقش تعیین‌کننده‌ای در دستیابی به بازیابی معنایی مؤثر دارد (۴۴). در مقابل، چانک‌بندی با اندازه ثابت، اگرچه از نظر پیاده‌سازی ساده‌تر است، اما می‌تواند منجر به ازهم‌گسیختگی محتوایی شود؛ وضعیتی که احتمال بازیابی ناقص یا ناهماهنگ اطلاعات را افزایش می‌دهد. این یافته با مطالعات پیشین نیز هم‌خوانی دارد که نشان داده‌اند حفظ ساختار منطقی اسناد فنی و استاندارد در فرایند پردازش، به بهبود کیفیت بازیابی اطلاعات و انسجام پاسخ‌های تولیدشده کمک می‌کند (۳۳، ۳۱). از

متمرکز بوده است. این الگوی سوالات بیانگر آن است که کاربران در جست‌وجوی تعاریف عمومی یا اطلاعات کلی نیستند، بلکه به دنبال پاسخ‌های ساختاریافته، مستند و قابل استناد برای هدایت اقدامات اجرایی خود هستند. غلبه سوالات مرتبط با مراحل عملی نمونه برداری و روش‌های آنالیز نشان می‌دهد که چالش اصلی در پایش مواجهه شغلی نه در دسترسی به استانداردها، بلکه در ترجمه دستورالعمل‌های رسمی به تصمیم‌های عملی و موقعیت‌محور نهفته است. این مشاهده با یافته‌های Heo و همکاران (۲۰۲۴) هم‌خوانی دارد که نشان دادند کیفیت و شفافیت روش‌های ارزیابی مواجهه، نقش تعیین‌کننده‌ای در کاهش عدم قطعیت‌های اجرایی و ارتقای کیفیت تصمیم‌گیری‌های سلامت دارد (۶). بر این اساس، نیاز کاربران بیش از آنکه معرفتی باشد، اجرایی و کاربردمحور است؛ به عبارت دیگر، مسئله اصلی دسترسی به دانش نیست، بلکه کاربست دقیق دانش استاندارد در شرایط واقعی است.

در چنین بستری، استفاده از معماری (RAG) صرفاً یک انتخاب فنی تلقی نمی‌شود، بلکه پاسخی ساختاری به ماهیت نیاز کاربران است. اتصال مدل زبانی به اسناد رسمی NIOSH امکان بازیابی هدفمند بخش‌های مرتبط و تولید پاسخ‌های مستند را فراهم می‌کند و از اتکای صرف به دانش ضمنی مدل جلوگیری می‌نماید (۴۲). این موضوع در حوزه‌های مقررات‌محور اهمیت ویژه‌ای دارد، زیرا پاسخ‌های غیرمستند یا حدسی می‌توانند پیامدهای عملی و حتی حقوقی به‌دنبال داشته باشند. علاوه بر این، تحلیل توزیع سوالات میان ماژول‌های پاسخ‌دهی (RAG)، LLM و رویکرد ترکیبی) نشان داد که هیچ راهکار تک ماژوله‌ای قادر به پوشش کامل دامنه نیازهای کاربران نیست. سوالات عمومی یا مفهومی می‌توانند به‌طور مؤثر توسط مدل زبانی مدیریت شوند، درحالی‌که سوالات فنی عمیق و وابسته به استاندارد نیازمند بازیابی دقیق و مستندسازی هستند. این یافته از طراحی معماری ترکیبی پشتیبانی می‌کند که در آن نوع سوال، تعیین‌کننده مسیر پردازش و پاسخ‌دهی است.

منظر نظری، نتایج این بخش نشان می‌دهد که کیفیت سازمان‌دهی داده‌های ورودی و نحوه ساختار بندی دانش، نقشی بنیادی در عملکرد سامانه‌های مبتنی بر RAG ایفا می‌کند و در بسیاری از موارد حتی می‌تواند تأثیرگذارتر از انتخاب صرف مدل زبانی باشد.

- نقش ریرنکر در افزایش دقت بازیابی

نتایج نشان داد که استفاده از ریرنکر به افزایش قابل توجه دقت پاسخ‌دهی سیستم منجر شده است. این یافته نشان می‌دهد که شباهت برداری اولیه در مرحله بازیابی، به‌تنهایی برای تشخیص ارتباط معنایی عمیق میان پرسش و اسناد کافی نیست و مرحله بازمرتب‌سازی مبتنی بر مدل‌های Cross-Encoder می‌تواند نقش کلیدی در پالایش نتایج بازیابی ایفا کند (۴۵). این نتیجه با یافته‌های مطالعات پیشین نیز هم‌راستا است که نشان داده‌اند ترکیب بازیابی برداری با مرحله بازرتبه‌بندی دقیق‌تر، می‌تواند ارتباط معنایی واقعی میان پرسش و اسناد را بهتر شناسایی کند و کیفیت پاسخ‌های نهایی را بهبود بخشد (۴۵، ۳۲). هرچند استفاده از ریرنکر با افزایش هزینه محاسباتی و زمان پاسخ‌گویی همراه است، بهبود حاصل‌شده در کیفیت پاسخ‌ها نشان می‌دهد که در حوزه‌های تخصصی و حساس مانند بهداشت شغلی، دقت اطلاعات از اولویت بالاتری نسبت به سرعت صرف برخوردار است.

- توازن میان دقت و فراخوانی در انتخاب پارامتر k

تحلیل مقدار بهینه پارامتر k نشان داد که انتخاب این پارامتر مستلزم ایجاد توازن میان افزایش فراخوانی اطلاعات و کنترل میزان نویز واردشده به سیستم است. افزایش مقدار k اگرچه احتمال از دست رفتن اطلاعات مرتبط را کاهش می‌دهد، اما هم‌زمان می‌تواند باعث ورود اسناد غیرمرتبط و کاهش انسجام پاسخ‌های تولیدشده شود. این مشاهده با یافته‌های Karpukhin و همکاران (۲۰۲۰) هم‌خوانی دارد که نشان داده‌اند افزایش بیش از حد تعداد اسناد بازیابی‌شده، در صورت نبود پالایش

مناسب، می‌تواند به کاهش کیفیت پاسخ‌های مدل زبانی منجر شود (۳۴). در پژوهش حاضر، انتخاب مقدار میانی برای k امکان دستیابی به پوشش مناسب اطلاعات موردنیاز مدل زبانی را فراهم کرد، بدون آن‌که حجم بالایی از داده‌های نامرتبط وارد فرایند تولید پاسخ شود. چنین توازنی برای جلوگیری از تولید پاسخ‌های پراکنده یا ناقص در سامانه‌های مبتنی بر RAG اهمیت ویژه‌ای دارد.

بحث پیرامون هدف چهارم: کاربردپذیری و تجربه کاربری ربات گفتگو هاوانا

نتایج حاصل از پرسشنامه CUQ نشان داد که «وضوح پاسخ‌ها» و «تناسب لحن» بالاترین امتیازها را از دیدگاه خبرگان دریافت کردند. این الگو بیانگر آن است که ربات HAVANA توانسته است با وجود محتوای فنی و پیچیده استانداردهای NIOSH، پاسخ‌هایی شفاف، منسجم و قابل فهم ارائه دهد. این یافته با مطالعات پیشین همچون پژوهش Borsci و همکاران (۲۰۲۰) همسو است که نشان دادند وضوح پاسخ و انسجام زبانی از عوامل کلیدی در پذیرش چت‌بات‌های تخصصی توسط کاربران حرفه‌ای محسوب می‌شود. این امر نشان می‌دهد که ترکیب معماری RAG با ساختار اسنادی استاندارد NIOSH توانسته است شفافیت اطلاعات و اعتمادپذیری فنی سامانه را تقویت کند (۱۸). در مقابل، امتیاز پایین‌تر مؤلفه «یادگیری و استمرار تعامل» نشان داد که مدیریت مکالمات طولانی، سناریوهای چندمرحله‌ای و پیگیری وضعیت تعاملی هنوز از چالش‌های رایج در چت‌بات‌های تخصصی به شمار می‌رود. همان‌طور که Holmes و همکاران (۲۰۱۹) نیز گزارش کرده‌اند، حفظ پیوستگی مکالمه و پشتیبانی از اصلاح تدریجی پرسش‌ها در حوزه‌های تخصصی از نقاط ضعف متداول رابط‌های مکالمه‌ای است. این نتیجه را می‌توان بازتاب مستقیم معماری فعلی HAVANA دانست که مبتنی بر یک مدل RAG بدون حافظه پایدار طراحی شده و هر پرسش را به طور مستقل پردازش می‌کند؛ بنابراین توانایی سامانه برای «یادگیری در طول تعامل» به‌صورت طبیعی محدود است (۴۶).

در کنار قضاوت حرفه‌ای کارشناسان بهداشت شغلی مورد استفاده قرار گیرد، نه جایگزین آن.

محدودیت‌ها

علی‌رغم نتایج به‌دست‌آمده، این پژوهش با چند محدودیت همراه بوده است که در تفسیر و تعمیم یافته‌ها باید مورد توجه قرار گیرند.

- نخست، دامنه محتوایی پایگاه دانش ربات به روش‌های استاندارد NIOSH محدود بوده و سایر مراجع بین‌المللی یا ملی مرتبط با پایش آلاینده‌های شغلی در این نسخه از ربات لحاظ نشده‌اند. این محدودیت می‌تواند موجب کاهش جامعیت پاسخ‌ها در شرایطی شود که چندین استاندارد موازی برای یک آلاینده یا روش نمونه‌برداری وجود دارد.

- دوم، حجم پنل خبرگان شرکت‌کننده در مراحل طراحی و ارزیابی، اگرچه از نظر تنوع تخصصی مناسب بوده است، اما از نظر تعداد محدود است. این مسئله می‌تواند بر قابلیت تعمیم نتایج کاربرپذیری و بازخوردهای کیفی اثرگذار باشد و نیاز به ارزیابی‌های گسترده‌تر در مطالعات آینده را برجسته می‌کند.

- سوم، ارزیابی عملکرد ربات عمدتاً در محیط‌های کنترل‌شده و مبتنی بر سناریوهای طراحی‌شده انجام شده و آزمون‌های میدانی گسترده در شرایط واقعی محیط کار، به دلیل محدودیت‌های زمانی و منابع انسانی، اجرا نشده است. در نتیجه، عملکرد ربات در شرایط پیچیده و متغیر محیط‌های صنعتی هنوز نیازمند بررسی‌های تکمیلی است.

- چهارم، کیفیت خروجی ربات به‌شدت به دقت و ساختار اسناد ورودی وابسته است. هرگونه ابهام، نقص یا ناسازگاری در ساختار روش‌های استاندارد می‌تواند بر چابک‌بندی، بازیابی و در نهایت پاسخ تولیدشده اثر منفی بگذارد. این وابستگی ذاتی، یکی از چالش‌های اصلی ربات‌های مبتنی بر بازیابی اسناد فنی محسوب می‌شود.
- در نهایت، اگرچه طراحی ربات بر کاهش تولید پاسخ‌های نادقیق یا حدسی متمرکز بوده است، اما مانند

این یافته نه تنها یک محدودیت، بلکه یک فرصت پژوهشی ارزشمند را برجسته می‌کند. پایین بودن امتیاز «یادگیری و بهبود در طول تعامل» نشان می‌دهد که نسخه‌های آینده HAVANA می‌توانند از معماری‌های «حافظه‌مند» و «زمینه‌آگاه» بهره ببرند؛ یعنی مدل‌هایی که از تاریخچه مکالمه، سطح دانش کاربر، و هدف مرحله‌ای تعامل آگاه بوده و این اطلاعات را در طول گفت‌وگو حفظ می‌کنند. پیشرفت‌های اخیر در مدل‌های زبانی بزرگ نشان می‌دهد که ترکیب بازیابی معنایی با لایه‌های حافظه مکالمه‌ای، ردیابی وضعیت گفتگو و شخصی‌سازی تدریجی می‌تواند به بهبود کیفیت تعاملات چندمرحله‌ای کمک کند، بدون آنکه دقت علمی یا قابلیت اتکای سیستم کاهش یابد. در نسخه‌های آینده، می‌توان معماری RAG را با اجزای رهگیری زمینه مکالمه و حافظه جلسه‌ای ادغام کرد تا سیستم بتواند تاریخچه تعاملات را حفظ کرده و پاسخ‌ها را بر اساس مسیر پیشین گفتگو تنظیم کند.

یافته‌های کیفی حاصل از فرم ثبت نظرات نیز این الگو را تقویت کرد. کاربران «دقت علمی»، «ارجاع‌پذیری» و «انطباق با استانداردهای رسمی» را مهم‌ترین نقاط قوت ربات معرفی کردند؛ نتیجه‌ای که با مرور نظام‌مند Schachner و همکاران (۲۰۲۰) درباره چت‌بات‌های حوزه سلامت نیز همخوان است. این امر نشان می‌دهد که اتصال مدل زبانی به منابع استاندارد از طریق معماری RAG نه تنها دقت اطلاعات را افزایش می‌دهد، بلکه اعتماد کاربران متخصص را نیز تقویت می‌کند. در مجموع، ترکیب طراحی کاربرمحور، چابک‌بندی ساختاریافته، بازیابی معنایی دقیق و بازمرتب‌سازی محتوا، چارچوبی مؤثر برای توسعه ربات‌های گفتگوی تخصصی در حوزه بهداشت شغلی ارائه می‌کند (۱۳). این مطالعه با محدودیت‌هایی نیز همراه بود. عملکرد ربات تا حد زیادی به کیفیت، ساختار و جزئیات اسناد ورودی وابسته است و ارزیابی انجام‌شده عمدتاً بر اساس سناریوهای شبیه‌سازی‌شده صورت گرفته است. بنابراین، HAVANA باید همچنان به‌عنوان ابزاری پشتیبان برای تصمیم‌گیری تخصصی و

سایر ربات‌های مبتنی بر مدل‌های زبانی، احتمال بروز خطاهای تحلیلی محدود، به‌ویژه در سناریوهای پیچیده یا سوالات چندمرحله‌ای، به طور کامل قابل حذف نیست. این موضوع اهمیت استفاده از ربات به عنوان ابزار پشتیبان تصمیم‌گیری و نه جایگزین قضاوت تخصصی انسانی را برجسته می‌سازد.

== پیشنهادات

یافته‌های این پژوهش چند مسیر مشخص برای توسعه‌های آتی و کاربردهای عملی ربات پیشنهادی ترسیم می‌کند.

- نخست، گسترش دامنه پایگاه دانش با ادغام سایر استانداردهای معتبر بین‌المللی نظیر OSHA، ISO و ACGIH می‌تواند جامعیت پاسخ‌ها را افزایش داده و امکان مقایسه تطبیقی بین روش‌های نمونه‌برداری و ارزیابی مواجهه را برای کاربران فراهم سازد. چنین توسعه‌ای به‌ویژه در محیط‌های صنعتی با الزامات چندگانه نظارتی اهمیت بالایی دارد.

- در گام بعدی، اجرای مطالعات میدانی در محیط‌های واقعی کار و مشارکت کاربران نهایی غیردانشگاهی (نظیر کارشناسان ایمنی و بهداشت در صنایع) می‌تواند به ارزیابی عملیاتی ربات در شرایط واقعی و شناسایی الگوهای استفاده بلندمدت منجر شود. نتایج این مطالعات می‌تواند مبنای بهبود طراحی مکالمات چندمرحله‌ای و افزایش پایداری تعامل کاربر با ربات قرار گیرد.

- از منظر فنی، توسعه راهکارهای پیشرفته‌تر برای مدیریت مکالمات طولانی، شامل حفظ زمینه بین نوبت‌های گفتگو و بهبود منطق تعامل بین ماژول‌های LLM و RAG، می‌تواند چالش‌های شناسایی‌شده در ارزیابی کاربردپذیری را کاهش دهد. همچنین بررسی استفاده از مدل‌های زبانی تخصصی‌تر یا تنظیم دقیق مدل‌ها بر روی متون تخصصی بهداشت حرفه‌ای، می‌تواند دقت تحلیلی پاسخ‌ها را در سناریوهای پیچیده افزایش دهد.

- از بعد طراحی داده، استانداردسازی نیمه‌خودکار

ساختار اسناد ورودی و توسعه ابزارهای هوشمند برای پایش کیفیت چانک‌ها، می‌تواند وابستگی ربات به کیفیت اولیه اسناد را کاهش دهد و پایداری عملکرد سیستم را در مقیاس‌های بزرگ‌تر تضمین کند.

- در سطح کاربردی، نتایج این پژوهش نشان می‌دهد که ربات‌های مبتنی بر RAG می‌توانند به عنوان ابزار پشتیبان تصمیم‌گیری در آموزش، ارزیابی اولیه مواجهه شغلی و مشاوره فنی مورد استفاده قرار گیرند. این ربات‌ها با کاهش زمان دسترسی به اطلاعات استاندارد و افزایش شفافیت پاسخ‌ها، ظرفیت ارتقای کارایی فرآیندهای بهداشت حرفه‌ای را دارند، مشروط بر آن که همچنان تحت نظارت و بازنگری متخصصان انسانی به کار گرفته شوند.

در مجموع، این پژوهش زمینه‌ای عملی برای توسعه نسل جدید ابزارهای هوشمند مبتنی بر اسناد استاندارد فراهم می‌کند که می‌توانند نقش مکمل متخصصان را در حوزه‌های تخصص‌محور و حساس ایفا نمایند و پل ارتباطی مؤثری میان دانش رسمی، فناوری هوش مصنوعی و نیازهای واقعی کاربران ایجاد کنند.

== نتیجه‌گیری

این پژوهش با هدف طراحی و ارزیابی یک ربات گفتگوی تخصصی مبتنی بر معماری (RAG) برای پشتیبانی از فرآیندهای نمونه‌برداری و آنالیز آلاینده‌های شغلی بر اساس استانداردهای NIOSH انجام شد. نوآوری اصلی این مطالعه در ترکیب مدل‌های زبانی بزرگ با یک پایگاه دانش ساختاریافته از دستورالعمل‌های استاندارد و توسعه یک سامانه کاربرمحور (HAVANA) است که قادر است پرسش‌های تخصصی کاربران را به پاسخ‌های مستند و مبتنی بر منابع رسمی تبدیل کند. یافته‌ها نشان داد که استفاده از معماری RAG امکان پیوند مؤثر میان توانایی‌های زبانی مدل‌های مولد و دانش استانداردهای تخصصی را فراهم می‌کند و در نتیجه می‌تواند دقت، قابلیت استناد و شفافیت پاسخ‌ها را در حوزه‌های مقررات‌محور مانند بهداشت حرفه‌ای افزایش دهد. مشارکت خبرگان

مفهوم انسان در حلقه تصمیم‌گیری قرار می‌گیرد. بدین معنا که خروجی‌های سامانه باید در کنار دانش، تجربه و قضاوت حرفه‌ای متخصصان مورد استفاده قرار گیرند. در نتیجه، تصمیم‌گیری نهایی در خصوص انتخاب روش نمونه‌برداری، تفسیر نتایج و ارزیابی مواجهه شغلی همچنان بر عهده کارشناسان و متخصصان حوزه بهداشت حرفه‌ای خواهد بود.

تشرک و قدردانی

این مقاله برگرفته از پایان‌نامه کارشناسی ارشد انجام‌شده در گروه مهندسی بهداشت حرفه‌ای و ایمنی کار، دانشکده بهداشت و ایمنی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران است (کد پایان‌نامه: ۴۳۰۱۱۱۵۲). این مطالعه با حمایت مالی دانشگاه علوم پزشکی شهید بهشتی و در قالب یک طرح پژوهشی مصوب انجام شده است (کد گرنت: ۴۳۰۱۸۶۷۵). در پایان، از مدل زبانی هوشمند Gemini به خاطر کمک در ویرایش و بهبود بیان برخی بخش‌های این مقاله قدردانی می‌شود. لازم به ذکر است که تمام محتوای نهایی توسط نویسندگان بازبینی، تأیید و مسئولیت آن کاملاً بر عهده آن هاست.

ملاحظات اخلاقی

پژوهش حاضر با تأیید کمیته اخلاق دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران (کد اخلاق IR.SBMU.PHNS.REC.1403.047) انجام شده است. کلیه مراحل تحقیق براساس اصول اخلاقی مندرج در بیانیه هلسینکی (ویرایش ۲۰۱۳) و آیین‌نامه اخلاق پژوهش‌های پزشکی کشور اجرا گردید. داده‌های کیفی مورد استفاده در این مطالعه از طریق مصاحبه‌های نیمه‌ساختاریافته با کارشناسان بهداشت حرفه‌ای جمع‌آوری شدند. پیش از شروع هر مصاحبه، اهداف پژوهش به‌طور شفاف به شرکت‌کنندگان توضیح داده شد و رضایت آگاهانه کتبی از آن‌ها اخذ گردید. شرکت‌کنندگان مطلع شدند که می‌توانند در هر مرحله از ادامه همکاری صرف‌نظر کنند

1. Human-in-the-loop

در مراحل طراحی نیز به شناسایی نیازهای واقعی کاربران و شکل‌گیری ساختار پرسش‌های کاربرمحور کمک کرد؛ نیازهایی که عمدتاً بر تصمیم‌گیری‌های عملی مانند انتخاب روش مناسب نمونه‌برداری، تنظیم شرایط اجرایی و تفسیر الزامات فنی استانداردها متمرکز بودند.

از منظر فنی، نتایج نشان داد که کیفیت ساختاردهی پایگاه دانش، چابک‌بندی مناسب اسناد استاندارد و بهینه‌سازی فرآیند بازبینی اطلاعات نقش تعیین‌کننده‌ای در عملکرد سامانه‌های مبتنی بر RAG دارند. همچنین ارزیابی کاربرپذیری نشان داد که ربات گفتگوی HAVANA از دیدگاه کاربران متخصص قابل قبول بوده و می‌تواند دسترسی سریع‌تر به اطلاعات استاندارد، کاهش ابهام در تفسیر دستورالعمل‌ها و تسهیل تصمیم‌گیری‌های تخصصی را در محیط‌های کاری پشتیبانی کند. بر این اساس، ربات‌های گفتگوی تخصصی مبتنی بر RAG می‌توانند به‌عنوان ابزارهای پشتیبان تصمیم‌گیری در حوزه‌های استانداردمحور مورد استفاده قرار گیرند؛ به‌ویژه در آموزش دانشجویان بهداشت حرفه‌ای، پشتیبانی از فعالیت‌های کارشناسان بهداشت حرفه‌ای و دسترسی سریع به الزامات فنی در محیط‌های صنعتی. با این حال، چنین سامانه‌هایی باید به‌عنوان ابزار مکمل در کنار قضاوت حرفه‌ای متخصصان در نظر گرفته شوند و نه جایگزین آن. در نهایت، گسترش پایگاه دانش با منابع استاندارد بیشتر، یکپارچه‌سازی دستورالعمل‌های بین‌المللی و ارزیابی عملکرد سامانه در محیط‌های واقعی کاری می‌تواند مسیرهای مهمی برای تحقیقات آینده و توسعه سامانه‌های هوشمند مشابه در سایر حوزه‌های تخصصی فراهم کند.

با توجه به حساسیت تصمیم‌گیری‌ها در حوزه بهداشت حرفه‌ای، لازم است تأکید شود که ربات HAVANA به‌عنوان یک ابزار پشتیبان تصمیم طراحی شده است و هدف آن تسهیل دسترسی به اطلاعات استاندارد و کمک به تفسیر اولیه دستورالعمل‌های NIOSH است، نه جایگزینی قضاوت تخصصی کارشناسان بهداشت حرفه‌ای. رویکرد به‌کارگرفته‌شده در این پژوهش در چارچوب

دقیق منابع مورد استفاده قرار گرفتند. تمامی یافته‌ها به‌صورت صادقانه، شفاف و بدون هیچ‌گونه دستکاری یا حذف انتخابی گزارش شده‌اند. نویسندگان هیچ تعارض منافع مالی یا غیرمالی در ارتباط با این پژوهش اعلام نمی‌دارند و تأکید دارند که هیچ نهاد یا سازمانی در تحلیل داده‌ها یا تفسیر نتایج دخالتی نداشته است.

تضاد منافع

نویسندگان تصریح می‌کنند که هیچ‌گونه تضاد منافی در مطالعه حاضر وجود ندارد.

REFERENCES

- Saxena V, Verma S. Managing VOC Hazards Proactively for a Safer, Healthier Workplace. BARC Newslett. 2025.
- Hao R, Lin X, Yao Z, Wang M, Sun J, Wang H. Emission characteristics, environmental impact, and occupational health risks assessment of volatile organic compounds (VOCs) in the hairdressing sector. Atmos Pollut Res. 2025;16(5):102471.
- Howard J. The National Institute for Occupational Safety and Health (NIOSH). 2024.
- Ashley K. Harmonization of NIOSH sampling and analytical methods with related international voluntary consensus standards. J Occup Environ Hyg. 2015;12(7):D107-D15.
- Garcia J-C, Varas M, Chang L, Moreira E, editors. Optimizing Manual Material Handling Safety: A Hybrid Approach Using NIOSH and Artificial Intelligence. E3S Web Conf. 2025: EDP Sciences.
- Heo S, Schuch D, Junger WL, Zhang Y, de Fatima Andrade M, Bell ML. The impact of exposure assessment on associations between air pollution and cardiovascular mortality risks in the city of Rio de Janeiro, Brazil. Environ Res. 2024;263:120150.
- DeBord DG, Shoemaker D, B'Hymer C, Snawder J. Application of biological monitoring methods for chemical exposures in occupational health. In: National Institute for Occupational Safety and Health, editor. Department of Health and Human Services CfDCAp. vol Chapter BI. 2022.
- Sarker IH. AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems. SN Comput Sci. 2022;3(2):158.
- Muñoz-Hernandez R, Rangel-Lara S. Artificial Intelligence in the Field of Ergonomics and Occupational Health for Risk Assessment. J Ind Eng Rev Ing Ind. 2025;9(21):1.
- Jung S, Qing L, Su B, Xu X. Toward a Fully Automated NIOSH Lifting Equation Using Computer Vision. Hum Factors. 2026;68(5):563-79.
- Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. Adv Neural Inf Process Syst. 2020;33:9459-74.
- Wang C, Liu X, Yue Y, Tang X, Zhang T, Jiayang C, et al. Survey on factuality in large language models: Knowledge, retrieval and domain-specificity. arXiv preprint arXiv:231007521. 2023.
- Schachner T, Keller R, v Wangenheim F. Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. J Med Internet Res. 2020;22(9):e20701.
- De Villiers C, Farooq MB, Molinari M. Qualitative research interviews using online video technology—challenges and opportunities. Meditari Accountancy Res. 2022;30(6):1764-82.
- Chuan C-H, Morgan S. Creating and evaluating chatbots as eligibility assistants for clinical trials: an active deep learning approach towards user-centered classification. ACM Trans Comput Healthcar. 2020;2(1):1-19.
- Varun S, Joseph L, editors. A Hybrid Deep Learning

- Algorithm for Improved ChatBot Accuracy and Relevance Through Advanced Retrieval-Augmented Generation. 2024 Int Conf Innov Comput Intell Commun Smart Electr Syst (ICSES). 2024: IEEE.
17. Zheng D, Yang J, Yong B, editors. Open Domain Question Answering Based on Retriever-Reader Architecture. Int Conf Bus Intell Inf Technol. 2021: Springer.
18. Borsci S, Malizia A, Schmettow M, Van Der Velde F, Tariverdiyeva G, Balaji D, et al. The chatbot usability scale: the design and pilot of a usability scale for interaction with AI-based conversational agents. Pers Ubiquitous Comput. 2022;26(1):95-119.
19. Lee H, Kim J, Kim S, Kong H-J, Joo H, Lee D, et al. Usability evaluation of user requirement-based teleconsultation robots: a preliminary report from South Korea. Methods Inform Med. 2020;59(02/03):086-95.
20. Ahmad M, Wilkins S. Purposive sampling in qualitative research: A framework for the entire journey. Qual Quant. 2025;59(2):1461-79.
21. Naz N, Gulab F, Aslam M. Development of qualitative semi-structured interview guide for case study research. 2022.
22. May CR, Albers B, Bracher M, Finch TL, Gilbert A, Girling M, et al. Translational framework for implementation evaluation and research: a normalisation process theory coding manual for qualitative research and instrument development. Implement Sci. 2022;17(1):19.
23. Lochmiller CR. Conducting thematic analysis with qualitative data. Qual Rep. 2021;26(6):2029-44.
24. Gao J, Choo KTW, Cao J, Lee RK-W, Perrault S. CoAICoder: Examining the effectiveness of AI-assisted human-to-human collaboration in qualitative analysis. ACM Trans Comput Human Interact. 2023;31(1):1-38.
25. Nicmanis M, Spurrier H. Getting Started with Artificial Intelligence Assisted Qualitative Analysis: An Introductory Guide to Qualitative Research Approaches with Exploratory Examples from Reflexive Content Analysis. Int J Qual Methods. 2025;24:16094069251354863.
26. Bloch T, Sacks R. Clustering information types for semantic enrichment of building information models to support automated code compliance checking. J Comput Civ Eng. 2020;34(6):04020040.
27. Bala MY, Damla K. A review of human-computer interaction design approaches towards information systems development. BRAIN Broad Res Artif Intell Neurosci. 2021;12(1):229-50.
28. Valencia ME. Qualitative inquiry and research design: Choosing among five traditions. Estud Gest. 2020(7):323-6.
29. Wambsganss T, Schmitt A, Mahning T, Ott A, Soellner S, Ngo NA, et al., editors. The potential of technology-mediated learning processes: A taxonomy and research agenda for educational process mining. Int Conf Inf Syst (ICIS). 2021.
30. Tuunanen T, Winter R, Brocke Jv. Dealing with complexity in design science research: a methodology using design echelons. MIS Quart. 2024;48(2):427-58.
31. Yu H, Gan A, Zhang K, Tong S, Liu Q, Liu Z, editors. Evaluation of retrieval-augmented generation: A survey. CCF Conf Big Data. 2024: Springer.
32. Salemi A, Zamani H, editors. Evaluating retrieval quality in retrieval-augmented generation. Proc 47th Int ACM SIGIR Conf Res Dev Inf Retrieval. 2024.
33. Bansal A. Optimizing RAG with hybrid search and contextual chunking. J Eng Appl Sci Technol. 2023;5(4):2-5.
34. Karpukhin V, Oguz B, Min S, Lewis PS, Wu L, Edunov S, et al., editors. Dense Passage Retrieval for Open-Domain Question Answering. EMNLP. 2020.
35. Chirkova N, Rau D, Déjean H, Formal T, Clinchant S, Nikoulina V. Retrieval-augmented generation in multilingual settings. arXiv preprint arXiv:240701463. 2024.
36. Chiu J, Shinzato K, editors. Cross-encoder data annotation for bi-encoder based product matching. Proc 2022 Conf Empir Methods Nat Lang Process: Ind Track. 2022.
37. Huang W, Xia F, Shah D, Driess D, Zeng A, Lu Y, et al. Grounded decoding: Guiding text generation with grounded models for embodied agents. Adv Neural Inf Process Syst. 2023;36:59636-61.
38. Boyd K, Potts C, Bond R, Mulvenna M, Broderick T, Burns C, et al., editors. Usability testing and trust analysis of a mental health and wellbeing chatbot. Proc 33rd Eur Conf Cogn Ergonomics. 2022.
39. Tabrizi A, Mehrabi Y, Dorfeshan F, Vahedi H, Farhang Dehghan S. Psychometric Validation of the Persian Short-Form Chatbot Usability Questionnaire: A Cross-Sectional Study. (Manuscript in preparation). 2026.

40. Dindler C, Smith R, Iversen OS. Computational empowerment: participatory design in education. *CoDesign*. 2020;16(1):66-80.
41. Stratton E, Choi I, Peters D, Calvo RA, Harvey SB, Glozier N. Co-designing a web-based decision aid tool for employees disclosure of mental health conditions: a participatory study design using employee and organizational preferences. *JMIR Formative Res*. 2020;4(11):e23337.
42. Arslan M, Ghanem H, Munawar S, Cruz C. A Survey on RAG with LLMs. *Procedia Comput Sci*. 2024;246:3781-90.
43. Jin Y, Coombs WT, Wang Y, van der Meer TG, Shivers BN. "READINESS": A keystone concept beyond organizational crisis preparedness and resilience. *J Contingent Crisis Manag*. 2024;32(1):e12546.
44. Zhu Y, Yuan H, Wang S, Liu J, Liu W, Deng C, et al. Large language models for information retrieval: A survey. *ACM Trans Inf Syst*. 2025;44(1):1-54.
45. Askari A, Aliannejadi M, Kanoulas E, Verberne S. Generating synthetic documents for cross-encoder re-rankers: A comparative study of chatgpt and human experts. *arXiv preprint arXiv:230502320*. 2023.
46. Holmes S, Moorhead A, Bond R, Zheng H, Coates V, McTear M, editors. Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? *Proc 31st Eur Conf Cogn Ergonomics*. 2019.