ISSN 2616-5783 Vol.8, Issue 11: 140-157, DOI: 10.25236/AJHSS.2025.081122

Applications of AI Chatbots in Occupational Safety and Health in Archaeology and User Experience

Mengxuan Zhang[1,a,*], Danni Lin[1], Chuqing Huang[1], Yating Zhao[1], Zhuojia Tan[1], Haibei Xie[2], Yan Zuo[2]

[1]*The University of Melbourne, Melbourne, Australia*
[2]*Department of Gynecology and Obstetrics Nursing, West China Second University Hospital, Chengdu, China*
[a]*zhangmengxuan007@gmail.com*
**Corresponding author*

Abstract: *This qualitative study evaluates user experiences with ArchHealth, a customized AI chatbot delivering occupational safety and health (OSH) information to archaeological professionals. Faced with unique hazards and inadequate traditional resources, participants (n=14) were interviewed after a 4-week pilot. Analysis revealed that ArchHealth successfully centralized fragmented OSH knowledge into a user-friendly interface, with users valuing its comprehensiveness and efficiency, which facilitated a journey from initial trial to routine use. While the tool demonstrates strong potential to bridge critical information gaps, future iterations require enhancements in localization, offline functionality, and multimedia support to maximize adoption and safety outcomes in field and lab contexts.*

Keywords: *Archaeology, Occupational Health and Safety, AI Chatbot, Large Language Model, User Experience, Qualitative Study*

1. Introduction

Occupational health and safety (OSH) in archaeology encompasses the identification, assessment, and mitigation of risks inherent to fieldwork, laboratory analysis, and museum curation. Defined as "the protection of workers' physical, mental, and social well-being from hazards arising in or from the workplace", archaeological OSH faces unique challenges due to its dynamic environments[1, 2]. Field sites, laboratories, and museums, which are the primary professional settings, expose workers to physical injuries (e.g., musculoskeletal strains from excavation), biological hazards (e.g., pathogen exposure in burial sites), chemical risks (e.g., solvents in artifact preservation), ergonomic strains (e.g., prolonged kneeling during surveys), and psychosocial stressors (e.g., isolation in remote fieldwork)[3-5].

The consequences of inadequate OSH guidance in archaeology are severe. Between 2018 and 2023, field schools reported a 15% increase in incidents ranging from heat exhaustion to equipment-related accidents, often linked to inconsistent safety protocols[3]. Project delays due to injuries cost an estimated $2.3 million annually in U.S.-based archaeological contracts, while psychosocial stressors contribute to a 22% attrition rate among early-career professionals[4, 5]. Traditional OSH resources, such as static data sheets, manuals, or infrequent training sessions, fail to address the real-time, context-specific needs of archaeologists. Kintigh et al.[6] critiques systemic limitations in archaeological practice that fragmented knowledge systems and non-standardized protocols, which are common in static OSH manuals, fail to address dynamic fieldwork risks. Ferreira et al.[7] systematically review occupational health risks among art conservators and restorers, who identify chemical exposures (e.g., solvents in artifact preservation), ergonomic strains from prolonged static postures, and biological hazards (e.g., mold spores in burial materials) as pervasive yet under-addressed risks. Notably, 68% of conservators reported inadequate access to real-time safety guidance, relying instead on outdated manuals or informal peer advice. They emphasize that generic safety protocols fail to account for context-specific scenarios, such as handling degraded materials in humid environments or adapting ergonomic practices to irregular workspaces. As a result, dynamic, interactive tools are needed to replace traditional static guidelines and adaptive, accessible tools that deliver timely and accurate OSH information tailored to archaeological workflows are urgently needed.

LLM AI tools are rapidly proliferating across a wide array of domains, demonstrating transformative potential in fields such as agriculture, medicine, information security, law, education, digital forensics, and autonomous systems. For example, OpenAI's GPT series and similar models have been applied to enhance efficiency and innovation in agriculture, support clinical decision-making and rare disease diagnosis in medicine, streamline legal research and democratize access to justice, and automate code generation in programming[8-13]. In education, LLMs are rapidly reshaping the domain by automating and enhancing a wide range of tasks, including content generation, grading, feedback provision, personalized learning, and curriculum development[14, 15]. However, adoption of such tools in archaeology remains nascent.

Current chatbot applications in OSH prioritize incident reporting over preventive education. For instance, a 2024 scoping review of IoT tools identified "reactive surveillance" as the dominant use case, with only 12% of systems offering proactive risk mitigation strategies[16]. This reactive paradigm neglects the preventive ethos of frameworks such as ISO 45001, which emphasizes "structured, preventive approaches to health and safety management" through policy alignment and continuous improvement[17]. Meanwhile, LLM customization techniques, such as fine-tuning on domain-specific corpora and reinforcement learning from human feedback, have enabled chatbots in construction to reduce safety violations by 41% by contextualizing regulatory knowledge[18]. Despite these successes, no AI-driven tools explicitly target archaeological OSH, leaving a critical gap in a field where 64% of professionals report inadequate access to scenario-specific safety resources[19]. The absence of AI-driven, adaptive solutions in archaeology perpetuates gaps in compliance and hazard responsiveness, underscoring the urgent need for tools such as ArchHealth to modernize safety practices.

The primary aim of this study was to explore user experiences of ArchHealth, a ChatGPT-based customized agent designed for archaeological OSH. Specifically, Objective 1 assessed the usability and utility of the chatbot in real-world professional workflows, examining ease of interaction and perceived value of guidance. Objective 2 sought to inform further refinement of ArchHealth by gathering qualitative feedback on content accuracy, interface design, and contextual relevance. Finally, Objective 3 extended the insights to inform development and adoption of similar AI-driven chatbots in specialized fields such as archaeology, leveraging lessons learned to guide future projects.

Our work makes four key contributions. First, it demonstrates that non-experts can build a domain-specific OSH chatbot using ChatGPT's Create agent and a curated knowledge base of archaeological OSH PDFs, supplemented by web search for general context. Second, it provides qualitative evidence on end-user acceptance and use of ArchHealth. Third, it offers practical guidelines for refining domain-specific AI tools based on user feedback. Finally, it highlights the potential for AI integration to enhance OSH practices in archaeology, informing policy, training, and future research on interactive digital assistants in niche professional domains. By filling a critical gap, namely the absence of AI-driven OSH chatbots in archaeology, our study advances occupational safety practice and models an innovative approach for specialized AI deployments. The insights gained here can inform educators, policymakers, and technologists aiming to leverage LLMs for domain-specific safety support, ultimately reducing injury risks and improving operational efficiency in archaeological and related fields.

2. Methods

2.1 Development and dissemination of ArchHealth

The ArchHealth chatbot was independently developed by the first author (MZ), a postgraduate student of the School of Historical and Philosophical Studies in the University of Melbourne, who routinely engage in cultural materials conservation, without prior programming or AI development experience. To begin, Zhang explored ChatGPT-4o’s Create agent feature, a built-in tool that enables users to configure a customized conversational agent by uploading domain-specific materials and defining tailored instructions without requiring programming expertise, and experimented with existing AI-driven chatbots to understand their capabilities and limitations. Through guided tutorials and trial interactions, Zhang learned how to navigate the Create agent interface, which automated much of the infrastructure setup. Once familiar with the platform, Zhang generated initial customization instructions using the interactive agent creation tool.

After producing a preliminary set of instructions, Zhang refined them using a structured prompt template recommended for prompt drafting. The template directed the agent to search uploaded materials first and to rely on web-based information when necessary. Zhang then conducted rudimentary tests on the chatbot's performance using 10 sample questions to verify basic functionality and to identify areas needing adjustment. See Multimedia Appendix 1 for the final version of the customization instructions.

ISSN 2616-5783 Vol.8, Issue 11: 140-157, DOI: 10.25236/AJHSS.2025.081122

Concurrently, Zhang collected twenty key PDF files, including excavation datasheets, laboratory safety manuals, museum handling protocols, and relevant regulatory guidelines, to serve as ArchHealth's knowledge base, in addition to its current knowledge base and access to knowledge on the Internet. Each PDF was uploaded through the Create agent interface (Figure 1).

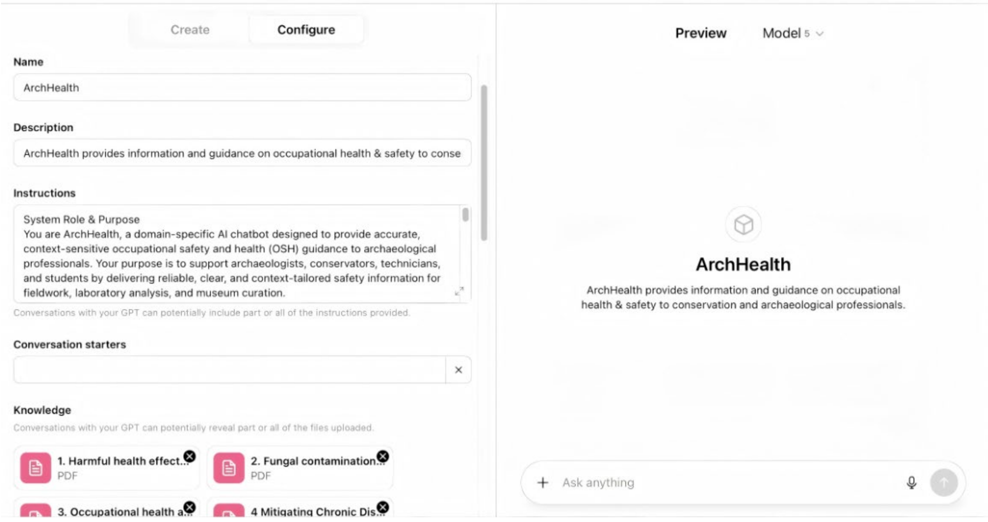

Figure 1: depicts the setup of ArchHealth, a ChatGPT-based agent tailored for archaeological OSH guidance using a domain-specific document library.

To assess performance, 100 frequently asked OSH questions were compiled from three sources: (1) examination problems routinely administered to archaeological students, which had established correct answers; (2) practice-related questions provided by archaeological professionals, where reference answers were available in institutional safety protocols; and (3) fact-based questions extracted directly from the uploaded OSH reference documents, where correct answers were explicitly available. For each question, a verified reference answer was prepared and independently cross-checked by two research team members (MZ, DL) with training in both occupational safety and archaeological practice. ArchHealth's responses were evaluated against the reference answers. Accuracy was defined as the proportion of chatbot responses that were fully aligned with the reference answers in terms of factual correctness, inclusion of essential elements, and consistency with established OSH best practices. To meet project requirements, the chatbot had to exceed 90% accuracy and produce responses with adequate detail and thoroughness. After five rounds of testing and incremental prompt refinements, ArchHealth achieved a 97% accuracy rate against expert-verified answers, surpassing the initial target (Figure 2).

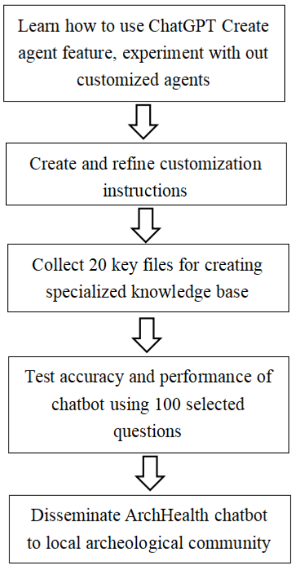

Figure 2: Development and testing of ArchHealth, which achieved 97% accuracy on 100 archaeological OSH questions prior to release.

Following successful testing, ArchHealth was disseminated to the Melbourne archaeological community. (Figure 3) Zhang shared a ChatGPT link of the chatbot through her professional network, with brief tutorial instructions. After four weeks, early-adopter were interviewed to gather real-world user experiences and insights.

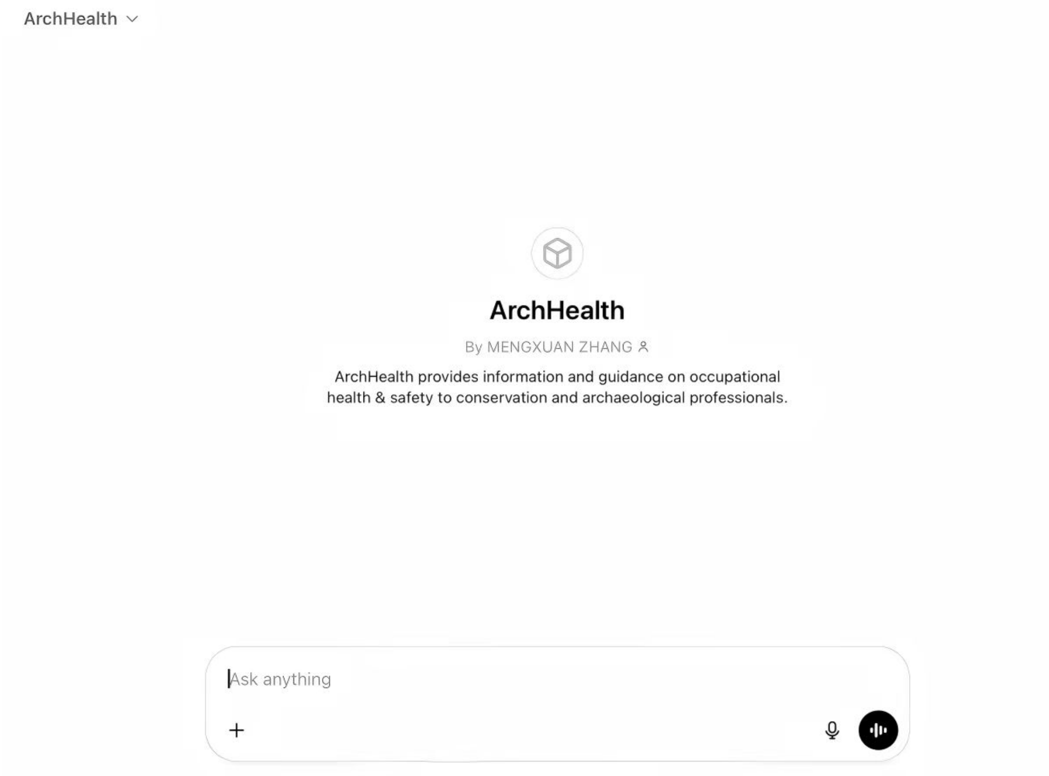

Figure 3: Screenshot of the ArchHealth chatbot interface provided to participants for user experience testing.

2.2 Study design and setting

A descriptive qualitative design guided this study to capture participants' firsthand experiences with ArchHealth without imposing predefined theoretical frameworks[20, 21]. This approach enabled exploration of usability, utility, and integration of the AI tool in professional workflows without forcing data into preset categories. The design allowed participants to share their perspectives in their own words. Such openness was essential to understand how ArchHealth functioned in real contexts. Notably, the descriptive qualitative approach prioritized exploration of user experiences as they emerged. This flexibility suited our study aim of examining ArchHealth's role across varied settings, including field, laboratory, museum, and government contexts. Participants encountered different workflows and encountered unique challenges that shaped their use of the chatbot[22, 23].

The descriptive qualitative study was designed through collaboration between researchers in Melbourne, Australia and Chengdu, China. While the Melbourne team possessed in-depth understanding of archaeological practice, exposure to technologies, and access to the local archeological community, they lacked experience in qualitative research. Consequently, colleagues in Chengdu, who had established proficiency with descriptive qualitative studies, guided the overall research design, data collection strategy, and analytical framework.

Together, the two teams developed a semi-structured interview guide, which the first author (MZ) pilot tested with three ArchHealth users. Based on feedback regarding question clarity and flow, the researchers refined the guide to ensure it elicited rich, detailed accounts of participants' experiences. The pilot testing feedback was not included in subsequent formal data analysis. See Multimedia Appendix 2 for the final version of the interview guide.

All members of the research team received training in descriptive qualitative methodology, particularly on Colaizzi's seven-step data analysis procedure. The first author (MZ) underwent additional instruction in conducting face-to-face, semi-structured interviews.

ISSN 2616-5783 Vol.8, Issue 11: 140-157, DOI: 10.25236/AJHSS.2025.081122

2.3 Sampling

Participants were selected through purposive sampling to ensure representation across diverse archaeological roles[24]. Eligibility criteria required an age of 20-60 years, active engagement in occupational OSH operations and decision-making in their respective settings, and basic familiarity with digital tools. Individuals with known psychological or cognitive conditions potentially compromising the findings of the study or unwilling to participate in interview were excluded.

Recruitment took place in the professional archaeological community in Melbourne, Australia, including university-affiliated researchers, cultural heritage conservators, museum staff, and archaeological field practitioners who were accessible through the Melbourne research team's professional networks. Potential participants were identified or recommended to the primary investigator (MZ), who assessed their eligibility based on the inclusion and exclusion criteria. Eligible individuals were contacted directly, provided with an introduction to the ArchHealth chatbot, and informed about the study's objectives and procedures. A follow-up call was scheduled for interview. In total, 30 archaeological professionals from the Melbourne community were approached. Four weeks after dissemination of the chatbot, a final group of participants was selected and invited for interviews.

Interviews continued until thematic saturation was achieved. That is, after 14 participants, no new insights emerged, confirming that the sample size was sufficient to capture the range of experiences with ArchHealth[25, 26]. Participants' demographic characteristics were documented to contextualize findings and ensure transferability of results[27].

2.4 Data collection and analysis

Between March 11 and March 27, 2025, the first author (MZ) conducted 14 in-person interviews. Several interview techniques were applied. Probing questions were used to encourage participants to elaborate on their initial responses and to provide specific examples. Strategic silence allowed participants time to reflect and add further detail. Clarification and paraphrasing were employed to confirm understanding and prompt additional insights. Follow-up questions were tailored to individual responses.

All interviews were audio recorded and transcribed verbatim within 24 hours. The transcriptions were independently verified by two other researchers, one from the Melbourne team and the other from the Chengdu team, for accuracy.

Data analysis followed Colaizzi's seven-step procedure[28, 29], implemented collaboratively by the Melbourne and Chengdu teams through a series of teleconferences. First, both research groups immersed themselves in the data by reading each transcript in its entirety to gain an overall sense of participants' experiences. Second, significant statements that directly addressed usability, utility, and contextual integration of ArchHealth were extracted and recorded. Third, the team formulated meanings from those statements by paraphrasing and interpreting participants' intended messages. Fourth, similar meanings were organized into clusters, which revealed preliminary thematic categories. During this step, researchers in Melbourne and Chengdu met weekly to compare coded statements and reconcile variations in interpretation, ensuring consistency across both sites. Fifth, an exhaustive description of each theme was written by synthesizing clustered meanings into comprehensive accounts. Sixth, researchers distilled the descriptions into a concise, fundamental structure that captured the essence of participants' experiences. Finally, to validate findings, the first author (MZ) returned to participants by phone to confirm that the thematic interpretations accurately reflected their perspectives.

The study and findings are reported according to the Standards for Reporting Qualitative Research (SRQR). See Multimedia Appendix 3 for the completed SRQR Checklist.

2.5 Study rigor

We implemented the following strategies to ensure study rigor[30-33]. First, credibility was enhanced through member checking. After initial data analysis, emerging themes were returned to participants for validation. Participants confirmed that the thematic descriptions reflected their experiences. Minor refinements were made where appropriate but no substantive changes to thematic content. We made peer debriefing regularly during teleconferences between the Melbourne and Chengdu teams, where researchers critically examined each other's interpretations of significant statements and codes, challenged assumptions, and clarified potential biases. We addressed dependability by maintaining a

ISSN 2616-5783 Vol.8, Issue 11: 140-157, DOI: 10.25236/AJHSS.2025.081122

detailed audit trail documenting key steps of data collection and analysis. Audio recordings, verbatim transcripts, and coding matrices were stored chronologically. An external audit was conducted by a senior qualitative methodologist, who reviewed a random selection of transcripts, codes, and thematic memos to confirm consistency between raw data and analytical outcomes. Confirmability was strengthened through reflexive journaling by the first author (MZ). Throughout the interview and analysis process, the first author (MZ) recorded personal reflections on how prior beliefs about AI tools might influence interpretation. During teleconference coding sessions, the reflections were disclosed, allowing co-researchers to account for and mitigate potential subjectivity. Transferability was promoted by providing description of participant contexts, roles, and work environments so readers can evaluate applicability to other archaeological communities.

2.6 Use of ChatGPT in Study and Preparation of Manuscript

We used ChatGPT to polish the English language in the final version of the manuscript. All other aspects of the study and manuscript preparation were carried out manually.

2.7 Ethical considerations, privacy protection, and data confidentiality

This study received exemption from ethical review by the Ethics Committee of the Second West China Hospital, Sichuan University, as it did not involve patients or patient data and posed minimal risk. Before data collection, each participant was provided with detailed information about the study's purpose, procedures, and voluntary nature; written informed consent was obtained. Interviews took place in private locations chosen by participants to protect their privacy. All identifying information was removed or de-identified from demographic forms, transcripts, and field notes, and each transcript was assigned a unique code (e.g., P1). Audio recordings, transcripts, and other study documentation were stored on a password-protected flash drive accessible only to the Melbourne research team. The files were shared with the Chengdu team via a password-protected cloud storage service (Aliyun); once downloaded by the lead Chengdu researcher (YZ), the online copies were promptly deleted. Access to all data remained restricted to designated members of the research teams.

3. Results

3.1 Participant demographics

Table 1: Demographic characteristics of participants (N = 14)

Characteristic	Summary statistics
Gender	Female: 11 (78.6%) Male: 3 (21.4%)
Age, year	Range: 21–51 Mean±SD: 32.4±8.6 Median: 28
Education level	Bachelor’s degree: 1 (7.1%) Master’s degree (completed): 8 (57.1%) Second Master’s (completed): 2 (14.3%) Master’s (currently enrolled): 1 (7.1%) Doctoral degree: 2 (14.3%)
Professional role(s)	Student (only): 7 (50.0%) Cultural heritage conservator/technician (only): 3 (21.4%) Museum technician: 1 (7.1%) Archaeological technician: 1 (7.1%) History professor: 1 (7.1%) Associate professor: 1 (7.1%) Dual role (Student + Conservator/Technician): 1 (7.1%)*
Years of experience, year	Range: 1–11 Mean (SD): 3.86 (3.31) Median: 2

Note: Participants in semi-structured interviews to solicit their user experiences with the ArchHealth chatbot. The study was a descriptive qualitative investigation conducted in Melbourne, Australia, between March 11 and March 27, 2025.

Fourteen participants were interviewed until data saturation was reached, with interview durations ranging from 15.5 to 26.5 minutes (average, 22.6 minutes). The cohort exhibited moderate diversity, although gender imbalance was evident: 11/14 participants (78.6%) were female. Ages ranged from 21 to 51 years (median, 28 years), concentrated in the late twenties and early thirties, capturing both early- and mid-career professionals. Educational attainment varied widely: one held a bachelor's degree; ten held master's degrees (including two with a second master's); one was a current master's student; and two held doctoral degrees (one in history, one in archaeology). Professionally, participants comprised students (n = 7), cultural heritage conservators or technical specialists (n = 4), one museum technician, one archaeological technician, a history professor, and an associate professor of archaeology, reflecting field, laboratory, museum, and academic settings. Years of experience in archaeology spanned one to eleven years (median, 2 years), with over half the group reporting one to two years (n = 8) and the most experienced member having over a decade in the field. Although female participants predominated, the sample offered sufficient variation in educational background, professional role, and experience to provide a multifaceted perspective on how ArchHealth might integrate into diverse workplace contexts. Table 1 summarizes participants' demographic characteristics.

3.2 Key Themes

3.2.1 Perceived benefits of ArchHealth

Participants consistently highlighted how ArchHealth streamlined access to occupational health information by consolidating diverse safety considerations into a single, intuitive resource. Rather than scouring multiple safety data sheets, protocols, or online manuals, users praised ArchHealth's ability to provide concise, context-specific recommendations tailored to archaeological and conservation settings. Its structured output, dividing responses into logical categories such as chemical hazards, personal protective equipment (PPE), and emergency procedures, allowed users to quickly identify relevant guidance without sifting through dense text. Moreover, ArchHealth's interactive nature, which prompted follow-up clarifications and offered ready-made templates (e.g., risk assessment checklists), was viewed as a significant time-saver and confidence builder. Novices particularly valued its nonjudgmental, step-by-step explanations, which served as a safe entry point into technical safety practices. Even experienced professionals found ArchHealth useful for validating decisions and uncovering overlooked hazards. By offering rapid, well-organized, and comprehensive advice, ArchHealth transformed what was previously a fragmented, time-consuming process into a cohesive, user-friendly experience, laying the foundation for deeper exploration of its specific benefits through the sub-themes that follow.

3.2.1.1 Comprehensiveness and breadth

Users lauded ArchHealth for integrating a wide array of safety topics, ranging from laboratory solvent handling to field excavation risks, into a cohesive framework. Unlike traditional approaches that required consulting separate chemical manuals, archaeological guidelines, and laboratory SOPs, ArchHealth bundled these diverse considerations into a single, coherent response. This breadth of coverage was particularly valuable for interdisciplinary projects involving lab analysis, fieldwork, and museum curation, where multiple hazard domains intersect. By offering a holistic perspective, ArchHealth reduced the likelihood of overlooking less obvious risks, such as mold growth in storage areas or noise exposure during artifact cleaning, and saved users significant research time.

"I asked about acetone in the lab and also dusty field conditions, and the chatbot (ArchHealth) gave me both PPE recommendations and ventilation advice in one response... Instead of flipping between MSDS, museum guidelines, and OSHA pages, I just typed 'lab and dig site safety,' and it covered everything." (P2)

"I was dealing with solvents in conservation and trench shoring in excavation, the AI (ArchHealth) addressed both chemical and physical hazards without me needing separate searches." (P9)

"It even mentioned latent mold issues in artifact storage while giving me respirator advice for site smoke. That breadth was impressive." (P7)

"the chatbot (ArchHealth)'s coverage went from solvent disposal to UV radiation in photogrammetry, which saved me hours comparing different manuals." (P14)

3.2.1.2 Structured, readable output

Another advantage of ArchHealth was its presentation of information in a clear, organized format. Participants noted that responses were broken into labeled sections such as "Chemical Hazards," "Recommended PPE," and "Emergency Procedures," facilitating rapid navigation. Bulleted lists, tables,

ISSN 2616-5783 Vol.8, Issue 11: 140-157, DOI: 10.25236/AJHSS.2025.081122

and concise headings made it easy to pinpoint critical details, especially when time was of the essence, such as before a field excursion or lab experiment. Users also valued the automatic generation of checklists and risk assessment templates embedded within the chat, which eliminated the need to create documents from scratch. The avoidance of jargon-laden paragraphs and the use of plain language further enhanced comprehension, even for those with limited safety training.

"When I asked about xylene, the chatbot (ArchHealth) gave me a table of PPE options and bullet points for spill response, so clear I could copy it directly into my lab notes." (P2)

"It generated a risk assessment template right in the chat. I didn't have to draft anything, just filled in the blanks. Instead of reading a 20-page manual, its two-page summary with bullets was exactly what I needed." (P10)

"The plain-language explanations, with numbered steps and subheadings, helped me feel sure I wasn't overlooking anything." (P11)

3.2.1.3 Rapid expansion and follow-up prompting

Participants appreciated ArchHealth's capacity to offer follow-up questions and dynamically expand on initial responses. After delivering a basic overview, such as general lab safety, ArchHealth frequently asked whether users wanted more details on disposal methods, local regulations, or sample documentation. This proactive prompting reduced the need for users to formulate every subsequent question, effectively guiding them through complex decision trees. Even when users posed vague or broad inquiries, ArchHealth suggested clarifications (e.g., specifying whether guidance was needed for museum storage versus excavation) to hone in on the precise context. Overall, this feature facilitated thorough exploration of safety topics, enabling users to build comprehensive plans through a structured, conversational workflow.

"It gave me basic respirator info, then asked if I needed a filtration chart. When I said yes, I got a detailed N95 vs. P100 comparison." (P11)

"I got a lab safety overview, then the chatbot (ArchHealth) offered to draft a risk assessment. Suddenly I had a solid template ready to customize... Even when I didn't know what to ask next, it (ArchHealth) suggested areas to explore, like 'Interested in field vs. lab PPE?', so I never felt lost." (P14)

3.2.1.4 Broad "safe entry" for non-experts

Novice users, particularly students and those new to conservation, emphasized that ArchHealth provided a welcoming, accessible introduction to occupational health concepts they might otherwise hesitate to explore. Participants with limited prior knowledge felt comfortable asking basic questions (e.g., "What PPE for solvent cleaning?" "How to handle moldy artifacts?") without fear of being judged or overwhelmed by technical jargon. ArchHealth's plain-language tone and step-by-step instructions demystified complex procedures, facilitating a gradual learning curve. Several users noted that before ArchHealth, they relied on generic internet searches that yielded inconsistent or low-quality results; in contrast, ArchHealth offered reliable, context-relevant guidance that built their confidence. Even experienced professionals sometimes used ArchHealth to validate their assumptions or discover lesser-known hazards.

"As a first-year conservator, I didn't know which solvents were hazardous. ArchHealth explained everything in plain English, so I didn't feel intimidated. Before, I'd Google 'artifact cleaning safety' and get conflicting advice. It (ArchHealth) gave me a clear, concise answer that I trusted." (P2)

"Even though I'm experienced, I still ask the chatbot (ArchHealth) to confirm rare procedures, like laser cleaning, because it explains things so clearly." (P12)

"I work part-time and hadn't dealt with erosion control. ArchHealth's explanation felt like a patient teacher rather than a textbook." (P14)

3.2.2 User journey and adoption patterns

Participants' engagement with ArchHealth typically unfolded in distinct stages: initial curiosity, exploratory experimentation, gradual trust-building, and eventual integration into routine workflows. Early impressions were formed by probing simple, low-stakes questions, such as general lab hazards or basic PPE recommendations, to evaluate ArchHealth's accuracy and relevance. Positive initial experiences, where accurate and context-specific guidance was provided, encouraged further exploration. As confidence grew, users began consulting ArchHealth more frequently, triggered by concrete tasks (e.g., prepping for excavation, drafting risk assessments) or unexpected workplace concerns (e.g.,

ISSN 2616-5783 Vol.8, Issue 11: 140-157, DOI: 10.25236/AJHSS.2025.081122

wildfire smoke infiltration, solvent shortages). Over time, many transitioned from sporadic use to habitual reliance, viewing ArchHealth as a virtual safety advisor on par with institutional experts. However, some remained cautious, continuing to use ArchHealth primarily for preliminary research before seeking human expert confirmation. Understanding these patterns illuminates how users move from tentative experimentation to habitual integration, as detailed in the following sub-themes.

3.2.2.1 Early experiences and first impressions

Initial user experiences with ArchHealth were characterized by cautious optimism. Many participants, already familiar with general-purpose ChatGPT, wondered whether a specialized safety version could truly match or exceed established resources. To test its reliability, they posed broad, non-critical questions (e.g., "What are common lab solvents' hazards?" "How to store flammable materials?") and assessed response accuracy against known standards or institutional guidelines. When ArchHealth provided clear, actionable, and accurate information on the first try, often generating entire risk assessment templates, users reported a surge in confidence that propelled them to further experimentation.

"I asked a simple question about ethanol hazards, and when it (ArchHealth) gave me a detailed PPE list, I thought, 'Okay, this might really work.'" (P3)

"My first query was 'soil dust exposure.' ArchHealth's instructions matched what my safety officer taught me, so I was immediately more confident." (P7)

"I tested it with a niche question, 'laser ablation safety.' It gave me more detail than I expected, and I was hooked after that." (P10)

Conversely, a few participants encountered overly generic or truncated answers during these initial trials, prompting them to refine their queries or compare results with other sources. Overall, this sub-theme highlights how first impressions, shaped by ArchHealth's ability to deliver credible guidance on a user's very first query, were critical to fostering continued engagement.

"At first, I was skeptical. I asked about respirators and got a brief answer. I had to rephrase to 'archaeology lab respirator guidance' to see its true depth." (P5)

3.2.2.2 Usage frequency and contextual triggers

As users transitioned beyond initial trials, ArchHealth usage became event-driven, tied to specific workplace tasks or concerns. Common triggers included: preparing for an upcoming excavation (prompting questions such as "What PPE for hot, dusty trenches?"), designing new laboratory protocols (e.g., "Draft a solvent disposal procedure"), or responding to sudden onsite issues (such as "How to mitigate wildfire smoke exposure?"). Participants noted that ArchHealth often became part of their pre-task checklist, consulted before each new experiment, artifact treatment, or field deployment. Some users even accessed ArchHealth during off-hours to preemptively learn about best practices for upcoming projects. Usage frequency varied, ranging from daily check-ins by those with heavy lab schedules to weekly consultations by field archaeologists between digs. Overall, ArchHealth's adoption was highly contextual, driven by real-world demands that shaped both the cadence and content of user interactions.

"I consult the chatbot (ArchHealth) right before setting up a new fume hood experiment to double-check my solvent disposal plan." (P1)

"Before a big dig, I'd ask ArchHealth every morning, 'What hazards to expect in this season's excavation?'" (P9)

"I don't use it (ArchHealth) every day, but whenever I start a new project, like laser cleaning, it's my first stop." (P4)

"Even on weekends, I'd ask ArchHealth for tips on storing artifacts in humid climates. It's become my go-to resource." (P12)

3.2.2.3 Learning curve and onboarding

While participants lauded ArchHealth's intuitive interface, they still faced a brief learning period to optimize question framing and interpret responses effectively. Early in their journey, users discovered that broad or ambiguous queries could yield generic answers, prompting them to rapidly iterate on question phrasing (e.g., specifying environment, chemical concentration, or local regulations). Feedback loops, where ArchHealth asked follow-up questions, sped up this onboarding, helping users refine inputs. A handful of participants suggested that an introductory walkthrough or built-in example prompts would further flatten the learning curve, enabling newcomers to understand how to extract maximal value

ISSN 2616-5783 Vol.8, Issue 11: 140-157, DOI: 10.25236/AJHSS.2025.081122

quickly. Once proficient, most users reported greater efficiency, as they naturally incorporated context-specific details (such as "Working in New South Wales, Australia, during summer") into their queries to unlock precise guidance.

"At first, my query 'lab hazards' was too vague. After a couple tries, I learned to say 'archaeology conservation lab hazards,' and the answers got so much better." (P6)

"I wish there were sample questions when I first logged in. It would've saved me trial and error instead of guessing how to ask... the chatbot (ArchHealth)'s follow-ups like 'Which solvent concentration?' taught me how to frame queries properly within a few chats." (P13)

"Once I started including specifics, like '4% formaldehyde in humid climates', I realized how custom the advice could be... After about three tries, I understood that mentioning local standards, like OSHA vs. EU regs, got me more region-specific answers." (P14)

3.2.3 Adoption drivers and barriers

Participants identified a combination of motivating factors and obstacles that influenced whether and how they embraced ArchHealth. Primary drivers included substantial time savings (collapsing hours of manual research into seconds), comprehensive coverage of multi-domain safety topics, and the ability to generate project-specific templates and checklists on demand. Trust-building elements, such as citing reputable sources when available and presenting logically organized responses, further encouraged adoption. Conversely, barriers emerged around concerns about context-specificity (e.g., local regulatory compliance, language preferences), occasional generic or incomplete answers, and the persistent need for human expert oversight in high-stakes scenarios. Organizational requirements, such as mandatory sign-offs from certified safety officers, also limited ArchHealth's utility as a standalone solution. Together, these drivers and barriers shaped participants' propensity to adopt ArchHealth, revealing a nuanced interplay between perceived utility and structural constraints. The sub-themes below elaborate on these factors.

3.2.3.1 Time-savings and practical utility

A dominant motivator for adoption was ArchHealth's ability to condense what previously took hours of research into a matter of seconds. Participants emphasized that tasks such as compiling hazard lists, drafting risk assessments, and locating relevant PPE guidelines, traditionally involving cross-referencing multiple documents and consulting experts, were now accomplished almost instantaneously. This time efficiency was particularly valuable for field archaeologists on tight excavation schedules and students juggling course workloads. Even when final approval by a safety officer was required, ArchHealth provided a robust starting point that greatly accelerated drafting processes. Several users noted that, in emergency situations (e.g., unexpected site contamination), rapid access to tailored guidance could mean the difference between a delayed response and immediate mitigation. As such, time-savings emerged as a powerful driver that spurred continued reliance on ArchHealth for practical, real-world tasks.

"I needed a solvent disposal plan before tomorrow's demo. ArchHealth gave me one in five seconds, I'd have spent two hours otherwise." (P2)

"Compiling a PPE list for dusty trenches used to take me half a day. Now I type 'dusty excavation PPE,' and I'm done." (P9)

"When I discovered a chemical spill, its (ArchHealth's) cleanup protocol saved me so much time; I could focus on the site rather than paperwork." (P5)

"Even drafting a basic risk assessment used to involve emailing three experts. It (ArchHealth) did it in one chat." (P12)

3.2.3.2 Trust-but-verify mindset

While ArchHealth's outputs were generally perceived as accurate, many participants adopted a "trust-but-verify" approach, using ArchHealth to gather baseline information, which they then cross-checked against official guidelines or consulted with experienced colleagues. For high-stakes decisions (e.g., working with unstable structures or handling pressurized gases), users felt that AI-generated advice was a helpful first step but not a substitute for human expertise. This cautious stance meant that ArchHealth functioned as a complementary tool rather than a sole authority. Participants described scenarios where ArchHealth's recommendations aligned perfectly with institutional policies, reinforcing trust, but also recounted occasional instances of conflicting or overly generalized advice that prompted verification. The "trust-but-verify" mindset thus served as both an adoption driver, since users felt comfortable

ISSN 2616-5783 Vol.8, Issue 11: 140-157, DOI: 10.25236/AJHSS.2025.081122

exploring ArchHealth's guidance, and a barrier, as it limited blind reliance and ensured continued human oversight in critical contexts.

"ArchHealth told me to use nitrile gloves for solvent work. I still checked the MSDS to confirm. It was a starting point, not the final word." (P1)

"For field shoring, I asked ArchHealth, then compared its advice to our university's excavation handbook. They matched, so I felt confident... Sometimes the chatbot (ArchHealth)'s generic wording makes me pause, like when it didn't specify a country's legal limit. So I always verify those details." (P7)

"I used ArchHealth for a respirator guide, but double-checked with our safety officer before actual purchase." (P8)

3.2.3.3 Localization and jurisdictional specificity

A recurring barrier to unqualified adoption was ArchHealth's occasional failure to account for region-specific regulations, permissible exposure limits, and best-practice standards. Participants operating in different countries, such as Australia, China, and various European nations, emphasized that local legal frameworks could differ significantly from general guidelines. When ArchHealth provided generic advice (e.g., defaulting to US OSHA limits), users felt compelled to verify or adjust recommendations to align with their jurisdiction. Bilingual participants also noted the benefit of dual-language terminology for chemicals and safety terms to reduce translation errors. The absence of precise localization sometimes undermined trust, particularly in contexts where noncompliance could carry legal or financial penalties. Thus, localization emerged as a critical factor that, when absent, limited ArchHealth's authority and adoption in diverse settings.

"For EU-based projects, its (ArchHealth's) generic references forced me to dig through my country's directives. It's a useful start, but not the final guide... When it (ArchHealth) cited US respirator standards, I had to manually find my local equivalent. That extra step sometimes delayed work." (P6)

"The chatbot (ArchHealth) suggested OSHA exposure limits, but in Australia our HWL values differ. I had to convert numbers myself." (P10)

"I wished the chemical names appeared in both English and Mandarin. A translation slip could have major consequences." (P13)

"Local waste disposal rules in China differ from what ArchHealth mentioned. I had to search the local government site afterward." (P14)

3.2.3.4 Preference for human expert oversight

Despite recognizing ArchHealth's strengths, many participants maintained that final safety decisions should involve a qualified human expert, such as a safety officer, lab manager, or senior conservator. Users felt that while ArchHealth offered reliable baseline recommendations, it could not fully replicate the nuanced judgment derived from years of field experience or specialized training. In scenarios involving unpredictable environmental factors (e.g., seismic activity at excavation sites) or fragile artifacts with unique conservation needs, human oversight was viewed as indispensable. This preference meant that ArchHealth's outputs were often used to inform discussions with supervisors rather than to replace them. While this barrier limited full reliance on the tool, it also established a productive collaboration: ArchHealth generated initial drafts and checklists, which were then refined and approved by human experts.

"On a dig in a region prone to flash floods, ArchHealth gave generic advice, but my field mentor's local insights were crucial. I trust the AI (ArchHealth) for mundane tasks, but when dealing with rare artifacts, I still run plans by a senior conservator." (P2)

"For laser cleaning an ancient manuscript, I used ArchHealth for basic steps, but my head conservator had to sign off on the protocol... ArchHealth's roadmap for solvent cleanup was great, but my lab manager tweaked it based on our facility's equipment layout." (P5)

"It's like consulting a knowledgeable intern. Helpful, but I wouldn't finalize anything without a pro supervising." (P11)

3.2.4 Feature and User Interface/User Experience (UI/UX) enhancements desired

Although ArchHealth's core functionality was widely praised, participants suggested several enhancements to better meet their needs. They wanted integrated multimedia, diagrams, photos, and videos, to demonstrate PPE use, ergonomics, and site setups alongside text advice. Field archaeologists

ISSN 2616-5783 Vol.8, Issue 11: 140-157, DOI: 10.25236/AJHSS.2025.081122

emphasized the importance of a mobile-friendly interface with voice input and offline caching for remote work. Improved localization, regional regulatory modules and bilingual terminology, was essential for diverse jurisdictions. Users also envisioned interactive, scenario-based workflows (e.g., decision trees for specific tasks) and a quick-access prompt library or toolbar with sample questions to simplify queries. Implementing these features would expand ArchHealth's appeal and usability across archaeological and conservation contexts.

3.2.4.1 Multimedia integration (diagrams, photos and videos)

Participants emphasized that visual aids, such as diagrams of safe excavation postures, photos of proper respirator use, and short videos on solvent disposal, are essential alongside ArchHealth's text. For spatially complex tasks (e.g., scaffold setup or laser cleaning), written instructions alone can be unclear. Embedding annotated images or brief clips within ArchHealth would help users translate guidance into safe practice without constantly leaving the app. Visual content also aids training, allowing students and new hires to follow step-by-step demonstrations rather than relying solely on text.

"A quick diagram showing how to tie off a harness for cliffside digs would be so much clearer than text alone. And seeing a short video of proper respirator seal checks would remove any guesswork when I'm in bright sunlight on site." (P3)

"Demonstrating how to lift heavy crates of artifacts with correct posture, an embedded photo or gif, would be invaluable." (P6)

"If it (ArchHealth) could show a 10-second clip of how to fill out a risk form, new interns wouldn't struggle with formatting." (P12)

3.2.4.2 Mobile and offline accessibility

Field archaeologists and traveling conservators stressed the need for a robust mobile interface suited to on-site use. They requested a responsive design for smartphones or tablets, with larger buttons and thumb-friendly menus for when typing is difficult (e.g., wearing gloves or in bright sunlight). Equally vital was an offline mode or lightweight cache to store accessed protocols and checklists, ensuring critical guidance remains available in remote areas with poor connectivity. Many envisioned a dedicated ArchHealth mobile app that syncs when online but retains core safety modules offline. Such improvements would allow field teams to access safety guidance anytime, reducing risks when lab or field conditions shift unexpectedly.

"On a dig site with zero bars, I had to wait until evening to check the chatbot (ArchHealth). Offline access would eliminate be great." (P11)

"If the chatbot (ArchHealth) cached my last five checks, I could use them in the field without worrying about connection." (P8)

"When I'm in a remote cave dig, I need safety protocols instantly. Offline mode would make it (ArchHealth) truly field-ready." (P10)

4. Discussion

Our study explored archaeological professionals' experiences with ArchHealth, a specialized AI-driven OSH guidance tool tailored to the unique demands of heritage work. The principal findings can be distilled into four interrelated themes: the perceived benefits of ArchHealth, users' journeys and adoption patterns, the drivers and barriers influencing adoption, and desired feature and UI/UX enhancements. Overall, participants expressed enthusiasm for ArchHealth's ability to centralize diverse safety information, ranging from chemical hazards to field excavation risks, into a single, coherent resource, thereby reducing the fragmentation and time-intensive nature of traditional OHS research. Early engagement was characterized by cautious experimentation, where positive initial experiences fostered increased trust and integration into routine workflows. Nonetheless, users maintained a "trust-but-verify" stance, reflecting the need to cross-check AI-generated advice against local regulations and institutional protocols. Adoption drivers included substantial time savings and practical utility, whereas barriers encompassed concerns about jurisdictional specificity, occasional generic responses, and the enduring necessity of human expert oversight. Finally, participants articulated a clear vision for future iterations of ArchHealth: multimedia integration, mobile and offline accessibility, enhanced localization, and interactive decision-making workflows. In what follows, we unpack these findings, situating them within existing literature, and offer possible explanations for observed patterns.

Users consistently highlighted that ArchHealth's comprehensiveness and structured output constituted a marked improvement over conventional OHS research methods, which often require navigating disparate manuals, safety data sheets, and regulatory websites. This reflects broader trends in digital health literature, where centralized, context-specific decision support systems have been shown to improve both efficiency and user confidence[34-36]. In our sample, novices particularly valued ArchHealth's step-by-step, jargon-free guidance, which aligns with findings by Cockburn, A. et al.[37] on the importance of user-friendly interfaces for non-expert adoption. More experienced professionals, by contrast, appreciated ArchHealth as a rapid validation tool, enabling them to uncover obscure hazards, such as latent mold growth in artifact storage or ultraviolet (UV) exposure during photogrammetry, that might otherwise be overlooked in domain-specific manuals. This dual appeal across experience levels suggests that ArchHealth effectively addresses both knowledge gaps and confirmation needs. The functions parallel roles identified in decision support literature, wherein systems both educate novices and augment experts[38-40]. Such a dual function likely contributes to ArchHealth's broad acceptability and reinforces that AI-driven safety tools can serve as both pedagogical and advisory resources[41, 42].

The second key theme, user journey and adoption patterns, reveals a recognizable progression from initial curiosity to habitual integration, punctuated by phases of trust-building and ongoing learning. Many participants, already familiar with general-purpose ChatGPT, began with low-stakes inquiries, which served as cognitive probes to gauge ArchHealth's domain expertise. When initial responses aligned with users' prior knowledge or institutional guidelines, confidence grew, prompting more complex, context-specific queries. This pattern echoes the Technology Acceptance Model (TAM) and its extensions, which posit that perceived ease of use and perceived usefulness drive user adoption[43-45]. Indeed, participants reported that once ArchHealth demonstrated accuracy and relevance in early trials, they felt comfortable relying on it for tasks such as drafting risk assessment templates or planning excavation PPE. The importance of positive first impressions is supported by Ursavaş and colleagues, who found that initial system performance strongly influences long-term adoption of AI tools in safety-critical domains[46].

However, the journey was not entirely frictionless. Participants described a brief learning curve during onboarding, wherein they had to learn how to optimize query framing to elicit the most precise advice. Vague or overly broad questions yielded generic responses, compelling users to iterate toward more specific phrasings. This underscores the interplay between user literacy and system design. While ArchHealth's follow-up prompting helped refine queries, prompting users to specify concentration levels or local regulations, some argued that a built-in tutorial or example prompts could accelerate proficiency. Such onboarding assistance aligns with best practices in human-computer interaction, which emphasize guided tours, engineered prompts, and context-aware suggestions to help users find the right "mental model" for interacting with AI[47, 48].

Regarding adoption drivers and barriers, time savings emerged as the most powerful motivator. Participants highlighted that tasks once requiring hours of cross-referencing multiple sources, compiling PPE lists, drafting solvent disposal protocols, or generating hazard checklists, were now accomplished in seconds. These findings align with a growing body of work demonstrating that digital decision support can drastically reduce cognitive workload and accelerate planning and decision-making[49-51]. Particularly in emergency scenarios, such as chemical spills or sudden wildfire smoke exposure, rapid access to reliable guidance can be critical, as noted by participants who contrasted ArchHealth's five-second risk assessment generation with traditional two-hour manual drafting. Time efficiency was especially valued by field archaeologists operating under tight excavation schedules or conservators balancing academic workloads, indicating that ArchHealth's value proposition extends across roles and contexts.

Nevertheless, a persistent "trust-but-verify" attitude tempered full reliance on ArchHealth. While participants generally found the advice accurate, they remained cautious about delegating final decisions to an AI. This cautious stance aligns with findings in other studies of AI-assisted decision support, where users often view AI recommendations as a helpful preliminary step but not a substitute for human expertise, especially in high-stakes contexts[52, 53]. For example, when ArchHealth provided default Occupational Safety and Health Administration (OSHA) exposure limits, participants operating in Australia or the European Union had to cross-reference local permissible exposure limits, which sometimes differed substantially. Similarly, participants in China emphasized the need for bilingual terminology to avoid translation errors when handling hazardous chemicals. The concerns reflect a broader challenge in AI deployment: ensuring localization and jurisdictional specificity. When AI tools default to U.S.-centric standards, they risk alienating users in other regulatory environments. Thus, while ArchHealth's comprehensive coverage was applauded, its occasional generic or missing localization eroded trust in contexts where legal or financial penalties could result from noncompliance.

Complementing localization concerns was the preference for human expert oversight. Despite ArchHealth's ability to generate baseline templates, many users conducted final reviews with certified safety officers, lab managers, or experienced conservators. This dynamic underscores that AI serves as an "augmented intelligence" rather than a replacement for professional judgment[54, 55]. In scenarios involving unpredictable environmental factors, such as seismic activity at excavation sites, or fragile artifacts requiring bespoke conservation methods, participants felt that AI could not match the nuanced situational awareness of seasoned practitioners. This finding resonates with prior work suggesting that human-AI collaboration yields optimal outcomes when each party's unique strengths are leveraged: AI for rapid, structured information retrieval and pattern recognition; humans for domain-specific contextual judgment and ethical considerations[56]. Ultimately, ArchHealth functioned as a complementary resource, offering initial drafts and checklists that human experts validated, adapted, and approved.

The final theme, feature and UI/UX enhancements, reflects participants' aspirations for ArchHealth's evolution. Multimedia integration topped the list, with users emphasizing that visual aids, such as annotated diagrams of excavation postures or photos of correct respirator donning, would bridge the gap between abstract guidance and field practice. This preference parallels research on safety training, which consistently finds that multimodal instruction (text + visuals + video) enhances comprehension and retention[57, 58]. Embedding short video clips or interactive diagrams within ArchHealth could reduce cognitive load and minimize the ambiguity inherent in purely text-based protocols. Moreover, participants noted that seamlessly integrated multimedia would support on-the-job training for students and new hires, providing step-by-step demonstrations that textual instructions alone cannot convey.

Another demand was for robust mobile and offline functionality. Field archaeologists and traveling conservators underscored that intermittent or absent connectivity in remote excavation sites renders web-only solutions impractical. Participants envisioned a dedicated ArchHealth mobile app, with responsive design elements, larger buttons, thumb-friendly navigation, and voice input, to accommodate challenging environmental conditions like wearing gloves or working under bright sunlight. Offline caching or lightweight storage of recently accessed protocols would enable uninterrupted access to critical guidance. The requests echo broader trends in mobile computing for rugged environments, where integration of an offline design and adaptive interfaces have been shown to improve usability and resilience[59, 60]. Implementing offline-capable features would not only expand ArchHealth's accessibility but also underscore its role as a mission-critical tool rather than a convenience application.

5. Implications and Recommendations for Practice

Our findings suggest several actionable steps for future practitioners and developers of ArchHealth and similar AI-driven chatbots. First, institutions should integrate ArchHealth into safety training and pre-task checklists, using its plain-language guidance to build foundational knowledge and serve as a rapid reference for both novices and experienced professionals. Second, developers must prioritize multimedia integration, adding annotated diagrams, photos, and brief videos, to translate abstract safety recommendations into clear, on-site practices, since multimodal instruction improves comprehension of complex tasks. Third, robust mobile and offline functionality is essential for fieldwork in remote locations; an "offline-first" app with responsive navigation and voice input would ensure continuous access to critical protocols. Fourth, localization should be enhanced by including region-specific regulatory modules and bilingual terminology to minimize translation errors and reduce the need for manual adjustments. Finally, onboarding can be smoothed by providing sample prompts and scenario-based decision trees, guiding users toward effective query phrasing to quickly extract context-specific guidance

6. Limitations

This study has several limitations. First, the sample exhibited a skewed gender distribution, with most participants being female, and a narrow age range concentrated in the late twenties and early thirties. Second, half of the participants were students, with relatively few mid-career or senior professionals, which may have shaped perspectives on trust, verification behaviors, and integration of ArchHealth into established professional routines. Third, recruitment was conducted through the Melbourne team's professional network, which may have introduced selection bias by favoring early adopters with more positive attitudes toward technology. Fourth, all participants were drawn from the Melbourne

ISSN 2616-5783 Vol.8, Issue 11: 140-157, DOI: 10.25236/AJHSS.2025.081122

archaeological community, limiting geographical diversity and potentially reflecting local organizational norms and cultural contexts. These may constrain the transferability of our findings. Future research should aim to recruit a more balanced sample across gender, age, experience level, and regional settings, and to use broader recruitment strategies to minimize bias and enhance the representativeness of user experiences. In addition, LLMs such as ChatGPT carry inherent technical risks, including correctness of response, the potential for hallucinated or incomplete responses, sensitivity to prompt phrasing, and the need for continuous updating and maintenance warrant systematic quantitative evaluation in future research.

7. Conclusions

ArchHealth demonstrates that a customized AI-driven chatbot can effectively bridge gaps in archaeological occupational health by centralizing diverse safety information, streamlining risk assessment processes, and fostering both novice learning and expert validation. Users reported substantial time savings and appreciated its structured, context-specific guidance, while maintaining a “trust-but-verify” approach that underscores the need for localized content and human oversight. Although Barriers such as regional regulatory specificity and interface enhancements remain, the positive user experiences suggest that AI tools such as ArchHealth can complement traditional safety practices, augmenting compliance and decision-making in archaeology. The findings support further refinement and broader deployment of domain-specific chatbots to enhance field, lab, and museum safety workflows.

Data Availability

The original interview recordings are not to be provided for privacy protection considerations. De-sensitized data analyzed can be provided on reasonable request to the correspondence author.

Authors' Contributions

MZ created the ArchHealth chatbot, conceptualized the study, and collected data. MZ and YZuo designed the study. DL, CH, and YZhao curated data. MZ, DL, CH, YZhao, ZT, HX, JZ, and YZuo performed formal analysis of data. MZ, DL, and ZT prepared the initial draft of manuscript. JZ and YZuo critically reviewed the final version. All authors have reviewed and approved the version for submission.

References

[1] Wang Y J, Chen H, Liu B, et al. A Systematic Review on the Research Progress and Evolving Trends of Occupational Health and Safety Management: A Bibliometric Analysis of Mapping Knowledge Domains [J]. Frontiers in Public Health, 2020(8).

[2] Koh D & Gan W H, Occupational Health[M]. Oxford Textbook of Global Public Health, 2021.

[3] Emerson M, Toward a Safe Archaeology Field School [J]. Advances in Archaeological Practice, 2021(9): 66-73.

[4] White W, Applying Risk Management Concepts from CRM and the Outdoor Recreation Industry to Academic Archaeology Projects [J]. Advances in Archaeological Practice, 2021(9): 61-65.

[5] Klehm C, Hildebrand E, and M S Meyers. Mitigating Chronic Diseases during Archaeological Fieldwork: Lessons from Managing Asthma, Diabetes, and Depression [J]. Advances in Archaeological Practice, 2021(9): 41-48.

[6] Kintigh K W, Jeffrey H A, Mary C B, et al. Grand challenges for archaeology [J]. Proc Natl Acad Sci U S A, 2014(111): 879-80. (in English)

[7] Ferreira M R, Brito A V, and Fernandes R J, Occupational Diseases in Art Conservators and Restorers: A Systematic Review [J]. Healthcare (Basel), 2025(13).

[8] Shaikh T , Rasool T, Veningston K and Yaseen S M. The role of large language models in agriculture: harvesting the future with LLM intelligence [J]. Progress in Artificial Intelligence, 2024.

[9] Yang X, Li T, Su Q, et al. Application of large language models in disease diagnosis and treatment [J]. Chinese Medical Journal, 2024(138): 130-142.

[10] Niyonkuru E, Caufield J H, Carmody L C, et al. Leveraging Generative AI to Accelerate Biocuration of Medical Actions for Rare Disease [J]. medRxiv, 2024.

[11] Siino M, Falco M, Croce D, and Rosso P. Exploring LLMs Applications in Law: A Literature Review on Current Legal NLP Approaches [J]. IEEE Access, 2025(13): 18253-18276.

[12] Padiu B, Iacob R, Rebedea T, and M Dascalu. To What Extent Have LLMs Reshaped the Legal Domain So Far? A Scoping Literature Review [J]. Information, 2024.
[13] Taeb M, Chi H, and S Bernadin. Assessing the Effectiveness and Security Implications of AI Code Generators [J]. Journal of The Colloquium for Information Systems Security Education, 2024.
[14] Yan L, Sha L, Zhao L,et al. Practical and ethical challenges of large language models in education: A systematic scoping review [J]. Br. J. Educ. Technol., 2023(55): 90-112.
[15] Alqahtani T, Badreldin H A, Alrashed M, et al.The Emergent Role of Artificial Intelligence, Natural Learning Processing, and Large Language Models in Higher Education and Research [J]. Research in social & administrative pharmacy : RSAP, 2023.
[16] Bouchikhi M El , Weerts S , and Clavien C. Behind the good of digital tools for occupational safety and health: A Scoping Review of Ethical Issues Surrounding the Use of the Internet of Things [J]. Frontiers in Public Health, 2024(12). (in English)
[17] Letuka S. Management of the Occupational Health and Safety System in Public Enterprises According to ISO 45001 [J]. Uprava, 2024(15): 49-62.
[18] Park J S, Lee D G, Jimenez J A, Lee S J and J-W Kim. Human-Focused Digital Twin Applications for Occupational Safety and Health in Workplaces: A Brief Survey and Research Directions [J]. Applied Sciences, 2023(13): 4598.
[19] Emerson M C. Toward a Safe Archaeology Field School: Insights into Policies, Procedures, and Team-Based Learning [J]. Advances in Archaeological Practice, 2021(9): 66-73.
[20] Doyle L, McCabe C, Keogh B, Brady A and McCann M. An overview of the qualitative descriptive design within nursing research [J]. Journal of Research in Nursing, 2019(25): 443-455.
[21] Villamin P, Lopez V, Thapa D and Cleary M. A Worked Example of Qualitative Descriptive Design: A Step-by-Step Guide for Novice and Early Career Researchers [J]. Journal of Advanced Nursing, 2024.
[22] Kim H, Sefcik J and Bradway C. Characteristics of Qualitative Descriptive Studies: A Systematic Review [J]. Research in Nursing & Health, 2017(40).
[23] Willis D, Sullivan-Bolyai S, Knafl K and Cohen M. Distinguishing Features and Similarities Between Descriptive Phenomenological and Qualitative Description Research [J]. Western Journal of Nursing Research, 2016(38): 1185-1204.
[24] Campbell S, Greenwood M, Prior S, et al. Purposive Sampling: Complex or Simple? Research Case Examples [J]. Journal of Research in Nursing, 2020(25): 652-661.
[25] Saunders B, Sim J, Kingstone T, et al. Saturation in Qualitative Research: Exploring Its Conceptualization and Operationalization [J]. Quality & Quantity, 2017(52): 1893-1907.
[26] Guest G, Namey E and Chen M. A Simple Method to Assess and Report Thematic Saturation in Qualitative Research [J]. PLoS ONE, 2020(15).
[27] Drisko J. Transferability and Generalization in Qualitative Research [J]. Research on Social Work Practice, 2024.
[28] Edward K & Welch T. The extension of Colaizzi's method of phenomenological enquiry [J]. Contemporary Nurse, 2011(39): 163-171.
[29] Abalos E, Rivera R, Locsin R and Schoenhofer S, Husserlian Phenomenology and Colaizzi's Method of Data Analysis: Exemplar in Qualitative Nursing Inquiry Using Nursing As Caring Theory [J]. 2016(20): 19-23.
[30] Mays N & Pope C. Rigour and Qualitative Research [J]. BMJ, 1995(311): 109-112.
[31] Cypress B. Rigor or Reliability and Validity in Qualitative Research: Perspectives, Strategies, Reconceptualization, and Recommendations [J]. Dimensions of Critical Care Nursing, 2017(36): 253.
[32] Morse J. Critical Analysis of Strategies for Determining Rigor in Qualitative Inquiry [J]. Qualitative Health Research, 2015(25): 1212-1222.
[33] Krefting L. Rigor in Qualitative Research: The Assessment of Trustworthiness [J]. The American Journal of Occupational Therapy, 1991(45): 214-222.
[34] Kawamoto K & McDonald C. Designing, Conducting, and Reporting Clinical Decision Support Studies: Recommendations and Call to Action [J]. Annals of Internal Medicine, 2020(172).
[35] Bertl M, Ross P, and Draheim D. Systematic AI Support for Decision-making in the Healthcare Sector: Obstacles and Success Factors [J]. Health Policy and Technology, 2023.
[36] Perfalk E, Bernstorff M, Danielsen A and Østergaard S. Patient Trust in the Use of Machine Learning-based Clinical Decision Support Systems in Psychiatric Services: A Randomized Survey Experiment [J]. European Psychiatry, 2024(67).
[37] Cockburn A, Gutwin C, Scarr J and Malacria S. Supporting Novice to Expert Transitions in User Interfaces [J]. ACM Computing Surveys (CSUR), 2014(47): 1-36.
[38] Nah F & Benbasat I. Knowledge-based Support in a Group Decision Making Context: An Expert-Novice Comparison [J]. J. Assoc. Inf. Syst., 2004(5): 5.
[39] Zohrevandi E, Westin C A L, Vrotsou K and Lundberg J. Exploring Effects of Ecological Visual

ISSN 2616-5783 Vol.8, Issue 11: 140-157, DOI: 10.25236/AJHSS.2025.081122

Analytics Interfaces on Experts' and Novices' Decision-Making Processes: A Case Study in Air Traffic Control [J]. Computer Graphics Forum, 2022(41): 453-464.
[40] Wang X & Liu X Q. Potential and limitations of ChatGPT and generative artificial intelligence in medical safety education [J]. World J Clin Cases, 2023(32): 7935-7939.
[41] El-Helaly M. Artificial Intelligence and Occupational Health and Safety, Benefits and Drawbacks [J]. Med Lav, 2024(115): e2024014.
[42] Pishgar M, Issa S F, Sietsema M, Pratap P and Darabi H. REDECA: A Novel Framework to Review Artificial Intelligence and Its Applications in Occupational Safety and Health [J]. Int J Environ Res Public Health, 2021(18): 13.
[43] Gupta S, Abbas A F & Srivastava R. Technology Acceptance Model (TAM): A Bibliometric Analysis from Inception [J]. Journal of Telecommunications and the Digital Economy, 2022(10): 77-106.
[44] Granić A & Marangunić N. Technology acceptance model in educational context: A systematic literature review [J]. British Journal of Educational Technology, 2019(50): 2572-2593.
[45] Davis F D. User acceptance of information technology: System characteristics, user perceptions and behavioral impacts [J]. International Journal of Man-Machine Studies, 1993(38): 475-487.
[46] Ursavaş Ö F, Yalçın Y, İslamoğlu H, Bakır-Yalçın E and Cukurova M. Rethinking the importance of social norms in generative AI adoption: investigating the acceptance and use of generative AI among higher education students [J]. Int J Educ Technol High Educ, 2025(22): 38.
[47] Cain W. Prompting Change: Exploring Prompt Engineering in Large Language Model AI and Its Potential to Transform Education [J]. TechTrends, 2023(68): 47-57.
[48] Singh V M, Aryasomayajula S A, Chatterjee A, Aydemir B & Amin R M. Improving AI-generated music with user-guided training [J]. ArXiv, 2025: abs/2506.04852.
[49] Naderpour M, Lu J & Zhang G. An Intelligent Situation Awareness Support System for Safety-Sritical Environments [J]. Decis. Support Syst., 2014(59): 325-340.
[50] Cox S, Caldwell N, Sevigny N, et al. Augmented Reality Clinical Decision Support Decreases Cognitive Workload on Surgical Tasks [C]. 2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR), 27 Jan 2025. pp. 394-398.
[51] Moncur B, Trigo M G & Mortara L. Augmented Reality to Reduce Cognitive Load in Operational Decision-Making [C]. 25th HCI International Conference. 2023.
[52] Inkpen K, Chappidi S, Mallari K, et al. Advancing Human-AI Complementarity: The Impact of User Expertise and Algorithmic Tuning on Joint Decision Making [J]. ACM Transactions on Computer-Human Interaction, 2022(30): 1-29.
[53] Zhang Q, Lee M & Carter S. You Complete Me: Human-AI Teams and Complementary Expertise [C]. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 27 Apr 2022.
[54] Lebovitz S,Lifshitz-Assaf H & Levina N. To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis [J]. Materials Engineering eJournal, 2021.
[55] Pasquale F. Professional Judgment in an Era of Artificial Intelligence and Machine Learning [J]. Boundary 2, 2019.
[56] Vaccaro M, Almaatouq A & Malone T. When Combinations of Humans and AI are Useful: A Systematic Review and Meta-Analysis [J]. Nat Hum Behav, 2024(8): 2293-2303.
[57] Luo H. Editorial: Advances in Multimodal Learning: Pedagogies, Technologies, and Analytics [J]. Front Psychol, 2023(14): 1286092.
[58] Chao Y. Applications of Multimodal Technology in Computer-Assisted Language Learning: Impacts on Cognitive Load and Attention Distribution [J]. Applied and Computational Engineering, 2025.
[59] Froiz-Míguez I, Fraga-Lamas P & Fernández-Caramé S. T. M. Design, Implementation, and Practical Evaluation of a Voice Recognition Based IoT Home Automation System for Low-Resource Languages and Resource-Constrained Edge IoT Devices: A System for Galician and Mobile Opportunistic Scenarios [J]. IEEE Access, 2023(11): 63623-63649.
[60] Gund P, Mane P, Padalikar H, Thorat S, Patil S & Dharmadhikari K. Exploring Students' Perspectives: Leveraging Smart Computing for Online and Offline Education Optimization [C]. 2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0, 5-7 June 2024.

Abbreviations

EU: European Union
GPT: Generative Pre-trained Transformer
HWL: hazardous workplace limit

ISSN 2616-5783 Vol.8, Issue 11: 140-157, DOI: 10.25236/AJHSS.2025.081122

MSDS: material safety data sheet
OSH: occupational safety and health
OSHA: Occupational Safety and Health Administration
PPE: personal protective equipment
UI: user interface
US: United States
UV: ultraviolet
UX: user experience