International Journal of Software Innovation
Volume 14 • Issue 1 • January-December 2026

Fostering Organizational Collective Intelligence From Human-Generative AI Interaction

Shoma Hino
http://orcid.org/0009-0003-2756-2812
Kobe University, Japan

Takuya Nakata
Kobe University, Japan

Sinan Chen
http://orcid.org/0000-0002-9898-7370
Kobe University, Japan

Sachio Saiki
Kochi University of Technology, Japan

Masahide Nakamura
http://orcid.org/0000-0002-1689-7620
Kobe University, Japan
Received: December 2nd, 2025 | **Accepted:** May 27th, 2026

ABSTRACT

Recent advancements in generative artificial intelligence have enabled its use for problem-solving across various fields. However, its effectiveness heavily depends on users' prompting ability, often limiting its potential. Additionally, valuable dialogue logs are frequently left unused. This study proposes a service that scores dialogue logs with generative artificial intelligence, accumulating and sharing them as contextualized problem-solving cases. The service was deployed within an organization for evaluation. Results revealed that the proposed service contributes to operational efficiency and promotes knowledge sharing. In a one-year experiment, 24 members accumulated 90 cases, and survey results indicated high satisfaction with the problem-solving support provided.

KEYWORDS

Collective Intelligence, Generative Artificial Intelligence, Application Development, Large Language Model, Human-Generative Artificial Intelligence, Prompt Engineering, ChatHubAI (CHAI), User Experiment, Assessment, Dify

INTRODUCTION

In modern organizations, the effective accumulation and sharing of knowledge (often referred to as organizational collective intelligence) is crucial for sustainable growth and competitive advantage. Traditionally, collective intelligence has been fostered through explicit documents, wikis, and forums. However, the rapid advancement of artificial intelligence (AI) has introduced new paradigms for knowledge creation and problem-solving.

DOI: 10.4018/IJSI.411903

Recent advancements in generative AI, represented by ChatGPT (OpenAI, n.d.) and Claude (Anthropic, n.d.), have drawn attention to dialogues with generative AI as a problem-solving approach in software development. Applications such as code generation, debugging, and testing assistance (Hagos et al., 2024; Kumar, 2024; Linkon et al., 2024; Senger, 2024) have rapidly become widespread. Large language models (LLMs; AI systems trained on vast amounts of text to understand and generate human language) have seen significant performance improvements since the advent of Transformers. Through the development of generative pre-trained (GPT)-based models, they have now reached a practical stage where they can be applied to complex development challenges (Orrù et al., 2023, Zhong et al., 2024).

The problem-solving capabilities of generative AI are not limited to micro-level issues, such as debugging errors encountered during programming. They also extend to macro-level challenges, such as objectively analyzing the direction of organizations (e.g., companies or research labs) through dialogues with generative AI (Alavi et al., 2024; Cress & Kimmerle, 2023; Storey, 2025). However, even if micro-level problems are resolved through dialogues with generative AI, the dialogue logs, which could be highly valuable to others within the same organization facing similar issues, are often left unused and rarely revisited.

This study targets software development organizations, where generative AI is frequently utilized for tasks such as code generation and debugging, making effective prompt engineering and knowledge sharing essential.

Moreover, the effectiveness of generative AI heavily depends on the user's ability to construct prompts, and there are many instances where its potential is not fully realized (Ravimohan & Sahasrabudhe, 2025). Effective utilization of generative AI requires appropriate prompt design. While services that provide templates for prompts are gradually increasing, few platforms evaluate and share which prompts were effective for specific problems.

To address these challenges, this study proposes a service that accumulates and shares problem-solving cases by saving dialogue logs with generative AI, including their context. By implementing this approach, dialogue logs generated daily through interactions with generative AI can be accumulated and shared as problem-solving cases, fostering collective intelligence within organizations (Zhang et al., 2025). This aims to extract value from dialogue logs as organizational assets, enabling sustainable growth through the enhancement of the organization's knowledge base.

The authors of this study developed a collective intelligence platform that effectively accumulates and shares problem-solving cases by saving dialogues with generative AI, including their context and usefulness, in the form of a web application. The proposed service is called ChatHubAI (CHAI). While previous studies on collective intelligence have focused on knowledge sharing through explicit documentation or forums, this approach uniquely leverages generative AI dialogue logs as dynamic, context-rich organizational assets.

Unlike enterprise versions of ChatGPT (OpenAI, n.d.), which primarily share chat history, or AI-integrated knowledge bases such as Confluence (Atlassian, n.d.), which focus on static documentation, CHAI captures the problem-solving process. It records both the final answer and the prompt engineering attempts, the context of the query (why and how), and the subjective evaluation of utility. This transforms transient interactions into a persistent knowledge asset, revealing the tacit knowledge of how a solution was reached. Furthermore, unlike prior prompt engineering research that emphasizes individual prompt optimization, this work introduces a mechanism for evaluating, accumulating, and sharing effective prompts and their outcomes within an organization. This dual perspective enables both the enhancement of collective intelligence and the systematic improvement of prompt engineering practices at the organizational level.

To evaluate the usefulness of the approach, this study set three related research questions and conducted experiments. Subsequently, it evaluated whether accumulating and sharing problem-solving cases based on dialogues with generative AI promoted operational efficiency and knowledge sharing,

as well as ensure reliability against misinformation. The evaluation was conducted through a Likert scale-based questionnaire survey and feedback from users who used the service.

Ultimately, this research aims to contribute to the advancement of organizational knowledge management and the cultivation of a knowledge sharing culture by providing a practical framework for leveraging generative AI dialogue logs as valuable organizational assets. The findings and methodology presented here are expected to serve as a foundation for future developments in AI-driven collective intelligence and knowledge sharing practices.

BACKGROUND

Collective Intelligence

Collective intelligence mechanisms refer to systems that integrate distributed information and cognitive abilities of multiple intelligent entities, such as humans or agents, through their interactions, thereby generating knowledge or intelligent behavior (Special Interest Group on Collective Intelligence Mechanism, 2021). Sharing collective intelligence allows the knowledge base of an organization to accumulate, enabling sustainable growth. This facilitates the acquisition of broader insights that individuals alone cannot achieve, leading to more comprehensive problem-solving and decision-making.

Within organizations that share objectives, similar problems are likely to arise. Therefore, if problem-solving cases related to such issues can be accumulated within the organization, it becomes possible to foster higher-quality organizational collective intelligence tailored to the organization. Furthermore, beyond individual challenges, sharing dialogue logs with generative AI regarding macro-level issues within the organization enables the alignment of organizational directions. This can enhance the transparency of decision-making processes within the organization and improve the overall quality of organizational decision-making.

Knowledge Management Theories

To ground this approach in established organizational theory, the study draws on two key frameworks: the socialization, externalization, combination, and internalization (SECI; a knowledge creation framework describing how tacit and explicit knowledge are transformed) model and communities of practice (CoP; a group of people who share knowledge, skills, and expertise around a common interest).

SECI Model

The SECI model explains how knowledge is created and converted in organizations (Nonaka & Takeuchi, 1995).

- **Socialization:** Sharing tacit knowledge through direct experience.
- **Externalization:** Articulating tacit knowledge into explicit concepts.
- **Combination:** Systematizing concepts into a knowledge system.
- **Internalization:** Embodying explicit knowledge into tacit knowledge

Dialogue logs with generative AI can be seen as a form of externalization, where a user's vague intent or problem is articulated into a concrete prompt. CHAI further supports Combination by organizing these logs into a searchable repository.

CoP

Wenger (1998) defined CoP as groups of people who share a concern or a passion for something they do and learn how to improve it as they interact. A key element of a CoP is the shared

repertoire of resources (e.g., experiences, stories, tools, and ways of addressing recurring problems). By sharing dialogue logs, CHAI contributes to this shared repertoire, allowing the organization to develop a collective understanding of how to effectively wield generative AI for problem-solving.

Generative AI

Generative AI technology generates text and code using large datasets and deep learning techniques. It is widely used in practical software development. In this study, the default model used for interacts with generative AI is ChatGPT-4o-latest (OpenAI, 2024).

The objective of this study is not to select the model with the highest performance on specific benchmarks, but rather to establish an operational infrastructure for continuously and reproducibly accumulating conversation logs for challenges in software development tasks. Therefore, in selecting the model, this study emphasized stable response quality for development tasks, stability of application programming interface specifications and behavior, and sustainable operational costs.

ChatGPT-4o was adopted as the default model for this study because it meets these conditions. While high-performance models such as Gemini exist, experimental models are susceptible to specification changes and do not necessarily align with the requirement of this study to ensure reproducibility in long-term operation.

FOCUSED CHALLENGES

This study focuses on three problems when accumulating and sharing problem-solving cases based on generative AI:

- **P1:** Disparity in problem-solving capabilities due to differences in prompt engineering skills for effectively utilizing generative AI.

The effective use of generative AI depends heavily on users' prompt engineering skills, which can vary (Przegalińska et al., 2025). As a result, some users can fully leverage the capabilities of generative AI, while others struggle to obtain useful outputs, leading to disparities in problem-solving abilities (Chatterjee et al., 2024; Zhuo et al., 2024; Knoth, N et al, 2024). To address this, it is important to create mechanisms that allow users to share and reference effective prompts and their results, thereby reducing these disparities and enabling more equitable use of generative AI.

- **P2:** Lack of mechanisms for effectively sharing and utilizing dialogue logs within an organization.

Although dialogue logs with generative AI can be valuable knowledge resources, they are often shared without sufficient context or structure, or they are not shared at all. This makes it difficult for others to understand or reuse past problem-solving cases, limiting the accumulation and dissemination of organizational knowledge (Kaczorowska-Spychalska et al., 2024). Establishing a system that enables structured accumulation, contextualization, and sharing of dialogue logs is essential for promoting knowledge reuse and fostering collective intelligence within organizations.

- **P3:** How to mitigate hallucinations in dialogue logs tasks using verifiable signals.

Generative AI is known to sometimes produce hallucinations or misinformation, which can be mistakenly accepted as fact if not properly checked (CNBC, 2023; Jobin et al., 2019). If unreliable information is included in shared dialogue logs, it can mislead users and undermine trust in the knowledge base. Therefore, it is necessary to implement mechanisms for evaluating, annotating,

and filtering dialogue logs to ensure their reliability and support effective factchecking (Rouzrokh et al., 2025).

Considering these issues, the challenges to be addressed can be summarized as follows. By saving problem-solving cases based on dialogue logs with generative AI in a format that includes their intent and is comprehensible, it is possible to assist other users in interpreting the prompts that enabled the problem-solving. This approach aims to avoid the division in problem-solving capabilities caused by differences in prompt engineering skills. Furthermore, it is necessary to establish a mechanism for effectively and reliably utilizing problem-solving cases within an organization.

PROPOSED METHOD

Goal and Key Idea

The objective of this research is to effectively utilize past dialogue logs with generative AI that have been endowed with signals from others regarding hallucinations and misinformation. This aims to avoid disparities in problem-solving capabilities while promoting the creation and sharing of new knowledge.

To achieve this objective, the idea is to develop a web application, CHAI, which accumulates and shares dialogues with generative AI as problem-solving cases, including their context and usefulness, thereby fostering organizational collective intelligence. Furthermore, this study evaluates the utility of CHAI by deploying it within an organization.

Regarding organizational collective intelligence, this study limits the sharing of problem-solving cases within the organization. This is because dialogues with generative AI may include non-public activities within the organization or personal information during problem-solving interactions. Thus, dialogue logs are kept private outside the organization.

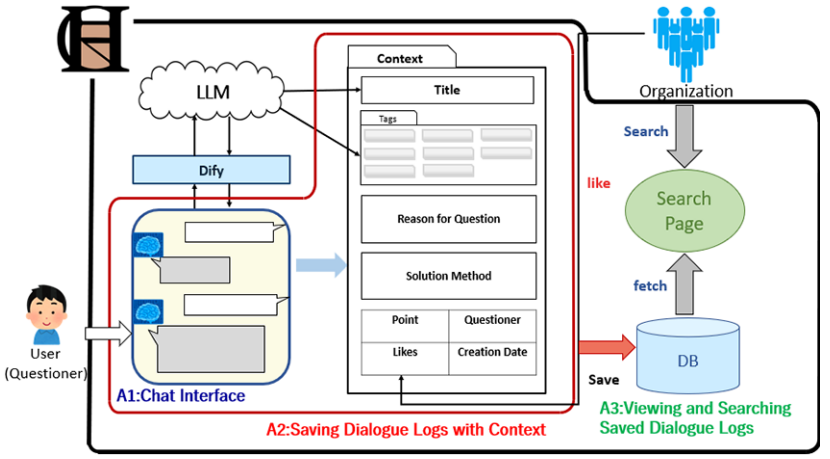
This study categorizes the functionalities of CHAI as follows:

- **A1:** Chat Interface
- **A2:** Saving Dialogue Logs With Context
- **A3:** Viewing and Searching Saved Dialogue Logs

Overall Architecture

The overall architecture of CHAI—incorporating approaches A1, A2, and A3—is shown in Figure 1.

Figure 1. Overall Architecture



In CHAI, interactions between an organization's user and generative AI are conducted through the A1: Chat Interface. The dialogue logs are then saved in the database with the context defined in through A2: Saving Dialogue Logs with Context. Subsequently, in A3: Viewing and Searching Saved Dialogue Logs, users can view and search the saved problem-solving cases.

By referencing the viewed problem-solving cases, users can enhance their knowledge base and refine their prompts. This enables more effective problem-solving through the chat interface using appropriate prompts. By iterating this cycle within the organization, mutual interactions occur as collective intelligence, fostering the development of organizational collective intelligence.

A1: Chat Interface

To effectively record dialogue logs generated through interactions with generative AI, a dedicated chat interface is indispensable. This study adopts the embedded chat interface provided by Dify—an open-source platform for building LLM-based applications (Dify, n.d.). This choice offers a user experience comparable to ChatGPT while maintaining the flexibility to connect with various LLMs beyond GPT-4o and incorporate retrieval-augmented generation-based extensions. Furthermore, the open-source nature of Dify makes it highly suitable for extensible development. By leveraging Dify as the foundation of the chat interface, this study aims to ensure adaptability to the rapid and continuous advancements in generative AI technologies.

A2: Saving Dialogue Logs With Context

When accumulating dialogue logs with generative AI as problem-solving cases, it is essential that others can understand the intent behind these logs. Therefore, it is necessary to save not only the questions and answers with the generative AI but also their context. To achieve this, the required context for treating dialogue logs with generative AI as problem-solving cases is comprehensively defined based on the 5W1H (+How many) framework (see Table 1).

Table 1. Context of Problem-Solving Cases

5W1H + How Many	Context
Who	Questioner
What	Title and Tags
Why	Reason for Question
When	Creation Date
Where	Organization Identification
How	Solution Method
How Many	Points, Likes

Questioner refers to the name of the person who generated the problem-solving case. By recording the information of the questioner, it becomes easier to understand their expertise and intent within the organization, facilitating the interpretation of dialogue logs. Additionally, not limited to CHAI, it becomes possible to provide direct feedback for problem-solving, creating opportunities for fostering collective intelligence.

Title is a summary of the dialogue log. Having a title allows users to infer the content of the dialogue log from the title, which can be helpful for organizing information even when many logs are accumulated.

Tags are multiple keywords related to the dialogue log. By combining this with search functionality, it becomes possible to efficiently aggregate logs related to specific themes or issues. Additionally, by saving the summary of the dialogue log as keywords, it becomes easier to grasp the content of the log.

Points represent a numerical evaluation of the usefulness of a log, with a maximum score of 10. Higher scores indicate greater usefulness, while lower scores suggest less usefulness. CHAI encourages saving dialogue logs regardless of their usefulness or lack thereof. While the value of useful logs has been discussed, even logs deemed unnecessary can be significant. For instance, they may highlight problems that generative AI struggles with or prompts that are difficult for AI to interpret. Accumulating such logs can improve problem-solving methods. Therefore, by introducing a scoring system that evaluates both usefulness and lack thereof, the contextual richness of the logs can be enhanced.

Reason for Question describes the background or purpose of the question, such as the problem the user aimed to address through generative AI. Clearly stating the reason for the question makes the intent of the dialogue easier to understand.

Solution Method describes how the problem was resolved—or not resolved—through dialogue with generative AI. While some responses from generative AI can directly solve problems, in cases where the responses are adapted or applied to solve the issue, failing to document how they were applied would result in the loss of critical contextual information about the solution. Therefore, this item is one of the most important components of the context.

Likes indicates how valuable the problem-solving case is to other users. In CHAI, each user can give one like per log. Logs with many likes often contain information that is useful to others and is considered highly referential. Additionally, the like feature enables both one-way and two-way collective intelligence.

Creation Date indicates when the problem-solving case was created. Understanding the timeline helps users grasp the freshness and context of the information.

Among these items, the Questioner, Title and Tags, and Creation Date can be set automatically without requiring the intent of the questioner. They can be derived from the saved dialogue logs. Therefore, these items were implemented to be generated automatically.

The automatic generation flow for Titles and Tags is as follows. First, the conversation log to be saved (question and answer columns) is input, and the LLM is instructed to generate title candidates that express the overall content of the entire log in a single sentence. The system generates a short list of effective keywords (tags) for searching. Next, the output format is fixed to JSON (title, tags[]) and restricted to parseable structured output only. Finally, searchability and aggregability are ensured by setting an upper limit on the number of tags, removing duplicates, and performing simple normalization of notation variations (lowercase conversion, symbol removal) before saving.

While these items are set for the entire log, each question and answer can also be assigned a necessity level with three categories: critical, necessary, and unnecessary. This implementation allows the user to specify critical parts of the dialogue log among the responses, making the intent of the saver clearer. Additionally, the unnecessary category hides specific parts of the dialogue, such as questions containing information that should not be disclosed within the organization, by making only those parts private.

A3: Viewing and Searching Saved Problem-Solving Cases

By implementing the functionality to view and search saved dialogue logs, it can promote knowledge sharing within the organization and foster collective intelligence.

In CHAI, saved problem-solving cases can only be viewed within the same organization. Searches can be conducted using questioner, title, and tags. Additionally, this search functionality sorts by title, tags, questioner, points, creation date, and number of likes. As shown in Figure 2, detailed filtering is also possible for each column using nine types of operators.

Figure 2. Filter Screen for Problem-Solving Cases

In the detailed view of problem-solving cases, questions and answers are displayed with highlights based on their necessity level, emphasizing critical parts. This detailed viewing and search functionality enables the promotion of knowledge sharing within the organization and the cultivation of collective intelligence.

IMPLEMENTATION

This section describes the implementation of CHAI. For its development, Java/SpringBoot was adopted for the backend. TypeScript/React/Next.js was used for the frontend. Furthermore, considering the potential for open-source development in the future, Docker (Docker, Inc., n.d.) was employed as the infrastructure foundation. This enabled the construction of the development environment using Docker Compose.

The login functionality implemented with Spring Security allows administrators to set an organization identification (ID) and password. Users can create unique user IDs using the organization ID; they can log in with their user ID and the organization password.

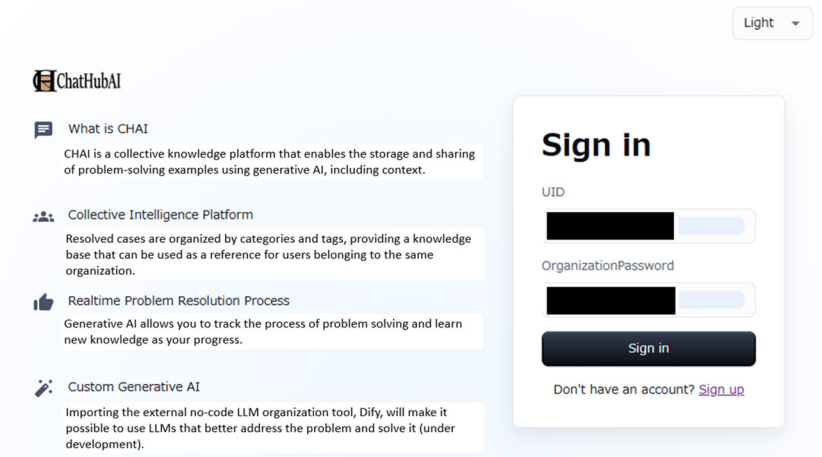
The generative AI model used for the chat interface is ChatGPT-4o. The problem-solving cases collected in this study aim to address issues across various domains. Therefore, ChatGPT-4o was adopted for the reasons described in the Background section (Chatbot Arena, n.d.). Additionally, LangChain (LangChain, Inc., 2025) was used for connecting with generative AI outside the chat interface. Specifically, it was used as an orchestration layer responsible for managing prompt templates for title and tag generation, obtaining structured output (JSON), retrying failed generations, and implementing fallbacks such as assigning default titles.

Based on these technologies, the functionalities described in A1, A2, and A3 were implemented.

Usage

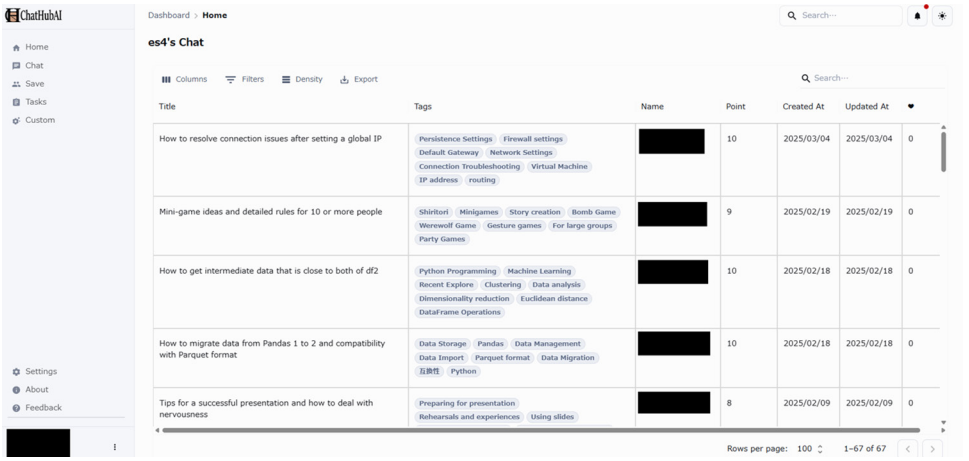
First, when accessing CHAI, the login screen (see Figure 3) is displayed. This screen conveys the purpose and features of CHAI, fostering awareness of collective intelligence cultivation. If the user has not registered, they can provide their user information along with the organization ID on the Sign-Up page. Afterward, they can log in by entering the registered user ID and the organization password on the Sign-In page. The organization ID and password are set by the organization's0 administrator.

Figure 3. CHAI Login Screen



After logging in, a screen displays the list of accumulated problem-solving cases within the organization (see Figure 4).

Figure 4. CHAI Log List Screen



On this screen, clicking on the title of a saved log displays the details of the corresponding problem-solving case. Additionally, a search function is implemented on this screen.

By default, the cases are sorted by creation date. However, sorting by the number of likes or points is also possible. Furthermore, detailed filtering for each column is available.

A chat screen is shown by clicking the chat button on the side menu in Figure 4 (see Figure 5).

Figure 5. CHAI Chat Screen

As the user continues to engage in problem-solving dialogues on the chat page, each dialogue is displayed as an unsaved chat (see Figure 6). These unsaved logs are not publicly accessible. By clicking on a log from this list and entering the context defined in the A2 section on the screen (see Figure 7), the problem-solving case can be saved. Notably, while the title and tags are automatically generated from the dialogue log, the user must manually input the other fields.

Figure 6. List of Unsaved Logs in CHAI

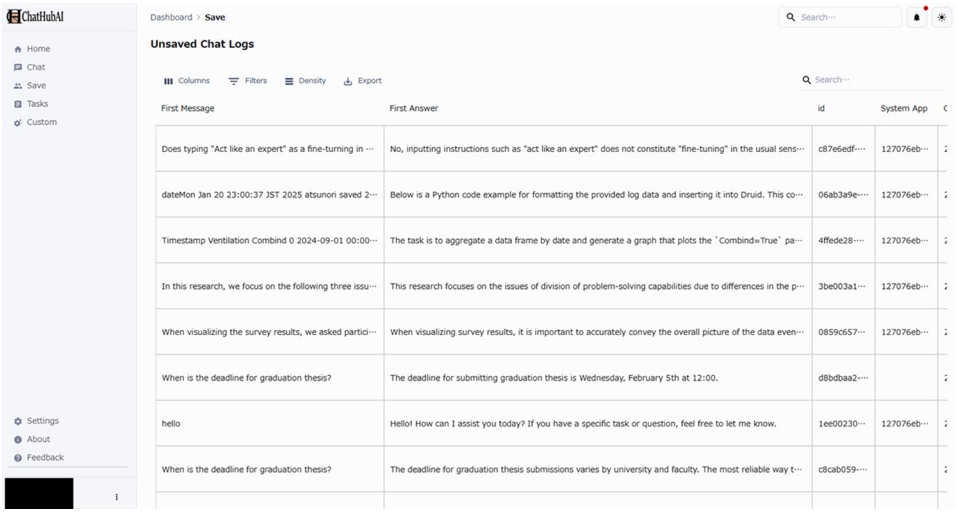
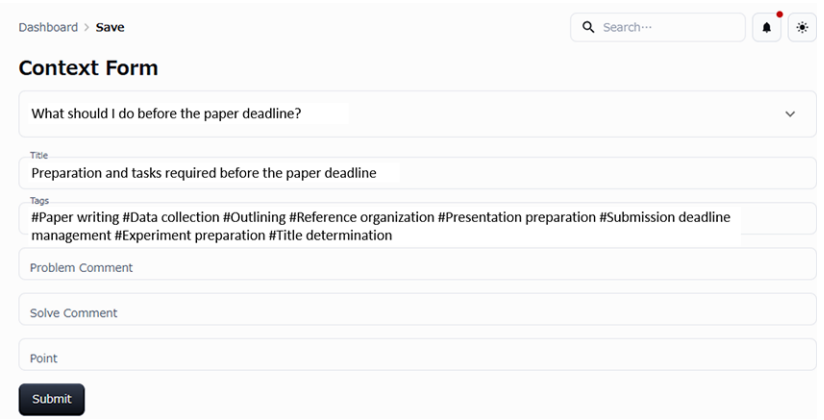


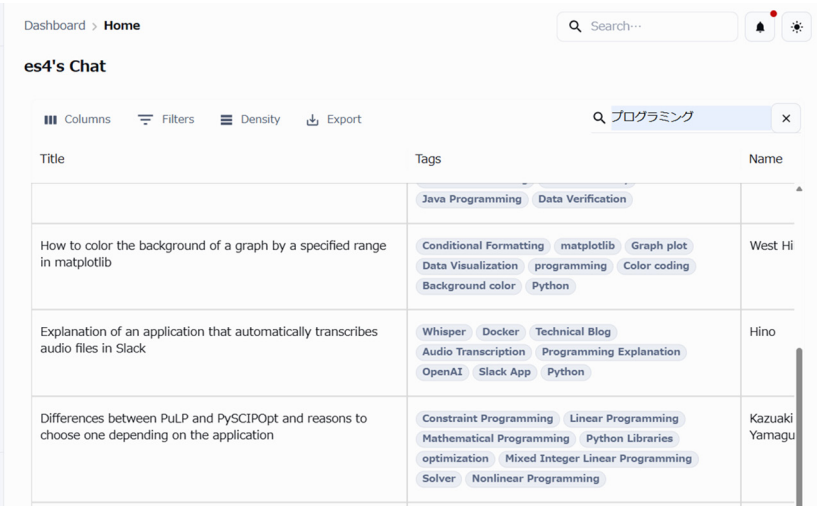
Figure 7. Context Input Screen in CHAI



The saved problem-solving cases are published within the organization on the problem-solving list page, enabling users to view the dialogue logs that led to the solution along with their context.

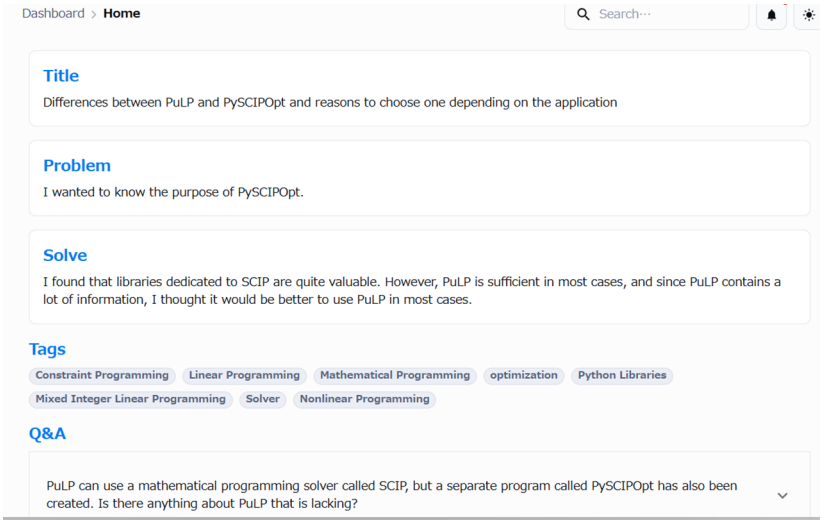
Figure 8 shows the search results when “programming” is entered in the search function of Figure 4. In this example, eight problem-solving cases were found. Among these, the log titled “Differences Between PuLP and PySCIPOpt and Reasons to Choose One Depending on the Application” received a score of eight points. Furthermore, since the author is an associate professor who conducts algorithm research in the author’s laboratory, the credibility of this log is considered high.

Figure 8. Search Results Screen in CHAI



Clicking on the title transitions to a detailed view of the saved problem-solving case, as shown in Figure 9. From the Solution Method section, it is possible to understand why PuLP was deemed the better choice and to trace the dialogue logs leading to this conclusion.

Figure 9. Saved Log Display in CHAI



EXPERIMENT

Overview of the Experiment

To evaluate the effectiveness of the proposed method, two studies were conducted. First, a comparative study verified the importance of context in understanding dialogue logs (V-A). Second, a

one-year pilot study evaluated the system's utility in a real-world software development environment (V-B).

V-A: Comparative User Study

To verify the hypothesis that "context is essential for understanding others' dialogue logs," this study conducted a controlled experiment comparing comprehension levels with and without context.

Experimental Design

Participants were divided into two groups:

1. **Group A (Context Group):** This group viewed dialogue logs within CHAI, including the Problem Description, Background, and Intent.
2. **Group B (No-Context Group):** This group viewed the raw chat history, mimicking standard ChatGPT history sharing.

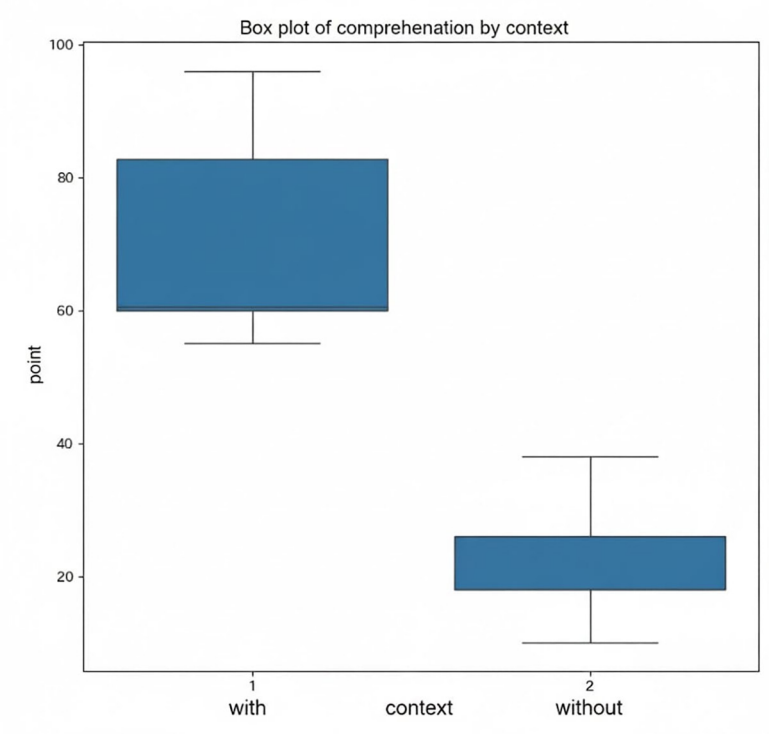
The subjects were 11 graduates and undergraduate students belonging to the authors' software research laboratory.

Both groups were asked to answer a test that assessed their understanding of dialogue log summarization, fact checking, inference, and problem-solving methods. This test was designed with reference to "The Effect of Modularization and Comments on Program Comprehension" (Woodfield et al., 1981). The comprehension scores (out of 100 points) for each group were measured.

V-A Results

The distribution of comprehension scores for both groups is shown in Figure 10. The results of the t-test were $t = 5.577$, $p = 0.000$.

Figure 10. Box Plot Comprehension by Context



The results indicate a significant difference between the two groups.

- **RQ1 (Difficulty Without Context):** The No-Context Group (Group B) showed significantly lower comprehension scores and a wider variance. This confirms that reading raw dialogue logs created by others is difficult and prone to misinterpretation.
- **RQ2 (Effectiveness of Context):** The Context Group (Group A) achieved consistently high scores. This demonstrates that supplementing dialogue logs with explicit context (Background, Intent) effectively assists in understanding others' problem-solving processes.

These findings justify the design of CHAI, which enforces the input of context rather than relying on simple chat history sharing.

V-B: Pilot Study

To verify the usefulness of accumulating and sharing problem-solving cases in a real-world setting, a pilot study was conducted using CHAI. For this evaluation, the following RQs were established:

- **RQ1:** To what extent does CHAI contribute to problem-solving capabilities and operational efficiency through the utilization of generative AI?
- **RQ2:** To what extent does CHAI's dialogue log-sharing feature promote knowledge sharing and knowledge reuse within an organization?
- **RQ3:** How does CHAI function as a reliable information resource in addressing concerns about hallucinations and misinformation in dialogue logs?

The participants were members of the laboratory (24 members) to which the authors belong. This laboratory focuses on software engineering and cloud computing research, making it a suitable environment for evaluating a system designed to support technical problem-solving. The laboratory was treated as a single organization; an organization ID and password were issued. A pilot study was conducted over one year, during which participants were encouraged to use CHAI for problem-solving activities. After the experiment, a survey was conducted to verify RQ1, RQ2, and RQ3.

The experiment period was set to one year (January 2025 to January 2026). This period observed the long-term usage and accumulation of collective intelligence.

Survey Implementation

To verify RQ1, RQ2, and RQ3, the following survey was conducted:

- **Q1:** To what extent did CHAI contribute to the progress of problem-solving?
- **Q2:** Did CHAI improve operational efficiency?
- **Q3:** To what extent did CHAI reduce the difficulty of prompt creation?
- **Q4:** Did the dialogue log-sharing feature help with knowledge sharing within the team?
- **Q5:** Were you able to derive new solutions or ideas from past logs?
- **Q6:** Did you feel that the dialogue logs provided by CHAI contained misinformation (hallucinations)?
- **Q7:** Do you think mechanisms such as scoring feature are sufficient to ensure reliability?
- **Q8:** Please share which features of CHAI were particularly helpful.
- **Q9:** What effects do you think this service brought on the organization as a whole?

RQ1 was evaluated using the responses to Q1, Q2, Q3, and Q8. RQ2 was evaluated using Q4, Q5, and Q9. RQ3 was evaluated using Q6 and Q7. For Q1 to Q7, a quantitative evaluation was conducted using a five- to four-point Likert scale. For Q8 and Q9, qualitative evaluations were conducted through free-text responses to collect both numerical evaluations and specific opinions on individual issues. The neutrality and comprehensiveness of these questions were confirmed by the review of multiple researchers.

V-B RESULT

As a result of the experiment, 90 problem-solving cases were accumulated over the one-year period (January 2025 to January 2026). Additionally, the number of participants who used CHAI (saved at least one log) was nine, including the author. Among these, six participants responded to the survey. The results related to RQ1 (Q1, Q2, Q3) are shown in Figure 11. Those related to RQ2 (Q4, Q5) are shown in Figure 12, and those related to RQ3 (Q6, Q7) are shown in Figure 13. The results of appropriately calculating Cronbach's alpha and correlation coefficient r for each research question (RQ) were as follows: for Q1 and Q2, $\alpha = 0.333$ and $r \approx 0.20$; for Q4 and Q5, $\alpha = 0.920$ and $r \approx 0.85$; and for Q6 and Q7, $\alpha = 0.585$ and $r \approx 0.42$.

Figure 11. Survey Results for Q1, Q2, and Q3

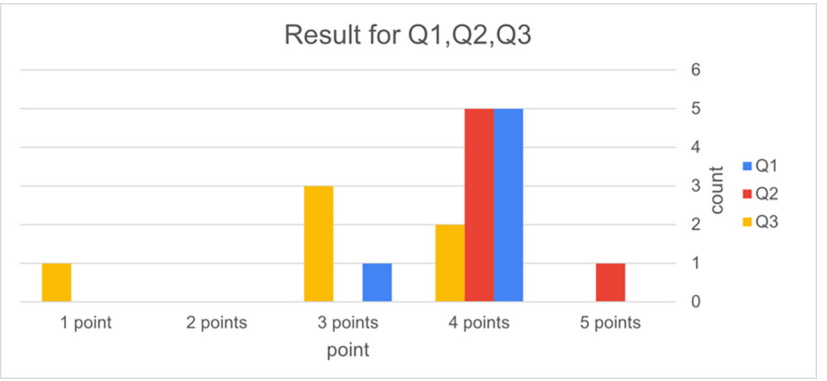


Figure 12. Survey Results for Q4 and Q5

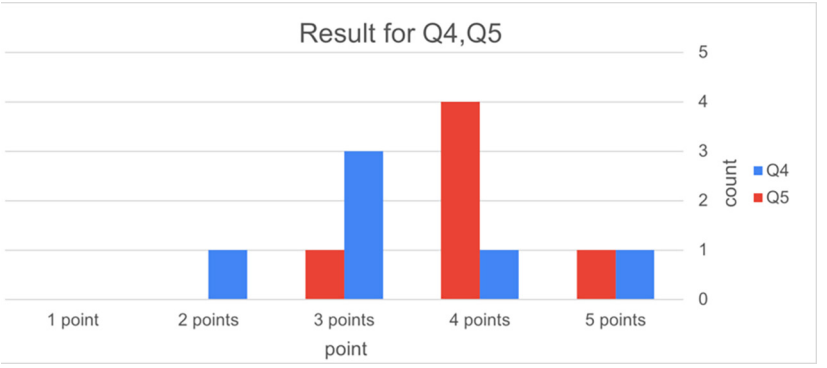


Figure 13. Survey Results for Q6 and Q7

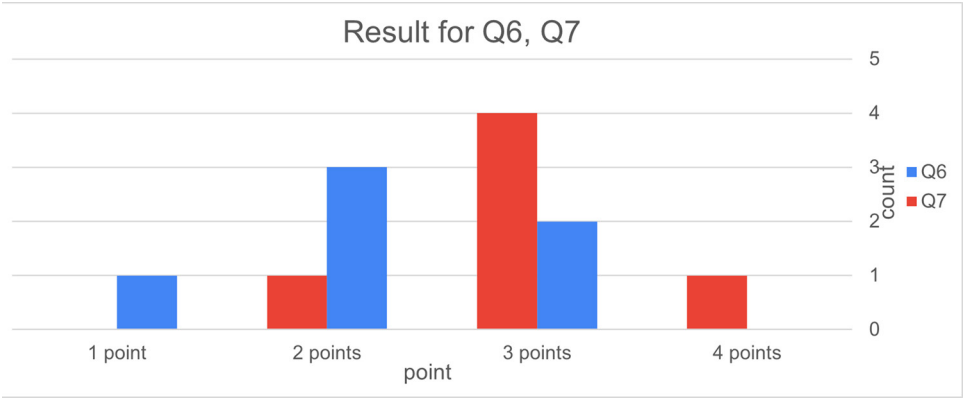


Figure 14 visualizes the frequency of 1,768 words. It was obtained by segmenting saved tags into individual words. It is ordered from highest to lowest frequency. Additionally, a word cloud of

the tags is shown in Figure 15. Figure 16 represents the co-occurrence relationships between these same segmented tag words as a network.

Figure 14. Tag Frequency

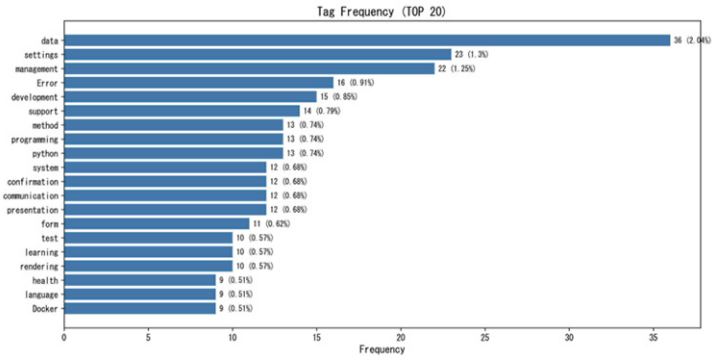


Figure 15. Word Cloud of Tags

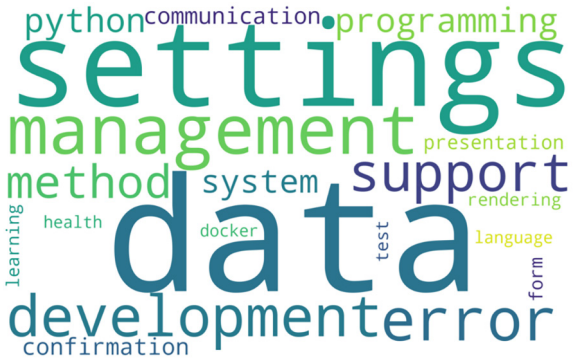
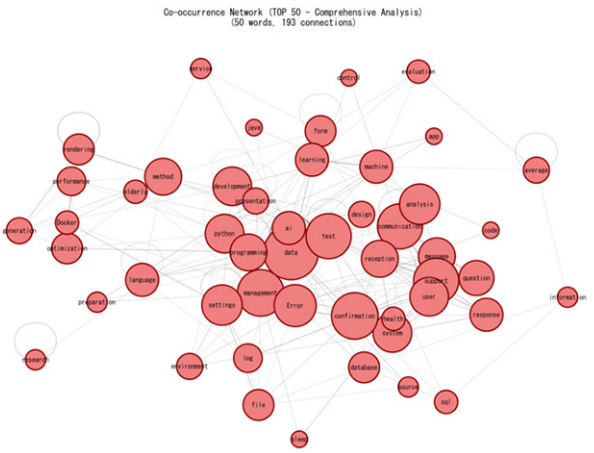


Figure 16. Co-Occurrence Relationships of Tags



Additionally, the free-text responses for Q8 and Q9 are shown in Tables 2 and 3, respectively. The results of the thematic analysis regarding the free-text responses are presented in Figure 17.

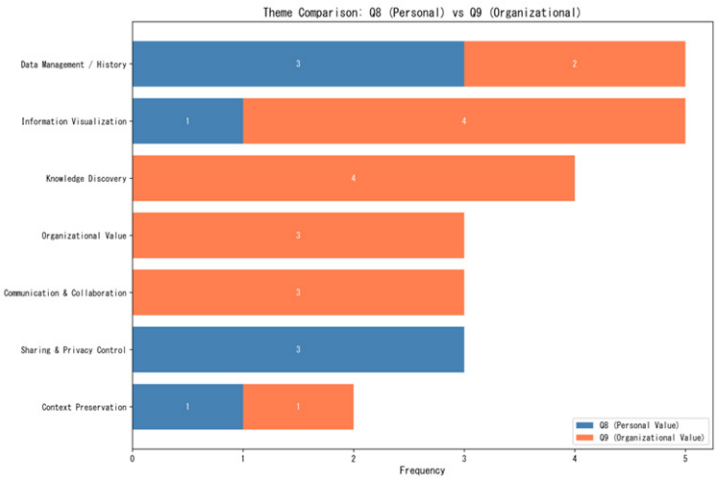
Table 2. Free-Text Responses for Q8

Q8. Please tell us about a feature that you found particularly useful in CHAI.
Ability to temporarily save unshared chats.
The fact that I can decide what to share. I like that I can ask personal questions.
Evaluation and sharing functions.
Ability to look back at past exchanges with ChatGPT in a continuous Q&A format.
It is useful to go back to your past chat history.

Table 3. Free-Text Responses for Q9

Q9. What effect do you think this service has had on the organization as a whole?
Effect of learning from the solutions of others in the organization using generative AI as a solution to their own problems.
I found it motivating to see what others were researching.
It's easier to see other people's questions.
It is of great significance to be able to look back on one's past questions, including their context. It is also a good opportunity for communication because you can learn what other people are having trouble with.
It helps to visualize what other people are working on. Even if you don't open the chat history itself, just looking at the titles and tags can be helpful.

Figure 17. Thematic Analysis of Fee-Text Answers



DISCUSSION

Evaluation of RQs

The evaluation of RQ1 is conducted based on Figure 11 and Table 2. Regarding Q1 and Q2, as can be seen from the graph, CHAI is evaluated as useful for problem-solving and operational improvement. From the free-text responses in Q8, it is also evident that the ability to trace back individual logs contributes to operational improvement. However, for Q3, the evaluation that CHAI reduced the difficulty of prompt creation was not as high, and the main point of appreciation was the ability to trace back one's own logs.

Next, the evaluation of RQ2 is conducted based on Figure 12 and Table 3. Regarding Q4 and Q5, the graphs show that CHAI received high evaluations. This suggests that the sharing of problem-solving cases in CHAI is considered effective for knowledge sharing and knowledge reuse within the organization. Furthermore, from the free-text responses in Q9, opinions such as "Being able to see what others are researching can boost motivation," "It helps visualize what others are working on," and "Sharing what others are struggling with could serve as a starting point for communication" indicate that CHAI has a strong impact on knowledge sharing within the organization.

Furthermore, the content of the knowledge base cultivated by CHAI can be observed from the word cloud in Figure 15, the word frequency bar graph in Figure 14, and the co-occurrence network in Figure 16. In terms of word frequency, "data," "setting," and "management" appear at the top, suggesting that many programming-related issues are registered. Moreover, in the co-occurrence network, "ai," "programming," and "test" are connected around "data." This finding indicates that CHAI is utilized for solving programming issues, including data processing and AI applications.

A thematic analysis was conducted on the free-text responses from Q8 (Table 2) and Q9 (Table 3) to provide deeper insights into the user experience. The responses were coded and categorized into three main themes.

- **Operational Efficiency Through Referencability:** Users frequently mentioned that by "looking back through logs" (Q8), they could reproduce successful prompt patterns without having to create them from scratch. Additionally, thematic analysis showed numerous mentions related to "data management" and "information visualization," indicating that CHAI enhances work convenience by allowing users to review their own and others' logs.

- **Motivation via Transparency:** Responses in Q9 highlighted that seeing others' logs "boosts motivation" and "visualizes what others are working on." This suggests that CHAI serves not just as a database but as a social awareness tool that fosters a sense of connection within the organization.
- **Knowledge Discovery:** Users often pointed out that shared logs serve as a "starting point for communication," with many descriptions related to "communication and collaboration." This indicates that the system promotes socialization in the SECI model by facilitating face-to-face discussions based on digital logs.

Furthermore, the evaluation of RQ3 is conducted based on Figure 12. Although hallucinations were infrequent during problem-solving, the mechanism for ensuring reliability by adding context helped mitigate the risk of hallucinations. However, "few hallucination reports" does not necessarily imply a low risk of hallucinations. The evaluation domains in this study (software development) consist largely of areas where outputs can be relatively easily verified through code execution, testing, static analysis, and document matching. This characteristic may partially influence the results.

Theoretical Implications and Applicability to Knowledge Management

This study offers several theoretical contributions to the field of organizational knowledge management. First, the CHAI platform operationalizes key concepts from knowledge management theory, such as the SECI model, by enabling the externalization and combination of tacit knowledge through the systematic capture and sharing of generative AI dialogue logs. Dialogue logs not only record explicit solutions but also preserve the context, reasoning, and iterative problem-solving processes that are often lost in traditional documentation. This supports the conversion of individual tacit knowledge into organizational explicit knowledge, facilitating organizational learning and innovation. Moreover, CHAI directly addresses the challenge of knowledge retention and transfer, which is critical for organizational resilience.

By accumulating a searchable, context-rich repository of problem-solving cases, organizations can mitigate the risk of knowledge loss due to staff turnover and support onboarding and upskilling of new members. The platform also enables the identification of knowledge gaps and recurring issues, informing targeted training and process improvement. From the perspective of knowledge sharing culture, CHAI fosters an environment of transparency, trust, and mutual support. The system's design encourages members to contribute their experiences, including both successes and failures, which is essential for building psychological safety and a sustainable knowledge sharing culture. Features such as peer evaluation, feedback, and recognition (e.g., likes, points) can further incentivize participation and reinforce positive sharing behaviors. This aligns with research on communities of practice and social capital, highlighting the importance of informal networks and shared norms in effective knowledge management.

The findings of this study have important implications for organizational knowledge management. By systematically accumulating and sharing dialogue logs with generative AI, CHAI enables organizations to capture not only explicit solutions but also the context and reasoning behind problem-solving processes.

This process aligns closely with the externalization phase of the SECI model. When a user crafts a prompt to solve a problem, they are externalizing their tacit understanding of the issue into explicit text. CHAI captures this externalized knowledge (the prompt and the context) and makes it available for combination with other users' knowledge. When other users read these logs and apply the prompt techniques to their own problems, internalization occurs, completing the knowledge creation spiral.

Furthermore, the structured sharing of dialogue logs encourages the development of CoPs. As members contribute to and reference the repository of effective prompts and problem-solving patterns, they build a shared repertoire of resources. This lowers the barrier to knowledge reuse and innovation, as the "practice" of using generative AI becomes a communal rather than individual capability.

The CHAI platform is not limited to a single laboratory or research group; its design and mechanisms are applicable to a wide range of organizations seeking to enhance their knowledge management practices. By fostering transparency, traceability, and collaborative problem-solving, CHAI can support organizations in various sectors to build robust knowledge bases and promote a culture of continuous learning and improvement. The integration of such systems can help organizations adapt to changing environments, retain critical know-how, and accelerate organizational learning across teams and departments.

Limitations

This study has several limitations. The issues related to these limitations and their improvement strategies are outlined here.

First, the study identified a low usage rate and a limited number of problem-solving cases. Only 9 out of 24 members created logs; 90 cases were accumulated in one year. This rate was due to limited user engagement and short experimental period. Instead of merely making abstract calls for increased motivation, future work will introduce the following:

- Weekly automatic notifications for "Useful Logs of the Week"
- Simplified input user interface to reduce save operations to within three clicks
- Monthly feedback based on the number of posts and reuses

These measures will increase triggers for posting behavior and promote continuous usage.

Second, the study had a small sample size and statistical power. The small number of participants ($n = 9$) in the pilot study (V-B) limits the statistical power required for inferential statistics (e.g., t-tests). Consequently, the analysis relies primarily on descriptive statistics and qualitative evaluation. The field deployment was conducted as an exploratory pilot study within a single laboratory to investigate real-world usage patterns and identify initial usability barriers before scaling. Future work should replicate the study with a larger sample size and conduct comparative studies across multiple organizations to enhance reliability and robustness.

Third, most saved cases had high scores. Low-score cases were rarely saved, limiting diversity. Thus, users may be reluctant to save less successful or incomplete cases. Users should be encouraged to save a wider range of cases, including those with low scores or unsuccessful outcomes, to enrich the knowledge base.

Fourth, the study had limited reference to prompt engineering techniques because the system did not support referencing specific prompt engineering techniques. There was a lack of detailed documentation or sharing of prompt strategies in the saved logs. Future studies should develop features to facilitate the sharing and referencing of effective prompt engineering methods within the organization.

Given the small number of users who created logs, a survey was conducted among members who did not save logs to investigate the reasons. The survey included the following questions:

- **Q11:** Do you use generative AI for problem-solving?
- **Q12:** What are the reasons for not using CHAI? (multiple answers allowed)
- **Q13**: What support do you think is necessary to effectively utilize CHAI? (multiple answers allowed)

The result of Q11 shows that 85.7% of the respondents answered "frequently" and the remaining 14.3% answered "occasionally." Other results are shown in Figures 18 and 19.

Figure 18. Survey Results for Q12

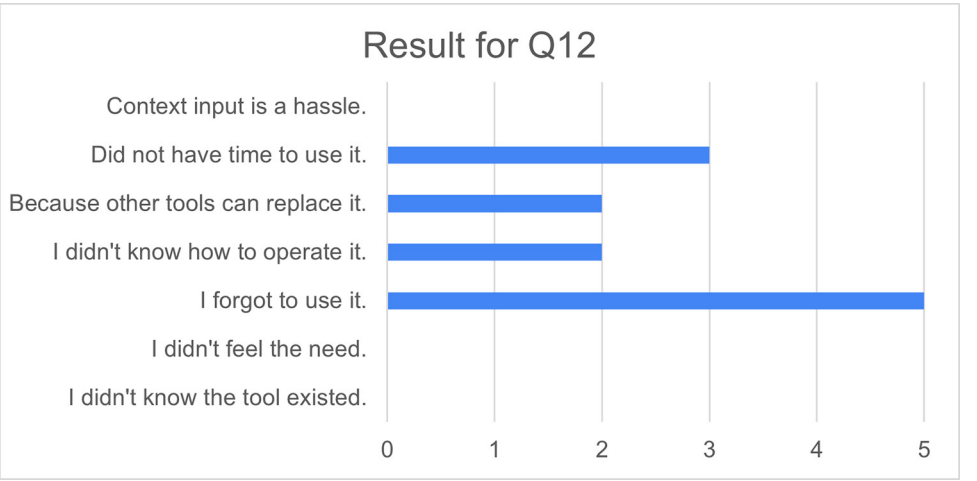
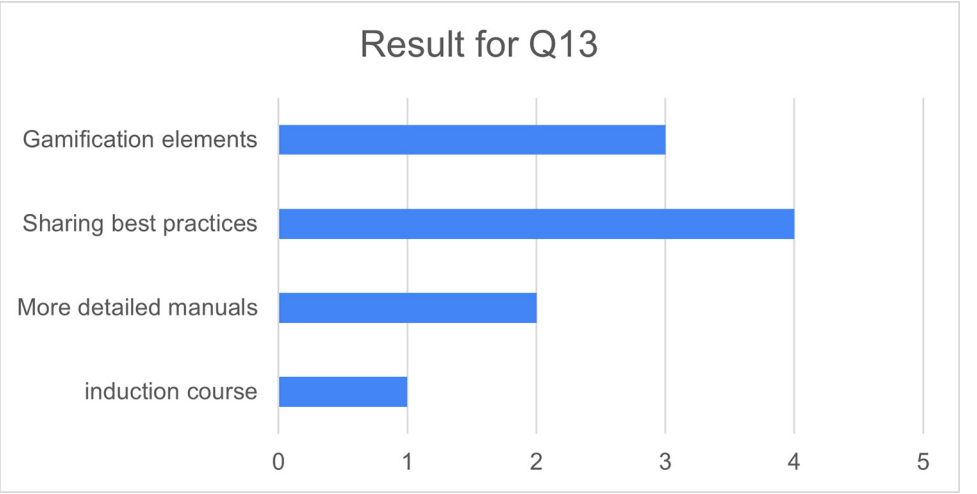


Figure 19. Survey Results for Q13



Looking at the results of Q11, Q12, and Q13, it can be observed that users who did not utilize CHAI often used other generative AI tools for problem-solving. Users cited reasons such as "forgot to use CHAI" and "did not have time to use it."

Additionally, as shown in Figure 18, the most frequently mentioned support needed to effectively utilize CHAI was "sharing best practices." In other words, encouraging the use of CHAI, it is important to demonstrate the advantages of fostering collective intelligence by accumulating and sharing dialogue logs with generative AI, including their context, as well as to present successful cases compared with simply using tools like ChatGPT.

However, as discussed in RQ1, despite the potential advantages of the proposed service, the current number of users and the limited accumulation of problem-solving cases suggest that, while CHAI can be effectively utilized for individual problem-solving, it has not yet demonstrated sufficient impact in enhancing problem-solving capabilities or improving work efficiency through

organizational interactions based on shared cases. Consequently, to highlight the advantages of the service—including successful use cases—it is necessary to increase both the number of users and the volume of accumulated cases. However, achieving such growth requires demonstrating the benefits of the service, thereby presenting a dilemma.

To address this dilemma, which lies in the early stages of fostering collective intelligence, it is necessary to promote the use of CHAI from perspectives beyond its current functionality as a tool for operational efficiency.

Regarding the quality of collective intelligence, as mentioned in the Discussion section, the average score of the logs is relatively high, indicating that problem-solving cases with high scores are being saved. However, this implies that problem-solving cases with low scores are not being saved. This suggests that a diverse range of problem-solving cases is not being accumulated, which is not ideal for fostering collective intelligence.

Additionally, concerning survey question Q3, "To what extent did CHAI reduce the difficulty of prompt creation?," the results were not as high compared to other indicators. This is likely due to the lack of information on how specific prompt engineering techniques influenced the output of generative AI, resulting in fewer opportunities for users to reference others' prompts. Therefore, a different approach may be necessary to save detailed methods for utilizing generative AI and effective prompts.

CONCLUSION

This study proposed and implemented CHAI, a service for accumulating and sharing problem-solving cases based on dialogue logs with generative AI. The main findings are as follows. CHAI enables the systematic collection and sharing of both explicit solutions and the contextual reasoning behind problem-solving, thereby supporting the creation and dissemination of organizational knowledge. The deployment and evaluation within a laboratory setting demonstrated that CHAI contributes to enhancing problem-solving capabilities, promotes knowledge sharing within the organization, and determines the trustworthiness of other's dialogue logs through verifiable signals. On the other hand, the number of participants and case studies is limited, and the current findings are at the stage of demonstrating the validity of the pilot study.

Looking ahead, CHAI may evolve from a repository of dialogue logs into a proactive organizational knowledge partner. Beyond simply increasing the number of cases, future work focuses on three key evolutionary stages.

The first stage is the transition from passive search to active recommendation. By implementing real-time retrieval-augmented generation, the system could proactively suggest relevant past problem-solving cases to users during their interactions, thereby preventing reinventing the wheel without explicit searching.

The second stage highlights the automated synthesis of organizational knowledge. Future work aims to develop mechanisms where generative AI autonomously aggregates scattered dialogue logs to construct and update structured knowledge bases, such as internal wikis or FAQs, dynamically reflecting the latest problem-solving trends.

Through these advancements, CHAI aims to not only support individual efficiency but also to continuously cultivate a self-evolving collective intelligence ecosystem.

FUNDING

This research was partially supported by JSPS KAKENHI Grant Numbers JP25H01167, JP25K02946, JP24K02765, JP24K02774, JP23K17006, JP23K28091, and JP23K28383.

COMPETING INTERESTS

The authors of this publication declare there are no competing interests.

CORRESPONDING AUTHOR

Correspondence concerning this article should be addressed to Shoma Hino, Department of Electrical and Electronic Engineering, Graduate School of Engineering, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan. Email: hino-s@es4.eedept.kobe-u.ac.jp

REFERENCES

Alavi, M., Leidner, D. E., & Mousavi, R. (2024). Knowledge management perspective of generative artificial intelligence (GenAI). *Journal of the Association for Information Systems*, *25*(1), 1–12. DOI: 10.17705/1jais.00859

Anthropic. (n.d.). *Claude*. https://claude.ai/new

Atlassian. (n.d.). *Confluence.* https://www.atlassian.com/software/confluence

Chatbot Arena. (n.d.). *Chatbot Arena.* https://lmarena.ai/?leaderboard

Chatterjee, A., Renduchintala, H. S. V. N. S. K., Bhatia, S., & Chakraborty, T. (2024). *POSIX: A prompt sensitivity index for large language models.* arXiv preprint. /arXiv.2410.02185DOI: 10.18653/v1/2024.findings-emnlp.852

CNBC. (2023, May 31). *OpenAI is pursuing a new way to fight A.I. "hallucinations."* https://www.cnbc.com/2023/05/31/openai-is-pursuing-a-new-way-to-fight-ai-hallucinations.html

Cress, U., & Kimmerle, J. (2023). Co-constructing knowledge with generative AI tools: Reflections from a CSCL perspective. *International Journal of Computer-Supported Collaborative Learning*, *18*(4), 607–614. DOI: 10.1007/s11412-023-09409-w

Dify. (n.d.). *Dify.* https://dify.ai/

Docker, Inc. (n.d.). *Docker.* https://www.docker.com/

Hagos, D. H., Battle, R., & Rawat, D. B. (2024). Recent advances in generative AI and large language models: Current status, challenges, and perspectives. *IEEE Transactions on Artificial Intelligence*, *5*(12), 5873–5893. DOI: 10.1109/TAI.2024.3444742

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. DOI: 10.1038/s42256-019-0088-2

Kaczorowska-Spychalska, D., Kotula, N., Mazurek, G., & Sułkowski, Ł. (2024). Generative AI as a source of change of knowledge management paradigm. *Human Technology*, *20*(1), 131–154. DOI: 10.14254/1795-6889.2024.20-1.7

Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence*, *6*, 100225. DOI: 10.1016/j.caeai.2024.100225

Kumar, P. (2024). Large language models (LLMs): Survey, technical frameworks, and future challenges. *Artificial Intelligence Review*, *74*(10), 260. Advance online publication. DOI: 10.1007/s10462-024-10888-y

LangChain, Inc. (2025, April 30). *LangChain*. https://www.langchain.com/

Linkon, A. A., Shaima, M., & Sarker, M. S. U. (2024). Advancements and applications of generative artificial intelligence and large language models on business management: A comprehensive review. *Journal of Computer Science and Technology Studies*, *6*(1), 225–232. DOI: 10.32996/jcsts.2024.6.1.26

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford University Press., DOI: 10.1093/oso/9780195092691.001.0001

Open, A. I. (2024). GPT-4o system card. arXiv. https://arxiv.org/abs/2410.21276

Open, A. I. (n.d.). *ChatGPT*. https://chatgpt.com/

Orrù, G., Piarulli, A., Conversano, C., & Gemignani, A. (2023). Human-like problem-solving abilities in large language models using ChatGPT. *Frontiers in Artificial Intelligence*, *6*, 1199350. Advance online publication. DOI: 10.3389/frai.2023.1199350 PMID: 37293238

Przegalińska, A. K., Triantoro, T., Kovbasiuk, A., Ciechanowski, L., Freeman, R. B., & Sowa, K. (2025). Collaborative AI in the workplace: Enhancing organizational performance through resource-based and task-technology fit perspectives. *International Journal of Information Management*, *29*, 102853. Advance online publication. DOI: 10.1016/j.ijinfomgt.2024.102853

Ravimohan A., & Sahasrabudhe, V. (2025). A proposed prompting protocol to get substantive enhancement of generative AI performance in an organizational context. *International Journal for Multidisciplinary Research*. DOI: 10.36948/ijfmr.2025.v07i02.43350

Rouzrokh, P., Khosravi, B., Faghani, S., Moassefi, M., Shariatnia, M. M., Rouzrokh, P., & Erikson, B. (2025). A current review of generative AI in medicine: Core concepts, applications, and current limitations. *Current Reviews in Musculoskeletal Medicine*, *1*(7), 246–266. https://link.springer.com/journal/43836. DOI: 10.1007/s12178-025-09961-y PMID: 40304941

Sengar, S. S., Hasan, A. B., Kumar, S., & Carroll, F. (2024). Generative AI: A systematic review and applications. *Multimedia Tools and Applications*, *93*. https://www.springer.com/journal/11042

Special Interest Group on Collective Intelligence Mechanism. (2021). *Collective knowledge mechanisms for society*. https://www.jst.go.jp/sis/scienceagora/2021/session/07-c15.html

Storey, V. (2025). Knowledge management in a world of generative AI: Impact and implications. *ACM Transactions on Management Information Systems*, *5*(3), 1–14. https://dl.acm.org/journal/tmis. DOI: 10.1145/3719209

Wenger, E. (1998). Communities of practice: Learning, meaning and identity. *Journal of Mathematics Teacher Education*, *6*(2), 185–194. DOI: 10.1023/A:1023947624004

Woodfield, S. N., Dunsmore, H. E., & Shen, V. Y. (1981). The Effect of Modularization and Comments on Program Comprehension. *Proceedings of the 5th International Conference on Software Engineering (ICSE '81)*, 215–223. DOI: 10.5555/800078.802534

Zhang, Q., Zuo, J., & Yang, S. (2025). Research on the impact of generative artificial intelligence (GenAI) on enterprise innovation performance: A knowledge management perspective. *Journal of Knowledge Management*, *29*(7), 1367–3270. DOI: 10.1108/JKM-10-2024-1198

Zhong, T., Liu, Z., Pan, Y., Zhang, Y., Zhang, Z., Zhou, Y., Liang, S., Wu, Z., Lyu, Y., Shu, P., Yu, X., Cao, C., Jiang, H., Chen, H., Li, Y., Zhao, H., Xu, S., Dai., H., Zhao, L., Zhang, R., … Liu, T. (2024). *Evaluation of OpenAI o1: Opportunities and challenges of AGI*. arXiv. https://arxiv.org/abs/2409.18486

Zhuo, J., Zhang, S., Fang, X., Duan, H., Lin, D., & Chen, K. (2024). ProSA: Assessing and understanding the prompt sensitivity of LLMs. *Findings of the Association for Computational Linguistics: EMNLP*, *2024*, 1950–1976. DOI: 10.18653/v1/2024.findings-emnlp.108