



Understanding user behaviors and satisfaction in LLM-assisted tasks: prompting strategies and user knowledge states among college students

Tzu-Yun Chien, Po-Yu Chen, and Muh-Chyun Tang

DOI: <https://doi.org/10.47989/ir31iConf64164>

Abstract

Introduction. This study examines how users interact with large language models (LLMs), focusing on the interplay between users' knowledge states and prompt strategies in shaping satisfaction.

Method. Data were collected from 39 students, yielding 187 valid task records. Participants reported knowledge states and satisfaction and submitted dialogues for analysis. Prompt strategies and the character of LLM were coded, and a *strategy matching rate* was proposed as a novel indicator of LLM application ability.

Analysis. Analyses combined descriptive statistics, correlation analyses, Mann-Whitney U tests, and mixed-effects logistic regression, with intercoder reliability established through iterative agreement.

Results. Directive, contextual, and iterative refinement strategies were most frequently observed, with notable discrepancies between self-reported and coded strategy use. Regression analyses showed that non-delegability and creativity increased the likelihood of satisfaction exceeding expectations, while topic familiarity and knowledge status effects reduced it. Chain-of-thought and self-consistency strategies positively predicted satisfaction, whereas directive and role prompting strategies had negative effects.

Conclusion(s). Satisfaction in LLM-assisted tasks emerges from the interaction of users' knowledge states, task characteristics, and prompting strategies. By introducing strategy matching rate as an empirical indicator of LLM application ability, this study contributes a practical approach to assessing AI literacy and advances understanding of human-LLM collaboration.

Introduction

Since their emergence, large language models (LLMs) have rapidly permeated education, knowledge work, and everyday life, becoming essential tools that assist users in creation, information retrieval, analysis, and decision-making. Previous research has largely examined LLM applications from the perspectives of AI literacy, satisfaction, and usage motivation. However, there remains a lack of empirical evidence regarding users' choices of prompt strategies in specific tasks and how such behaviors are shaped by their knowledge status and task characteristics from users.

This study seeks to integrate perspectives from human-computer interaction and information retrieval theory to investigate LLM users' prompt strategies, character of LLM, and satisfaction during tasks. A total of 187 task records were collected from 39 students. Through survey responses and dialogue coding, we examined the consistency between self-reported and actual strategies, constructing strategy matching rate as a novel indicator of LLM application ability. Moreover, this study incorporates topic familiarity, non-delegability, and knowledge status as task characteristics from users to explore their influence on users' strategy choices, character of LLM adoption, and satisfaction.

Based on this research design, the study addresses the following research questions:

RQ1: What is the tendency of strategy used by LLM users?

RQ2: How does users' knowledge state, such as topic familiarity, non-delegability and knowledge dependency, and meta-cognition influence prompting behaviors and prompting strategy?

RQ3: Do variables mentioned above influence user satisfaction?

By combining qualitative and quantitative analyses, this study extends the application of AI literacy and information retrieval theories in LLM usage contexts and offers new insights for understanding the roles of knowledge status and prompt strategies in human-LLM interaction research.

Literature review

Purposes of using LLMs

LLMs are increasingly integrated into professional and everyday contexts. Their main purposes include advice seeking, information retrieval, entertainment, creative ideation, problem solving, writing assistance, and API-based task execution (Wang et al., 2024). In professional domains, such as medical libraries, they support reference services, research, and cataloguing (Ju & Stewart, 2024). For everyday use, LLMs facilitate information seeking, task efficiency, social interaction, and creative outputs such as code, music, or text (Divekar et al., 2025; Ju & Stewart, 2024; Kim et al., 2025), thereby advancing users' intellectual development (Song et al., 2025). They are also used in language learning, programming, and knowledge acquisition, where they can refine users' cognitive models and enhance reflective thinking (Ali et al., 2025; Holstein et al., 2025). Moreover, there are usage in domains such as news copywriting and character development (Kim et al., 2025), as well as in research and extended forms of creative writing (Eigner & Händler, 2024).

User-LLM Interaction Patterns and the Character of LLM

Studies show that user *satisfaction* is central to continued LLM use. For instance, Divekar et al. (2025), drawing upon Bloom's taxonomy (as cited in Krathwohl, 2002), categorised users' tasks with LLMs into six cognitive levels, ranging from lower to higher order: Remember, Understand, Apply, Analyse, Evaluate, and Create. Personalisation and perceived usefulness foster sustained engagement, while conversational ability is less predictive (Ju & Stewart, 2024). Beyond task

execution, LLMs enable dynamic dialogue, companionship, feedback, and problem explanation (Eigner & Händler, 2024; Kacperski et al., 2025). Compared with search engines, LLMs can directly interpret natural language queries, thereby reducing cognitive load by generating synthesised answers and supporting decision-making through multi-perspective responses (Kim et al., 2025; Stadler, 2024). Phillips et al. (2024) identified four interaction modes—Digital Assistant, Co-pilot, Team Member, and Digital Influencer—emphasising how the *character of LLM* and degrees of *non-delegability* shape behaviors. Similarly, Zhu et al. (2024) showed that human-led or balanced collaborations with LLMs were perceived as beneficial. Building on this, the present study investigates whether interaction modes influence user *satisfaction*.

Knowledge state of users

Research highlights two key user characteristics: *topic familiarity* and AI literacy. In the domain of information retrieval, higher familiarity leads to stricter relevance judgments and lower *non-delegability* (Tang et al., 2013), while low familiarity increases overtrust (Nourani et al., 2020). LLM studies confirm that mismatches in topic familiarity between users and LLMs reduce *satisfaction* (Palta et al., 2025). AI literacy further influences *satisfaction* (Lee & Park, 2023) and *prompt strategy* use, with high-literacy users better able to identify and correct errors (Kim, 2024). Yet, reliance on *self-reported* measures risks overestimation due to the simplicity of LLM interfaces (Divekar et al., 2025). This study therefore combines self-reported variables and alternative measures of LLM *application ability*. Within information retrieval theory, *knowledge status* is viewed as tacit, dynamic, and externalised through interaction. Saracevic (1996) described relevance as situational judgment, while Belkin's (1980) '*anomalous state of knowledge*' and Ingwersen's (1992) models emphasised iterative clarification. Extending these insights, this study examines whether *knowledge status* similarly shapes behaviours and *prompt strategies* in LLM use. Metacognition further enriches this perspective. Defined as the ability to reflect on and regulate learning (Schraw, 1994), it enables users to select appropriate *prompt strategies* to address information needs. Studies in reading show that metacognitive awareness can be measured (Mokhtari & Reichard, 2002) and improved through instruction (McCabe, 2011), underscoring its relevance to LLM interaction.

Research method

In this study, 39 students enrolled in a course in the Department of Library and Information Science were asked to complete reflection records on their experiences using LLMs, including 7 from non-LIS majors, 2 LIS graduate students, and 30 LIS undergraduate students. Among them, 14 were male and 25 were female.

Major	count	percentage
Not LIS	7	17.95%
LIS	32	82.05%
Total	39	100.00%

Table 1. Majors of students participated.

Gender	count	percentage
female	25	64.10%
male	14	35.90%
Total	39	100.00%

Table 2. The gender of students participated.

At the beginning of the semester, students received basic training in prompt strategies and were provided with simple operational guidelines, drawing upon instructions from OpenAI and Google's prompt engineering documentation.

Each student was required to submit at least five records of their LLM usage experiences and reflections. Submissions included either a publicly accessible link to the dialogue transcripts or uploaded files, accompanied by responses to nine self-reported survey questions. These questions addressed the use of prompt strategies during the task, the relationship between the task and the user's knowledge status, and satisfaction with the model's responses. Data collection took place between March 1 and June 22, 2025. In total, 199 records were submitted, of which 187 task records remained valid after excluding entries with inaccessible transcripts. These 187 records were transcribed at the dialogue level, including both user prompts and model responses, and qualitatively coded using grounded theory (Glaser & Strauss, 2017). The coding scheme covered two dimensions: (1) prompt strategies employed by the user in a single dialogue, and (2) the character of LLM adopted in that dialogue. Inter-coder reliability for strategy coding reached 59.06% in the first round and improved to 70.56% in the second round. For the character of LLM coding, reliability was 59.78% in the first round and increased to 66.40% in the second. Each round was followed by discussion, and final coding was established through consensus, resulting in 1,913 coded dialogue instances.

To capture users' self-reflected evaluations of knowledge states of users, the study adapted measures from Tang et al. (2013), designing nine items that covered aspects such as creativity, personal knowledge status, and dependency on individual expertise.

Original Questions	Concept
1. What prompting strategies have you used for this session? Choose as many as applicable.	7 kinds of Strategies
2. To what extent does this request rely on your personal context, domain knowledge, or evolving goals—making it harder to delegate to someone unfamiliar with your work?	non-delegability
3. How much does this request depend on iterative exploration guided by your own knowledge and personal interests, rather than a universal approach anyone could follow?	knowledge dependency
4. Does the request involve generating creative or novel solutions?	creativity
5. How familiar are you with the topic or domain related to this request?	topic familiarity
6. How strongly do you feel your expertise level (or lack thereof) shaped your approach to prompting and interpreting the LLM's responses?	knowledge status effect
7. To what extent is the knowledge you're seeking publicly documented and likely included in the LLM's training data, and how much might be local, private, or otherwise outside publicly accessible sources.	data publicity
8. Before asking questions to LLM, you expect LLM to give you satisfactory answers.	pre-expectation
9. How satisfied are you with the LLM's responses to your task? If you were dissatisfied, what factors contributed to that?	satisfaction

Table 3. Users' self-reflected evaluations of task characteristics.

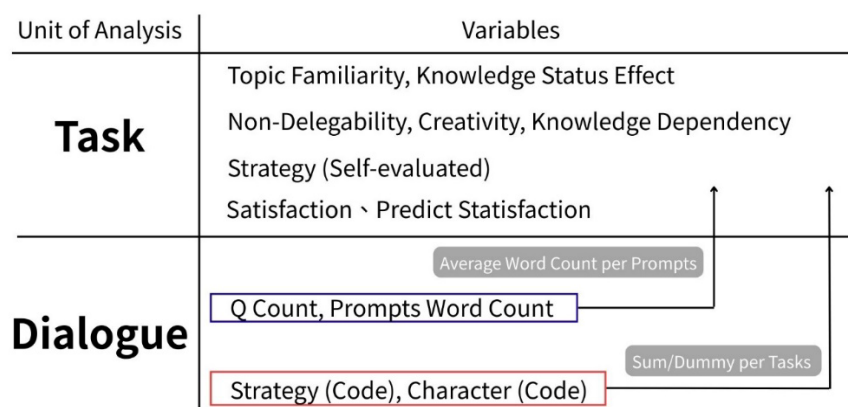


Figure 1. Unit of analysis of variables.

The above variables are from two kinds of analysis units (see Figure 1.):

1. Task unit: Knowledge states of users and user's self-evaluated prompt strategies by tasks.
 - Knowledge status effect: Drawing from discussions of knowledge status in the information retrieval domain and integrating perspectives from metacognition, this construct seeks to examine whether LLM users also resolve their information needs through dynamic interactions with the system.
 - Strategy matching rate: The degree of difference between users' self-reported evaluations of their employed prompt strategies and the strategies coded in this study.
 - Topic familiarity: The extent to which users are familiar with the subject matter involved in a given task.
 - Knowledge dependency: The extent to which the task relies on personal knowledge and interests.
 - Creativity: The degree to which the task entails creative attributes, such as literary composition.
 - Non-delegability: The extent to which the task cannot be delegated to others. The more this variable goes, the harder this task be delegated to others.
 - Satisfaction, predict satisfaction: Users' final satisfaction with the responses generated by the LLM, as well as their anticipated satisfaction prior to engaging in the task.
2. Dialogue Unit: Analysis by dialogues in one single task.
 - Q Count: Turns of questions users asked in one task.
 - Prompt word counts: Word counts in every question the user asked in one single task.
 - Strategy (Code)、 Character (Code): Made by Coders after evaluating each dialogue. Made into dummy variables and aggregate variables in order to compare with user self-evaluated ones.

Ethics statement

This study was conducted within the context of a university course and involved an optional bonus assignment. Students were informed in advance that their submitted materials might be used for research purposes, and participation in the study was entirely voluntary.

The assignment encouraged students to use generative AI tools for knowledge management in authentic learning contexts. Grading was based solely on students' demonstrated understanding

of knowledge management concepts and learning engagement and was completed prior to any research analysis. The criteria for grading were independent of the research variables and analytical outcomes. For research purposes, all collected materials were de-identified before analysis, and no personally identifiable information was accessible to the researchers during the analytical process. To promote transparency, aggregated descriptive statistics of the questionnaire results were shared with participating students after data collection. According to institutional and national guidelines, this study was classified as a minimal-risk educational study and therefore did not require formal ethics review by an Institutional Review Board.

Results

User self-reflection

The descriptive statistics of topic familiarity, knowledge status, knowledge dependency, creativity, non-delegability, data publicity, pre-expectation, and satisfaction are presented in Table 4.

Users reported varying levels of self-reported topic familiarity with the subjects involved in their tasks, self-reported of the creativity of the task was even more divergent. In contrast, knowledge status, non-delegability, knowledge dependency, data publicity, pre-expectation, and satisfaction were predominantly distributed in the higher range of the scale (>4). Notably, the average values of pre-expectation and actual satisfaction were comparable. For most tasks, there was no significant difference between anticipated and actual satisfaction. Moreover, the number of tasks in which actual satisfaction exceeded pre-expectation was greater than those in which it fell below expectations. This suggests that in most reflective accounts, users tended to experience satisfaction that was either consistent with or slightly higher than anticipated.

Variable	Mean	Std. Dev.	Min	Max
Topic familiarity	4.343915	1.407891	1	7
Knowledge status effect	5.380952	1.078246	2	7
Knowledge dependency	4.994709	1.548841	1	7
creativity	4.534392	1.555749	1	7
non-delegability	5.306878	1.233834	2	7
Data publicity	4.793651	1.489239	1	7
Pre expectation	5.21164	0.9440023	3	7
Response satisfaction	5.497354	1.089722	2	7

Table 4. Variable distribution of self-reflected evaluations users.

Use of prompt strategies and the character of LLM in tasks

Drawing upon the guidelines provided by Google Cloud and OpenAI, the study identified seven types of prompt strategies that users could select during self-reported reflection:

1. **S1: Chain of thoughts prompting**
2. **S2: Few-shot prompting**
3. **S3: Instruction or directive prompting**
4. **S4: Role prompting**
5. **S5: Self-consistency or 'double-check' prompting**
6. **S6: Contextual or narrative prompting**
7. **S7: Iterative refinement prompting**

In subsequent manual coding, researchers further classified interactions based on the purposes and modes of LLM use as summarised by Phillips et al. (2024) and Divekar et al. (2025). Specifically, seven categories of the character of LLM within tasks were defined: collaborator (C1), consultant (C2), Idea generator (C3), subordinate/ghostwriter (C4), dictionary and search engine (C5), Instructions (C6), and Partner for chat (C7). Both prompt strategies and the character of LLM were coded at the dialogue level. The coding results were then transformed into binary matrices and aggregated counts, providing the basis for advanced analysis and interpretation.

Usage of prompt strategies across 187 tasks

Figure 2 illustrates the use of seven prompt strategies across 187 tasks. The blue bars represent users' self-reported responses, while the red bars indicate the results of manual coding. The findings reveal that for S1-S5, the number of tasks where users self-reported employing the strategy was consistently higher than the number identified through coding, with the largest discrepancy observed in Few-Shot Prompting (S2). Conversely, for Contextual or Narrative Prompting (S6) and Iterative Refinement Prompting (S7), the number of tasks in which strategies were self-reported was lower than the number identified by coders.

When aggregated at the dialogue level, as shown in Figure 3, Instruction or Directive Prompting (S3), S6, and S7 emerged as the three most frequently applied prompt strategies overall. Whether individual prompt strategies aligned between self-reported and coded results will be examined in subsequent analyses.

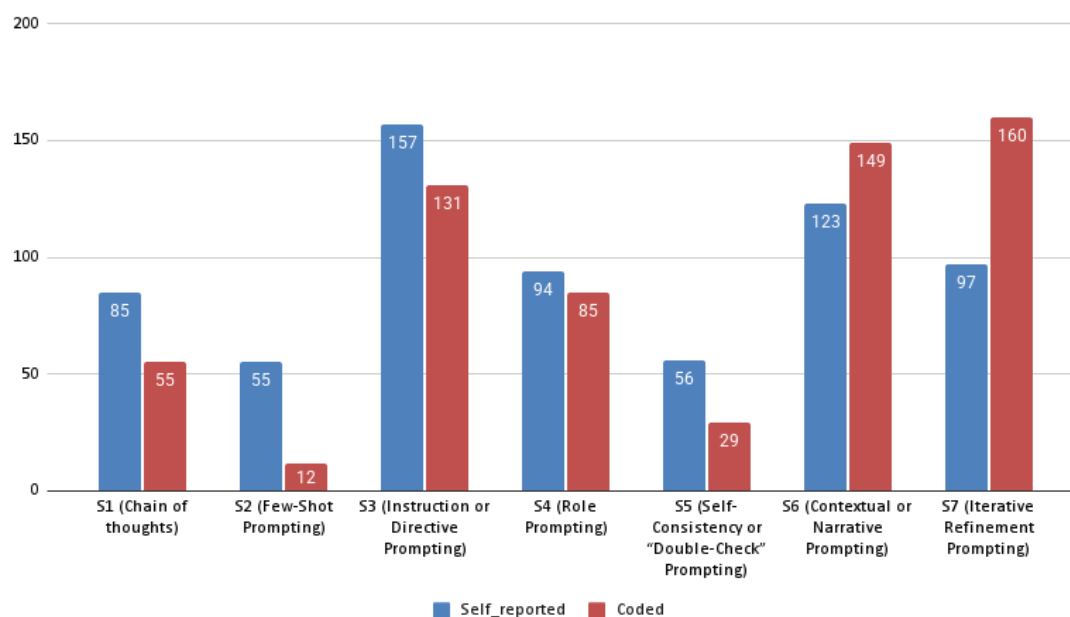


Figure 2. The use of seven prompt strategies in tasks.

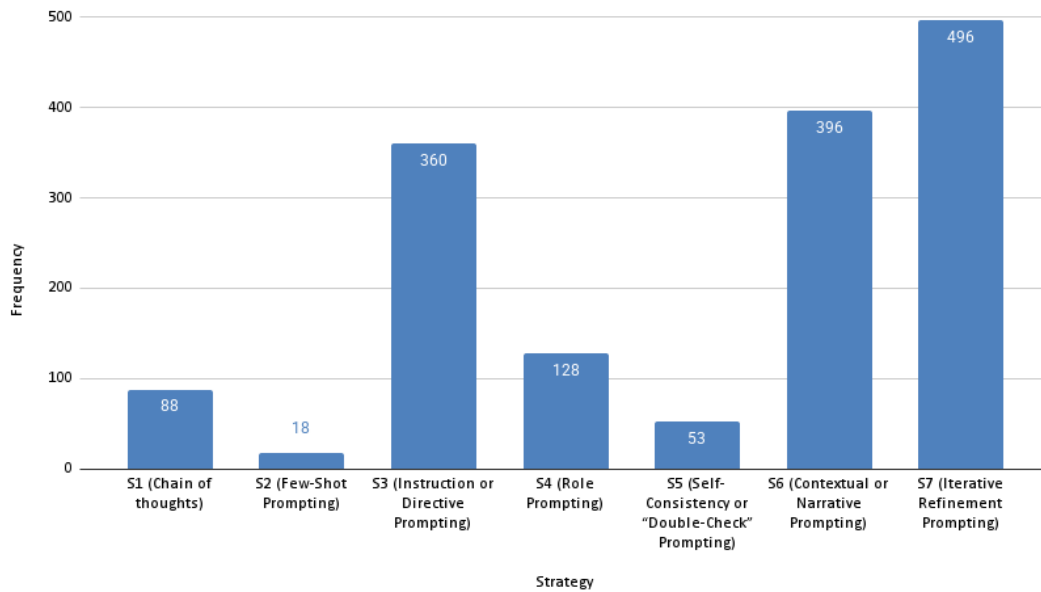


Figure 3. Frequency of seven prompt strategies in dialogue level.

Differences between users' self-reported prompt strategy use and coding results

By comparing users' self-reported prompt strategy use with the results of manual coding, four categories were identified: **Not used**, where both self-reports and coding indicated the absence of a strategy; **Self-rated use**, where users reported employing a strategy but coders did not identify it; **No conscious use**, where users did not self-report a strategy but coders identified its presence; and **Conscious use**, where both self-reports and coding confirmed the strategy's application.

The strategy most frequently self-reported by users but not confirmed through coding was S3, followed by S6, S7, and Role Prompting (S4). Strategies most often applied unconsciously—detected through coding but not reported by users—were Chain of thoughts (S1), S2, S3, and Self-Consistency or 'Double-Check' Prompting (S5). The strategy most consistently verified across both self-reports and coding was S7, followed by S6 (see Table 5 and Figure 4).

As shown in Figure 4, although S7 ranked third among strategies most frequently misperceived by users as applied (after S3 and S6), it also had the highest percentage of agreement between self-reports and coding. This suggests that users who consciously recognised their use of iterative refinement prompting demonstrated greater accuracy in defining and applying this strategy compared to other prompt strategies. This finding corresponds with McCabe's (2011) observation regarding students' awareness of learning strategies, and it resonates with the self-awareness experiences that users supplemented in the qualitative questions of this study's questionnaire.

Strategy	Not used		Self-rated use		No conscious use		Conscious use	
S1	85	45.45%	38	20.32%	47	25.13%	17	9.09%
S2	128	68.45%	8	4.28%	47	25.13%	4	2.14%
S3	13	6.95%	114	60.96%	43	22.99%	17	9.09%
S4	90	48.13%	82	43.85%	12	6.42%	3	1.60%
S5	122	65.24%	20	10.70%	36	19.25%	9	4.81%
S6	21	11.23%	106	56.68%	17	9.09%	43	22.99%
S7	17	9.09%	87	46.52%	10	5.35%	73	39.04%

Table 5. Differences in strategies between self-assessment and coding results of tasks.

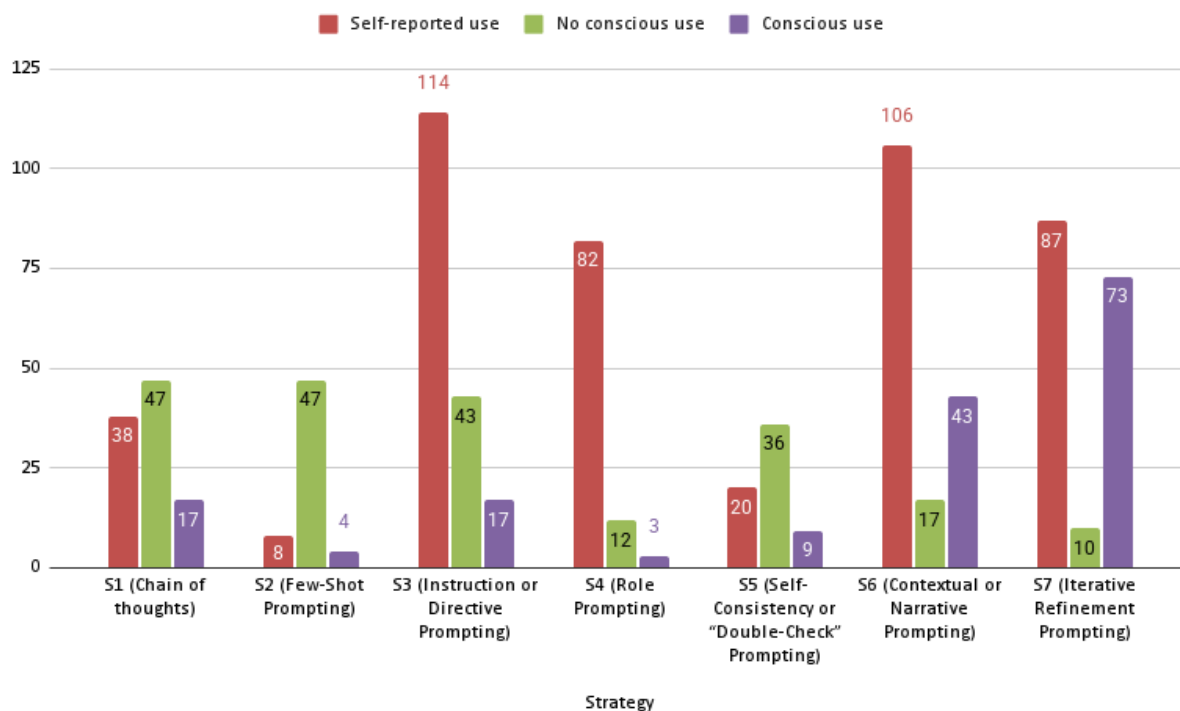


Figure 4. Differences in strategies between self-assessment and coding results in tasks.

Character of LLM across 187 tasks

To better understand how users positioned the LLM during task execution, researchers examine the roles the LLM assumed across all 187 tasks. Figure 5 present the cumulative frequency of the task characters of LLM across 187 tasks. Consultant (C2) and Subordinate/Ghostwriter (C4) were observed with significantly higher frequency—both in terms of presence/absence and cumulative counts—compared to other roles. This indicates that users tended to perceive the LLM either as a consultant providing professional knowledge and advice or as a tool delegated to complete designated tasks on their behalf.

Given that the number of conversations varied across tasks, researchers calculated the proportion of each character of LLM within individual tasks to identify the main character. As shown in Table 6, nearly 60% of the tasks primarily positioned the LLM as C2 (Consultant). This finding suggests that the majority of users, during their interactions, regarded the LLM as a consultative partner capable of offering domain-specific expertise and professional insights.

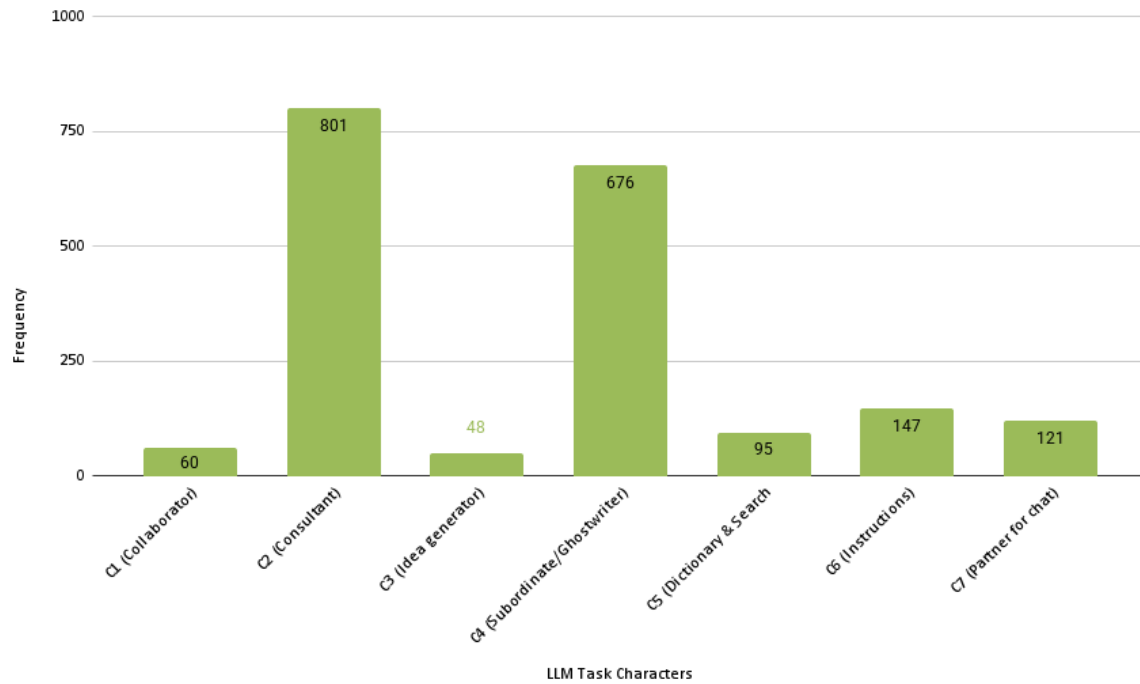


Figure 5. Frequency of the task characters of LLM across 187 tasks.

Main character	C1	C2	C3	C4	C5	C6	C7	Total
Frequency	4	110	3	54	3	7	6	187
Percentage	2.14%	58.82%	1.60%	28.88%	1.60%	3.74%	3.21%	100.00%

Table 6. Number of each main character of LLM within tasks.

Distribution of prompt strategy usage across tasks with different main characters

Figure 6 illustrate the application of seven prompt strategies across groups categorised by the main character of LLM. Among these, tasks in which C2 (Consultant) and C4 (Subordinate/Ghostwriter) served as the main character were the most frequent; accordingly, the strategy counts for these two groups were also the highest in Figure 6. Within both groups, S3, S6, and S7 consistently emerged as the three most frequently employed strategies.

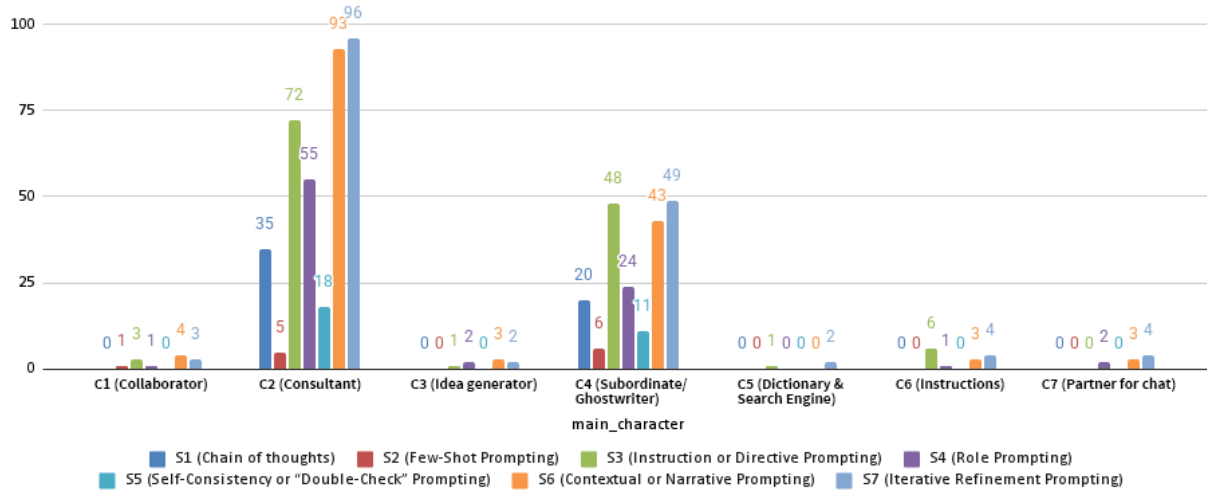


Figure 6. Frequency of seven self-reported prompt strategies across main task characters.

User characteristics within tasks

Using strategy matching rate to indicate users' LLM application ability

In this study, we deliberately avoid relying on self-reported questionnaires to measure users' LLM application ability by operationalising users' understanding of prompt strategy use by comparing the agreement between their self-reported strategies for a given task and the strategies identified through researcher coding. The *strategy matching rate* is computed via the Jaccard similarity coefficient. Specifically, we first transform the self-reported strategies into a set S_{self} and the coded strategies into a set S_{coded} . We then calculate the sizes of the intersection and union of these two sets; dividing the former by the latter yields a ratio in $[0,1]$, where larger values indicate stronger agreement:

$$strategy_matching_rate = \frac{|S_{self} \cap S_{coded}|}{|S_{self} \cup S_{coded}|}$$

Hereafter, we use the *strategy matching rate* as the indicator of users' LLM application ability for subsequent, more fine-grained analyses.

Analysis of users' LLM application ability and adoption of the character of LLM

This study divided all tasks into high- and low-scoring groups based on the median value of the *strategy matching rate*. For each of the seven characters of LLM (C1–C7), we calculated the average frequency of use at the task level (count variable, coded as non-negative integers).

Because the distributions of character usage frequencies were skewed and contained many zero values, the Mann–Whitney U test was employed to compare differences between the two groups. The results indicated that, with the exception of C2, no significant differences were observed across characters (all p -values $> .05$). For C2, however, the average usage frequency in the high-scoring group was significantly greater than in the low-scoring group ($p = .007 < .05$).

This finding suggests that the intensity of reliance on C2 (Consultant) was higher among users with stronger LLM application ability, indicating that this role may serve as a key marker distinguishing between the two groups.

role (Count)	Average usage count of	Average usage count of	difference	p-value
	high-matching-rate group	low-matching-rate group		
C1 (Collaborator)	0.368	0.272	0.097	0.142
C2 (Consultant)	4.853	3.696	1.157	0.007**
C3 (Idea generator)	0.316	0.196	0.12	0.436
C4 (Subordinate/Ghostwriter)	4.242	2.967	1.275	0.704
C5 (Dictionary & Search Engine)	0.347	0.674	-0.327	0.398
C6 (Instructions)	0.789	0.783	0.007	0.578
C7 (Partner for chat)	0.442	0.859	-0.417	0.9

Table 7. Outcomes of the mixed-effects logistic regression model

Analysis of knowledge states of users and prompt content

We first conducted correlation analyses on variables related to prompts, which are further underlined in the correlation matrix below, including total word count, number of conversational turns (Q count), and average word count per query, and the knowledge states of users collected in task units (see Figure 7).

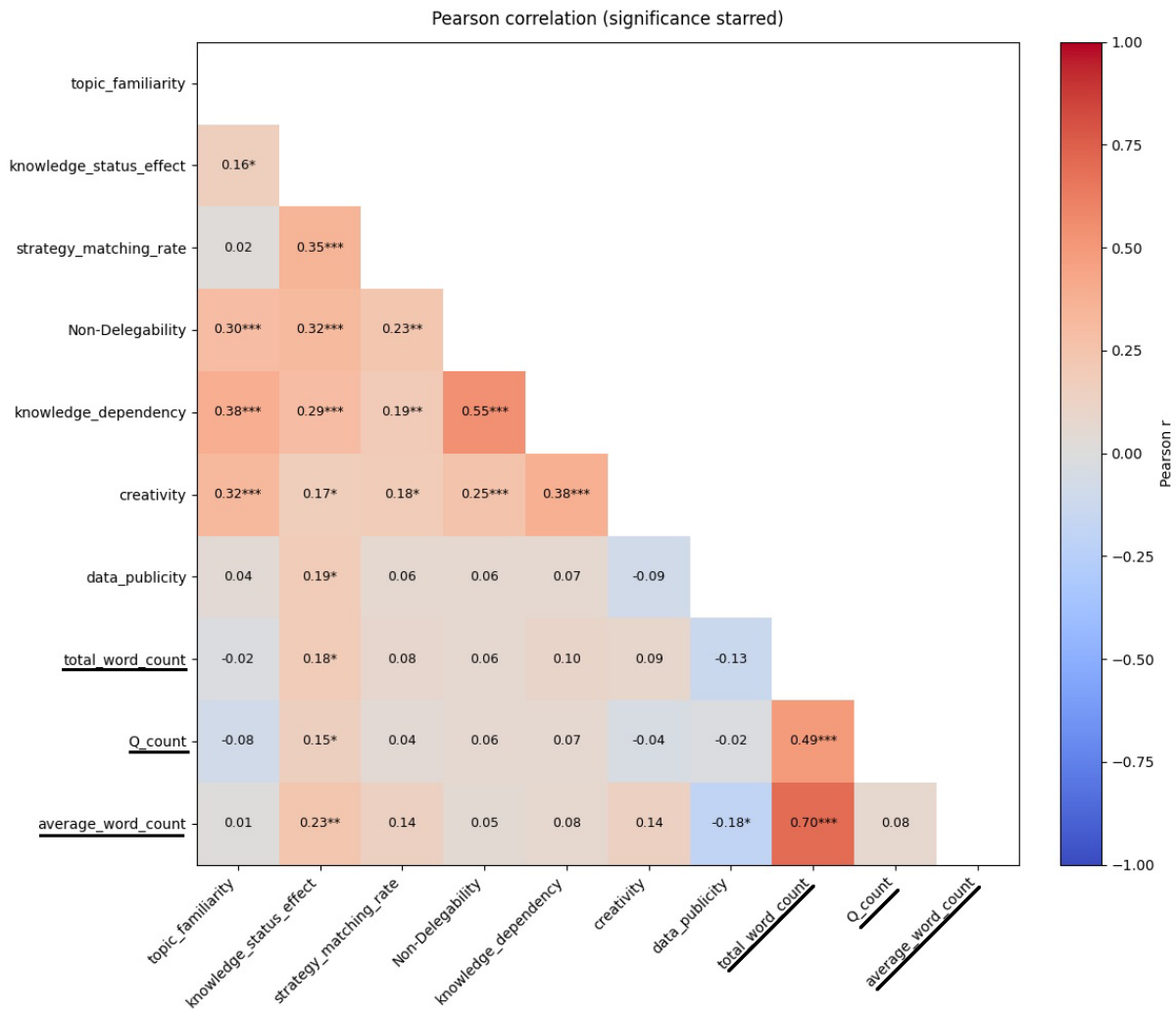


Figure 7. Correlation matrix of knowledge states of users and prompt content.

We found that users' knowledge status in a given task was significantly correlated with total word count, number of turns, and average word count. This finding aligns with prior literature in the information retrieval domain, suggesting that LLM users address the demands of their current knowledge status through sustained interaction and more elaborated textual exchanges. The correlation between knowledge status and data publicity may reflect users' subjective perceptions regarding the openness of data involved in the task. From a metacognitive perspective, this could be interpreted as users' awareness that insufficiently public data might not be fully captured by the LLM, thereby limiting its ability to provide accurate responses. Similarly, the correlation between knowledge status and strategy matching rate can also be explained from a metacognitive standpoint: users' metacognition is closely related to their LLM application ability in selecting and employing prompt strategies to achieve task goals, which resonates with prior scholarly findings. Furthermore, topic familiarity was significantly correlated with creativity and knowledge dependency. This may be interpreted as users, once recognising that their domain knowledge is sufficient, are more capable of performing creative tasks in that domain and more likely to perceive such tasks as reliant on personal knowledge and interests.

The correlation between knowledge dependency and creativity may indicate that creative tasks inherently rely on personal knowledge. The significant correlation between non-delegability and knowledge dependency suggests that when tasks are embedded within the context of personal

knowledge and interests, they are less likely to be delegated to others. The correlation between non-delegability and topic familiarity further indicates that tasks with specific contextual demands—thus difficult to delegate—may require higher levels of domain knowledge, which corresponds with findings in information retrieval research. Finally, the correlation between non-delegability and strategy matching rate may suggest that in such types of tasks, users need to demonstrate a more advanced understanding of LLM application ability in order to apply prompt strategies effectively.

Analysis of knowledge states of users and self-reported prompt strategies

We then conducted correlation analyses and the knowledge states of users collected in task units, and LLM prompt strategies. Since our primary aim was to examine how task characteristics from users influenced their choice of prompt strategies, we adopted self-reported strategies (converted into dummy variables) rather than coder-interpreted strategies. Second, as the strategy matching rate is derived from the comparison between self-reported strategies and coder-interpreted strategies, it was not appropriate to include it in the present correlation analysis. The resulting correlation matrix is presented in Figure 8.

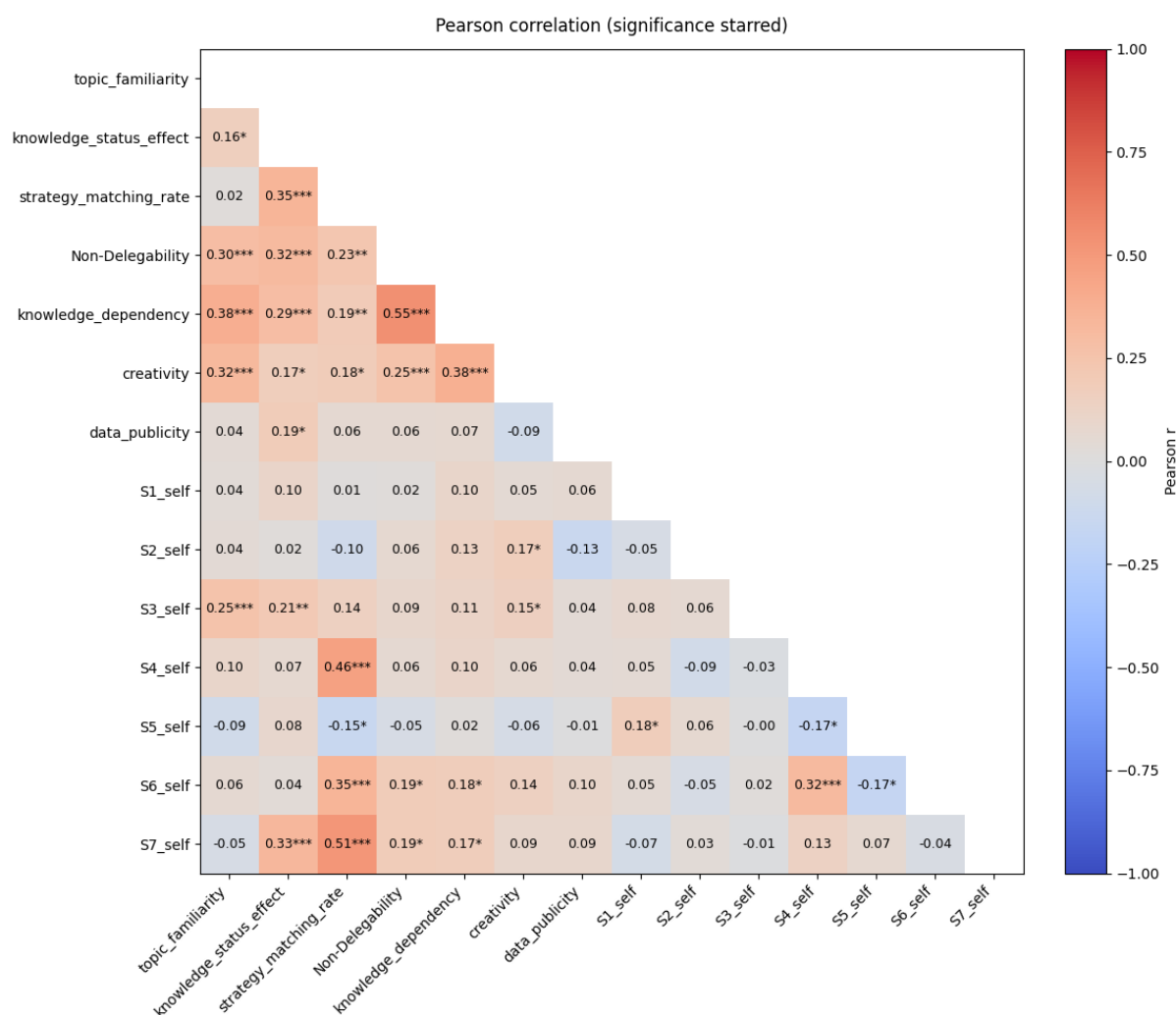


Figure 8. Correlation matrix of knowledge states of users and self-reported prompt strategies.

According to the correlation matrix, topic familiarity was significantly correlated with S3, suggesting that higher levels of domain knowledge may enable users to incorporate more explicit instructions into their prompts. Notably, the knowledge status effect was significantly correlated

with S7, which once again echoes findings from prior information retrieval studies: users externalise their knowledge status through sustained interaction with the system. In the context of LLM use, when users become aware that their current knowledge status cannot be clearly expressed through a single query, they are more likely to employ iterative strategies in interacting with the LLM.

Moreover, non-delegability was significantly correlated with both S6 and S7. When a task is context-specific and difficult to delegate, users tend to employ S6 by explaining the problem's background and context to the LLM, while also relying on S7 to resolve the issue through multiple rounds of engagement. Similarly, when a task depends on personal knowledge and interests, users are also required to gradually articulate their information needs to the LLM through these strategies. Finally, with respect to creativity, the correlation matrix suggests that users may be inclined to adopt S4 when engaging in creative tasks. At the same time, they may also attempt to employ explicit instructions to reduce the potential unpredictability inherent in creative outputs.

Effects of prompt strategies and knowledge states of users on satisfaction

Defining satisfaction groups to represent satisfaction groups

The final aim of this study was to examine whether knowledge states of users and the use of prompt strategies influenced satisfaction. Before conducting the analysis, we categorised each task into groups based on responses from the self-reported reflection survey. Specifically, if the difference between satisfaction and pre-expectation was positive, the task was assigned to the first group, indicating that the user's final satisfaction exceeded expectations. Conversely, if the difference was negative, the task was assigned to the second group, indicating that final satisfaction fell below expectations. We define these categories collectively as the *satisfaction group*, which served as the dependent variable in subsequent analyses.

Analysing the influence of prompt strategies and task characteristics from users on satisfaction groups

To predict the binary outcome of the *satisfaction group*, we employed a mixed-effects logistic regression model. A random intercept was specified for each user to account for the non-independence of multiple task records from the same individual. The model was estimated using the **BinomialBayesMixedGLM** procedure in *statsmodels* with a Bayesian approximation method.

Variable	Coef.	OR	OR_CI_low	OR_CI_high
Intercept	-0.14	0.87	0.6	1.25
Topic familiarity	-0.09	0.92	0.85	0.99
Knowledge status effect	-0.21	0.81	0.76	0.87
Strategy matching rate	-0.05	0.95	0.52	1.72
Non-delegability	0.2	1.23	1.15	1.31
Knowledge dependency	-0.34	0.71	0.66	0.76
creativity	0.36	1.43	1.32	1.54
Data publicity	0.07	1.07	0.99	1.16
S1_count	0.74	2.1	1.44	3.06
S2_count	0.66	1.93	0.75	4.97
S3_count	-0.25	0.78	0.68	0.9
S4_count	-0.59	0.55	0.39	0.78

Table 8. Outcomes of the mixed-effects logistic regression model.

Both *topic familiarity* (OR = 0.92, 95% CI = [0.85, 0.99]) and the *knowledge status effect* (OR = 0.81, 95% CI = [0.76, 0.87]) showed significant negative effects. This suggests that users with higher familiarity with the topic were more likely to be dissatisfied with LLM outputs, consistent with Palta et al. (2025), while those who heavily relied on knowledge exploration or whose interactions were strongly shaped by their *knowledge status* experienced greater difficulty in achieving outcomes that exceeded expectations.

Non-delegability (OR = 1.23, 95% CI = [1.15, 1.31]) positively predicted the likelihood of actual *satisfaction* surpassing anticipated levels, which shows that although the task is hard to delegate to other, the satisfaction may still exceed expectation. In contrast, *knowledge dependency* (OR = 0.71, 95% CI = [0.66, 0.76]) was negatively associated with outcomes, implying that tasks requiring personal knowledge or interests are less likely to produce unexpectedly high *satisfaction*. Conversely, *creativity* (OR = 1.34, 95% CI = [1.25, 1.52]) exerted a positive effect, indicating that LLMs are particularly effective in enhancing user experience when supporting creative tasks.

In terms of prompt strategy use, chain-of-thought prompts (S1; OR = 2.10, 95% CI = [1.44, 3.06]) and repetitive confirmation prompts (S5; OR = 1.80, 95% CI = [1.21, 2.68]) were positively associated with outcomes exceeding expectations, suggesting that these strategies enhance users' experience of task completion. By contrast, explicit instruction (S3; OR = 0.78, 95% CI = [0.68, 0.90]) and role specification (S4; OR = 0.55, 95% CI = [0.39, 0.78]) had negative effects, implying that overreliance on these strategies may hinder users from achieving higher-than-anticipated *satisfaction*. Other variables, such as *strategy matching rate* and data publicity, did not reach statistical significance.

Taken together, the results of the mixed-effects logistic regression demonstrate that whether users' satisfaction exceeds their expectations is shaped not only by knowledge states of users (*topic familiarity*, *knowledge status*, *non-delegability*, *knowledge dependency*, and *creativity*), but also the choice of prompt strategies.

Qualitative evidence

We also collected reasons for users' explanations of LLM responses, in which users explained why the satisfaction exceeded or fell behind expectation. Most users stressed the evaluation of the model's generated output as the primary basis for their satisfaction ratings. However, some users noted that they believed the generated responses were influenced by their own knowledge status or the prompt strategies they adopted. This observation resonates with findings from the literature review and the hypotheses of this study.

Furthermore, when the model's responses did not align with expectations, some users reflected on whether this discrepancy might be attributable to their own knowledge status:

I felt that the problem was that I was not entirely certain about what outcome I wanted at the moment, which led to more diffuse responses from the LLM... However, within these diffuse responses, I still found the answers I was seeking, and through them, I was able to further explore other resources to deepen my understanding.

I could not fully determine the reliability of the recipes provided by ChatGPT or Gemini, which may be related to my limited cooking skills and lack of experience to judge their quality.

I cannot be completely certain whether the vocabulary information and usage offered by the AI are entirely accurate, and I am reluctant to verify them one by one. Nevertheless, I still tend to trust its linguistic ability and knowledge...

The above outcomes show the existence of the user's metacognitive, which may further affect the final satisfaction as we have proved in the statistical outcome.

Conclusion

This study examined how users employ prompt strategies, how their knowledge status and other knowledge states of users influence prompting behaviors, and how these factors collectively shape satisfaction with LLM-assisted tasks. Results from 187 tasks showed that while directive (S3), contextual (S6), and iterative refinement (S7) were the most frequently applied strategies, discrepancies existed between self-reported and coded data, particularly in the over-reporting of directive prompting. At the level of the character of LLM, users predominantly positioned the system as a consultant (C2) or subordinate/ghostwriter (C4), reflecting its dual role as both advisor and executor.

Findings further demonstrated that knowledge states of users play a significant role: higher topic familiarity facilitated directive prompting, knowledge status effect was strongly linked to iterative refinement, and non-delegability encouraged both contextual and iterative prompting. Satisfaction outcomes were shaped not only by these characteristics but also by strategy choice. Non-delegability increased the likelihood of exceeding expectations, while heavy reliance on knowledge status reduced it. Strategically, chain-of-thought (S1) and self-consistency (S5) enhanced satisfaction, whereas directive (S3) and role prompting (S4) had negative effects. Taken together, these findings highlight that satisfaction in LLM-assisted tasks emerges from the interplay of knowledge states, task contexts, and strategy use. By introducing strategy matching rate as an indicator of LLM application ability, this study also contributes a novel approach to evaluating users' capacity to apply prompt strategies, offering implications for AI literacy research and future investigations of human-LLM collaboration.

Acknowledgements

The authors would like to thank the reviewers for their constructive comments and valuable suggestions. We also sincerely thank the students who participated in the course assignments and assessments that formed the basis of this study. Their thoughtful engagement and reflections made this research possible.

About the authors

Tzu-Yun Chien is a master's student in the Department of Library and Information Science at National Taiwan University. She is interested in human-computer interaction, AI collaboration, and information behavior and decision-making. She can be reached at r13126006@ntu.edu.tw

Po-Yu Chen is a master's student in the Department of Library and Information Science at National Taiwan University. He focuses his studies on information retrieval, human-computer interaction and music streaming information behavior. He can be contact at r12126016@ntu.edu.tw

Muh-Chyun Tang is a Professor in the Department of Library and Information Science at National Taiwan University. He is the Principal Investigator of the Self, Network & Algorithm Lab (SNA Lab), where his research focuses on how human psychology, social networks, and algorithmic systems co-evolve in digital environments. His research interests include recommender systems, human-AI interaction, cognitive bias, network analysis, and music and self-identity.

References

- Ali, S. J., Reinhartz-Berger, I., & Bork, D. (2025). How are LLMs Used for Conceptual Modeling? An Exploratory Study on Interaction Behavior and User Perception. In W. Maass, H. Han, H. Yasar, & N. Multari (Eds.), *Conceptual Modeling* (pp. 257–275). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-75872-0_14
- Belkin, N. (2014). Anomalous States of Knowledge as A Basis for Information Retrieval. *Canadian Journal of Information Science*, 133–143.
- Divekar, R. R., Gonzalez, L., Guerra, S., Boos, N., & Divekar, R. (2025). Can Generative AI Replace Search Engines for Learning? Understanding Student Preferences, Use and Perceived Proficiency in Using AI. *TechTrends*. <https://doi.org/10.1007/s11528-025-01095-9>
- Eigner, E., & Händler, T. (2024). Determinants of LLM-assisted Decision-Making (No. arXiv:2402.17385). arXiv. <https://doi.org/10.48550/arXiv.2402.17385>
- Glaser, B., & Strauss, A. (2017). *Discovery of grounded theory: Strategies for qualitative research*. Routledge.
- Holstein, J., Diener, M., & Spitzer, P. (2025). From Consumption to Collaboration: Measuring Interaction Patterns to Augment Human Cognition in Open-Ended Tasks (No. arXiv:2504.02780). arXiv. <https://doi.org/10.48550/arXiv.2504.02780>
- Ju, B., & Stewart, J. B. (2024). Empowering Users with ChatGPT and Similar large language models (LLMs): Everyday Information Needs, Uses, and Gratification. *Proceedings of the Association for Information Science and Technology*, 61(1), 172–182. <https://doi.org/10.1002/pra2.1018>
- Kacperski, C., Ulloa, R., Bonnay, D., Kulshrestha, J., Selb, P., & Spitz, A. (2025). Characteristics of ChatGPT users from Germany: Implications for the digital divide from web tracking data. *PLOS ONE*, 20(1), e0309047. <https://doi.org/10.1371/journal.pone.0309047>
- Kim, S., Eun, J., Park, Y. E., Lee, K., Lee, G., & Lee, J. (2025). PromptPilot: Exploring User Experience of Prompting with AI-Enhanced Initiative in LLMs. *International Journal of Human-Computer Interaction*, 1–23. <https://doi.org/10.1080/10447318.2025.2489030>
- Kim, Y., Lee, J., Kim, S., Park, J., & Kim, J. (2024). Understanding Users' Dissatisfaction with ChatGPT Responses: Types, Resolving Tactics, and the Effect of Knowledge Level. *Proceedings of the 29th International Conference on Intelligent User Interfaces*, 385–404. <https://doi.org/10.1145/3640543.3645148>
- Lee, S., & Park, G. (2023). Exploring the Impact of ChatGPT Literacy on User Satisfaction: The Mediating Role of User Motivations. *Cyberpsychology, Behavior, and Social Networking*, 26(12), 913–918. <https://doi.org/10.1089/cyber.2023.0312>
- McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & cognition*, 39(3), 462–476.
- Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of educational psychology*, 94(2), 249.
- Open AI, & Shieh, J. (2024). Best practices for prompt engineering with the OpenAI API.
- Palta, S., Chandrasekaran, N., Rudinger, R., & Counts, S. (2025). Speaking the Right Language: The Impact of Expertise Alignment in User-AI Interactions (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2502.18685>

- Phillips, N., Kalvapalle, S., & Kennedy, M. (2024). Beyond the Turing Test: Exploring the implications of generative AI for category construction. *Organisation Theory*, 5(3), 26317877241275113. <https://doi.org/10.1177/26317877241275113>
- Prompt Engineering for AI Guide. (n.d.). Google Cloud. <https://cloud.google.com/discover/what-is-prompt-engineering?hl=en>
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary educational psychology*, 19(4), 460-475.
- Song, Y., Huang, L., Zheng, L., Fan, M., & Liu, Z. (2025). Interactions with generative AI chatbots: Unveiling dialogic dynamics, students' perceptions, and practical competencies in creative problem-solving. *International Journal of Educational Technology in Higher Education*, 22(1), 12. <https://doi.org/10.1186/s41239-025-00508-2>
- Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior*, 160, 108386. <https://doi.org/10.1016/j.chb.2024.108386>
- Tang, M.-C., Liu, Y.-H., & Wu, W.-C. (2013). A study of the influence of task familiarity on user behaviors and performance with a MeSH term suggestion interface for PubMed bibliographic search. *International Journal of Medical Informatics*, 82(9), 832-843. <https://doi.org/10.1016/j.ijmedinf.2013.04.005>
- Wang, J., Ma, W., Sun, P., Zhang, M., & Nie, J.-Y. (2024). Understanding User Experience in Large Language Model Interactions (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2401.08329>
- Zhu, G., Sudarshan, V., Kow, J. F., & Soon Ong, Y. (2024). Human-Generative AI Collaborative Problem Solving Who Leads and How Students Perceive the Interactions. 2024 IEEE Conference on Artificial Intelligence (CAI), 680-686. <https://doi.org/10.1109/CAI59869.2024.00133>

© [CC-BY-NC 4.0](#) The Author(s). For more information, see our [Open Access Policy](#).

Appendix

Prompt strategies for task

code	Strategy	Definition	Keyword
S1	Chain of thoughts	Breaking down complex requests into numbered or clearly structured steps helps ensure nothing is overlooked. By guiding the model through a series of smaller tasks, users can produce more thorough and accurate responses. This approach also makes the reasoning process clearer and more transparent.	'Follow the procedures...' 'One step at a time, and help me deal with it first...'
S2	Few-Shot Prompting	Provide the model with a few examples of the input-output relationship before giving the final query. It helps steer the model toward a specific style, tone, or format by showing exactly what kind of output the user wants.	'Please output the result in the format of...'
S3	Instruction Directive Prompting or	Explicitly instruct the model on how to format or structure its response. This includes telling the model about its role, style, or constraints on the answer (e.g., length limit, bullet points, formality). It reduces ambiguity and gives users more control over the final output.	'The word limit is...' 'Please explain in a way that even elementary school students can understand.'
S4	Role Prompting	Assign the model a specific persona or role—like a teacher, doctor, or historian—and ask it to respond from that perspective. It helps the model align its tone, vocabulary, and content with the role users have described, often producing more contextually relevant and authoritative answers.	'You are a...(role)'
S5	Self-Consistency or 'Double-Check' Prompting	After the model provides an answer, encourage it to verify or reflect on its correctness, sometimes asking it to explain why its solution is correct or to reevaluate potential errors. This strategy reduces mistakes and helps identify inconsistencies in the model's output.	'Please explain the reason for your answer.' 'Your answer is different from the actual rules...' 'Please look back.'

S6	Contextual Narrative Prompting	or Give the model a backstory or detailed context to help it generate more accurate or coherent responses, especially useful for creative writing or scenario-based queries. Providing the model with additional context helps it maintain consistency and accuracy when dealing with complex narratives or specialised knowledge.	'I need you to do this because of (background)...'
S7	Iterative Refinement Prompting	Iteratively refine the prompt based on the model's previous output. users might ask follow-up questions or request modifications, clarifications, or expansions. This allows the user to home in on exactly the answer or format the user needs without starting from scratch. It also helps correct misunderstandings or incomplete answers.	'Please further explain... Please focus on the ○○ aspect.'
S0	N/A	None of the strategies above are used.	Say 'thank you' to the LLM or unfinished query

Characters of LLM plays in the task

code	Definition	Description
C1	Collaborator	Users and AI perform tasks together (e.g., writing or planning).
C2	Consultant	Ask the AI for its opinions, suggestions, or how to proceed.
C3	Idea generator	Inspirational and brainstorming purposes (e.g., providing list of points, examples, or naming).
C4	Subordinate/Ghost writer	The AI completes an entire paragraph for the user, performing tasks on their behalf.
C5	Dictionary & Search Engine	Short queries targeting concepts or knowledge.
C6	Instructions	Ask the AI to explain a process, rule, or usage, such as 'Explain how to calculate the Pearson correlation coefficient.'
C7	Partner for chat	Non-task communication, such as expressing emotions or passing time.