RESEARCH-ARTICLE

When Efficiency Meets Fulfillment: Understanding Long-Term LLM Integration in Knowledge Work

CHARLOTTE KOBIELLA, Ludwig-Maximilians-University Munich, Munich, Bayern, Germany

TEODORA MITREVSKA, Ludwig-Maximilians-University Munich, Munich, Bayern, Germany

ALBRECHT SCHMIDT, Ludwig-Maximilians-University Munich, Munich, Bayern, Germany

FIONA DRAXLER, University of Mannheim, Mannheim, Baden-Wurttemberg, Germany

CHIWORK '25: Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work (June 2025)
https://doi.org/10.1145/3729176.3729180
ISBN: 9798400713842

When Efficiency Meets Fulfillment: Understanding Long-Term LLM Integration in Knowledge Work

Charlotte Kobiella
LMU Munich
Munich, Germany
kobiella@cdtm.de

Teodora Mitrevska
LMU Munich
Munich, Germany
teodora.mitrevska@ifi.lmu.de

Albrecht Schmidt
LMU Munich
Munich, Germany
albrecht.schmidt@ifi.lmu.de

Fiona Draxler
University of Mannheim
Mannheim, Germany
fiona.draxler@uni-mannheim.de

Abstract

Large Language Models (LLMs) are transforming knowledge work across industries. While current HCI research emphasizes productivity, users' subjective experience of integrating LLMs into daily work has received less attention. We conducted a follow-up study of our 2023 investigation by employing a two-week diary study with 15 participants, exploring how prolonged LLM use shapes perceived productivity, accomplishment, and self-efficacy as psychological dimensions of work. Our findings reveal higher ratings of perceived accomplishment with prolonged LLM use. Efficiency emerged as a main driver not only for productivity but also for enhancing users' sense of accomplishment and self-efficacy, suggesting that thoughtful LLM integration can create more meaningful work experiences. Our research advances our understanding of technology adoption and adaptation, providing insights for developing tools and processes that honor personal fulfillment while leveraging technological advancement.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; **User studies**.

Keywords

Large Language Models, adoption, adaptation, productivity, self-efficacy, sense of accomplishment

ACM Reference Format:
Charlotte Kobiella, Teodora Mitrevska, Albrecht Schmidt, and Fiona Draxler. 2025. When Efficiency Meets Fulfillment: Understanding Long-Term LLM Integration in Knowledge Work. In *CHIWORK '25: Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work (CHIWORK '25), June 23–25, 2025, Amsterdam, Netherlands.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3729176.3729180

CHIWORK '25, Amsterdam, Netherlands

ACM ISBN 979-8-4007-1384-2/25/06
https://doi.org/10.1145/3729176.3729180

1 Introduction

Since OpenAI's launch of ChatGPT to the general public in November 2022, LLMs have garnered significant attention and have been adopted across various domains (e.g., [36, 60, 67]). While the long-term impact of these tools is still unfolding, recent observations indicate notable changes in work practices and interactions with technology [17, 66]. Simultaneously, the rapid integration of LLMs into everyday tasks has also raised critical questions about job displacement, deskilling, and potential adverse cognitive effects [1–3, 8, 12, 59].

As LLMs continue to disrupt traditional workflows, it becomes increasingly crucial to understand how individuals work, learn, and interact with LLMs beyond mere productivity gains, and how this develops with sustained use. While initial research has primarily focused on quantitative productivity metrics [44] and is now moving to industry-specific applications [28, 31, 43], there is a notable gap in our understanding of the psychological impacts of LLM adoption: How do users' perceptions and sense-making processes evolve as they become accustomed to these powerful AI (artificial intelligence) tools? This subjective perspective is essential for understanding long-term effects on workplace well-being and can inform sustainable human-centered technology integration.

To address these critical questions, we conducted a follow-up study with 15 out of 21 participants from our initial 2023 investigation [34] to explore the longitudinal effects of AI usage. Our research employed a mixed-methods approach, combining a two-week diary study with in-depth interviews to capture both quantitative measures and qualitative insights that contextualize participants' experiences and thought processes.

Our findings reveal a nuanced picture of LLM adoption and its psychological impacts. Notably, we observed that participants now report feeling more accomplished when using LLMs compared to the previous year. This increase appears to stem from a growing understanding of the LLMs' capabilities and limitations and subsequent adaptation of interaction strategies ultimately leading to perceived efficiency gains. As users become more familiar with LLMs over time, their initial feelings of inferiority appear to diminish, replaced by a more confident and strategic approach to AI collaboration. This also highlights the importance of longitudinal observations and studies in understanding the impact of disruptive technologies on human cognition and work practices beyond novelty effects.

2 Related Work

AI and LLMs have been widely discussed since the launch of ChatGPT. Since then, academics and practitioners have been exploring ways to interact effectively with LLMs. From prompt engineering tutorials to cheat sheets, productivity is often seen as one of the main drivers of LLM adoption [17, 24]. In our research, we aim to provide a long-term view of what it means to adopt and adapt to technology as highly disruptive as LLMs from a subjective user-centric perspective building on our study conducted in 2023 [34] (see section 2.4 for an overview of our 2023 results).

2.1 Technology Adoption, Adaptation, and Appropriation

Technology adoption is the process of how individuals and organizations embrace and implement new technological innovations [25]. Technology adoption involves the initial acceptance of technology and the subsequent learning curve and adaptation required to integrate it effectively into existing practices. Multiple factors influence this adoption process, including the expected performance benefits, the perceived ease of use, the availability of supporting resources and infrastructure, and the impact of social norms and peer influence.

One model that tries to explain technology adoption is the Technology Acceptance Model (TAM) developed by Davis [25, 26]. The original model suggests that two primary factors influence an individual's intention to use a system: Perceived Usefulness (PU) and Perceived Ease of Use (PEOU). PU describes the degree to which a person believes that using a particular system would enhance their job performance, while PEOU describes the degree to which they believe that using the system would be free of effort. The extension of the original TAM, TAM2, was developed by Venkatesh and Davis in 2000 [63, 64]. It builds on the original TAM by incorporating additional determinants of perceived usefulness and usage intentions. Key added factors in TAM2 include social influence processes, such as subjective norms, and cognitive instrumental processes, such as job relevance.

While TAM and TAM2 focus on initial acceptance and use of technology, researchers have also explored how users adapt and appropriate technology over time. Technology adaptation refers to the process by which users modify their use of technology to better suit their needs or context [7]. This concept recognizes that users do not always use technology exactly as designed but may find creative ways to apply it to their specific situations. Technology appropriation, a related concept, describes how users take possession of a technology's capabilities and features, making the technology their own through use [18]. This process involves users exploring, evaluating, and adopting or rejecting various features of a technology. Salovaara and Tamminen [50] argue that the concept of appropriation is particularly important for understanding post-adoption usage of mature technologies. They suggest that users often find innovative ways to use technologies beyond their original design intent and that this process of appropriation can lead to different usage patterns and perceptions among users. This perspective challenges the notion of a uniform acceptance of technology and instead proposes a more heterogeneous view of technology use. These perspectives on adaptation and appropriation complement the acceptance models by highlighting that technology adoption is not a one-time event but an ongoing process of negotiation between users and technology [57].

As adoption is a process, and it is not enough to probe it once, our research aims to provide the first step of a long-term view of LLM adoption and adaptation as we interact with this highly disruptive technology.

2.2 The Role of (Technological) Self-Efficacy

Self-efficacy, a concept originating from Bandura's Social Cognitive Theory [4, 6], also plays a significant role in technology adoption, adaptation, and appropriation. In the context of technology use, self-efficacy refers to an individual's belief in their ability to use technology successfully to accomplish specific tasks. Compeau and Higgins [23, p. 192] introduced the concept of computer self-efficacy, defining it as "a judgment of one's capability to use a computer". Their research demonstrated that individuals with higher computer self-efficacy are more likely to adopt and use new technologies and are more resilient when facing difficulties in technology use. In terms of technology adoption, Venkatesh and Davis [65] incorporated self-efficacy into an extended version of TAM. They found that self-efficacy significantly influences perceived ease of use, which in turn affects intention to use technology. This suggests that individuals who feel more confident in their ability to use technology are more likely to perceive it as easy to use and, consequently, are more likely to adopt it. Regarding technology adaptation and appropriation, Jasperson et al. [32] argue that self-efficacy plays a crucial role in post-adoptive behaviors. Users with higher self-efficacy are more likely to explore and experiment with technology features, leading to more innovative and extensive use. This aligns with the concept of technology appropriation, where users adapt technology to fit their needs and contexts. Bandura [5] posited that self-efficacy is not static but can change over time through experience, vicarious learning, social persuasion, and emotional states. In the context of technology use, this suggests that as users gain experience with technology, their self-efficacy may increase, potentially leading to more advanced or creative uses of the technology.

Building on this foundation, our research investigates how interactions with LLMs may influence users' self-efficacy in technology use.

2.3 The Current Adoption of LLMs in Knowledge Work & Education

The adoption of LLMs like ChatGPT in knowledge work and education represents a new frontier in technology acceptance and appropriation. These AI-powered tools are rapidly changing how people approach tasks in professional and educational settings, presenting unique challenges and opportunities for understanding technology adoption. As the participants in our study are knowledge workers and students alike, we want to highlight the recent adoption of LLMs in these two domains.

In education, the adoption of LLMs presents both opportunities and challenges [33, 42]. Mollick and Mollick [41] explored how educators are dealing with AI writing tools used by students, highlighting the need for new pedagogical approaches and assessment methods. For example, while LLMs may lower the cognitive load required for tasks such as collecting information, this can affect learners' active engagement with the material and, thus, learning [52]. This underscores how the adoption of LLMs may require not just individual adaptation but systemic changes in educational practices. One approach examined in recent work is using LLMs for giving students formative feedback, supporting self-regulated learning at a large scale [53].

In the realm of knowledge work, LLMs have been picked up early in different domains such as software engineering [20], law [54], consumer science [47], and natural sciences [40, 61]. Recent Human-Computer Interaction (HCI) research focuses on professionals' perception of generative artificial intelligence (GenAI) in different domains such as User Experience and Interaction Design [35, 36, 60], healthcare [56], programming [9, 29, 66], and data science [28, 68], providing us with a snapshot of AI adoption (for an overview of AI perception across seven industries, see also [67]).

Sun et al. [58] showed that users in creative fields develop personalized prompting strategies based on mental models of how GenAI works and continuously approach each new context with different learning and adaptation strategies. In a three-week study, Long et al. [39] analyzed how users adopt, adapt, and appropriate an AI-based workflow. Interestingly, their research also showed that users' perceived usefulness (one of the key components of the TAM) increases after familiarization with the tool.

The concept of self-efficacy appears particularly relevant in the context of LLM adoption. Liang et al. [37] found that individuals with higher levels of general technology self-efficacy were more likely to experiment with and adopt LLMs in their educational tasks. However, they also identified a need for a new construct of "AI interaction self-efficacy", reflecting the unique skills required to effectively prompt and interpret LLM outputs.

2.4 Our Previous Findings on LLMs' Impact on Perceived Productivity and Accomplishment

Our 2023 study investigated how ChatGPT affects young professionals' perceptions of productivity and sense of accomplishment through a two-week diary study with 21 participants [34]. Our results revealed a complex relationship between AI tool usage and user experience. Many participants reported enhanced perceptions of perceived productivity and accomplishment, primarily driven by greater creative output and satisfaction from efficient tool utilization. However, others experienced decreased perceived productivity and accomplishment due to a diminished sense of ownership, a perceived lack of challenge in the work process, and acceptance of mediocre results despite time constraints. The study identified that the suitability of LLMs varied significantly by task type. LLMs proved especially valuable for comprehending broad subject domains, generating creative solutions, and uncovering new information. Conversely, participants found it less suitable for research tasks due to hallucination issues requiring extensive validation. Participants developed strategies to preserve their sense of accomplishment, particularly by maintaining ownership through post-processing of AI-generated content and developing effective prompting techniques that enhanced their sense of competence.

With our follow-up research, we contribute to the body of HCI research by providing the first step of a long-term view of LLM adoption and adaptation. We are particularly interested in users' *perceived* productivity, accomplishment and self-efficacy as important factors of technology adoption, and thus focus on the following research questions:

- **RQ1** Which factors influence users' perceived productivity, accomplishment, and self-efficacy after sustained use of LLMs?
- **RQ2** How have users' self-reported productivity and accomplishment changed compared to August 2023?
- **RQ3** How have users' interaction strategies with LLMs evolved through sustained use?

3 Methodology

Our research is a follow-up study to our 2023 study [34], examining how the perception of LLMs among young professionals has evolved and consolidated over the past year. As in 2023, our study combines a diary study to track LLM usage with short surveys and an exit interview for more detailed qualitative insights.

3.1 Participants

We invited the 21 participants who had participated in our earlier study [34], and 15 agreed to join. All participants were enrolled in a master's level educational program focusing on entrepreneurship and innovation. Participants' ages ranged from 21 to 28 years (M = 24, SD = 1.73), with 47% identifying as female and 53% as male. The sample included diverse nationalities and academic fields, with all participants having at least ten weeks of full-time equivalent work experience in technology or business, from which 80% had more than one year of full-time work experience. At the time of our study, most participants were either full-time students (47%) or students with a part-time job (33%). We chose not to apply a threshold for LLM use throughout the year to capture a full spectrum of user experiences and adoption patterns. This allowed us to observe not only trends among active users, but also reasons for decreased use or disengagement, which are equally valuable for understanding the broader societal integration of LLM technology. Our participants reported a high level of expertise in interacting with LLMs ($M = 4.13, SD = 1.06$), high familiarity with LLMs ($M = 4.87, SD = 0.99$), and using LLMs often in their day-to-day tasks ($M = 4.53, SD = 0.99$). All three measures were part of the demographics survey and were measured on 6-point scales (1 being *novice/not familiar at all/never* and 6 being *expert/very familiar/daily*). Table 1 shows a detailed overview of the participants' demographics. Following Broberg [16] and Paavola and Hakkarainen [46], we consider the students in our sample a subset of knowledge workers since they actively engage in knowledge creation through information processing and cognitive tools, particularly in innovation-focused educational environments like their master's program where participants combine academic work with professional experience.

3.2 Procedure

Participants were onboarded to our study via email. We informed the participants about the study's purpose and data usage and asked for their consent. Our study consisted of a two-step approach. First, we conducted a 2-week feedback diary study [19] and second, we

Table 1: The participants' demographics for the diary study (N = 15).

Gender	
Female	7 (47%)
Male	8 (53%)
Age	
∅	24
SD	1.73
Nationality	
Germany	9
Austria	1
Egypt	1
Italy	1
Poland	1
Taiwan	1
Uzbekistan	1
Highest Completed Education	
High School	3 (20%)
Bachelor Degree	8 (53%)
Master Degree	4 (27%)
Primary Field of Study	
Business and Economics	7
Engineering and Technology	1
Business Information Systems	1
Mathematics and Computer Science	2
Natural Sciences	3
Arts and Design	1
Work Experience	
More than 24 months (2+ years)	4 (27%)
13-24 months	8 (53%)
7-12 months	2 (13%)
1-6 months	1 (7%)
Occupation (at the time of the study)	
Student	7 (47%)
Student and working part-time	5 (33%)
Full-time employee	2 (13%)
Unemployed	1 (7%)

held semi-structured interviews (exit interviews) with each participant after the diary study was completed. The diary study surveys were sent via email every business day at 6:00 p.m. (Monday to Friday). We opted for daily surveys to preserve contextual richness and capture real-time interactions [11]. Participants were not given specific instructions regarding how or when to use LLMs during the diary period as our aim was to capture participants' real-life, self-directed interactions with LLMs in their natural work and daily contexts. To be eligible for the study, participants were required to complete at least seven out of ten weekday logs, with the option to report both LLM and non-LLM technology use. All participants fulfilled this condition. We introduced participants to the study on a rolling basis. All participants concluded the study between July and August 2024 and received EUR 20 upon study completion. The study was approved by the ethics committee within our University Faculty.

3.3 Measures

The diary study surveys included both closed- and open-ended questions. Closed-ended questions consisted of pre-defined answer choices and 6-point Likert scales assessing three dimensions: perceived productivity, perceived accomplishment, and perceived self-efficacy. These dimensions were introduced to participants with the following definitions: *productivity* as "doing more in less time", *accomplishment* as "the feeling of satisfaction or fulfillment that comes from successfully completing a task or achieving a goal", and *self-efficacy* as "a person's belief in their capacity to do tasks and succeed." Participants rated their immediate perceptions using 6-point scales: productivity (from *not productive at all* to *very productive*), accomplishment (from *not accomplished at all* to *very accomplished*), and self-efficacy (from *not at all effective/skilled* to *very effective/skilled*). To assess perceived improvements attributed to LLM use, additional 6-point scales (from *strongly disagree* to *strongly agree*) were used for each dimension.

We used open-ended questions for qualitative insights into the dimensions' ratings. Table 2 offers an excerpt of our diary study questions and structure for the three dimensions.

Finally, we used a semi-structured guide for the exit interviews to probe deeper into identified themes and patterns. Interviews were recorded and transcribed using Notta[1] to facilitate accurate data capture for subsequent analysis. The diary study and exit interview questionnaires are included in the supplementary material.

3.4 Data Analysis

3.4.1 Diary Study. The analysis of the diary study consisted of three main steps: (1) descriptive and inferential analyses of the 2024 quantitative data, (2) thematic analysis of the 2024 data, and (3) significance tests comparing the quantitative 2023 data [34] and the new 2024 data.

For (1), we provide an overview of the completed diary surveys per participant and visualize the distribution of self-assessment data collected through the diary study. We also calculated Spearman rank correlations between the dimensions of perceived productivity, accomplishment, and self-efficacy, as well as the perceived impact of LLM usage on each of these dimensions.

For (2), we analyzed the open-ended responses using thematic analysis for the dimensions of perceived productivity, perceived accomplishment, and perceived self-efficacy. We imported all open-ended responses into the qualitative research software Condens[2]. For each dimension, we followed Braun and Clarke's guidelines [15]. One researcher generated the initial codes, capturing meaningful aspects of participants' explanations for their ratings and subsequently analyzed and grouped the codes into overarching themes based on patterns and relationships between the codes. To ensure the reliability of the analysis, a second researcher independently reviewed the codes and resulting themes to validate the interpretation and thematic structure. Any disagreements were discussed until a consensus was reached. We excluded codes with too few occurrences, and our analysis ultimately resulted in four themes across the dimensions of perceived productivity, perceived accomplishment and perceived self-efficacy.

[1]https://www.notta.ai/en (last accessed: 01/28/2025)
[2]https://app.condens.io (last accessed: 01/18/2025)

Table 2: The dairy study questions for the three dimensions of perceived productivity, perceived accomplishment, and perceived self-efficacy.

Theme	Rating (closed-ended question)	Influence of LLMs (closed-ended question)	Reasoning (open-ended question)
Perceived Productivity	How productive do you feel by your work done with the LLM today?	Did the quantity of your work improve by using the LLM?	Why do you feel so?
Perceived Accomplishment	How accomplished do you feel by your work done with the LLM today?	Did your perception of accomplishment improve by using an LLM?	Why do you feel so?
Perceived Self-efficacy	How skilled/effective did you feel while completing your tasks?	Did your perception of self-efficacy improve by using an LLM?	Why do you feel so?

For (3), we used Shapiro-Wilk tests to evaluate the normality of our data and subsequent paired-sample t-tests for the differences between the perceived productivity and perceived accomplishment ratings, applying a significance threshold of $\alpha < 0.05$. We used Cohen's d to interpret the effect sizes [21].

3.4.2 Exit Interviews. Similar to the open-ended survey questions, we analyzed the exit interviews following the diary study using Condens. On average interviews lasted 21 minutes ($SD = 4:38$). We employed a collaborative, open coding approach to analyze the semi-structured interview transcripts. Two researchers independently conducted initial open coding on half of the transcripts. A separate codebook was developed for the analysis of the interviews, as it was expected that the diary codebook is unsuitable for this task. In the diary study, the participants provided immediate day-to-day reflections, in comparison to the exit interviews, where the participants reflected on their overall experience with LLMs in the past study period. Following an initial analysis, the researchers discussed preliminary findings and refined the coding scheme. Each researcher then independently coded the remaining half of the transcripts using the refined coding framework. Through an iterative process of discussion and comparison, we refined the code collection by merging similar codes and categorizing the rest of the codes [10]. This process resulted in 38 codes, grouped into 5 categories pertaining to our research questions.

4 Results

We gathered data from two instances: the two-week diary study with ten consecutive short surveys and the exit interviews. The diary study entries (114 reports by 15 participants) captured participants' perceived productivity and perceived accomplishment ratings for individual days, enabling comparisons with our 2023 data [34] as well as a more nuanced picture for these dimensions. The participants filled out the diary study surveys over ten business days. We probed deeper and collected context for the experiences of the diary study in our exit interviews ($N = 15$) after the diary study period.

4.1 Findings from the Diary Study

In total, 15 participants from the 2023 study concluded our 2024 follow-up study. On average, participants filled out 7.6 ($MIN = 7$, $MAX = 9$) of the diary study surveys and reported using LLMs on average on 4.4 ($MIN = 0$, $MAX = 6$) of the ten business days. Figure 1 shows the distribution of diary study entries per participant. 66 of those 114 entries (58%) are entries where participants used LLMs, whereas 48 (42%) are attributed to days where participants did not use any LLMs. While most participants reported using LLMs regularly during the two-week diary period, one participant (P4) did not log any LLM use. In the exit interview, P4 explained that their time during the study was largely occupied with organizing a large event, which did not require LLM support. However, they emphasized that under normal circumstances, they use LLMs frequently in their day-to-day work.

For most of the entries, participants used ChatGPT-4o (79% of all diary study entries), followed by ChatGPT-4 (17%), Claude 3.5 Sonnet (3%), and Perplexity (2%). In our study, participants mostly used LLMs for improving text (29%), understanding complex or new topics (20%), programming (15%), research (14%), and writing text (13%).

Similar to our 2023 findings, we looked at perceived productivity and perceived accomplishment (see table 2). As perceived accomplishment is a feeling that is measured *after* a lot of work or effort is achieved, we were also interested in whether the use of LLMs impacts participants' feelings of how well they can achieve something *prior to and during* a task. We thus included a question on the users' self-efficacy with *"How skilled/effective did you feel while completing your tasks?"*. All three items were rated on a 6-point scale (1-6), with 1 being *"not at all productive/accomplished/skilled"* and 6 being *"completely productive/accomplished/skilled"*. Figure 2 shows the scale chart for all three items. In summary, participants reported a high sense of perceived productivity ($M = 4.82$ and $SD = 0.95$), perceived accomplishment ($M = 4.55$ and $SD = 0.85$), and self-efficacy ($M = 4.53$ and $SD = 0.87$). Appendix A shows the per-participant scale responses (M and SD) for each of the dimensions.

Our Spearman rank correlation analysis revealed strong positive relationships between perceived self-efficacy and perceived accomplishment ($r_s(66) = 0.70$, $p < .001$), and moderate positive

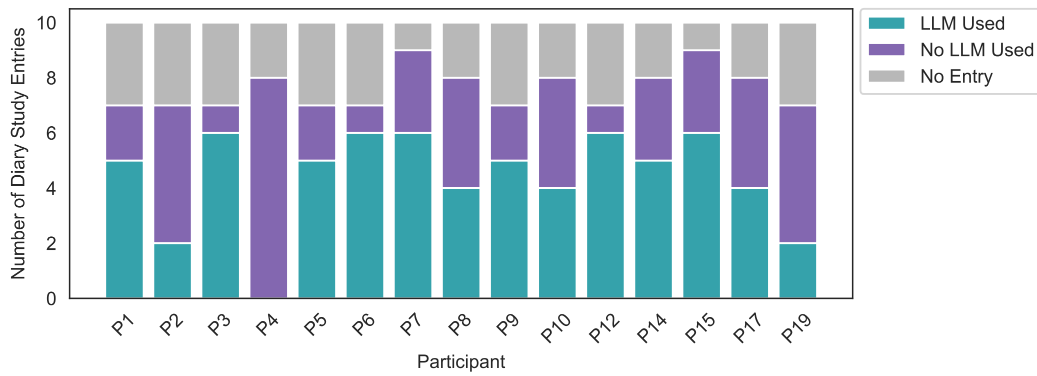

Figure 1: Distribution of diary study entries across participants, showing the number of days when LLMs were used, not used, or when no entry was recorded. Each participant had a maximum of 10 possible diary entries.

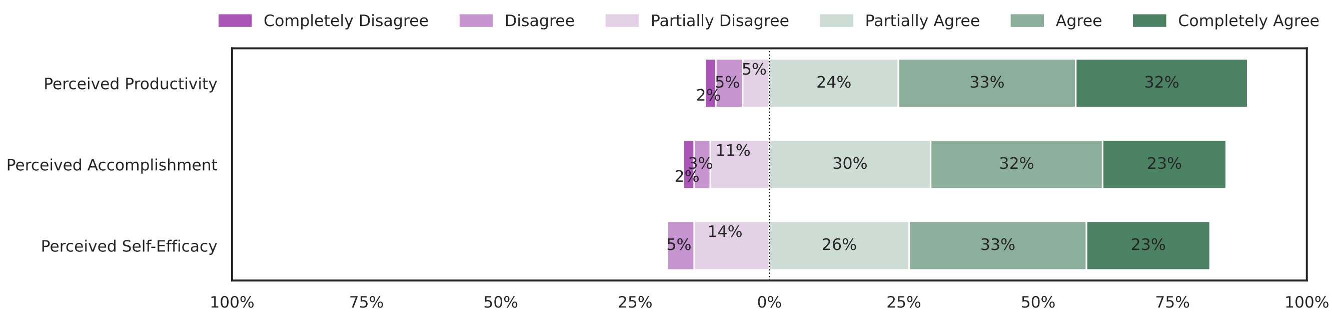

Figure 2: Overview of scale ratings for the survey items of *Perceived Productivity, Perceived Accomplishment,* and *Perceived Self-efficacy.*

correlations between perceived productivity and both perceived accomplishment ($r_s(66) = 0.52$, $p < .001$) and perceived self-efficacy ($r_s(66) = 0.56$, $p < .001$).

We further evaluated the impact of LLM usage on the dimensions of perceived productivity, perceived accomplishment, and perceived self-efficacy. The Spearman rank correlations showed significant strong correlations between LLM usage and perceived productivity ($r_s(66) = 0.68$, $p < .001$), perceived accomplishment ($r_s(66) = 0.74$, $p < .001$), and perceived self-efficacy ($r_s(66) = 0.70$, $p < .001$) Figure 3 shows the Spearman rank correlations for the three dimensions.

Additionally, participants provided reasoning in the form of open text fields during the diary study for the three dimensions of perceived productivity ($N = 66$), perceived accomplishment ($N = 65$), and self-efficacy ($N = 63$). The questions we posed were *"Did the quantity of your work improve by using the LLM?"*, *"Did your perception of accomplishment improve by using an LLM?"*, and *"Did your perception of self-efficacy improve by using an LLM?"*. In very few instances, participants did not share any reasoning. In the following, we present our thematic analysis results for the open text fields of these three dimensions. We add exemplary participant statements to illustrate the themes. Numbers in brackets indicate the participants' ratings from 1 to 6 for each quote, and letters in subscript indicate the dimension for which the statement was made (i.e., *P* for perceived productivity, *A* for perceived accomplishment, and *S* for self-efficacy). Figure 4 provides an overview of the four main themes and which dimensions they relate to.

Contextual Usefulness. Participants consistently noted that the usefulness of LLMs depended heavily on the nature of the task. For perceived productivity, LLMs were described as particularly helpful for programming, writing, information synthesis, and quickly gaining a high-level understanding. One participant shared, *"Instead of having to think and search [...], you can ask ChatGPT and it will save you time. [...] With ChatGPT (and its interactiveness) you can quickly get all your questions answered."* (P6, 6_P). Another reported, *"I had a very tedious task of manually searching hundreds of company names in the database. Using LLMs saved me probably like an hour."* (P3, 5_P). However, LLMs were less useful for very specific queries or tasks requiring deep domain expertise: *"It improved a bit, but most of my work today involved manual little adjustments to the structure and content of my research – AI cannot help with that, as I am the only one that is this fully immersed in all the data."* (P14, 4_P)

Although reported by less participants, the effect of LLMs on perceived accomplishment (P1, P5, P3, P7) and self-efficacy (P3, P5, P6, P8, P19) also varied based on the nature and complexity of

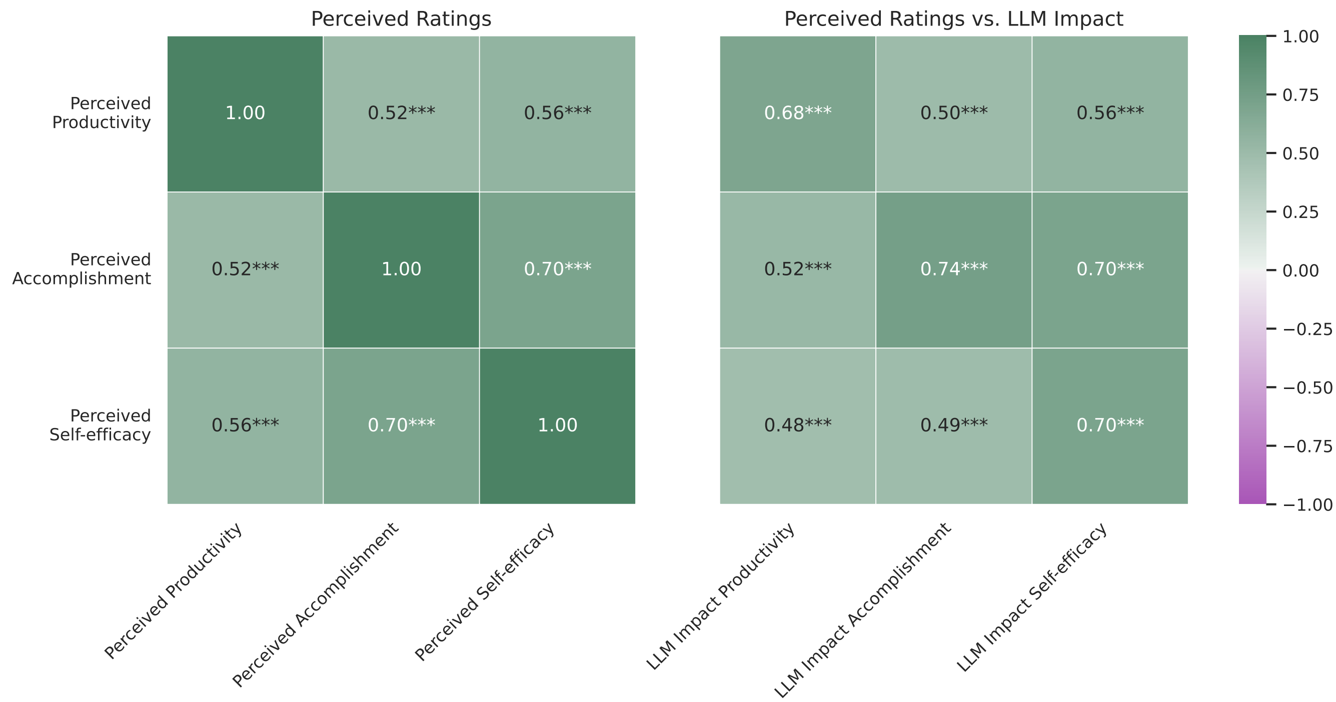

Figure 3: Spearman rank correlation tables showing the correlation coefficient ($N = 66$) for perceived ratings (left) and perceived ratings vs. LLM impact (right). The color of the squares indicates the height and direction of the correlations. Significance levels are represented as follows: *p <. 05, **p < .01, ***p <. 001.

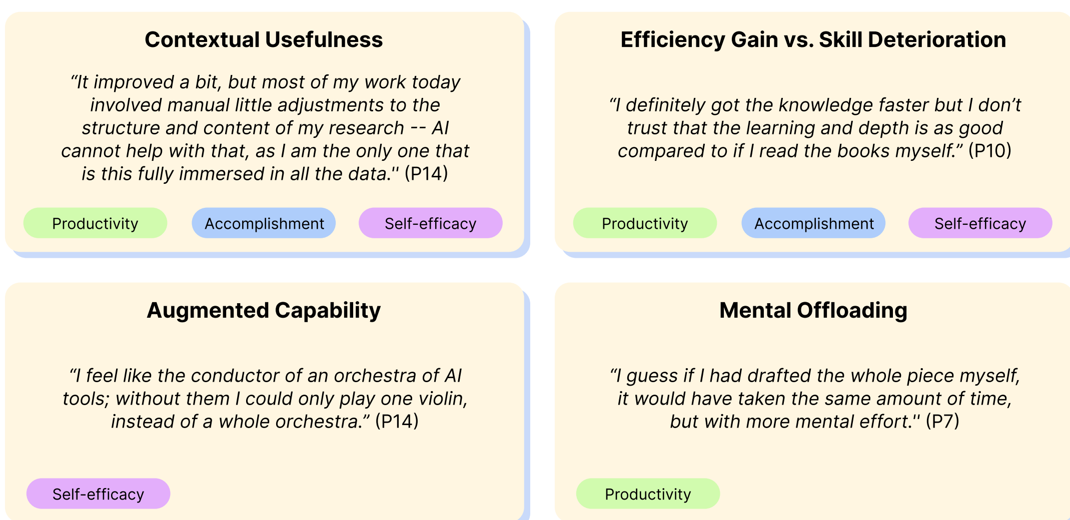

Figure 4: The four themes elicited from the open-ended diary study questions.

the task. Similar to the dimension of perceived productivity, LLMs proved to be beneficial for tasks involving information synthesis, programming, writing, and learning but were less effective or even disruptive in tasks requiring personal expertise: *"It actually made some parts of my work better and others worse so I had to put in additional effort. So my sense of accomplishment went down."* (P1, 2_A) and similarly, *"The tasks were very high-level and ChatGPT wouldn't be able to solve them."* (P3, 3_S)

In some cases, using LLMs actually increased time investment, with P6 (2_S) observing they *"lost time"* due to poor prompting, while P8 (4_S) found that for certain tasks, *"a Google search would have led to the same results today."* This suggests that the utility of LLMs is highly contextual and not universally beneficial across all task types.

Efficiency Gain vs. Skill Deterioration. A prevalent theme across all dimensions was the trade-off between efficiency and preserving the depth of understanding or long-term skill development. In terms of perceived productivity, many participants (P5, P7, P10, P14, P17) described LLMs as tools for improving efficiency, even if this did not always translate into more output: *"I didn't do more than usual but I did the task faster or at least I felt that way."* (P5, 4_P). Higher productivity ratings were often associated with both time savings and tangible output gains (P3, P5, P6, P8, P10, P15, P17), especially for structured or repetitive work: *"I was able to get more features done in less time. I also believe the quality improved, especially related to stuff like automatic documentation, test creation, etc."* (P15, 6_P). Several participants also pointed to goal attainment as a key driver of perceived accomplishment: the ability to *"get more work done"* (P1, 6_A) or achieve goals *"better and faster"* (P14, 5_A) was consistently tied to a sense of making progress and succeeding in one's tasks (P1, P3, P7, P8, P10, P14, P15, P17).

This boost in efficiency also contributed to perceived self-efficacy, especially when LLMs helped participants overcome blockers or finish work they might otherwise have struggled with (P5, P6, P9, P10, P18). As one participant explained, *"LLMs help to get work done faster. Therefore, you get further in the same amount of time."* (P9, 4_S). Others credited the tools with enabling fast comprehension of complex material: *"That otherwise would have taken me way longer to grasp."* (P14, 4_A), or with supporting more advanced problem-solving: *"ChatGPT was a tremendous help in analyzing data in Excel, as you can get help with the formulas by uploading screenshots."* (P9, 4_A).

However, these gains were often accompanied by concerns around reduced learning and skill deterioration. Several participants worried that relying on LLMs would erode their abilities over time or hinder deeper engagement with content (P6, P10, P12): *"I definitely got the knowledge faster but I don't trust that the learning and depth is as good compared to if I read the books myself."* (P10, 3_A), and *"I guess it is nice that for some niche tasks you can still rely on yourself. I feel like this is also crucial to staying smart, taking time to think about things."* (P6, 4_A). These reflections extended to self-efficacy, where some feared that convenience was coming at the cost of personal competence. As P12 put it, *"losing the skill to do it myself"* (P12, 3_S), while P10 described their experience as *"taking the easy route"* (P10, 4_S). Even participants with high self-efficacy scores expressed ambivalence, questioning their reliance on AI tools: *"Many people before ChatGPT also needed to pass this course and they understood it without these tools. So I guess it makes you think. . . "* (P6, 6_S).

Augmented Capability. Several participants (P1, P3, P5, P6, P7, P12, P14, P15) described how LLMs enhanced their sense of self-efficacy by enabling them to complete tasks they found difficult or time-consuming. One participant metaphorically described feeling *"like the conductor of an orchestra of AI tools; without them I could only play one violin, instead of a whole orchestra"* (P14, 4_S), highlighting the transformative impact on their capabilities. The tools helped participants overcome specific challenges, with P15 (4_S) noting how it helped them *"mitigate my weakness with writing."* In professional situations, LLMs provided valuable support, as exemplified by P7 (5_S) who described a boost in presentation confidence *"the content and bullet points ChatGPT provided, helped me feel more confident that I could run the session smoothly."*

Mental Offloading. In terms of perceived productivity, participants noted reduced mental effort. This theme was particularly pronounced in the higher ratings for P1, P8, and P10. P10 (6_P) mentioned to be *"less frustrated trying to figure it out than by usual Google search"* and P1 (6_P) highlighted, *"While the LLM was correcting my errors, I could focus on other tasks and could take breaks."* However, prompting inefficiencies seemed to overcast the reduced cognitive load for P7 (2_P): *"It took a while to prompt ChatGPT to do exactly what I needed. I guess if I had drafted the whole piece myself, it would have taken the same amount of time, but with more mental effort"*. This theme was not prevalent for perceived accomplishment or self-efficacy.

4.2 Comparison to 2023 Findings

As we were interested in how the participants' perceived productivity and personal accomplishment have changed over the course of a year, we compare the results of the 2024 data to our 2023 results [34]. We omit P4 and P5 for this analysis as we specifically want to look at the within-subject changes, and both participants reported not having used LLMs during the diary study in 2024 and 2023, respectively. Figure 5 shows the boxplots of selected response-scale options for perceived productivity and perceived accomplishment per year. Whereas the mean for perceived productivity only slightly increased from $M_{2023} = 4.53$ ($SD_{2023} = 1.26$) to $M_{2024} = 4.84$ ($SD_{2024} = 0.96$), we see a more pronounced effect for the perceived accomplishment ratings: $M_{2023} = 3.97$ ($SD_{2023} = 1.04$) compared to $M_{2024} = 4.53$ ($SD_{2024} = 0.85$).

Initial Shapiro–Wilk tests for normality confirmed that the data for the differences per participant for both dimensions is normally distributed. A subsequent paired samples t-test revealed that there is no significant difference between the within-subject perceived productivity ratings of 2023 and 2024 ($p = .266$). However, the data shows a significant difference for the within-subject perceived accomplishment ratings of 2023 and 2024 ($t(12) = 2.619$, $p < .05$). For measuring effect size, we used Cohen's d interpretation, where .2 indicates a small effect, .5 indicates a medium effect, and .8 indicates a large effect [22]. Our results showed a medium effect size (Cohen's $d = .726$). We report a summary of our statistical analyses in Table 3.

4.3 Exit Interviews

In the following, we present the findings from the exit interviews, where participants reflected on their overall experience of using LLMs. We first contextualize our findings on the participants' increased sense of accomplishment (see section 4.2) before highlighting prominent themes that gave us additional insights into the diary study. Additional insights were collected on the theme *Interactions*, which refers to the changes in interactions with LLMs since the participants have adopted them. The themes *LLM Strengths* and *LLM Limitations* refer to the use cases where the participants reported satisfactory/unsatisfactory collaboration with the LLM. *Concerns* refer to the fears and use cases where the participants are

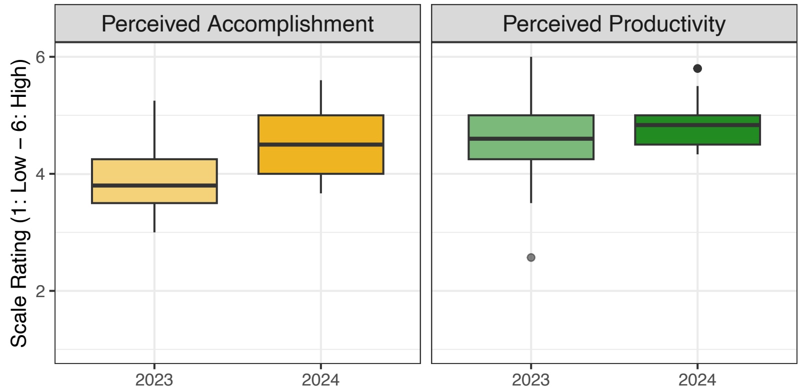

Figure 5: Boxplots for the dimensions of perceived accomplishment (left) and perceived productivity (right) for the 2024 and 2023 study, averaged by participant ($N = 13$).

Table 3: Paired samples t-test results comparing participants' ratings of perceived productivity and perceived accomplishment for 2023 and 2024. We assessed normality using the Shapiro-Wilk test.

Item	**Paired Samples t-test**			**Effect Size**	**Normality**	
	df	t	p	Cohen's d	W	p
Perceived Productivity	12	1.166	.266	–	.932	.364
Perceived Accomplishment	12	2.619	.022*	.726	.943	.503

hesitant to use LLMs, and *Future vision of LLMs* refers to the users' expectations for modeling future interactions with LLMs.

Increased Accomplishment. In the exit interviews, participants reported two key factors for their increased sense of accomplishment over time: improved LLM output quality and enhanced user proficiency, both ultimately increasing users' output. All participants, except for P7 and P19, emphasized how their growing expertise in LLM interaction positively impacted their sense of accomplishment. P14 noted, *"I know better how to deal with it and how to get what I want"*, while P10 mentioned feeling *"more confident in my abilities and the output of the LLM"*. P2 specifically highlighted the satisfaction derived from precise prompt engineering: *"you're also more accomplished because [...] you managed to write correctly what you want."* Notably, the sense of accomplishment of P6 matured from initial hesitation to viewing LLMs as capability enhancers: *"at the beginning you had the stigma of 'Okay, this thing is doing my work, and I don't have that much of a part in it'. Whereas now, I think, the mindset is that this is a tool that I can use."*

Interactions. The change in interactions with LLMs was one prominent theme that was mentioned in most interviews. Here we identified one larger group of participants (P3, P5, P7, P8, P10, P15, P19), which reported that they narrowed down the use cases in which they interact with LLMs, and another smaller group (P1, P4, P9, P17) which reported that they expanded the variety of LLM use cases. The participants who narrowed down the use cases often reported that they had become proficient in using LLMs for specific tasks and had obtained a sense of how to generate satisfactory output. For example, P15 said, *"I very much know which tasks work well and which don't, and I already know when I encounter one that works very well, then I know how to write the prompt so that it kind of works."* Several participants expanded on this claim, stating that they now have *"developed a better feeling for when it's helpful [...] and when it's not"* (P4) but also that LLMs have improved at generating relevant output. Additionally, participants tend to rely on LLMs in cases where they know in advance that little content post-processing will be necessary. The group that expanded the use cases reported higher ambition in learning how to write better prompts: *"The more often you do it, you kind of see in which situations ChatGPT needs more detailed input to actually give you great results and where very short, straightforward prompts are enough."* (P9) Once they have mastered prompting for certain tasks, they would move on to expand the variety of tasks. P1 reported, *"I think in the beginning I was only doing like grammar and spelling things, but now I'm also doing more stuff like, hey, brainstorm this and that, like, more creative tasks or, like, giving a text and asking for a complete list about covering the whole topic."*

LLM Strengths and Limitations. Extending on the change in interactions, we also noted trends in which use cases the participants found useful and which ones they discarded due to unsatisfactory output quality. For example, programming tasks were one use case where most participants reported reliable and frequent use, exhibiting similar feelings as P3: *"LLMs are better at programming [than researching]. I used it a lot already for programming in the last year, and so I kind of figured out what I can ask, and it will return* *some code that I basically have certainty about its correctness."* Data analysis and learning were also among the use cases where the participants found LLMs to be particularly useful. Interestingly, there also seems to be a trend in outsourcing tasks to LLMs that the participants did not want to do, such as finding a recipe to cook something with the ingredients they already had available or formatting text, noting it down as a *"stupid"* (P5) task they don't want to spend time on. Participants also outsourced tasks to the LLM that they did not enjoy, did not know how to do, or did not see value in learning – effectively using the LLM as an assistant. In contrast, they tended to avoid scenarios in which the LLM repeatedly produced unsatisfactory output. For example, creative tasks or brainstorming were tasks that the participants would rather not delegate to the LLM. Some of them reported that the lack of original output discouraged them from further using the LLMs for such tasks, and they would rather do the task by themselves. Use cases where extra context needed to be given to the LLM to complete a task were also avoided. For instance, P1 stated, *"I think what stops me from feeling full productiveness is that currently, LLMs are not able to capture full workflows. For example, I want to do a systematic literature review, and I want to search all the studies, and I want to come up with the research question and the methodology. I think this is something that hinders me to come to 100% productivity."*

Concerns. Here, some participants also reported fears and general privacy concerns related to prolonged usage of LLMs. Some of them (P5, P7) fear that outsourcing tasks to LLMs might prevent our development in writing skills and creative thinking *"like it will do all the thinking for us in a way that we as humans can easily become a lot dumber than we have been before"* (P5). Similarly, others reported privacy concerns related to either revealing details about themselves to the LLM or analyzing documents that can reveal the personal details of others. Interestingly, some of them speculate that what they share with the LLM could cause "data privacy problems", but they do not know the "rules" for sure (P7). Lastly, participants seem torn between wishing that the LLM had more data about them (so they do not need to provide context each time) and being concerned that behind the LLM stands a private company with access to the data they shared (P2, P5).

Future vision. To take a step forward, we also asked how participants envision the development of interactions with LLMs. Here, the vast majority of participants (P1, P2, P4, P5, P6, P7, P8, P10, P19) expressed a vision for more implicit interactions where the LLM would understand their wishes without too much context and would rather be in the role of an autonomous assistant that lessens the workload on the human. For example, P6 expressed *"I think it will do everything for you. It will schedule your calendar invites with your friends or with your colleagues. It will look up all the data for you that you need. I think it will be completely integrated in our everyday lives."* On the other hand, P5's statement *"So my sense of accomplishment would probably be 100% when the LLM knows me from one word or two words, and it's my soulmate in a way. So it knows what I want, it knows what I think"* expresses a vision where the LLM is tailored to how the specific user thinks to serve them in the best way.

5 Discussion

Our study revealed how factors such as efficiency, adaptation, and personal interaction patterns influence users' perceived productivity, accomplishment, and self-efficacy (RQ1). A comparison with our earlier study [34] also showed a significant increase in perceived accomplishment within one year while perceived productivity remained similarly high (RQ2). Users in our study showed different ways of mastering LLMs, with one group expanding their use cases while the other narrowed them down to ensure efficiency gains (RQ3). These findings provide a snapshot of AI adoption's long-term (psychological) impact in knowledge work, moving beyond initial productivity metrics to understand how users' perceptions, capabilities, and interaction methods evolve over time. In the following sections, we discuss our findings in light of technology adoption and adaptation, why this is important, and how longitudinal observations like our study can advance the HCI field.

5.1 Technology Adoption and Adaptation

Our results confirm and extend previous work on technology adoption and AI integration in knowledge work. Building on our 2023 study [34], we find that initial feelings of inferiority can transform into confidence and strategic AI use over time. This echoes established technology adoption models such as TAM2 [63, 64], which emphasize the role of experience in shaping perceived usefulness and ease of use. TAM2 also highlights the influence of social factors, drawing from the Theory of Reasoned Action (TRA) to show how expectations from others can shape behavioral intentions [64]. In line with this, our participants reported that uncertainty around using LLMs in 2023 often stemmed from not knowing whether use was permitted or professionally appropriate.

By 2024, however, the narrative had shifted. Some participants described a growing sense of pressure to adopt LLMs, saying they felt they were "falling behind" without them. LLMs had become an everyday tool used to enhance productivity and cognitive reach. This transformation illustrates both the social normalization of the technology and a shift in perceived utility driven by firsthand experience.

Yet LLM adoption does not merely follow familiar trajectories. Compared to earlier disruptive technologies such as smartphones, personal computers, or the internet, LLMs display uniquely rapid and widespread uptake. Already in 2023, 42% of a representative US sample reported having used an LLM [27]; by 2025, that number had risen to half of all US adults [48]. Adoption is especially high among younger users, highly educated individuals, and knowledge workers, with usage exceeding 80% among researchers [38], and substantial uptake across higher education [55].

Several features help explain this accelerated adoption curve. First, LLMs support hybrid, boundary-blurring usage patterns: the same tool is used for learning, ideation, communication, and formal work tasks [13, 14, 38, 49, 55]. Second, their impact on productivity is both immediate and well-documented [30, 44, 45, 66], offering clear incentives to integrate them into workflows [34]. And crucially, unlike earlier technologies that required specialized skills or onboarding, LLMs are accessible through natural language – a modality familiar to virtually everyone. This dramatically lowers the barrier to entry and broadens the adoption base.

Finally, the unusually fast pace of LLM development introduces a dynamic element to the adoption process. Capabilities evolve rapidly, continuously unlocking new use cases. Unlike past technologies that demanded users adapt to new technical paradigms, LLMs adapt to users, aligning with innate human communication patterns. This "conversational interface" not only accelerates adoption but also sustains it, as users can integrate new functionality without relearning how to interact with the system.

Our qualitative findings from the exit interviews also show how users *adapt* their LLM usage by either expanding or narrowing down their use cases. This is similar to patterns identified by Long et al. [39], who examined how users interact with and customize an AI workflow for writing science communication tweets over time. Their study shows that after a short familiarization phase, users' perceived system usefulness even increased with time, indicating that the positive impression was not just a novelty effect. They attribute this to the workflow supporting cognitively challenging tasks and enabling better human-AI collaboration through customization. Our findings suggest a similar understanding of the familiarization phase in human-LLM collaboration. As users adapt to LLMs as a new technology, e.g., by experimenting with their use cases (for further exploration of knowledge work use cases, see for example [13]), their prompting strategies and how they integrate these into their workflows [67], they also see their personal accomplishment increase. It is important to mention that over time, the participants gained a "gut feeling" for which use cases they are to expect a satisfactory output and for which cases not. For example, when participants felt the LLM lacked the context to perform well, they would rather not delegate that task to the LLM. Their positive attitude towards perceived productivity and accomplishment was often powered by their routine use cases, where they have mastered the prompting and were guaranteed a satisfactory result. While they found certain use cases to work better than others, throughout this process, the participants also became aware of the limitations of this technology, which helped them to better understand their role in these interactions. While last year, participants were asking themselves *"If the machine is as good as me, then what use am I?"* [34], this year's findings suggest a more positive notion with participants emphasizing that exploring the capabilities and the boundaries of LLMs enables them to see their own value in task achievement and envisioning the LLM as a supportive assistant.

5.2 Efficiency as a Mediating Variable for Perceived Productivity, Accomplishment, and Self-Efficacy

Our qualitative analysis suggests that perceived efficiency is closely associated with users' self-reported productivity, sense of accomplishment, and self-efficacy. While efficiency gains are an expected driver for productivity improvements, their strong influence on perceived accomplishment and self-efficacy is notable. Participants frequently described perceived productivity and accomplishment as interrelated aspects of their work experiences – an observation that is further supported by patterns in their quantitative survey responses. These correlations appear to be reinforced by users' evolving interaction strategies with LLMs. Many participants reported refining their use cases and developing a more nuanced understanding of when and how to effectively apply LLMs which may help explain their higher self-reported sense of accomplishment. However, this efficiency-driven approach presents a paradox: while users report higher productivity and accomplishment, they express concerns that increased efficiency through LLM use might compromise their skill retention, potentially affecting their long-term self-efficacy. These findings also intersect with emerging perspectives in technology design that challenge traditional productivity-focused frameworks. For instance, Somanath et al. [51] advocate for rethinking workplace technologies not just as tools for enhancing output, but as systems that support employee well-being and happiness. While our study did not directly measure happiness, participants' reflections on accomplishment and the potential loss of skill suggest that users are already implicitly negotiating what "meaningful" outcomes look like. In this sense, our findings offer a complementary view of how knowledge workers evaluate the role of AI in their work – not only in terms of what gets done, but how it feels to do it.

5.3 Methodological Approach: A Longitudinal Perspective on LLM Adoption

To capture how AI adoption unfolds over time, we conducted a follow-up study that builds on a our previous investigation from 2023 [34]. By re-engaging participants from our initial study one year later, we were able to track how their perceptions, behaviors, and integration strategies with LLMs evolved. This design allowed us to move beyond one-off impressions and document a more dynamic learning process.

Long-term approaches are critical in understanding how users adapt to emerging technologies. As our findings show, the initial novelty or inferiority effects reported in 2023 gave way to more strategic, confident, and context-specific usage patterns by 2024. Participants refined their use cases, customized workflows, and developed a sense of agency in positioning themselves as decision-makers, with the LLMs acting as assistants.

Our approach aligns with prior research on long-term technology adoption. For example, van den Hooff's longitudinal study of email integration in organizational contexts [62] illustrates how digital tools are gradually woven into work practices over time. Email, initially used for simple messaging, eventually became essential for complex, collaborative professional workflows. A similar evolution was evident in our participants' LLM use: from experimentation and curiosity to task-specific, efficiency-driven applications.

However, LLM adoption presents unique characteristics that set it apart from previous technologies. While email integration occurred gradually over decades, LLM development and uptake have progressed at a much faster pace. This accelerated timeline not only amplifies the importance of longitudinal studies but also necessitates more frequent data collection to capture fast-moving changes in usage and perception.

Our follow-up study revealed notable shifts even within a single year. While most participants stuck to scenarios that had proven effective, they increasingly emphasized autonomy and control in their interactions with LLMs. Many described delegating tedious or repetitive tasks to LLMs, which allowed them to focus more on the aspects of their work they found meaningful.

Participants also expressed clear hopes for future iterations of LLMs: they wanted systems that could handle mundane tasks (e.g., scheduling) more autonomously, yet still reflect their preferences and retain a sense of user agency. Overall, our findings highlight the value of longitudinal research in understanding not just whether people adopt AI tools, but how they integrate them into their daily lives.

5.4 Limitations and Future Work

Our study involved a small sample of young professionals and may not be representative of all knowledge workers or LLM users. With an average age of 24 years, our sample is slightly younger than the average LLM user. For example, the majority (44.1%) of OpenAI's user base[3] is between 25 and 34 years old. This opens up space for further research for the adoption of LLMs by diverse generations of knowledge workers. Our sample also consisted of 47% female participants, which could further skew the results since Draxler et al. [27] showed that female users tend to use LLMs less often compared to male users, which was also the case for our sample. On average, female participants in our study reported using LLMs for 51.36% days of the two-week study, while male participants reported using LLMs for 64.36% of the days. Further, our study could be impacted by a self-selection bias, meaning that only people who are interested in joining again joined the study, which may have skewed our sample to more tech-savvy participants and people who potentially use LLMs regularly. While the knowledge workers from our sample reported similar use case scenarios, the findings might be different for a different assortment of use cases or even different groups of workers.

Our study primarily examined ChatGPT and other LLMs. The findings may not generalize to all types of AI assistants or future LLM iterations. This opens up space for closer observation of which systems the users reach out to and why. For instance, it would be interesting to see if they use or create custom instances of GPT for certain purposes. Further, the study was conducted during a period of rapid AI advancement and increasing public awareness. It is unclear to what extent the development of LLMs had an impact, since the technological improvements could also increase people's perceived performance. As GenAI tools and LLMs are a topic of interest in the general public, our results may also be influenced by broader societal trends and media narratives around AI. However, our qualitative data from the surveys and interviews also shows that participants improved their own skills in prompting, workflow integration, and evaluation of suitable use cases, depicting users as active agents in human-AI collaboration.

6 Conclusion

The rapid integration of LLMs into knowledge work influences productivity but has also raised concerns about job displacement, deskilling, and potential adverse cognitive effects. This technological acceleration demands a deeper understanding of how individuals make sense of these powerful AI tools beyond mere productivity gains.

[3]https://www.statista.com/statistics/1470825/openai-com-user-age-worldwide/ (last accessed: 02/04/2025)

In the field of HCI, there is a notable gap in our understanding of the fundamental psychological impacts of LLM adoption, particularly regarding users' self-efficacy and sense of accomplishment over time. Most existing research focuses on quantitative productivity metrics and industry-specific applications, neglecting the long-term psychological effects of AI integration in knowledge work.

To address this gap, our study investigated how users' perceptions towards their perceived productivity and accomplishment evolve as they become accustomed to LLMs, and how the integration of these tools influences users' self-efficacy. We conducted a follow-up study to our 2023 study [34] to explore first long-term effects. Our research employed a mixed-methods approach, combining a two-week dairy study with in-depth interviews with 15 participants from the 2023 study. With this, we capture both quantitative measures and qualitative insights into participants' experiences and thought processes.

Our findings reveal three key phenomena. First, perceived efficiency emerges as a crucial mediating variable between LLM use and users' sense of accomplishment and self-efficacy. Here, we observe that positive perceptions of efficiency gains align with the growing proficiency in using LLMs. Second, we observe distinct adaptation patterns in how users integrate LLMs into their workflows, with some expanding their use cases while others deliberately narrow them to optimize perceived efficiency. This divergence suggests that successful LLM integration is highly individualized and evolves with experience. Third, our comparison with 2023 data reveals a significant increase in perceived accomplishment with similarly high perceived productivity measures. This suggests that as users develop more sophisticated mental models of LLMs' capabilities and limitations, they move beyond initial feelings of inferiority toward more strategic and confident AI collaboration.

Our research makes significant contributions to the HCI field by providing a nuanced, longitudinal perspective on AI integration in knowledge work. By analyzing how users adopt and adapt to AI technologies, we challenge simplistic narratives about automation and replacement. Our findings also highlight the importance of longitudinal studies and investigations in understanding the long-term impacts of disruptive technologies on human cognition and work practices. As LLM capabilities continue to evolve, so too will user strategies, expectations, and definitions of what constitutes "productive" or "meaningful" work.

References

[1] Daron Acemoglu and Pascual Restrepo. 2018. The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment. *American Economic Review* 108, 6 (June 2018), 1488–1542. doi:10.1257/aer.20160696

[2] Daron Acemoglu and Pascual Restrepo. 2019. Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives* 33, 2 (May 2019), 3–30. doi:10.1257/jep.33.2.3

[3] Sayed Fayaz Ahmad, Heesup Han, Muhammad Mansoor Alam, Mohd. Khairul Rehmat, Muhammad Irshad, Marcelo Arraño-Muñoz, and Antonio Ariza-Montes. 2023. Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities and Social Sciences Communications* 10, 1 (June 2023), 311. doi:10.1057/s41599-023-01787-8

[4] Albert Bandura. 1986. The Explanatory and Predictive Scope of Self-Efficacy Theory. *Journal of Social and Clinical Psychology* 4, 3 (Sept. 1986), 359–373. doi:10.1521/jscp.1986.4.3.359

[5] Albert Bandura. 1997. *Self-efficacy: The exercise of control.* Freeman, New York, NY, USA.

[6] Albert Bandura. 1999. Social Cognitive Theory: An Agentic Perspective. *Asian Journal of Social Psychology* 2, 1 (April 1999), 21–41. doi:10.1111/1467-839X.00024

[7] Izak Benbasat and Henri Barki. 2007. Quo vadis TAM? *J. AIS* 8 (April 2007). doi:10.17705/1jais.00126

[8] James E. Bessen, Maarten Goos, Anna Salomons, and Wiljan Van Den Berge. 2019. *Automatic Reaction - What Happens to Workers at Firms that Automate?* Technical Report. Boston University School of Law. https://www.ssrn.com/abstract=3328877

[9] Avinash Bhat, Disha Shrivastava, and Jin L.C. Guo. 2024. Do LLMs Meet the Needs of Software Tutorial Writers? Opportunities and Design Implications. In *Designing Interactive Systems Conference*. ACM, IT University of Copenhagen Denmark, 1760–1773. doi:10.1145/3643834.3660692

[10] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. Qualitative HCI Research: Going Behind the Scenes *(Synthesis Lectures on Human-Centered Informatics)*. Springer International Publishing, Cham. doi:10.1007/978-3-031-02217-3

[11] Niall Bolger, Angelina Davis, and Eshkol Rafaeli. 2003. Diary Methods: Capturing Life as it is Lived. *Annual Review of Psychology* 54, 1 (Feb. 2003), 579–616. doi:10.1146/annurev.psych.54.101601.145030

[12] Tamer Boyaci, Caner Canyakmaz, and Francis de Véricourt. 2022. Human and machine: The impact of machine input on decision-making under cognitive limitations. *ESMT Berlin Working Paper No. 20-02 (R1)* (Nov. 2022). doi:10.2139/ssrn.3740508

[13] Michelle Brachman, Amina El-Ashry, Casey Dugan, and Werner Geyer. 2024. How Knowledge Workers Use and Want to Use LLMs in an Enterprise Context. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/3613905.3650841

[14] Michelle Brachman, Amina El-Ashry, Casey Dugan, and Werner Geyer. 2025. Current and Future Use of Large Language Models for Knowledge Work. doi:10.48550/arXiv.2503.16774 arXiv:2503.16774 [cs].

[15] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. doi:10.1191/1478088706qp063oa

[16] A. Broberg. 1999. Learners as knowledge workers - Some implications. In *FIE'99 Frontiers in Education. 29th Annual Frontiers in Education Conference. Designing the Future of Science and Engineering Education. Conference Proceedings (IEEE Cat. No.99CH37011*, Vol. 2. 13a2_25–13a2_30. doi:10.1109/FIE.1999.841717 ISSN: 0190-5848.

[17] Erik Brynjolfsson, Danielle Li, and Lindsey Raymond. 2025. Generative AI at Work*. *The Quarterly Journal of Economics* 140, 2 (May 2025), 889–942. doi:10.1093/qje/qjae044

[18] Jennie Carroll. 2004. Completing Design in Use: Closing the Appropriation Cycle. In *ECIS 2004 Proceedings*, Vol. 44. Association for Information Systems. http://aisel.aisnet.org/ecis2004/44

[19] Scott Carter and Jennifer Mankoff. 2005. When participants do the capturing: the role of media in diary studies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. Association for Computing Machinery, New York, NY, USA, 899–908. doi:10.1145/1054972.1055098

[20] Davide Castelvecchi. 2022. Are ChatGPT and AlphaCode going to replace programmers? *Nature* (Dec. 2022). doi:10.1038/d41586-022-04383-z

[21] Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2 ed.). Routledge, New York. doi:10.4324/9780203771587

[22] Jacob Cohen. 1995. The earth is round (p<.05): Rejoinder. *American Psychologist* 50, 12 (1995), 1103–1103. doi:10.1037/0003-066X.50.12.1103 Place: US Publisher: American Psychological Association.

[23] Deborah R. Compeau and Christopher A. Higgins. 1995. Computer Self-Efficacy: Development of a Measure and Initial Test. *MIS Quarterly* 19, 2 (1995), 189–211. doi:10.2307/249688 Publisher: Management Information Systems Research Center, University of Minnesota.

[24] Michele Cremaschi, Max Dorfmann, and Antonella De Angeli. 2024. A Steampunk Critique of Machine Learning Acceleration. In *Designing Interactive Systems Conference*. ACM, IT University of Copenhagen Denmark, 246–257. doi:10.1145/3643834.3660688

[25] Fred D. Davis. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* 13, 3 (1989), 319–340. doi:10.2307/249008 Publisher: Management Information Systems Research Center, University of Minnesota.

[26] Fred D. Davis, Richard P. Bagozzi, and Paul R. Warshaw. 1989. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science* 35, 8 (1989), 982–1003. https://www.jstor.org/stable/2632151 Publisher: INFORMS.

[27] Fiona Draxler, Daniel Buschek, Mikke Tavast, Perttu Hämäläinen, Albrecht Schmidt, Juhi Kulshrestha, and Robin Welsch. 2023. Gender, Age, and Technology Education Influence the Adoption and Appropriation of LLMs. doi:10.48550/arXiv.2310.06556 arXiv:2310.06556 [cs].

[28] Ian Drosos, Advait Sarkar, Xiaotong Xu, Carina Negreanu, Sean Rintel, and Lev Tankelevitch. 2024. "It's like a rubber duck that talks back": Understanding Generative AI-Assisted Data Analysis Workflows through a Participatory Prompting Study. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work (CHIWORK '24)*. Association for Computing Machinery, New York, NY, USA, 1–21. doi:10.1145/3663384.3663389

[29] Molly Q Feldman and Carolyn Jane Anderson. 2024. Non-Expert Programmers in the Generative AI Future. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work (CHIWORK '24)*. Association for Computing Machinery, New York, NY, USA, 1–19. doi:10.1145/3663384.3663393

[30] Brian S. Freeman, Kendall Arriola, Dan Cottell, Emmett Lawlor, Matt Erdman, Trevor Sutherland, and Brian Wells. 2024. Evaluation of Task Specific Productivity Improvements Using a Generative Artificial Intelligence Personal Assistant Tool. doi:10.48550/arXiv.2409.14511 arXiv:2409.14511 [cs].

[31] Toshali Goel, Orit Shaer, Catherine Delcourt, Quan Gu, and Angel Cooper. 2023. Preparing Future Designers for Human-AI Collaboration in Persona Creation. In *Proceedings of the 2nd Annual Meeting of the Symposium on Human-Computer Interaction for Work (CHIWORK '23)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3596671.3598574

[32] Jon (Sean) Jasperson, Pamela E. Carter, and Robert W. Zmud. 2005. A Comprehensive Conceptualization of Post-Adoptive Behaviors Associated with Information Technology Enabled Work Systems. *MIS Quarterly* 29, 3 (2005), 525–557. doi:10.2307/25148694 Publisher: Management Information Systems Research Center, University of Minnesota.

[33] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneci. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences* 103 (April 2023), 102274. doi:10.1016/j.lindif.2023.102274

[34] Charlotte Kobiella, Yarhy Said Flores López, Franz Waltenberger, Fiona Draxler, and Albrecht Schmidt. 2024. "If the Machine Is As Good As Me, Then What Use Am I?" – How the Use of ChatGPT Changes Young Professionals' Perception of Productivity and Accomplishment. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–16. doi:10.1145/3613904.3641964

[35] Peter Kun, Matthias Anton Freiberger, Anders Sundnes Løvlie, and Sebastian Risi. 2024. GenFrame – Embedding Generative AI Into Interactive Artifacts. In *Designing Interactive Systems Conference*. ACM, IT University of Copenhagen Denmark, 714–727. doi:10.1145/3643834.3660750

[36] Jie Li, Hancheng Cao, Laura Lin, Youyang Hou, Ruihao Zhu, and Abdallah El Ali. 2024. User Experience Design Professionals' Perceptions of Generative Artificial Intelligence. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–18. doi:10.1145/3613904.3642114

[37] Jing Liang, Lili Wang, Jia Luo, Yufei Yan, and Chao Fan. 2023. The relationship between student interaction with generative artificial intelligence and learning achievement: serial mediating roles of self-efficacy and cognitive engagement. *Frontiers in Psychology* 14 (Dec. 2023). doi:10.3389/fpsyg.2023.1285392 Publisher: Frontiers.

[38] Zhehui Liao, Maria Antoniak, Inyoung Cheong, Evie Yu-Yen Cheng, Ai-Heng Lee, Kyle Lo, Joseph Chee Chang, and Amy X. Zhang. 2024. LLMs as Research Tools: A Large Scale Survey of Researchers' Usage and Perceptions. doi:10.48550/arXiv.2411.05025 arXiv:2411.05025 [cs].

[39] Tao Long, Katy Ilonka Gero, and Lydia B Chilton. 2024. Not Just Novelty: A Longitudinal Study on Utility and Customization of an AI Workflow. In *Designing Interactive Systems Conference*. ACM, IT University of Copenhagen Denmark, 782–803. doi:10.1145/3643834.3661587

[40] Ali Madani, Ben Krause, Eric R. Greene, Subu Subramanian, Benjamin P. Mohr, James M. Holton, Jose Luis Olmos, Caiming Xiong, Zachary Z. Sun, Richard Socher, James S. Fraser, and Nikhil Naik. 2023. Large language models generate functional protein sequences across diverse families. *Nature Biotechnology* 41, 8 (Aug. 2023), 1099–1106. doi:10.1038/s41587-022-01618-2

[41] Ethan R. Mollick and Lilach Mollick. 2022. New Modes of Learning Enabled by AI Chatbots: Three Methods and Assignments. doi:10.2139/ssrn.4300783

[42] Ethan R. Mollick and Lilach Mollick. 2023. Assigning AI: Seven Approaches for Students, with Prompts. *SSRN Electronic Journal* (2023). doi:10.2139/ssrn.4475995

[43] Michael Muller, Justin D. Weisz, Stephanie Houde, and Steven I. Ross. 2024. Drinking Chai with Your (AI) Programming Partner: Value Tensions in the Tokenization of Future Human-AI Collaborative Work. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work (CHIWORK '24)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3663384.3663390

[44] Shakked Noy and Whitney Zhang. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *American Association for the Advancement of Science* 381, 6654 (July 2023), 187 – 192. doi:10.1126/science.adh2586

[45] OECD. 2024. *The impact of Artificial Intelligence on productivity, distribution and growth: Key mechanisms, initial evidence and policy challenges*. OECD Artificial Intelligence Papers 15. doi:10.1787/8d900037-en Series: OECD Artificial Intelligence Papers Volume: 15.

[46] Sami Paavola and Kai Hakkarainen. 2005. The Knowledge Creation Metaphor – An Emergent Epistemological Approach to Learning. *Science & Education* 14, 6 (Aug. 2005), 535–557. doi:10.1007/s11191-004-5157-0

[47] Justin Paul, Akiko Ueno, and Charles Dennis. 2023. ChatGPT and consumers: Benefits, Pitfalls and Future Research Agenda. *International Journal of Consumer Studies* 47, 4 (2023), 1213–1225. doi:10.1111/ijcs.12928

[48] Lee Rainie. 2025. Close encounters of the AI kind: The increasingly human-like way people are engaging with language models. (March 2025). https://imaginingthedigitalfuture.org/wp-content/uploads/2025/03/ITDF-LLM-User-Report-3-12-25.pdf

[49] Paavo Ritala, Mika Ruokonen, and Laavanya Ramaul. 2023. Transforming boundaries: how does ChatGPT change knowledge work? *Journal of Business Strategy* 45, 3 (July 2023), 214–220. doi:10.1108/JBS-05-2023-0094 Publisher: Emerald Publishing Limited.

[50] Antti Salovaara and Sakari Tamminen. 2009. Acceptance or Appropriation? A Design-Oriented Critique of Technology Acceptance Models. In *Future Interaction Design II*. 157–173.

[51] Sowmya Somanath, Bahare Bakhtiari, and Regan L Mandryk. 2024. Design Fiction on Capturing, Amplifying, and Instilling Happiness in Work. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. ACM, Newcastle upon Tyne United Kingdom, 1–17. doi:10.1145/3663384.3663395

[52] Matthias Stadler, Maria Bannert, and Michael Sailer. 2024. Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior* 160 (Nov. 2024), 108386. doi:10.1016/j.chb.2024.108386

[53] Steffen Steinert, Karina E. Avila, Stefan Ruzika, Jochen Kuhn, and Stefan Küchemann. 2024. Harnessing large language models to develop research-based learning assistants for formative feedback. *Smart Learning Environments* 11, 1 (Dec. 2024), 62. doi:10.1186/s40561-024-00354-1

[54] Chris Stokel-Walker. 2023. Generative AI Is Coming For the Lawyers. *Wired* (Feb. 2023). https://www.wired.com/story/chatgpt-generative-ai-is-coming-for-the-lawyers/

[55] Jérémie Sublime and Ilaria Renna. 2024. Is ChatGPT Massively Used by Students Nowadays? A Survey on the Use of Large Language Models such as ChatGPT in Educational Settings. doi:10.48550/arXiv.2412.17486 arXiv:2412.17486 [cs].

[56] Hyewon Suh, Aayushi Dangol, Hedda Meadan, Carol A. Miller, and Julie A. Kientz. 2024. Opportunities and Challenges for AI-Based Support for Speech-Language Pathologists. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work (CHIWORK '24)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3663384.3663387

[57] Heshan Sun. 2012. Understanding User Revisions When Using Information System Features: Adaptive System Use and Triggers. *MIS Quarterly* 36, 2 (2012), 453–478. doi:10.2307/41703463 Publisher: Management Information Systems Research Center, University of Minnesota.

[58] Yuan Sun, Eunchae Jang, Fenglong Ma, and Ting Wang. 2024. Generative AI in the Wild: Prospects, Challenges, and Strategies. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–16. doi:10.1145/3613904.3642160

[59] Daniel Susskind. 2017. *A Model of Technological Unemployment*. Technical Report. Oxford Handbook of AI Governance. https://www.semanticscholar.org/paper/A-Model-of-Technological-Unemployment-Susskind/ca10d575bc84eb16f0d5fda58bd1cf06c26adb7f

[60] Macy Takaffoli, Sijia Li, and Ville Mäkelä. 2024. Generative AI in User Experience Design and Research: How Do UX Practitioners, Teams, and Companies Use GenAI in Industry?. In *Designing Interactive Systems Conference*. ACM, IT University of Copenhagen Denmark, 1579–1593. doi:10.1145/3643834.3660720

[61] Darren Shu Jeng Ting, Ting Fang Tan, and Daniel Shu Wei Ting. 2023. ChatGPT in ophthalmology: the dawn of a new era? *Eye* (June 2023), s41433–023–02619–4. doi:10.1038/s41433-023-02619-4

[62] B van den Hooff. 2005. A learning process in email use - a longitudinal case study of the interaction between organization and technology. *Behaviour & Information Technology* 24, 2 (March 2005), 131–145. doi:10.1080/01449290512331321893

[63] Viswanath Venkatesh. 2000. Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. *Information Systems Research* 11 (Dec. 2000), 342–365. doi:10.1287/isre.11.4.342.11872

[64] Viswanath Venkatesh and Fred Davis. 2000. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science* 46 (Feb. 2000), 186–204. doi:10.1287/mnsc.46.2.186.11926

[65] Viswanath Venkatesh and Fred D. Davis. 1996. A Model of the Antecedents of Perceived Ease of Use: Development and Test. *Decision Sciences* 27, 3 (1996), 451–481. doi:10.1111/j.1540-5915.1996.tb00860.x

[66] Thomas Weber, Maximilian Brandmaier, Albrecht Schmidt, and Sven Mayer. 2024. Significant Productivity Gains through Programming with Large Language Models. *Proc. ACM Hum.-Comput. Interact.* 8, EICS (June 2024), 256:1–256:29. doi:10.1145/3661145

[67] Allison Woodruff, Renee Shelby, Patrick Gage Kelley, Steven Rousso-Schindler, Jamila Smith-Loud, and Lauren Wilcox. 2024. How Knowledge Workers Think Generative AI Will (Not) Transform Their Industries. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–26. doi:10.1145/3613904.3642700

[68] Qian Zhu, Dakuo Wang, Shuai Ma, April Yi Wang, Zixin Chen, Udayan Khurana, and Xiaojuan Ma. 2024. Towards Feature Engineering with Human and AI's Knowledge: Understanding Data Science Practitioners' Perceptions in Human&AI-Assisted Feature Engineering Design. In *Designing Interactive Systems Conference*. ACM, IT University of Copenhagen Denmark, 1789–1804. doi:10.1145/3643834.3661517

A Per-Participant Scale Responses

Table 4: Mean (SD) scores per participant for perceived productivity, accomplishment, and self-efficacy.

Participant	Perceived Productivity	Perceived Accomplishment	Perceived Self-efficacy
P1	5.80 (0.45)	5.60 (0.89)	6.00 (0.00)
P2	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)
P3	4.83 (0.75)	4.50 (1.76)	4.83 (0.98)
P5	4.60 (0.89)	4.80 (0.84)	4.80 (0.84)
P6	4.33 (2.25)	5.33 (1.63)	5.17 (1.60)
P7	4.33 (1.51)	3.67 (0.82)	4.17 (1.33)
P8	5.00 (0.82)	4.00 (0.00)	4.25 (0.96)
P9	5.20 (0.45)	4.80 (0.84)	5.00 (0.71)
P10	5.50 (0.58)	4.00 (0.82)	4.00 (0.82)
P12	4.50 (1.38)	4.17 (0.98)	3.50 (1.22)
P14	4.40 (0.89)	4.20 (0.84)	4.40 (0.89)
P15	4.50 (1.52)	4.67 (1.37)	4.33 (0.82)
P17	5.00 (1.15)	5.00 (1.15)	4.50 (1.29)
P19	4.50 (0.71)	4.00 (0.00)	3.50 (0.71)
Overall	4.82 (0.95)	4.55 (0.85)	4.53 (0.87)