The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers

Hao-Ping (Hank) Lee
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
haopingl@cs.cmu.edu

Advait Sarkar
Microsoft Research
Cambridge, United Kingdom
advait@microsoft.com

Lev Tankelevitch
Microsoft Research
Cambridge, United Kingdom
levt@microsoft.com

Ian Drosos
Microsoft Research
Cambridge, United Kingdom
t-iandrosos@microsoft.com

Sean Rintel
Microsoft Research
Cambridge, United Kingdom
serintel@microsoft.com

Richard Banks
Microsoft Research Cambridge
Cambridge, United Kingdom
rbanks@microsoft.com

Nicholas Wilson
Microsoft Research
Cambridge, United Kingdom
niwilson@microsoft.com

Abstract

The rise of Generative AI (GenAI) in knowledge workflows raises questions about its impact on critical thinking skills and practices. We survey 319 knowledge workers to investigate 1) when and how they perceive the enaction of critical thinking when using GenAI, and 2) when and why GenAI affects their effort to do so. Participants shared 936 first-hand examples of using GenAI in work tasks. Quantitatively, when considering both task- and user-specific factors, a user's task-specific self-confidence and confidence in GenAI are predictive of whether critical thinking is enacted and the effort of doing so in GenAI-assisted tasks. Specifically, higher confidence in GenAI is associated with less critical thinking, while higher self-confidence is associated with more critical thinking. Qualitatively, GenAI shifts the nature of critical thinking toward information verification, response integration, and task stewardship. Our insights reveal new design challenges and opportunities for developing GenAI tools for knowledge work.

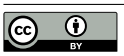
CCS Concepts

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

Keywords

Critical thinking, Generative AI tools, Knowledge worker, Bloom's taxonomy, Survey

ACM Reference Format:
Hao-Ping (Hank) Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. 2025. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/3706598.3713778



CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713778

1 Introduction

Generative AI (**GenAI**) tools, defined as any *"end user tool [...] whose technical implementation includes a generative model based on deep learning"*,[1] are the latest in a long line of technologies that raise questions about their impact on the quality of human thought, a line that includes writing (objected to by Socrates), printing (objected to by Trithemius), calculators (objected to by teachers of arithmetic), and the Internet.

Such consternation is not unfounded. Used improperly, technologies can and do result in the deterioration of cognitive faculties that ought to be preserved. As Bainbridge [7] noted, a key irony of automation is that by mechanising routine tasks and leaving exception-handling to the human user, you deprive the user of the routine opportunities to practice their judgement and strengthen their cognitive musculature, leaving them atrophied and unprepared when the exceptions do arise.

In response, research has begun looking closely at how different activities are impacted by GenAI and the extent to which cognitive offloading [8] occurs, and whether this may be an undesirable thing. Some work has focused, for instance, on studying the effects of GenAI use on memory (e.g., [1, 106]) and on creativity (e.g., [28, 100]). Moreover, design research has also been developing interventions that *improve* the ability of people to think in certain ways (e.g., [24]). We review these lines of work in Section 2.

In this paper, we focus on a higher-level concept that captures another aspect of thought considered desirable and worthy of preservation: *critical thinking* (defined in Section 2). The effect of the use

[1]While there is no broad consensus on how to define this now-common term, for clarity we adopt this definition, a rationale for which is given in [115].

of GenAI tools on critical thinking, as a direct object of inquiry, has not yet been explored.

Moreover, we focus on critical thinking for *knowledge work* (as conceptualised by Drucker [30] and Kidd [67]). Much research on the effect of GenAI on thinking skills is focused on educational settings, where concern for skill cultivation is most acute (e.g., the effect of GenAI code completion tools on programming and computer science education [107]). As previously noted [116, 119], critical thinking has been operationalised in detail in certain specific disciplines, such as academic history, clinical psychology, and nursing. But the ostensible shifts in critical thinking behaviours brought about by GenAI extend to a broad set of professions and knowledge workflows — GenAI tools are now widely used in knowledge work [13] — and little is known about the critical thinking demands of these. We lack broad-based empirical examples of what *kinds* of knowledge work activities are considered by professionals to require critical thinking.

Recent work has motivated the need for critical thinking support in AI-assisted knowledge work [116, 119]. It is motivated primarily by the observation of the tendency of AI-assisted knowledge workflows to be subject to "mechanised convergence" [114], i.e., that users with access to GenAI tools produce a less diverse set of outcomes for the same task, compared to those without. This tendency for convergence reflects a lack of personal, contextualised, critical and reflective judgement of AI output and thus can be interpreted as a deterioration of critical thinking.

However, we lack direct empirical evidence for an interpretation that posits a connection between mechanised convergence and critical thinking. Output diversity is a *proxy* for critical thinking, and a flawed one. For instance, users who reuse GenAI output without editing it may have nonetheless performed a critical, reflective judgement in forming the decision not to edit it. Such reflective thinking is invisible to measures that focus only on the ultimate artefact produced. Without knowing how knowledge workers enact critical thinking when using GenAI and the associated challenges, we risk creating interventions that do not address workers' real needs.

In this paper, we aim to address this gap by conducting a survey of a professionally diverse set of knowledge workers ($n = 319$), eliciting detailed real-world examples of tasks (936) for which they use GenAI, and directly measuring their perceptions of critical thinking during these tasks: when is critical thinking necessary, how is critical thinking enacted, whether GenAI tools affect the effort of critical thinking, and to what extent (Section 3). We focus on "enaction" (i.e., actions that are signals or manifestations) of critical thinking rather than critical thinking *per se*, because critical thinking itself as a pure mental phenomenon is difficult for people to self-observe, reflect on, and report.

Concretely, we aim to answer two research questions:

RQ1 When and how do knowledge workers perceive the enaction of critical thinking when using GenAI?

RQ2 When and why do knowledge workers perceive increased/decreased effort for critical thinking due to GenAI?

With respect to RQ1 (Section 4), the study reveals that knowledge workers engage in critical thinking when using GenAI tools primarily to ensure the quality of their work. They define critical thinking as setting clear goals, refining prompts, and assessing AI-generated content to meet specific criteria and standards. Their reflective approach involves verifying outputs against external sources and their own expertise, especially in tasks that require higher accuracy.

The data identify key motivators for critical thinking: the desire to enhance work quality, avoid negative outcomes, and develop skills. However, several barriers inhibit this reflective process, including lack of awareness, limited motivation due to time pressure or job scope, and difficulty improving AI responses in unfamiliar domains. Surprisingly, while AI can improve efficiency, it may also reduce critical engagement, particularly in routine or lower-stakes tasks in which users simply rely on AI, raising concerns about long-term reliance and diminished independent problem-solving.

Regarding RQ2 (Section 5), GenAI tools appear to reduce the perceived effort required for critical thinking tasks among knowledge workers, especially when they have higher confidence in AI capabilities. However, workers who are confident in their own skills tend to perceive greater effort in these tasks, particularly when evaluating and applying AI responses.

The data shows a shift in cognitive effort as knowledge workers increasingly move from task execution to oversight when using GenAI. While this shift "from material production to critical integration" has been observed in prior studies [114], such studies are typically controlled studies in narrow domains with small participant samples. Our data provides complementary evidence that this also occurs in real-world use of GenAI tools, across a wide variety of tasks and professions. For tasks like knowledge retrieval, AI reduces effort by automating information gathering, but workers must now invest more in verifying the accuracy of AI outputs. Similarly, while AI simplifies content creation, workers still need to spend time aligning outputs with specific needs and quality standards.

Our paper makes the following contributions:

- We review the literature on interaction design interventions for critical thinking, and studies of the effects of automation on knowledge workflows (Section 2).
- We describe the development and deployment of a survey for gathering empirical evidence for knowledge workers' experiences and perceptions of the effect of GenAI on critical thinking (Section 3). We find that GenAI tools reduce the perceived effort of critical thinking while also encouraging over-reliance on AI, with confidence in the tool often diminishing independent problem-solving. As workers shift from task execution to AI oversight, they trade hands-on engagement for the challenge of verifying and editing AI outputs, revealing both the efficiency gains and the risks of diminished critical reflection (Sections 4 and 5).
- Drawing from our survey insights, we highlight how the use of GenAI tools creates new challenges for critical thinking. We outline implications for designing GenAI to support knowledge workers to enhance their awareness, motivation, and ability to think critically (Section 6).

2 Related Work

2.1 Critical thinking

We adopt the definition of critical thinking developed by Bloom et al. [12, 54], a hierarchical taxonomy that characterises student learning objectives into six types: knowledge (recall of ideas), comprehension (demonstrating understanding of ideas), application (putting ideas into practice), analysis (contrasting and relating ideas), synthesis (combining ideas), and evaluation (judging ideas through criteria).

This definition of critical thinking is not uncontested. There are multiple alternative frameworks [36–38, 104], and critical thinking is sometimes also referred to as reflective thinking [26], though not all scholars conflate them. There have been multiple proposals for connecting and reconciling this multiplicity of frameworks [32, 74, 96].

We adopt the Bloom et al. framework for multiple reasons. First, as one of the earliest frameworks, it has strong support in the research literature and wide adoption in education systems — its definition of critical thinking has been widely influential, and has withstood severe criticism and scrutiny [40]. Second, it is relatively simple, having only six core dimensions (as opposed to, for instance, the nuanced Paul-Elder framework [104] which consists of eight "elements of thought", ten "intellectual standards", and eight "intellectual virtues"). The simplicity of the Bloom et al. framework — its small set of dimensions with clear definitions — renders it more suitable as the basis of a survey instrument.

Critical thinking skills can be developed in sequential stages [70, 98, 104]. Despite concerns about whether critical thinking can be taught [138], research in education has developed a number of approaches to teaching critical thinking [104, 139], such as structured argumentation exercises [25, 70, 72, 133]. Critical thinking can be measured through self-, peer-, or expert evaluation [66], using a range of questionnaires [35, 65, 73, 145, 146], justified multiple choice questions, structured essays, protocols for whole-portfolio assessment, task observation, and peer interaction [34, 105]. In our study, we apply a one-item five-point scale assessment for each of the six cognitive activities associated with critical thinking (six items in total, see Section 3.1.3), similar to previous work (e.g., Alaoutinen and Smolander [3]).

2.2 Design research for critical and reflective thinking

Previous research has investigated how interaction design can encourage critical or reflective thinking. Various dimensions of the space of critical thinking interventions have been explored. For instance, whether the system should be proactive, i.e., introduce critical thinking prompts without an explicit user request [69, 109]. Or the extent to which user participation and engagement is important in creating critical thinking outcomes, e.g., presenting AI explanations as questions rather than statements improves logical discernment [24], questions also improve critical reading [110, 142], attention checks promote systematic thinking [49], conflict-filled discussion induces critical thinking [78], and in general increased engagement results in behavioural changes [82, 92]. Research has explored the effectiveness of gamification of critical thinking [31, 91, 129]. Research has also explored the extent to which interventions ought to be presented in an agentised or anthropomimetic manner [99, 131, 141].

There are domains and activities, some of which are relevant to common knowledge workflows, where critical thinking interventions have been heavily studied. For example, design for critical thinking can aid in the prevention and verification of misinformation, e.g., through structured thinking aids [50, 51], analytical thinking nudges [143], worksheets and group discussion [136], and gamification [129]. Or in writing, ideation and argumentation tools, such as through visualising argument structure [126, 133], reflecting on future scenarios [132], ideation and evaluation support [45], assessing risks in research impact statements [94]. Another common area for reflective thinking interventions is in mental health and wellbeing, e.g., to support cognitive reappraisal [71], reduce compulsive smartphone use [80, 81], improve time management [55], create journaling prompts [97], encourage reflection on book highlights [61], support prayer [75], coaching for leadership growth [4], and reflection on cherished objects [57]. Critical thinking interventions have also been explored in data analysis [44, 48].

Overreliance, defined as "users accepting incorrect recommendations, i.e., making errors of commission" [102], is closely related to (the lack of) critical thinking. Buçinca et al. [17] found that "cognitive forcing functions" such as requiring the user to wait before receiving AI output, or to make interactive updates to AI output, significantly reduce overreliance compared to simpler AI explanations. Though there is overlap, overreliance is not strictly the same problem as (and is perhaps a special case of) a lack of critical thinking. A lack of critical thinking may also manifest through accepting a solution that merely meets a baseline aspirational threshold [6, 119] — in such cases, the AI solution is correct (albeit potentially of poor quality) and therefore not overreliance, strictly speaking.

Collectively, these can inform design interventions to support critical thinking for knowledge workers. Still, these systems and tools do not engage with how the need for critical thinking support changes due to shifts in workflow caused specifically by the use of GenAI. We also lack empirical foundations for understanding how knowledge workers enact critical thinking in real-world GenAI workflows.

2.3 Effects of automation on thinking and knowledge workflows: writing and memory

Effects on writing. Generative AI tools like Copilot and ChatGPT can boost writing productivity by assisting with tasks such as content generation, idea creation, and stylistic editing, helping both expert and novice writers [18, 84, 112, 135]. However, there are concerns that novice writers may become overly reliant on these tools, potentially impairing their long-term skill development by bypassing critical writing processes such as constructing logical arguments and understanding subject matter [14, 53, 63, 64]. To mitigate this, using GenAI for individualised, content-focused feedback may help novice writers develop writing skills while improving productivity [58, 86, 144]. Although human feedback has traditionally been necessary for effective self-improvement, the integration of AI into tools like Microsoft Word could democratise access to writing skill development by providing consistent, low-cost feedback [2, 123]. Early studies suggest that AI-generated

feedback can improve writing quality and logical structure, especially for lower-performing students and less confident English learners [79, 101, 128, 135]. Thus, equipping AI tools with better feedback mechanisms could foster long-term writing skill development while addressing inequalities in access to writing education [2, 79], and enable humans and AI to interact over time to maximise both productivity and learning outcomes [128, 135].

Effects on memory. While GenAI and conversational search engines can streamline tasks like literature reviews, some fear that outsourcing this work could harm our ability to learn and remember, in what is sometimes referred to as "digital amnesia" [47, 111], though evidence for this effect is largely inconclusive [19, 21, 27, 127]. Research shows that summarising material and follow-up writing practice enhance memory by integrating new knowledge with existing knowledge [62, 93, 134], but real-world summary writing is often passive and ineffective [15, 16, 41, 121, 140], and thus may not improve recall in comparison to simply re-reading the text [124]. GenAI tools like ChatGPT and Copilot can mitigate these drawbacks, especially for less experienced learners, by providing high-quality summaries upon which collaborative, self-monitored writing tasks can be conducted [120, 125]. Cognitive science shows that effective learning requires "grounding" information through multiple perspectives and examples [10, 11, 68], and GenAI can offer personalised analogies to aid this process [77, 90].

In summary, previous work has defined critical thinking and investigated ways to develop and measure this skill in educational settings. Separately, design research has investigated ways of developing technology that induces critical reflection. It has also been found that AI tools can significantly impact common knowledge workflows, such as writing. However, there is a gap in understanding knowledge workers' perceptions of how GenAI affects their enaction of critical thinking, and the effort of doing so, across a broad range of use cases. This is the gap we address with our survey.

3 Method

To answer our research questions — when and how knowledge workers perceive the enaction of critical thinking when using GenAI (RQ1), and when and why do knowledge workers perceive increased/decreased effort for critical thinking due to GenAI (RQ2) — we conducted an online survey on the Prolific platform[2] to study knowledge workers' experiences with critical thinking when using GenAI tools for their work.

To ensure participants fully understood the scope and meaning of our questions on critical thinking, as part of the survey study onboarding, they were introduced to the concept of critical thinking in the context of using GenAI through concrete examples of how critical thinking could be applied at various levels of Bloom's taxonomy (e.g., checking the tone of generated emails, verifying the accuracy of code snippets, and assessing potential biases in data insights). These examples served to sensitise participants to the various dimensions of critical thinking while avoiding conceptualising critical thinking too narrowly. These acted as "cognitive priming", helping participants better understand the concept of critical thinking, thus soliciting better recognition of critical thinking behaviours in participants' daily GenAI use.

In total, we received 319 survey responses, in which participants shared a total of 936 real-world examples where they used a GenAI tool for their work, and shared how critical thinking played a role in these tasks.

To answer **RQ1**, we created an explanatory regression model with a dependent variable measuring *whether participants perceived the enaction of critical thinking* when using GenAI tools for the tasks they shared, and independent variables corresponding to two sets of factors that we hypothesised might correlate with the tendency to engage with tasks critically: 1) **task factors**: measures about the task at hand — e.g., task type, confidence in doing the task. 2) **User factors**: measures about users — e.g., age, gender, occupation, tendency to reflect in work, and trust in GenAI. In addition, we analysed participants' motivators and inhibitors for critical thinking from their free-text responses.

To answer **RQ2**, we create explanatory regression models with dependent variables measuring *whether participants perceived different cognitive activities constituting critical thinking (e.g., breaking down a problem, putting together ideas) to be more or less effortful* when using a GenAI tool for the tasks compared to when not using one. Independent variables included the same set of factors as for RQ1 above. We also analysed participants' free-text responses to understand why they perceived these cognitive activities as more or less effortful due to GenAI.

3.1 Survey Design

To model the relationship between task and user factors as they relate to critical thinking activities, we designed a survey as follows (see Appendix A.1 for the complete survey).

3.1.1 Task-Related Factors. Prior studies have shown that knowledge workers apply GenAI tools for a range of tasks and express different needs while doing these tasks [13], and that their perceived confidence in themselves and AI doing the tasks can influence their use and reliance on the tool [20, 22, 83, 130]. We hypothesised that factors relating to the user's task, including task type, confidence in themselves, and AI doing the task, could affect their critical thinking.

Task type. Brachman et al. [13] classify knowledge workers' current usage of GenAI tools into nine types (See Table 1), grouped into three major categories: 1) for **creation**, 2) to find or work with **information**, 3) to get **advice**. This taxonomy offers clear distinctions among the major categories of task type, which we hypothesised would correlate with users' critical thinking due to differing objectives and requirements. We follow Brachman et al. [13]'s taxonomy and operationalise their task type categorisation in our survey, focusing on the major categories. For each GenAI tool use example, participants were first asked to describe in detail the task they did (i.e., *Please tell us: 1) what you were trying to achieve, 2) in what GenAI tool, and 3) how you used the GenAI tool, including any prompts.*). Then, they were asked to pick one of the nine task types that best described their task. Using this information, we

[2] https://prolific.co/

Table 1: Categories and sub-categories for GenAI tool usage [13].

Category	Sub-category	Description
Creation	Artefact	Generate a new artefact to be used directly or with some modification
	Idea	Generate an idea, to be used indirectly
Information	Search	Seek a fact or piece of information
	Learn	Learn about a new topic more broadly
	Summarise	Generate a shorter version of a piece of content that describes the important elements
	Analyse	Discover a new insight about information or data
Advice	Improve	Generate a better version
	Guidance	Get guidance about how to make a decision
	Validation	Check whether an artefact satisfies a set of rules or constraints

classified each example as creation, information, or advice, per the Brachman et al. [13] taxonomy.

Task confidence. Guided by prior studies on user confidence in AI-assisted decision-making [20, 85, 130], for each self-reported task we consider three aspects of user confidence: 1) **confidence in self** (i.e., *How confident are you in your ability to do this task without GenAI?*), 2) **confidence in GenAI** (i.e., *How confident are you in the ability of GenAI to do this task?*), and 3) **confidence in evaluation** (i.e., *How confident are you, in the course of your normal work, in evaluating the output that AI produces for this task?*). Participants rated each aspect of confidence on a five-point scale ranging from *"not at all confident"* (1) to *"extremely confident"* (5).

3.1.2 User factors. We hypothesised that participants' general tendency to reflective thinking and trust in GenAI would affect their baseline critical thinking awareness and practice, and adapted validated instruments from prior work to measure this.

Tendency to reflect on work. We use Kember et al. [65]'s Reflective Thinking Inventory to measure participants' baseline tendency to think reflectively. Reflective thinking is closely related to critical thinking (Section 2) and the Kember et al. inventory can be interpreted as a proxy for the disposition to think critically [38].

Trust in generative AI. We measure participants' overall trust in GenAI, which has been shown to correlate with users' attitudes and adoption of the use of the technologies [43, 76]. To that end, we adapted the six-item Propensity to Trust Technology scale [56], replacing the word "technology" with "GenAI".

Gender, age, and occupation. We collect demographic information, including gender, age range and occupation. For occupation, participants self-selected the most appropriate occupation category from the Occupational Information Network (O*NET)'s occupational listings[3]. We classify occupations as being **in risk of automation** based on the economic analyses of Ghosh et al. [42], including the categories of Office and Administrative Support, Sales and Related, Computer and Mathematical, Business and Financial Operations, and Arts, Design, Entertainment, Sports, and Media.

3.1.3 Critical Thinking, Associated Cognitive Activities, and Effort.

Perceived enaction of critical thinking. A key dependent variable of **RQ1** — *when* knowledge workers perceive the enaction to think critically — was answered using a pair of questions, first asking whether participants perceived that they had *performed* critical thinking for that task (a binary yes/no question), followed by a free text question asking them to justify their response. If participants answered "yes" to the first question, they were asked to elaborate why and how they enacted critical thinking in free text (i.e., *Please share one real-world example when you applied the critical thinking tactic(s) to this task, and explain why you did critical thinking.*), as well as the challenges, if any, they faced while doing so (i.e., *When applying this critical thinking tactic during your use of GenAI tool, have you ever encountered any challenges and obstacles?*). If the participants answered "no" to the question, they were asked to elaborate on why they did not think critically for the task, in free text.

Perceived effort in critical thinking: Bloom's taxonomy. As discussed in Section 2, we selected Bloom's taxonomy as the framework to operationalise the measurement of critical thinking activities [12]. The taxonomy includes six different levels of cognitive activities: Knowledge (i.e., recall), Comprehension (i.e., organising/translating ideas), Application (i.e., problem-solving), Analysis (i.e., breaking down a problem), Synthesis (i.e., putting together ideas), and Evaluation (i.e., evaluating and quality checking). See Table 2 for more details.

For each task example, participants were asked if, and how much, the use of the GenAI tool changed the effort of critical thinking activities compared to when they did not use the AI tool. We used the five-point scale *"much less effort"*, *"less effort"*, *"about the same"*, *"more effort"*, to *"much more effort"* (which we code as integers ranging between −2 and +2). Participants could choose "N/A" if they thought that a cognitive activity was not relevant to the task. Finally, participants were asked to elaborate in free-text why they had marked any critical thinking activities as requiring more or less effort with GenAI.

3.2 Study Setup and Recruitment

We recruited participants through the Prolific platform who self-reported using GenAI tools at work at least once per week. This criterion ensured the study focused on knowledge workers with direct, ongoing experience integrating GenAI tools into their day-to-day work tasks. We received 333 responses but excluded 14 from

[3]A list of 23 occupation categories listed as "Major Group" in https://www.onetcenter.org/taxonomy/2019/structure.html

Table 2: Cognitive activities defined in Bloom's taxonomy [12].

Cognitive activity	Description
Knowledge	Recognising or remembering facts, terms, basic concepts, or answers
Comprehension	Organising, summarising, translating, generalising, giving descriptions, and stating the main ideas
Application	Using acquired knowledge to solve problems in new situations
Analysis	Examining and breaking information into component parts, determining how the parts relate to one another, identifying motives or causes, making inferences, and finding evidence to support generalisations
Synthesis	Building a structure or pattern from diverse elements; putting parts together to form a whole or bringing pieces of information together to form a new meaning
Evaluation	Presenting and defending opinions by making judgements about information, the validity of ideas, or quality of work based on a set of criteria

Table 3: Participant demographics.

Dimension	Sub-dimension	Participants
Gender	Man	159 (49.84%)
	Woman	153 (47.96%)
	Non-binary/gender diverse	5 (1.57%)
	Prefer not to say	2 (0.63%)
Age	18-24	86 (26.96%)
	25-34	143 (44.83%)
	35-44	62 (19.44%)
	45-54	21 (6.58%)
	55+	7 (2.19%)
GenAI tool use* (top 5)	ChatGPT	309 (96.87%)
	Microsoft Copilot (website)	74 (23.20%)
	Gemini (website)	69 (21.63%)
	Copilot in Microsoft products (e.g., Word)	60 (18.81%)
	Gemini in Google products (e.g., Google Slides)	49 (15.36%)
Occupation (top 5)	Computer and Mathematical	59 (18.50%)
	Arts, Design, Entertainment, Sports, and Media	44 (13.79%)
	Office and Administrative Support	38 (11.91%)
	Business and Financial Operations	35 (10.97%)
	Educational Instruction and Library	23 (7.21%)
Country of residency (top 5)	United Kingdom	37 (11.60%)
	Canada	25 (7.84%)
	United States	20 (6.27%)
	South Africa	18 (5.64%)
	Poland	17 (5.33%)

*participants selected all the GenAI tools they use at work

the analysis due to low response quality (i.e., low-effort free-text responses). For the remaining 319 responses, participants spent an average of 43.19 minutes (STD=23.13) in completing the survey. The 319 participants (159 men, 153 women, 5 non-binary/gender diverse, 2 prefer not to say) came from diverse age groups, occupations, and countries of residence (see Table 3). Participants were compensated with GBP £10 for completing the study. Our study protocol was approved by our institution's ethics and compliance review board. All participants were briefed and signed a consent form.

3.3 Analysis Procedure

In our survey, participants were asked to share three real examples of their GenAI tool use at work. To increase the variety of examples collected, participants were asked to think of three different examples, one for each task type: Creation, Information, and Advice (see Section 3.1.1). Then, participants were asked to share an example of each task type in detail. The order of task types was randomised to avoid order and fatigue effects. For each example, as mentioned, we measure participants' perceived enaction of critical thinking, perceived effort in critical cognitive activities, and perceived confidence. All participants shared three examples. However, they were allowed to skip any task type they did not have experience of and substitute another task type — e.g., a participant could share two examples about Creation and one example about Advice, if they had no experience of an Information task.

After participants shared three examples of using GenAI tools, the survey assessed their overall reflective thinking tendency, trust in GenAI, and demographic details such as gender, age group, and occupation.

We employed quantitative and qualitative analyses, guided by our research questions. Both **RQ1** — when and how do knowledge workers perceive the enaction of critical thinking when using GenAI? — and **RQ2** — when and why do knowledge workers perceive increased/decreased effort for critical thinking due to GenAI? — were answered via both quantitative and qualitative analysis (See Figure 1 for an overview of our approach).

3.3.1 Dataset Cleaning and Overview. Our 319 participants shared a total of 957 real-world examples of their use of GenAI tools at work. We removed 11 examples lacking sufficient detail to analyse (e.g., brief or vague examples like "To build my portfolio."). We also removed 11 examples for which a participant shared duplicated or non-GenAI tool use examples in their responses.

We retained 936 examples, including 374 (39.96%) related to Creation, 303 (32.37%) related to Information, and 259 (27.67%) related to Advice. Our participants self-reported to have enacted critical thinking for 555 (59.29%) of the examples they shared, and perceived critical thinking activities, overall, to require less effort when using a GenAI tool compared to when not using one (see *DV distribution* in Table 4).

3.3.2 Quantitative Analysis. To model the relationship between task and user factors (independent variables) with (1) a binary measure of users' perceived enaction of critical thinking and (2) six five-point scales of users' perceived effort in cognitive activities associated with critical thinking, we respectively fit (1) one random-intercepts logistic regression model and (2) six random-intercepts linear regression models. To account for repeated measures, we include Participant ID as a random intercept term. For all categorical variables, we selected the most common factor level as the baseline reference. To correct for multiple comparisons, we apply the Benjamini–Hochberg procedure [9] with a total of 98 hypothesised predictors across the seven models, yielding a corrected p-value

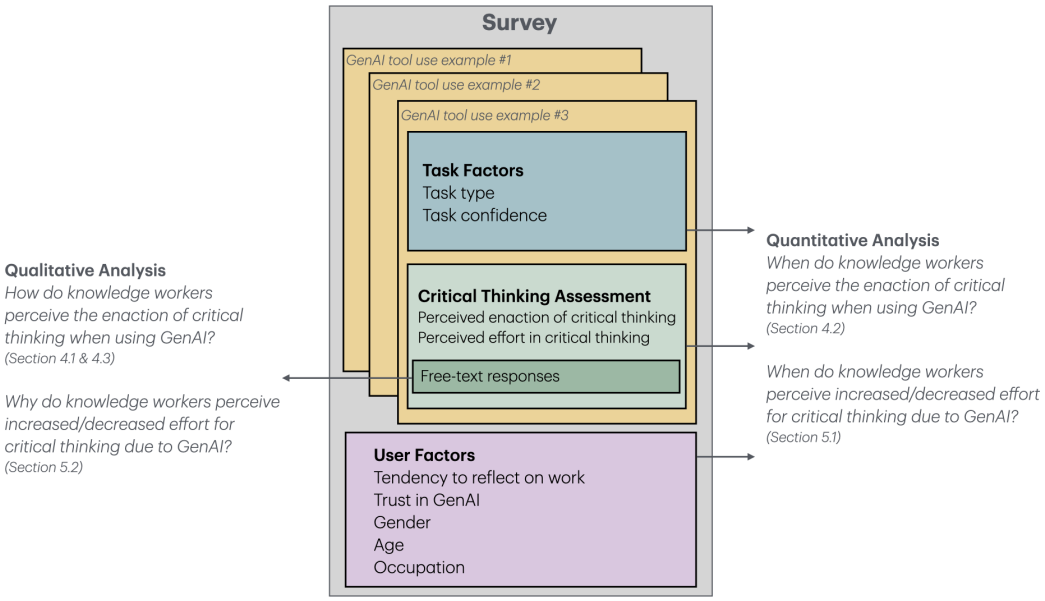


Figure 1: Schematic overview of the survey design and our corresponding analysis approach.

threshold of 0.007. We adjust the p-values accordingly and report significant effects based on these corrected values.

Table 4 summarises the seven models and reports the corrected p-values. For interpretability, we computed z-scores to standardise each numeric user factor (i.e., overall tendency to reflect, overall trust in GenAI). Thus, a positive coefficient implies the increase in log odds (in the logistic regression model) or the value (in the linear regression models), for every one standard deviation increase of that factor. A negative coefficient implies the opposite. For confidence scales (i.e., confidence in self, confidence in GenAI, confidence in evaluation), a positive coefficient is the increase in log odds/values for every one-point increase above the base score (1: not at all confident), and a negative coefficient implies the opposite. For categorical and binary factors (i.e., task type, gender, age group, occupation in risk of automation), the coefficient is the predicted difference in log odds/increase of the values for a given factor level relative to a baseline level. Positive coefficients imply increased log odds/values relative to the reference level and vice versa.

3.3.3 Qualitative Analysis. Guided by our research questions, we open-coded [23] participants' free-text responses on i) why they did or did not think critically when using GenAI tool for the task, ii) why they perceived more or less effort to perform critical thinking activities with the GenAI tool. One researcher performed the initial coding on 50 survey responses in discussion with three other researchers to iteratively construct a codebook. Another researcher joined the coding process when the initial codebook was constructed, and was trained with the initial codebook. The two researchers then coded the remaining 269 survey responses. All research team members regularly met and discussed emerging themes during the coding process. Disagreements were negotiated and resolved at each stage, using negotiated agreement best practices [87]. We report our findings in Sections 4 and 5, and include the codebook in Appendix Table 5. We also report on how frequently participants discussed the identified themes.

4 Findings for RQ1: *When and how do knowledge workers perceive the enaction of critical thinking when using GenAI?*

To answer RQ1, we investigated how knowledge workers define critical thinking (Section 4.1), and when (Section 4.2) and why (Section 4.3) they enact critical thinking in their use of GenAI tools. Qualitatively, we found that knowledge workers view critical thinking as ensuring the objectives and quality of their work. Through our quantitative analysis of *when* knowledge workers do critical thinking, we found their confidence in themselves doing and evaluating the task, and their general tendency to reflect on work strongly correlated with their perceived enaction of critical thinking. We also found a negative correlation between the perceived enaction of critical thinking and their confidence in AI doing the task. Finally, we qualitatively analysed participants' free-text responses to understand why they do or do not enact critical thinking, identifying three key motivators (work quality, potential negative outcomes, skill development) and three inhibitors (awareness, motivation, ability) for critical thinking.

4.1 How knowledge workers enact critical thinking

We first explored knowledge workers' definition and perceived enaction of critical thinking by examining the activities they describe as performing critical thinking. While our participants worked across diverse occupations, the common denominator was that they viewed critical thinking in their GenAI tool use as cognitive

activities performed to ensure the quality of AI responses, and intentionality while using the tools.

We mapped our findings to each phase of knowledge workers' GenAI tool workflow. We classified knowledge workers' critical thinking practices into 1) goal and query formation, 2) inspect response, and 3) integrate response. Our analysis is based primarily on workflow characterisations from previous work [29, 46, 130], though more general frameworks for human cognitive problem solving [137] and problem solving with AI [60, 89, 108] are also related.

4.1.1 Goal and query formation. During goal and query formation, participants enact critical thinking through prompt optimisation to produce the responses they desire. They also enact critical thinking by "taking a step back" to consolidate their goals and queries to the tools. These phenomena correspond to the goal and query formulation phases in the *iterative goal satisfaction* framework proposed by Drosos et al. [29].

Form goal (6/319). Before engaging with a tool, knowledge workers reflect on their goals, needs and intents, and identify a need for assistance where the GenAI tool could be applied. For example, when P140 tried to learn the functionality of a code snippet through ChatGPT, he saw critical thinking as the need to *"analyze what my goal was and how I was going to achieve it... I had to first learn what was I going to use in order to make progress."* Similarly, participants defined critical thinking as setting clear goals in mind before using GenAI tools to generate images (e.g., P14) and ideas for a report (e.g., P2).

Formation of intentions applies to other computational tools and is not unique to GenAI. However, as emphasised in the generative AI metacognitive framework proposed by Tankelevitch et al. [130], critical thinking in the form of goal setting is particularly relevant due to its direct connection with the process of "forming queries" — users must first establish clear goals to effectively generate queries for the tool.

Form query (30/319). Some knowledge workers enacted critical thinking by creating or revising prompts to GenAI tools to get the desired response. With a goal in mind, knowledge workers create queries that further clarify the final deliverables for the tool. For example, when P97 tried to create an art piece for her website, *"[I] was reflective when it came to giving the correct prompts, in order to get the correct result a correct description needs to be given."*

The process of iterating on a prompt may help clarify knowledge workers' goals and provide an opportunity for enacting critical thinking. For instance, when a teacher (P19) generated an image with DALL-E for her presentation about hand washing at school: *"I noticed it was missing soap dispensers. So I changed my prompt to include them and tried again... By thinking about what the image really needed to show, I got a much better result from the AI for my presentation."*

4.1.2 Inspect response. Prior work has identified the work of understanding and evaluating GenAI output as a key aspect of working with GenAI [29, 46, 114, 130]. Participants also enacted critical thinking by assessing if a GenAI output meets certain criteria and standards, or if the information it contains is verified or verifiable. They applied multiple types of quality criteria and verification approaches.

Ensure quality through objective criteria (125/319). When applicable, knowledge workers evaluate the GenAI output with objective criteria (which we define as those that are straightforward to articulate and apply[4]), such as if the output complies with their queries, or if the generated artefact is functional (e.g., generated code compiles without errors). For example, when P278 prepared a specification document for her client with ChatGPT, *"I had to make sure each piece of text generated met the requirements of the client based on criteria [in the prompt] like colour palette, and people in photos - male/female, skin tone, etc."* Similarly, when asking for a content summary, knowledge workers ensure the response is *"properly taking all info into account"* (P177) and check *"whether the AI added irrelevant content and if it changed up my main point of the letter"* (P144). Artefacts such as program code can be tested for quality using other software tools such as compilers, or runtime environments such as browsers. For example, P308 asked Claude to write code for her web application, and had *"to make sure it runs without error and then observed how it functioned."*

Ensure quality through subjective standards (77/319). Knowledge workers also evaluate GenAI output through response-specific subjective quality standards, some of which reflect what Paul and Elder [103] refer to as "intellectual standards" in thinking. Some participants evaluated the real-world **feasibility** of any suggestions. For example, when P297 looked into her social service work for people with mental health disorders and learning disabilities, she had to *"really think about whether the answer the GenAI tool gave me would be easily transferrable to real life situations in social care... not every company has the budget and necessary equipment to provide this most of the times."* Others evaluated the internal **logic** of the AI response. For instance, when a forex and commodities trader (P10) used ChatGPT to *"generate recommendations for new resources and strategies to explore to hone my trading skills, I evaluated whether the stated ideas flowed logically."* Participants also evaluated the **relevance** of the AI response, to see how well it matches *"with my presentation on Kaizen methods on performance management"* (P188) or whether it is appropriately *"in a manner that address the needs of the target job role and attract attention of the recruiter"* (P123).

Verify information by assessing referenced sources (23/319). Participants were generally aware of the issues of hallucination in GenAI, and manually **verify sources** that are directly referenced in GenAI output to ensure they are real and reputable. This is especially true when users request high stakes information, such as advice for medical symptoms (e.g., P5), or the references need to be verifiable for the task to progress, e.g., in P213's job search: *"I was looking for a full-stack role and there was no such role at the company [websites] the GenAI listed"*.

Verify information by cross-referencing external sources (114/319). More commonly, knowledge workers cross-referenced information in the GenAI output against reputable, external sources, to validate it. For tasks within their domain knowledge, our participants relied

[4]We acknowledge that this is a necessary oversimplification, and there are degrees of subjectivity in every criterion.

on their **own knowledge** to identify biases and limitations of the AI response, as noted by P133: *"the AI may suggest repertoire [for the concert I direct], but it sometimes is very American-centric. I often have to use my judgment to come up with a repertoire that fits my reality."* For responses involving technical and professional details, participants cross-referenced **technical or formal documentations** such as official manuals, guidelines, and reports to verify the reliability of the responses. For example, a nurse (P250) verified a ChatGPT-generated educational pamphlet for newly diagnosed diabetic patients by cross-checking with the diabetes management guidelines from her hospital. Similarly, participants verified AI responses with a more general **web search** for information accessible from online forums (e.g., Quora, YouTube, Wikipedia) and other websites. While less common, participants also shared other external sources for cross-referencing, such as responses of other GenAI tools, other task-specialised tools (e.g., language translation), and consulting human domain experts.

4.1.3 Integrate response. Prior work has suggested GenAI requires knowledge workers to perform "critical integration" [114]: the work of editing and incorporating GenAI output into a broader workflow. Qualitatively, we observe that participants integrate GenAI output to their tasks in two distinct ways: they focus either on the *content* — selecting and manipulating a part of the output for use — or *form* — modifying style, wording, tone, etc.

Integrate partial response (36/319). GenAI excels at generating large amounts of information that appear relevant, and not all of it is useful. Participants viewed the process of selectively incorporating the relevant parts of GenAI output into their tasks as critical thinking. For example, when P188 used ChatGPT to help her summarise her past work as an auditor for her resume: *"some of the information provided did not particularly relate to my role and even to the country I was working in. So rather than copying over everything, I had to critically evaluate what would apply, the regulations mentioned - do they apply to the country I work in."*

Modify style to be appropriate for the task (45/319). Finally, participants reflected not only on *what* to incorporate from the response, but also *how* to incorporate it. They might add a "personal touch", or adjust the tone to align the response with their intended style. For example, when P210 used ChatGPT to revise his paper abstract, he had to rephrase the output with a scientific tone because *"often the AI writes awful stuff like "our groundbreaking and fundamental analysis shows..." that sounds too emphatic and does not fit the scientific style."* Participants also attempted to make the GenAI output read less "AI-generated" and more personal, as P254 noted: *"I did make sure it [email composed by ChatGPT] read properly and made sense and did sound like an email that I had composed myself and that a colleague would send."*

4.2 When knowledge workers perceive the enaction of critical thinking

Over 936 GenAI tool use examples, participants self-reported having enacted some critical thinking activity (see Section 4.1) for approximately 60% (555 out of 936) of them. Both knowledge workers' task confidence and their tendency to reflect on work are associated with *when* they perceive the enaction of critical thinking during GenAI tool use (see *Perceived Enaction of Critical Thinking* in Table 4). We discuss key findings for each type of factor, in turn.

4.2.1 Task Factors. While prior work suggests that knowledge workers employ task-dependent strategies for GenAI tool use [13], we did not find a main effect on perceived critical thinking for task type (Creation, Advice, Information). Instead, users' perceptions of confidence — in themselves and in AI doing the task — significantly correlated with their perceived enaction of critical thinking. In line with recent projections that more accessible GenAI tools may exacerbate the risks of technology over-reliance [29, 102, 130], our results provide empirical evidence that knowledge workers' confidence in AI doing the tasks indeed *negatively* correlates with their enaction of critical thinking (β=-0.69, $p < 0.001$). Nevertheless, we also found that knowledge workers' confidence in doing the task themselves (β=0.26, $p = 0.026$) and evaluating AI responses (β=0.31, $p = 0.046$) both *positively* correlate with their enaction of critical thinking. These findings suggest that a reflective approach toward the use of GenAI tools, which can lead to what prior work refers to as "pathways to non-reliance on AI" [20], is more likely to occur when knowledge workers have more confidence in doing the task without AI, or in evaluating AI responses. Our qualitative analysis (see Section 4.3) finds that participants enacted critical thinking when trying to improve the quality and mitigate the negative consequences of AI responses.

4.2.2 User Factors. We also found that knowledge workers' overall tendency to reflect on their work had a positive effect on perceived enaction of critical thinking (β=0.52, $p < 0.001$). This suggests that knowledge workers who already engage in critical thinking in their work are likely to continue doing so even when using GenAI tools. However, in contrast to knowledge workers' confidence in AI doing the task at hand (i.e., *Confidence in AI*, above), which negatively correlated with their perceived enaction of critical thinking, we did not find a significant correlation between knowledge workers' *overall trust in GenAI* and their perceived enaction of critical thinking. A possible explanation is that users' reliance and confidence on AI, as well as their perceived enaction of critical thinking, might vary across tasks; accordingly, the variance that would have been explained by the general user-level factor may already be well captured by the task-level confidence factors.

4.3 Motivators and inhibitors for the perceived enaction of critical thinking

We analysed participants' free-text responses about why they engaged in or prioritised critical thinking (or did not do so) when using GenAI tools for work. We found that enaction of critical thinking was motivated by improvement in work quality, avoidance of negative outcomes, and skill development. We found many inhibitors for the enaction of critical thinking related to awareness (e.g., reliance on AI), motivation (e.g., lack of time), and ability (e.g., barriers to improving GenAI output).

4.3.1 Critical thinking motivators.

Work quality (74/319). As shown in Section 4.1, participants' critical thinking actions were often performed to improve the quality of the work artefact being produced. A key motivator for critical

Table 4: Non-standardised coefficients of the mixed-effects regressions modeling knowledge workers' perceived enaction of critical thinking and perceived effort in cognitive activities when using generative AI tools.

	Perceived Enaction of Critical Thinking (N=930)	Knowledge (N=782)	Comprehension (N=849)	Application (N=768)	Analysis (N=753)	Synthesis (N=825)	Evaluation (N=764)
DV distribution Binary: True/False 5-Likert: -2/-1/0/1/2	551/ 379	236/326/ 191/19/10	333/335/ 139/30/12	227/305/ 201/23/12	219/320/ 184/16/14	267/359/ 156/30/13	177/246/ 221/84/36
(Pseudo) r square/ conditional r square	0.43	0.33	0.32	0.37	0.43	0.38	0.44
Task Factors							
Task type: *Creation*	0 r	0 r	0 r	0 r	0 r	0 r	0 r
Task type: *Advice*	0.12 ($p = 0.829$)	0.04 ($p = 0.816$)	0.00 ($p = 0.994$)	0.06 ($p = 0.713$)	-0.03 ($p = 0.865$)	-0.04 ($p = 0.839$)	-0.18 ($p = 0.127$)
Task type: *Information*	0.32 ($p = 0.364$)	-0.13 ($p = 0.223$)	-0.03 ($p = 0.865$)	0.08 ($p = 0.474$)	-0.14 ($p = 0.127$)	0.01 ($p = 0.967$)	0.14 ($p = 0.248$)
Confidence in self	**0.26*** ($p = 0.026$)	0.02 ($p = 0.713$)	0.02 ($p = 0.779$)	**0.08*** ($p = 0.029$)	0.07 ($p = 0.121$)	0.01 ($p = 0.965$)	**0.10*** ($p = 0.027$)
Confidence in AI	**-0.69***** ($p < 0.001$)	**-0.11*** ($p = 0.029$)	**-0.13*** ($p = 0.014$)	-0.09 ($p = 0.128$)	**-0.15**** ($p = 0.003$)	**-0.12*** ($p = 0.026$)	**-0.23***** ($p < 0.001$)
Confidence in evaluation	**0.31*** ($p = 0.046$)	0.00 ($p = 0.994$)	0.00 ($p = 0.97$)	-0.06 ($p = 0.364$)	0.10 ($p = 0.06$)	-0.03 ($p = 0.795$)	-0.01 ($p = 0.967$)
User Factors							
Gender: *Man*	0 r	0 r	0 r	0 r	0 r	0 r	0 r
Gender: *Woman*	0.33 ($p = 0.38$)	0.03 ($p = 0.865$)	-0.03 ($p = 0.865$)	-0.02 ($p = 0.967$)	-0.15 ($p = 0.248$)	-0.14 ($p = 0.29$)	-0.21 ($p = 0.127$)
Gender: *Non-binary*	1.11 ($p = 0.517$)	0.26 ($p = 0.713$)	0.03 ($p = 0.967$)	0.14 ($p = 0.865$)	-0.51 ($p = 0.338$)	-0.25 ($p = 0.718$)	-0.45 ($p = 0.495$)
Age group: *25-34*	0 r	0 r	0 r	0 r	0 r	0 r	0 r
Age group: *18-24*	0.14 ($p = 0.849$)	0.08 ($p = 0.713$)	0.06 ($p = 0.783$)	0.04 ($p = 0.865$)	0.01 ($p = 0.967$)	0.06 ($p = 0.795$)	-0.01 ($p = 0.967$)
Age group: *35-44*	0.31 ($p = 0.59$)	0.00 ($p = 0.994$)	-0.11 ($p = 0.589$)	-0.01 ($p = 0.967$)	-0.06 ($p = 0.841$)	-0.02 ($p = 0.967$)	-0.05 ($p = 0.865$)
Age group: *45-54*	-0.28 ($p = 0.804$)	0.18 ($p = 0.529$)	0.24 ($p = 0.378$)	0.25 ($p = 0.38$)	0.23 ($p = 0.451$)	0.17 ($p = 0.589$)	0.24 ($p = 0.474$)
Age group: *55+*	-0.96 ($p = 0.474$)	-0.11 ($p = 0.865$)	0.04 ($p = 0.967$)	-0.23 ($p = 0.713$)	-0.25 ($p = 0.713$)	0.04 ($p = 0.967$)	-0.57 ($p = 0.29$)
Occupation's risk of automation: *Low*	0 r	0 r	0 r	0 r	0 r	0 r	0 r
Occupation's risk of automation: *High*	0.17 ($p = 0.74$)	0.04 ($p = 0.829$)	0.10 ($p = 0.451$)	0.15 ($p = 0.248$)	0.03 ($p = 0.865$)	0.19 ($p = 0.116$)	0.16 ($p = 0.29$)
Tendency to reflect	**0.52***** ($p < 0.001$)	-0.01 ($p = 0.967$)	0.06 ($p = 0.378$)	0.05 ($p = 0.511$)	0.01 ($p = 0.967$)	0.06 ($p = 0.392$)	0.05 ($p = 0.59$)
Trust in GenAI	-0.01 ($p = 0.967$)	**-0.12*** ($p = 0.029$)	-0.08 ($p = 0.223$)	**-0.17**** ($p = 0.002$)	**-0.12*** ($p = 0.046$)	-0.05 ($p = 0.499$)	**-0.24***** ($p < 0.001$)

Significance: *p<.05; **p<.01; ***p<.001; r: reference

thinking is to think of ways to improve AI responses. Participants shared several examples of when the AI response fell short of their standards, and motivated critical revision. For instance, when P92 generated content with ChatGPT for his company website: *"the output is way too cookie cutter, full of cliché [text] and boring. I have to edit it a lot to get something out of it that I could ever give to my bosses."* GenAI output can be too shallow and generic for participants' tasks, motivating them to think critically about the depth and specificity of the work. As P133 noted when using ChatGPT to write an executive summary: *"the AI does not understand the niche type of work I do. I have to adapt the output to fit my needs."*

Potential negative outcomes (116/319). Participants shared that their critical thinking was driven by the potential negative outcomes of their use of GenAI. They wished to avoid harm to their work, such as program code that produces wrong outcomes (e.g., P210), outdated information (e.g., P240), or faulty mathematical formulas (P155). This is especially the case when GenAI is applied in high-stakes scenarios and workplaces. For example, P267 used ChatGPT to help her write the pharmacist continuing professional development (CPD) documents, *"the entry is to be submitted for review so I would to double check to be sure otherwise I might have to face suspension."*

Social conflict was another undesirable outcome that motivated critical thinking about GenAI output. For example, P101 reported to a younger supervisor with a different ethnic background. Thus, when preparing work presentations and emails with ChatGPT, he must *"always consider that hierarchy, age, respect for even Chinese festivals, [which] are culturally really important for them."*

Skill development (13/319). Finally, knowledge workers are incentivised to improve skills and learn best practices for their work, even when assisted by GenAI tools. Participants were motivated to enact critical thinking about GenAI output as a means to learn about the task and not simply rely on AI in the long run. For example, when P154 asks ChatGPT for solutions to the issue in a code snippet, *"I make sure that I understood how it works and can do it by myself next time."* Likewise, P176 used ChatGPT to improve an important email draft to sound more professional, and he decided to *"read and break down all the suggested corrections to improve my email writing style"*. This helped improve his writing style, and his later emails *"required less correction."*

4.3.2 Critical thinking inhibitors. In this section, we organised the findings by highlighting the three types of critical thinking barriers introduced by the use of GenAI tools — i.e., awareness, motivation, and ability.

Awareness barriers. Potential downstream harms of GenAI responses can motivate critical thinking (see Section 4.3.1), but only if the user is consciously aware of such harms. Our analysis finds, however, that GenAI tools create obstacles for knowledge workers to be aware of the need for critical thinking, especially when the tasks are perceived to be less important, and when users trust and rely on GenAI tools.

Some participants shared examples in which they thought critical thinking was unnecessary because their **use of GenAI tool is secondary** (14/319) to their goals. P147 used *"Dall-E for indirect purposes (visual reference), [so] there's no need to over-correct what the AI outputs."* Likewise, participants do not enact critical thinking when the **task is perceived to be trivial and insignificant** (55/319), such as writing social media posts (P239) and meeting minutes summary (P271).

Complementing our quantitative findings, knowledge workers' **trust and reliance on GenAI** (83/319) doing the task can discourage them from critically reflecting on their use of the tools. Users often adopt a mental model that assumes AI is competent for simple tasks. This was influenced by users' prior experience with GenAI tools, where the AI had proven trustworthy for specific tasks, as P289 noted: *"With straightforward factual information, ChatGPT usually gives good answers."* For instance, P275 remarked: *"It's a simple task [make a passage professional] and I knew ChatGPT could do it without difficulty, so I just never thought about it, as critical thinking didn't feel relevant."* This mental model, however, can lead to overestimating AI capabilities. Some users, like P185, believed the information provided by GenAI tools was always truthful and of high quality, while others (e.g., P143, P236) assumed the outputs would consistently and accurately reflect referenced data sources. Complementary to the perception of AI as being competent and capable, some participants expressed self-doubt in their ability to perform tasks independently, such as verifying grammar in text (P101) or composing legal letters (P204). This self-doubt led them to accept GenAI outputs by default — a phenomenon corroborated by prior studies [117].

Overreliance on computing technology is not a novel phenomenon; however, GenAI tools can exacerbate the associated risks. Indeed, such reliance may be tolerable for low-stakes tasks, like grammar checking, but it can lead to significant negative outcomes in high-stakes contexts, like drafting legal documents (e.g., [118]). While critical thinking may not be necessary for low-stakes tasks, it is risky for users to only apply critical thinking in high-stakes situations. Without regular practice in common and/or low-stakes scenarios, cognitive abilities can deteriorate over time [5], and thus create risks if high-stakes scenarios are the only opportunities available for exercising such abilities. This phenomenon is well-documented, as in Bainbridge's "Ironies of Automation" [7], and has been recently revisited in the context of GenAI by Simkute et al. [122] as the "Ironies of Generative AI".

Motivation barriers. Knowledge workers also discussed how prioritising critical thinking in their work might be misaligned with their overall task motivations or job objectives. For example, participants discussed a **lack of time** (44/319) for critical thinking at work. For instance, a sales development representative (P295) noted that *"[t]he reason I use AI is because in sales, I must reach a certain quota daily or risk losing my job. Ergo, I use AI to save time and don't have much room to ponder over the result."* Even when time was not constrained, knowledge workers often lacked incentives to engage in critical thinking when it is perceived as **not part of their job responsibilities** (11/319). P232, who used ChatGPT to write the company's marketing campaigns: *"verification and rewriting is handled by another part of the team. The team is able to verify, sense check and modify the content of the landing pages as they see fit."*

Ability barriers. Participants face obstacles to enacting critical thinking, specifically in verifying and improving GenAI output, even if they are otherwise motivated to do so. Participants report **barriers to inspect AI responses** (58/319), such as not possessing enough domain knowledge. As P290 noted: *"in cases where you don't know the specific topic [e.g., translation and math problems], it's hard to determine whether the AI is giving the correct answer or not."*

Even if knowledge workers identify limitations in the GenAI output, they encounter **barriers in revising queries and improving the response** (72/319). For example, P239 received negative feedback from colleagues for a document that ChatGPT helped her write, but *"I'm not sure how I could have improved the text that ChatGPT wrote."* Also, GenAI tools can be "stubborn" and do not follow through with users' revised prompts, as P208 shared when asking GenAI to fix an error in his code: *"it repeatedly recommended the wrong solution despite asking for a different suggestion."*

5 Findings for RQ2: *When and why do knowledge workers perceive increased/decreased effort for critical thinking due to GenAI?*

To answer **RQ2**, we report a descriptive analysis of participants' perceived effort in cognitive activities associated with critical thinking,

as defined by Bloom's taxonomy (Section 5.1) — i.e., recall (Knowledge), organising/translating ideas (Comprehension), problem solving (Application), breaking down a problem (Analysis), putting together ideas (Synthesis), and evaluating and quality checking (Evaluation). We complement this with an analysis of participants' free text elaborations on why they perceived an increase or decrease in effort due to GenAI, observing three qualitative shifts in critical thinking effort (Section 5.2).

A perceived reduction in effort when using GenAI may be due to participants 1) enacting the same "amount" of critical thinking but feeling supported by GenAI, 2) offloading the work of critical thinking to GenAI, 3) enacting "less" critical thinking overall, or 4) conflating reduction in cognitive effort in general, with reduction in critical thinking effort specifically. We address each of these interpretations in context.

5.1 When knowledge workers perceive increased/decreased effort for critical thinking due to GenAI

In the majority of examples, knowledge workers perceive decreased effort for cognitive activities associated with critical thinking when using GenAI compared to not using one — examples that were reported as "much less effort" or "less effort" comprise 72% in Knowledge, 79% in Comprehension, 69% in Application, 72% in Analysis, 76% in Synthesis, and 55% in Evaluation dataset (See Figure 2). Moreover, knowledge workers tend to perceive that GenAI reduces the effort for cognitive activities associated with critical thinking when they have **greater confidence in AI doing the tasks** and possess **higher overall trust in GenAI** (see Table 4).

5.1.1 Task Factors. We found that knowledge workers' *confidence in AI* doing the tasks *negatively* correlated with perceived effort for five of the six cognitive activities (all except Application). The higher the participant's *confidence in AI*, the greater is their perceived reduction in effort for Knowledge (β=-0.11, p = 0.029), Comprehension (β=-0.13, p = 0.014), Analysis (β=-0.15, p = 0.003), Synthesis (β=-0.12, p = 0.026), and Evaluation (β=-0.23, p < 0.001). Moreover, knowledge workers' *confidence in themselves* doing the task correlates positively with perceived effort in Application (β=0.08, p = 0.029) and Evaluation ((β=0.10, p = 0.027). We qualitatively analyse participant rationales in the next section in more detail, but one explanation for why knowledge workers' confidence in AI and in themselves had the opposite effects on perceived effort in these cognitive activities is the following. GenAI tools can decrease knowledge workers' cognitive load by automating a significant portion of their tasks, but as knowledge workers have more confidence in doing the task themselves, they employ more engaged practices in steering AI responses, especially when applying (Application) and evaluating (Evaluation) AI responses.

These findings, along with our quantitative findings for **RQ1**, reveal a connection between knowledge workers' self-confidence and confidence in AI and their perceived critical thinking during GenAI tool use: 1) **a higher confidence in GenAI is associated with less critical thinking even though it is perceived as less effort to do so**, and 2) **a higher self-confidence is associated with more critical thinking even though it is perceived as more effort to do so**. We discuss this in more detail in Section 6.1.1.

5.1.2 User Factors. In contrast to our findings about knowledge workers' perceived enaction of critical thinking (see Section 4.2), we found no significant correlation between their overall tendency to reflect and perceived effort of critical thinking for any cognitive activities. This suggests that knowledge workers who do (or do not) tend to reflect on their work do not necessarily perceive a higher or lower effort of critical thinking with GenAI. However, knowledge workers' *overall trust in GenAI* was negatively correlated with perceived effort for four of the six cognitive activities — i.e., higher trust in the technology is associated with less perceived effort for Knowledge (β=-0.12, p = 0.029), Application (β=-0.17, p = 0.002), Analysis (β=-0.12, p = 0.046), and Evaluation (β=-0.24, p < 0.001). Thus, knowledge workers with higher levels of trust in GenAI — generally or for specific tasks — perceive engaging in critical thinking activities to be less effortful. A possible explanation, supplemented with our qualitative analysis in RQ1 (see Section 4.3.2), is that trust and reliance on GenAI inhibit the enaction of critical thinking, i.e., users underinvest in critical thinking when using GenAI.

5.2 Why knowledge workers perceive increased/decreased effort for critical thinking due to GenAI

To understand *why* participants perceived an increase or decrease in the effort of critical thinking due to GenAI, we analysed the free-text responses in which they were asked to elaborate, mapping the responses onto the six cognitive activities.

We found that GenAI tools shift the effort of critical thinking in three distinct ways: for Knowledge and Comprehension, the effort shifts from information gathering to information verification; for Application, effort shifts from problem-solving to AI response integration; and for Analysis, Synthesis, and Evaluation, effort shifts from task execution to task stewardship.

5.2.1 Knowledge & Comprehension: From information gathering to information verification. Efforts invested in Knowledge (e.g., retrieving relevant information) and Comprehension (understanding that information) often go hand in hand when using GenAI tools. In general, participants perceived less effort in retrieving and curating task-relevant information, because GenAI automates the process. However, they perceived more effort in verifying the information in the AI response.

Participants perceived less effort to **fetch task-specific information at scale, and in real-time** (111/319). For instance, P232 shared that her market research results through ChatGPT *"are immediate and at a sufficient level of detail for me to get to grips with the basics of the industries. I would otherwise have to read a lot of press reports and subscribe to multiple newsletters."*

GenAI tools are perceived to **organise and present information in a readable format** (87/319). For example, P86 compared his experience in searching in a web browser with that in ChatGPT: *" Research using Google is time-consuming; even clicking on a couple of websites takes more time than asking a single question to an LLM. Also, the LLM produces organized answers... the tools*

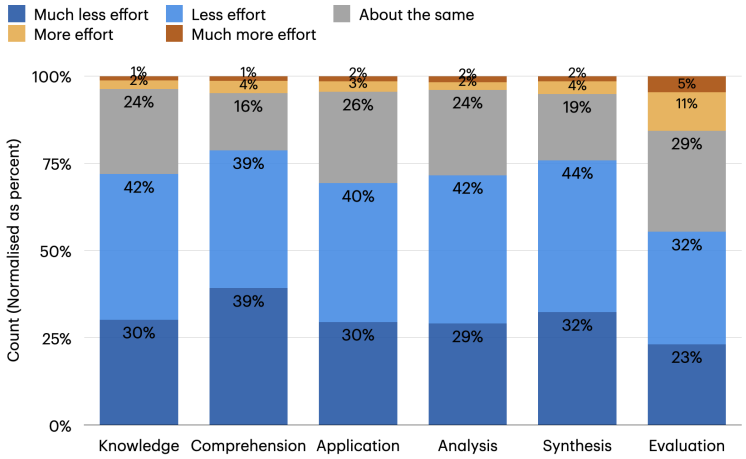


Figure 2: Distribution of perceived effort (%) in cognitive activities (based on Bloom's taxonomy) when using a GenAI tool compared to not using one.

and techniques were categorized by type, and a dotted list was produced for each." Participants find it less effort to re-structure and summarise information in GenAI tools. E.g., P137 tried to update protocol documents to comply with a new standard: *"I did not have to check the templates one by one... Questions I had related to the procedures were answered by the GenAI, and it helped me to know better this new standard."*

However, many participants shared examples when they perceived more effort in information retrieval because the **AI response can be wrong and needs verification** (56/319). For example, when a lawyer (P147) used ChatGPT to find relevant laws for a legal case, he noticed *"AI tends to make up information to agree with whatever points you are trying to make, so it takes valuable time to manually verify."*

5.2.2 Application: From problem-solving to response integration. GenAI can contextually apply knowledge to users' specific questions and examples, reducing perceived effort for Application overall. However, users must instead spend effort integrating GenAI output, in form and content (as mentioned in Section 4.1.3).

Participants perceived less effort in problem-solving and question answering because GenAI tools **provide personalised solutions to their problems** (77/319). For example, P154 compared his experience in reviewing code with and without ChatGPT: *"trying to understand how something works or understanding the problem is the main challenge. People have to "google" a lot. Find the correct information and then try to find people facing similar problems. That takes a lot of effort. GPT simply answers those very fast and easily and mostly correctly."*

With in-context learning, GenAI can also **apply users' examples to new context** (9/319). For example, participants used GenAI tools to generate text, guided with examples: *"company has a set out list of possible scenarios and how we can address them, all I have to do is feed it to the AI, and it would generate a set response based on the data given"* (P268).

Despite the ability for contextual tailoring, participants still reported an increased effort in having to **apply the responses** (19/319) to their tasks and to meet specific needs. For example, when P51 wrote a promotional blog post for their product launch, *"the AI-generated content required substantial editing to align with specific marketing guidelines and tone preferences. This editing process could be time-consuming, particularly when ensuring that technical details were accurate and comprehensible to our target audience."* Additional application effort is incurred when knowledge workers integrate AI-generated content with content from other sources, or misjudge the extent to which GenAI output will be contextualised to their scenario. As P36 noted *"the extra effort in determining that the code generated matched my existing code, and making subsequent alterations to make it fit was more effort than just doing it myself in the end."*

5.2.3 Analysis, Synthesis, and Evaluation: From task execution to task stewardship. Participants perceived these activities, overall, to require less effort due to GenAI tools. Specifically, GenAI helps knowledge workers scaffold complicated tasks and information; it helps knowledge workers automate artefact creation; and it helps form feedback cycles that knowledge workers otherwise do not have access to. Nevertheless, knowledge workers perceived increased effort spent on AI stewardship — translating intentions into queries, steering AI responses, and assessing if the AI response meets their quality standards for work, while retaining accountability for the work.

Analysis. Participants reported reduced effort when GenAI tools helped to **scaffold complicated tasks and information** (48/319). For instance, P203 used ChatGPT to write a complex Slack message to an unfamiliar colleague, and *"GenAI broke down the problem."* This helped her think Analytically, to derive criteria such as to *"make sure the message structure is to the point and understandable to someone who doesn't have the same background knowledge"* as well as *"ensure that I am not missing elements or being confusing with examples."*

However, GenAI tools also require users to **articulate their needs and translate intentions into a query** (45/319), which

was perceived to increase Analysis effort. As mentioned in Section 4.1.1, revising queries is a critical thinking activity specific to GenAI use. P24 described several phases of image generation prompting, saying *"Image generation requires more effort for everything except the actual image generation. I have to think of what I want to be drawn, then on how the AI wants it described, then correct it when it makes wacky outputs."*

Synthesis. Participants perceived less effort when GenAI **automates the creation process** (129/319), such as drafting documents, responding to emails, or generating code.

However, participants noted that the reduced effort in Synthesis could lead to less critical engagement with the task. For instance, P131, when generating advising campaigns for her business, remarked having *"to read what ChatGPT generates and make sure that it's what I want, but not to [let it] think the whole idea."* Moreover, participants perceived it to be more effort to constantly **steer AI responses** (48/319), which incurs additional Synthetic thinking effort due to the cost of developing explicit steering prompts. For example, P110 tried to use Copilot to learn a subject more deeply, but realised: *"its answers are prone to several [diversions] along the way. I need to constantly make sure the AI is following along the correct 'thought process', as inconsistencies evolve and amplify as I keep interacting with the AI."*

Evaluation. Finally, critical thinking is perceived to be less effort because GenAI tools **provide personalised feedback loops for tasks** (40/319) that users otherwise do not have access to. For example, to edit text P313 said he previously *"would often go through multiple rounds of checks by others [humans] for feedback"*, but with GenAI could do so *"on my own time"* by asking the *"AI to do alternate versions, and compare what I like and don't"*.

In certain cases where GenAI is perceived to have a strength relative to the user's own capability (e.g., in spelling or grammar in a non-native language), **GenAI responses are perceived to make few mistakes** (19/319). Thus, participants perceived a reduced effort needed for Evaluation, as P239 noted: *"I can be confident that everything is spelt correctly, I don't need to second guess myself... I can get the reassurance I need without having to bother another person to check it for me."*

Those cases notwithstanding, as noted in Section 4.1.2, participants needed to **evaluate AI-generated content** (42/319) through several objective and subjective criteria, and reported increased effort in doing so.

6 Discussion

6.1 Implications for Designing GenAI Tools That Support Critical Thinking

6.1.1 Self-Confidence and Task Confidence. Task confidence appears to significantly influence knowledge workers' perceived enaction of critical thinking and the effort they invest in it. Specifically, a user's confidence in GenAI is predictive of the extent to which critical thinking is exercised in GenAI-assisted tasks. Both our quantitative and qualitative results suggest that higher confidence in GenAI is associated with less critical thinking, as GenAI tools appear to reduce the perceived effort required for critical thinking tasks among knowledge workers. Conversely, with the important caveat that users' self-confidence is a subjective measure of their knowledge, experiences, and abilities on the tasks [20, 59, 85], higher self-confidence is associated with more critical thinking, even though workers who are confident in their own skills tend to perceive greater effort in these tasks, particularly when evaluating and applying AI responses.

Our analysis does not establish causation. However, based on our evidence, it is possible that fostering workers' domain expertise and associated self-confidence may result in improved critical thinking when using GenAI. Task confidence significantly influences how users engage with AI tools, particularly in the context of human-AI "collaboration" (notwithstanding objections to that term [113]). Previous frameworks have categorised human-AI collaborations by how often the user or the AI initiates an action [95], and which entity takes on a "supervisory" role [88]. Our findings shed light on this issue in the context of GenAI-assisted knowledge work. High task confidence is associated with users' ability to delegate tasks effectively, fostering better stewardship while maintaining accountability. Conversely, lower self-confidence may lead users to rely more on AI, potentially diminishing their critical engagement and independent problem-solving skills. This reliance on AI can be seen as a form of cognitive offloading [8], where users depend on AI to perform tasks they feel less confident in handling themselves.

Confidence in AI is associated with reduced critical thinking effort, while self-confidence is associated with increased critical thinking effort. This duality indicates that design strategies should focus on balancing these aspects. The aims are both to improve the quality of AI-assisted tasks and also to empower users to develop their skills and maintain a balanced "relationship" with AI. To address task confidence recalibration, AI tools could incorporate feedback mechanisms that help users gauge the reliability of AI outputs, when to trust the AI and when to apply their critical thinking skills. This aligns with the goals of explainable AI [33]. Moreover, the user should remain responsible and accountable for the outcome. AI tools must support users in actively and critically customising and refining AI-generated content. Tools may incorporate explicit controls for users to regulate the extent of AI assistance, depending on their confidence levels and the task's complexity.

6.1.2 Awareness, Motivation, and Execution of Critical Thinking. Our study identifies key motivators for and inhibitors of critical thinking among knowledge workers using GenAI. The design implications are clear: critical thinking interventions for GenAI tools should aim to enhance and leverage motivators while mitigating and avoiding inhibitors.

One design approach is to enhance *awareness* of critical thinking opportunities. Our findings indicate that knowledge workers tend to forgo critical thinking for tasks perceived as unimportant or secondary, while engaging in it when aiming to improve task quality or avoid negative outcomes. This suggests a need for both proactive and reactive critical thinking interventions. Proactive systems take the initiative [52] to interrupt the user to highlight the need and opportunity for critical thinking in situations where it is likely to be overlooked; a reactive approach would allow the user to explicitly request critical thinking assistance when it is consciously needed.

Another approach is to increase the *motivation* to think critically. Our study reveals that knowledge workers often neglect critical thinking when they perceive it as outside their job scope, but engage in it when aiming to improve their professional skills. Thus, critical thinking interventions for GenAI tools could be positioned as contributing to long-term skill development and professional growth, as opposed to an extraneous "co-auditing" [46] task that is only relevant on a task-by-task basis.

Finally, design could aim to enhance the *ability* to execute critical thinking. We find that knowledge workers often refrain from critical thinking when they lack the skills to inspect, improve, and guide AI-generated responses. GenAI tools could incorporate features that facilitate user learning, such as providing explanations of AI reasoning, suggesting areas for user refinement, or offering guided critiques. The tool could help develop specific critical thinking skills, such as analysing arguments [72], or cross-referencing facts against authoritative sources. This would align with the motivation-enhancing approach of positioning AI as a partner in skill development.

6.2 Shifts in Critical Thinking Due to Generative AI

Critical thinking in knowledge work involves a range of cognitive activities, such as analysis, synthesis, and evaluation. We observed that the use of GenAI tools shifts the knowledge workers' perceived critical thinking effort in three ways. Specifically, for recall and comprehension, the focus shifts from information gathering to information verification. For application, the emphasis shifts from problem-solving to AI response integration. Lastly, for analysis, synthesis, and evaluation, effort shifts from task execution to task stewardship.

The use of GenAI in knowledge work creates new cognitive tasks for knowledge workers. The task of response integration is a prime example. Knowledge workers must assess AI-generated content to determine its relevance and applicability to their specific tasks, often modifying the style and tone to align with the intended purpose and audience.

Conversely, some cognitive tasks become less necessary due to GenAI. For instance, information gathering has been significantly reduced. GenAI tools automate the process of fetching and curating task-relevant information, making it less effortful for knowledge workers. As a result, the cognitive load associated with searching for and compiling information has decreased.

Some cognitive tasks remain, but have evolved in their nature due to GenAI. One such is information verification; cross-referencing AI-generated outputs with external sources and their own expertise to ensure accuracy and reliability. Workers have always needed to verify the information they work with, but as a tool, GenAI has its own particular strengths and failure modes when it comes to correctness, accuracy, and bias.

With GenAI, knowledge workers also shift from task execution to oversight, requiring them to guide and monitor AI to produce high-quality outputs — a role we describe as "stewardship". It is not that execution has disappeared altogether, nor is having high-level oversight on a task an entirely new cognitive role, but there is a shift from the former to the latter. Unlike in human-human collaboration, in a human-AI "collaboration", the responsibility and accountability for the work still resides with the human user despite the labour of material production being delegated to the GenAI tool, which makes stewardship strike us as a more appropriate metaphor for what the human user is doing, than teammate, collaborator, or supervisor.

In light of these changes, training knowledge workers to think critically when working with GenAI should focus on developing skills in information verification, response integration, and task stewardship. Training programs should emphasise the importance of cross-referencing AI outputs, assessing the relevance and applicability of AI-generated content, and continuously refining and guiding AI processes. Additionally, a focus on maintaining foundational skills in information gathering and problem-solving would help workers avoid becoming overreliant on AI [102].

6.3 Limitations

Our study has limitations that warrant consideration and offer avenues for future research. Firstly, we observed that participants occasionally conflated reduced effort in *using* GenAI with reduced effort in *critical thinking* with GenAI. This misconception may stem from the infrequent contemplation of critical thinking in their daily tasks (regardless of whether they use GenAI), potentially leading to inaccurate self-reporting. This conflation often occurred when participants were satisfied with AI-generated responses, suggesting that when AI produces expected outcomes, users may engage in less critical evaluation. Future studies could employ alternative measures of critical thinking, such as think-aloud protocols or task-based assessments, to better differentiate between effort reduction and critical thinking processes.

Secondly, we assess users' subjective task confidence following prior work on AI-assisted decision-making [20, 59, 85]. Still, one's subjective self-confidence may not always be well-calibrated with respect to objective expertise on tasks [39, 130]. Future work should explore this subjective/objective distinction in the context of critical thinking with GenAI in knowledge work.

Thirdly, our survey was conducted exclusively in English, with participants required to be fluent English speakers. This approach ensured consistency in data collection and feasibility of analysis by our English-speaking research team, but has no representation of non-English speaking populations or multilingual contexts. Future research could explore cross-linguistic and cross-cultural perspectives on GenAI usage and critical thinking.

Fourthly, our sample was biased towards younger, more technologically skilled participants who regularly use GenAI tools at work at least once per week. This demographic skew may not fully represent the broader population of knowledge workers, potentially overlooking the experiences and perceptions of older or less tech-oriented professionals.

Lastly, GenAI tools are constantly evolving, and the ways in which knowledge workers interact with these technologies are likely to change over time. We adopted the task taxonomy due to Brachman et al. [13] to capture relatively stable and coarse-grained characteristics of tasks without overcomplicating our explanatory models. Future work on different goals can expand our measures

with more detailed categorisation and/or task-specific measurements (e.g., task difficulty and skill). To that end, our study provides a valuable baseline for understanding critical thinking in the context of current GenAI tools. In future work, longitudinal studies tracking changes in AI usage patterns and their impact on critical thinking processes would be beneficial. Additionally, developers of GenAI tools can deploy telemetry, within-tool surveys, or experience sampling to their users, to gain more insight into how specific tools can evolve to better support critical thinking in different tasks.

7 Conclusion

We surveyed 319 knowledge workers who use GenAI tools (e.g., ChatGPT, Copilot) at work at least once per week, to model how they enact critical thinking when using GenAI tools, and how GenAI affects their perceived effort of thinking critically. Analysing 936 real-world GenAI tool use examples our participants shared, we find that knowledge workers engage in critical thinking primarily to ensure the quality of their work, e.g. by verifying outputs against external sources. Moreover, while GenAI can improve worker efficiency, it can inhibit critical engagement with work and can potentially lead to long-term overreliance on the tool and diminished skill for independent problem-solving. Higher confidence in GenAI's ability to perform a task is related to less critical thinking effort. When using GenAI tools, the effort invested in critical thinking shifts from information gathering to information verification; from problem-solving to AI response integration; and from task execution to task stewardship. Knowledge workers face new challenges in critical thinking as they incorporate GenAI into their knowledge workflows. To that end, our work suggests that GenAI tools need to be designed to support knowledge workers' critical thinking by addressing their awareness, motivation, and ability barriers.

Acknowledgments

We thank members of the Tools for Thought group at Microsoft Research (https://aka.ms/toolsforthought) and the Calc Intelligence group at Microsoft Research (https://aka.ms/calcintel) for their guidance and discussions throughout our study. We thank our participants for their time, and our reviewers for their helpful feedback.

References

[1] Muhammad Abbas, Farooq Ahmed Jam, and Tariq Iqbal Khan. 2024. Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education* 21, 1 (2024), 10.

[2] Sophie Abel, Kirsty Kitto, Simon Knight, and Simon Buckingham Shum. 2018. Designing personalised, automated feedback to develop students' research writing skills. In *ASCILITE 2018 Conference Proceedings*. University of Technology Sydney.

[3] Satu Alaoutinen and Kari Smolander. 2010. Student self-assessment in a programming course using bloom's revised taxonomy. In *Proceedings of the fifteenth annual conference on Innovation and technology in computer science education*. ACM, Bilkent Ankara Turkey, 155–159. https://doi.org/10.1145/1822090.1822135

[4] Riku Arakawa and Hiromu Yakura. 2024. Coaching Copilot: Blended Form of an LLM-Powered Chatbot and a Human Coach to Effectively Support Self-Reflection for Leadership Growth. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 2, 14 pages. https://doi.org/10.1145/3640794.3665549

[5] Winfred Arthur Jr, Winston Bennett Jr, Pamela L Stanush, and Theresa L McNelly. 1998. Factors that influence skill decay and retention: A quantitative review and analysis. *Human performance* 11, 1 (1998), 57–101.

[6] Florian M Artinger, Gerd Gigerenzer, and Perke Jacobs. 2022. Satisficing: Integrating two traditions. *Journal of Economic Literature* 60, 2 (2022), 598–635.

[7] Lisanne Bainbridge. 1983. Ironies of automation. In *Analysis, design and evaluation of man–machine systems*. Elsevier, 129–135.

[8] Nathaniel Barr, Gordon Pennycook, Jennifer A Stolz, and Jonathan A Fugelsang. 2015. The brain in your pocket: Evidence that Smartphones are used to supplant thinking. *Computers in Human Behavior* 48 (2015), 473–480.

[9] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* 57, 1 (1995), 289–300.

[10] Jeffrey R Binder and Rutvik H Desai. 2011. The neurobiology of semantic memory. *Trends in cognitive sciences* 15, 11 (2011), 527–536.

[11] Jeffrey R Binder, Rutvik H Desai, William W Graves, and Lisa L Conant. 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral cortex* 19, 12 (2009), 2767–2796.

[12] Benjamin S Bloom, Max D Engelhart, Edward J Furst, Walquer H Hill, David R Krathwohl, et al. 1956. *Taxonomy of educational objetives: the classification of educational goals: handbook I: cognitive domain*. Technical Report. New York, US: D. Mckay.

[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. https://doi.org/10.1145/3613905.3650841

[14] Arthur Brookes and Peter Grundy. 1990. *Writing for Study Purposes: A Teacher's Guide to Developing Individual Writing Skills*. Cambridge University Press, 40 West 20th St.

[15] Ann L Brown and Jeanne D Day. 1983. Macrorules for summarizing texts: The development of expertise. *Journal of verbal learning and verbal behavior* 22, 1 (1983), 1–14.

[16] Ann L Brown, Jeanne D Day, and Roberta S Jones. 1983. The development of plans for summarizing texts. *Child development* (1983), 968–979.

[17] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 188 (apr 2021), 21 pages. https://doi.org/10.1145/3449287

[18] Oğuz "Oz" Buruk. 2023. Academic Writing with GPT-3.5: Reflections on Practices, Efficacy and Transparency. *arXiv preprint arXiv:2304.11079* (2023). https://doi.org/10.48550/arXiv.2304.11079

[19] Michael Castelluccio. 2022. IS DIGITAL AMNESIA REAL? *Strategic Finance* 104, 3 (2022), 57–58.

[20] Valerie Chen, Q. Vera Liao, Jennifer Wortman Vaughan, and Gagan Bansal. 2023. Understanding the Role of Human Intuition on Reliance in Human-AI Decision-Making with Explanations. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (Sept. 2023), 1–32. https://doi.org/10.1145/3610219

[21] Ooi Yan Chiew, An Qi Lai, and Wen Xin Liew. 2020. *Digital technology overuse as a predictor of digital amnesia and productivity*. Ph. D. Dissertation. UTAR.

[22] Leah Chong, Guanglu Zhang, Kosa Goucher-Lambert, Kenneth Kotovsky, and Jonathan Cagan. 2022. Human confidence in artificial intelligence and in themselves: The evolution and impact of confidence on adoption of AI advice. *Computers in Human Behavior* 127 (Feb. 2022), 107018. https://doi.org/10.1016/j.chb.2021.107018

[23] Juliet Corbin and Anselm Strauss. 2015. *Basics of qualitative research*. Vol. 14. sage.

[24] Valdemar Danry, Pat Pataranutaporn, Yaoli Mao, and Pattie Maes. 2023. Don't Just Tell Me, Ask Me: AI Systems that Intelligently Frame Explanations as Questions Improve Human Logical Discernment Accuracy over Causal AI explanations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 352, 13 pages. https://doi.org/10.1145/3544548.3580672

[25] Martin Davies. 2011. Concept mapping, mind mapping and argument mapping: what are the differences and do they matter? *Higher education* 62 (2011), 279–301.

[26] John Dewey. 1910. *How We Think*. D.C. Heath & Co., Publishers, Boston.

[27] Amir Dirin, Ari Alamäki, and Jyrki Suomala. 2019. Digital amnesia and personal dependency in smart devices: A challenge for AI. *Proceedings of Fake Intelligence Online Summit 2019* (2019).

[28] Anil R Doshi and Oliver P Hauser. 2024. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances* 10, 28 (2024), eadn5290.

[29] 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*. ACM, Newcastle upon Tyne United Kingdom, 1–21. https://doi.org/10.1145/3663384.3663389

[30] Peter F. Drucker. 1959. *Landmarks of Tomorrow.* Harper.

[31] Wantong Du, Zhiying Zhu, Xinhui Xu, Haoyuan Che, and Shi Chen. 2024. CareerSim: Gamification Design Leveraging LLMs For Career Development Reflection. 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, Article 71, 7 pages. https://doi.org/10.1145/3613905.3650928

[32] Christopher P Dwyer, Michael J Hogan, and Ian Stewart. 2014. An integrated critical thinking framework for the 21st century. *Thinking skills and Creativity* 12 (2014), 43–52.

[33] Upol Ehsan, Philipp Wintersberger, Q Vera Liao, Elizabeth Anne Watkins, Carina Manger, Hal Daumé III, Andreas Riener, and Mark O Riedl. 2022. Human-Centered Explainable AI (HCXAI): beyond opening the black-box of AI. In *CHI conference on human factors in computing systems extended abstracts.* 1–7.

[34] Robert H Ennis. 1993. Critical thinking assessment. *Theory into practice* 32, 3 (1993), 179–186.

[35] Noreen C Facione, Peter A. Facione, and Carol A. Sanchez. 1994. Critical Thinking Disposition as a Measure of Competent Clinical Judgment: The Development of the California Critical Thinking Disposition Inventory. *Journal of Nursing Education* 33, 8 (10 1994), 345–350. https://ezp.lib.cam.ac.uk/login?url=https://www.proquest.com/scholarly-journals/critical-thinking-disposition-as-measure/docview/1026710544/se-2 Copyright - Copyright SLACK INCORPORATED Oct 1994; Last updated - 2023-02-22; CODEN - JNUEAW.

[36] Peter Facione. 1990. Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report). (1990).

[37] Peter A Facione et al. 2011. Critical thinking: What it is and why it counts. *Insight assessment* 1, 1 (2011), 1–23.

[38] Peter A Facione, Carol A Sanchez, Noreen C Facione, and Joanne Gainen. 1995. The disposition toward critical thinking. *The Journal of General Education* 44, 1 (1995), 1–25.

[39] Daniela Fernandes, Steeven Villa, Salla Nicholls, Otso Haavisto, Daniel Buschek, Albrecht Schmidt, Thomas Kosch, Chenxinran Shen, and Robin Welsch. 2024. AI Makes You Smarter, But None The Wiser: The Disconnect Between Performance and Metacognition. https://doi.org/10.48550/arXiv.2409.16708 arXiv:2409.16708.

[40] Mary Forehand. 2010. Bloom's taxonomy. *Emerging perspectives on learning, teaching, and technology* 41, 4 (2010), 47–56.

[41] Ruth Garner and Joseph L McCaleb. 1985. Effects of text manipulations on quality of written summaries. *Contemporary Educational Psychology* 10, 2 (1985), 139–149.

[42] Bhaskar Ghosh, Karthik Narain, Lan Guan, and Jim Wilson. 2023. AI for everyone. https://www.accenture.com/us-en/insights/technology/generative-ai

[43] Ella Glikson and Anita Williams Woolley. 2020. Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals* 14, 2 (July 2020), 627–660. https://doi.org/10.5465/annals.2018.0057

[44] Katrin Glinka and Claudia Müller-Birn. 2023. Critical-Reflective Human-AI Collaboration: Exploring Computational Tools for Art Historical Image Retrieval. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 263 (oct 2023), 33 pages. https://doi.org/10.1145/3610054

[45] Andreas Göldi, Thiemo Wambsganss, Seyed Parsa Neshaei, and Roman Rietsche. 2024. Intelligent Support Engages Writers Through Relevant Cognitive Processes. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24).* Association for Computing Machinery, New York, NY, USA, Article 1047, 12 pages. https://doi.org/10.1145/3613904.3642549

[46] Andrew D. Gordon, Carina Negreanu, José Cambronero, Rasika Chakravarthy, Ian Drosos, Hao Fang, Bhaskar Mitra, Hannah Richardson, Advait Sarkar, Stephanie Simmons, Jack Williams, and Ben Zorn. 2023. Co-audit: tools to help humans double-check AI-generated content. http://arxiv.org/abs/2310.01297 arXiv:2310.01297 [cs].

[47] Chris Greenwood and Matthew Quinn. 2017. Digital amnesia and the future tourist. *Journal of Tourism Futures* 3, 1 (2017), 73–76.

[48] Galen Harrison, Kevin Bryson, Ahmad Emmanuel Balla Bamba, Luca Dovichi, Aleksander Herrmann Binion, Arthur Borem, and Blase Ur. 2024. JupyterLab in Retrograde: Contextual Notifications That Highlight Fairness and Bias Issues for Data Scientists. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24).* Association for Computing Machinery, New York, NY, USA, Article 475, 19 pages. https://doi.org/10.1145/3613904.3642755

[49] David J. Hauser and Norbert Schwarz. 2015. It's a Trap! Instructional Manipulation Checks Prompt Systematic Thinking on "Tricky" Tasks. *Sage Open* 5, 2 (2015), 2158244015584617. https://doi.org/10.1177/2158244015584617 arXiv:https://doi.org/10.1177/2158244015584617

[50] Adrian Holzer, Sten Govaerts, Samuel Bendahan, and Denis Gillet. 2015. Towards Mobile Blended Interaction Fostering Critical Thinking. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Copenhagen, Denmark) *(MobileHCI '15).* Association for Computing Machinery, New York, NY, USA, 735–742. https://doi.org/10.1145/2786567.2793695

[51] Adrian Holzer, Nava Tintarev, Samuel Bendahan, Bruno Kocher, Shane Greenup, and Denis Gillet. 2018. Digitally Scaffolding Debate in the Classroom. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18).* Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188499

[52] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems.* 159–166.

[53] C.W. Howell. 2023. So I followed @GaryMarcus's suggestion and had my undergrad class use ChatGPT for a critical assignment... https://twitter.com/cwhowell123/status/1662501821133254656. Retrieved July 10, 2023.

[54] William Huitt. 2011. Bloom et al.'s taxonomy of the cognitive domain. *Educational psychology interactive* 22 (2011), 1–4.

[55] Jovan Jeromela and Owen Conlan. 2023. Voicing Suggestions and Enabling Reflection: Results of an Expert Discussion on Proactive Assistants for Time Management. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) *(CUI '23).* Association for Computing Machinery, New York, NY, USA, Article 48, 6 pages. https://doi.org/10.1145/3571884.3604317

[56] Sarah A. Jessup, Tamera R. Schneider, Gene M. Alarcon, Tyler J. Ryan, and August Capiola. 2019. The Measurement of the Propensity to Trust Automation. In *Virtual, Augmented and Mixed Reality. Applications and Case Studies*, Jessie Y.C. Chen and Gino Fragomeni (Eds.). Vol. 11575. Springer International Publishing, Cham, 476–489. https://doi.org/10.1007/978-3-030-21565-1_32 Series Title: Lecture Notes in Computer Science.

[57] Jun Li Jeung and Janet Yi-Ching Huang. 2024. Unlocking Memories with AI: Exploring the Role of AI-Generated Cues in Personal Reminiscing. 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, Article 356, 6 pages. https://doi.org/10.1145/3613905.3650979

[58] Hong Jiao and Robert W Lissitz. 2020. *Application of Artificial Intelligence to Assessment.* IAP.

[59] Heather Johnston, Rebecca F. Wells, Elizabeth M. Shanks, Timothy Boey, and Bryony N. Parsons. 2024. Student perspectives on the use of generative artificial intelligence technologies in higher education. *International Journal for Educational Integrity* 20, 1 (Feb. 2024), 2. https://doi.org/10.1007/s40979-024-00149-4

[60] Srecko Joksimovic, Dirk Ifenthaler, Rebecca Marrone, Maarten De Laat, and George Siemens. 2023. Opportunities of artificial intelligence for supporting complex problem-solving: Findings from a scoping review. *Computers and Education: Artificial Intelligence* 4 (2023), 100138.

[61] Sol Kang and William Odom. 2024. On the Design of Quologue: Uncovering Opportunities and Challenges with Generative AI as a Resource for Creating a Self-Morphing E-book Metadata Archive. 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, Article 255, 16 pages. https://doi.org/10.1145/3613905.3650909

[62] Jeffrey D Karpicke and Janell R Blunt. 2011. Retrieval practice produces more learning than elaborative studying with concept mapping. *Science* 331, 6018 (2011), 772–775.

[63] Ronald T Kellogg. 2008. Training writing skills: A cognitive developmental perspective. *Journal of Writing Research* 1, 1 (2008), 1–26. https://doi.org/10.17239/jowr-2008.01.01.1

[64] Ronald T Kellogg and Bascom A Raulerson. 2007. Improving the writing skills of college students. *Psychonomic Bulletin & Review* 14, 2 (2007), 237–242. https://doi.org/10.3758/BF03194058

[65] David Kember, Doris YP Leung, Alice Jones, Alice Yuen Loke, Jan McKay, Kit Sinclair, Harrison Tse, Celia Webb, Frances Kam Yuet Wong, Marian Wong, et al. 2000. Development of a questionnaire to measure the level of reflective thinking. *Assessment & evaluation in higher education* 25, 4 (2000), 381–395.

[66] David Kember, Jan McKay, Kit Sinclair, and Frances Kam Yuet Wong. 2008. A four-category scheme for coding and assessing the level of reflection in written work. *Assessment & evaluation in higher education* 33, 4 (2008), 369–379.

[67] Alison Kidd. 1994. The marks are on the knowledge worker. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 186–191.

[68] Markus Kiefer and Friedemann Pulvermüller. 2012. Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *cortex* 48, 7 (2012), 805–825.

[69] Minyeong Kim, Jiwook Lee, Youngji Koh, Chanhee Lee, Uichin Lee, and Auk Kim. 2024. Interrupting for Microlearning: Understanding Perceptions and Interruptibility of Proactive Conversational Microlearning Services. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24).* Association for Computing Machinery, New York, NY, USA, Article 570, 21 pages. https://doi.org/10.1145/3613904.3642778

[70] Patricia M King. 1997. The reflective judgment model: Transforming assumptions about knowing. *College student development and academic life: psychological, intellectual, social, and moral issues* 4 (1997), 141.

[71] Alexandra Kitson, Petr Slovak, and Alissa N. Antle. 2024. Supporting Cognitive Reappraisal With Digital Technology: A Content Analysis and Scoping Review of Challenges, Interventions, and Future Directions. In *Proceedings of the CHI*

Conference on Human Factors in Computing Systems (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 694, 17 pages. https://doi.org/10.1145/3613904.3642488

[72] Charles W Kneupper. 1978. Teaching argument: An introduction to the Toulmin model. *College Composition and Communication* 29, 3 (1978), 237–241.

[73] Aleksander Kobylarek, Kamil Błaszczyński, Luba Ślósarz, and Martyna Madej. 2022. Critical Thinking Questionnaire (CThQ)–construction and application of critical thinking test tool. *Andragogy Adult Education and Social Marketing* 2, 2 (2022), 1–1.

[74] Deanna Kuhn. 1993. Connecting scientific and informal reasoning. *Merrill-Palmer Quarterly (1982-)* (1993), 74–103.

[75] Soonho Kwon, Dong Whi Yoo, and Younah Kang. 2024. Spiritual AI: Exploring the Possibilities of a Human-AI Interaction Beyond Productive Goals. 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, Article 299, 8 pages. https://doi.org/10.1145/3613905.3650743

[76] Alain Lacroux and Christelle Martin-Lacroux. 2022. Should I Trust the Artificial Intelligence to Recruit? Recruiters' Perceptions and Behavior When Faced With Algorithm-Based Recommendation Systems During Resume Screening. *Frontiers in Psychology* 13 (July 2022). https://doi.org/10.3389/fpsyg.2022.895997 Publisher: Frontiers.

[77] George Lakoff and Mark Johnson. 2008. *Metaphors we live by*. University of Chicago press.

[78] Sunok Lee, Dasom Choi, Minha Lee, Jonghak Choi, and Sangsu Lee. 2023. Fostering Youth's Critical Thinking Competency About AI through Exhibition. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 451, 22 pages. https://doi.org/10.1145/3544548.3581159

[79] Young-Ju Lee. 2020. The Long-Term Effect of Automated Writing Evaluation Feedback on Writing Development. *English Teaching* 75, 1 (2020), 67–92.

[80] Zhuoyang Li, Minhui Liang, Ray Lc, and Yuhan Luo. 2024. StayFocused: Examining the Effects of Reflective Prompts and Chatbot Support on Compulsive Smartphone Use. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 247, 19 pages. https://doi.org/10.1145/3613904.3642479

[81] Zhuoyang Li, Minhui Liang, Hai Trung Le, Ray Lc, and Yuhan Luo. 2023. Exploring Design Opportunities for Reflective Conversational Agents to Reduce Compulsive Smartphone Use. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) *(CUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 37, 6 pages. https://doi.org/10.1145/3571884.3604305

[82] Jingxian Liao, Mrinalini Singh, and Hao-Chuan Wang. 2023. DeepThinkingMap: Collaborative Video Reflection System with Graph-based Summarizing and Commenting. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (Minneapolis, MN, USA) *(CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 369–371. https://doi.org/10.1145/3584931.3607501

[83] Zhuoran Lu and Ming Yin. 2021. Human Reliance on Machine Learning Models When Performance Feedback is Limited: Heuristics and Risks. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–16. https://doi.org/10.1145/3411764.3445562

[84] Tiago Lubiana, Rafael Lopes, Pedro Medeiros, Juan Carlo Silva, Andre Nicolau Aquime Goncalves, Vinicius Maracaja-Coutinho, and Helder I Nakaya. 2023. Ten Quick Tips for Harnessing the Power of ChatGPT/GPT-4 in Computational Biology. *arXiv preprint arXiv:2303.16429* (2023). https://doi.org/10.48550/arXiv.2303.16429

[85] Shuai Ma, Xinru Wang, Ying Lei, Chuhan Shi, Ming Yin, and Xiaojuan Ma. 2024. "Are You Really Sure?" Understanding the Effects of Human Self-Confidence Calibration in AI-Assisted Decision Making. http://arxiv.org/abs/2403.09552 arXiv:2403.09552 [cs].

[86] Jill Burstein McCaffrey, Brian Riordan, and Daniel. 2020. Expanding Automated Writing Evaluation. In *Handbook of Automated Scoring*. Chapman and Hall/CRC.

[87] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 72 (nov 2019), 23 pages. https://doi.org/10.1145/3359174

[88] Nathan J McNeese, Beau G Schelble, Lorenzo Barberis Canonico, and Mustafa Demir. 2021. Who/what is my teammate? Team composition considerations in human–AI teaming. *IEEE Transactions on Human-Machine Systems* 51, 4 (2021), 288–299.

[89] Lucas Memmert and Eva Bittner. 2022. Complex problem solving through human-AI collaboration: literature review on research contexts. (2022).

[90] Lotte Meteyard, Sara Rodriguez Cuadrado, Bahador Bahrami, and Gabriella Vigliocco. 2012. Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex* 48, 7 (2012), 788–804.

[91] Josh Aaron Miller, Kutub Gandhi, Matthew Alexander Whitby, Mehmet Kosa, Seth Cooper, Elisa D. Mekler, and Ioanna Iacovides. 2024. A Design Framework for Reflective Play. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 519, 21 pages. https://doi.org/10.1145/3613904.3642455

[92] Richard L Miller and William Wozniak. 2001. Counter-attitudinal advocacy: Effort vs. self-generation of arguments. *Current Research in Social Psychology* 6, 4 (2001), 46–55.

[93] C Donald Morris, Barry S Stein, and John D Bransford. 1979. Prerequisites for the utilization of knowledge in the recall of prose passages. *Journal of Experimental Psychology: Human Learning and Memory* 5, 3 (1979), 253.

[94] Anwesha Mukherjee, Vagner Figueredo De Santana, and Alexis Baria. 2023. ImpactBot: Chatbot Leveraging Language Models to Automate Feedback and Promote Critical Thinking Around Impact Statements. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 388, 8 pages. https://doi.org/10.1145/3544549.3573844

[95] Michael Muller and Justin Weisz. 2022. Extending a human-ai collaboration framework with dynamism and sociality. In *Proceedings of the 1st Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 1–12.

[96] Jennifer Wilson Mulnix. 2012. Thinking critically about critical thinking. *Educational Philosophy and theory* 44, 5 (2012), 464–479.

[97] Subigya Nepal, Arvind Pillai, William Campbell, Talie Massachi, Eunsol Soul Choi, Xuhai Xu, Joanna Kuc, Jeremy F Huckins, Jason Holden, Colin Depp, Nicholas Jacobson, Mary P Czerwinski, Eric Granholm, and Andrew Campbell. 2024. Contextual AI Journaling: Integrating LLM and Time Series Behavioral Sensing Technology to Promote Self-Reflection and Well-being using the MindScape App. 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, Article 86, 8 pages. https://doi.org/10.1145/3613905.3650767

[98] Quoc Dinh Nguyen, Nicolas Fernandez, Thierry Karsenti, and Bernard Charlin. 2014. What is reflection? A conceptual analysis of major definitions and a proposal of a five-component model. *Medical education* 48, 12 (2014), 1176–1189.

[99] Oda Elise Nordberg and Frode Guribye. 2023. Conversations with the News: Co-speculation into Conversational Interactions with News Content. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) *(CUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 32, 11 pages. https://doi.org/10.1145/3571884.3597123

[100] Shakked Noy and Whitney Zhang. 2023. *Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence*. Technical Report. Working Paper.

[101] Lorena Parra G. and Ximena Calero S. 2019. Automated Writing Evaluation Tools in the Improvement of the Writing Skill. *International Journal of Instruction* 12, 2 (2019), 209–226.

[102] Samir Passi and Mihaela Vorvoreanu. 2022. Overreliance on AI literature review. *Microsoft Research* (2022).

[103] Richard Paul and Linda Elder. 2020. *The miniature guide to critical thinking concepts and tools* (8th edition ed.). Rowman & Littlefield, Lanham, Md. OCLC: on1132213785.

[104] Richard W Paul, Linda Elder, and Ted Bartell. 1997. California teacher preparation for instruction in critical thinking: Research findings and policy recommendations. (1997).

[105] Sheila A. Paul. 2014. Assessment of critical thinking: A Delphi study. *Nurse Education Today* 34, 11 (2014), 1357–1360. https://doi.org/10.1016/j.nedt.2014.03.008

[106] Nikolaos Pellas. 2023. The Effects of Generative AI Platforms on Undergraduates' Narrative Intelligence and Writing Self-Efficacy. *Education Sciences* 13, 11 (2023), 1155.

[107] James Prather, Brent N Reeves, Juho Leinonen, Stephen MacNeil, Arisoa S Randrianasolo, Brett A Becker, Bailey Kimmel, Jared Wright, and Ben Briggs. 2024. The Widening Gap: The Benefits and Harms of Generative AI for Novice Programmers. In *Proceedings of the 2024 ACM Conference on International Computing Education Research-Volume 1*. 469–486.

[108] Sebastian Raisch and Kateryna Fomina. 2023. Combining human and artificial intelligence: Hybrid problem-solving in organizations. *Academy of Management Review* ja (2023), amr–2021.

[109] Leon Reicherts, Gun Woo Park, and Yvonne Rogers. 2022. Extending Chatbots to Probe Users: Enhancing Complex Decision-Making Through Probing Conversations. In *Proceedings of the 4th Conference on Conversational User Interfaces* (<conf-loc>, <city>Glasgow</city>, <country>United Kingdom</country>, </conf-loc>) *(CUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 2, 10 pages. https://doi.org/10.1145/3543829.3543832

[110] Liam Richards Maldonado, Azza Abouzied, and Nancy W. Gleason. 2023. ReaderQuizzer: Augmenting Research Papers with Just-In-Time Learning Questions

to Facilitate Deeper Understanding. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (Minneapolis, MN, USA) *(CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 391–394. https://doi.org/10.1145/3584931.3607494

[111] S James Robert, S Kadhiravan, and Dean McKay. 2024. The development and validation of digital amnesia scale. *Current Psychology* (2024), 1–10.

[112] Christoph Saffer. 2023. Boosting Productivity using GPT-4: Writing Articles and Coding efficiently. https://medium.com/@ChristophSaffer/boosting-productivity-using-gpt-4-writing-articles-and-coding-efficiently-ab0ddb955c2c. Retrieved July 10, 2023.

[113] Advait Sarkar. 2023. Enough With "Human-AI Collaboration". In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 415, 8 pages. https://doi.org/10.1145/3544549.3582735

[114] Advait Sarkar. 2023. Exploring Perspectives on the Impact of Artificial Intelligence on the Creativity of Knowledge Work: Beyond Mechanised Plagiarism and Stochastic Parrots. In *Proceedings of the 2nd Annual Meeting of the Symposium on Human-Computer Interaction for Work* (Oldenburg, Germany) *(CHIWORK '23)*. Association for Computing Machinery, New York, NY, USA, Article 13, 17 pages. https://doi.org/10.1145/3596671.3597650

[115] Advait Sarkar. 2023. Will Code Remain a Relevant User Interface for End-User Programming with Generative AI Models?. In *Proceedings of the 2023 ACM SIGPLAN International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software* (Cascais, Portugal) *(Onward! 2023)*. Association for Computing Machinery, New York, NY, USA, 153–167. https://doi.org/10.1145/3622758.3622882

[116] Advait Sarkar. 2024. AI Should Challenge, Not Obey. *Commun. ACM* 67, 10 (Sept. 2024), 18–21. https://doi.org/10.1145/3649404

[117] Advait Sarkar. 2024. Intention Is All You Need. *Proceedings of the 35th Annual Conference of the Psychology of Programming Interest Group (PPIG 2024)* (2024).

[118] Advait Sarkar. 2024. Large Language Models Cannot Explain Themselves. In *ACM CHI 2024 Workshop on Human-Centered Explainable AI (HCXAI)*.

[119] Advait Sarkar, Xiaotong (Tone) Xu, Neil Toronto, Ian Drosos, and Christian Poelitz. 2024. When Copilot Becomes Autopilot: Generative AI's Critical Risk to Knowledge Work and a Critical Solution. In *Proceedings of the Annual Conference of the European Spreadsheet Risks Interest Group (EuSpRIG 2024)*.

[120] Ulrike Schultze. 2000. A confessional account of an ethnography about knowledge work. *MIS quarterly* (2000), 3–41.

[121] Carol Sherrard. 1986. Summary writing: A topographical study. *Written Communication* 3, 3 (1986), 324–343.

[122] Auste Simkute, Lev Tankelevitch, Viktor Kewenig, Ava Elizabeth Scott, Abigail Sellen, and Sean Rintel. 2024. Ironies of Generative AI: Understanding and Mitigating Productivity Loss in Human-AI Interaction. *International Journal of Human–Computer Interaction* (2024), 1–22.

[123] Jared Spataro. 2023. Introducing Microsoft 365 Copilot – your copilot for work. https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/. Retrieved July 10, 2023.

[124] Arie S Spirgel and Peter F Delaney. 2016. Does writing summaries improve memory for text? *Educational Psychology Review* 28 (2016), 171–196.

[125] Tom Stafford, Herman Elgueta, and Harriet Cameron. 2014. Students' engagement with a collaborative wiki tool predicts enhanced written exam performance. *Research in Learning Technology* 22 (2014).

[126] Na Sun, Chien Wen (Tina) Yuan, Mary Beth Rosson, Yu Wu, and Jack M. Carroll. 2017. Critical Thinking in Collaboration: Talk Less, Perceive More. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 2944–2950. https://doi.org/10.1145/3027063.3053250

[127] Shakti Swaminathan et al. 2020. Digital amnesia: The smart phone and the modern Indian student. *Journal of Humanities and Social Sciences Studies* 2, 3 (2020), 23–31.

[128] Elham Tajik and Fatemeh Tajik. 2023. A comprehensive Examination of the potential application of Chat GPT in Higher Education Institutions. (2023). https://doi.org/10.36227/techrxiv.22589497.v1

[129] Haoheng Tang and Mrinalini Singha. 2024. A Mystery for You: A fact-checking game enhanced by large language models (LLMs) and a tangible interface. 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, Article 631, 5 pages. https://doi.org/10.1145/3613905.3648110

[130] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The Metacognitive Demands and Opportunities of Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 680, 24 pages. https://doi.org/10.1145/3613904.3642902

[131] Thitaree Tanprasert, Sidney S Fels, Luanne Sinnamon, and Dongwook Yoon. 2024. Debate Chatbots to Facilitate Critical Thinking on YouTube: Social Identity and Conversational Style Make A Difference. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 805, 24 pages. https://doi.org/10.1145/3613904.3642513

[132] Jordi Tost, Marcel Gohsen, Britta Schulte, Fidel Thomet, Mattis Kuhn, Johannes Kiesel, Benno Stein, and Eva Hornecker. 2024. Futuring Machines: An Interactive Framework for Participative Futuring Through Human-AI Collaborative Speculative Fiction Writing. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 42, 7 pages. https://doi.org/10.1145/3640794.3665904

[133] Chun-Yen Tsai, Chih-Neng Lin, Wen-Ling Shih, and Pai-Lu Wu. 2015. The effect of online argumentation upon students' pseudoscientific beliefs. *Computers & Education* 80 (2015), 187–197. https://doi.org/10.1016/j.compedu.2014.08.018

[134] David Wade-Stein and Eileen Kintsch. 2004. Summary Street: Interactive computer support for writing. *Cognition and instruction* 22, 3 (2004), 333–362.

[135] Thiemo Wambsganss, Christina Niklaus, Matthias Cetto, Matthias Söllner, Siegfried Handschuh, and Jan Marco Leimeister. 2021. ArgueTutor: An Adaptive Dialog-Based Learning System for Argumentation Skills. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM.

[136] Ge Wang, Jun Zhao, Konrad Kollnig, Adrien Zier, Blanche Duron, Zhilin Zhang, Max Van Kleek, and Nigel Shadbolt. 2024. KOALA Hero Toolkit: A New Approach to Inform Families of Mobile Datafication Risks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 226, 18 pages. https://doi.org/10.1145/3613904.3642283

[137] Yingxu Wang and Vincent Chiew. 2010. On the cognitive process of human problem solving. *Cognitive systems research* 11, 1 (2010), 81–92.

[138] Daniel T. Willingham. 2008. Critical Thinking: Why Is It So Hard to Teach? *Arts Education Policy Review* 109, 4 (2008), 21–32. https://doi.org/10.3200/AEPR.109.4.21-32 arXiv:https://doi.org/10.3200/AEPR.109.4.21-32

[139] Donna Wilson and Marcus Conyers. 2016. *Teaching students to drive their brains: Metacognitive strategies, activities, and lesson ideas*. Ascd.

[140] Peter N Winograd. 1984. Strategic difficulties in summarizing texts. *Reading research quarterly* (1984), 404–425.

[141] ShunYi Yeo, Gionnieve Lim, Jie Gao, Weiyu Zhang, and Simon Tangi Perrault. 2024. Help Me Reflect: Leveraging Self-Reflection Interface Nudges to Enhance Deliberativeness on Online Deliberation Platforms. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 806, 32 pages. https://doi.org/10.1145/3613904.3642530

[142] Kangyu Yuan, Hehai Lin, Shilei Cao, Zhenhui Peng, Qingyu Guo, and Xiaojuan Ma. 2023. CriTrainer: An Adaptive Training Tool for Critical Paper Reading. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 44, 17 pages. https://doi.org/10.1145/3586183.3606816

[143] Liudmila Zavolokina, Kilian Sprenkamp, Zoya Katashinskaya, Daniel Gordon Jones, and Gerhard Schwabe. 2024. Think Fast, Think Slow, Think Critical: Designing an Automated Propaganda Detection Tool. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 491, 24 pages. https://doi.org/10.1145/3613904.3642805

[144] Olaf Zawacki-Richter, Victoria I Marín, Melissa Bond, and Franziska Gouverneur. 2019. Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education* 16, 1 (2019), 39. https://doi.org/10.1186/s41239-019-0171-0

[145] Esperanza Zuriguel-Pérez, Anna Falcó-Pegueroles, Juan Roldán-Merino, Sandra Agustino-Rodriguez, Maria del Carmen Gómez-Martín, and Maria Teresa Lluch-Canut. 2017. Development and psychometric properties of the nursing critical thinking in clinical practice questionnaire. *Worldviews on Evidence-Based Nursing* 14, 4 (2017), 257–264.

[146] Esperanza Zuriguel-Pérez, María-Teresa Lluch-Canut, Montserrat Puig-Llobet, Luis Basco-Prado, Adrià Almazor-Sirvent, Ainoa Biurrun-Garrido, Mariela Patricia Aguayo-González, Olga Mestres-Soler, and Juan Roldán-Merino. 2022. The nursing critical thinking in clinical practice questionnaire for nursing students: A psychometric evaluation study. *Nurse Education in Practice* 65 (2022), 103498.

A Appendix

A.1 Survey Questions

A.1.1 Task Questions.

1. Please share one more specific real-world example of the way you used GenAI tool while doing work. Please tell us: 1) what you were trying to achieve, 2) in what GenAI tool, and

3) how you used the GenAI tool, including any prompts (it may help to look at your GenAI tool chat history, or if you cannot recall the exact prompts you used, please include a rough equivalent). [Free response]

2. For the specific example you share, what best describes this task?
   (a) Generate something (e.g., text, Python code, or image) to be used directly.
   (b) Generate something (e.g., text, Python code, or image) to be used with some modification.
   (c) Generate an idea to be used indirectly (e.g., use a chatbot to generate product ideas to help you think, but you won't use the text in a document).
   (d) Seek a fact or piece of information (e.g., find specific instructions for a tool, or search my document for relevant passages).
   (e) Learn about a new topic more broadly (e.g., how can I get a job as a software engineer).
   (f) Generate a shorter version of a piece of content that describes the important elements (e.g., summarise text from external websites).
   (g) Discover a new insight about information or data (e.g., analyse a spreadsheet or CSV file for business insights).
   (h) Generate a better version (e.g., re-write text that was too long or complex).
   (i) Get guidance about how to make a decision (e.g., try to figure out the ideal amount of time a project should take).
   (j) Check whether an artefact satisfies a set of rules, constraints, quality checks, or formatting requirements (e.g., document checking to ensure all required elements are included).
   (k) Other: [Free response]
3. Did GenAI make your work easier or harder? When you used a GenAI tool for this task, did you have to put in more effort or less effort for the following activities you may have performed during the task, compared to when you did not use GenAI? (1: Much less effort; 5: Much more effort; N/A: this activity is not relevant to the task):
   (a) Recall: Recognizing or remembering facts, terms, basic concepts, or answers
   (b) Organising/translating ideas: Organizing, summarising, translating, generalising, giving descriptions, and stating the main ideas
   (c) Problem solving: Using acquired knowledge to solve problems in new situations
   (d) Breaking down a problem: Examining and breaking information into component parts, determining how the parts relate to one another, identifying motives or causes, making inferences, and finding evidence to support generalisations
   (e) Putting together ideas: Building a structure or pattern from diverse elements; putting parts together to form a whole or bringing pieces of information together to form a new meaning
   (f) Evaluating and quality checking: Presenting and defending opinions by making judgments about information, the validity of ideas, or quality of work based on a set of criteria
4. If you selected "more effort" or "much more effort" for any of the activities above, please explain why those activities require more effort with GenAI, compared to when you did not use GenAI. [Free response]
5. If you selected "less effort" or "much less effort" for any of the activities above, please explain why those activities require less effort with GenAI, compared to when you did not use GenAI. [Free response]
6. Have you ever done any reflective/critical thinking (e.g., reflect on your use and the outputs you got from LLM tools) when doing this task with GenAI tool? [Yes/No]
7. *(If selected Yes in Q6)*
   (a) What type of reflective/critical thinking tactic(s) did you do to for this task in GenAI? (select all that apply)
      (i) Reflecting on facts or basic concepts, and cross-check AI output with other sources. (Example: After the AI generates a summary of a historical event, you verify the dates and key figures by looking them up on trusted websites or in textbooks.)
      (ii) Reflecting on organisation, summarization, and generalisation. Consider whether the AI output is well structured and formatted, whether it is too long/short, etc. (Example: You receive a report from the AI and check if the sections are clearly divided, headings are properly used, and the summary accurately reflects the main points without missing any critical information.)
      (iii) Reflecting on how knowledge is applied, such as considering whether AI correctly understood and applied any high-level concepts in your work, and reflecting on your own application of knowledge. (Example: When the AI writes a technical explanation, you review it to ensure that it correctly applies industry-specific terminology and concepts, such as proper use of scientific methods or legal principles.)
      (iv) Reflecting on individual elements and their relationship. Thinking about whether the AI output flows logically, whether different claims are coherent with each other, etc. (Example: The AI creates a persuasive essay, and you evaluate whether each argument builds logically on the previous one and whether there are any contradictions or gaps in the reasoning.)
      (v) Reflecting on how ideas are combined to form new meaning. (Example: The AI proposes a new business strategy by combining market analysis, customer feedback, and competitor data. You assess whether these elements are integrated in a way that offers a novel and feasible approach.)
      (vi) Reflecting on the quality of the work, such as making sure the work meets objective standards and expectations in your workplace, and also deciding what quality standards matter and when to apply them.(Example: You review an AI-generated project proposal to ensure it meets your company's standards for clarity, thoroughness, and professionalism, and aligns with the objectives of the task.)

(vii) Other: [Free response]
(b) Please share one real-world example when you applied the critical thinking tactic(s) to this task, and explain why you did critical thinking. [Free response]
(c) When applying this critical thinking tactic during your use of GenAI tool, have you ever encountered any challenges and obstacles? [Free response]
(d) How did you learn to apply critical/reflective thinking when using GenAI? (select all that apply)
(i) Informally, at school or university (e.g., learnt from peers, or picked it up over time)
(ii) Through formal training at school or university (e.g., took a course)
(iii) Informally, at my workplace (e.g., learnt from colleagues, or picked it up over time)
(iv) Through formal training after school or university (e.g., took a professional development seminar)
(v) Other: [Free response]
8. *(If selected No in Q6)*
(a) What prevented you from applying critical thinking strategies when doing this task with GenAI? (select all that apply)
(i) Not enough time in my schedule
(ii) Not prioritised by management
(iii) Not sure how to verify information
(iv) Not sure how to improve AI suggestions quality
(v) It didn't occur to me
(vi) The task doesn't require critical thinking
(vii) Other: [Free response]
(b) Please tell us why you chose the answer(s) above: [Free response]
9. Why do you use GenAI for this task? (select all that apply)
(a) It helps me save time
(b) It helps me do the work better
(c) It helps me make progress when I am stuck
(d) It helps me be more creative and get more ideas
(e) It helps me do things that I don't have the expertise to do myself
(f) Other reason: [Free response]
10. Would you like GenAI to automate this task entirely? [Yes/No]
11. Your confidence in doing the task with and without GenAI (1: Not at all confident; 5: Extremely confident):
(a) How confident are you in your ability to do this task without GenAI?
(b) How confident are you in the ability of GenAI to do this task?
(c) How confident are you, in the course of your normal work (e.g., accounting for time and resource demands of your task), in evaluating the output that AI produces for this task?

A.1.2 User Questions.

1. What GenAI tools do you use in your work? (check all that apply)
(a) ChatGPT
(b) Gemini website
(c) Gemini in Google products such as Gmail, Google Slides
(d) Copilot website (formally known as Bing Chat)
(e) Microsoft 365 Copilot (embedded with Office apps such as Word)
(f) Claude.ai
(g) DeepAI AI Chat
(h) Pi.ai
(i) Perplexity.ai
(j) Dall-E
(k) Stable Diffusion
(l) Midjourney
(m) Other: [Free response]
2. What is your age range?
(a) 18-24
(b) 25-34
(c) 35-44
(d) 45-54
(e) 55 or over
(f) Prefer not to say
3. What is your gender identity?
(a) Man
(b) Woman
(c) Non-binary/gender diverse
(d) Prefer not to say
4. What is currently your primary country of residence? [Free response]
5. What is your job title? [Free response]
6. Which of these best describes your work?
(a) Military
(b) Transportation and Material Moving
(c) Production
(d) Installation, Maintenance, and Repair
(e) Construction and Extraction
(f) Farming, Fishing, and Forestry
(g) Office and Administrative Support
(h) Sales and Related
(i) Personal Care and Service
(j) Building and Grounds Cleaning and Maintenance
(k) Food Preparation and Serving Related
(l) Protective Service
(m) Healthcare Support
(n) Healthcare Practitioners and Technical
(o) Arts, Design, Entertainment, Sports, and Media
(p) Educational Instruction and Library
(q) Legal
(r) Community and Social Service
(s) Life, Physical, and Social Science
(t) Architecture and Engineering
(u) Computer and Mathematical
(v) Business and Financial Operations
(w) Management
(x) Other: [Free response]
7. To what extent do you agree with the following statements, regarding your daily work? (1: Strongly disagree; 5: Strongly agree)
(a) I sometimes question the way others (e.g., your colleagues) do something and try to think of a better way.

(b) I like to think over what I have been doing and consider alternative ways of doing it.
(c) I often reflect on my actions to see whether I could have improved on what I did.
(d) I often re-appraise my experience so I can learn from it and improve for my next performance.

8. For the below listed items, please read each statement carefully. Using the 5-point scale ranging from 1 (Strongly disagree) to 5 (Strongly agree), select the answer that most accurately describes your feelings.

(a) Generally, I trust GenAI.
(b) GenAI helps me solve many problems.
(c) I think it's a good idea to rely on GenAI for help.
(d) I don't trust the information I get from GenAI.
(e) GenAI is reliable.
(f) I rely on GenAI.

Table 5: Codebook for the qualitative analysis.

RQ1: How do knowledge workers perceive the enaction of critical thinking when using GenAI?	
Goal and query formation	**Critical thinking motivators**
Form goal	Work quality
Form query	Potential negative outcomes
	Skill development
Inspect response	
Ensure quality through objective criteria	**Critical thinking inhibitors**
Ensure quality through subjective standards	Awareness barriers
Verify information by assessing referenced sources	*- use of GenAI tool is secondary*
Verify information by cross-referencing external sources	*- task is perceived to be trivial and insignificant*
	- trust and reliance on GenAI
Integrate response	Motivation barriers
Integrate partial response	*- lack of time*
Modify style to be appropriate for the task	*- not part of their job responsibilities*
	Ability barriers
	- barriers to inspect AI responses
	- barriers in revising queries and improving the response

RQ2: Why do knowledge workers perceive increased/decreased effort for critical thinking due to GenAI?	
Knowledge & Comprehension	**Analysis, Synthesis, and Evaluation**
Fetch task-specific information at scale, in real-time	Scaffold complicated tasks and information
Organise and present information in a readable format	Automate the creation process
AI response can be wrong and needs verification	Provide personalised feedback loops for tasks
	GenAI responses are perceived to make few mistakes and easy to review
Application	Users need to articulate the need and translate intentions into a query
Provide personalised solutions to their problems	Users need to steer AI responses
Apply users' examples to new context	Users need to evaluate AI-generated content
Users need to apply the responses to their tasks	