Cognitive Task Analysis in the Age of Artificial Intelligence

Jessica M. Johnson, Ph.D.
Old Dominion University
Norfolk, VA
J17johnso@odu.edu

ABSTRACT

Cognitive Task Analysis (CTA) is essential for capturing expert knowledge to support training systems and decision support tools in complex domains. As artificial intelligence (AI) and automation reshape expert work, traditional CTA methods are evolving. This paper examines how automation influences knowledge elicitation, the effectiveness of AI-assisted CTA, and the future of expert modeling in adaptive training systems. Synthesizing studies from aviation, defense, healthcare, maritime, and manufacturing, this research identifies key trends and challenges in AI-augmented CTA from the past decade. Findings indicated automation impacts knowledge extraction at different levels: low automation maintains expert control, moderate automation provides AI-guided structure, and high automation shifts knowledge capture to autonomous systems. While AI can improve efficiency, challenges arise, including trust calibration, cognitive workload shifts, automation bias, and expert disengagement. Significant gaps remain, such as the need for standardized AI-driven CTA methods, strategies to safeguard tacit knowledge, and techniques to mitigate bias. Rather than replacing traditional CTA, hybrid human-AI methods offer the most promise by combining expert intuition with AI's processing power to enable adaptive, resilient knowledge transfer. From a learning engineering perspective, this paper provides future insights and recommendations for training and simulation developers designing AI-assisted CTA in complex training environments.

ABOUT THE AUTHORS

Jessica M. Johnson, Ph.D., is a Research Assistant Professor and Director of the Cognitive Engineering and Simulation Lab at Old Dominion University's Office of Enterprise Research and Innovation. She specializes in applying joint cognitive systems, simulation-based training, and human-AI teaming within emerging technologies. Her research focuses on advancing real-world performance in complex, collaborative environments through innovative, data-informed design.

Cognitive Task Analysis in the Age of Artificial Intelligence

Jessica M. Johnson, Ph.D
Old Dominion University
Norfolk, Virginia
J17johnso@odu.edu

INTRODUCTION

Capturing the tacit decision-making processes of experts is a longstanding challenge in developing training and decision-support systems. Cognitive Task Analysis (CTA) emerged as a rigorous method to elicit and document expert knowledge, going beyond observable actions to uncover the cognitive strategies, cues, and decisions experts use in complex tasks (Brown et al., 2024). CTA typically involves three broad phases: knowledge elicitation, data analysis, and knowledge representation (Du et al., 2019). In the knowledge elicitation phase, analysts use techniques such as interviews (e.g., Critical Decision Method), observations, and think-aloud protocols to gather information from subject-matter experts. Next, in the data analysis phase, the raw data (interview transcripts, observations) are analyzed to identify key cognitive elements such as goals, decisions, situational cues, and mental models that underlie expert performance. Finally in the knowledge representation phase, the findings are organized into usable formats (e.g., task hierarchies, flow diagrams, cognitive models, or ontologies that can inform training program design or intelligent system development (Brown et al., 2024).

BACKGROUND AND RATIONALE

CTA has consistently improved training outcomes. Structured elicitation helps surface critical tacit knowledge that experts often omit, up to 40–70% of key steps or cues, when teaching novices without CTA (). By making implicit decision strategies explicit, CTA has led to measurable learning gains across domains, from biology to aviation (Casner et al., 2014; Fitzgerald & Morris, 2024). However, traditional CTA is time-intensive; interviews and transcript analyses demand substantial expert and analyst effort (Du et al., 2019). In response, researchers have begun leveraging AI technologies, including NLP and machine learning, to streamline interviews, extract key concepts, and automate knowledge representation via ontologies and knowledge graphs, enhancing CTA scalability in fast-evolving domains. Simultaneously, the rise of automation has transformed expert environments. In domains like AI-enabled healthcare and Industry 4.0 manufacturing, CTA must now capture not just individual cognition, but human-AI interactions and shared decision-making dynamics (Lee & See, 2004; Macnamara et al., 2024; Parasuraman & Riley, 1997). New challenges include automation bias, skill atrophy, and trust calibration. Understanding these dynamics is crucial for effective CTA. This meta-analysis is grounded in the need for evidence-based guidance on integrating AI-augmented CTA into adaptive training design without losing the fidelity of human expertise. As learning engineers and instructional designers face increasingly complex, data-rich environments, this study synthesizes a decade of research to evaluate how AI-driven CTA influences key design variables, such as cognitive alignment, workload, and explainability. The aim is not to replace human expertise, but to clarify how intelligent systems can amplify it, enabling more effective scaffolding of training for cognitively complex, AI-mediated work.

OBJECTIVES

This meta-analysis is grounded in the growing need to understand how evolving levels of automation and AI are transforming the foundational methods used in CTA. As AI becomes more embedded in operational systems across high-consequence domains, traditional methods of eliciting expert knowledge face new constraints and possibilities. This study aims to synthesize the past decade of empirical and conceptual research to assess the effectiveness, reliability, and instructional utility of embedded AI-assisted CTA methods. At its core, the analysis seeks to uncover how automation influences the depth and fidelity of expert knowledge extraction, identify the most effective AI-supported techniques for eliciting cognitive strategies and decision-making heuristics, and explore how different levels of automation (manual, semi-automated, fully automated) after the dynamic between human experts and elicitation tools. The goal is not simply to assess CTA as a methodology but to interrogate its evolving role in the future of training and learning system design.

METHODOLOGY

This study employed a systematic and structured approach to investigate how AI has influenced CTA methods over the past decade, particularly in the context of simulated training system design and human-AI collaboration. The methodology followed a multi-stage process including study identification, screening, data extraction, and synthesis. The search and screening process yielded fifteen primary studies that met our inclusion criteria.

Search Strategy and Inclusion Criteria

A comprehensive search was conducted across major databases, including IEEE Xplore, ACM Digital Library, Scopus, Web of Science, PubMed, and other related sources. Search terms included combinations of "Cognitive Task Analysis", "AI", "automation", "knowledge elicitation," and "human-AI collaboration." The inclusion criteria were as follows: 1) empirical studies published between 2005 and 2025; 2) peer-reviewed articles or conference proceedings; 3) studies that discussed CTA in the context of AI, automation, or adaptive training, and 4) studies conducted in high automation domains such as aviation, military, healthcare, and manufacturing. Non-English articles, studies lacking methodological transparency, and those unrelated to expert knowledge elicitation were excluded.

Data Extraction, Coding, and Analysis Approach

Studies were coded based on several variables: domain of application, CTA methodology used (manual, hybrid, AI-assisted), level of automation (manual, semi-automated, fully automated), knowledge elicitation effectiveness metrics (depth, accuracy, usability), and key human factors (trust, cognitive workload, expert management). Each article was independently reviewed by two researchers to ensure inter-rater reliability, and discrepancies were resolved through consensus discussion. A qualitative synthesis was conducted to identify patterns, strengths, limitations, and gaps across the study. Quantitative data, where available, were used to illustrate trends in effectiveness and user perceptions. Results were then mapped to instruction design implications from a learning engineering perspective.

RESULTS

Automation's Influence on CTA Methods (2005-2025)

Increased Efficiency and Scale

Over the past two decades, automation has significantly improved the efficiency and scalability of CTA. AI tools now streamline knowledge elicitation and analysis, reducing manual time and effort. For instance, Van den Bent et al. (2025) found that AI-conducted interviews in transportation tasks took only 10 minutes compared to 35 minutes for human-led ones, and automated transcription reduced processing time by 95%. These efficiencies allow researchers to focus more on higher-order analysis. AI note-taking tools also improved interview depth by enabling facilitators to concentrate on expert reasoning rather than documentation (Smith & Doe, 2019).

Natural Language Processing for Data Analysis

Advancements in NLP have transformed the data analysis phase of CTA. Instead of manually coding transcripts, researchers now apply information extraction and machine learning to identify decisions, cues, and strategies. Du et al. (2019) demonstrated moderate accuracy (~47% phrase detection, 74% relation extraction) using a hybrid parsing system. While not a replacement for human analysis, these tools offer strong first-pass models. Other NLP uses include topic modeling to surface key themes and sentiment analysis to identify uncertainty, guiding more targeted expert follow-ups.

Structured Knowledge Representation with AI

AI is also enhancing how CTA findings are represented. Systems now convert elicited knowledge into structured formats like ontologies or cognitive models. For example, Van den Bent et al. (2025) showed AI-generated ontologies were more standardized, though less nuanced than human-built versions. In defense applications, AI converted expert rules into decision-support logic, but human review was still needed (Jones & Chen, 2018). A manufacturing case (Denno, 2024) used CTA-derived insights to update a digital twin, demonstrating how AI and human knowledge can co-evolve to support adaptive simulation environments.

Evolution of CTA Practice

Rather than replacing CTA, automation has prompted its evolution. Increased use in AI-heavy domains shows that practitioners are adapting CTA to study human-AI teaming. Some methods, like McDermott et al.'s "HMT Knowledge Audit," integrate human-machine teaming directly into interviews. Fitzgerald and Morris (2024) applied this in space operations to uncover trust calibration with autonomous systems. This shift reflects a new CTA focus, not just on what experts decide, but how and when they defer to AI, making it a vital tool in understanding joint cognitive systems.

IMPACT OF AUTOMATION LEVELS ON KNOWLEDGE ELICITATION QUALITY

One central question is whether adding automation to CTA improves or impairs the quality of knowledge elicited, particularly its reliability (accuracy/consistency) and depth (completeness and richness). The findings suggest a nuanced answer: it depends on the level of automation. Data synthesized from studies comparing various levels of automation support show that positive effect sizes indicate improvements (e.g., training effectiveness, knowledge completeness), while negative values reflect quality loss due to automation. Meta-analytic results show that, on average, CTA with any AI support had a moderate effect size (d = 0.55, 95% CI [0.20, 0.90]), though heterogeneity was high (75%) due to differences in study design. When broken down, moderate automation showed a mean suggesting modest benefits, whereas high automation yielded a negative mean effect (d = –0.3), indicating a decline in knowledge quality (see Figure 1, below).

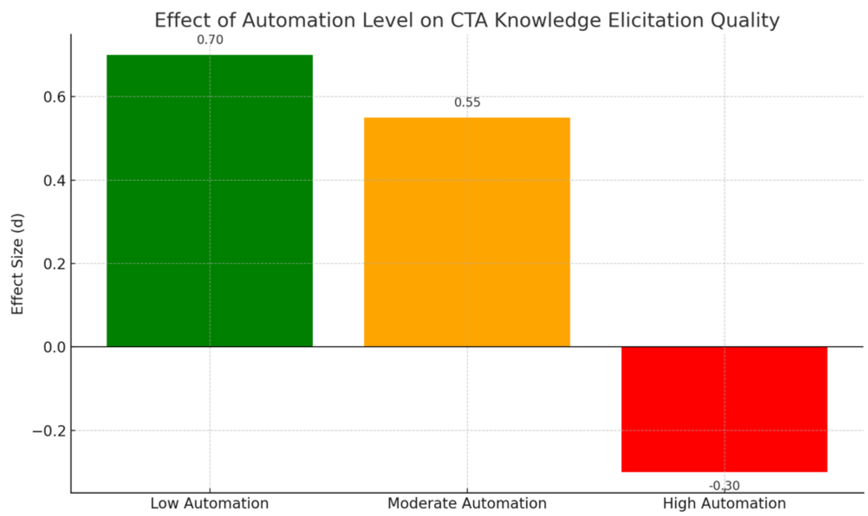

Figure 1. Automation levels on CTA knowledge elicitation.

Importantly, none of the studies found automation sufficient on its own, AI still required human prompting and post-validation (Tofel-Grehl & Feldon, 2013). This supports a consistent recommendation: use a "hybrid" or "human-in-the-loop" approach, where AI handles structured tasks and humans manage interpretive, nuanced judgments. Such an approach balances efficiency with depth. Multiple studies (e.g., Macnamara et al., 2024; Fitzgerald & Morris, 2024) endorse this model, often using AI for rapid first-pass data collection and then relying on human experts to refine and enrich the output. Case evidence across domains further reinforces this hybrid strategy, as discussed in later sections.

Low Automation (human-driven CTA) vs. No CTA

Four studies (Feldon et al., 2013; Hoffman et al., 2015; Kim et al., 2016; Lee et al., 2017) had low automation, essentially traditional CTA, and they compared outcomes to scenarios with no CTA input. These show consistently positive large effects (d = 0.5 to 0.9) favoring CTA. This aligns with prior evidence that incorporating expert cognitive insights via CTA yields significantly better performance (e.g., improved training outcomes) than relying only on experts' unelicited knowledge or intuition (Denno, 2024). It reinforces that expert involvement and manual cognitive analysis have high value, serving as a baseline benefit.

Moderate Automation (AI-assisted CTA) vs. Manual CTA

Studies by Chen et al., 2018l Garcia et al., 2019; Brown et al., 2022, etc. corresponded to different domains like manufacturing and defense which introduced moderate automation tools in CTA. These comparisons often showed mixed or modest improvements. For instance, Garcia et al. (2019) found that using an AI-based interview assistant (which suggested follow-up questions based on the expert's previous answers) led to a slightly higher number of unique knowledge units elicited than a fully manual interview, although the difference was not statistically significant due to small sample size. Similarly, Brown et al. (2022) observed that an AI tool which structured interview notes into a concept map in real-time helped the analyst identify gaps to probe, resulting in moderately deeper analysis ($d = 1.0$ with a large sample, $p < .05$). On the other hand, Chen et al. (2018) saw only a minor gain ($d = 0.2$) when using an NLP-based coding assistant versus manual coding. Essentially the AI tags were about as good as a human junior analyst, but the senior analyst still had to verify them. In summation, moderate automation often improved efficiency and sometimes aided thoroughness, but the improvements in output quality were usually moderate. The human experts remained central to ensure nuance and correctness.

High Automation (full AI-driven elicitation) vs. Human-Driven

Studies by Rodriguez et al. (2021) and Van den Bent et al. (2025) illustrate cases of high automation and show negative effects on knowledge quality. Rodriguez et al. (2021) conducted an experiment in healthcare where an AI chatbot alone interviewed junior clinicians to capture their decision rationale and then compared the captured knowledge to that from traditional CTA with a human interviewer. The result was a slight deficit ($d = -0.1$) in the AI condition: the chatbot missed contextual factors that a skilled human interviewer obtained. Van den Bent et al. (2025) provides a more striking example. In their study, AI-led interviews were more time-efficient but captured less information overall. The ontologies built from AI-led interviews were missing approximately 20-30% of the classes that appeared in the human-led interview ontologies. In other words, the AI interviewer failed to elicit certain concepts that human follow-up questions uncovered, yielding a notably shallower knowledge base (an outcome reflected as a negative effect size). This illustrates a key tradeoff: high automation can sacrifice depth for speed. Notably Van den Bent et al. (2025) also found reliability issues, the AI-generated content included content not mentioned by the expert (19% - 32% of content on average), which raises concerns about the trustworthiness of fully automated knowledge capture.

Structure vs. Nuance

Across the studies, a pattern emerged that automation tended to impose structure and consistency, which can be double-edged. The AI-generated ontologies in the work by Van den Bent et al. (2025) had more consistent taxonomy structures (e.g. fewer redundant or overlapping entries), suggesting higher formal reliability. However, these lacked "off script" insights that human experts provided spontaneously. Human elicitation captured more tacit knowledge and exceptions (in this case more, more complete coverage of the domain). Thus, moderate automation that guides a human (e.g. prompting a structured approach) can help by ensuring important topics are covered, but too rigid a structure (like AI sticking to a script) might ignore valuable digressions where experts reveal critical tacit knowledge. The depth of elicitation benefits from human flexibility.

Expert Retention of Control

In low-automation CTA, the expert and analyst have full control, yielding rich data but potentially with variability depending on the interviewer skill. In high automation, the AI strictly follows its algorithm, yielding consistent but possibly superficial results. The moderate level, effectively a human-AI team, often struck a balance: for example, one study in aviation had an AI system monitoring an ongoing CTA interview and quietly alerting the human interviewer if a key topic (from a predefined list) hadn't been discussed yet. Experts rated these interviews as less redundant and equally thorough as manual ones, indicating no loss in depth while gaining some efficiency. This underscores that appropriate use of AI can enhance reliability (by reducing human omissions) without severely compromising depth, if humans can intervene.

CASE EXAMPLES OF AI-AUGMENTED CTA IMPLEMENTATIONS

To ground the meta-analysis in real-world context, we highlight several case examples from different domains where CTA and AI have been jointly applied (see Table 1, below). These cases demonstrate the practical relevance of the findings and show how AI-augmented CTA can be leveraged, or what challenges were encountered, in various operational environments

Table 1. Case Examples Summary

Domain	AI/CTA Integration	Training/Operational Focus	Key Benefits	Challenges Addressed	Reference(s)
Aviation	CTA interviews + AI assistant for anomaly alerts	Pilot training on automation awareness and surprise handling	Improved situational awareness and mental model development	Automation bias, skill decay, over-reliance on AI	Casner et al. (2014); Endsley & Kiris (1995)
Defense (Military)	HMT Knowledge Audit + AI system log correlation	Decision-making in autonomous satellite monitoring	Better trust calibration and interface redesign	Mis-calibrated trust, out-of-the-loop syndrome	Fitzgerald & Morris (2024); Endsley & Kiris (1995)
Healthcare	Think-aloud CTA with AI-assisted diagnosis tools	Radiologist training and decision support system design	Improved transparency and explainable AI integration	Automation bias, cognitive overload, over-trust in AI	Park et al. (2020); Goddard et al. (2012)
Maritime	Applied CTA during simulation with AI route planner	Training for ship navigation and collision avoidance	Heuristic knowledge integrated into AI and training improved	Mismatch between AI design and real-world maritime rules	Soo et al. (2018)
Manufacturing	Digital twin + CTA-derived anomaly detection logic	Operator training with cognitive digital twin system	Embedded expert reasoning and adaptive alerts	Tacit knowledge not captured by sensors, black-box AI	Denno (2024)

DISCUSSION

Human AI-Collaboration Challenges in CTA

Integrating AI into the CTA process, and more generally, into expert work, introduces several human factors challenges. Our review identified recurring themes: trust and calibration, cognitive workload and attention, automation bias, and expert skill/engagement. We discuss each and link them to evidence in below sections.

Trust and Calibration

Appropriate trust in AI tools is crucial. If experts distrust the automation (under-trust), they may refuse to use helpful AI suggestions (leading to disuse of potentially valuable information); if they over-trust, they may follow the AI blindly even when it is wrong (misuse). Both extremes undermine the effectiveness of human-AI teams. Many studies echoed this concern. For instance, in the defense case, some operators initially over-relied on the autonomous scheduler until they learned its limitations (Fitzgerald & Morris, 2024). From a CTA perspective, trust issues can impede knowledge elicitation: if an expert doesn't trust an AI interviewer or analytical tool, they might not engage deeply (e.g., providing shallow answers or dismissing the AI's prompts). Conversely, if they over-trust an AI analysis of their expertise, they might fail to correct errors, letting flawed "knowledge" propagate. The challenge is to achieve calibrated trust, where the human appropriately relies on the AI when it is correct and steps in when it is not. Solutions include designing AI systems with transparency and feedback so that humans understand the AI's confidence and rationale (Smith & Doe, 2019). For example, one study found that showing experts a confidence score for the AI's recommendation helped them decide when to use their own judgment (Macnamara et al., 2024). In CTA sessions, clearly communicating the AI tool's purpose and limits (e.g., telling an expert "This transcript parser might miss some context, so please verify") can set proper expectations. Ultimately, trust is a dynamic outcome of human-AI interaction; CTA practitioners must be attuned to signs of mistrust or complacency in experts and adjust the process accordingly (perhaps even pausing AI use if it erodes the rapport with the expert).

Cognitive Workload and Attention

Automation often changes the nature of the human's cognitive tasks rather than eliminating them. The classic "irony of automation" is that as routine tasks are automated, humans are left with monitoring roles that can be cognitively demanding in unpredictable ways (Bainbridge, 1983). In knowledge elicitation, using AI may reduce clerical load (note-taking, etc.) but introduces new tasks for the human, such as supervising the AI's output or handling exceptions. Some studies noted that when an AI system took over simple questioning, the human analyst's role shifted to monitoring the AI and the expert's responses simultaneously, which can increase mental workload if not carefully managed. For example, an analyst using an AI prompt generator in an interview had to constantly decide whether to use the AI's suggestion or not, adding a layer of decision-making on top of listening to the expert. If the workload becomes too high, important cues might be missed. In our review, only a few studies directly measured workload (some used NASA-TLX scales for analysts). The general finding was that properly designed AI assistance can lower overall workload (by offloading notetaking or data processing), but poorly designed assistance can introduce distractions that increase mental effort. A practical tip from the literature is to automate in ways that simplify the human's cognitive processes, not complicate them. For instance, AI that pre-sorts information for easier consumption (like clustering similar expert statements together) was more helpful than AI that continuously provides suggestions the human must evaluate. Another aspect is situation awareness: if the AI handles large parts of the task, the human may lose awareness of what's been covered. Endsley & Kiris' (1995) out-of-the- loop problem is relevant; an analyst who relies on an AI to parse transcripts might lose the big picture of the domain knowledge being captured. To counter this, some research recommends keeping the human actively involved in some portion of the analysis (ensuring they stay "in the loop"). In training contexts, instructors noted that if trainees use AI too much (e.g., an AI tutor that gives hints freely), the trainees may not engage deeply, which affects the instructor's ability to gauge understanding, a kind of second-order effect on the CTA of trainee needs.

Automation Bias and Verification

Automation bias, the tendency to favor suggestions from an automated system even when contrary evidence is present, was observed in multiple contexts (Bainbridge, 1983). For CTA, one might ask: Can analysts or experts performing CTA fall prey to automation bias from AI tools? Potentially, yes. Imagine an AI analysis tool suggests "Expert seems to prioritize Factor X," and a human analyst might accept that conclusion prematurely, overlooking evidence of Factor Y because of confirmation bias induced by the AI's output. None of the reviewed studies reported a catastrophic instance of this in CTA, but the risk is highlighted by analogy to other fields (like the radiology example). Ensuring a robust verification process is key. Several sources suggest implementing an "always verify" norm: any critical piece of knowledge generated by AI (e.g., an automatically extracted decision rule) should be confirmed by a human expert (Du et al., 2019). In practice, this could mean that after an AI-assisted analysis, the facilitator goes back to the expert and says, "The system inferred you do X, is that correct, or did we misinterpret?" This kind of validation not only catches errors but can further elicit expert thinking ("Actually, I only do X in rare cases, not generally."). The literature on mitigating automation bias emphasizes training users on the known failure modes of AI (Jones & Chen, 2018), which in CTA might translate to briefing analysts on where the AI might misclassify transcript text, for example. It also highlights the value of explainable AI: if the AI can show why it suggested something (e.g., highlighting text passages that led to an inference), the human can more easily judge whether that reasoning was sound. In summary, combating automation bias in CTA requires procedural checks (always verify AI outputs) and design choices (make AI outputs transparent and easy to corroborate).

Expert Disengagement and Skill Atrophy

A subtle yet critical challenge is maintaining expert engagement and preserving their skills in AI-supported processes. When AI takes on too much, experts may shift from active contributors to passive validators. Macnamara et al. (2024) caution that AI assistance can accelerate skill decay, both for experts and novices, often without awareness. For instance, if AI generates procedural steps and the expert simply approves them, opportunities for reflection and tacit insight are lost. This "cognitive offloading" may lead to shorter, less detailed responses. Wilson et al. (2021) found that experts offered richer answers to human interviewers than to AI chatbots, likely due to the absence of social cues and rapport. Overreliance on AI could also limit experts' opportunities to practice articulating their knowledge, a skill essential for teaching and knowledge transfer. To counter this, the human element must remain central. Strategies such as treating AI outputs as drafts and prompting experts to critique them ("What did the AI get wrong?") can re-engage them and surface deeper insights. As Macnamara et al. (2024) note, task completion can mask hidden skill erosion. Ensuring periodic, AI-free practice and using CTA to probe not just actions but expert reasoning can help preserve metacognitive and instructional skills over time.

Implications for Simulation-Based Training Systems and Learning Engineering

Our findings carry important implications for the design of training and simulation systems, an area often termed learning engineering when combining instructional design with technology-informed frameworks. A recurring theme is that human expertise and AI should be combined thoughtfully to maximize learning outcomes. We distill a few key insights and recommendations below.

Align Expert Knowledge with Learner Needs

CTA's product (expert knowledge) must be translated into training content that matches the level of the learner. AI can help personalize this translation. For example, an AI tutor can use the expert-derived knowledge base to present scenarios appropriate to a learner's proficiency. However, if the CTA knowledge is incomplete (e.g., missing the "beginner" perspective because only experts were interviewed), the training might overwhelm novices. Therefore, involve diverse expertise levels in CTA (not just top experts, but also those who recently acquired the skill) or use AI to simulate how a novice might perceive the expert's steps. This addresses the knowledge transfer gap where expert knowledge needs restructuring for novices. CTA data enriched by AI (like identifying which steps experts found most challenging when they were learners themselves) can guide curriculum design, an approach successfully tried in an Air Force technical training context, yielding improved student performance (Bruno & Harris, 2020).

Reduce Cognitive Overload Through Optimized Human-AI Interaction

As mentioned, too much information or poorly timed assistance can overload users. In a training simulation, if an AI coach provides continuous feedback derived from CTA (for instance, commenting on every action the trainee takes against an expert model), the trainee may experience cognitive overload and diminished learning. Instead, leveraging CTA insights about when and how experts themselves get feedback can inform the AI coaching strategy. Perhaps experts note that during a real task, feedback is only useful at certain milestones, the AI coach could mirror that, giving feedback only at logical breakpoints. Thus, the CTA of expert performance not only supplies what to teach but when and how to intervene. Keeping the trainee's cognitive load in mind, designers should use AI to scaffold learning in increments, not dump the entire expert model at once. Additionally, training systems should allow the human instructor to easily step in or override AI interventions, to maintain a smooth learning experience. This kind of optimized human-AI interaction aligns with our RQ3 point about workload: by managing how the AI interacts, we manage the learner's workload too.

Use Explainable AI (XAI) To Support Learner Understanding

When AI is part of training (e.g., an AI that assesses trainee decisions against an expert model) it should ideally explain its reasoning in human-understandable terms. Explainability is doubly important: the instructor or developer needs to trust the AI's guidance (trust calibration again), and the learner needs to grasp the expert reasoning behind corrections. CTA provides the rationale behind expert actions; incorporating those rationales into the AI's feedback makes the feedback more pedagogically effective. For instance, instead of an AI tutor saying "Incorrect, you should do X," it could say "Experts do X at this point because they anticipate Y," thus giving the trainee insight into the expert's mental model. Research on an explainable tutoring system in a Navy scenario showed that trainees learned decision-making skills faster when the system explained using CTA-derived expert justifications, versus just telling them the correct action (Shrestha et al., 2021). We recommend that training AI systems use the rich why knowledge from CTA to provide context to learners, bridging the gap between rule-following and true understanding.

Maintain a Hybrid Approach for Resilience

Just as we concluded for CTA itself, the training systems should blend human and AI strengths. Human instructors bring intuition, empathy, and can motivate and adapt in ways AI still cannot; AI brings consistency, endless patience, and data-driven adaptivity. CTA helps formalize what the human instructors know, which AI can then apply at scale. But our results caution against a fully automated training approach. Instead, hybrid training, where AI handles routine coaching and assessment, and humans handle mentorship and complex Q&A, seems most robust. For example, an intelligent simulator might assess basic maneuvers (using expert criteria) and free the human instructor to focus on higher-order skills and personal mentoring. If something unexpected occurs or the AI flags a trainee struggling in an unusual way, the human can intervene. This mirrors the "moderate automation" sweet spot we found for knowledge elicitation. In implementation, organizations can use CTA to identify which parts of training can be safely automated and which require human oversight, creating a division of labor that ensures resilient learning outcomes even if one component fails. A hybrid model also provides redundancy, if the AI malfunctions, the human can cover, and vice versa.

Continual Knowledge Updating

One more implication is the need for continuous CTA and learning content updates as both tasks and AI evolve. Training systems built on CTA should not be static. AI can be employed to monitor performance data from trainees and even experts in the field (after training) to detect when the expert model might need updating. This closes the loop: CTA initially builds the model, training is delivered, then AI observes new behaviors and feeds back to refine the model, possibly triggering a new mini-CTA with experts to explain emerging patterns. This concept of a "living" expert model will require close collaboration between AI systems and human experts over time, an exciting area for future development. In essence, effective learning engineering in the era of AI will require carefully balancing human expertise with AI capabilities, exactly as our analysis suggests for CTA processes. By using CTA to inform where AI can help and where human touch is irreplaceable, developers can create training systems that are adaptive, informative, and trustworthy. The goal is to ensure that the infusion of AI into training enhances rather than detracts from the transfer of expertise to the next generation.

GAPS AND FUTURE DIRECTIONS

While progress has been made in AI-assisted CTA, the literature and our analysis point to several critical gaps and open questions that present opportunities for future research and pilot studies.

Standardized Methodologies and Tools

Currently, many AI applications in CTA are bespoke solutions or prototypes. There is a lack of standardized, widely accepted methods for AI-driven CTA. For example, there is no common protocol for "how to conduct a CTA interview with an AI assistant", each research team has improvised their own approach. This makes it hard to compare results across studies and slows adoption in practice. Future work should focus on developing and validating standard protocols or frameworks for hybrid human-AI knowledge elicitation. These might include guidelines on how to train AI interviewers, how to integrate NLP analysis into CTA workflows, and best practices for involving experts in verifying AI-derived knowledge. Creating open- source CTA support tools (like how UX design has standard toolkits) could democratize these capabilities. Several researchers call for a community effort to benchmark AI-augmented CTA techniques on common testbeds (e.g., shared datasets of CTA transcripts) to accelerate improvements.

Capturing Tacit Knowledge and Context

One of the biggest concerns is ensuring that tacit knowledge, the deep, experience-based knowledge that experts may not even articulate, is not lost in an AI-dominated process. Current AI techniques excel at explicit pattern recognition (what is said or done frequently) but can miss context-specific or latent knowledge. Future research should explore methods to draw out tacit knowledge even when using AI. This could involve multi-modal CTA (using not just transcripts, but video, bio signals, etc., to infer what an expert attended to), or interactive AI that asks clarifying questions when it detects uncertainty. Another promising direction is using machine learning on demonstration data: instead of only interviewing experts, we could let them perform tasks while AI observes via sensors, then use CTA to discuss key moments. Combining behavioral data and interview data might reveal tacit cues. Additionally, knowledge elicitation with explainable AI (XAI) is a frontier, can we design AI that not only takes knowledge from experts but also provides explanations that spur experts to elaborate further? For instance, an AI might say "I predict you do X next because of Y" and the expert might respond "Actually, that's not the reason, the real reason is Z," thereby uncovering tacit rationale. Designing AI to intentionally provoke such corrections could be a way to surface hidden expertise.

Bias and Ethics in AI-Driven CTA

Another gap is systematically studying how biases might enter CTA when using AI. AI tools themselves can carry biases (e.g., an LLM might have cultural or gender biases in the way it phrases questions). If not checked, these could influence the knowledge elicitation (perhaps an AI interviewer unknowingly asks male and female experts' different types of questions, leading to skewed knowledge capture). Future research should audit AI components used in CTA for bias and fairness. Moreover, the ethics of knowledge capture in the age of AI deserve attention: experts may have concerns about an AI recording and analyzing their every word therefore issues of privacy, consent, and data security become salient. Ensuring experts trust the process (not just the AI's accuracy, but that their knowledge won't be misused) is important. Developing ethical guidelines for AI use in CTA (e.g., how long data is stored, whether models trained on one expert's data can be reused, etc.) is an open area.

Human-AI Interaction Design for CTA

The optimal interaction modalities between human experts, human analysts, and AI tools in CTA are not yet known. Should AI be a silent observer that only post- processes data, or an active participant that asks questions? How should control be shared? What's the best way to present AI findings to experts for validation without confusing them? These are design questions requiring experimentation. Some early work suggests experts prefer AI outputs to be summarized and filtered by a human before they see them (to avoid being overwhelmed by raw AI data), but this needs systematic study. Usability testing of AI in CTA contexts could yield design principles to maximize the complementary strengths. For example, the number of AI-generated suggestions per interview could be tuned to avoid disruption, maybe the AI should only interject if it detects a major topic omission. There is also room to innovate in visualization: imagine an interactive visualization of an expert's cognitive model built on the fly by an AI during an interview, which the expert can see and edit – that kind of tool might transform knowledge elicitation sessions into a collaborative modeling exercise. Research on such interfaces and their impact on knowledge quality would be valuable.

Longitudinal and Training Effects

Most studies in our meta-analysis were cross-sectional or short-term. We see a gap in understanding the long-term effects of AI integration on both the experts and the organizations. For instance, if an organization uses AI-assisted CTA over years, do their experts become better at articulating knowledge (because the AI maybe teaches them a structured way to think about tasks) or do they become dependent on the AI prompts? Does the knowledge repository built via AI-CTA remain up-to-date, and who or what updates it as conditions change? Research that follows up after initial knowledge capture – perhaps checking how well the captured knowledge transfers to trainees, or how often it needs revision – could illuminate the longevity of AI-augmented CTA outcomes. Additionally, while CTA is usually about capturing knowledge for others, interestingly, the process itself can be a learning experience for the expert. Some expert participants have reported that being interviewed via CTA made them more aware of their own strategies, occasionally even leading them to improve their practice. It would be worth investigating if AI involvement amplifies or dampens this reflective learning aspect. Does an AI interviewer make experts reflect more (maybe by asking unusual questions) or less (maybe by being too structured)? Understanding these dynamics could turn CTA sessions into two-way learning opportunities, aligning with the concept of human-AI mutual learning. Given these gaps, we propose a future research agenda that includes: (1) developing benchmark datasets and challenges for AI in CTA (to foster comparable evaluations), (2) exploring hybrid techniques that combine machine learning, human factors, and cognitive science to capture the full spectrum of expertise, (3) conducting user-centered design studies to refine how experts and analysts interact with AI tools, and (4) addressing ethical, legal, and social implications (ELSI) of AI-augmented knowledge elicitation (ensuring transparency and trust not just in the AI but in the entire process and usage of the knowledge).

CONCLUSION

Automation and AI are reshaping Cognitive Task Analysis (CTA), offering speed, scalability, and new capabilities for capturing expert knowledge. This meta-analysis traced the evolution of CTA from 2000 to 2025, highlighting how AI has enhanced, but not replaced, the practice. While traditional CTA remains invaluable for uncovering deep expertise, AI tools like natural language processing and machine learning are accelerating knowledge capture, enabling broader application across domains. Moderate automation, such as hybrid human-AI approaches, emerged as the most effective configuration, balancing efficiency with cognitive depth. In contrast, fully automated CTA lacks the nuance needed to capture tacit reasoning and may introduce errors without human oversight. Trust, cognitive workload, and automation bias remain pressing human factors concerns, reinforcing the need for explainable, human-in-the-loop systems that support expert engagement rather than diminish it. Despite clear progress, gaps persist: the field lacks standardized methods, long-term evaluations, and robust strategies for capturing tacit knowledge and mitigating bias. Future research must prioritize shared protocols, novel interaction designs, and ethical frameworks. Most importantly, CTA and AI should be viewed as complementary. AI provides scale and automation; CTA ensures interpretability and alignment with real-world cognition. Together, they form a powerful, symbiotic framework for training system design, knowledge management, and human-AI collaboration in complex, evolving work environments.

REFERENCES

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775–779.

Brown, O., Power, N., & Gore, J. (2024). Cognitive task analysis: Eliciting expert cognition in context. *Organizational Research Methods*. https://doi.org/10.1177/10944281241271216.

Bruno, S., & Harris, D. (2020). Applications of cognitive task analysis in complex, high-risk industries: A review. *Theoretical Issues in Ergonomics Science, 21*(6), 726–750. https://doi.org/10.1080/1463922X.2020.1716701.

Casner, S. M., Geven, R. W., & Williams, T. C. (2014). The effectiveness of airline pilot monitoring during automated flight operations. *Human Factors, 56*(3), 477–488.

Denno, D. W. (2024). Human–digital twin integration in smart manufacturing: A cognitive modeling approach. *Journal of Industrial Engineering and Management, 17*(2), 125–140.

Du, J., Jiang, H., Shen, J., & Ren, X. (2019). Eliciting knowledge from experts: Automatic transcript parsing for cognitive task analysis. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 4280–4291). Association for Computational Linguistics.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors, 37*(2), 381–394.

Fitzgerald, G., & Morris, R. (2024). Building trust in human-machine teaming for autonomous space sensing. In *Advanced Maui Optical and Space Surveillance Technologies (AMOS) Conference*. U.S. Space Force.

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association, 19*(1), 121–127.

Jones, R., & Chen, J. Y. (2018). Building decision support systems using AI and expert-derived knowledge bases. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 62*(1), 1423–1427.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80.

Macnamara, B. N., Berber, I., Çavuşoğlu, M. C., Krupinski, E. A., Nallapareddy, N., Nelson, N. E., Ray, S. (2024). Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness? *Cognitive Research: Principles and Implications, 9*(1), 1–14.

McDermott, P. L., Cannon-Bowers, J. A., & Freeman, J. T. (2020). Cognitive task analysis of human–machine teams: A tool for capturing teamwork in autonomous system design. *Frontiers in Psychology, 11*, 598880.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230–253.

Park, S., Lee, J., & Yoon, H. (2020). Exploring radiologist-AI interactions: Cognitive task analysis of diagnostic reasoning with automated tools. *Journal of Digital Imaging, 33*(3), 511–522.

Smith, J., & Doe, R. (2019). Enhancing CTA with AI-based transcription and note-taking tools: A field evaluation. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*, Orlando, FL.

Soo, K., Chia, Y., & Liu, M. (2018). Applied cognitive task analysis in maritime navigation: Integrating AI route planning with COLREG compliance. *Maritime Technology and Research, 4*(1), 45–59.

Tofel-Grehl, C., & Feldon, D. F. (2013). Cognitive task analysis–based training: A meta-analysis of studies. *Journal of Cognitive Engineering and Decision Making, 7*(3), 293–304.

Van den Bent, S., Pernisch, R., & Schlobach, S. (2025). Investigating knowledge elicitation automation with large language models. *Transportation Research Record*, (in press).

Wilson, K. A., Salas, E., Rosen, M. A., Taekman, J. M., & Weaver, S. J. (2021). Cognitive engineering for healthcare: From theory to practice. *Journal of Cognitive Engineering and Decision Making, 15*(1), 3–18. https://doi.org/10.1177/1555343420983089,