POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

The Product Manager and the AI Buffer: A Multi-Level Framework Linking Cognition to Organizational Practice

Tesi di Laurea Magistrale in Management Engineering

Author: **Duccio Profeti** **Edoardo Pinzauti**

Student ID: 10700827 YYYYY
Advisor: Prof. Sergio Terzi, Politecnico di Milano
Co-advisors: Prof. Patrick Rau, Tsinghua University
Academic Year: Academic Year: 2026–27

Abstract

Generative AI is increasingly embedded in knowledge-intensive work, producing summaries, explanations, and recommendations that persist alongside source material. These AI-generated external representations can reduce effort and accelerate decisions, but they may also reshape what people encode, how they retrieve information, and how confidently they act on it. Because research on cognition and organizational adoption is often disconnected, the mechanisms that link individual memory effects to professional decision-making and governance remain underspecified.

This thesis develops an integrative, multi-level framework that treats AI outputs as part of a task's cognitive infrastructure. It introduces the *AI Buffer Model*, which characterizes how persistent AI representations interact with memory encoding, cue-dependent retrieval, source monitoring, confidence calibration, and cognitive load. The model is embedded within an *AI-Augmented Decision Framework* that links these cognitive mechanisms to decision processes, role shifts, and accountability structures in organizations.

Two complementary studies test the framework. Study 1 is a controlled experiment ($N = 36$) in which participants read short expository articles with AI-generated summaries that vary in *timing* (pre-reading, synchronous, post-reading) and *structure* (integrated paragraph vs. segmented bullets). Study 2 is a cross-national survey of product managers ($N = 74$; USA and China) examining AI use, perceived impacts on workflow and decision support, and governance practices.

Results show that AI summaries improve recognition performance but do not strengthen generative free recall. Benefits are largest when summaries are available before reading and can function as encoding scaffolds, whereas segmented summaries increase false-lure endorsement (OR 5.93), indicating elevated source-monitoring risk when representations are fragmented. Field evidence shows perceived productivity and decision-support gains, but also frequent governance ambiguity (e.g., unclear ownership and missing guidelines). Integrating both studies, the thesis argues that representational design parameters—timing, structure, and provenance cues—are not cosmetic interface choices: they shape cognitive mechanisms that scale into artifact-centric work practices and governance needs. Practical implications include using AI early for framing, preferring integrated and traceable artifacts, and institutionalizing verifica-

tion routines and clear accountability for AI-mediated evidence.

Keywords: generative AI, external representations, cognitive offloading, source monitoring, product management, governance

Abstract in lingua italiana

L'intelligenza artificiale generativa e' sempre piu' integrata nel lavoro ad alta intensita' di conoscenza, producendo riassunti, spiegazioni e raccomandazioni che permangono accanto al materiale originale. Queste rappresentazioni esterne generate dall'IA possono ridurre lo sforzo e accelerare le decisioni, ma possono anche modificare cio' che le persone codificano, come recuperano le informazioni e con quanta sicurezza agiscono. Poiche' la ricerca sui meccanismi cognitivi e quella sull'adozione organizzativa sono spesso disconnesse, i nessi tra effetti sulla memoria individuale, processi decisionali professionali e governance restano sottospecificati.

Questa tesi sviluppa un framework integrativo multi-livello che considera gli output dell'IA parte dell'infrastruttura cognitiva del compito. Introduce l'*AI Buffer Model*, che descrive come rappresentazioni IA persistenti interagiscano con codifica, recupero basato su indizi, monitoraggio della fonte, calibrazione della fiducia e carico cognitivo. Il modello e' inserito in un *AI-Augmented Decision Framework* che collega tali meccanismi ai processi decisionali, ai cambiamenti di ruolo e alle strutture di accountability nelle organizzazioni.

Due studi complementari testano il framework. Lo Studio 1 e' un esperimento controllato ($N = 36$) in cui i partecipanti leggono brevi articoli espositivi con riassunti generati dall'IA che variano per *tempistica* (pre-lettura, sincrono, post-lettura) e *struttura* (paragrafo integrato vs. punti elenco segmentati). Lo Studio 2 e' un sondaggio cross-nazionale su product manager ($N = 74$; USA e Cina) che analizza utilizzo dell'IA, impatti percepiti su workflow e supporto decisionale, e pratiche di governance.

I risultati mostrano che i riassunti IA migliorano il riconoscimento, ma non rafforzano il richiamo libero generativo. I benefici sono massimi quando i riassunti sono disponibili prima della lettura e fungono da scaffold di codifica, mentre i riassunti segmentati aumentano l'accettazione di false esche (OR 5.93), evidenziando un rischio di monitoraggio della fonte quando le rappresentazioni sono frammentate. I dati sul campo indicano benefici percepiti in termini di produttivita' e supporto decisionale, ma anche frequente ambiguita' di governance (ad es., ownership poco chiara e linee guida mancanti). Integrando i due studi, la tesi sostiene che i parametri di progettazione—tempistica, struttura e indizi di provenienza—non sono scelte cosmetiche: plasmano meccanismi cognitivi che scalano in pratiche di lavoro centrate sugli artefatti e in esigenze di governance. Le implicazioni pratiche includono usare l'IA precocemente per

inquadrare i problemi, preferire artefatti integrati e tracciabili, e istituzionalizzare routine di verifica e accountability esplicite per le evidenze mediate dall'IA.

Parole chiave: intelligenza artificiale generativa, rappresentazioni esterne, offloading cognitivo, monitoraggio della fonte, product management, governance

Contents

List of Figures

List of Tables

List of Acronyms / Glossary

AI	Artificial Intelligence
LLM	Large Language Model
PM	Product Manager / Product Management
PRD	Product Requirements Document
MCQ	Multiple-Choice Question
NLP	Natural Language Processing
RPA	Robotic Process Automation
QA	Quality Assurance
UHI	Urban Heat Island
GDPR	General Data Protection Regulation
EUV	Extreme Ultraviolet Lithography
SoC	System-on-a-Chip

1 | Introduction

Generative AI systems are increasingly embedded in everyday information work. Two use cases are especially salient for learning and decision-intensive knowledge work: (i) AI-generated summaries and explanations that accompany reading and study, and (ii) AI-generated analyses and recommendations that accompany professional judgment. In both cases, the AI produces *external representations*—persistent, content-bearing artifacts that can guide attention, provide cues, and reshape how information is encoded, retrieved, and acted upon.

Despite this shared reliance on AI-generated representations, prior research is often fragmented across disciplinary boundaries. Cognitive science has focused on memory, attention, and learning mechanisms, while organizational and information-systems research has emphasized adoption, productivity, and role change. Generative AI challenges this separation because it inserts persistent representations directly into everyday cognitive tasks: changes in encoding, retrieval, and confidence at the individual level can propagate to decision quality, accountability, and governance at the organizational level.

These developments sit on top of a longer trajectory of AI research and deployment. Historical and bibliometric accounts describe multiple waves of AI evolution—from early symbolic systems to contemporary deep-learning approaches and the emergence of new "generations" of AI capabilities—that have expanded the feasibility of natural-language interfaces and large-scale decision support [21, 74, 77, 94].

From a cognitive perspective, external representations can support performance by reducing extraneous processing costs, but they can also change encoding strategies via cognitive offloading [71, 80]. From a decision-making perspective, the same representations can shape reliance: users may defer to AI outputs without fully scrutinizing their provenance or quality, creating epistemic and governance risks [67, 95]. These tensions are particularly relevant for roles such as product management, where decisions integrate heterogeneous evidence (user research, analytics, technical constraints, and stakeholder input) and where generative AI is increasingly used to summarize, prioritize, and communicate [28, 31, 34, 63].

Across sectors, AI is framed as both an operational technology (automation of routine processes) and a strategic capability (augmenting scenario exploration, planning, and coordina-

tion). Industry and practitioner sources document diffusion of AI applications across corporate functions—from customer service and operations management to finance and human resources—and emphasize that value capture requires changes in workflows, roles, and governance rather than model deployment alone [9, 23, 30, 52, 57, 58, 62, 70].

This thesis addresses these issues through a *multi-level perspective* that links cognitive mechanisms to organizational practice. At the cognitive level, Study 1 uses a controlled reading-and-memory experiment in which AI assistance is operationalized as pre-generated summaries that vary in *structure* (integrated vs. segmented) and *timing* (pre-reading vs. synchronous vs. post-reading). At the organizational level, Study 2 examines how product managers integrate generative AI tools into decision-intensive workflows and how they perceive impacts on efficiency, judgment, and ethical responsibility. The thesis proposes the *AI Buffer* as a unifying concept: AI-generated external representations can function as persistent buffers that support cue-based cognition while also introducing source-monitoring and reliance vulnerabilities.

Context and Motivation Learning from text depends on how readers allocate attention during encoding, how they manage limited working-memory resources, and which cues are available at retrieval [7, 20, 88]. In real-world settings, learners frequently rely on external representations—notes, outlines, highlighted passages, and search engines—to support these processes. Generative AI introduces a new, scalable form of external representation: systems can produce structured summaries instantly and present them alongside source material.

At the same time, external representations can shift effort away from internal encoding. The "Google effect" shows that when information is easy to re-access, people are more likely to remember where it is than what it is [29, 71, 80]. In generative-AI settings, this creates a core tension: summaries may support understanding and recognition by providing cues, but they may also encourage shallow processing, reduce detail-rich encoding, and increase vulnerability to source confusion when AI content blends with the original material [15, 37].

Beyond individual learning, the same dynamics play out in professional contexts where decisions depend on accurate, contextual recall. Knowledge-management research suggests that AI can help organizations retrieve prior decisions and maintain continuity across projects [66]. However, AI-advised decision-making also raises risks of epistemic dependence and opaque reasoning [67, 95]. Product management provides a concrete setting in which these issues are amplified: product managers are expected to synthesize noisy information into decisions and to communicate rationale clearly across teams, making AI-provided summaries and recommendations both useful and potentially hazardous [31, 34, 63].

Research Problem and Gap Despite rapid adoption, there remains limited integrated evidence on *how* AI-generated external representations shape cognition and decision-making.

At the cognitive level, evidence on generative AI is still emerging and often treats "AI assistance" as a unitary intervention. Yet design choices matter: *when* a summary is available and *how* it is

structured can change attention allocation, encoding depth, and the balance between internal reconstruction and reliance on external cues. Controlled studies suggest that conversational AI can amplify false memories under certain conditions [15], and reviews warn about over-reliance and reduced cognitive engagement [93], but there is still a need for causal evidence that isolates interface-relevant dimensions (timing, format) and measures both performance and epistemic risk.

At the organizational level, many accounts focus on productivity or adoption narratives, while fewer connect observed reliance patterns to cognitive mechanisms such as offloading and source monitoring. For decision-intensive roles such as product management, the open question is not only whether AI improves efficiency, but whether it changes how decisions are formed, justified, and governed—including how trust, transparency, and accountability are maintained when AI outputs shape reasoning [67, 86].

The central gap motivating this thesis is therefore the lack of a *multi-level framework* that links (i) mechanisms of AI-assisted encoding and retrieval to (ii) patterns of reliance and judgment in professional decision-making contexts.

Research Objectives and Multi-Level Perspective Following common master-thesis conventions, the work is organized around concrete objectives that span the two levels of analysis and their integration.

- **O1 — Cognitive baseline:** quantify memory differences between AI-assisted reading and unaided reading under matched constraints (Study 1).
- **O2 — Cognitive design dimensions:** test how summary *timing* and *structure* affect recall quality, recognition accuracy, and behavioral indicators of reliance (Study 1).
- **O3 — Epistemic risk:** measure susceptibility to false-lure endorsement and source-monitoring errors when AI-generated content is present (Study 1).
- **O4 — Organizational practice:** characterize how product managers deploy generative AI tools in decision-intensive workflows, including perceived benefits, risks, and ethical responsibilities (Study 2).
- **O5 — Cross-level integration:** develop an integrated account linking AI-generated external representations to both memory mechanisms and professional decision processes, and derive propositions that can guide safer tool design and governance.

Research Questions Building on the research gap identified in the previous section, this thesis investigates how artificial intelligence reshapes knowledge-intensive decision-making by simultaneously affecting individual cognitive processes and organizational practices. To address this objective, the study adopts a multi-level perspective and is guided by the following research questions:

- **RQ1 – Cognitive Level (Study 1). RQ1:** How do AI-generated external representations influence core human memory processes during knowledge-intensive tasks?

  - *Memory encoding* (attention allocation, depth of processing).
  - *Recall and recognition* (quantity and quality of retained information).
  - *Source monitoring* (discriminating article content from AI-generated content).
  - *Confidence calibration* (alignment between confidence and accuracy).
  - *Cognitive load* (perceived effort and split-attention costs).

  Conceptually, RQ1 corresponds to the AI Buffer Model formulated as a research question: it asks how persistent AI-generated representations shape encoding, retrieval, and metacognitive monitoring.
- **RQ2 – Organizational Level (Study 2). RQ2:** How does the adoption of AI tools shape decision-making practices, role configuration, and perceived responsibility in product management?
  - Role transformation and reallocation of tasks toward more strategic work.
  - AI-supported decision-making (sensemaking, prioritization, and communication).
  - Governance and ethics (accountability, transparency, oversight practices).
  - Cross-context comparison (e.g., differences across organizational and cultural settings such as USA–China).

  Importantly, RQ2 is not treated as a mere "application" of Study 1, but as an autonomous analytical level that characterizes AI adoption in a decision-intensive managerial domain.
- **RQ3 – Integrative Level (Cross-study). RQ3:** How can changes in organizational decision-making practices enabled by AI be explained through underlying cognitive mechanisms of AI-assisted memory?

RQ3 is the key integrating question of the thesis: rather than reporting two disconnected sets of findings, the thesis aims to explain observed organizational patterns (Study 2) using a coherent account of memory mechanisms identified under controlled conditions (Study 1). This integrative aim motivates the conceptual framework in Chapter 3 and the dedicated cross-study integration logic developed later in the thesis.

Contribution of the Thesis The thesis contributes at three levels: cognitive theory, organizational understanding, and integrative framing.

Cognitive contribution. Study 1 provides design-sensitive evidence on AI-assisted reading by measuring both performance (recall and recognition) and epistemic risk (false-lure endorsement), explicitly manipulating timing and structure of AI-generated summaries.

Organizational contribution. Study 2 synthesizes patterns of AI use in product management and clarifies how AI tools are positioned as automation and augmentation within decision-intensive workflows, including perceived shifts in responsibility and ethical vigilance.

Integrative contribution. The thesis advances the *AI Buffer* as a compact bridge concept: AI-generated representations can function as persistent external buffers that support cue-based cognition while also altering source monitoring and reliance conditions. This framing supports

actionable implications for AI interface design, training, and governance in knowledge work.

Scope, Definitions, and Assumptions This thesis focuses on AI-generated external representations and their consequences under two complementary methodological lenses.

Key definitions. *Generative AI* refers to systems that produce novel text based on learned patterns from data. *External representations* refer to persistent artifacts (summaries, notes, recommendations) that users can consult during a task. *Cognitive offloading* refers to shifting cognitive work from internal memory to external aids [71]. *Source monitoring* refers to processes by which people attribute information to its origin [37]. In this thesis, the *AI Buffer* is defined as an AI-generated representation that persists during a task and can shape encoding, retrieval, and metacognitive monitoring.

Study boundaries. Study 1 is intentionally scoped to controlled laboratory conditions, pre-generated summaries (rather than open-ended dialogue), and within-session memory measures. Study 2 is scoped to product management as a decision-intensive domain and emphasizes professional practice and governance rather than direct cognitive performance measurement.

Assumptions. The thesis treats AI summaries as fallible representations rather than authoritative ground truth and assumes that interface properties (timing, structure, provenance cues) can change cognitive and organizational outcomes by shaping attention, cue availability, and reliance norms.

Structure of the Thesis The remainder of the thesis follows the structure below:

- **Chapter 2 (Literature Review)** synthesizes human-memory foundations and research on cognitive offloading, external representations, and human–AI reliance in knowledge work and managerial contexts.
- **Chapter 3 (Conceptual Framework and Hypotheses)** introduces the AI Buffer model and derives cross-level propositions linking cognitive mechanisms to decision processes.
- **Chapter 4 (Methodology)** details Study 1 (controlled experiment) and Study 2 (field study on product managers), including instruments and ethical considerations.
- **Chapters 5–7 (Analysis, Results, Integration)** report analytical strategy and findings for each study and then integrate evidence across levels.
- **Chapters 8–9 (Discussion and Conclusion)** interpret implications, limitations, and future research directions, and summarize the thesis contributions.
- **Appendices** provide study materials, instruments, and supplementary outputs.

2 | Literature Review

This chapter reviews prior work that motivates the thesis structure introduced in Chapter 1. It first establishes foundations of AI in knowledge-intensive work, then examines AI as a form of cognitive support through the lenses of distributed cognition and cognitive offloading. Building on this, it reviews the foundations of human memory and metacognition, synthesizes research on how AI-generated representations influence memory processes, and examines confidence, source monitoring, and cognitive load in assisted cognition. The chapter then bridges individual memory mechanisms to decision-making processes and reviews human–AI decision-making dynamics including trust, reliance, and accountability. It closes with AI in professional and managerial contexts, a focused analysis of product management as a decision-intensive domain, and a synthesis that identifies the multi-level research gap motivating the conceptual framework developed in Chapter 3.

2.1. Foundations of AI and Generative AI in Knowledge-Intensive Work

The increasing adoption of Artificial Intelligence across organizational contexts has profoundly altered how knowledge-intensive work is structured and performed. While early implementations focused on automating narrowly defined tasks, recent advances have expanded AI's role toward supporting human reasoning, interpretation, and decision-making. Understanding this transition is essential for situating the cognitive and organizational implications of AI, particularly in professional roles where performance depends on the integration of complex information over time rather than on execution speed alone.

From Automation to Augmentation Historically, AI systems were primarily conceived as tools for automation: the value of AI resided in its ability to replace human labor in repetitive, rule-based, or highly standardized tasks. Organizational benefits were framed in terms of efficiency gains, cost reduction, and scalability. This perspective implicitly assumed that tasks could be decomposed into discrete operations and that optimal performance could be achieved by minimizing human involvement [44].

However, this automation-centered view has proven increasingly inadequate for describing AI use in knowledge-intensive work. Such work is characterized by ambiguity, contextual dependence, and the need to integrate heterogeneous sources of information. Decision quality in these settings depends less on computational precision than on interpretation, judgment, and the ability to maintain continuity across decisions made at different points in time. Fully automating these processes is neither feasible nor desirable, as it would require formalizing tacit knowledge and contextual understanding that are difficult to codify.

As a response, the notion of AI *augmentation* has gained prominence. Rather than replacing human decision-makers, augmentative AI systems aim to enhance human cognitive capabilities by modifying the informational environment in which decisions are made [44]. This shift reframes AI as a complement to human intelligence, emphasizing collaboration rather than substitution. Research on decision support systems suggests that AI is most effective when it assists humans in complex, non-routine tasks rather than when it attempts to fully automate them [54]. In these contexts, AI contributes value by enabling better sense-making and prioritization rather than by executing decisions independently.

Generative AI as a Representation Generator Generative AI systems differ from earlier decision-support tools because they can produce novel, fluent representations on demand: summaries, outlines, explanations, and draft rationales that can persist through the course of a workflow. Rather than merely retrieving information, they actively construct external representations that reorganize and reinterpret existing knowledge.

The cognitive significance of external representations has long been recognized. Research on text comprehension and knowledge construction demonstrates that structured representations can profoundly influence understanding by guiding attention, emphasizing relationships, and reducing the need for internal reconstruction [89]. Generative AI extends this principle by dynamically adapting representations to user goals and contextual constraints. These artifacts can reduce the effort required to integrate large volumes of information, but they also have the potential to steer attention toward what is represented—and away from what is omitted or de-emphasized. For this thesis, the central implication is that AI assistance is not only a source of information; it is also a mediator of encoding and retrieval by shaping which cues are available and how they are structured.

AI as a Cognitive Artifact in Knowledge Work A useful way to conceptualize generative AI in professional contexts is as a *cognitive artifact*: a tool that supports, enhances, or transforms cognitive processes by externalizing information and restructuring tasks [64]. Classic examples include written notes, diagrams, and computational tools, all of which alter how problems are approached and solved. Generative AI represents an advanced form of cognitive artifact: unlike static artifacts, it is adaptive and interactive, responding to user input and generating new representations in real time. By externalizing intermediate representations that would

otherwise need to be maintained internally, AI systems can reduce demands on working memory and enable users to focus on higher-level reasoning.

This perspective aligns with theories of distributed and extended cognition, which argue that cognitive processes are not confined to the individual mind but are distributed across internal and external resources [18, 35]. Within these frameworks, tools that are reliably integrated into task execution become part of the cognitive system itself. When used consistently, AI-generated artifacts can function as part of a distributed cognitive system, supporting continuity across time and across collaborators. This motivates treating AI outputs as external representations that can become integrated into reasoning, planning, and decision justification—not merely as optional add-ons.

Importantly, the integration of AI as a cognitive artifact raises questions about agency and control. While externalizing cognitive processes can enhance performance, it may also obscure the boundary between human understanding and machine-generated representations. This ambiguity underscores the need to examine not only the benefits of AI augmentation but also its effects on cognitive ownership and responsibility.

AI Literacy, Trust, and Effective Use The value and risks of AI assistance depend on how users interpret and evaluate outputs. The concept of *AI literacy* captures this dimension by encompassing technical understanding, interpretative skills, and ethical awareness related to AI use [91]. Effective use requires the ability to recognize uncertainty, anticipate failure modes, and calibrate when to verify, revise, or defer to AI-generated content.

Empirical research indicates that users with higher levels of AI literacy are better equipped to recognize the limitations of AI systems and to avoid inappropriate reliance on their outputs [92]. Conversely, insufficient understanding may lead to over-trust in AI-generated representations, particularly when outputs are fluent and coherent. Because generative outputs are often fluent, over-trust can arise even when accuracy is variable.

Trust calibration plays a central role in effective AI augmentation. Appropriate trust enables users to leverage AI support without surrendering critical judgment. Excessive skepticism may prevent users from benefiting from AI assistance, while excessive trust may lead to uncritical acceptance. Research on human–AI collaboration suggests that AI systems are most effective when they are perceived as cognitive partners rather than as authoritative decision-makers [75]. This creates a motivation for the later focus on reliance, confidence calibration, and source monitoring as mechanisms through which AI-generated representations can alter downstream decision quality and accountability.

Positioning AI within Knowledge-Intensive Work In roles such as product management, AI is frequently deployed to synthesize feedback, draft documentation, support roadmap planning, and structure strategic analysis. These tasks sit at the intersection of memory and decision processes: they require maintaining representations of prior assumptions and trade-offs while

integrating new evidence. By reshaping how information is externalized and accessed, AI systems influence not only task performance but also underlying cognitive processes such as memory encoding, retrieval, and confidence. These effects have implications that extend beyond individual cognition to decision-making and organizational outcomes.

This thesis therefore treats generative AI as a persistent representational layer that can support efficiency and coordination while also introducing epistemic risks when users substitute AI-generated representations for internally reconstructed knowledge. The following section builds on this foundation by examining AI explicitly as a form of cognitive support, focusing on distributed cognition and cognitive offloading.

2.2. AI as Cognitive Support

The growing presence of AI systems in knowledge-intensive work has prompted a reconceptualization of how cognition itself is understood in organizational contexts. Rather than treating AI as an external aid that merely accelerates task execution, recent perspectives emphasize its role in reshaping how cognitive processes are distributed between internal mental resources and external representational structures. Conceptualizing AI as a form of cognitive support allows for a more precise analysis of its effects on memory, learning, and decision-making.

From Internal Cognition to Distributed Cognition Classical cognitive models have traditionally located cognition within the individual mind, treating perception, memory, and reasoning as internal processes. External tools were typically considered facilitators of input or output but not as integral components of cognition itself. However, research in cognitive science has increasingly challenged this assumption by demonstrating that cognitive processes often emerge from interactions between individuals and their environments.

The theory of *distributed cognition* argues that cognition is not confined to the individual but is spread across people, artifacts, and representational systems [35]. In this view, external representations are not merely aids to cognition; they actively shape how cognitive tasks are performed by structuring information and guiding action. Empirical studies of complex systems—such as navigation and collaborative problem-solving—show that performance depends on how information is distributed across internal and external resources. Cognitive artifacts, such as charts, instruments, or written procedures, enable individuals to offload memory demands and coordinate actions over time; they do not simply store information but transform tasks by making certain operations easier and others less necessary [64].

The *extended mind hypothesis* further develops this argument by proposing that external resources can become part of the cognitive system itself when they are reliably and transparently integrated into task execution [18]. Under such conditions, the functional distinction between internal and external representations becomes less relevant than their role in enabling successful

cognition.

Generative AI extends this logic by producing artifacts that are adaptive and task-specific: the system can generate outlines, decision options, and draft rationales tailored to a user's goals. Over time, these artifacts can become a stable part of a workflow, shaping which information is repeatedly accessed and how it is organized for later reuse. This adaptivity has important implications for the distribution of cognition. When AI-generated representations are persistently available, users can rely on them as part of their cognitive workspace—reducing demands on working memory and enabling more complex reasoning, but also raising questions about how cognitive responsibility is shared between human and artificial components.

Cognitive Offloading and Epistemic Dependence External representations can offload cognitive work by reducing the need to store and manipulate information internally. *Cognitive offloading* refers to the strategic use of external artifacts to store information or to reduce internal processing requirements [71]. Offloading can be adaptive in environments where working-memory resources are limited and tasks are time-constrained [7].

However, offloading can also change encoding strategies: if information is easy to re-access, users may prioritize knowing *where* it is over learning *what* it is. The most visible modern example is internet search: when information is reliably accessible, people encode less of the content itself and more of the location or access strategy [80]. Meta-analytic evidence suggests that frequent search behavior is associated with reduced recall of content alongside improved memory for where information can be found [29]. These findings clarify an important trade-off: external availability can improve task performance while weakening internal knowledge structures if it displaces deep processing [20].

The introduction of AI systems intensifies this dynamic by extending offloading beyond simple information storage and retrieval. Generative AI systems not only provide access to information but also perform interpretative and integrative functions. As a result, users may offload not only memory but also aspects of reasoning and synthesis. When individuals repeatedly access information through external cues, they may become more dependent on those cues for successful retrieval, reducing their ability to reconstruct information independently. The concept of *epistemic dependence* captures this phenomenon: in AI-supported tasks, epistemic dependence may increase when users accept AI-generated representations as authoritative without critically engaging with them [93].

Crucially, the effects of cognitive offloading depend on how AI systems are used. When AI is employed to scaffold reflection, prompt elaboration, or encourage active engagement with information, offloading can support deeper processing rather than replace it. In contrast, passive consumption of AI-generated outputs may reduce cognitive effort and weaken memory formation. This distinction highlights the importance of interaction design and user behavior in determining cognitive outcomes. Later sections synthesize evidence on these trade-offs and

specify how the thesis operationalizes them via timing and structure manipulations.

AI as a Cognitive Partner in Knowledge-Intensive Work Unlike static artifacts, generative AI can behave like an interactive partner: it can propose interpretations, ask clarifying questions (in dialog settings), and generate alternative framings that influence sense-making. Research on human–AI teaming emphasizes that effective collaboration requires appropriate coordination, role allocation, and mutual understanding between human and artificial agents [75]. Rather than functioning as autonomous decision-makers, AI systems are most effective when they support human judgment by providing relevant information, alternative perspectives, or decision cues.

Even when the system is not interactive—as in Study 1's pre-generated summaries—its outputs can still function as "partner-like" cues that guide attention and reshape what is treated as central or peripheral. In knowledge-intensive roles, AI as a cognitive partner may influence not only how decisions are made but also how cognitive effort is distributed over time. By maintaining external representations of prior analyses, assumptions, or intermediate conclusions, AI systems can support continuity across decisions made at different moments. This motivates a mechanistic focus on how AI-generated representations alter attention allocation during encoding and cue availability at retrieval.

Implications for Studying Memory and Decision-Making Treating AI outputs as cognitive supports suggests that evaluation should go beyond accuracy of the output itself to include downstream cognitive and metacognitive outcomes. Key questions include whether AI support improves comprehension while weakening detail-rich encoding, whether it increases confidence without improving calibration, and whether it blurs source boundaries by mixing generated content with original material. These outcomes matter not only for individual learning but also for decision processes in organizations, where traceability, justification, and accountability depend on what people remember and how confidently they act on it.

2.3. Foundations of Human Memory

Knowledge-intensive work places sustained demands on human memory. Professionals operating in complex organizational environments are required to integrate information from multiple sources, maintain continuity across decisions made at different points in time, and recall prior assumptions, constraints, and rationales. Understanding the limitations and properties of human memory is essential for explaining both the appeal and the risks of AI-based cognitive support.

Theoretical Foundations of Human Memory The study of human memory has evolved through several major theoretical frameworks that describe how information is encoded, stored, and retrieved. The earliest comprehensive account is the *modal model* proposed by Atkinson and

Shiffrin [3], which conceptualizes memory as a system composed of three structural stores: sensory memory, short-term memory (STM), and long-term memory (LTM). Information enters through high-capacity but rapidly decaying sensory registers, is filtered into STM via attentional control, and may be consolidated into LTM through rehearsal and elaborative processing. This model introduced the influential distinction between temporary and durable memory systems and highlighted the central role of control processes—particularly attention and rehearsal—in regulating information flow.

A major theoretical shift occurred with Baddeley's reconceptualization of STM as *working memory* [7]. Rather than a passive buffer, working memory is understood as an active cognitive workspace composed of multiple subsystems (the phonological loop, visuospatial sketchpad, episodic buffer, and a central executive) that enable the integration of multimodal information. This framework supports reasoning, decision-making, and sustained mental effort. Working memory constraints are not merely quantitative but also qualitative, affecting the ability to manage competing demands and maintain task-relevant information. When tasks exceed working memory capacity, individuals must rely on external supports or risk cognitive overload.

Building on these models, Tulving introduced the distinction between *episodic* and *semantic* memory [87]. Episodic memory refers to the recollection of personally experienced, context-rich events, whereas semantic memory stores generalized knowledge about the world. In knowledge-intensive work, both systems are engaged: professionals draw on semantic memory when applying domain knowledge and on episodic memory when recalling prior decisions, discussions, and contextual constraints. Tulving further proposed the *encoding specificity* principle [88], which posits that retrieval is most successful when the cues available at encoding are reinstated at recall. This principle is central for understanding how external aids—including AI-generated summaries—may facilitate or distort later recall by changing which cues are available and salient.

Craik and Lockhart's *Levels of Processing* framework [20] emphasized that the durability of a memory trace depends not on which store it resides in, but on the *depth* and *semantic elaboration* of processing applied to the information. Deep, meaning-oriented encoding produces more persistent memory traces than shallow, perceptual processing—a finding repeatedly confirmed by subsequent behavioral and neuroscientific studies. In knowledge-intensive tasks, deep processing often requires time, attention, and active engagement—resources that are frequently scarce.

Neurophysiological research has extended classical models by illuminating the biological basis of encoding and retrieval. Single-neuron recording studies show that hippocampal and medial-temporal lobe neurons respond selectively to abstract concepts, forming building blocks of episodic and semantic representations [73]. Electrophysiological evidence further demonstrates that stronger neural activation during encoding predicts greater likelihood of later retrieval

success—the subsequent-memory effect [14].

Memory research has also embraced ecological and contextual perspectives. Studies using immersive virtual environments indicate that spatially rich, multisensory contexts enhance episodic encoding [68], while "real-life" neuroscience approaches argue that memory should be studied in dynamic, socially embedded settings rather than in isolation [76]. At a systems level, contemporary accounts emphasize that memory emerges from distributed interactions across cortical–hippocampal networks [61], and neuroimaging evidence suggests that retrieving a memory can reactivate cortical patterns present during encoding [48]. Taken together, these perspectives motivate two principles used throughout this thesis: (i) encoding depth and attentional control strongly shape memory outcomes, and (ii) memory is cue-dependent, reconstructive, and sensitive to external context.

Metacognition, Confidence, and Source Monitoring

Beyond structural limitations, effective use of memory in professional contexts depends on *metacognitive* processes: individuals' awareness and regulation of their own cognitive processes, including judgments about what they know and how confident they are in that knowledge. In decision-making contexts, metacognitive accuracy is essential for assessing the reliability of recalled information and for determining when additional information is needed.

Confidence judgments play a central role. Individuals routinely assess their confidence in memories and use these assessments to guide decisions. However, confidence is not always well calibrated to accuracy: individuals may be highly confident in inaccurate memories or uncertain about accurate ones, particularly when contextual cues are ambiguous or misleading [41].

A key mechanism underlying these phenomena is *source monitoring*. Source monitoring refers to the process by which individuals attribute memories to their origins, distinguishing between information derived from direct experience, external sources, or inference [37]. Accurate source monitoring is critical for maintaining ownership and accountability in decision-making, as it allows individuals to trace the provenance of their knowledge. Failures of source monitoring can lead to misattributions, such as confusing internally generated thoughts with externally provided information [47]. In organizational contexts, such misattributions may have significant consequences: decisions based on incorrectly attributed information may be defended with unwarranted confidence, obscuring responsibility and hindering learning from errors.

A complementary tradition in cognitive psychology emphasizes that remembering is not a verbatim replay but a *constructive* process shaped by schemas and expectations. In his seminal work, Bartlett [10] argued that recall reflects reconstruction guided by prior knowledge, which can yield distortions that are coherent but not faithful to the original experience. This reconstructive property is central for AI-assisted cognition: when users encounter fluent AI-generated representations, those representations can become integrated into the schema that guides later

reconstruction, increasing the risk that plausible but incorrect details are remembered as "having been in the source."

From a systems perspective, episodic remembering is distinguished from semantic knowledge because it carries contextual tags (where, when, and how information was acquired) that enable source attribution and confidence judgments. Reviews of episodic-memory mechanisms highlight that contextual reinstatement and source cues are key determinants of retrieval fidelity [37, 87]. In this thesis, this motivates treating source monitoring and misinformation susceptibility as first-class outcomes rather than as peripheral errors.

How Readers Learn from Expository Texts Humans do not memorize information in a vacuum; when reading expository texts (such as scientific articles or technical reports), they actively construct a mental representation of the content that integrates new information with prior knowledge. According to discourse comprehension theory, readers form a hierarchical *situation model* of a text, which entails establishing local coherence between consecutive ideas and global coherence across the entire discourse [89]. In building this situation model, the reader must identify the text's structure, discern relationships between key concepts, and continuously update their understanding as new information is introduced. This process places substantial demands on attention and working memory.

Two lines of learning research are particularly relevant for interpreting AI summaries in reading contexts. First, learning improves when readers actively organize and elaborate content rather than passively re-reading it, consistent with the Levels of Processing framework [20]. Second, even when verbatim recall does not increase, organizational supports (e.g., structured previews or graphic organizers) can improve conceptual integration and performance on applied questions [81].

Because of these cognitive demands, the way an expository text is structured—and the presence of guiding cues or previews—can significantly influence learning outcomes. Pre-reading aids, often termed *advance organizers* [5, 6], activate relevant schemas and highlight the organizational framework of the forthcoming content. Empirical studies find that students who receive a conceptual outline or summary before reading complex text often achieve better comprehension and organization of recall [33, 56]. Structural cues such as informative headings guide attention and improve memory for the organization of ideas [51].

Importantly, these aids often improve *integration* and comprehension even when total free recall of facts is unchanged [33, 81]. This distinction between organizational understanding and verbatim detail is relevant for evaluating AI-generated summaries: summaries may strengthen gist and coherence while potentially weakening detail-rich encoding.

Memory Constraints in Knowledge-Intensive Work The limitations of human memory have concrete implications for knowledge-intensive work. Professionals are often required to make decisions based on incomplete information, under time pressure, and with limited opportu-

nity for rehearsal or reflection. One critical challenge is the *temporal dispersion of decisions*: in many organizational settings, decisions are made incrementally over extended periods, and maintaining continuity requires recalling not only factual information but also the rationale behind prior choices. Episodic memory, however, is particularly vulnerable to decay and contextual mismatch, making such continuity difficult to sustain. Another challenge arises from *information overload*: the sheer volume of information encountered in knowledge-intensive roles exceeds the capacity of working memory, forcing individuals to prioritize and selectively encode information. These selection processes are influenced by attentional constraints and task demands, which may not align with future retrieval needs.

These constraints help explain the appeal of AI-based cognitive support systems. By externalizing information and maintaining persistent representations, AI systems can alleviate some of the burdens imposed by memory limitations. However, as subsequent sections show, this alleviation comes with trade-offs that affect how information is encoded, retained, and retrieved. Because the thesis links cognitive constructs (encoding, retrieval, source monitoring) to organizational practices (decision workflows, governance), it requires a coherent mapping between levels and paradigms. This motivates the thesis's multi-method strategy: rather than assuming one-to-one equivalence between laboratory tasks and professional practice, the thesis uses Study 1 as mechanistic evidence and Study 2 as contextual evidence and integrates them via the AI Buffer Model.

2.4. Cognitive Offloading and External Representations

External representations are a long-standing feature of human cognition: people routinely use notes, outlines, and digital tools to reduce memory demands and to coordinate complex tasks. Offloading can be adaptive in environments where working-memory resources are limited and tasks are time-constrained [7, 71].

The most visible modern example is internet search. When information is reliably accessible, people encode less of the content itself and more of the location or access strategy [80]. Meta-analytic evidence suggests that frequent search behavior is associated with reduced recall of content alongside improved memory for where information can be found [29]. These findings clarify an important trade-off: external availability can improve task performance while weakening internal knowledge structures if it displaces deep processing [20].

This trade-off has motivated broader arguments that pervasive digital scaffolding may reshape cognition over time by changing the balance between internal memory and external retrieval habits. The "online brain" perspective synthesizes evidence that ubiquitous access technologies can shift attentional strategies and memory reliance [27, 71, 80]. While such shifts can be adaptive in complex environments, they can also create dependence on external systems

for knowledge continuity and reduce opportunities for effortful encoding and self-generated retrieval.

Generative AI extends offloading beyond retrieval by producing *transformed* representations (summaries, explanations, and drafts). This transformation can function as a scaffold for comprehension and decision-making, but it also changes source-monitoring conditions: when AI-generated content is fluent and plausible, users may later confuse its origin or accept it with insufficient scrutiny [37, 93]. Accordingly, the effects of generative AI cannot be understood solely as "more information access"; they depend on how AI representations are timed, structured, and integrated into the user's workflow.

2.5. Cognitive Effects of AI-Generated Representations on Memory

The integration of AI systems into knowledge-intensive work has introduced new forms of interaction between humans and external representations. Unlike traditional informational tools, AI systems—particularly generative AI—do not merely store or retrieve information but actively generate representations that shape how information is processed. These representations influence memory at multiple stages, including encoding, maintenance and consolidation, and retrieval. Understanding these effects is essential for assessing both the benefits and risks of AI-supported cognition. This section synthesizes findings from cognitive psychology, learning sciences, and recent studies on AI-assisted tasks to analyze how AI-generated representations affect human memory processes.

Effects on Encoding: Attention Allocation and Depth of Processing Memory encoding refers to the processes through which information is initially processed and transformed into a form that can be stored in long-term memory. Encoding quality is a critical determinant of later recall and understanding. Cognitive research has consistently shown that encoding is not automatic; it depends on attentional allocation, elaboration, and depth of processing [20].

AI systems can influence encoding by restructuring information and guiding attention toward specific elements. Summaries and structured outlines generated by AI may reduce the effort required to identify key points, thereby facilitating initial comprehension. Research on advance organizers suggests that providing structured representations before or during learning can enhance encoding by activating relevant schemas and directing attention to important relationships [5, 6]. Empirical studies on instructional design support this view: pre-instructional strategies such as headings or conceptual overviews improve recall by helping learners organize incoming information [33, 56], and structured cues embedded in texts can guide readers toward deeper processing of relevant content [51].

However, the benefits of such support are not unambiguous. While structured representations can facilitate deep processing, they may also reduce the need for users to actively construct their own mental models. If AI-generated summaries *replace* rather than *scaffold* elaborative processing, users may engage in shallower encoding. This risk is particularly salient when AI outputs are consumed passively, without critical engagement.

Research on generative AI in learning contexts illustrates this tension. Bai et al. [8] found that students who interact with a large language model by asking questions and explaining answers can achieve better learning results than those who study without AI. The dialogue can prompt deeper semantic processing, consistent with the Levels of Processing framework. Similarly, Haider et al. [32] report improvements in recall and problem-solving in AI-assisted learning settings. However, when users rely heavily on AI-generated explanations without active engagement, they may achieve higher immediate performance but exhibit weaker long-term retention, consistent with reduced depth of processing. Complementing behavioral evidence, preliminary neuroscience work suggests that AI-assisted writing can modulate neural engagement associated with attention and semantic processing [60].

Thus, AI systems can both enhance and impair encoding. They enhance encoding when they scaffold attention and encourage elaboration; they impair it when they substitute for active cognitive engagement.

Effects on Maintenance and Consolidation Once information has been encoded, its retention depends on processes of maintenance and consolidation. Maintenance involves keeping information active in working memory through rehearsal, while consolidation refers to the stabilization of memory traces in long-term memory over time. Both processes are sensitive to cognitive load and attentional resources.

Working memory is severely capacity-limited, and excessive cognitive load can impair both maintenance and consolidation [7]. Cognitive load theory emphasizes that learning and retention are optimized when extraneous load is minimized, allowing resources to be allocated to germane processing [83]. AI systems can reduce cognitive load by externalizing information that would otherwise need to be maintained internally, freeing cognitive resources for higher-level reasoning and integration.

However, the externalization of rehearsal also introduces trade-offs. When rehearsal is delegated to external systems, internal consolidation processes may be weakened. Research on split attention and redundancy effects suggests that learning can be impaired when learners rely too heavily on external representations without integrating them into internal memory structures [69]. Similarly, passive interaction with instructional materials has been shown to result in weaker retention compared to active engagement [81]. Active learning research further supports this distinction: engaging learners in activities that require explanation, retrieval, or application leads to stronger memory consolidation than passive exposure to information [25].

AI systems that encourage active interaction—such as prompting users to refine, critique, or extend generated outputs—may therefore support consolidation more effectively than systems that simply present finished representations. The Memoro system, for instance, offers a real-time dialogue interface that prompts users to restate and verify information during study [59]. This resembles retrieval practice: periodic prompting and feedback can strengthen retention by reactivating information and integrating it with existing knowledge. AI systems can also support *metacognition*—the monitoring and regulation of one's own cognitive processes. Sun et al. [82] show that generative AI can improve metacognitive calibration in creative tasks by prompting evaluation of alternatives and reflection on uncertainty. In learning contexts, similar prompts may improve confidence judgments by making uncertainty explicit and by encouraging verification behaviors.

Effects on Retrieval: Cueing and Dependence Memory retrieval involves accessing stored information using available cues. Successful retrieval depends on the overlap between encoding and retrieval contexts, as described by the encoding specificity principle [88]. External cues can facilitate retrieval by reinstating aspects of the original encoding context.

AI systems can serve as powerful retrieval cues by re-presenting information in structured or familiar formats. By maintaining external representations of prior analyses or decisions, AI systems reduce the need for internal reconstruction during retrieval. This cueing logic suggests that buffer access can improve recognition of summary-aligned propositions even if free recall remains unchanged. At the same time, reliance on external cues may reduce the need to reconstruct information from internal traces, which could weaken detail-rich recall when the buffer is unavailable or when retrieval requires reconstruction beyond the buffer's coverage.

Research on cognitive offloading indicates that individuals who frequently rely on external systems for retrieval may become less able to recall information independently [80]. Studies on digital information access suggest that repeated reliance on external cues can reshape retrieval strategies, making individuals more dependent on the availability of those cues [29]. Generative AI amplifies these dynamics by providing not only retrieval cues but also reconstructed representations of information. While this can improve immediate performance, it may weaken internal retrieval pathways and reduce the robustness of memory.

The trade-off between immediate accessibility and long-term robustness is central to understanding AI's impact on retrieval. By maintaining task-relevant representations in an accessible form, the buffer stabilizes recall across time and context; however, externally supported recall introduces dependence that may prove fragile when external representations are unavailable or inaccurate.

Risks: Benefits, Costs, and the Broader Cognitive Landscape The benefits of AI assistance are coupled with significant risks. One concern is that AI tools can encourage users to invest less effort in understanding and remembering, since fluent answers and summaries are available

on demand. This may produce an *illusion of competence*: a user appears capable with AI support but has not internalized the information [65]. Such risk is consistent with the offloading literature and with theories that emphasize deep processing and active engagement for durable memory [7, 20].

At a broader level, pervasive AI tools may shift cognitive habits by reallocating effort from memorizing details toward managing interfaces and interpreting AI outputs [27]. AI can also shape the emotional and contextual aspects of memory: AI-curated experiences may heighten arousal or personalization, which can strengthen certain memories but can also narrow exposure and reduce diversity of encoded experiences [11]. Understanding when AI functions as a beneficial extension versus a detrimental replacement of human memory is a central challenge; this thesis approaches the problem by examining design-sensitive mechanisms, particularly timing and structure of AI-generated summaries.

2.6. Confidence, Source Monitoring, and Cognitive Load in Assisted Cognition

AI-generated summaries can alter not only what is learned, but also the conditions under which confidence is formed and sources are discriminated. This section synthesizes three theoretical strands that motivate the key design manipulations in Study 1: (i) cognitive load and split attention during learning, (ii) how timing of access changes encoding and cue availability, and (iii) how source-monitoring demands shape vulnerability to plausible but incorrect information.

Foundational Theories Motivating Timing and Structure

The experimental manipulations in this thesis are grounded in well-established findings in the learning sciences. The AI-generated summary functions as a compact surrogate for an article's key propositions. The *timing* of access to this summary is manipulated to test whether seeing a summary *before* reading helps to organize and scaffold encoding of the full text, as opposed to seeing it mid-way or only after reading. The *structure* of the summary is manipulated to test whether presenting it as an integrated paragraph versus as segmented bullet points affects coherence building and source attributions.

Advance organizers and pre-reading scaffolds. Ausubel's theory of advance organizers predicts that providing a high-level conceptual framework *prior* to learning enhances integration of new knowledge by activating relevant schemas and guiding attention during encoding [5, 6]. Evidence on pre-instructional strategies shows that previews (e.g., outlines or summaries) can improve comprehension and organization [33, 56], and headings can guide attention and improve recall of structure [51]. This suggests that pre-reading access to AI summaries may be especially beneficial when the summary acts as an encoding scaffold rather than as a post-hoc reminder.

Mid-reading interruptions and divided attention. Presenting an AI summary synchronously may introduce costs via task-switching and disrupted coherence: readers may need to pause, reorient, and rebuild a situation model. Task interruption and divided attention can impose extra working-memory demands, potentially weakening integration of information and increasing reliance on surface-level cues [7].

Cognitive load, split attention, and integrated presentation. Cognitive load theory emphasizes limited working-memory resources and predicts that formats that minimize extraneous processing demands yield better learning [83]. The split-attention effect shows that when related information is separated, learners must expend effort to integrate it, which can impair learning [16]. Applied evidence confirms that integration vs. separation can produce measurable performance differences [69]. In this context, an integrated paragraph summary may reduce unnecessary scanning and facilitate verification against the source text, whereas segmented bullet points may encourage piecemeal processing and increase coordination demands.

Memory Distortion and Epistemic Risk

The Source Monitoring Framework [37] emphasizes that remembering involves determining where information came from (e.g., "Did I read this in the article, or did it come from the summary?"). Source attributions rely on contextual and qualitative features of a memory; when different sources overlap and memories lack distinctive details, misattribution becomes more likely. The *misinformation effect* illustrates this risk: misleading post-event information can be incorporated into memory and later misattributed to the original event [49, 50]. Related work shows that suggestibility increases when sources share overlapping details and contextual tags are weak [47].

Beyond individual paradigms, reviews of misinformation and memory distortion emphasize that false memories are not rare anomalies but systematic outcomes of normal reconstructive processes under conditions of suggestion, source-confusability, and fluency [15, 49]. For AI-assisted cognition, this is a critical framing: when AI outputs are plausible and presented in ways that blur provenance cues, they can function as post-event information that competes with or overwrites source material in memory.

False-memory research further underscores how gist-based representations can generate confident errors. In the DRM paradigm, people falsely recall theme-consistent lure words [72]. Fuzzy Trace Theory explains this by positing that both gist and verbatim traces are encoded, but reliance on gist can dominate when verbatim traces are weak [12]. Summaries emphasize gist and can thus increase vulnerability to gist-consistent false recognitions, especially if AI output is plausible and fluent. This motivates measuring epistemic risk (e.g., false-lure endorsement) in Study 1 and motivates design choices that support provenance discrimination.

AI-generated representations may exacerbate these vulnerabilities. Generative AI systems can produce fluent and coherent outputs that appear authoritative, even when they contain inac-

curacies or hallucinations. When users integrate such outputs into their understanding, they may inadvertently encode false information, leading to distorted memories. Recent studies on AI-assisted tasks indicate that users may develop inflated confidence in AI-supported knowledge, even when accuracy is compromised [27, 93]. This combination of fluency, confidence, and source confusion represents a significant cognitive risk, particularly in high-stakes decision-making environments.

2.7. From Individual Memory to Decision-Making

Decision-making in knowledge-intensive work is deeply intertwined with human memory. Decisions rarely rely on immediate perception alone; rather, they depend on the ability to recall prior information, integrate past experiences, and assess the reliability of available knowledge. As such, memory processes—particularly encoding quality, retrieval reliability, and metacognitive awareness—constitute foundational inputs to decision-making. This section bridges the cognitive foundations reviewed above to the decision-level dynamics that the thesis examines.
Memory Constraints and Decision Quality Classical theories of decision-making often assume that individuals can access relevant information and evaluate alternatives rationally. However, the concept of *bounded rationality* captures a more realistic insight: decision-makers operate under constraints of limited attention, memory, and computational capacity [78]. Rather than optimizing, individuals *satisfice*, relying on heuristics and incomplete representations of complex problems.
Memory limitations play a central role in these constraints. Decisions are shaped not by objective information sets, but by what decision-makers can recall and reconstruct at the moment of choice. When relevant information is poorly encoded or difficult to retrieve, decision quality suffers regardless of the availability of external data. The effort required to retrieve and integrate information further influences outcomes: individuals tend to rely on information that is more readily accessible, even when it is not the most relevant [41]. This *accessibility bias* means that memory salience, rather than objective importance, often guides decision-making. Consequently, external representations that increase accessibility can have a disproportionate impact on decisions.
AI systems can alter this dynamic by reshaping information accessibility. By externalizing information and providing structured representations, AI reduces the cognitive effort required to recall and integrate information. This reduction may improve decision quality by enabling decision-makers to consider a broader range of relevant factors. At the same time, it may also bias decisions toward information emphasized by AI-generated representations [95].
Decision Speed and Cognitive Offloading In addition to decision quality, memory constraints influence decision speed. Knowledge-intensive work often requires timely decisions under uncer-

tainty, where prolonged deliberation may be impractical. AI systems can significantly enhance decision speed by providing readily accessible summaries, recommendations, or prior rationales, reducing the time required to reconstruct decision contexts.

However, faster decisions are not always better decisions. The reduction of cognitive effort may come at the cost of reduced deliberation, increasing the risk of superficial reasoning. The distinction between fast, intuitive processing and slow, deliberative processing highlights this trade-off [41]. AI-supported decision-making may inadvertently favor faster, less reflective modes of reasoning if users rely uncritically on AI outputs. The balance between speed and deliberation is particularly important in organizational contexts, where decisions often have long-term implications.

Decision Ownership, Confidence, and Accountability Beyond quality and speed, decision-making involves issues of ownership and accountability. Decisions are not merely cognitive outcomes; they are socially situated acts for which individuals and organizations must take responsibility. Memory plays a critical role by supporting confidence judgments and source attribution.

Confidence in decisions often reflects confidence in the underlying knowledge. When decision-makers are uncertain about the provenance or reliability of recalled information, their confidence may be miscalibrated. Metacognitive processes such as source monitoring are therefore essential for maintaining decision ownership [37]. Accurate source monitoring allows individuals to distinguish between information derived from personal experience and information obtained from external sources.

AI systems complicate this process by introducing new sources of information that are neither fully internal nor fully external in traditional terms. AI-generated representations may be integrated into users' understanding in ways that blur the distinction between human judgment and machine-generated input. Over time, this blurring can lead to source confusion, making it difficult to attribute decisions to specific informational origins. If AI is perceived as an authoritative source, decision-makers may defer judgment and attribute responsibility to the system, undermining accountability. Conversely, if AI is treated as a cognitive partner rather than a decision-maker, human ownership of decisions can be preserved.

Maintaining decision ownership in AI-augmented environments thus requires explicit attention to cognitive and organizational design. Users must remain aware of the role AI plays in shaping their decisions, and organizations must establish norms that clarify responsibility. These considerations form a critical link between individual cognition and organizational governance.

2.8. Human–AI Decision-Making (Trust, Reliance, Bias, Accountability)

When AI outputs support decisions, the central behavioral question becomes not only *accuracy* but also *calibrated reliance*: users must decide when to accept an AI representation, when to verify it, and when to disregard it. In educational contexts, over-reliance can reduce active engagement and long-term skill development [93]. In professional contexts, similar dynamics can create a dependence loop in which decision-makers defer to AI recommendations even when they could reason independently, producing a knowledge imbalance and increasing epistemic risk [95]. These dynamics are more likely when AI outputs are fluent, time pressure is high, and verification costs are non-trivial.

Work on AI assistants and user interpretations further suggests that interface framing can shape expectations about agency and competence, influencing both trust and perceived responsibility—a dynamic that becomes more salient as AI outputs are embedded into workflows and presented with conversational fluency [43].

Importantly, "trust" and "reliance" should be distinguished. Users may express general trust in AI systems while still choosing not to rely on them in a given task, or they may rely heavily because the workflow makes verification costly even when trust is low. In assisted cognition, reliance becomes a form of *resource allocation*: deciding whether to invest effort in independent reasoning, to cross-check sources, or to offload to an external representation [71]. This implies that improving outcomes is not only about increasing AI accuracy, but also about shaping the conditions of calibrated reliance (feedback, provenance cues, and low-friction verification).

Reliance is also mediated by skill and literacy. Because generative AI outputs are probabilistic and can contain plausible errors, effective use requires AI literacy: the ability to interpret outputs, recognize uncertainty, and choose appropriate validation strategies [91]. Without such literacy, users may adopt uncritical patterns of use (e.g., accepting fluent summaries as ground truth) or may underuse AI due to uncertainty about when it is appropriate. This thesis therefore treats reliance not as a stable attitude, but as a dynamic interaction between task constraints, user competencies, and representational design.

Reliance is also shaped by how AI is embedded in team settings. In organizational decision-making, the relevant unit is often not the individual user but the *human–AI–team* system: AI outputs are interpreted, debated, and propagated through artifacts that multiple people consume. Conceptual and empirical work on combining human and artificial intelligence for strategic decisions highlights both complementarities (speed of synthesis, scenario exploration) and coordination risks when AI recommendations are treated as authoritative without scrutiny [39, 40, 86]. Research on human–AI collaboration emphasizes that effective collaboration re-

quires appropriate coordination, role allocation, and mutual understanding between human and artificial agents; AI systems are most effective when they support human judgment rather than replace it [75]. In collaborative work, the benefits of AI assistance can also depend on experience and role configuration: evidence on human–AI teaming suggests that workers with different levels of experience may collaborate with AI differently, which can shape whether AI augments judgment or induces over-reliance [92].

Reliance is also shaped by accountability and transparency constraints. If an AI system influences decisions but its rationale is opaque, organizations may struggle to audit errors, detect bias, or assign responsibility. The "black-box" problem is therefore not only a technical issue but a governance issue: opaque systems can encourage uncritical acceptance while making failures hard to diagnose [67]. For roles that mediate between stakeholders and technical systems—such as product management—this elevates the importance of provenance cues, documentation practices, and norms of verification when AI-generated summaries and recommendations enter decision workflows [31, 34, 63].

Finally, human–AI decision-making has an explicit ethical dimension. When AI representations influence judgments, errors can be difficult to detect and responsibility can be difficult to assign, especially when AI content is reused and propagated through documents, dashboards, and presentations. Accountability is not only about explaining model outputs; it is also about maintaining a decision trace that links claims to sources and indicates which parts of a decision were AI-mediated [67]. This is directly relevant for roles like product management, where decisions must be justified to stakeholders and later revisited.

2.9. AI in Professional and Managerial Contexts

AI–Augmented Memory in Organizations and Applied Systems Beyond individual cognition, the integration of AI into memory processes has significant implications for organizations and knowledge-intensive work. Enterprises increasingly rely on digital ecosystems in which AI systems support information recall, contextual understanding, and knowledge retrieval, effectively functioning as external, collective memory infrastructures.

AI tools can act as cognitive extensions in professional environments by organizing personal and collective knowledge into searchable, semantically structured archives. By surfacing prior meeting notes, decision histories, or relevant research exactly when needed, such systems externalize elements of working memory and help maintain continuity in fast-paced workflows [27, 59]. This extends cognitive offloading [71, 80] into organizational settings where information fragmentation is a constant challenge.

This perspective aligns with work on AI-assisted knowledge management, which emphasizes complementarity between AI systems and human expertise [66]. AI can rapidly organize and

filter information, but it is most effective when guided by humans who provide context and interpretive insight. Together, they can sustain an organizational memory that is more comprehensive and accessible than unaided human memory, yet more flexible and meaningful than data left to accumulate without human interpretation.

Applied AI–memory systems illustrate how algorithmic tools can function as external supports for encoding, retrieval, and metacognitive regulation in real-world settings. In healthcare and neurotechnology, AI-guided stimulation approaches have been shown to enhance episodic recall in clinical contexts [45]. In education, intelligent tutoring systems improve learning outcomes by tailoring practice and feedback [2, 53], and active learning improves durable retention relative to passive reading [25]. Generative AI tutors can leverage these principles by sustaining reflective dialogue and prompting retrieval and self-explanation [1]. In everyday settings, augmented reality interfaces are emerging as a platform for on-demand contextual recall; the "AR Secretary Agent" combines wearable AR glasses with an LLM-based assistant to provide real-time memory augmentation [24].

AI Adoption, Digital Transformation, and Role Reconfiguration At the organizational level, AI is increasingly framed as a general-purpose technology that reshapes business processes, role boundaries, and strategic decision-making rather than as a point solution. Strategy and management research discusses how AI can be combined with human judgment in organizational decisions, potentially improving speed and breadth of analysis while creating new coordination and accountability challenges [42, 46, 86]. AI is increasingly treated as a strategic capability for competitive positioning and business model innovation [36]. Industry reports similarly emphasize that value creation depends on how organizations "rewire" processes and capabilities around AI deployment rather than on model access alone [58].

This transformation is visible across business functions, where AI is adopted for customer-facing and back-office work. Practitioner overviews document AI use in customer support and engagement [22], operations management [57], financial services [17], scenario analysis in corporate finance [9], and human-resource management practices [62]. In many cases, adoption is paired with automation technologies such as RPA and workflow tooling, which shift the division of labor by externalizing routine decision steps into software artifacts [23, 55].

Such changes create new demands for organizational learning and training, including "hands-on" AI skill development in business education and professional contexts [26, 91]. Reviews and agendas highlight AI's role in strategic decision-making and the emergence of new business-model configurations enabled by algorithmic capabilities [4, 19, 38–40, 79].

AI-Augmented Organizational Memory and Knowledge Management

While previous sections have focused on individual and role-level cognition, the implications of AI-supported cognition extend to the organizational level. Organizations rely on memory to ensure continuity, coordination, and learning over time. Decisions, routines, and strategies

are embedded in documents, processes, and shared representations that collectively constitute *organizational memory*—the stored information from an organization's history that can be brought to bear on present decisions [90]. Unlike individual memory, organizational memory is not bounded by biological constraints, but it is shaped by how information is encoded, stored, and retrieved within organizational structures.

AI systems introduce new possibilities for enhancing organizational memory by enabling semantic indexing, contextual retrieval, and dynamic summarization of organizational knowledge. By generating structured representations of past decisions, discussions, and outcomes, AI can support continuity by making prior knowledge more accessible and interpretable. Recent research suggests that AI-enabled systems can improve knowledge retrieval and reuse by aligning information with users' current needs [66]. In this sense, AI functions as an active mediator of organizational memory rather than as a passive storage mechanism.

This generative capacity positions AI as an active participant in organizational memory formation. Rather than merely storing past information, AI systems influence how organizational history is interpreted and recalled. Over time, repeated reliance on AI-generated representations may shape shared understanding and norms within the organization. Just as individuals rely on external artifacts to support memory, organizations may rely on AI systems to maintain continuity and coherence. While this reliance can enhance efficiency and reduce knowledge loss, it also introduces dependencies that may affect organizational learning and adaptability.

A key concern is *epistemic vulnerability*: when AI-generated representations become central to organizational memory, inaccuracies or omissions may propagate across decisions. Because AI outputs often appear coherent and authoritative, errors may be difficult to detect, particularly when original sources are not revisited. Algorithmic opacity exacerbates these challenges: AI systems often operate as black boxes, making it difficult for users to understand how representations are generated or which sources are prioritized [67]. Research on AI governance emphasizes the need for clear norms and structures that define how AI systems are used and how responsibility is allocated [21]. Effective governance requires transparency regarding AI contributions, mechanisms for auditing AI-generated outputs, and training that fosters critical engagement rather than passive reliance. Maintaining redundancy and human oversight is therefore essential for mitigating epistemic vulnerability.

Best practices in AI-augmented knowledge management include designing systems that preserve links to original sources, encourage user reflection, and support traceability of decision rationales. Moreover, fostering AI literacy at the organizational level is critical: users must understand the strengths and limitations of AI systems to effectively integrate them into knowledge workflows [91].

2.10. Product Management as a Decision-Intensive Domain

Product management is a decision-intensive role that integrates heterogeneous evidence (user feedback, market signals, analytics, technical feasibility, and stakeholder constraints) into product choices under uncertainty and time pressure. In agile settings, product managers coordinate discovery and prioritization work while aligning cross-functional teams around goals and trade-offs [85]. These responsibilities make the role highly dependent on *external representations*: roadmaps, requirement documents, research summaries, and internal knowledge bases stabilize collective understanding over time.

The Product Manager Role as an Information-Integration Problem The core challenge of product management lies in integrating diverse and often conflicting information streams. PMs must synthesize inputs from engineering teams, user research, market analysis, business strategy, and organizational leadership. Classic accounts emphasize the PM's responsibility for maintaining a coherent product vision while navigating trade-offs among competing priorities [13]. This responsibility requires not only technical understanding but also the ability to recall prior decisions, assumptions, and rationales. Continuity over time is critical: decisions made at one stage of product development often constrain future options, and failure to remember or reconstruct earlier reasoning can lead to inconsistency and suboptimal outcomes.

From a cognitive perspective, this continuity places significant demands on episodic and semantic memory. These demands are exacerbated by the fragmented and fast-paced nature of product work. The PM role therefore illustrates how memory constraints translate into organizational challenges: when information is poorly integrated or when decision rationales are lost over time, organizations risk repeating mistakes, misaligning teams, or pursuing inconsistent strategies.

AI Use-Cases in Product Management From a decision-process perspective, product management can be described as a recurring cycle of (i) problem framing (defining user needs and business objectives), (ii) option generation (alternative solutions and feature sets), (iii) prioritization under constraints (engineering capacity, time-to-market, risk), and (iv) evaluation and iteration after release. Each step relies on artifacts that function as shared memory objects: PRDs, user stories, sprint backlogs, dashboards, meeting notes, and stakeholder narratives. These artifacts do not merely store information; they coordinate attention, align mental models across functions, and preserve rationale so that decisions remain revisitable when assumptions change.

This artifact-centric nature makes product management a sensitive domain for AI-generated external representations. If AI becomes the first producer of the artifacts that others rely on—

summaries of research, synthesized user feedback themes, competitive analyses, and executive-ready narratives—then AI outputs can shape the initial frame of the decision problem and thereby influence downstream choices even when humans remain nominally in control.

Generative AI is increasingly positioned as an additional representational layer in product-management workflows, supporting tasks such as summarizing customer feedback, drafting product narratives, and accelerating synthesis across fragmented information sources [31, 34, 63]. Practitioner and industry sources describe how AI is inserted into the everyday artifact pipeline of product work: automating documentation and coordination tasks (e.g., drafting PRDs, writing user stories, generating meeting summaries, and preparing stakeholder updates), as well as use cases in discovery and synthesis (e.g., clustering feedback themes, proposing hypotheses, and generating experiment variants). In software-heavy settings, analyses of generative AI in product development highlight potential acceleration of time-to-market and changes in cross-functional collaboration, as AI-generated artifacts flow between product, design, and engineering teams [28].

These uses are best understood as a mix of *automation* (reducing repetitive synthesis work) and *augmentation* (supporting judgment with condensed representations). While automation benefits can be captured through time saved, augmentation benefits depend on whether the representations preserve key assumptions, uncertainties, and trade-offs that support judgment. This distinction motivates Study 2's focus on both perceived efficiency gains and perceived decision-making changes (Section 6.2).

However, AI adoption also shifts part of the PM role toward *evaluation and governance* of AI outputs: product managers may need to detect when AI-generated summaries are incomplete or biased, maintain provenance cues, and ensure that accountability for decisions remains clear [67]. At the same time, because product decisions are accountable to multiple stakeholders, AI-generated representations can create organizational risks that mirror cognitive risks. If AI outputs are reused across documents without clear provenance, errors can propagate and responsibility can diffuse. This makes governance mechanisms (documentation, traceability, and explicit validation steps) part of the representational system itself rather than an external compliance layer.

Cross-Cultural Adoption and Organizational Context AI adoption in product management does not occur in a vacuum; it is shaped by organizational and cultural contexts. Differences in risk tolerance, governance practices, and attitudes toward automation influence how AI systems are integrated into workflows. Cross-cultural studies suggest that variations in organizational norms and managerial practices can lead to different patterns of AI use: cultures that emphasize efficiency and standardization may be more inclined toward automated workflows, while those that prioritize individual judgment may favor augmentation-oriented approaches. These differences have implications for how cognitive offloading and epistemic dependence manifest

in practice.

AI literacy plays a moderating role in these dynamics. Users with higher AI literacy are better equipped to critically evaluate AI outputs, calibrate trust, and integrate AI support into decision-making processes without relinquishing control [91, 92]. Conversely, limited AI literacy may exacerbate over-reliance or misuse, particularly in high-pressure decision environments. Because product management combines heavy information processing with high-stakes communication and justification, it provides a natural applied setting for studying how AI-generated representations shape both cognition and reliance.

These accounts align with a broader business-operations framing in which AI supports both operational efficiency and strategic exploration [4, 84]. For product managers, the implication is that AI becomes part of the decision infrastructure: it can influence which options are generated, which evidence is highlighted, and how trade-offs are communicated, with downstream implications for responsibility allocation and governance when AI-generated content is reused across the organization. Organizational maturity, regulatory expectations, and cultural norms can shape how much AI is used, which tasks are automated versus augmented, and how governance is implemented. This motivates explicit cross-context comparison within Study 2 (e.g., USA–China) as part of RQ2's scope.

2.11. Synthesis and Research Gap

Across cognitive and organizational perspectives, prior work converges on a common pattern: external representations can enhance performance by providing cues and reducing load, but they can also induce cognitive offloading and source confusion when they displace deep processing [20, 37, 71, 80]. Generative AI intensifies this trade-off because it produces fluent, transformed representations whose provenance may be ambiguous and whose errors can be incorporated into memory and judgment [15, 93]. In organizations, AI-augmented knowledge management promises continuity and efficiency [66], yet can create epistemic dependence and accountability challenges when AI outputs shape professional decisions [67, 95].

The Cognitive Gap Many cognitive studies operationalize "AI assistance" as a coarse treatment (AI vs. no AI) without testing how representation *design features*—timing, structure, and provenance cues—shape outcomes. This makes it difficult to derive actionable design guidance: if outcomes depend on *how* AI is embedded into the task, then binary comparisons can obscure mechanisms and lead to inconsistent conclusions. Moreover, performance benefits are often reported without parallel measurement of epistemic risk (e.g., misinformation acceptance, source confusions, or miscalibrated reliance), even though these risks may be more consequential in decision settings than small changes in accuracy. There is limited understanding of how specific configurations of AI support causally affect memory encoding, consolidation, retrieval,

confidence, and source monitoring.

The Organizational Gap On the organizational side, the emerging literature often documents adoption, perceived productivity gains, and general attitudes, but less frequently connects observed practices to cognitive mechanisms. Research on AI adoption in managerial and professional roles frequently relies on perceptual or performance-based measures, such as self-reported productivity gains or role changes, while abstracting away from the cognitive mechanisms that drive observed outcomes. Questions such as how AI-generated representations affect decision continuity, accountability, and organizational memory over time, or how they influence the preservation and reuse of decision rationales, remain underexplored. Without addressing these issues, organizational research risks treating AI as a black box that produces outcomes without explaining the processes through which they arise.

The Integrative Gap The most significant gap is integrative in nature. Cognitive studies rarely connect interface-level design choices (timing, structure, provenance cues) to professional reliance patterns, while organizational accounts rarely link observed AI use to mechanisms such as cue-dependent retrieval and source monitoring. Although cognitive and organizational perspectives on AI have developed in parallel, they are rarely connected through a unified framework. This separation obscures the pathways through which AI-supported cognition influences organizational performance and governance. Bridging this gap requires a multi-level approach that explicitly links cognitive mechanisms—such as memory encoding, retrieval reliability, confidence, and source monitoring—to decision-making processes and organizational outcomes.

Positioning of the Thesis This thesis addresses these gaps with a multi-level design: Study 1 provides causal evidence on how timing and structure of AI-generated summaries affect memory and epistemic risk in a controlled reading task, while Study 2 characterizes how product managers deploy AI-generated representations in decision-intensive work [31, 34, 63, 85]. These empirical investigations are unified by an integrative theoretical framework that connects AI systems to cognitive buffering mechanisms and decision-making outcomes. By explicitly linking micro-level cognitive processes to meso-level organizational dynamics, the thesis contributes to a more comprehensive understanding of AI augmentation in knowledge-intensive work. The next chapter introduces the *AI Buffer Model* as an integrative framework to connect these levels and to derive propositions for safer, more effective AI-assisted cognition and decision-making.

3 | Conceptual Framework and Hypotheses

This chapter develops the integrative theoretical framework that connects the cognitive and organizational levels of analysis in this thesis. Building on the literature review, it articulates a multi-level structure in which distinct components operate at different levels of analysis while remaining tightly interrelated.

The chapter proceeds in four main steps. First, it positions AI-generated external representations as a form of cognitive infrastructure (Section 3.1). Second, it introduces the *AI Buffer Model*, which formalizes the cognitive mechanisms through which AI-generated representations interact with human memory during task execution (Section 3.2). Third, it traces how these cognitive mechanisms shape decision-making processes and organizational outcomes through the *AI-Augmented Decision Framework* (Sections 3.3 and 3.4). Finally, it derives testable hypotheses and propositions aligned with RQ1–RQ3 (Section 3.6).

Without an explicit account of how AI-induced cognitive changes translate into decision-making dynamics, organizational analyses risk treating AI as a black box that produces outcomes without explaining the mechanisms that generate them. By making explicit the links between cognition, decision-making, and organization, this chapter provides the conceptual foundation for the empirical studies presented in the subsequent chapters.

3.1. Positioning: AI-Assisted External Representations as Cognitive Infrastructure

Across domains, humans use external representations to stabilize information, reduce memory demands, and coordinate complex tasks. In cognitive terms, this practice is often described as *cognitive offloading*—the strategic use of external artifacts to store information or reduce internal processing requirements [71]. Digital search technologies provide an emblematic case: when information is reliably accessible, people may encode less of the content itself and more of the access strategy [80]. This does not imply that offloading is inherently harmful; rather,

it highlights that cognition adapts to the structure of the environment and the availability of external resources.

Generative AI extends external representation beyond retrieval by producing *transformed* representations: summaries, explanations, and recommendations that compress, reorganize, and reframe source material. Compared with notes or static documents, these representations are often fluent, context-rich, and immediately reusable. As a result, they can become part of the *cognitive infrastructure* of a task—not merely an output to be read, but an always-available representational layer that shapes what users attend to, what they encode, and which cues are available during retrieval. From the perspective of encoding specificity [88], the introduction of a persistent AI representation can change both encoding conditions (what cues are present while learning) and retrieval conditions (what cues are present when remembering), with direct implications for memory performance and error.

AI-generated representations also introduce distinctive epistemic risks. When an external representation is plausible and blends seamlessly with source content, users may later confuse its origin. The Source Monitoring Framework emphasizes that remembering involves attributing content to its source; when contextual tags are weak, source confusions and misattributions become more likely [37]. The risk is amplified when AI content includes subtle inaccuracies or leading suggestions: generative systems can introduce details that are later incorporated into a user's recollection or judgment [15]. Accordingly, AI-assisted external representations should be conceptualized not only as aids, but also as potential drivers of distortion, over-reliance, and miscalibrated confidence [93].

Finally, the notion of cognitive infrastructure has an organizational dimension. Knowledge-intensive work is mediated by shared artifacts—documents, roadmaps, dashboards, and decision logs. When AI-generated summaries and recommendations enter these artifact ecosystems, they can reshape collective memory and decision processes. Organizations may benefit from improved knowledge continuity and retrieval [66], while simultaneously facing new challenges of accountability, transparency, and epistemic dependence [67, 95]. This thesis therefore treats AI-assisted external representations as a multi-level phenomenon: they influence individual cognition (RQ1), organizational practice in product management (RQ2), and the link between these levels (RQ3).

3.2. The AI Buffer Model

The increasing integration of AI systems into knowledge-intensive work has intensified the need for cognitive models capable of explaining how externally generated representations interact with human cognition. While existing research has addressed related phenomena such as cognitive offloading, distributed cognition [35], and extended memory, these perspectives often

remain fragmented or underspecified when applied to AI-supported tasks. In particular, they frequently fail to distinguish between static external aids and adaptive, generative systems that dynamically reshape information during task execution.

The AI Buffer Model addresses this gap by providing a formal cognitive framework that conceptualizes AI-generated external representations as an intermediate cognitive layer. This layer persists throughout task execution and actively interacts with core memory processes, including encoding, retrieval, confidence calibration, and cognitive load management. By doing so, the model offers a principled account of how AI systems influence cognition without assuming either full automation or passive tool use.

Crucially, the AI Buffer Model does not treat AI systems as autonomous decision-makers, nor does it reduce them to mere storage devices. Instead, it situates AI-generated representations within the cognitive loop, emphasizing their role as epistemic scaffolds that reshape how users engage with information over time.

Definition and Scope

The **AI Buffer** is defined as a set of AI-generated external representations that persist during task execution and are intentionally integrated into the user's cognitive process. These representations may take the form of summaries, structured outlines, reformulations, contextual prompts, synthesized explanations, or dynamically updated representations of task-relevant information.

What distinguishes the AI Buffer from traditional external memory aids is not its physical location or medium, but its *functional integration* within cognition. Static artifacts such as notes or documents typically require deliberate retrieval and interpretation. In contrast, AI-generated representations are produced in response to user intent, adapted to task context, and continuously consulted during ongoing cognitive activity.

The AI Buffer does not aim to replace internal memory or to function as a long-term archival system. Rather, it operates as an intermediate representational layer that supports cognition by maintaining task-relevant information in a structured, accessible, and context-sensitive form. From a theoretical standpoint, the AI Buffer builds upon the tradition of distributed cognition [35], which conceptualizes cognitive processes as distributed across internal and external resources. However, the model extends this tradition by focusing specifically on AI-generated representations that are adaptive, generative, and persistently available throughout task execution—properties that distinguish the AI Buffer from earlier forms of external cognition and make it particularly relevant for contemporary AI-supported work.

Importantly, the AI Buffer is knowingly externalized. Users are aware that representations originate from an AI system, yet these representations become tightly coupled with internal cognition through repeated interaction. This coupling creates a hybrid cognitive system in which internal and external representations jointly shape reasoning and memory.

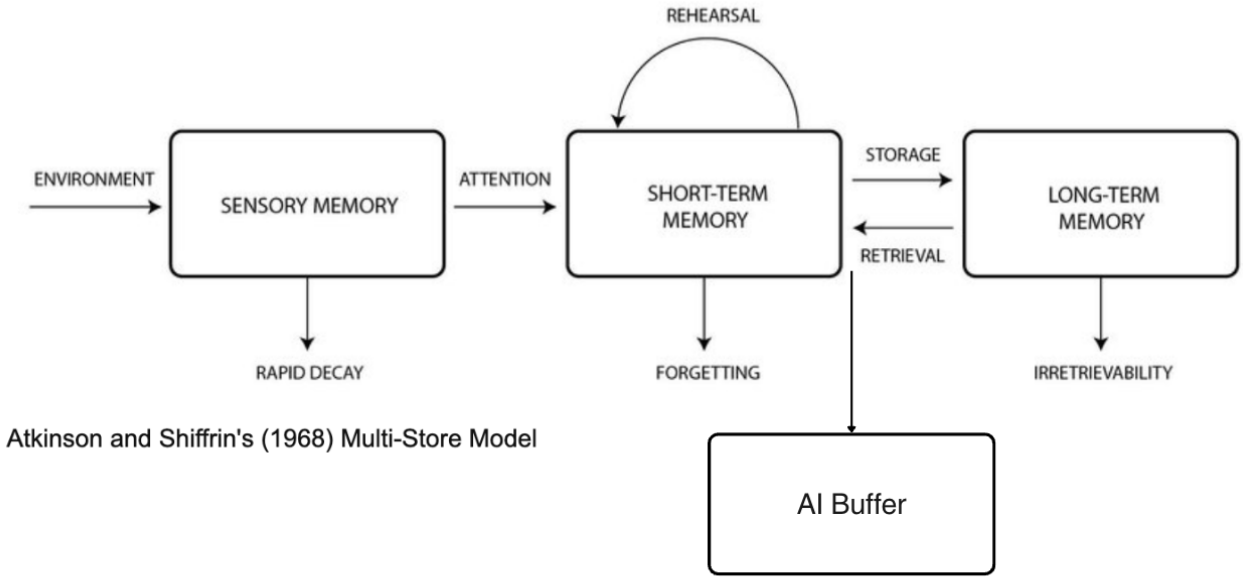


Figure 3.1: The AI Buffer Model: AI-generated external representations (the buffer) interact with human cognition and mediate downstream decision processes and organizational outcomes.

Model Components and Design Levers

The AI Buffer Model distinguishes three core elements:

- **Human memory system.** A capacity-limited working-memory workspace and long-term memory stores that support encoding, consolidation, and retrieval [7].
- **AI-generated representation (the buffer).** A persistent external representation (e.g., a summary) that can be consulted during task execution and that provides semantic content and retrieval cues.
- **Task and interface context.** The constraints and interaction structure that govern how the buffer is accessed (e.g., timing of availability, integrated vs. segmented formatting) and how the user allocates attention between sources.

In Study 1, two interface-level levers operationalize properties of the AI Buffer:

- **Timing of access:** whether the buffer is available before reading (pre-reading), during reading (synchronous), or after reading (post-reading). Timing changes whether the buffer functions as an advance organizer, an interruption, or a review aid [5, 6].
- **Structure of presentation:** whether the buffer is presented as an integrated paragraph or as segmented bullet points. Structure changes coherence cues and coordination demands, and is expected to affect split-attention costs and verification behavior [16, 83].

Other properties (e.g., factual accuracy, provenance cues, and the degree of interactivity) are treated as important contextual variables, but are held constant or discussed as boundary

conditions in the present work.

Cognitive Functions and Mechanisms Aligned with RQ1

RQ1 asks how AI-generated external representations influence core human memory processes. Within the AI Buffer Model, these influences are expressed through five tightly linked constructs: encoding, recall/recognition, source monitoring, confidence calibration, and cognitive load.

Encoding modulation. Encoding refers to the transformation of information into a form that can be stored in memory. Encoding depth depends on attentional allocation and elaborative processing [20]. The AI Buffer modulates encoding by guiding attention toward specific elements and by structuring content in ways that highlight relationships, hierarchies, and relevance.

When the buffer functions as an advance organizer, it may activate schemas and provide a high-level framework that guides attention toward central propositions [5, 6]. By externalizing intermediate representations, the buffer can reduce the cognitive effort required to identify salient information, supporting deeper encoding when users allocate freed attentional resources to interpretation and integration rather than surface-level processing.

However, encoding modulation is not uniformly beneficial. When a buffer is perceived as a reliable substitute for internal encoding, it may encourage shallow processing and cognitive offloading: users may allocate less working-memory effort to constructing a detailed internal representation because the external representation remains available [71, 80]. The model explicitly accounts for this duality by treating encoding modulation as contingent on interaction patterns rather than as an inherent benefit of buffering.

Recall and recognition. At retrieval, a buffer can serve as a cue source that reinstates aspects of the encoding context. Encoding specificity predicts better retrieval when cues present at encoding are available at recall [88]. This cueing logic suggests that buffer access can improve recognition of summary-aligned propositions even if free recall remains unchanged. At the same time, reliance on external cues may reduce the need to reconstruct information from internal traces, which could weaken detail-rich recall when the buffer is unavailable or when retrieval requires reconstruction beyond the buffer's coverage.

By maintaining task-relevant representations in an accessible form, the buffer stabilizes recall across time and context, supporting continuity in long-term tasks where internal memory decay would otherwise undermine coherence. However, externally supported recall introduces dependence: improved recall through buffering may come at the cost of reduced robustness when external representations are unavailable or inaccurate.

Source monitoring. Because AI-generated content is fluent and semantically plausible, it can blur boundaries between what was read in the original text and what was supplied by the buffer. The Source Monitoring Framework predicts greater misattribution when sources share overlapping content and when contextual cues are weak [37]. In AI-assisted settings, this risk

extends to both inadvertent inaccuracies and to structurally plausible but incorrect details. The potential for suggestion-driven errors is supported by evidence that conversational AI can increase false memories under certain conditions [15]. In the AI Buffer Model, source monitoring is therefore a central epistemic-risk channel, not a peripheral outcome.

Source monitoring also plays a critical role in decision ownership. Accurate source monitoring enables decision-makers to distinguish between internally generated knowledge and externally provided input. When AI-generated representations become tightly integrated into cognitive processes, this distinction may blur, leading to uncertainty about the origin of specific judgments or memories.

Confidence calibration. Confidence is a metacognitive judgment about the reliability of one's own memory or decision. AI-generated representations often enhance perceived fluency by presenting information in a coherent, structured, and articulate form. This fluency can inflate confidence independently of actual understanding, leading to overconfidence [41]. However, fluency is not equivalent to correctness: if the buffer introduces inaccuracies or if the user misattributes buffered content to the source, confidence can become miscalibrated. This is especially problematic when high confidence drives uncritical reliance on AI-generated representations [93].

The model incorporates this ambivalence by treating confidence calibration as a modulated outcome shaped by the interaction between AI-generated representations and user metacognition: the buffer can support confidence by clarifying complex information, but it can also distort confidence when fluency masks uncertainty or gaps in knowledge.

Cognitive load redistribution. Working memory capacity is limited, and excessive cognitive load impairs both performance and learning [7]. The buffer can reduce load by compressing information and by offering structured cues, but it can also add load by creating split attention and task-switching demands. Cognitive load theory predicts that formats imposing extraneous coordination demands reduce learning efficiency [83]. The split-attention effect predicts that separated information sources require extra integration effort, whereas integrated formats reduce this burden [16]. Accordingly, a segmented summary may increase coordination costs relative to an integrated summary, while synchronous access may introduce interruption costs during reading.

The model treats cognitive load redistribution as a *reallocation* rather than a simple reduction, emphasizing trade-offs between immediate performance and long-term learning. This redistribution can free cognitive resources for higher-level reasoning and integration, but it also reduces internal rehearsal and maintenance, which are important for consolidation.

Temporal Dynamics and Interaction Effects

A key contribution of the AI Buffer Model lies in its attention to temporal dynamics. Buffering effects are not static; they evolve over time as users interact repeatedly with AI-generated

representations. Early interactions may enhance performance and reduce cognitive strain, while prolonged reliance may reshape memory strategies and metacognitive habits.

The model therefore conceptualizes buffering as a dynamic process that unfolds across repeated task episodes. Over time, users may adapt their encoding strategies, reduce internal rehearsal, and increasingly rely on external representations. These adaptations can alter the balance between internal and external cognition, with implications for learning, expertise development, and resilience.

Interaction patterns also matter. Reflective interaction with the buffer—such as questioning, revising, or extending AI-generated representations—may support deeper processing and mitigate dependence. Passive consumption, by contrast, may amplify shallow encoding and over-reliance. The model explicitly accommodates these interaction effects, positioning them as boundary conditions rather than exceptions.

Boundary Conditions and Cognitive Risks

The AI Buffer Model does not assume that buffering is universally beneficial; it specifies conditions under which buffering supports cognition and conditions under which it undermines it.

Memory distortion and false memories. Human memory is susceptible to suggestion and reconstruction errors [10, 50]. AI-generated representations that contain inaccuracies or hallucinations may be inadvertently integrated into memory, leading to distorted recall. Because AI outputs often appear fluent and authoritative, such distortions may remain undetected [15].

Source confusion. When AI-generated representations are tightly integrated into cognitive processes, source monitoring may blur, leading to uncertainty about the origin of specific judgments or memories [37].

Epistemic dependence. Repeated reliance on buffering may lead to epistemic dependence. External support can gradually replace internal memory robustness, increasing vulnerability to system failures or errors [80]. The model treats epistemic dependence as an emergent property of interaction patterns rather than as an inherent flaw of AI systems.

Individual differences. Prior knowledge can reduce reliance on the buffer because users can evaluate and integrate information more independently. Conversely, high time pressure or low familiarity can increase reliance because verification costs are higher. At the organizational level, AI literacy—the ability to interpret, evaluate, and appropriately use AI outputs—is expected to moderate reliance behaviors and perceived benefits [91]. These factors are relevant for both Study 1 (as individual differences) and Study 2 (as professional competencies and organizational practices).

3.3. From Cognitive Buffering to Decision Processes

While RQ1 focuses on memory mechanisms, knowledge-intensive work is ultimately evaluated through decisions and actions. In most knowledge-intensive tasks, decision-makers do not operate on the basis of complete, immediately available information. Instead, they rely on reconstructed representations of prior analyses, experiences, and assumptions accumulated over time. This dependence is well captured by the notion of bounded rationality, which emphasizes that decision-makers operate under constraints of limited attention, working memory capacity, and cognitive resources [78]. Memory limitations thus function as a structural filter, determining the informational basis upon which decisions are formed.

AI-generated representations intervene precisely at this level. By externalizing intermediate representations and maintaining them persistently throughout task execution, AI systems reduce the need for internal reconstruction. In doing so, they reshape the cognitive conditions under which decisions are made, altering both the accessibility and the perceived reliability of information. Crucially, this intervention does not replace human judgment, but modifies the informational landscape within which judgment operates.

The AI Buffer Model bridges cognition to decisions through three pathways:

(1) Evidence selection and salience. By compressing and reordering information, the buffer makes certain propositions more salient than others. This can improve efficiency (faster access to core points), but also introduce representational bias: what is summarized and how it is framed can shape which evidence is considered. When users offload heavily, the buffer becomes a primary evidence source rather than a secondary aid [71, 80]. Deeper and more organized encoding increases the likelihood that relevant information will be accurately recalled and appropriately integrated during decision-making, thereby supporting higher decision quality. However, when AI outputs substitute for cognitive effort, encoding may remain shallow, weakening the informational foundation of subsequent decisions.

(2) Epistemic control and verification. Decisions in organizations often require traceability: the ability to justify a choice by pointing to sources. Source monitoring failures can therefore become governance failures. If users misattribute AI content to authoritative sources, decision rationales may be built on unverified claims. This risk is magnified when AI systems are opaque or when their reasoning is difficult to audit [67]. Persistent external representations reduce reliance on internal reconstruction, supporting consistency across decisions made at different points in time. At the same time, externally anchored consistency introduces dependence: consistency achieved through buffering may be fragile if external representations are inaccurate or unavailable [80].

(3) Confidence-driven reliance. Confidence affects whether people seek additional information, challenge a recommendation, or defer to an automated output. AI-generated represen-

Cognitive mechanism (RQ1)	Decision-process implication	Organizational risk / design need
Offloading and cue-dependent retrieval	Faster access to "good enough" evidence; less internal reconstruction	Over-reliance under time pressure; need for verification norms
Source monitoring and misinformation	Misattribution of AI claims as authoritative sources	Accountability failures; need for provenance cues and audit trails
Confidence calibration	Greater decisiveness; reduced information-seeking when confidence is high	Overconfidence in flawed outputs; need for calibration feedback
Cognitive load and split attention	Reduced effort when summaries compress information; increased effort when switching sources	Interface design matters (integration, timing); need to minimize extraneous load

Table 3.1: How AI-assisted memory mechanisms can translate into decision-process dynamics.

tations often enhance fluency and coherence, which may inflate subjective confidence independently of actual knowledge [41]. If AI buffers increase confidence without improving accuracy, they can encourage over-reliance. Misaligned confidence can lead to premature commitment or insufficient scrutiny. At scale, this can produce a "knowledge imbalance" in AI-advised decision-making where decision-makers become dependent on representations they cannot fully evaluate [95].

Decision ownership and accountability. Beyond cognitive performance, decision-making involves issues of ownership and accountability. Memory plays a central role by supporting source monitoring and retrospective explanation. AI buffering complicates these dynamics by introducing representations that blur the boundary between internal cognition and external support. Decision-makers may experience source ambiguity, leading to uncertainty about whether a particular judgment originated from personal analysis or from system-generated input. Over time, this ambiguity may weaken the sense of ownership over decisions [37]. Preserving decision ownership therefore requires maintaining metacognitive awareness of the role played by AI-generated representations. Without such awareness, authority may implicitly shift toward the system, undermining accountability. These issues are particularly salient in organizational contexts, where decisions have long-term consequences and must be justified retrospectively.

Table 3.1 summarizes how the cognitive mechanisms identified in the AI Buffer Model translate into decision-process dynamics and organizational considerations.

3.4. The AI-Augmented Decision Framework

The AI Buffer Model provides a detailed account of how AI-generated external representations interact with human memory and metacognitive processes during task execution. However, understanding the broader implications of these mechanisms requires situating them within a structure that explains how cognitive effects translate into decisions and, ultimately, into organizational outcomes. The *AI-Augmented Decision Framework* fulfills this role by integrating the AI Buffer Model into a multi-level decision-oriented perspective.

Crucially, the two models are not alternative or competing frameworks. The AI Buffer Model specifies the cognitive mechanisms through which AI systems influence human cognition, while the AI-Augmented Decision Framework specifies how these mechanisms propagate upward through decision processes to shape organizational outcomes. In this sense, the AI Buffer Model provides the cognitive foundation upon which the AI-Augmented Decision Framework is built.

Framework Overview

The AI-Augmented Decision Framework is structured around four analytically distinct but interdependent levels:

AI Systems

↓

AI Buffer (Cognitive Level)

↓

Decision Processes

↓

Organizational Outcomes

At the first level, **AI systems** generate external representations in response to user intent and task context. These representations include summaries, structured reformulations, contextual cues, and synthesized explanations. Importantly, these outputs are not treated as final decisions or recommendations, but as informational structures intended to support human cognition.

At the second level, these representations function as an **AI Buffer**. As described in Section 3.2, AI-generated representations persist during task execution and interact with memory encoding, recall reliability, confidence calibration, and cognitive load distribution. This level captures how AI systems reshape the cognitive environment in which individuals think, remember, and interpret information.

At the third level, cognitive buffering mechanisms influence **decision processes**. Decisions are formed on the basis of information that is accessible, salient, and trusted at the moment of choice. By modulating memory and metacognition, the AI Buffer reshapes the informational inputs and constraints under which decisions are made, affecting how alternatives are evaluated,

how strongly judgments are held, and how decisions are justified.

At the fourth level, repeated decision processes accumulate into **organizational outcomes**. Decisions are embedded in routines, roles, coordination mechanisms, and governance structures. Over time, the cognitive effects of AI buffering shape how organizations maintain continuity, align actions, and assign responsibility.

The framework emphasizes a directional but non-deterministic logic. AI systems do not directly produce organizational outcomes. Instead, they intervene at the cognitive level, altering the conditions under which decisions are made. Organizational outcomes emerge from the aggregation of AI-augmented decisions within specific contexts.

Organizational Outcomes

Within the framework, organizational outcomes are derived from systematic changes in decision processes induced by cognitive buffering.

Decision quality over time emerges from the interaction between encoding modulation and recall reliability. When AI buffering supports deeper encoding and stabilizes access to prior analyses, decision-makers are better able to integrate relevant information across time. However, when buffering substitutes for active engagement, shallow encoding may undermine the informational basis of future decisions.

Strategic coherence follows from the stabilization of decision rationales. By maintaining persistent representations of prior assumptions, analyses, and trade-offs, AI buffering can support alignment across decisions made at different points in time or by different actors. At the same time, externally anchored coherence may increase dependence on AI-mediated representations, making strategic alignment vulnerable to errors embedded in those representations.

Role transformation arises from the redistribution of cognitive load associated with information integration and recall. As AI buffering externalizes memory-intensive operations, roles may shift toward higher-level judgment, coordination, and meta-decision-making. This transformation does not imply a reduction in responsibility; on the contrary, greater emphasis is placed on evaluating inputs, managing uncertainty, and ensuring alignment across stakeholders.

Accountability and governance are shaped by changes in confidence calibration and source monitoring. When AI-generated representations influence judgments, organizations must ensure that decision rationales remain transparent and that responsibility remains clearly assigned. If source monitoring is weakened and confidence becomes miscalibrated, decision ownership may become diffuse.

Organizational Layer: AI Adoption in Product Management

RQ2 focuses on product management as a decision-intensive domain in which AI adoption is expected to reshape both work practices and responsibility structures. Product managers routinely integrate heterogeneous evidence—user research, market signals, analytics, technical feasibility, and stakeholder constraints—into prioritization and strategy under uncertainty and

time pressure. In agile settings, they coordinate discovery and delivery while aligning cross-functional teams around goals and trade-offs [85]. These responsibilities depend heavily on external representations: roadmaps, PRDs, experiment results, meeting notes, and decision logs stabilize shared understanding and enable coordination across time.

Generative AI tools are increasingly positioned as an additional representational layer in these workflows [31, 34, 63]. Typical uses include summarizing customer feedback, drafting product narratives, accelerating synthesis across fragmented sources, and supporting decision preparation (e.g., "pros/cons" lists, alternatives, and risk summaries). These uses combine *automation* (reducing repetitive synthesis work) and *augmentation* (supporting judgment with condensed representations). However, they also shift part of the PM role toward evaluation and governance of AI outputs: product managers must detect when AI-generated summaries are incomplete or biased, maintain provenance cues, and ensure that accountability for decisions remains clear [67].

The organizational layer therefore foregrounds three constructs that parallel the cognitive layer:

- **Workflow reconfiguration:** which tasks are delegated to AI representations and which remain human-led.
- **Role configuration and skill shift:** movement toward strategic, interpretive, and oversight responsibilities; increasing importance of AI literacy [91].
- **Governance and perceived responsibility:** who is accountable for AI-influenced decisions, and what oversight practices are adopted to prevent uncritical acceptance and mitigate bias [95].

Finally, adoption is expected to be context-sensitive. Organizational maturity, regulatory expectations, and cultural norms can shape how much AI is used, which tasks are automated versus augmented, and how governance is implemented. This motivates explicit cross-context comparison within Study 2 (e.g., USA–China) as part of RQ2's scope.

Product management exemplifies the characteristics that make knowledge-intensive roles both promising and sensitive contexts for AI augmentation: sustained information integration, temporal dispersion of decisions, and high memory load. In such roles, AI buffering can significantly reshape the cognitive environment by stabilizing access to prior analyses, reducing reconstruction effort, and supporting integration across information sources. At the same time, the risks identified by the framework—over-reliance, miscalibrated confidence, and source ambiguity—are particularly consequential when decisions have long-term strategic implications.

Feedback Loops and Multi-Level Integration

The core contribution of this thesis is an integrated framework that connects cognitive mechanisms of AI-assisted memory (RQ1) to organizational decision-making practices in product management (RQ2), and uses the former to explain the latter (RQ3).

At the **cognitive level**, the AI Buffer Model specifies how AI-generated external representa-

tions shape encoding, retrieval, source monitoring, confidence calibration, and cognitive load. These mechanisms determine not only whether people remember more or less, but also what *kind* of knowledge is retained (gist vs. detail), how confidently it is held, and whether its source is correctly attributed.

At the **decision-process level**, these cognitive outcomes translate into observable behaviors: how evidence is selected, how options are generated and evaluated, and how rationales are communicated. For example, cue-driven recognition may speed up decision preparation, whereas source confusions can undermine traceability and learning from past decisions.

At the **organizational level**, repeated use of AI-generated representations can stabilize into norms of reliance, documentation, and accountability. Over time, teams may increasingly treat AI outputs as default inputs to decisions, potentially creating epistemic dependence and shifting expertise toward those who can evaluate and govern AI systems [95]. In parallel, organizations may invest in knowledge management practices that embed AI into collective memory systems [66].

Crucially, the framework includes **feedback loops**. Organizational norms and training influence individual reliance and verification behavior (e.g., AI literacy and governance practices), while individual experiences of accuracy, effort reduction, and error influence organizational adoption decisions. These loops imply that AI's impact is not static: the same tool can produce different outcomes depending on governance maturity, interface design, and task context.

3.5. Implications for Empirical Research

The AI-Augmented Decision Framework provides a multi-level account of how AI-generated external representations influence cognition, decision processes, and organizational outcomes. A central implication is that AI-related effects cannot be adequately captured at a single level of analysis. Cognitive buffering mechanisms operate at the individual level, decision processes mediate these effects at the meso level, and organizational outcomes emerge over time through repeated decision episodes. Capturing this progression requires empirical strategies capable of isolating micro-level mechanisms while also assessing their relevance in real-world organizational contexts. For this reason, the thesis adopts a multi-method empirical design composed of two complementary studies.

Cognitive-level measurement (Study 1). The framework suggests that the timing, persistence, and structure of AI-generated representations are critical variables. Cognitive effects are expected to differ depending on whether AI support is provided during encoding, during recall, or continuously throughout task execution. Similarly, the structure of AI-generated representations—whether they scaffold elaboration or substitute for cognitive effort—is expected to shape both performance and memory outcomes. These considerations motivate an

experimental approach focused on isolating the causal impact of AI buffering configurations on memory-related outcomes.

Importantly, the framework highlights that performance improvements alone are insufficient indicators of cognitive benefit. Enhanced task performance may coexist with weakened internal memory or miscalibrated confidence. Study 1 operationalizes the cognitive mechanisms via complementary measures (Chapter 4): free recall quality (internal reconstruction), recognition accuracy (cue-supported retrieval), false-lure endorsement and attribution outcomes (source monitoring / epistemic risk), confidence ratings (calibration), and behavioral process logs (attention allocation and reliance proxies).

Organizational-level measurement (Study 2). The framework emphasizes that AI effects are mediated by patterns of use rather than by system capabilities alone. The same AI system may produce different decision outcomes depending on how it is integrated into workflows, how frequently it is consulted, and how decision-makers interpret its outputs. Study 2 operationalizes the organizational layer via survey evidence on AI use-cases, perceived efficiency gains, changes in decision practice, and perceived governance and responsibility implications (Chapter 6).

Linking experimental and survey evidence. Within the framework, experimental and survey-based methods target different levels of the same underlying process. Experimental evidence provides insight into how AI buffering configurations affect cognitive mechanisms under controlled conditions. Survey evidence captures how similar mechanisms are experienced, interpreted, and enacted in real organizational settings. For example, experimentally observed effects on confidence calibration can be linked to survey measures of decision confidence and reliance. By grounding both studies in a shared theoretical framework, the thesis positions them as complementary perspectives on the same phenomenon rather than independent findings.

3.6. Research Hypotheses and Propositions

This section formalizes testable expectations derived from the conceptual framework. Hypotheses are stated for the controlled experiment (Study 1; RQ1), while propositions are stated for the organizational analysis (Study 2; RQ2) and the cross-study integration (RQ3).

Study 1 (RQ1): Cognitive-Level Hypotheses

1. **H1 (AI Buffer vs. No-AI; recall/recognition).** Relative to unaided reading, access to an AI Buffer (operationalized as an AI-generated summary) is expected to improve recognition performance for core propositions, with weaker or no improvement expected for free recall quality.
2. **H2 (Timing; encoding scaffold vs. interruption).** Pre-reading access to the AI Buffer is expected to yield stronger encoding and higher recognition accuracy than synchronous or post-reading access, because it can function as an advance organizer rather

than as an interruption or a post-hoc reminder [5, 6].

3. **H3 (Structure; split-attention and coherence).** Integrated-paragraph buffers are expected to reduce split-attention costs and support more coherent mental-model construction than segmented bullet-point buffers, resulting in higher recall quality and/or recognition accuracy [16, 83].
4. **H4 (Epistemic risk; source monitoring).** The presence of an AI Buffer is expected to increase vulnerability to source-monitoring errors and false recognitions for plausible but non-present details. This risk is expected to be higher under segmented structure than integrated structure, because segmentation can promote decontextualized processing and weaker provenance cues [37].
5. **H5 (Confidence calibration).** AI Buffer access is expected to increase subjective confidence in memory judgments. However, confidence may be less well calibrated to accuracy under high reliance conditions, especially for buffer-consistent false lures [93].

Study 2 (RQ2): Organizational-Level Propositions

1. **P1 (Role transformation).** AI adoption in product management is expected to shift effort away from routine synthesis and coordination toward more strategic, interpretive, and oversight activities [34, 63, 92].
2. **P2 (Decision-support centrality).** AI-generated external representations are expected to become central inputs to decision preparation and communication (e.g., summarizing evidence, framing options, and drafting rationale), increasing the speed of decision cycles while changing how evidence is selected and narrated [28, 31, 63].
3. **P3 (Governance and accountability).** As AI outputs influence product decisions, perceived responsibility for verification, transparency, and ethical oversight is expected to increase, and governance mechanisms are expected to become more salient in PM practice [67].
4. **P4 (AI literacy as a capability).** Higher AI literacy is expected to be associated with more calibrated reliance and more effective integration of AI into PM workflows [91].
5. **P5 (Contextual moderation).** Adoption patterns, perceived responsibility, and governance emphasis are expected to vary across organizational and cultural contexts (e.g., USA–China), reflecting differences in norms, constraints, and maturity of AI deployment [42, 58].

Cross-Study Integration (RQ3): Integrative Propositions

1. **I1 (Offloading → reliance norms).** When AI Buffers reliably reduce cognitive effort, users and teams are expected to offload more frequently, which can scale into organizational norms of reliance under time pressure [71, 95].
2. **I2 (Source monitoring → accountability).** Source-monitoring difficulties at the individual level are expected to manifest as accountability and traceability challenges at

the organizational level, increasing the need for provenance cues, documentation, and verification practices [37, 67].

3. **I3 (Confidence calibration → trust and escalation).** Miscalibrated confidence in AI representations is expected to contribute to over-trust and insufficient escalation or verification, motivating organizational interventions that improve calibration (training, feedback, and auditability) [93].

Chapter summary. This chapter has developed an integrative theoretical foundation for understanding how AI systems reshape cognition, decision-making, and organizational outcomes in knowledge-intensive work. The AI Buffer Model formalizes the cognitive mechanisms—encoding modulation, recall reliability, source monitoring, confidence calibration, and cognitive load redistribution—through which AI-generated external representations interact with human memory. The AI-Augmented Decision Framework then traces how these buffering mechanisms propagate upward from individual cognition through decision processes to organizational outcomes such as strategic coherence, role transformation, and governance challenges. Together, the two models provide a coherent, multi-level lens that connects interface-level design choices to measurable cognitive and organizational outcomes, and motivates a multi-method empirical design combining experimental and survey-based approaches. The following chapters present the methodology and results of the two empirical studies that operationalize this framework.

4 | Methodology and Research Design

This chapter describes the research design and methods used to answer the three research questions introduced in Chapter 1. The thesis adopts a multi-level approach: Study 1 provides controlled evidence on how AI-generated external representations affect memory mechanisms (RQ1), Study 2 provides field evidence on how AI adoption reshapes decision practices and perceived responsibility in product management (RQ2), and the two are integrated conceptually and analytically to explain organizational patterns through cognitive mechanisms (RQ3).

4.1. Research Design Overview

The central methodological challenge of this thesis is that the phenomenon of interest is inherently multi-level. At the individual level, AI-generated representations can change attention allocation, encoding depth, cue availability, and source monitoring, with downstream effects on recall, recognition, confidence, and perceived cognitive load. At the organizational level, the same representational layer can become embedded in workflows, shaping how evidence is synthesized, how decisions are justified, and how responsibility is distributed in practice. No single method is sufficient to capture both levels with high validity.

Accordingly, the thesis combines two complementary designs:

- **Internal validity (Study 1).** A controlled experiment isolates the causal effect of AI-generated summaries as an external representation by manipulating key design properties of the AI Buffer (timing and structure) under matched task constraints.
- **Ecological validity (Study 2).** A field study captures how AI tools are adopted and governed in a real managerial role (product management), including role configuration, decision support practices, and perceived accountability.

The two studies are linked by a shared conceptual object: *AI-assisted external representations*. In Study 1, this object is operationalized as pre-generated summaries that persist during a reading-and-memory task. In Study 2, the object is operationalized as generative AI tools

Component	RQ	Design and unit of analysis	Primary measurements
Study 1	RQ1	Controlled experiment; individuals performing a timed reading-and-memory task	Free recall, recognition, false-lure endorsement/source monitoring, confidence ratings, cognitive-load ratings, behavioral logs
Study 2	RQ2	Field study; product managers reporting AI adoption and decision practices in context	AI tool use, workflow shifts, decision-support patterns, governance/ethics and perceived responsibility, USA–China comparison, open-ended responses
Integration	RQ3	Cross-study mapping (mechanisms → practices)	Mechanism-based explanation of observed organizational patterns and boundary conditions

Table 4.1: Multi-level research design linking RQ1–RQ3.

that produce summaries, analyses, and drafts within product-management workflows. The conceptual bridge is provided by the AI Buffer Model introduced in Section 3.2.

Replication and documentation. All participant-facing materials are reproduced in the Appendices (Study 1 stimuli and tests in Appendix A; Study 2 full questionnaire in Appendix B). The quantitative and qualitative analysis strategies are specified in Chapter 5 (Sections 5.4 and 5.5), so that the separation between "what was done" (this chapter) and "how it was analyzed" (Chapter 5) remains explicit.

4.2. Study 1: Controlled Experiment (Cognitive Level – RQ1)

Study 1 answers RQ1: *How do AI-generated external representations influence core human memory processes during knowledge-intensive tasks?* The study uses a controlled reading paradigm in which an AI-generated summary constitutes a persistent external representation (an AI Buffer) that participants can consult before, during, or after reading. The design targets five operational dimensions aligned with RQ1: memory encoding, recall and recognition, source monitoring, confidence calibration, and cognitive load.

Participants and SettingThirty-six adults were retained for analysis ($N_{\text{AI}} = 24$, $N_{\text{NoAI}} = 12$), including university students and early-career professionals. Participants were 18–35 years old ($M = 24.33$, $SD = 3.83$) and were balanced overall by gender (18 women, 18 men). All participants reported Chinese as their native language.

The study was delivered in a browser-based environment. Participant-facing materials were presented in Simplified Chinese to match participants' native language, while English descriptions

Table 4.2: Participant demographics by group (Study 1).

	AI ($n = 24$)	No-AI ($n = 12$)	Total ($N = 36$)
Age, M (SD)	24.50 (4.31)	24.00 (2.76)	24.33 (3.83)
Age range	18–35	20–30	18–35
Gender (F/M)	13/11	5/7	18/18
Native language (Chinese), n (%)	24 (100%)	12 (100%)	36 (100%)

are used in this thesis for exposition.

Participants received a base compensation of 60 RMB plus a performance-based bonus of up to 60 RMB linked to recognition-test accuracy.

Experimental Design and ConditionsAI assistance was operationalized as an AI-generated summary of each article. Within the AI-assisted sample, the core design forms a 2×3 matrix:

- **Summary structure (between-subjects):** integrated paragraph vs. segmented bullet points.
- **Summary timing (within-subjects):** pre-reading vs. synchronous during reading vs. post-reading.

A separate No-AI baseline group provides an external comparison but does not include the timing manipulation, yielding an asymmetric mixed design that balances experimental control and feasibility.

Structure manipulation. The integrated condition presents the buffer as a coherent paragraph-style summary. The segmented condition presents the same informational content as a set of bullet points. The manipulation targets coherence cues and coordination demands: integrated presentation should reduce split-attention and verification costs relative to segmented presentation [16, 83].

Timing manipulation. Timing determines whether the buffer functions primarily as an encoding scaffold (pre-reading), as an interruption and on-demand aid during comprehension (synchronous), or as a review cue after initial encoding (post-reading). This manipulation is theoretically grounded in advance-organizer logic and interruption/switching-cost accounts [5, 6].

Materials and Task Procedure

Materials

Texts. Participants read three expository articles of approximately 1,300 words each (topics included urban heat islands, CRISPR gene editing, and semiconductor supply chains). Articles were selected to be comparable in difficulty and conceptual density and were administered in Simplified Chinese using fixed, pre-generated stimulus content. English descriptions are used throughout the thesis for exposition, and study materials are provided in the Appendices.

AI summaries (structure and incompleteness). Summaries were generated and refined *in ad-*

vance and were approximately 250 words. They were identical in informational content across the two AI structure conditions, differing only in formatting (integrated paragraph vs. segmented bullets). Summaries were intentionally incomplete (approximately 15–20% of article information omitted) to discourage reliance on summaries alone and to ensure that some recognition items required article-based encoding.

Example stimulus and summary output (illustrative). To make the manipulation concrete, Appendix A provides full stimuli and summary outputs. In brief, the integrated condition presents a single consolidated representation aligned with the reading flow, whereas the segmented condition splits content across separated units. This difference is intended to manipulate the availability of provenance cues and the ease of cross-checking during encoding and retrieval.

False lures. To assess susceptibility to misinformation and source-monitoring errors, each article's summary contained two plausible but incorrect statements (*false lures*). These lure concepts also appeared as distractor options in the recognition test; selecting them was counted as false-lure acceptance.

Platform and implementationThe experiment was delivered via a custom browser-based application with a Python (Flask) backend and HTML/CSS/JavaScript front-end pages. The platform was designed to support (i) strict enforcement of timing constraints across conditions, (ii) automated counterbalancing and randomization, and (iii) fine-grained behavioral logging of reading and summary exposure. All stimuli (articles, summaries, and MCQs) were preloaded, ensuring identical content across participants and removing runtime variability due to on-the-fly AI generation.

Procedure and timing constraintsParticipants completed three article blocks. In each block, participants encountered the summary according to their assigned timing condition, read the article under a fixed time window, and then completed memory assessments and post-block ratings. Article order was randomized across the three articles (3! possible orders). For AI participants, timing order was implemented by sampling one of the six possible permutations of the three timing conditions (Table 4.3), so that each participant experienced each timing condition exactly once.

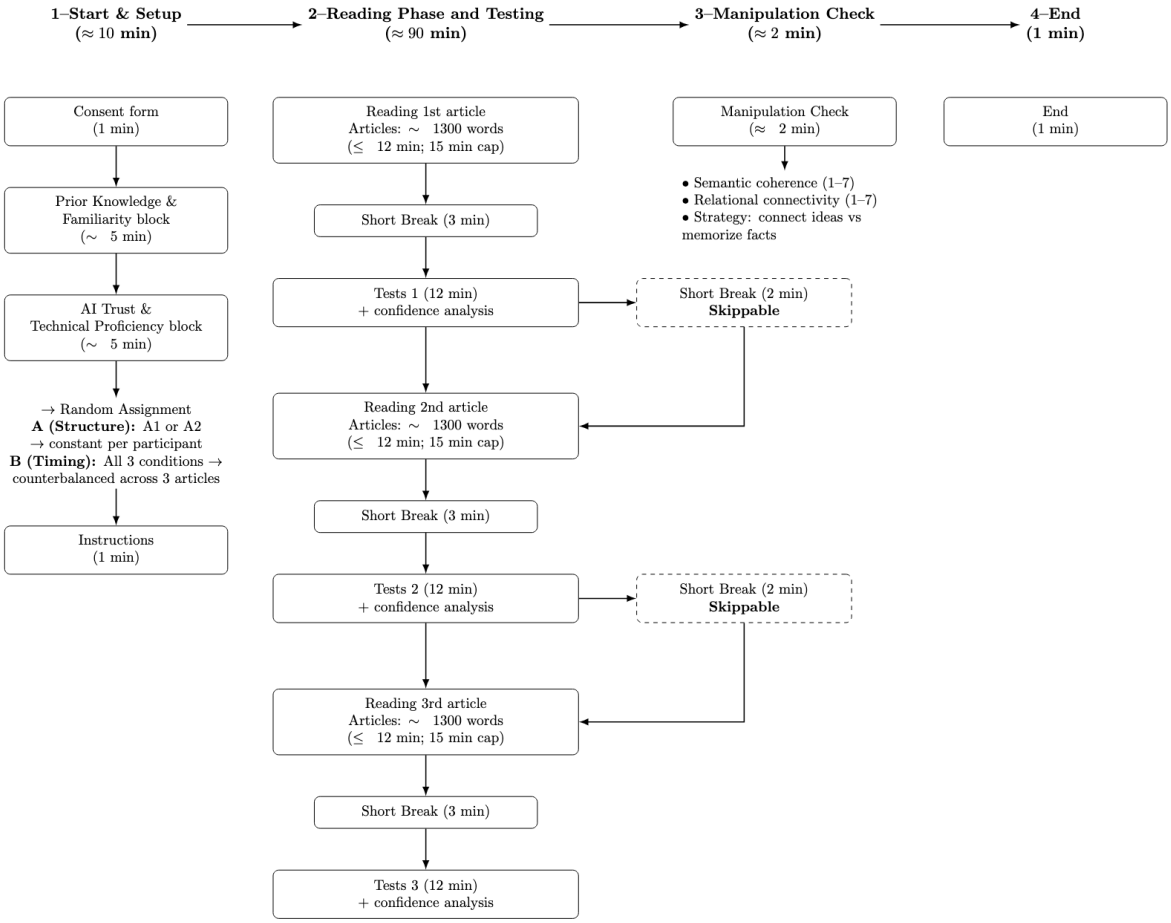


Figure 4.1: Experimental procedure and timing flow used in Study 1.

Schedule	Timing order across the three article blocks
S1	Pre-reading → Synchronous → Post-reading
S2	Pre-reading → Post-reading → Synchronous
S3	Synchronous → Pre-reading → Post-reading
S4	Synchronous → Post-reading → Pre-reading
S5	Post-reading → Pre-reading → Synchronous
S6	Post-reading → Synchronous → Pre-reading

Table 4.3: Counterbalancing schedule for timing conditions in Study 1 (AI group). Timing order was sampled per participant from the six possible permutations; article order was randomized independently.

After each reading phase, a short enforced break preceded testing to reduce immediate carryover and to standardize the transition to retrieval. Free recall lasted up to 5 minutes and the MCQ

Condition	Summary timing	Summary access	Reading cap
No-AI	–	None	15 min
Pre-reading	Before reading	Separate page, $\leq$ 3 min (continue early allowed)	12 min
Synchronous	During reading	On-demand open/close during reading	15 min
Post-reading	After reading	Separate page, $\leq$ 3 min (continue early allowed)	12 min

Table 4.4: Implementation of timing conditions and exposure windows (Study 1).

block lasted up to 7 minutes, with automatic submission at timeout. Back-navigation was disabled to prevent revisiting prior pages or materials.

Measures and Operationalization of RQ1. Study 1 operationalizes RQ1 through complementary behavioral, performance, and self-report measures aligned with the AI Buffer Model (Section 3.2).

Table 4.5: Measures overview (summary).

Construct	Operationalization	Purpose in the thesis
Recognition accuracy	MCQ accuracy by condition and item type	Cue-supported retrieval performance
Free recall	Recall scoring (overall and by timing)	Generative retrieval without cues
Source monitoring	False-lure endorsement / attribution outcomes	Provenance and misattribution risk
Process measures	Time allocation (summary vs reading)	Attention allocation mechanism
Calibration	Confidence vs accuracy (miscalibration)	Distinguish reliance from overconfidence

Free recall (generative retrieval). After reading each article, participants produced a written free recall. Free recall was scored as an index of episodic–semantic reconstruction of expository text rather than verbatim reproduction. In line with constructive memory theory [10], responses were evaluated for fidelity to the original conceptual structure (and penalized for plausible distortions), not for length or fluency.

Each recall received a single ordinal score from 0 to 10 based on five rubric dimensions: (i) factual accuracy, (ii) mechanistic reconstruction, (iii) structural alignment, (iv) specificity vs. gist, and (v) source-monitoring integrity (penalizing false-lure intrusions). To improve scoring consistency, anchor bands were used (9–10 high fidelity; 7–8 mostly accurate; 5–6 gist-level;

3–4 headline-level; 1–2 fragmentary).

To reduce variability associated with qualitative rating, free-recall responses were initially scored using an LLM rater applying the pre-specified rubric, and all scores were subsequently reviewed by a human rater (100% of responses) to confirm rubric adherence and correct questionable cases.

Recognition (cue-supported retrieval). Recognition was measured as proportion correct on 14 multiple-choice items per article. Items were categorized by information source (summary-sourced vs. article-only) to separate recognition of summary-covered information from comprehension of the underlying text. Misinformation susceptibility was indexed by the number of false-lure options selected (0–2 per article; 0–6 overall).

Source monitoring and epistemic risk. Source monitoring is assessed primarily through false-lure endorsement and related misattributions, motivated by the Source Monitoring Framework [37] and evidence that fluent AI interaction can amplify false memories under some conditions [15].

Confidence and cognitive load. Participants provided recall confidence and post-block ratings of mental effort and perceived difficulty using 7-point scales. Confidence calibration is analyzed by relating confidence to accuracy, testing whether AI assistance increases confidence without commensurate improvement in correctness.

Trust, dependence, and behavioral logs (process measures). AI-group participants provided baseline (general) measures of AI trust and AI dependence and post-block (state) ratings of trust and dependence. Behavioral measures were computed from server logs, including reading time, summary view time, number of summary openings, and interaction timestamps. These process indicators are used to triangulate how timing and structure shape reliance and attention allocation.

Data Quality and Exclusion LogicTo preserve internal validity, sessions were excluded prior to analysis if participants failed to complete the full protocol or showed clear non-compliance (e.g., implausibly short reading times indicating rushing or non-substantive responses). The resulting dataset is structured at the block level (participant $\times$ article), yielding 108 block-level observations (36 participants $\times$ 3 articles). This structure supports within-participant comparisons across timing conditions in the AI group while retaining a No-AI baseline for reference.

4.3. Study 2: Field Study on Product Managers (Organizational Level – RQ2)

Study 2 answers RQ2: *How does the adoption of AI tools shape decision-making practices, role configuration, and perceived responsibility in product management?* The study uses a cross-national field survey to capture how product managers incorporate AI tools into their workflow, how they perceive role transformation and skill shifts, and how they understand governance and ethical responsibility when AI influences decisions.

Design and SampleThe field study is based on a structured survey administered in two languages (English and Chinese) to support participation across contexts. The final dataset comprises 74 validated responses, evenly split across the United States ($n = 37$) and China ($n = 37$). This balanced structure enables both (i) estimation of overall patterns of AI adoption in product management and (ii) a comparative lens on how adoption and responsibility perceptions vary across contexts (USA–China), directly addressing the cross-context component of RQ2.

The focus on the United States and China is motivated by their central roles in the global AI ecosystem and by differences in organizational environments that may shape adoption (e.g., governance norms, tool ecosystems, and data-driven practices). The comparative design is therefore not treated as a mere "case split", but as a way to identify which adoption patterns are robust across contexts and which patterns are context-contingent. This logic directly supports RQ2 and also provides a foundation for the integrative explanation in RQ3.

Responses were screened for completeness and coherence (e.g., removing incomplete submissions and invalid patterns). The thesis treats the resulting dataset as an analytic snapshot of AI adoption in a decision-intensive managerial role rather than as a representative census of all product managers.

Survey Structure and Measurement StrategyThe survey combines Likert-scale items, multiple-choice items, and open-ended questions. This combination is intentional: structured items quantify adoption patterns and perceived impacts, while open-ended responses capture contextual nuance (e.g., examples of AI-supported decisions, perceived risks, and organizational constraints).

At a high level, the instrument is organized around four measurement domains aligned with RQ2. The complete instrument (including a reference Simplified Chinese translation) and a construct-to-item mapping used for analysis are provided in Appendix B and table B.2:

- **AI tool usage and task coverage:** frequency and type of AI applications used in daily PM work, including automation tools, analytics, and generative AI for drafting and synthesis.
- **Role configuration and workflow reallocation:** perceived changes in responsibilities,

time saved through AI support, and whether time is reallocated toward more strategic work.

- **Skill shift and AI literacy:** perceived need for new competencies (e.g., interpreting AI outputs, prompt practices, data literacy) and organizational support for upskilling.
- **Governance, ethics, and responsibility:** perceptions of accountability, transparency, bias management, and customer-trust implications when AI informs decisions or product features.

Operationally, the survey is structured in four sections (A–D). Section A measures AI tool usage and organizational adoption context (e.g., which tools are used, who drives adoption, and what goals motivate integration). Section B measures workload reduction, tasks that become more complex due to AI, and how time savings are reinvested (e.g., strategic planning, innovation). Section C focuses on skill development, confidence, and training resources. Section D focuses on governance and ethics (e.g., responsibility for AI-mediated decisions, guideline structures, and ethical concerns). This structure is aligned with the conceptual framework's organizational layer and supports both descriptive analysis and cross-context comparisons.

In terms of measurement design, Section A combines (i) frequency scales for distinct AI application categories (e.g., automation and workflow tools, analytics, NLP-based feedback analysis, generative AI drafting tools, and chatbots) with (ii) adoption-maturity items (e.g., duration of use) and (iii) open-ended prompts for tool names and examples. Sections B–D use a mix of Likert-scale judgments and categorical items to capture both intensity (how strongly an impact is perceived) and structure (who owns responsibility, whether guidelines exist, what forms of training are available). This combination supports the thesis's interpretation that AI adoption is not only a level of usage, but a configuration of practices and governance structures [42, 58, 66].

Planned analysis logic (brief). The survey analysis combines descriptive summaries with inferential comparisons aligned with RQ2. Likert-type items are evaluated against neutral midpoints to test whether perceived impacts differ from "no change", and USA–China differences are tested via independent-samples comparisons; categorical distributions are compared using χ^2-type tests, and open-ended responses are thematically coded. Full details are provided in Section 5.5.

4.4. Data Collection Instruments and Materials

Across both studies, the thesis uses instruments designed to capture behavior, judgment, and subjective experience around AI-generated external representations.

Study 1 instruments (experiment). Study 1 uses (i) expository reading materials, (ii) AI-generated summaries in two presentation formats (integrated vs. segmented), (iii) a free-recall

prompt, (iv) a recognition test per article including false-lure distractors, and (v) post-block rating scales for perceived cognitive load and confidence. The experimental platform records fine-grained interaction logs (e.g., time spent reading and time spent consulting the summary) to provide behavioral indicators of reliance. Verbatim stimuli, MCQs, and rating items are included in Appendix A.

Study 2 instruments (field survey). Study 2 uses a bilingual structured survey comprising Likert-scale items, multiple-choice items, and open-ended questions. The instrument measures frequency of AI use across task categories, perceived role transformation, time savings and reallocation, skill development and training, and governance/ethics perceptions. The bilingual format is intended to reduce language barriers and support valid cross-context comparison. The full questionnaire is reproduced in Appendix B.

Integration materials. For RQ3, the key "instrument" is the mapping logic specified by the conceptual framework in Chapter 3. Cognitive constructs from Study 1 (offloading, cue-dependent retrieval, source monitoring, confidence calibration, cognitive load) are used as explanatory lenses for interpreting adoption and governance patterns observed in Study 2.

4.5. Ethical Considerations

Both studies involve minimal-risk procedures and follow standard research ethics principles: voluntary participation, informed consent, the right to withdraw, and privacy-preserving data handling.

Study 1 (experiment). Participants were informed about the study procedures, data collection, and compensation, and provided consent before participation. Personal identifiers were not stored with behavioral data. Because the experiment includes plausible but incorrect statements in AI summaries to test source monitoring and misinformation susceptibility, the study requires careful debriefing: participants should be informed after participation that some summary content was intentionally inaccurate for research purposes, to avoid leaving lasting misconceptions.

Study 2 (field survey). Survey participation was voluntary and responses were collected and analyzed in aggregated form. The survey was designed to avoid collection of sensitive corporate information (e.g., proprietary product details). Any open-ended responses are treated as confidential and are reported only in anonymized, non-identifying form.

Data management. Across studies, datasets are stored in access-controlled locations and are used only for research purposes. Reporting focuses on patterns rather than on identifying individuals or organizations, consistent with the thesis's goal of developing generalizable insights about AI-assisted cognition and decision-making.

5 | Data Analysis

This chapter specifies how data from Study 1 (controlled experiment; RQ1) and Study 2 (field study on product managers; RQ2) are transformed into analytic constructs, prepared for analysis, and analyzed. It also defines the cross-study integration logic used to address RQ3 and summarizes how validity, reliability, and robustness are handled across the full thesis.

5.1. Software, Reproducibility, and Reporting Standards

All analyses were executed using reproducible scripts (Study 1 experimental dataset and Study 2 survey dataset), with a fixed analysis pipeline from raw inputs to final tables/figures. Analyses were conducted primarily in **R** (version `4.3.2`) using packages for data wrangling and visualization (e.g., `tidyverse`/`ggplot2`) and for inferential modeling and contrasts (e.g., `lme4`, `lmerTest`, `emmeans`). All data transformations were scripted to avoid manual edits. When random procedures were used (e.g., resampling, bootstrap confidence intervals), a fixed random seed was set and reported.

To support transparency, the thesis follows a consistent reporting standard across outcomes: (i) descriptive statistics (means/SDs or proportions), (ii) inferential tests aligned with the measurement scale, (iii) effect sizes with confidence intervals when appropriate, and (iv) explicit sample sizes per analysis (including any exclusions due to missingness or compliance screening). Where multiple comparisons are performed within a family of tests (e.g., timing contrasts), p-values are adjusted using a principled correction procedure (e.g., Holm), and both raw and adjusted p-values are reported when space permits.

5.2. Constructs and Operationalization

This section defines the constructs used in each study and specifies how they map to the three research questions. The goal is to make the link from theory (AI Buffer mechanisms) to measured variables explicit, so that subsequent modeling choices and result interpretation follow transparently.

5| Data Analysis

The guiding principle for operationalization is alignment with the three research questions. RQ1 requires measuring memory mechanisms under AI-assisted external representations (the AI Buffer). RQ2 requires measuring how AI adoption reshapes decision practices and perceived responsibility in product management. RQ3 requires a mapping between the constructs: organizational patterns are interpreted through cognitive mechanisms.

Study 1 Constructs (RQ1). Study 1 operationalizes the AI Buffer as an AI-generated summary and defines experimental conditions via two design factors:

- **AI Buffer presence (between):** No-AI baseline vs. AI-assisted.
- **AI Buffer properties (AI participants):** *timing* (pre-reading, synchronous, post-reading) and *structure* (integrated paragraph, segmented bullet points).

Study 1 outcomes map directly to the RQ1 components:

- **Recall and recognition.** Free-recall responses capture reconstruction from internal traces, while recognition items capture cue-supported memory performance.
- **Source monitoring.** False-lure endorsement in recognition provides an index of epistemic risk and misattribution under AI assistance, motivated by source-monitoring theory [37].
- **Confidence calibration.** Confidence ratings are used to assess whether AI assistance increases confidence and whether confidence aligns with accuracy (calibration).
- **Cognitive load.** Post-block ratings of mental effort and perceived difficulty operationalize subjective cognitive load under each condition.
- **Behavioral reliance (supporting indicator).** Platform logs (e.g., time spent viewing the summary, number of openings, and reading time) operationalize reliance behavior and exposure to the AI Buffer.

Outcome definitions and interpretation rules (Study 1). To increase interpretability and replicability, outcomes are defined with explicit scoring rules. Unless otherwise stated, outcomes are computed at the block level (participant $\times$ article): (i) *recall quality* is scored using a predefined rubric (higher values indicate more accurate and complete reconstruction of article content); (ii) *recognition accuracy* is computed as the proportion of correct responses (and also modeled at item level as a binary correct/incorrect outcome when mixed-effects specifications are used); (iii) *false-lure endorsement* captures acceptance of plausible-but-incorrect statements and is interpreted as higher epistemic risk / weaker source monitoring; (iv) *confidence calibration* is assessed by relating confidence to correctness (e.g., calibration curves or regression of correctness on confidence), such that miscalibration corresponds to elevated confidence without commensurate accuracy; (v) *cognitive load* is interpreted such that higher values indicate greater subjective effort and/or perceived difficulty.

Scoring details (Study 1). Recall quality is scored as a single ordinal rubric score from 0 to 10 per article, based on predefined dimensions (including factual accuracy, structural alignment,

and penalties for false-lure intrusions). Recognition accuracy is computed as the proportion correct out of 14 multiple-choice items per article and is additionally modeled at item level as a binary response. False-lure endorsement is measured as the number of false-lure options selected (0–2 per article; 0–6 overall); higher values indicate increased misattribution risk. Confidence ratings are recorded on a 7-point scale and summarized per block; calibration analyses relate confidence to correctness (e.g., correctness regressed on confidence and/or calibration curves).
Study 2 Constructs (RQ2). Study 2 operationalizes product-management adoption of AI tools through survey measures that cover four domains:

- **AI usage and task coverage.** Frequency and types of AI tools used in daily PM work (automation, analytics, generative drafting/synthesis, decision support).
- **Workflow and role configuration.** Perceived task reallocation and role transformation (e.g., shift toward strategic work; changes in coordination, analysis, and leadership responsibilities).
- **Perceived responsibility and governance.** Accountability, transparency, oversight practices, fairness/bias concerns, and customer-trust implications when AI informs product decisions [67].
- **Context and moderation.** Region (USA vs. China) as a key grouping variable, allowing cross-context comparison aligned with RQ2's comparative component.

When multiple items measure a shared latent concept (e.g., perceived responsibility or AI literacy), items are combined into composite indices by averaging standardized item scores, subject to internal-consistency checks (see Section 5.7).
Cross-Study Constructs (RQ3). To answer RQ3, Study 2 patterns are interpreted through the cognitive mechanisms measured in Study 1. The mapping uses the AI Buffer Model constructs from Chapter 3:

- **Offloading and cue-dependent retrieval** (effort substitution and reliance norms).
- **Source monitoring and epistemic risk** (traceability and accountability).
- **Confidence calibration** (trust and escalation behavior).
- **Cognitive load and split attention** (interface costs and workflow friction).

5.3. Data Preparation and Quality Checks

Data preparation follows two goals: (i) ensure the dataset represents compliant task execution and valid responses, and (ii) ensure constructs reflect the intended operational definitions. The section is organized by study and specifies exclusion rules, missing-data handling, and harmonization steps so that the analytic datasets can be reconstructed from raw inputs.
Study 1 Data PreparationStudy 1 data are structured at the **block level** (participant × article), enabling within-person comparisons across timing conditions for AI participants while retaining

a No-AI baseline group.

Preparation includes:

- **Session completeness.** Exclude incomplete sessions and blocks with missing core outcomes (recall or recognition).
- **Compliance screening.** Flag and exclude clearly non-compliant behavior (e.g., implausibly short reading time suggesting rushing, non-substantive recall text).
- **Derived variables.** Compute block-level indices: recall score, recognition accuracy, false-lure endorsement count, confidence measures, cognitive-load ratings, and behavioral reliance metrics from logs.
- **Randomization checks.** Verify balance in article assignment and timing order (AI group), and confirm that key participant characteristics do not differ systematically across between-subject structure conditions.

Compliance thresholds (Study 1). Compliance thresholds are defined *a priori* in rule-based terms (e.g., reading time below a fixed minimum or below a low percentile of the distribution; recall entries that are empty, near-empty, or non-responsive). Exclusions are reported transparently by step (screening flow), and all main analyses are repeated in sensitivity form under a plausible alternative threshold to verify robustness. In practice, "implausibly short" reading time is operationalized using a conservative cutoff (e.g., below the 5th percentile of observed reading times or below an absolute minimum consistent with reading the material), and "non-substantive" recall is defined by empty entries or minimal content (e.g., fewer than `10` meaningful tokens) that cannot plausibly reflect genuine recall.

Study 2 Data PreparationStudy 2 data are structured at the **respondent level** (one row per product manager). Preparation includes:

- **Validity screening.** Retain only complete and internally consistent responses; remove duplicate or corrupted entries when identifiable.
- **Missing data handling.** For item non-response, apply listwise deletion for analyses requiring the item, and report the effective sample size for each test. If missingness is non-trivial for a composite construct, compute the index only when at least 70% of the items for that construct are present, and report the effective n for each composite.
- **Coding and harmonization.** Harmonize categorical response labels across languages; code open-ended responses into qualitative themes for interpretive analysis.
- **Cross-context consistency.** Ensure that the USA and China cohorts are treated symmetrically in variable coding, scaling, and aggregation to support valid comparison.

Cross-language harmonization and measurement comparability (Study 2)

Because Study 2 uses a bilingual instrument and a comparative design (USA vs. China), data preparation includes explicit checks for cross-language and cross-context comparability. First, response options and category labels are harmonized so that identical constructs map to iden-

tical codes across languages. Second, for any multi-item index (e.g., responsibility/governance perceptions), internal consistency is inspected overall *and* within each region to ensure that composites reflect a coherent latent construct in both contexts.
Finally, descriptive distribution checks are used to flag items that may be vulnerable to translation ambiguity or systematic interpretation differences (e.g., extreme-response styles). When such patterns appear, results are either (i) interpreted with caution and explicitly framed as context-sensitive, or (ii) supported with robustness checks using alternative operationalizations (e.g., binary recodes or nonparametric comparisons).

5.4. Analysis Strategy for Study 1 (Experiment)

This section defines the analytic choices used to estimate the effects of timing and structure on each memory mechanism. The goal is to align model specification with the experimental design and with the RQ1 decomposition.
The Study 1 analysis strategy mirrors the experimental design and the RQ1 decomposition into memory mechanisms. Because AI participants contribute repeated measures across three timing conditions, analyses account for within-participant dependence.
Analyses were conducted in **R** using a combination of mixed-design ANOVA and mixed-effects modeling. For ANOVA-style tests, sum-to-zero contrasts and Type-III tests were used to align inference with the factorial design. For mixed-effects models, random intercepts were specified for participants and (where applicable) for articles to account for repeated measurement and stimulus heterogeneity. Post-hoc contrasts were estimated using marginal-means frameworks with Holm correction for multiple comparisons.
Diagnostics, assumption checks, and sensitivity analyses
Model assumptions and fit are evaluated to ensure that inference is not driven by artefacts of specification. For ANOVA-style analyses, residual distributions and influential observations are inspected, and (where relevant) sphericity is assessed; if sphericity violations are detected, corrected degrees of freedom are used. For mixed-effects models, convergence diagnostics are checked and random-effects structures are simplified only when necessary to achieve stable estimation.
For count and binomial outcomes, dispersion and boundary behavior are explicitly checked (e.g., overdispersion, rare-event patterns, or quasi-separation). If diagnostics indicate mismatch, alternative specifications are used (e.g., binomial instead of Poisson for bounded counts, or robust/penalized alternatives when needed), and the choice is documented.
Sensitivity analyses are used to test robustness of conclusions under reasonable alternatives, including: (i) alternative exclusion thresholds for non-compliance (e.g., reading-time based filters), (ii) alternative outcome operationalizations (e.g., item-level vs. block-level recognition),

and (iii) inclusion of exposure/reliance covariates (e.g., summary viewing time) to test whether performance effects persist beyond differential interaction with the AI Buffer.

Primary comparisonsThe primary comparisons are:

- **AI vs. No-AI (baseline effect).** Compare AI-assisted blocks vs. No-AI blocks on recognition accuracy, recall quality, cognitive load, confidence, and epistemic risk.
- **Timing effects (AI only).** Compare pre-reading vs. synchronous vs. post-reading timing within the AI group.
- **Structure effects (AI only).** Compare integrated vs. segmented structures between AI participants.
- **Timing × structure (AI only).** Test whether timing effects differ by summary structure.

Models and outcome-specific approachesAnalyses are performed using regression models appropriate to outcome type:

- **Recall quality.** Modeled as a continuous or ordinal outcome at the block level (depending on score distribution), with predictors for timing and structure and participant-level random effects.
- **Recognition accuracy.** Modeled either at the item level (binary correct/incorrect) using logistic mixed-effects models, or at the block level using binomial or proportion models, with predictors for timing and structure and participant-level random effects. When recognition items can be mapped to the AI summary versus the source article (via pre-specified item tagging), items are partitioned into summary-aligned vs. article-only subsets to distinguish buffer-supported performance from article-dependent performance.
- **Source monitoring / false-lure endorsement.** Modeled as a count (e.g., 0–2 per article) or as a binary item-level event using count models or logistic mixed-effects models. The main inferential focus is on whether segmented structure increases false-lure endorsement relative to integrated structure.
- **Confidence and calibration.** Confidence is modeled as an outcome and also used as a predictor of accuracy to test calibration. Calibration is assessed by relating confidence to correctness (e.g., via regression or correlation-based indices) and by comparing calibration across conditions.
- **Cognitive load.** Mental-effort and difficulty ratings are modeled as continuous outcomes; timing is expected to increase load when it introduces interruptions and split attention, while integrated structure is expected to reduce extraneous load.
- **Behavioral reliance.** Summary view time and opening counts are analyzed descriptively and via mixed-effects models to test whether timing and structure change reliance behavior. These metrics are also used as covariates or sensitivity controls to assess whether performance effects persist beyond differential exposure.

Inference and reporting
Unless otherwise stated, tests are two-sided with $\alpha = 0.05$. Results are reported with (i) point estimates, (ii) effect sizes, and (iii) uncertainty intervals appropriate to the outcome and model class. For ANOVA-style tests, partial η^2 is reported alongside test statistics. For regression and mixed-effects models, effects are reported as unstandardized coefficients (or odds ratios for logistic models) with confidence intervals; when helpful, marginal means and contrasts are reported to align interpretation with the experimental design.
Because multiple outcomes are analyzed, multiplicity is handled at the level of clearly defined "families" of tests (e.g., timing contrasts for a given outcome). Within each family, p-values are adjusted (e.g., Holm), and the family definition is stated so that the reader can interpret inferential strength without ambiguity.
Interpretation prioritizes mechanism-level coherence over isolated significance: conclusions focus on whether recognition, source monitoring, confidence, cognitive load, and reliance metrics tell a consistent story about the AI Buffer properties (timing and structure), rather than on maximizing the count of statistically significant effects.

5.5. Analysis Strategy for Study 2 (Field Study)

This section specifies how survey items and open-ended responses are analyzed to characterize adoption patterns and perceived responsibility in product management. The goal is to report both overall patterns and USA–China differences in a way that is consistent across domains and reproducible.
The Study 2 analysis strategy is designed to characterize adoption patterns, role transformation, and perceived responsibility in product management (RQ2) and to support cross-context comparison (USA–China).
Following the logic of Section 4.3, each survey item is interpreted through two linked lenses: an *overall* lens (what pattern is present across the full sample) and a *comparative* lens (whether the USA and China cohorts differ). This dual emphasis ensures that the study can report both (i) general adoption patterns that plausibly reflect role-level dynamics in product management and (ii) context-contingent differences that may reflect organizational environments and governance norms [42, 58].
Quantitative analysis of structured itemsFor Likert-scale items, the analysis proceeds in three steps:

- **Descriptive statistics.** Compute means, standard deviations, and distribution plots for each item and for composite indices (where applicable).
- **Overall effects.** Test whether the overall sample indicates above-neutral endorsement of key statements (e.g., increased strategic focus, increased governance responsibility) using

one-sample tests against the neutral midpoint (typically 3 on a 1–5 scale).

- **Cross-context comparison.** Compare the USA and China cohorts using independent-samples tests for continuous-like outcomes and contingency analyses for categorical outcomes.

Because Study 2 includes many single-item Likert measures, Likert responses are treated as approximately continuous for primary analyses (means-based descriptives; one-sample tests against the neutral midpoint; and between-region comparisons), which is common when the goal is to summarize central tendency and when distributions are not severely skewed. To ensure conclusions are not artefacts of scale assumptions, key inferences are checked for directional consistency using robustness alternatives (e.g., rank-based or ordinal comparisons) when items show marked non-normality or ceiling/floor effects.

Given the breadth of survey items, inference is interpreted at the level of conceptually grouped families (e.g., "workload reduction", "decision practices", "governance/responsibility", "training/readiness"). Within each family, p-values are adjusted (e.g., Holm) to reduce false positives; patterns are emphasized when they replicate across multiple indicators within the same conceptual domain rather than relying on a single item.

Families are defined *ex ante* based on the questionnaire structure, for example: (i) adoption intensity and tool portfolio, (ii) workload/task-shift effects, (iii) decision-making practices, (iv) training/readiness and AI literacy, and (v) governance/responsibility and ethical concerns.

For categorical or multiple-choice items (e.g., tool selection, governance ownership, training modes), chi-square tests are used to evaluate (i) whether observed distributions differ from uniform or expected baselines (goodness of fit) and (ii) whether response distributions differ by region (tests of independence). When distributions are sparse, categories are consolidated to preserve interpretability.

Qualitative analysis of open-ended responsesOpen-ended responses are analyzed using thematic coding to identify recurring patterns relevant to RQ2, including: examples of AI-supported decisions, perceived risks (bias, opacity, over-reliance), governance practices, and skill-development narratives. Themes are compared across contexts (USA–China) to identify convergences and divergences, and qualitative insights are used to contextualize quantitative patterns.

To improve transparency and replicability, coding follows a staged approach: (i) open coding to identify candidate themes, (ii) consolidation into a codebook aligned with the measurement domains in Section 4.3, and (iii) axial coding to connect themes to role transformation and governance narratives. Where themes differ by region, interpretation is tied back to the comparative logic of RQ2 (context-contingent adoption patterns).

Coding reliability and transparency are supported through an audit trail: the evolving codebook, representative excerpts for each theme, and notes on boundary cases are retained to document interpretive decisions. When codes require subjective judgment (e.g., distinguishing

"over-reliance" from "efficiency" narratives), ambiguous cases are revisited and resolved using the codebook definitions to maintain internal consistency. Qualitative results are reported with brief illustrative excerpts (anonymized) and are used primarily to (i) explain why quantitative patterns may arise and (ii) surface mechanisms that are not captured by fixed-response items. Coding was performed by a single analyst, with a structured second-pass review to reduce drift and to ensure consistent application of code definitions across contexts.

5.6. Cross-Study Integration Logic (RQ3)

RQ3 asks how changes in organizational decision-making practices enabled by AI can be explained through underlying cognitive mechanisms of AI-assisted memory. The integration logic follows a mechanism-to-practice mapping approach:

1. **Extract mechanism signatures from Study 1.** Identify which AI Buffer properties (timing, structure) change (i) performance (recall/recognition), (ii) epistemic risk (false-lure endorsement/source monitoring), and (iii) metacognitive and effort signals (confidence and cognitive load).
2. **Extract practice patterns from Study 2.** Identify how AI tools are used in product management (automation vs. augmentation vs. decision support), what role shifts are reported, and how responsibility and governance are perceived and implemented.
3. **Map patterns to mechanisms.** Interpret practice patterns through the AI Buffer mechanisms:
    - Workflow acceleration and increased strategic focus are mapped to *offloading and cue-supported cognition* (effort substitution and faster evidence access).
    - Governance emphasis and accountability concerns are mapped to *source monitoring* and *epistemic risk* (traceability failures under misattribution).
    - Trust and reliance narratives are mapped to *confidence calibration* (when confidence is increased by fluency rather than by accuracy).
    - Adoption friction and interface concerns are mapped to *cognitive load* and *split attention* (coordination costs and interruptions).
4. **Specify boundary conditions.** Use Study 2 context differences (USA–China) and reported skill differences (AI literacy and training) as candidate moderators of the mechanism-to-practice link, consistent with the framework in Section 3.4.

The goal of integration is explanatory coherence: the thesis does not treat Study 1 and Study 2 as two separate projects, but as complementary evidentiary sources that jointly support the AI Buffer Model as an account of how AI-generated external representations reshape knowledge work.

5.7. Validity, Reliability, and Robustness

Because the thesis combines experimental and field designs, validity and robustness are addressed at multiple levels.

Study 1**Internal validity.** Randomization and controlled timing constraints support causal interpretation of timing and structure effects. Compliance screening reduces noise from noncompliance.

Construct validity. RQ1 constructs are measured with complementary outcomes (recall, recognition, false lures, confidence, cognitive load, logs), reducing reliance on any single metric. False-lure endorsement provides a direct operationalization of epistemic risk rooted in source-monitoring theory [37].

Reliability. When qualitative scoring is required (e.g., recall quality), reliability is supported by a pre-specified rubric and by systematic review procedures. Sensitivity checks include verifying that results do not hinge on a small number of ambiguous scoring cases.

Robustness. Robustness checks include alternative model specifications (e.g., item-level vs. block-level recognition models), inclusion/exclusion of covariates (e.g., reading time and summary exposure), and alternative exclusion thresholds for compliance.

Study 2**External validity.** The field study measures AI adoption in a real decision-intensive role. The cross-national design increases generalizability by comparing two influential contexts (USA and China).

Measurement reliability. For multi-item constructs, internal consistency is assessed (e.g., Cronbach's alpha) and items are combined only when coherence is acceptable. For qualitative themes, coding consistency is supported through structured codebooks and review.

Cross-context comparability. The bilingual instrument design is intended to reduce language barriers and support semantic equivalence across cohorts. Analyses check that observed differences are not driven solely by systematic missingness or coding artifacts.

Cross-study robustnessCross-study claims are treated as *explanatory* rather than purely statistical. Robustness is evaluated by triangulation: an integrative claim is considered stronger when (i) it is consistent with Study 1 mechanisms, (ii) it aligns with Study 2 reported practices, and (iii) it remains plausible under alternative boundary conditions (e.g., differences in AI literacy, governance maturity, or time pressure).

6 | Results

6.1. Study 1 Results (RQ1)

This section reports the outcomes of the controlled experiment and addresses **RQ1**: *How do AI-generated external representations influence core human memory processes during knowledge-intensive tasks?* Results are organized around the mechanisms specified in the AI Buffer Model (encoding scaffolding, cue-dependent retrieval, source monitoring, confidence calibration, and cognitive load; Section 3.2).

Sample overview and outcome structureThe experimental dataset includes $N = 36$ participants (AI: $n = 24$; No-AI: $n = 12$), each completing three reading blocks with comprehension and memory assessments, yielding 108 block-level observations. Within the AI group, the experimental manipulation followed a 2 (summary structure: integrated vs segmented) $\times$ 3 (summary timing: pre-reading vs synchronous vs post-reading) mixed design (details in Section 4.2).

Reporting conventionsFor each outcome, we report descriptive statistics (mean and standard deviation) followed by inferential tests. For tests within the AI group, timing is treated as a within-participant factor and structure as a between-participant factor. Where post-hoc timing contrasts are reported, p-values are Holm-corrected within the relevant family of comparisons. Unless stated otherwise, analyses use the complete-case dataset described in Sections 4.2 and 5.4.

Recognition and learning outcomes (MCQ accuracy)

AI vs No-AI baseline. Across all multiple-choice items, the AI condition outperformed the No-AI condition (AI: $M = 0.598$, $SD = 0.085$; No-AI: $M = 0.510$, $SD = 0.098$), $F(1, 34) = 7.86$, $p = .008$, $\eta_p^2 = 0.188$. This establishes a global recognition benefit associated with AI summaries. In absolute terms, this corresponds to an 8.8 percentage-point gain in recognition accuracy (Cohen's $d \approx 0.98$).

Interpretation. This baseline effect indicates that AI summaries primarily support performance under cue-rich recognition conditions. Because recognition can be improved by the availability of well-formed retrieval cues, this gain should not be interpreted as evidence of durable internalization on its own. The free-recall analyses reported below therefore serve as a complementary

test of whether AI assistance strengthens generative retrieval without external cues.

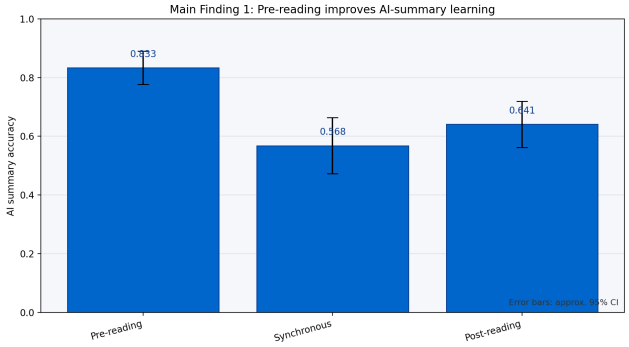
Group	Overall MCQ accuracy (Mean ± SD)	N
AI-assisted	0.598 ± 0.085	24
No-AI	0.510 ± 0.098	12

Table 6.1: Overall MCQ accuracy by AI availability (Study 1).

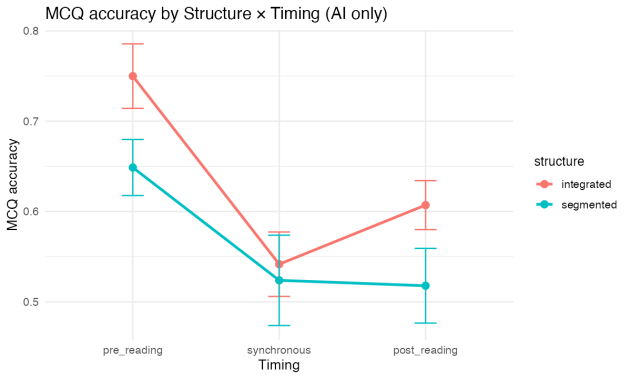
Timing effects within the AI group. When recognition is indexed specifically to information covered by the AI summary (*AI-summary-sourced* questions), pre-reading access yields the highest accuracy. AI-summary-sourced accuracy by timing is: pre-reading $M = 0.833$ ($SD = 0.141$), synchronous $M = 0.568$ ($SD = 0.239$), and post-reading $M = 0.641$ ($SD = 0.196$). A mixed ANOVA (structure × timing; AI group) shows a strong timing main effect, $F(2, 44) = 14.00$, $p < .001$, $\eta_p^2 = 0.389$, with no structure main effect and no interaction. Holm-corrected comparisons indicate pre-reading outperforms synchronous ($p = .00052$, $d_z = 0.91$) and post-reading ($p = .00057$, $d_z = 0.87$); the synchronous vs post-reading contrast is not significant ($p = .148$).

Interpretation. The timing pattern is consistent with a sequencing mechanism: when the summary is available before reading, it can function as an advance organizer that guides what is attended to and how information is structured during encoding. This interpretation is strengthened by the fact that the effect is concentrated in AI-summary-sourced items rather than reflecting a broad uplift in comprehension across all question types.

Overall MCQ accuracy (all questions). The same timing ordering appears in overall MCQ accuracy (pre-reading $M = 0.699$, $SD = 0.125$; synchronous $M = 0.533$, $SD = 0.147$; post-reading $M = 0.562$, $SD = 0.127$), with a significant timing effect, $F(1.77, 38.87) = 11.77$, $p < .001$.



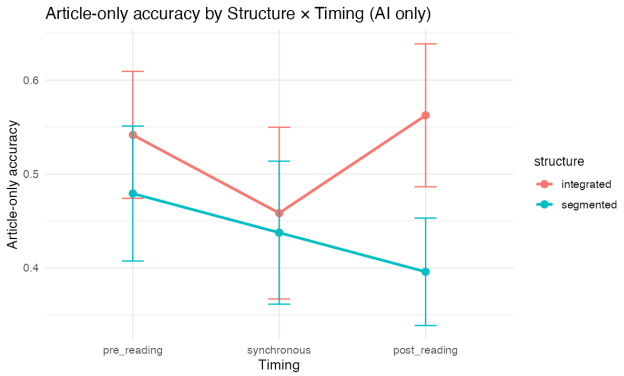
(a) AI-summary-sourced accuracy by timing (AI group).



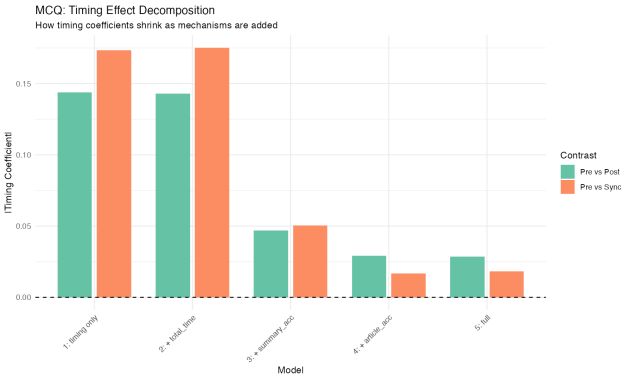
(b) Overall MCQ accuracy by structure and timing (AI group).

Figure 6.1: Recognition outcomes in Study 1. Error bars indicate ±SE.

Boundary condition: article-only comprehension. Timing does not meaningfully affect accuracy on questions that are not covered by the AI summary (article-only accuracy), indicating that the timing manipulation primarily changes learning for summary-covered information rather than broadly improving comprehension of the underlying article.



(a) Article-only accuracy by timing (AI group).



(b) Decomposition of timing effects into summary- vs article-sourced accuracy (AI group).

Figure 6.2: Timing diagnostics for Study 1: timing primarily affects learning of summary-covered content.

Mechanism-oriented robustness checks for the timing advantageTo test whether the timing advantage is explained by summary-aligned encoding and/or by increased exposure to the summary, two complementary model-based checks were conducted.

Summary-alignment mechanism. AI-summary-sourced accuracy strongly predicts overall MCQ accuracy in a mixed-effects model controlling for timing and structure ($\beta = 0.472$, $p < .001$). When AI-summary-sourced accuracy is included as a predictor, the pre-reading advantage on overall MCQ accuracy shrinks substantially (pre–synchronous: $\Delta = 0.172 \rightarrow 0.043$; pre–post: $\Delta = 0.143 \rightarrow 0.044$) and is no longer significant after Holm correction ($p = .352$). Model fit improves markedly ($\chi^2(1) = 46.38$, $p < .001$), supporting the interpretation that timing primarily increases learning of summary-covered content rather than broadly raising comprehension.

Exposure control (summary time). Controlling for log-transformed summary exposure reduces the timing estimates only modestly, and the timing contrasts remain significant, suggesting that pre-reading advantages are not explained solely by "more time" on the summary. Summary

time itself is a positive predictor of AI-summary-sourced accuracy ($\beta = 0.062$, $p = .031$).

Model	Pre–Synchronous estimate	Pre–Post estimate
Base (timing only)	0.271***	0.214***
+ log(summary_time)	0.258***	0.164**
Reduction	**5%**	**23%**

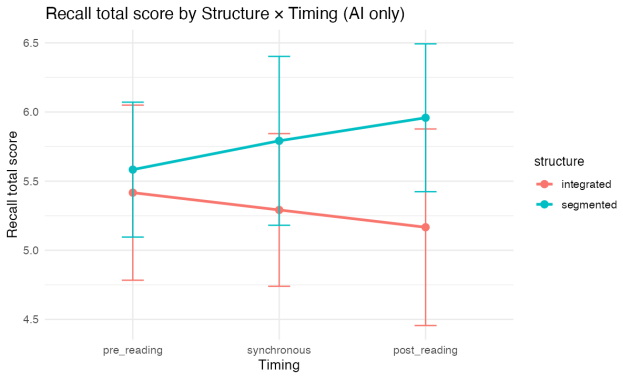
Table 6.2: Timing contrasts on AI-summary-sourced accuracy with and without controlling for summary exposure.

Note: *** $p < .001$, ** $p < .01$.

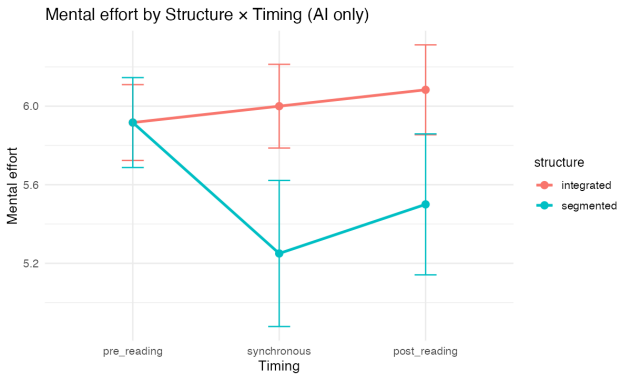
Generative retrieval: free recall is unaffected by timing

In contrast to recognition performance, free recall does not vary by timing in the AI group: pre-reading $M = 5.50$ ($SD = 1.92$), synchronous $M = 5.54$ ($SD = 1.99$), and post-reading $M = 5.56$ ($SD = 2.17$). A mixed ANOVA shows no timing effect, $F(1.86, 40.88) = 0.03$, $p = .969$. Additionally, an AI vs No-AI comparison indicates no meaningful recall difference (AI: $M = 5.535$; No-AI: $M = 5.403$; $p > .50$). Together, these findings indicate that AI assistance primarily affects cue-supported recognition rather than internally generated retrieval.

Interpretation. The absence of timing and baseline differences in free recall suggests that the recognition gains are not driven by a uniformly stronger memory trace. Instead, AI assistance likely changes the organization and availability of cues that benefit recognition while leaving generative reconstruction largely unchanged. This distinction matters for downstream tasks that require explanation, transfer, or reconstruction without access to external supports.



(a) Recall total score by timing (AI group).



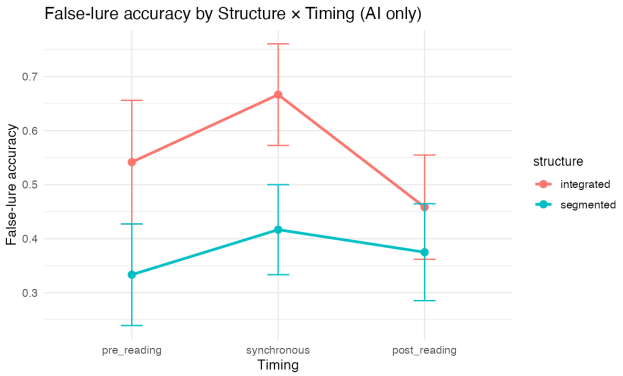
(b) Mental effort by condition (AI and No-AI).

Figure 6.3: Secondary outcomes in Study 1: recall and cognitive load. Error bars indicate ±SE.

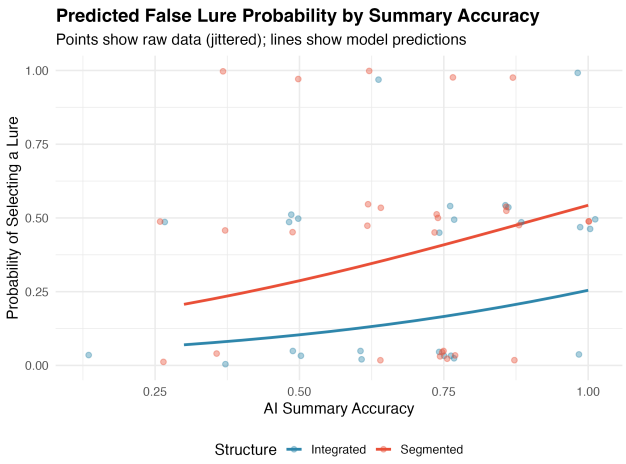
Source monitoring and misinformation: structure shapes false-lure endorsement

Susceptibility to false AI-generated claims is primarily driven by summary structure. In the AI group, segmented summaries increase false-lure endorsement relative to integrated summaries. Participants exposed to segmented summaries select more false lures (integrated: $M = 0.58$, $SD = 0.69$; segmented: $M = 1.06$, $SD = 0.79$), and show lower false-lure accuracy (integrated: $M = 0.556$, $SD = 0.354$; segmented: $M = 0.375$, $SD = 0.302$). A binomial GLMM predicting false-lure endorsement yields an odds ratio of 5.93 for segmented structure (95% CI $[1.63, 21.5]$; $p = .007$).

Interpretation. The structure effect suggests that representational fragmentation can weaken source-monitoring cues: when outputs are segmented, claims may be encoded with less contextual binding to their origin, increasing later endorsement of false lures. This complements the calibration analyses by implying that misinformation vulnerability here is better explained by degraded provenance cues and cross-checking opportunities than by a general increase in confidence.



(a) False-lure accuracy by structure and timing (AI group).

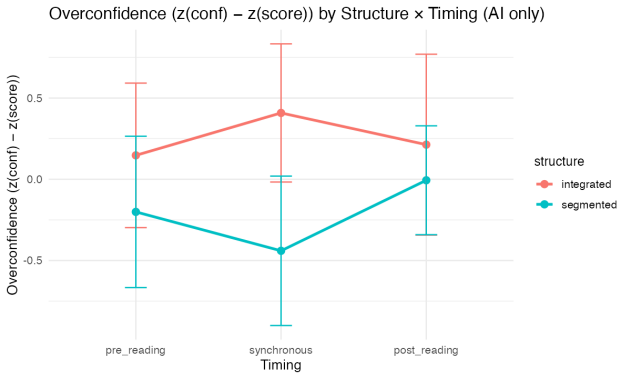


(b) False-lure endorsement probability by structure (AI group).

Figure 6.4: Source-monitoring outcomes in Study 1. Error bars indicate ±SE.

Confidence calibration

Metacognitive calibration (confidence relative to actual performance) is imperfect overall, but AI does not worsen overconfidence. Recall confidence is similar across AI availability (AI: $M = 4.25$; No-AI: $M = 4.08$; scale 1–7), and an independent-samples t-test on overconfidence indicates no difference between groups, $t(31.7) = 0.16$, $p = .873$.



(a) Overconfidence by group (AI vs No-AI).

(b) Recall calibration by group.

Figure 6.5: Confidence calibration in Study 1. Error bars indicate ±SE.

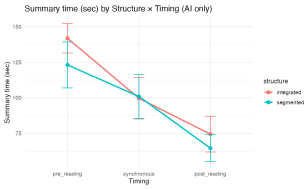
Process evidence: timing redistributes attention without increasing total time

The timing manipulation changes how attention is allocated between summary and article. In the AI group, pre-reading produces the highest summary viewing time (mean 132.5 seconds) and summary share (24.9%), while post-reading is lowest (mean 69.5 seconds; 13.7%). Importantly, total time-on-task remains stable across timing conditions (approximately 531–536 seconds), indicating a redistribution of attention rather than more time spent overall.

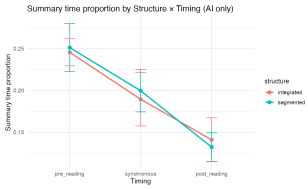
Interpretation. Because total time-on-task is stable, the most parsimonious account is a redistribution of attention rather than an effort increase. Pre-reading appears to shift time toward the external representation at a moment when it can shape subsequent processing, providing process-level support for the encoding-scaffold interpretation of the timing advantage reported above.

Timing (AI group)	Summary time (s)	Reading time (min)	Total time (s)	Summary share ('
Pre-reading	132.5	6.72	535.8	24.9
Synchronous	100.3	7.19	531.6	19.5
Post-reading	69.5	7.69	530.8	13.7

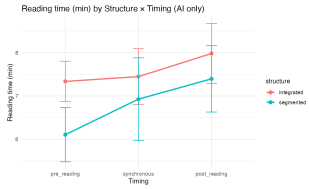
Table 6.3: Time allocation by summary timing (Study 1, AI group; means).



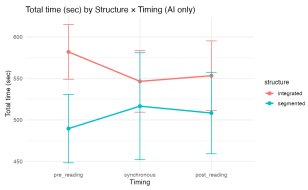
(a) Summary time (seconds).



(b) Summary share of total time.



(c) Reading time (minutes).

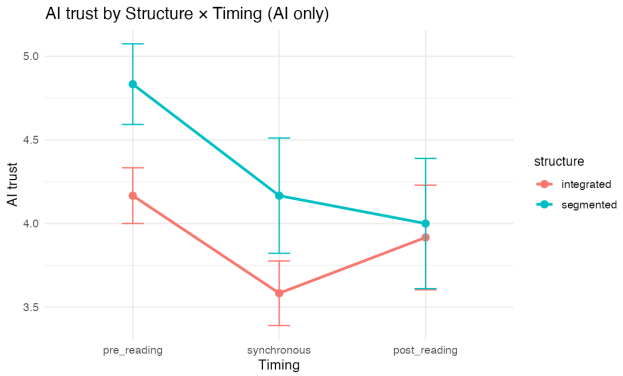


(d) Total time (seconds).

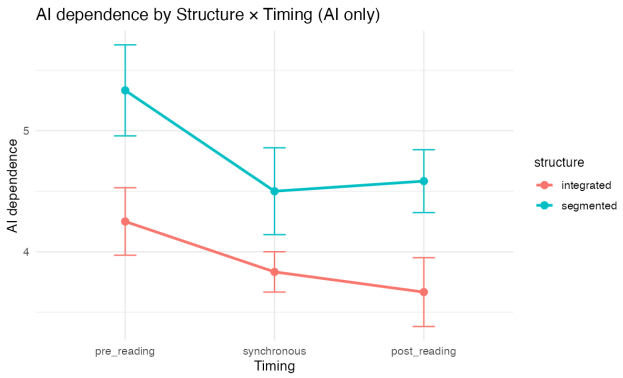
Figure 6.6: Process measures in Study 1 (AI group). Error bars indicate ±SE.

Subjective reliance: trust and dependence

Self-reported reliance measures track the timing manipulation more than the structure manipulation. Pre-reading tends to yield higher trust and dependence ratings than post-reading, consistent with the process evidence that pre-reading increases summary engagement. Mixed-design analyses show a timing effect for trust ($F(2, 43) = 7.90$, $p = .001$) and both structure and timing effects for dependence (structure: $F(1, 22) = 6.21$, $p = .021$; timing: $F(2, 43) = 7.74$, $p = .001$). These measures are reported as complementary signals alongside behavioral logs; full post-hoc contrasts are reported in Appendix A.



(a) Trust by structure and timing (AI group).

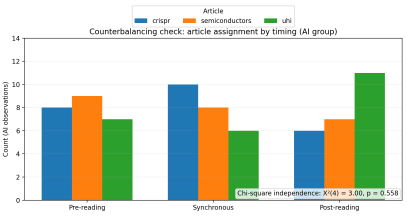


(b) Dependence by structure and timing (AI group).

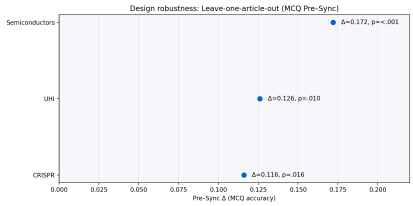
Figure 6.7: Subjective reliance in Study 1 (AI group). Error bars indicate ±SE.

Robustness checks and supplementary diagnostics

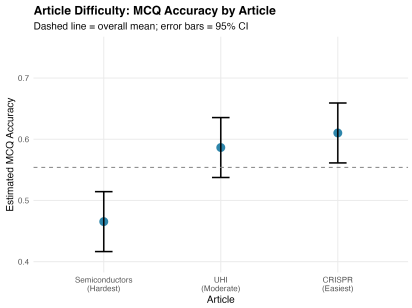
Robustness checks support the stability of the main conclusions. Counterbalancing diagnostics indicate that timing conditions were distributed across articles as intended; leave-one-article-out analyses show that the timing advantage for AI-summary-sourced accuracy persists when any single article is removed. Supplementary plots also characterize potential stimulus heterogeneity (e.g., article difficulty) and the decomposition of timing effects.



(a) Counterbalancing check: timing by article.



(b) Leave-one-article-out robustness (timing effect).



(c) Article difficulty diagnostic.

Figure 6.8: Robustness and diagnostic checks for Study 1.

6.2. Study 2 Results (RQ2)

This section reports results from the field study and addresses **RQ2**: *How does the adoption of AI tools shape decision-making practices, role configuration, and perceived responsibility in product management?* The final dataset comprises $N = 74$ validated responses, evenly split between the United States ($n = 37$) and China ($n = 37$) (Section 4.3).

Reporting conventions (Study 2)Unless otherwise stated, Likert items (1–5) are summarized with mean (SD) overall and by region, and tested (i) against the neutral midpoint (3) using one-sample t-tests, and (ii) between USA and China using independent-samples t-tests. Categorical items are summarized with response frequencies and tested using chi-square goodness-of-fit and independence tests (USA vs China). Alongside p-values, we report effect sizes (Cohen's d for mean contrasts; Cramérś V for chi-square tests) where relevant.

Adoption patterns and tool portfolio

Respondents report broad AI adoption across multiple tool categories. Table 6.4 reports mean usage intensity for the main tool portfolio items (1–5; higher indicates more frequent use). Overall, the highest-intensity item is the *generative writing assistant* ($M = 3.22$, $SD = 1.10$), while several automation and analytics tools cluster around moderate usage (approximately $M \approx 2.8$–3.0).

Tool / use case (1–5)	Mean	SD
RPA bots	2.99	0.95
Intelligent scheduling assistant	2.94	1.02
Customer service AI chatbot	2.88	1.17
Customer feedback analysis tool	2.84	1.15
Market research automation	2.84	1.04
Generative writing assistant	3.22	1.10
AI-driven product testing	2.80	1.01
Ideation assistant	2.66	1.02
Competitor analysis platform	2.58	0.98
Predictive analytics tool	2.96	1.09

Table 6.4: Study 2 AI tool portfolio usage intensity (overall descriptives; 1–5).

Expectations are generally met or exceeded ("AI tools met initial expectations": $M = 3.52$, $SD = 1.08$) and respondents anticipate continued growth in AI usage over the next 12 months ($M = 3.66$, $SD = 0.90$).

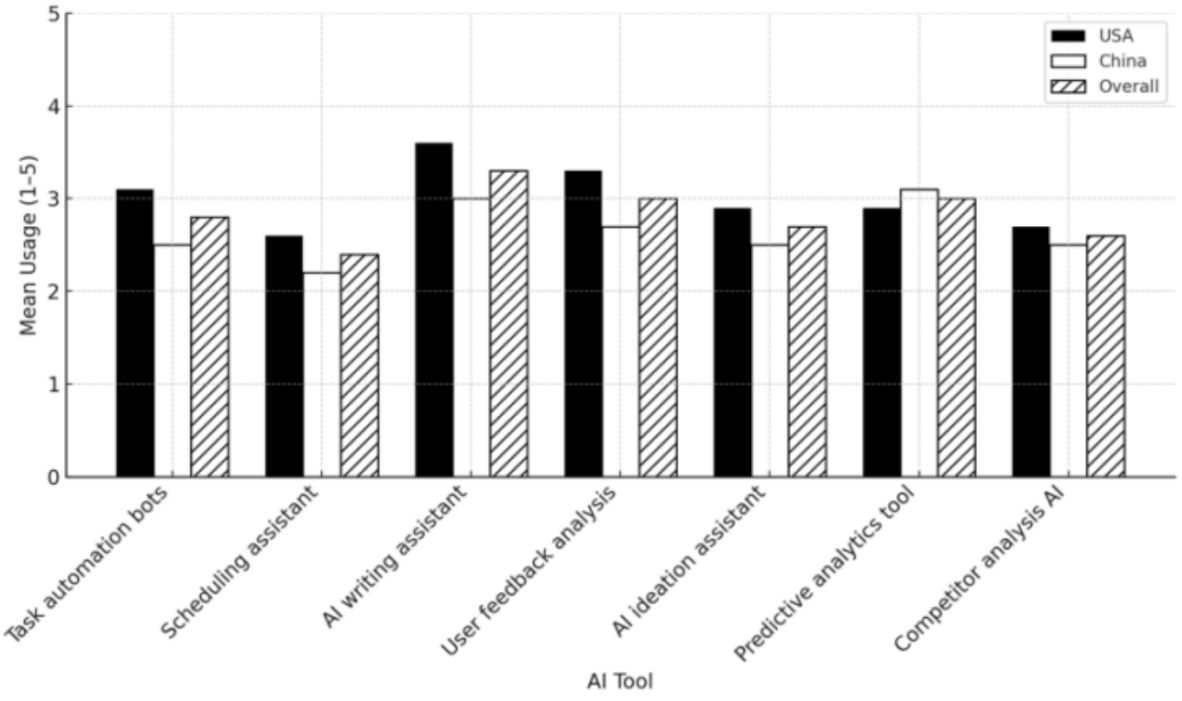


Figure 6.9: Self-reported AI tool usage intensity in product management (Study 2; error bars show ±SD).

Strategic landscape of AI adoption and organizational determinants

Table 6.5 summarizes adoption maturity and forward expectations. Respondents report being within the "1–3 years" range on average for AI usage in product management (scale coded 1–5), indicate positive perceived effectiveness of AI tools, and anticipate increased usage over the next 12 months. No meaningful USA–China differences emerge for these strategic-orientation indicators.

Variable (1–5)	**Mean**	**SD**	$t(73)$ **vs 3**	p
Duration of AI usage in PM	3.35	0.89	-5.66	$< .001$
AI tools are effective for PM decisions	3.38	0.92	4.15	$< .001$
Expected growth in AI usage (12 months)	3.66	0.90	6.15	$< .001$

Table 6.5: Strategic orientation toward AI integration (Study 2): one-sample tests against the neutral midpoint (3).

Organizational determinants (who drives AI implementation decisions, integration goals, perceived barriers, and tool selection criteria) are summarized in Table 6.6. Distributions are non-uniform (overall goodness-of-fit tests $p < .001$), indicating concentration of decision ownership and selection logic rather than diffuse, evenly distributed practices.

Categorical item	χ^2 **(overall)**	p **(overall)**	χ^2 **(USA vs CN)**	p **(USA vs CN)**
Decision-making group for AI implementation	108.99	$< .001$	22.27	$< .001$
AI integration goals	10.58	.102	0.19	.666
Preferred AI tool category	22.04	$< .001$	9.70	.0078
AI adoption barriers	7.32	.292	2.00	.157
Tool selection criteria	16.84	.0048	0.06	.804

Table 6.6: Strategic drivers and organizational determinants of AI adoption (Study 2): goodness-of-fit (overall) and USA–China independence tests.

Workload shifts and role reconfiguration

Workload reduction items are measured on a 1–5 scale (higher values indicate stronger perceived reduction). Table 6.7 shows strong perceived reductions for routine administrative work, customer-feedback analysis, documentation/content creation, and task/backlog management. Routine communications are the only domain with a strong *negative* deviation below the midpoint.

Task (1–5)	Mean	SD	$t(73)$ vs 3	p
Routine administrative tasks (reporting, status updates)	3.88	1.11	6.49	$< .001$
Customer feedback analysis (NLP-supported synthesis)	3.84	1.05	6.56	$< .001$
Market research and competitor analysis	2.97	0.73	-0.49	.625
AI-driven QA / product testing	2.97	0.75	1.18	.241
Documentation / content creation	3.70	0.93	6.48	$< .001$
Routine communications	2.22	0.85	-7.95	$< .001$
Task/backlog management	3.53	1.01	4.49	$< .001$

Table 6.7: Workload reduction across product-management tasks due to AI integration (Study 2): one-sample tests against midpoint (3).

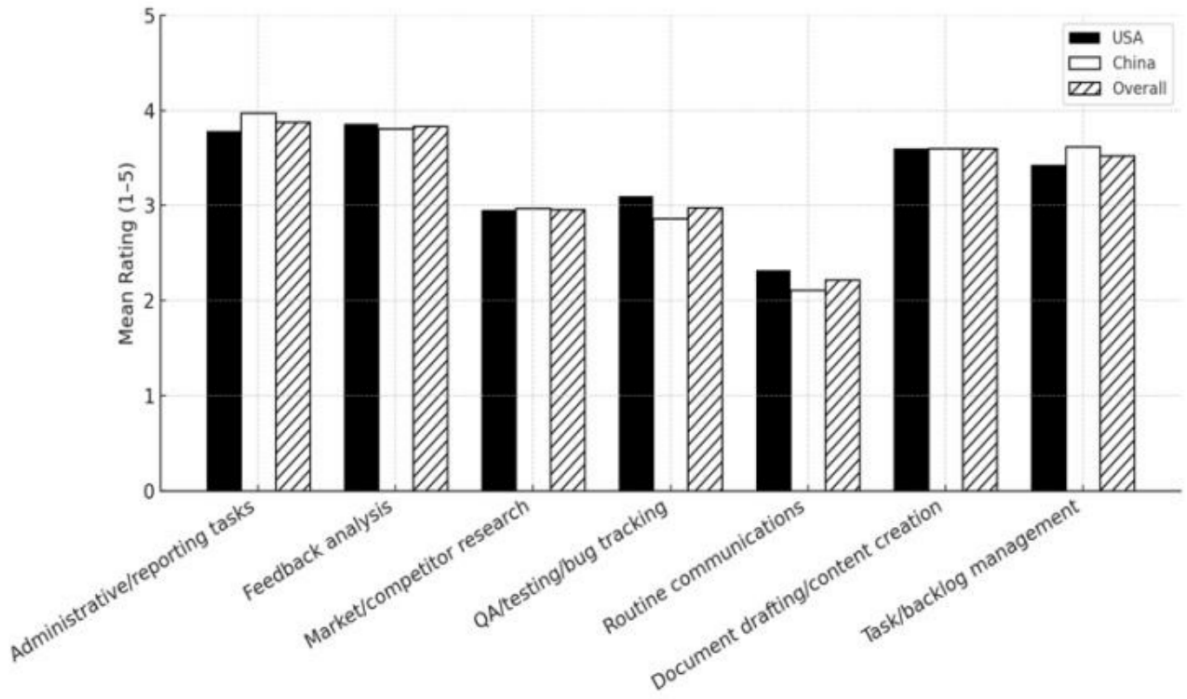


Figure 6.10: Perceived workload reduction by product-management task (Study 2; error bars show ±SD).

Task complexity increases and time reinvestment patterns

AI adoption does not only reduce workload; it also shifts work toward validation and governance.

Item	Most frequent response	%	*N*
Task complexity increased (top)	Human-in-the-loop checks	75.7%	74
Time reallocated to (top)	Product innovation	85.1%	74

Table 6.8: Top complexity increase and top time-reinvestment category (Study 2). Full distributions by region are reported in Appendix C.

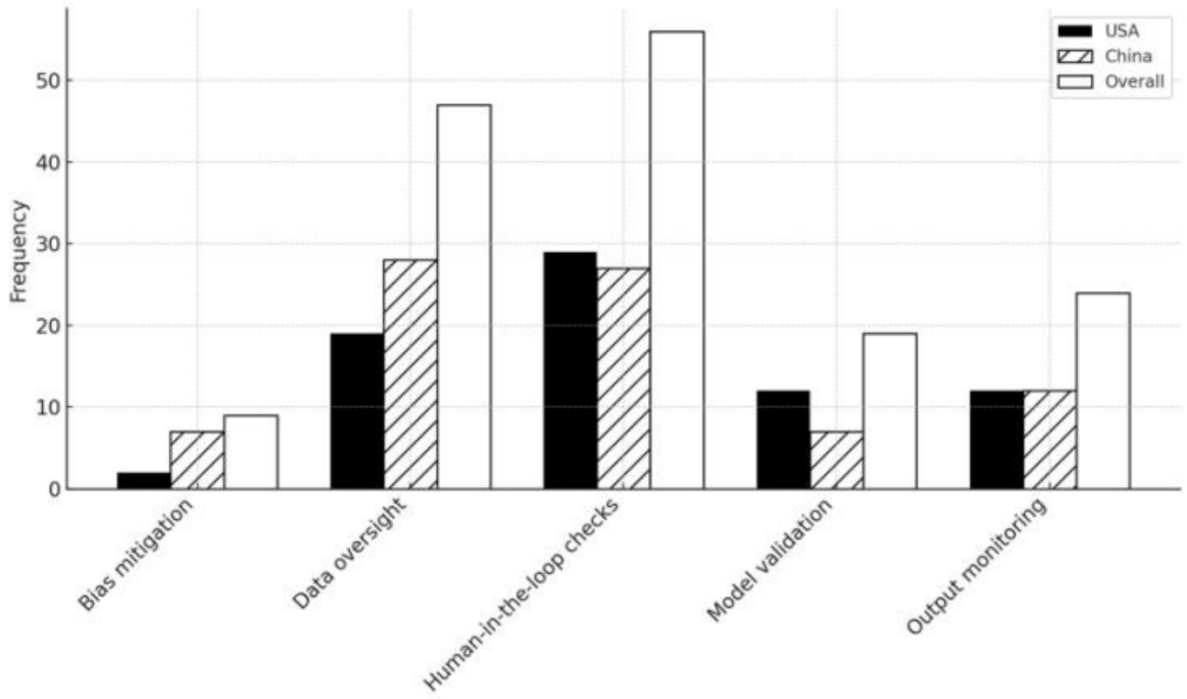


Figure 6.11: Tasks reported as more complex due to AI adoption (Study 2; overall and by region).

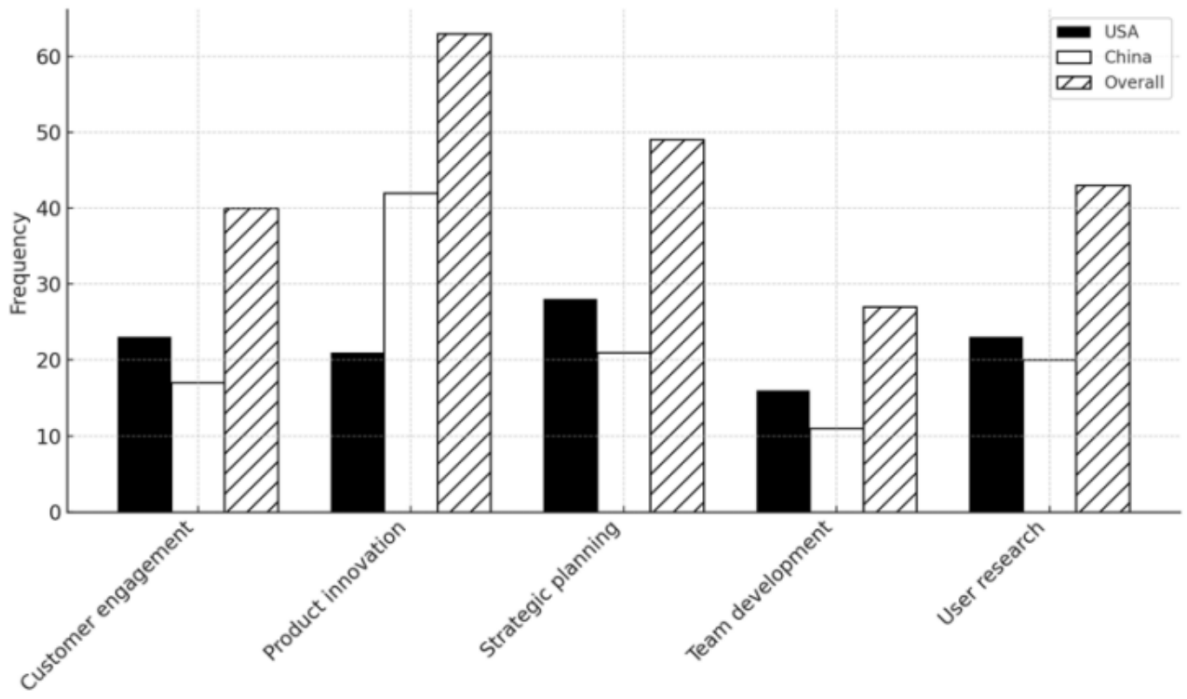


Figure 6.12: Tasks prioritized with time saved through AI (Study 2; overall and by region).

Decision-making support and strategic impact

Strategic impact items are measured on 1–5 scales (higher values indicate more positive perceived impact), except for reported strategic time gained, which is measured in hours per week. Table 6.9 reports overall endorsement and USA–China contrasts.

Outcome	**Mean**	**SD**	$t(73)$	p	p **(USA vs CN)**
Overall productivity impact (1–5)	3.47	1.10	4.77	$< .001$	0.999
Perceived decision-making improvement (1–5)	3.45	1.05	3.34	.0013	.0104
Strategic time gained (hours/week)	0.38	0.50	5.33	$< .001$	0.751
AI tools meeting expectations (1–5)	3.52	1.08	4.14	$< .001$	0.770

Table 6.9: Strategic impacts of AI integration (Study 2): one-sample tests and USA–China comparisons.

Interpretation. Means around 3.4–3.5 indicate moderately positive perceived impact rather than uniformly transformative effects, and the standard deviations suggest substantial heterogeneity across respondents. The significant USA–China difference in perceived decision improvement, alongside similar productivity ratings, is consistent with contextual differences in how AI is embedded into decision routines; however, because these outcomes are self-reported,

the contrast should be interpreted as perceived value rather than objective performance differences.

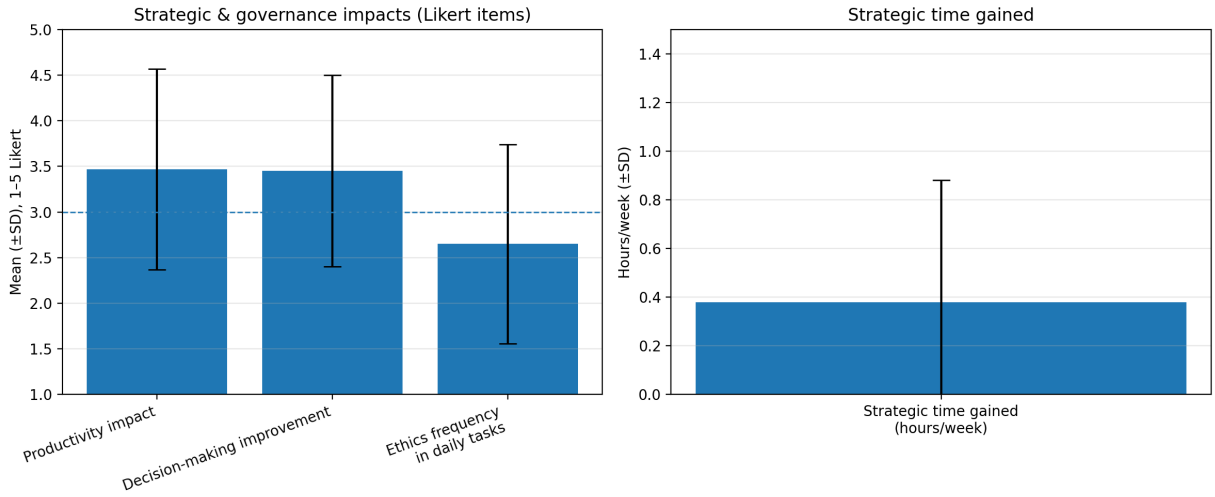


Figure 6.13: Strategic and governance impacts of AI adoption (Study 2; error bars show ±SD).

Confidence, training, and skill development

To capture adoption maturity, we report perceived confidence using AI tools, the availability of structured training, and which skill domains respondents perceive as most improved due to AI integration. These outcomes are reported overall and contrasted across USA and China.

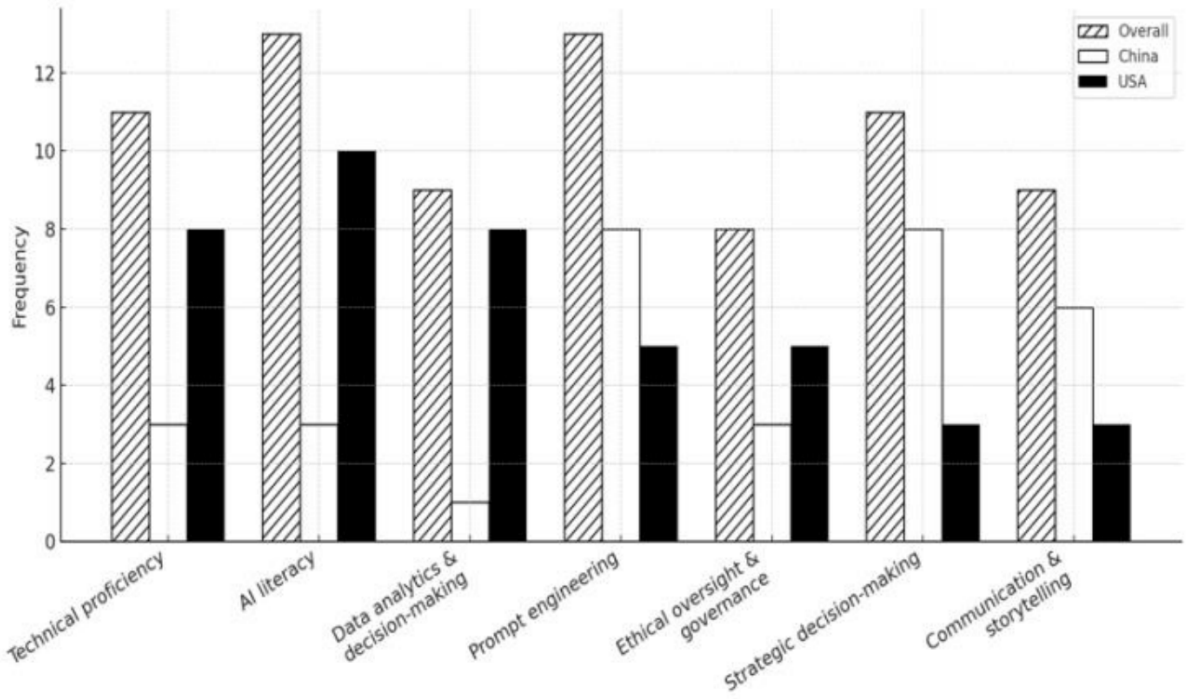


Figure 6.14: Top skills improved due to AI integration (Study 2; overall and by region).

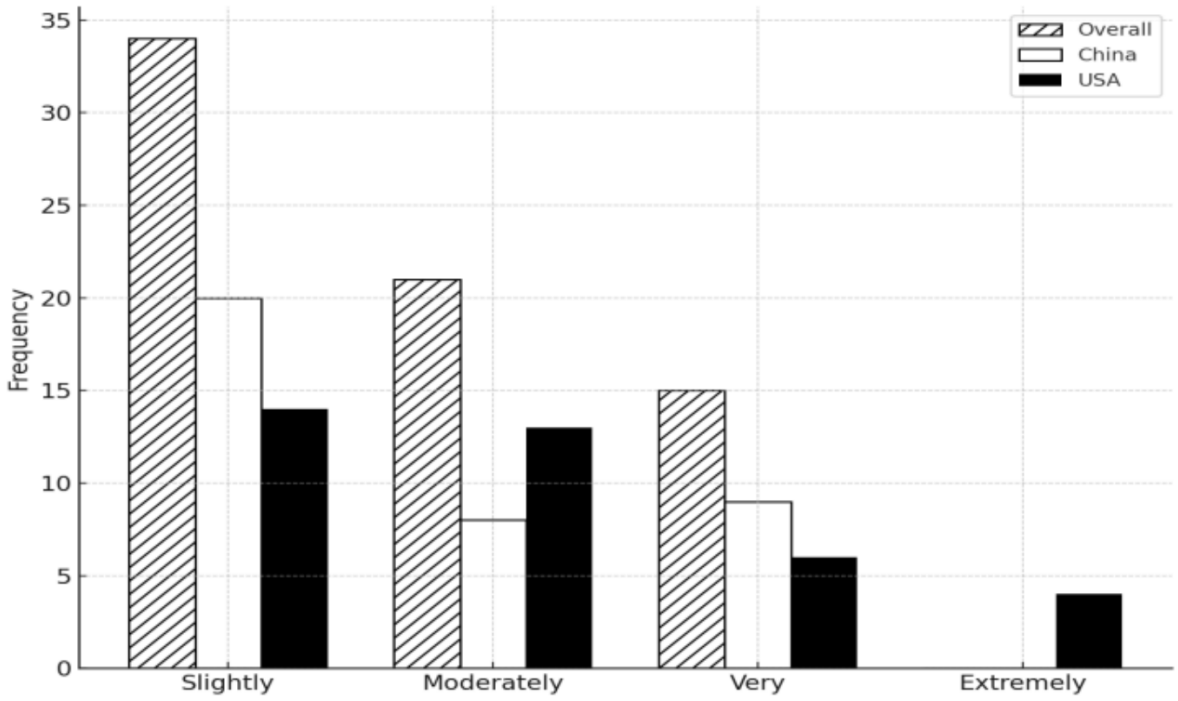


Figure 6.15: Confidence in effectively leveraging AI tools (Study 2; overall and by region).

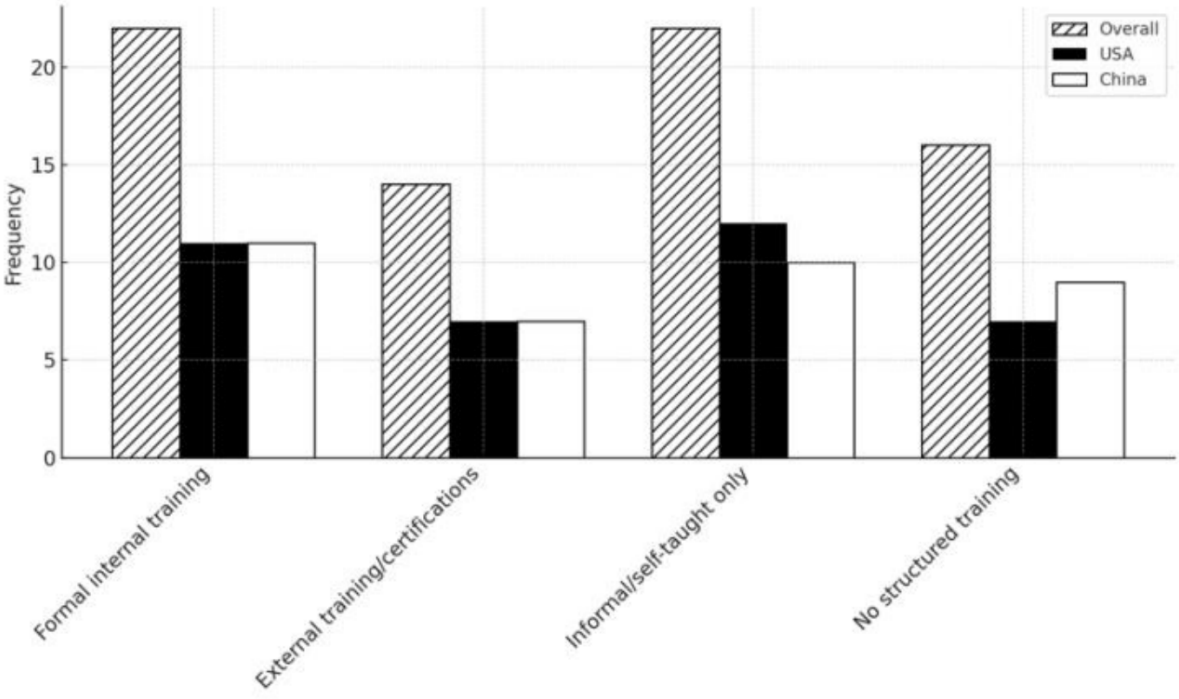


Figure 6.16: Structured AI training offered by organizations (Study 2; overall and by region).

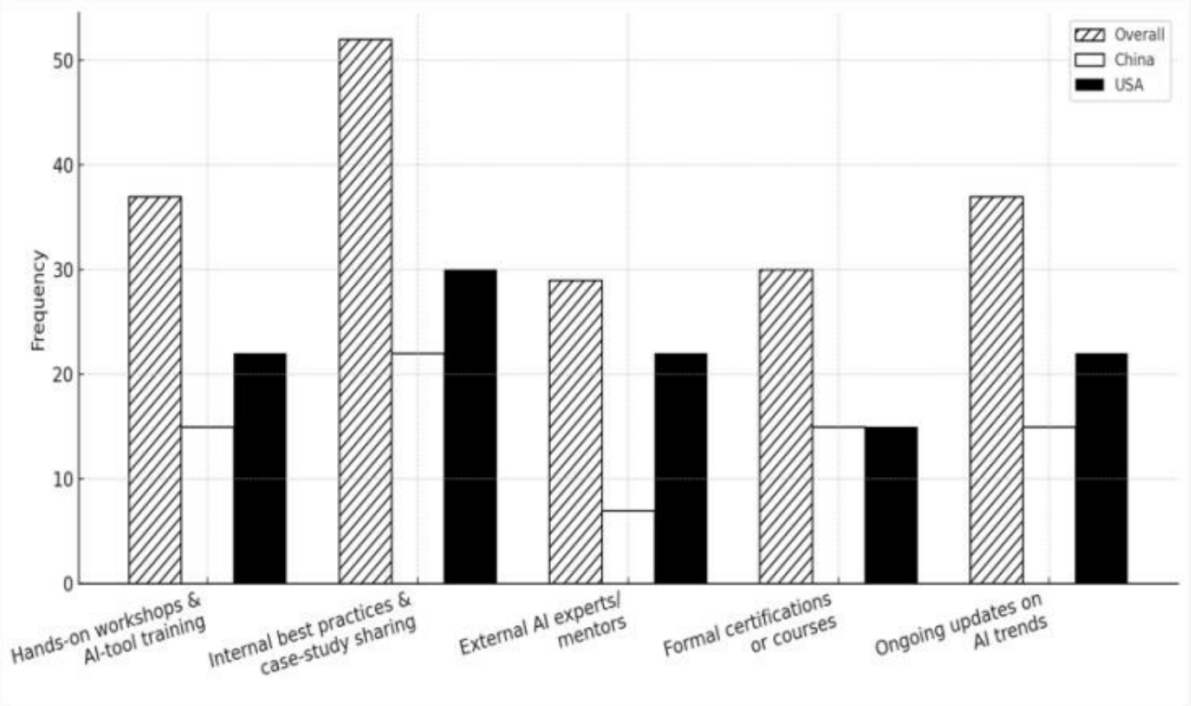


Figure 6.17: Preferred resources for AI skill development (Study 2; overall and by region).

Takeaway. These patterns characterize whether adoption is supported mainly by informal learning, formal training provision, or mixed models—and whether the USA and China cohorts report different upskilling trajectories.

Responsibility, governance, and ethics

Ethical issues are reported as occurring less than "occasionally" on average (Figure 6.13). However, governance and accountability remain unresolved for many respondents. The most frequent response for ethical responsibility indicates no clear ownership (33.8%), and the most frequent response for the existence of ethical AI guidelines indicates no clear guidelines (36.5%).

Item	**Most frequent response**	**%**	*N*
AI tools used by PMs	ChatGPT	78.4%	74
AI integration goals (top)	Reducing operational costs	52.7%	74
Tool selection criteria (top)	Cost-effectiveness	58.1%	74
Task complexity increased (top)	Human-in-the-loop checks	75.7%	74
Time reallocated to (top)	Product innovation	85.1%	74
Ethical responsibility (top)	No clear ownership	33.8%	74
Ethical guidelines (top)	No clear guidelines exist	36.5%	74

Table 6.10: Study 2 highlights for categorical items (most frequent response).

Interpretation. These modal responses suggest that adoption has outpaced institutionalization: AI tools are used widely, yet many respondents cannot identify a clear owner or rule-set governing ethical use. In artifact-mediated workflows, this ambiguity is consequential because responsibility for verifying AI-mediated evidence can become diffuse, increasing the risk that errors persist unchallenged as artifacts circulate within teams.

Ethical challenges and perceived concerns

Beyond governance structure, respondents report concrete ethical challenges and concerns associated with increased AI use. We summarize the response distributions overall and by region (full frequency tables are provided in Appendix C).

Ethical challenge	**USA**	**China**	**Overall**
Privacy and protection of customer data	11	12	23
Potential bias in AI-driven decisions	11	8	19
Accountability for AI decisions or errors	5	9	14
Transparency and explainability	7	5	12
Compliance with regulatory standards	2	1	3
No ethical challenges encountered	1	2	3

Table 6.11: Ethical challenges encountered due to AI tool use (Study 2; counts).

Greatest ethical concern	**USA**	**China**	**Overall**
Accuracy & reliability	7	10	17
Privacy & data security	9	2	11
Accountability & responsibility	7	2	9
Job displacement	4	2	6
Overuse & creativity loss	4	1	5
No concern stated	2	13	15
Other	4	7	11

Table 6.12: Greatest ethical concerns about increased AI use in product management (Study 2; counts).

Cross-national governance patternsTo contextualize the “no clear ownership” and “no clear guidelines” patterns, Tables 6.13 and 6.14 report the full response distributions by region. Two patterns are salient. First, the U. S. cohort reports substantially higher ambiguity in responsibility ownership (18 vs. 7 reporting “no clear ownership”), while the China cohort more frequently attributes responsibility to product managers individually. Second, the U. S.

cohort reports more formal internal guideline structures (18 vs. 4), while the China cohort reports higher reliance on informal guidance.

Responsibility structure	China	USA	Overall	% Overall
Formal organizational guidelines/policies	7	6	13	17.6%
Product managers individually	11	0	11	14.9%
Dedicated AI/Ethics committee	4	6	10	13.5%
Cross-functional team/collaboration	7	6	13	17.6%
No clear ownership/responsibility defined	7	18	25	33.8%

Table 6.13: Responsibility for ethical AI decisions by region (Study 2; counts and overall percentages).

Guideline structure	China	USA	Overall	% Overall
Formal internal AI ethics guidelines exist	4	18	22	29.7%
No clear guidelines exist	16	11	27	36.5%
Informal ethical guidance only	12	7	19	25.7%
External standards followed (e.g., GDPR, EU AI Act)	4	0	4	5.4%

Table 6.14: Existence of ethical AI guidelines in product management by region (Study 2).

Qualitative themes from open-ended responses

Open-ended responses provide mechanism-level context for the structured patterns reported above. Three recurring themes are salient.

Theme 1: AI as a drafting and synthesis layer. Respondents describe using AI primarily to compress information and produce first drafts (e.g., summarizing user feedback, drafting PRDs), consistent with the quantitative concentration of benefits in documentation tasks.

Theme 2: Verification as the new bottleneck. A common narrative is that time saved upstream is partially reallocated to human-in-the-loop checking, validation, and alignment work. This aligns with the high frequency of "human-in-the-loop checks" as the task that became more complex.

Theme 3: Governance ambiguity and escalation uncertainty. Many respondents report unclear responsibility ownership and missing guidelines, often framing ethics as a cross-functional problem without a clear escalation path. Where regional differences emerge, they are reported explicitly (USA vs. China) to support the comparative logic of RQ2.

Representative excerpts (anonymized) are reported in Appendix B to illustrate each theme and to document coding boundaries.

Robustness and sensitivity checks (Study 2)Key directional findings are checked for robustness using ordinal or rank-based alternatives when Likert items show ceiling/floor effects or non-normality. Regional comparisons are additionally repeated after consolidating sparse categories in categorical items. Robustness outcomes are summarized in Appendix B.

6.3. Cross-Study Integration (RQ3)

This section addresses **RQ3**: *How can changes in organizational decision-making practices enabled by AI be explained through underlying cognitive mechanisms of AI-assisted memory?* The cross-study logic treats Study 1 as mechanistic evidence about *how* external representations shape memory and metacognition, and Study 2 as field evidence about *where* those mechanisms become operational in decision-intensive work settings.

Integrated interpretation across levelsThe combined findings support three integration claims consistent with the AI Buffer Model and the multi-level framework introduced in Section 3.4.

Claim 1: "AI-first framing" is effective when it shapes encoding, not when it replaces it. Study 1 shows a strong timing advantage for pre-reading summaries on recognition outcomes, supported by process evidence that timing redistributes attention toward the summary without increasing total time. Study 2 indicates that PMs use AI to reduce routine workload and report improved decision-making. Together, these results suggest that organizational benefits are most plausible when AI is used early to frame problems and structure attention, while humans still perform critical integration and judgement.

Claim 2: Fragmentation increases misattribution risk, motivating governance mechanisms. Study 1 demonstrates that segmented presentations substantially increase false-lure endorsement (odds ratio 5.93), pointing to a source-monitoring vulnerability when AI-provided information is fragmented. Study 2 shows that accountability and ethical governance are often unclear and that guidelines are frequently absent. The integration implication is that the *provenance* and *structure* of AI outputs are part of decision governance: integrated, traceable artifacts are expected to reduce misattribution and support responsibility assignment.

Claim 3: Reliance is not purely "overconfidence"; it is a shift in cognitive infrastructure. Study 1 finds that AI does not worsen overconfidence but does change reliance and attention allocation (trust and dependence track timing). Study 2 complements this by showing that AI adoption reallocates work toward higher-level activities while introducing new responsibilities (e.g., human-in-the-loop checks). The integrated interpretation is that reliance operates through infrastructural shifts in the external representation: who curates it, when it is consulted, and how it is audited.

Cognitive mechanism (Study 1)	Empirical pattern (Study 1)	Organizational implication (Study 2)
Encoding scaffold (advance organizer)	Pre-reading yields highest AI-summary-sourced accuracy; timing is a strong driver of recognition	Field reports are consistent with higher value when AI is used early to frame options and guide subsequent work, rather than as a late-stage "answer engine"
Cue-dependent retrieval and offloading	MCQ improves while free recall is unchanged (recognition–recall dissociation)	Decision practices may become artifact-driven (easy retrieval of AI-framed options) without equivalent deep internalization; verification practices become essential
Source monitoring	Segmented structure increases false-lure endorsement (OR 5.93)	Fragmented AI usage across tools/channels can increase misattribution and accountability diffusion; integrated documentation and provenance controls are plausible risk mitigations
Confidence calibration	AI does not increase overconfidence relative to No-AI	Self-reported confidence is an unreliable governance signal; explicit checks (e.g., references, validation, sign-off) remain necessary
Load allocation and reliance	Pre-reading increases summary engagement without increasing total time; reliance measures track timing	Adoption shifts time from operational to strategic work, but also shifts responsibility to monitoring, auditing, and escalation of AI outputs

Table 6.15: Cross-study integration linking cognitive mechanisms (Study 1) to product-management practices (Study 2).

6.4. Results chapter summary

Across Study 1 and Study 2, three results are central.

RQ1 (mechanisms). AI summaries improve cue-supported recognition (especially when ac-

cessed pre-reading), while free recall remains largely unchanged. Fragmented (segmented) summaries increase false-lure endorsement, indicating higher misattribution risk under weaker source-monitoring cues.

RQ2 (field patterns). Product managers report broad AI adoption, with strongest perceived workload reduction in routine and documentation tasks, and positive perceived impact on productivity and decision quality. Governance remains a salient gap, with frequent reports of unclear responsibility and missing guidelines.

RQ3 (integration). Field-level shifts (automation + new monitoring responsibility) align with mechanism-level signatures: early AI-framing supports encoding and decision structuring, while fragmented AI artifacts elevate epistemic risk and motivate stronger provenance and accountability practices.

7 | Discussion

7.1. Discussion Overview

This discussion interprets the findings through the thesis' central claim: AI systems increasingly function as *external representations* that reshape cognition and work not only by "adding information," but by changing the timing, structure, and provenance of the cues people use to encode, retrieve, and justify decisions. Across the two studies, the same representational parameters that improve short-term performance can also introduce new epistemic and governance risks when provenance cues are weakened or responsibilities are diffuse.

A core theme is the distinction between *performance with support* versus *internalization without support*. Study 1 shows that AI summaries can improve cue-supported recognition, especially when presented early enough to act as an encoding scaffold, yet this does not automatically translate into better generative recall. This pattern is consistent with the AI Buffer Model's prediction that external representations primarily strengthen cue availability and retrieval support rather than building durable, self-generated traces.

Study 2 extends this cognitive account to organizational settings. Product managers report that AI improves productivity and perceived decision-making, particularly for routine drafting and synthesis. At the same time, the field patterns show that many organizations have not institutionalized ownership and guidelines for AI-mediated work. Taken together, the studies support a multi-level interpretation: as AI artifacts become shared memory objects, the quality of provenance cues and the clarity of accountability structures become central determinants of whether AI assistance yields robust organizational value or scalable epistemic risk.

The remainder of this chapter proceeds in three steps. First, it answers the research questions by interpreting the key results in mechanism terms. Second, it articulates theoretical and practical implications as designable parameters of AI-assisted cognition and work. Third, it clarifies boundary conditions, limitations, and future work needed to validate the proposed multi-level account.

7.2. Answering the Research Questions

Key quantitative takeaways.

- **Study 1 (recognition):** overall MCQ accuracy improves with AI (0.598 vs 0.510; $\Delta = 0.088$; $F(1, 34) = 7.86$, $p = .008$, $\eta_p^2 = 0.188$).
- **Study 1 (timing):** for AI-summary-sourced items, pre-reading outperforms synchronous and post-reading (0.833 vs 0.568 vs 0.641; $F(2, 44) = 14.00$, $p < .001$, $\eta_p^2 = 0.389$).
- **Study 1 (epistemic risk / source monitoring):** segmented summaries increase false-lure endorsement (OR 5.93, 95% CI $[1.63, 21.5]$, $p = .007$).
- **Study 2 (field patterns):** productivity impact $M = 3.47$ and decision improvement $M = 3.45$ (USA vs CN: $p = .0104$), alongside governance ambiguity (no clear ownership 33.8%; no clear guidelines 36.5%).

Interpretive takeaways (what the numbers imply).

- **AI helps most when it sets the frame early:** the pre-reading advantage suggests that summaries work as scaffolds that guide later attention and encoding, not merely as extra content added at the end.
- **Not all "better performance" is deeper memory:** the recognition–recall dissociation implies that AI assistance can raise accuracy in cue-rich tests without strengthening generative retrieval.
- **Format is a safety parameter:** segmented representations can raise misattribution/false-lure risks, indicating that interface structure affects source monitoring and epistemic reliability.
- **Organizational adoption shifts the unit of memory:** in the field, AI use concentrates value in artifacts (drafts, briefs, summaries), which increases the importance of governance (ownership, guidelines, verification norms) because artifacts become the decision trace.

RQ1: How do AI-generated external representations influence core human memory processes?
Study 1 provides converging evidence that AI-generated summaries function as *external representations* that selectively reshape memory performance depending on *when* and *how* they are presented (Section 6.1). Three implications are central for RQ1.
First, AI assistance improves cue-supported recognition, but does not strengthen generative retrieval. The AI condition outperforms No-AI on MCQ accuracy (Table 6.1), yet free recall remains largely unchanged across timing conditions and does not meaningfully differ between AI and No-AI (Section 6.1). This pattern is consistent with the AI Buffer Model's prediction that external representations primarily support *cue-dependent retrieval* and may not translate into stronger internally generated retrieval traces (Section 3.2). In other words, AI changes the *availability* of retrieval cues more than the *durability* of self-generated memory. This recognition

advantage is reflected in the MCQ difference summarized above (AI: 0.598 vs No-AI: 0.510).
Second, timing shapes encoding by changing attention allocation, not time-on-task. Pre-reading access yields the highest recognition performance on summary-covered information (Section 6.1 and fig. 6.1), and process measures indicate that pre-reading increases summary engagement without increasing total time (Section 6.1 and table 6.3). This is visible in the timing pattern summarized above (pre-reading highest; synchronous lowest). Summary time is highest in pre-reading (mean 132.5 seconds; 24.9% of total) and lowest post-reading (mean 69.5 seconds; 13.7%), while total time-on-task remains approximately 531–536 seconds. Together, these findings support an *encoding scaffold* interpretation: the external representation is most effective when encountered early enough to structure subsequent processing, consistent with advance organizer and cognitive load accounts [5, 83].
Third, structure shapes epistemic risk through source monitoring. Integrated summaries reduce false-lure endorsement relative to segmented summaries (Section 6.1 and fig. 6.4). In a binomial GLMM, segmented structure increases false-lure endorsement (OR 5.93, 95% CI $[1.63, 21.5]$, $p = .007$). This supports the claim that the structure of an external representation can either preserve or degrade cues that enable source attribution and cross-checking [16, 37]. Importantly, these risks are not well explained by a simple "overconfidence" mechanism: calibration is imperfect overall, but AI does not increase overconfidence (Section 6.1 and fig. 6.5).
Alternative explanations and boundary conditions for RQ1. Two alternative interpretations are worth addressing. First, the recognition gain could reflect a "test alignment" effect: MCQs reward cue availability and partial familiarity, so improvements may overestimate deeper understanding. The lack of free-recall gains supports this interpretation, suggesting that AI assistance changes retrieval conditions more than the underlying memory trace. Second, timing effects could partially reflect how summaries steer perceived relevance: pre-reading may act as a relevance filter that increases selective attention during reading, whereas post-reading may be treated as a redundant recap.
These boundary conditions imply that AI assistance may be most beneficial when (i) users must later recognize or verify information under cue-rich conditions, and less beneficial when (ii) tasks require generative explanation, transfer, or reconstruction without external cues. They also suggest that the quality and faithfulness of summaries should matter: if summaries omit key causal relations or introduce subtle distortions, early exposure could scaffold attention toward a misleading frame, making the "scaffold" mechanism double-edged in high-stakes contexts.
Design implications for AI-assisted memory. Viewed mechanistically, the results identify two design levers. The first is *sequencing*: enabling a brief, high-level preview before deep work to structure encoding. The second is *provenance-preserving integration*: integrated representations that keep source cues and reduce split attention should be treated as a safety feature, because they support cross-checking and source attribution. This reframes interface choices (segmented

vs integrated presentations) as epistemic risk controls, not merely usability preferences.

Overall, RQ1 is answered by a mechanism-specific conclusion: AI-generated representations can improve recognition outcomes when they function as early scaffolds, but they can also increase misattribution risks when fragmented, and they do not automatically improve generative recall.

RQ2: How does AI adoption reshape decision-making practices, roles, and responsibility in product management?

Study 2 suggests that AI adoption in product management is best understood as a dual shift: a *task shift* toward automation of operational work and a *governance shift* toward monitoring, accountability, and escalation (Section 6.2).

Task shift and role reconfiguration. Respondents report the strongest workload reductions for routine work and document drafting/content creation (Figure 6.10). In parallel, respondents report positive productivity and decision-making impacts (Figure 6.13). This aligns with the thesis positioning of AI as an external representational layer that accelerates synthesis and documentation, enabling more time for prioritization and strategic work (Section 2.10). Routine administrative tasks ($M = 3.88$, $SD = 1.11$) and documentation/content creation ($M = 3.70$, $SD = 0.93$) are significantly above the midpoint (both $p < .001$), while routine communications show the weakest reduction ($M = 2.22$, $SD = 0.85$, $p < .001$ below midpoint) and market research/competitor analysis is near-neutral ($M = 2.97$, $SD = 0.73$, $p = .625$) (Table 6.7).

Decision support with cross-context variation. Perceived decision-making improvement is positive overall and differs across contexts, with Chinese respondents reporting higher perceived gains than U. S. respondents (Section 6.2). This suggests that the organizational value of AI may depend on local adoption maturity, tool ecosystems, and governance constraints, even when overall usage patterns converge. The USA–China difference is significant for perceived decision-making improvement ($p = .0104$), while productivity impact shows no regional difference ($p = 0.999$) and strategic time gained also does not differ ($p = 0.751$) (Table 6.9).

Responsibility and governance are frequently unresolved. Despite widespread tool usage (e.g., ChatGPT as the most frequently reported tool), survey highlights show that many respondents report unclear ethical ownership and missing guidelines (Table 6.10). In practical terms, adoption changes not only *what* PMs do, but *who* is accountable for failures and for the verification of AI-mediated evidence—a core organizational dimension of assisted decision-making. The modal responses are “no clear ownership” (33.8%) and “no clear guidelines exist” (36.5%). By region, “no clear ownership” is higher in the USA (18) than China (7), while formal internal guidelines are more often reported in the USA (18) than China (4) (Section 6.2). Regional differences are moderate in magnitude (responsibility: Cramérś $V = 0.48$; guidelines: $V = 0.46$).

Interpreting the regional difference (USA vs China) without overclaiming. The higher perceived decision-improvement in China relative to the USA can be interpreted in at least two plausible ways that do not require assuming objective performance differences. One possibility is *baseline*

opportunity: if teams differ in prior tooling maturity or documentation routines, the same AI capability could feel more transformative where baseline support is lower. Another possibility is *workflow embedding*: decision support may be perceived as higher when AI is integrated into standardized routines (templates, review flows, mandated briefs), whereas ad-hoc use yields uneven benefits even if productivity gains remain similar. Because Study 2 is cross-sectional and perception-based, these interpretations should be treated as hypotheses for future multi-site validation rather than causal claims.

Failure modes: when task automation becomes decision fragility. The pattern that AI reduces routine drafting and increases perceived decision support also implies a risk: when teams increasingly rely on AI-generated artifacts as the substrate for decision conversations, errors and omissions can propagate if verification norms are weak. In practical terms, the organizational question becomes less "Should we use AI?" and more "What routines ensure that AI-mediated evidence is checked, documented, and attributable?" This is precisely where the reported lack of ownership and guidelines becomes consequential: missing governance structures turn local cognitive reliance into system-level accountability gaps.

In sum, RQ2 is answered by a practice-level conclusion: AI adoption is associated with operational efficiency and perceived decision support, but also with a redistribution of responsibility toward evaluation and governance activities that many organizations have not yet institutionalized.

RQ3: How can organizational changes be explained through cognitive mechanisms of AI-assisted memory?

RQ3 is addressed by integrating the mechanisms observed in Study 1 with the adoption patterns observed in Study 2 (Section 6.3 and table 6.15). Three bridges are particularly salient.

Bridge 1: Early framing translates into organizational value. The pre-reading advantage in Study 1 indicates that external representations deliver the highest benefits when they shape encoding and attention allocation. In product management, this mapping is consistent with workflows in which AI is used early to frame decision spaces (e.g., synthesizing signals into initial hypotheses), after which humans perform integration and judgement. The implication is not that AI should "replace" decision-making, but that its organizational value is highest when it functions as cognitive infrastructure for problem framing.

Bridge 2: Artifact-driven work emerges from cue-dependent retrieval. Study 1 shows a recognition–recall dissociation, consistent with AI supporting cue-based retrieval more than deep internalization. Study 2 respondents report that AI reduces documentation and routine workload and improves perceived decision support. Together, these findings imply that organizations may increasingly rely on AI-produced artifacts as shared memory objects. This strengthens the case for systematic documentation practices (what was generated, why it was trusted, and how it was validated), because the artifact itself becomes part of the decision trace. Mechanistically,

this implies artifact-mediated work: performance improves when AI-produced cues are present, without necessarily increasing internalization when they are absent.

Bridge 3: Fragmentation produces misattribution risk and accountability diffusion. Structure-driven false memory effects in Study 1 show that fragmented AI representations increase misattribution risk. Study 2 respondents report that accountability is often unclear and governance policies are frequently missing. The integrated interpretation is that fragmentation at the cognitive level (weak source cues) can scale into fragmentation at the organizational level (diffuse ownership). This links the AI Buffer Model's source-monitoring pathway to governance requirements for traceability and responsibility assignment [67]. In combination, Study 1's structure effect (segmented outputs → higher misattribution; OR 5.93) and Study 2's governance ambiguity (no clear ownership 33.8%, no guidelines 36.5%) support the idea that representational fragmentation is a scalable risk—from cognitive provenance failures to organizational accountability diffusion.

Multi-level account: from representational parameters to organizational routines

The cross-study integration can be stated as a multi-level chain: *representational parameters* (timing, structure, provenance cues) shape *individual memory mechanisms* (scaffolding, cue-dependent retrieval, source monitoring), which in turn shape *work practices* (artifact-centric coordination, reduced drafting time), and ultimately shape *governance needs* (ownership, verification standards, auditability). This chain yields testable propositions that can guide future validation:

1. **Sequencing proposition:** workflows that use AI summaries as *pre-reading frames* will show stronger downstream decision quality than workflows that introduce AI primarily post hoc, mediated by improved attention allocation and more coherent problem framing.
2. **Artifact dependence proposition:** as AI use increases, teams will rely more on AI-produced artifacts as shared memory objects; decision trace quality will depend on whether artifacts include provenance and explicit verification notes.
3. **Provenance proposition:** interface structures that preserve source cues (integrated, traceable outputs) will reduce misattribution and rework relative to fragmented, channel-scattered outputs.
4. **Governance amplification proposition:** the organizational cost of weak source monitoring increases with artifact centrality: the more a team coordinates via AI artifacts, the more misattribution risk scales into accountability diffusion.

This strengthens the thesis' explanatory claim by making the integration falsifiable: future studies can manipulate workflow sequencing and representation structure in field settings while measuring both cognitive outcomes (source accuracy, recall) and organizational outcomes (decision audit quality, rework, escalation rates).

Taken together, the thesis answers RQ3 by showing how micro-level cognitive mechanisms

(scaffolding, cue dependence, source monitoring, and reliance) provide a coherent explanation for macro-level changes in decision practices and accountability boundaries.

7.3. Theoretical Implications

The thesis contributes to theory at three levels: cognition, human–AI interaction, and organizational decision-making.

External representations as designable cognitive infrastructure. Study 1 shows that timing and structure are not superficial interface features: they are parameters that modulate memory outcomes. The pre-reading advantage is consistent with the idea that an AI-generated representation can act as an advance organizer that structures encoding [5], while the segmented-structure risk connects to split-attention and source-monitoring accounts [16, 37]. This supports a design-oriented view of AI assistance: the cognitive consequences of "using AI" depend on how the representation is integrated into the task pipeline.

AI-assisted memory is cue-dependent and may not generalize to durable internalization. The recognition–recall dissociation refines cognitive offloading arguments by distinguishing between (i) enhanced performance under cue-rich retrieval conditions and (ii) durable generative retrieval without cues. This aligns with the broader perspective that external memory aids can change what is encoded and how it is retrieved, rather than uniformly strengthening memory [71, 80, 88].

From individual reliance to organizational responsibility. Study 2 extends the cognitive story into organizational contexts by showing that AI adoption reorganizes work toward artifact-mediated decision processes while simultaneously creating governance gaps. The cross-study integration provides theoretical support for the thesis claim that reliance is not reducible to subjective trust alone; it is a structural shift in how cognitive work is distributed across people, tools, and documentation (Section 3.4).

7.4. Practical Implications

The results suggest actionable implications for AI tool design, for product managers, and for organizations. To avoid duplicating the cross-study synthesis elsewhere, we consolidate them here into a single operational framework: four design principles, a minimal workflow template, and evaluation metrics beyond productivity.

Design principles (from mechanisms to deployment)

1. **Principle 1: Sequence for scaffolding (timing).** If the goal is better comprehension and reliable downstream decisions, AI representations should appear early enough to structure attention and encoding. This follows directly from the Study 1 timing advan-

tage (pre-reading as an encoding scaffold) and suggests providing pre-reading outlines, hypotheses, or "signal maps" *before* deep work begins. Post hoc summaries remain useful as documentation, but they are less likely to shape encoding and can encourage superficial confirmation.

2. **Principle 2: Integrate to protect provenance (structure).** Representations should preserve source cues and reduce split attention. Integrated views (e.g., a single consolidated brief with clear attributions) should be preferred over segmented outputs scattered across tools and channels. This principle operationalizes Study 1's source-monitoring risk: fragmented structure weakens provenance cues and increases misattribution/false-lure endorsement.
3. **Principle 3: Make verification explicit (process).** Because AI can improve cue-supported performance without guaranteeing durable internalization (recognition–recall dissociation in Study 1), validation must be treated as a separate step rather than an implicit expectation. Tools and routines should prompt explicit checks for high-stakes claims (source links, counter-evidence, uncertainty notes) rather than relying on users to "remember to verify".
4. **Principle 4: Institutionalize accountability for AI-mediated evidence (governance).** As adoption shifts work toward artifact-centric coordination (Study 2), organizations must specify ownership for evidence quality: who defines verification standards, who escalates incidents, and who maintains the decision trace. Without these roles, cognitive provenance failures can scale into organizational accountability diffusion, consistent with the governance ambiguity reported in Study 2.

A practical workflow template for product teams

A minimal workflow consistent with the thesis findings has four stages:

1. **Frame (pre-reading):** generate an outline/hypothesis map to guide attention.
2. **Ground:** consult primary sources and annotate which claims are supported.
3. **Consolidate (integrated artifact):** produce one brief that includes attributions and open questions.
4. **Decide and log:** record the decision with a short justification and verification notes.

This template preserves the productivity advantages reported in Study 2 while reducing the misattribution and provenance risks indicated by Study 1.

Evaluation metrics beyond productivity

To assess whether AI adoption is improving decisions rather than only accelerating output, organizations can track:

- **Decision trace completeness:** presence of sources, assumptions, alternatives, and open questions;
- **Rework rates:** revisions due to incorrect or missing evidence in AI-mediated artifacts;

- **Escalation frequency and resolution time:** how often uncertain claims trigger review and how quickly they are resolved;
- **Post-mortem outcomes:** whether failures are traceable to provenance gaps, verification omissions, or accountability breakdowns.

These measures align evaluation with the cognitive risks (cue-dependence, source-monitoring failures) and the organizational risks (governance ambiguity) identified in this thesis.

7.5. Competing Explanations and Boundary Conditions

Several factors could moderate the observed effects and should be treated as boundary conditions rather than nuisances. First, **summary quality and faithfulness** likely determine whether early scaffolding helps or misleads: a high-level preview is beneficial when it is accurate and well-calibrated, but it could bias attention toward incorrect frames when hallucinations or omissions occur. Second, **expertise** may flip the timing effect: novices may benefit more from pre-reading scaffolds, whereas experts may prefer later summaries that compress and confirm already-formed mental models.

Third, **task stakes and verification norms** should moderate both memory and governance outcomes. In low-stakes tasks, cue-dependent retrieval may be "good enough" and the cost of misattribution limited. In high-stakes tasks, the same misattribution risk can become unacceptable, making provenance cues and accountability assignments central. Finally, **organizational maturity** likely determines whether AI adoption produces robust value: when teams have templates, review routines, and audit trails, AI artifacts can accelerate work while remaining inspectable; when such routines are absent, artifacts can become persuasive but weakly grounded substitutes for evidence.

7.6. Limitations

Several limitations constrain the interpretation and generalization of the findings.

Study 1 limitations. The experiment uses a controlled task environment with a limited sample and a finite set of stimuli. The AI assistance is operationalized as summaries rather than fully interactive dialogue, and the outcomes focus on short-term performance. These choices strengthen internal validity but limit conclusions about long-term retention, expert populations, and interactive AI workflows. A further concern is that the controlled setting may *underestimate* reliance dynamics that emerge with repeated use in real work, where AI outputs become habitual and retrieval increasingly depends on external cues. Conversely, the short retention interval may *overestimate* benefits for recognition if gains decay over time or if users do not revisit the same representational scaffold.

Study 2 limitations. The field study relies on self-reported perceptions in a cross-sectional survey. Causal claims about productivity or decision quality cannot be inferred, and results may reflect selection bias (e.g., respondents who adopt AI more actively may be more likely to participate) and measurement limitations in cross-cultural survey design. Because the measures capture perceived impacts, estimates may reflect optimism, novelty, or organizational narratives about AI. At the same time, self-reports may *under-detect* governance failures, since accountability breakdowns often become visible only during incidents, audits, or escalations.
Cross-study limitations. The multi-level integration is explanatory rather than a statistical mediation test: the studies involve different tasks and populations, and cognitive variables are not directly measured in the field context. The integration therefore provides a coherent account of plausible mechanisms, but it should be treated as a framework for future multi-level validation.

7.7. Future Research Directions

The thesis motivates several directions for future work.

- **Longitudinal cognitive outcomes:** test whether timing and structure effects persist over longer retention intervals and in repeated-use settings where offloading may accumulate.
- **Interactive AI and provenance interventions:** evaluate whether inline citations, uncertainty cues, and structured verification prompts reduce false-lure endorsement without reducing productivity.
- **Decision-quality measures in the field:** complement self-reports with behavioral logs, decision audit outcomes, and objective performance metrics in product teams.
- **Multi-level causal designs:** run field experiments that manipulate AI representation design in real workflows (e.g., pre-reading vs post-reading briefs; integrated vs fragmented outputs) while measuring both cognitive and organizational outcomes.
- **Boundary conditions:** examine moderators such as domain expertise, time pressure, governance maturity, and regulatory context to specify when AI functions as beneficial cognitive infrastructure versus a source of epistemic risk.

Taken together, the discussion supports a single integrated implication: early scaffolding improves cue-supported performance, fragmented representations increase misattribution risk, and organizational adoption concentrates work around AI artifacts while leaving governance gaps unresolved. The design and deployment of AI assistance therefore need to co-evolve: representation choices that preserve provenance cues (timing, integration, traceability) must be matched by organizational routines that make verification and ownership explicit.

The framework above summarizes how representational design and organizational routines must

co-evolve for AI-assisted work to remain both efficient and accountable.

8 | Conclusion

This thesis examined how AI-generated outputs function as *external representations* that reshape both individual cognition and organizational decision-making. Across a controlled experiment (Study 1) and a field survey of product managers (Study 2), the results support a coherent account: AI assistance tends to improve performance when it provides useful retrieval cues and early framing, but it can also introduce scalable epistemic and governance risks when provenance cues are weak or when responsibility for verification is unclear.

8.1. Summary of Answers to the Research Questions

RQ1: How do AI-generated external representations influence core human memory processes?Study 1 shows that AI-generated summaries primarily support *cue-dependent* performance. Recognition improves with AI assistance, particularly when summaries are presented *before* reading and can function as encoding scaffolds. However, this benefit does not automatically translate into stronger generative recall. In addition, representation *structure* matters: segmented formats increase false-lure endorsement, consistent with a source-monitoring pathway in which fragmented outputs degrade provenance cues and cross-checking opportunities.
RQ2: How does AI adoption reshape decision-making practices, roles, and responsibility in product management?Study 2 suggests that adoption is associated with a *task shift* (reduced workload for drafting and routine documentation) and a *governance shift* (greater need for monitoring, accountability, and escalation). PMs report positive impacts on productivity and perceived decision support, with contextual variation across regions. At the same time, the field data indicate substantial governance ambiguity, including unclear ownership and missing guidelines in many organizations.
RQ3: How can organizational changes be explained through cognitive mechanisms of AI-assisted memory?Integrating the studies supports a multi-level explanation: representational parameters (timing, structure, and provenance cues) shape memory mechanisms (scaffolding, cue-dependence, and source monitoring), which in turn shape organizational practices (artifact-centric coordination and reliance on AI outputs as shared memory objects). As AI artifacts

become part of the decision trace, cognitive provenance failures can scale into organizational accountability diffusion. The thesis thus links micro-level memory mechanisms to macro-level governance needs.

8.2. Contributions

The thesis contributes at four levels.

Conceptual contribution. It articulates a design-oriented view of AI assistance: the cognitive and organizational consequences of "using AI" depend on *how* external representations are introduced (timing), *how* they are presented (structure), and *what provenance cues* they preserve. This strengthens the claim that AI-assisted work is not merely an efficiency gain but a restructuring of cognitive infrastructure.

Empirical contribution. Study 1 provides causal evidence that pre-reading summaries improve recognition for summary-covered information and that segmented structure increases false-lure endorsement. Study 2 documents convergent adoption patterns in product management alongside common governance gaps, highlighting that perceived benefits coexist with unresolved responsibility structures.

Methodological contribution. By combining controlled cognitive measurement with a field study, the thesis demonstrates a practical approach to studying human–AI interaction across levels of analysis: experimental manipulation to identify mechanisms and survey-based evidence to identify where those mechanisms may matter in real decision work.

Practical contribution. The findings translate into actionable guidance for designing AI tools and deploying them in product organizations: support early framing, preserve provenance cues, and institutionalize accountability for AI-mediated evidence.

8.3. Actionable Takeaways

The results suggest a small set of robust, implementation-oriented takeaways:

1. **Use AI early for framing, not late for justification:** summaries function best as pre-reading scaffolds that structure subsequent attention.
2. **Prefer integrated representations over fragmented ones:** fragmentation should be treated as an epistemic risk factor because it weakens source monitoring cues.
3. **Treat AI artifacts as part of the decision record:** document what was generated, what was verified, and why it was trusted.
4. **Separate drafting from validation:** make verification steps explicit for high-stakes claims rather than implicit in individual judgment.
5. **Assign ownership:** define who is accountable for checking AI-mediated evidence and

how escalation works when uncertainty is high.

6. **Measure success beyond speed:** evaluate AI adoption using decision trace quality (auditability, provenance, rework) as well as productivity.

8.4. Closing Statement

Taken together, the thesis argues that the central question is not whether AI "helps" or "harms" memory and decision-making in the abstract. Instead, the outcomes depend on designable parameters of external representations and on the organizational routines that govern their use. When AI systems are deployed as cognitive infrastructure that supports early framing and preserves provenance, they can improve cue-supported performance and accelerate knowledge work. When outputs are fragmented and accountability is unclear, the same systems can amplify misattribution and diffuse responsibility. Future research and practice should therefore treat provenance and governance as first-class design requirements for AI-assisted work.

Bibliography

[1] F. Abrar, S. Khan, A. Rauf, and S. Fatima. Cognitive development in the age of AI: How AI tools influence problem solving and creativity in psychological terms. *Journal of Cognitive Science and AI Research*, 12(2):45–62, 2025.

[2] F. Alshaikh and N. Hewahi. AI and machine learning techniques in the development of intelligent tutoring systems: A review. In *Proceedings of the 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, pages 273–278. IEEE, 2021.

[3] R. C. Atkinson and R. M. Shiffrin. Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence, editors, *The Psychology of Learning and Motivation*, volume 2, pages 89–195. Academic Press, 1968.

[4] Y. Atsmon. Artificial intelligence in strategy, 2023. URL `https://www.mckinsey.com`. Accessed 2025-06-04.

[5] D. P. Ausubel. The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology*, 51(5):267–272, 1960.

[6] D. P. Ausubel. *Educational Psychology: A Cognitive View*. Holt, Rinehart & Winston, 1968.

[7] A. D. Baddeley. Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63:1–29, 2012. doi: 10.1146/annurev-psych-120710-100422.

[8] L. Bai, X. Liu, and J. Su. ChatGPT: The cognitive effects on learning and memory. *Brain-X*, 4(1):12–25, 2023. doi: 10.1016/j.brax.2023.100079.

[9] K. Bailey. How AI transforms scenario analysis in corporate finance, 2025. URL `https://corporatefinanceinstitute.com`. Accessed 2025-06-04.

[10] F. C. Bartlett. *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press, Cambridge, UK, 1932.

[11] H. Beyaria, T. N. Hashem, and O. Alrusaini. Neuromarketing: Understanding the effect of emotion and memory on consumer behavior by mediating the role of artificial intelligence and customers' digital experience. *Journal of Project Management*, 9:323–336, 2024. doi: 10.5267/j.jpm.2024.9.001.

[12] C. J. Brainerd and V. F. Reyna. *The Science of False Memory*. Oxford University Press,

New York, NY, 2005.

[13] M. Cagan. *Inspired: How to Create Tech Products Customers Love*. Wiley, Hoboken, NJ, 2nd edition, 2017.

[14] J. B. Caplan, M. G. Glaholt, and A. R. McIntosh. EEG activity underlying successful study of associative and order information. *Journal of Cognitive Neuroscience*, 21(6):1050–1062, 2009. doi: 10.1162/jocn.2009.21090.

[15] S. Chan, P. Pataranutaporn, A. Suri, W. Zulfikar, P. Maes, and E. F. Loftus. Conversational AI powered by large language models amplifies false memories in witness interviews. arXiv preprint arXiv:2408.04681, 2024.

[16] P. Chandler and J. Sweller. The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, 62(2):233–246, 1992.

[17] D. Chlouverakis and K. Karapanos. How artificial intelligence is reshaping the financial services industry. Technical report, 2024.

[18] A. Clark and D. J. Chalmers. The extended mind. *Analysis*, 58(1):7–19, 1998. doi: 10.1093/analys/58.1.7.

[19] J. Corbo and O. Fountaine. It's time for businesses to chart a course for reinforcement learning, 2021. URL `https://www.mckinsey.com`. Accessed 2025-06-04.

[20] F. I. M. Craik and R. S. Lockhart. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6):671–684, 1972. doi: 10.1016/S0022-5371(72)80001-X.

[21] Y. K. Dwivedi and V. Dutot. Evolution of artificial intelligence research in Technological Forecasting and Social Change: Research topics, trends, and future directions. *Technological Forecasting and Social Change*, 2023.

[22] A. D. Echegu. *Artificial Intelligence (AI) in Customer Service: Revolutionising Support and Engagement*. Kiu Publication Extension, 2024.

[23] A. Edlich and F. Iqbal. How bots, algorithms, and artificial intelligence are reshaping the future of corporate support functions, 2018. URL `https://www.mckinsey.com`. Accessed 2025-06-04.

[24] R. A. El Haddad, Z. Wang, Y. Shin, R. Liu, Y. Wang, and C. Yu. AR secretary agent: Real-time memory augmentation via LLM-powered augmented reality glasses, 2025.

[25] S. Freeman, S. L. Eddy, M. McDonough, and et al. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23):8410–8415, 2014.

[26] A. Freire. Reshaping business education with hands-on artificial intelligence. In *Business School Internationalisation in a Changing World*. Routledge, 2024.

[27] M. Gerlich. AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1):6, 2025. doi: 10.3390/soc15010006.

[28] C. Gnanasambandam and M. Hall. How generative AI could accelerate software product time to market, 2024. URL `https://www.mckinsey.com`. Accessed 2025-06-04.

[29] C. Gong and Y. Yang. Google effects on memory: A meta-analytical review of the media effects of intensive internet search behavior. *Frontiers in Public Health*, 12:1332030, 2024. doi: 10.3389/fpubh.2024.1332030.

[30] M. Goodwin and C. Quintero-Valencia. What is artificial intelligence (AI) in business, 2024. URL `https://www.ibm.com/topics/artificial-intelligence`. Accessed 2025-06-04.

[31] A. Gupta. The top use cases of AI for product managers, 2024. URL `https://www.productgrowth.io`. Accessed 2025-06-04.

[32] Z. Haider, A. Zummer, A. Waheed, and M. Abuzar. Study how AI can be used to enhance cognitive functions, such as memory or problem-solving, and the psychological effects of these enhancements. *Bulletin of Business and Economics*, 13(3):256–263, 2024. doi: 10.61506/01.00485.

[33] J. Hartley and I. K. Davies. Preinstructional strategies: The role of pretests, behavioral objectives, overviews and advance organizers. *Review of Educational Research*, 46(2):239–265, 1976.

[34] P. Huryn. How to automate your work as a product manager, 2024. URL `https://www.productcompass.com`. Accessed 2025-06-04.

[35] E. Hutchins. *Cognition in the Wild*. MIT Press, Cambridge, MA, 1995.

[36] M. Iansiti and K. R. Lakhani. *Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World*. Harvard Business Review Press, Boston, MA, 2020.

[37] M. K. Johnson, S. Hashtroudi, and D. S. Lindsay. Source monitoring. *Psychological Bulletin*, 114(1):3–28, 1993.

[38] P. K. Jorzik. AI-driven business model innovation: A systematic review and research agenda. *Technological Forecasting and Social Change*, 2024.

[39] S. Kaggwa and O. Fonsah. AI in decision making: Transforming business strategies. *International Journal of Research and Scientific Innovation*, 2024.

[40] S. Kaggwa and T. Fonsah. AI in decision making: Transforming business strategies. *International Journal of Research and Scientific Innovation*, 2024.

[41] D. Kahneman. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York, 2011.

[42] M. Kamariotou and F. Kitsios. Artificial intelligence and business strategy towards digital transformation: A research agenda. *Sustainability*, 2021.

[43] A. Kaplan and M. Haenlein. Siri, siri, in my hand: Who's the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 2018.

[44] A. Kaplan and M. Haenlein. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1):15–25, 2019. doi: 10.1016/j.bushor.2018.08.004.

[45] M. T. Kucewicz, G. A. Worrell, and N. Axmacher. Direct electrical brain stimulation of human memory: Lessons learnt and future perspectives. *Brain*, 146(5):1681–1695, 2023. doi: 10.1093/brain/awad020.

[46] F. Li and Y. Zhao. Transforming organisations through AI: Emerging strategies for navigating the future of business. *Journal of Financial Transformation*, 2025.

[47] D. S. Lindsay and M. K. Johnson. The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, 17(3):349–358, 1989.

[48] J. Liu, H. Zhang, T. Yu, and G. Xue. Transformative neural representations support long-term episodic memory. *Science Advances*, 7(8):eabc5546, 2021. doi: 10.1126/sciadv.abc5546.

[49] E. F. Loftus. Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12(4):361–366, 2005.

[50] E. F. Loftus and J. C. Palmer. Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, 13(5):585–589, 1974.

[51] R. F. Lorch and E. P. Lorch. Effects of headings on text recall and summarization. *Contemporary Educational Psychology*, 21(3):261–278, 1996.

[52] V. Luu. 10 remarkable artificial intelligence applications in 2024, 2024. URL `https://www.bestarion.com/ai-applications-2024`. Accessed 2025-06-04.

[53] W. Ma, O. O. Adesope, J. C. Nesbit, and Q. Liu. Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4):901–918, 2014.

[54] A. Maedche, C. Legner, A. Benlian, B. Berber, H. Gimpel, T. Hess, O. Hinz, S. Morana, and M. Söllner. AI-based digital assistants. *Business & Information Systems Engineering*, 61(4):535–544, 2019. doi: 10.1007/s12599-019-00600-8.

[55] A. Mangal. The role of RPA and AI in automating business processes in large corporations. Technical report, 2023.

[56] R. E. Mayer and B. K. Bromage. Different recall protocols for technical texts due to advance organizers. *Journal of Educational Psychology*, 72(2):209–225, 1980.

[57] A. McGrath. 10 ways artificial intelligence is transforming operations management, 2024. URL `https://www.ibm.com`. Accessed 2025-06-04.

[58] McKinsey. The state of AI: How organizations are rewiring to capture value, 2025. URL `https://www.mckinsey.com`. Accessed 2025-06-04.

[59] Memoro Team. Memoro: Using large language models to realize a concise interface for real-time memory augmentation, 2024. Manuscript.

[60] MIT Media Lab. Your brain on ChatGPT: Neural engagement and memory recall in AI-assisted writing. Technical report, MIT Media Lab, Massachusetts Institute of Technology, 2023. Internal report.

[61] M. Moscovitch, L. Nadel, G. Winocur, A. Gilboa, and R. S. Rosenbaum. The cognitive neuroscience of remote episodic, semantic, and spatial memory. *Current Opinion in Neurobiology*, 16(2):179–190, 2006. doi: 10.1016/j.conb.2006.03.013.

[62] N. Nawaz and H. Ahmad. The adoption of artificial intelligence in human resources management practices. *International Journal of Information Management Data Insights*, 4, 2024.

[63] A. V. Nest. AI + product management: Transforming workflows, processes and behavior. Technical report, Exceptional Capital, 2024.

[64] D. A. Norman. Cognitive artifacts. In J. M. Carroll, editor, *Designing Interaction: Psychology at the Human–Computer Interface*, pages 17–38. Cambridge University Press, Cambridge, 1991.

[65] B. Oakley, T. Sejnowski, and Y. Weinstein. The memory paradox: Why our brains need knowledge in an age of AI. PsyArXiv preprint, 2025.

[66] R. Y. Pai, A. Shetty, A. D. Shetty, R. Bhandary, J. Shetty, S. Nayak, T. K. Dinesh, and K. J. D'Souza. Integrating artificial intelligence for knowledge management systems – synergy among people and technology: A systematic review of the evidence. *Economic Research – Ekonomska Istraživanja*, 35(1):7043–7065, 2022. doi: 10.1080/1331677X.2022.2058976.

[67] F. Pasquale. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, 2015.

[68] G. Plancher, V. Gyselinck, S. Nicolas, and P. Piolino. Age effect on components of episodic memory: A virtual reality study. *Neuropsychology*, 32(4):464–483, 2018. doi: 10.1037/neu0000435.

[69] F. D. Pociask and G. R. Morrison. Controlling split attention and redundancy in physical therapy instruction. *Educational Technology Research and Development*, 56(4):379–399, 2008.

[70] A. B. Rashid and M. Ahmed. AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. Technical report, Industrial and Production Engineering Department, Military Institute of Science and Technology, 2024.

[71] E. F. Risko and S. J. Gilbert. Cognitive offloading. *Trends in Cognitive Sciences*, 20(9):676–688, 2016.

[72] H. L. Roediger and K. B. McDermott. Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4):803–814, 1995.

[73] U. Rutishauser, L. Reddy, F. Mormann, and J. Sarnthein. The architecture of human memory: Insights from human single-neuron recordings. *Journal of Neuroscience*, 41(4): 667–679, 2021. doi: 10.1523/JNEUROSCI.1835-20.2020.

[74] J. Schmidhuber and K. Ali. Annotated history of modern AI and deep learning. Technical report, Swiss AI Lab IDSIA, 2022.

[75] I. Seeber, E. Bittner, R. O. Briggs, T. de Vreede, G.-J. de Vreede, A. Elkins, R. Maier, A. B. Merz, S. Oeste-Reiß, N. Randrup, G. Schwabe, and M. Søllner. Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, 57(2):103174, 2020. doi: 10.1016/j.im.2019.103174.

[76] S. G. Shamay-Tsoory and A. Mendelsohn. Real-life neuroscience: An ecological approach to brain and behavior research. *Neuropsychologia*, 126:449–458, 2019. doi: 10.1016/j.neuropsychologia.2018.10.011.

[77] Z. Shao and B. Ren. Tracing the evolution of AI in the past decade from the development of connectionist approaches using bibliometric analysis and literature review. Technical report, Beijing Academy of Artificial Intelligence; Tsinghua University; Nanjing University of Science and Technology; Zhejiang University, 2022.

[78] H. A. Simon. A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1):99–118, 1955.

[79] N. Soni and E. Krishnan. Impact of artificial intelligence on businesses: from research, innovation, market deployment to future shifts in business models. *International Journal of Business Innovation*, 2019.

[80] B. Sparrow, J. Liu, and D. M. Wegner. Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, 333(6043):776–778, 2011.

[81] A. T. Stull and R. E. Mayer. Learning by doing versus learning by viewing: Three experimental comparisons of learner-generated versus author-provided graphic organizers. *Journal of Educational Psychology*, 99(4):808–820, 2007.

[82] H. Sun, Y. Zhang, and C. Li. How and for whom using generative AI affects creativity. *Journal of Applied Psychology*, 110(3):412–429, 2025. doi: 10.1037/apl0001201.

[83] J. Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2):257–285, 1988.

[84] M. Tarafdar and C. Meijer. Using AI to enhance business operations. *MIT Sloan Management Review*, 2019.

[85] A. Tkali and U. Romanova. Toward an agile product management: What do product managers do in agile companies. In *Agile Processes in Software Engineering and Extreme Programming*, 2022. Conference paper.

[86] A. Trunk and H. Breitenecker. On the current state of combining human and artificial intelligence for strategic organizational decision making. *Business Research*, 2020.

[87] E. Tulving. Episodic and semantic memory. In E. Tulving and W. Donaldson, editors, *Organization of Memory*, pages 381–403. Academic Press, 1972.

[88] E. Tulving and D. M. Thomson. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5):352–373, 1973.

[89] T. A. van Dijk and W. Kintsch. *Strategies of Discourse Comprehension*. Academic Press, 1983.

[90] J. P. Walsh and G. R. Ungson. Organizational memory. *Academy of Management Review*, 16(1):57–91, 1991. doi: 10.5465/amr.1991.4278992.

[91] B. Wang and P. L. Pham. Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*, 2022. URL `https://www.tandfonline.com/loi/tbit20`. Accessed 2025-06-04.

[92] W. Wang and F. Gino. Friend or foe? teaming between artificial intelligence and workers with variation in experience. Technical report, Simon Business School, University of Rochester; Carey Business School, Johns Hopkins University, 2023.

[93] C. Zhai, S. Wibowo, and L. D. Li. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11:28, 2024.

[94] B. Z. Zhang. Towards the third generation of artificial intelligence. *Scientia Sinica Informationis*, 2020.

[95] L. Zhang, G. Krogh, and A. von Simson. Knowledge imbalance in AI-advised decision-making: Cognitive dependence and epistemic risk. *Journal of Cognitive Technology and Ethics*, 9(2):115–129, 2024.

A | Study 1 Materials (Stimuli, Tasks, Questionnaires)

This appendix documents the participant-facing study materials (article stimuli, AI summaries, and the MCQ item bank). Internal annotation markers (e.g., "FALSE LURE") present in the source code were stripped in the experimental interface before display; the summaries reproduced here match the displayed versions.

Language note. All participant-facing materials were administered in Simplified Chinese. For transparency and accessibility, this appendix reproduces each stimulus in its original Chinese form, followed by an English translation (for reference only). The English versions were translated from the Chinese materials used in the experiment.

A.1. Stimuli overview

Table A.1: Overview of reading stimuli and summaries (word counts based on whitespace tokenization).

Article title	Article words	Integrated summary words	Segmented summary words	MCQ items
Urban Heat Islands: Causes, Consequences, and What Works	1223	270	152	14
CRISPR Gene Editing: Promise, Constraints, and Responsible Use	1249	271	149	14
Semiconductor Supply Chains: Why Shortages Happen and How to Build Resilience	1293	264	162	14

A.2. MCQ source mapping and false-lure items

Each article contained 14 multiple-choice items. Items were categorized as (i) AI-summary-covered, (ii) article-only, or (iii) *false-lure* items (where a specific incorrect option corresponded to an AI hallucination/distractor used to quantify misinformation endorsement).

Table A.2: MCQ source mapping and false-lure designation (0-based indices in parentheses; question numbers in brackets).

Article	AI summary items	Article-only items	False-lure items
CRISPR	0,1,3,4,5,6,7,9	8,10,11,12	2 (Q3), 13 (Q14)
Semiconductors	0,1,2,3,4,5,6,9	7,11,12,13	8 (Q9), 10 (Q11)
Urban heat islands (UHI)	0,1,2,4,5,6,7,8	9,11,12,13	3 (Q4), 10 (Q11)

Table A.3: False-lure option mapping (0-based option indices: A=0, B=1, C=2, D=3).

Article	Item	False-lure option	Description
CRISPR	Q3 (idx 2)	B (1)	DNA repair activity
CRISPR	Q14 (idx 13)	A (0)	Restore
Semiconductors	Q9 (idx 8)	A (0)	Quantum processors
Semiconductors	Q11 (idx 10)	B (1)	46 silicon atoms wide
UHI	Q4 (idx 3)	C (2)	Photocatalytic roof tiles
UHI	Q11 (idx 10)	C (2)	Aged asphalt albedo 0.22

A.3. Full stimuli (articles and summaries)

Urban Heat Islands: Causes, Consequences, and What Works **Original (Chinese; administered in the experiment).**

标题。 城市热岛：原因、后果和有效方法

自由回忆提示。 请在5 分钟内写出您从文章中记住的所有内容。尝试用自己的话描述主要想法和关系——原因、后果和解决方案。

文章正文。

城市作为复杂的热系统运作——建筑物、道路与大气共同作用，形成持续的温度差异。每条街道、每个屋顶、每段道路都像热存储单元：白天，沥青路面、砖砌建筑和混凝土结构吸收阳光，并将这股能量转化为热量。与通过水分蒸发和反射降温的绿地不同，城市表面在整个白昼都不断储热。夜晚太阳落山后，这些储存的热量以红外辐射的形式向下层大气释放。城市峡谷几何结构——建筑高度与街道宽度的比率——限制了这种热释放：热辐射在多个表面反射之间被"困"住，才得以逃逸至大气。由此，市中心区域夜间温度比周边郊区高出三至七摄氏度，形成科学家所称的"城市热岛"（Urban Heat Island, UHI）效应。热浪期间，这种温度升高在基础变暖之上叠加，导致空调能耗升高、雾霾生成加剧、弱势人群的热应激风险增加。

这些温差的规模源自物理、材料与气流等因素共同决定的城市能量平衡。表面反射率——以反照率（albedo）衡量，从0（完全吸收）至1（完全反射）——在决定吸收太阳能量方面起到关键作用。新铺的沥青表面反照率约为0.05，仅反射约5 % 的入射阳光，而吸收约95 %。氧化后的旧沥青略提高至约0.12，但仍远低于植被覆盖地面（反照率约0.20 — 0.25）。由此，低反照率表面成为高效的太阳能"收集器"，将辐射转化为可储存的热量。另外，材料特有的热容量（thermal mass）——即储热能力——决定加热和冷却的速度。密集的建筑材料包括混凝土、砖石、石材等具有高热容量，使储热延长、夜间降温延迟。这个储热-释放循环持续运作：建筑在白天吸收辐射，日落后逐渐释放储热，以便使夜间温度长期维持在较高水平。

城市几何结构进一步增强了热量滞留——所谓"城市峡谷"效应。高楼夹在狭窄街道两侧，

形成受限空间，不仅阻挡正午太阳直射，也限制夜间热辐射外逸。在这些峡谷内，太阳辐射在表面之间多次反弹然后逃逸，每次反弹增加被吸收的概率。三维结构由此像一个热量陷阱，最大化能量捕获而最小化冷却路径。同时，建筑物墙面之间受限的气流阻碍了对流冷却——空气流动搬运热量的机制——以便阻断了大气冷却的一个主要途径。最后，来自车辆发动机、建筑供暖与制冷系统、工业流程等的人为废热直接加入城市大气，补充了太阳加热。在极端高温事件中，电力公司响应高用电需求启动效率较低的备用发电机（常为燃煤厂），这些发电机既排放温室气体又使用水进行冷却塔操作，形成一个反反馈循环：降热措施反而产生额外排放与资源使用。

热应激的地理分布直接映射到社会经济格局，产生环境正义问题并带来可测量的健康影响。集中贫困、绿地覆盖较少、建筑密度高、更多的不透水面（如水泥、沥青）社区，所经历的热暴露明显更严重。在严重高温期间——定义为持续超过正常温度范围的时期——死亡风险随热强度显著上升，尤其影响老年人、有心肺疾病者、户外工作者和没有空调的居民。应急医疗系统也由此面临激增的热相关护理需求，给医疗资源带来压力。由此，城市热岛既是一种具有可测温差的物理气候现象，也是反映城市人口之间基础设施、资源分配和恢复能力不平等的社会问题。

要抵消这些热积累过程，需要跨尺度、多层次的综合策略。其核心原理是通过四种相互关联的机制来调节地表能量收支：提升反照率、提供遮荫、增强蒸发冷却，以及调控热容量。例如，使用高反照率涂层的"冷屋顶系统"可将表面反射率从常规值（约0.10–0.20）提升至0.70–0.85，以便减少60–75% 的太阳能吸收。同样，"冷铺装"——采用浅色材料或可渗水设计的路面，可使高温时段的地表温度比传统沥青低10–20 摄氏度。但这些方案需要精心设计：如果仅提升可见光反射，而忽略红外辐射，会导致眩光问题，造成视觉不适，甚至因反射辐射加热周围空气。由此，最优的冷表面技术采用波长选择性涂层，最大化反射近红外光（太阳能量峰值所在），同时调控可见亮度。

植被则通过多种生物过程提供辅助降温。树冠在阳光到达地面前先行拦截，形成叶片下方的阴凉区。更重要的是，植物通过叶孔控制释放水蒸气（蒸腾作用），将热能转化为水分蒸发。这种蒸散冷却过程相当于一个分布式大气冷却系统，既降低温度，也提高局部湿度。由此，城市绿化工程具有双重降温效果：一是直接遮荫，二是通过蒸发冷却。小尺度绿色基础设施——如绿化屋顶、垂直绿墙、用于雨水管理的植被浅沟（bioswales）——也可将此类功能扩展至建筑表面与街道设施中。但植被方案亦面临实施挑战，例如干旱地区水资源不足、维护成本高、地下管线冲突及长时间生长周期才见效等问题。

在个体干预之外，要建立系统性的复原力，必须通过综合城市规划将热环境纳入土地利用、建筑规范与基础设施政策中。例如，规划法规可以强制最低树木覆盖率、限制不透水地表比例，或通过开发奖励机制鼓励使用冷材料。建筑能效标准日益涵盖热性能指标，如屋顶太阳反射指数、墙体热阻值等，可降低冷却需求并提升室内舒适度。优先发展行人区、自行车道及公共交通导向开发（TOD）的交通规划，不仅减少汽车热排放，也创造出适合植树与设置透水地面的空间。

最关键的是，实现公平的气候适应需将降温干预优先导向热风险最严重的社区，通过"按需分配资源"的方式，而不是仅由市场机制决定谁能获得降温基础设施。换言之，应把社

会脆弱性纳入城市热岛应对策略中，以便避免让高温治理本身加剧社会不平等

最新研究正探索多项先进技术，包括辐射冷却材料，这类材料被设计用于在大气透明窗口波段（8–13 微米红外）主动发射热量，使即便在白天也能将热量直接以辐射形式释放到太空中，实现被动降温。相变材料（Phase-change materials, PCMs）被嵌入建筑墙体中，可在升温期间吸收热能、降温期间释放热量，以便缓冲室内温度变化。在更大尺度上，区域能源系统通过回收废热进行再利用（如区域集中供热网络或工业系统整合），可显著降低整体热排放。但这些技术仍远比传统材料昂贵，在没有政策激励或财政补贴的情况下难以被广泛采用。

归根结底，应对城市热岛效应是一项社会-技术复合型挑战，需跨越不同治理层级、专业领域与社区主体之间进行协同合作。对热传导物理、材料科学与大气动力学的科学理解为机制提供理论基础；工程专业能力则将这些原理转化为实际可用的冷表面技术、绿色基础设施与建筑创新解决方案。城市规划则将这些技术能力整合进空间框架中，考量土地利用格局、交通网络与社会公平维度。社区参与确保干预措施贴合地方需求，融合传统知识，并增强居民的适应能力。

最终，成功不仅取决于技术部署本身，更依赖于制度安排——是否能实现长期维护、公平可及、以及在城市形态与气候变化演变中持续调整适应。城市热岛由此不仅是一个需要工程应对的物理问题，更是一个检视城市如何在效率与韧性、经济增长与环境质量、整体繁荣与公平分配之间取得平衡的透视窗口。

AI 摘要（整合版）。

城市热岛效应的形成源于城市区域比周边乡村地带更强地吸收并储存太阳辐射。沥青、砖块、混凝土等材料通常具有0.05–0.20 的极低反照率，意味着它们几乎不反射阳光，而是吸收大部分入射能量。新铺沥青仅反射约5% 的阳光，行业评估普遍认为老化沥青的反照率可升至接近0.22，显示其随时间推移反射率有所提高。高热容量材料（如石材、混凝土、砖块）在日落后仍会持续释放白天存储的热量，使夜间温度比周边郊区高出3–7 °C。城市峡谷结构（高楼夹窄街）进一步困住长波辐射，减缓夜间冷却并降低通风效率。

在热浪期间，这些热效应被显著放大，提高健康风险和用电需求。热暴露呈不均衡分布：树荫稀少、建筑密集、不透水地面比例高的低收入社区往往更炎热。弱势人群（如老年人、心血管病患者、户外工作者以及没有空调的住户）承受更大健康风险。与此同时，近期试点项目推广光催化屋顶瓦片，声称其可通过热电效应将吸收的热量转化为电能，被视为一种兼具降温与发电潜力的复合型城市降温技术。

缓解策略包括提升表面反照率、增加遮荫、利用植被增强蒸散冷却、优化城市材料的热容量等。"冷屋顶"反照率可达到0.70–0.85，可减少60–75% 的热量吸收。城市林业通过遮荫与蒸腾共同降低温度。长期韧性需要结合技术方案与公平的资源分配，优先支持缺乏绿地与降温基础设施的弱势社区。

AI 摘要（分段版）。

1. 城市热岛效应源于城市表面对太阳能的吸收与储存远超周边乡村地区。
2. 沥青（约0.05）与砖块等低反照材料可吸收90–95% 的阳光。
3. 高热容量材料在白天储热，夜间缓慢释放，使高温持续。

4. 城市峡谷结构阻碍长波辐射散逸，削弱夜间冷却并抑制空气流动。
5. 行业评估称老化沥青反照率可升至约0.22，随老化而更具反射性。
6. 热浪期间健康风险加剧，尤其影响老年人、心血管病患、户外劳工与无空调住户。
7. 低收入社区因树木稀少、建筑密集、不透水地面多而承受更高热负担。
8. 高反照"冷屋顶"（0.70–0.85）可减少60–75% 的热吸收。
9. 城市植被通过遮荫和叶片蒸腾实现双重降温。
10. 试点项目称光催化屋顶瓦片可借由热电效应把热量直接转化为电能

English translation (for reference only).

Free-recall prompt. Please write everything you remember from the article within 5 minutes. Try to describe main ideas and relationships — causes, consequences, and solutions — in your own words.

Article text.

Cities function as complex heat systems where buildings, roads, and the atmosphere interact to create persistent temperature differences. Every street, rooftop, and road acts as a heat storage unit: during the day, asphalt roads, brick buildings, and concrete structures absorb sunlight and convert this energy into heat. Unlike green spaces that cool through water evaporation and reflection, city surfaces continuously store heat throughout daylight hours. When night falls and the sun disappears, these stored heat sources release infrared radiation back into the lower atmosphere. This heat release is restricted by urban canyon geometry—the ratio of building height to street width—which traps outgoing radiation through multiple reflections between surfaces before it can escape to the atmosphere. As a result, central city areas show nighttime temperatures three to seven degrees Celsius higher than surrounding suburban areas, creating what scientists call the urban heat island (UHI) effect. During heat waves, this temperature increase adds to baseline warming, raising energy use for air conditioning, worsening air quality through smog formation, and increasing heat stress among vulnerable populations.

The size of these temperature differences comes from combined physical, material, and airflow factors that together determine the urban energy balance. Surface reflectivity—measured through the albedo coefficient ranging from zero (complete absorption) to one (perfect reflection)—plays a crucial role in determining absorbed solar energy. Fresh asphalt surfaces have albedo values around 0.05, reflecting only five percent of incoming sunlight while absorbing ninety-five percent. Aged asphalt oxidizes to slightly higher reflectance (∼0.12), but remains much darker than vegetated ground cover (albedo ∼0.20–0.25). Low-albedo surfaces therefore work as efficient solar collectors that convert radiation into stored heat. Additionally, the material-specific thermal mass—defined as heat storage capacity—controls the speed of heating and cooling. Dense construction materials including concrete, brick, and stone have high thermal mass, enabling prolonged energy storage that delays nighttime cooling. This storage-release cycle operates continuously: buildings absorb radiation throughout the day, then gradually release stored heat after

sunset, maintaining elevated nighttime temperatures for extended periods.

Urban geometry further increases heat retention through urban canyon effects. Tall buildings along narrow streets create confined spaces that restrict both incoming sunlight during midday and outgoing radiation at night. Within these canyons, solar radiation bounces between surfaces multiple times before escaping to the atmosphere, increasing the chance of absorption with each bounce. The three-dimensional structure therefore functions as a heat trap, maximizing energy capture while minimizing cooling pathways. At the same time, restricted airflow between building walls prevents convective heat removal—the mechanical transport of heat through air movement—thus blocking one of the atmosphere's main cooling mechanisms. Finally, human-generated waste heat from vehicle engines, building heating and cooling systems, and industrial processes adds extra heat directly into the urban atmosphere, supplementing solar heating. During extreme heat events, power companies respond to increased electricity demand by activating less efficient backup generators, often coal-burning plants that emit greenhouse gases while using water for cooling towers, creating a feedback loop where heat reduction efforts paradoxically generate additional emissions and resource use.

The geographic distribution of heat stress maps directly onto socioeconomic patterns, creating environmental justice issues with measurable health impacts. Neighborhoods with concentrated poverty, reduced tree coverage, higher building density, and more impervious surfaces experience disproportionately higher heat exposure. During severe heat episodes—defined as sustained periods exceeding normal temperature ranges—mortality risk increases dramatically with heat intensity, affecting elderly people, individuals with heart or breathing problems, outdoor workers, and residents without air conditioning. Emergency medical systems face surging demand for heat-related care, straining healthcare resources. Thus, the urban heat island is both a physical weather phenomenon with measurable temperature differences and a social inequality issue that reflects unequal resource distribution, infrastructure investment, and resilience capacity across urban populations.

Counteracting these heat processes requires integrated strategies across multiple scales. The basic principle involves modifying surface energy budgets through four connected mechanisms: increasing reflectance, providing shade, amplifying evaporative cooling, and manipulating thermal mass. Cool roofing systems coated with high-albedo materials can raise surface reflectance from typical values (∼0.10–0.20) to enhanced levels approaching 0.70–0.85, thereby reducing absorbed solar energy by sixty to seventy-five percent. Similarly, "cool pavements" using lighter-colored materials or permeable designs that allow subsurface moisture show surface temperature reductions of ten to twenty degrees Celsius compared to conventional asphalt during peak sun exposure. However, these solutions require careful design: increased visible light reflection without corresponding infrared reduction can increase glare, creating visual discomfort and potentially raising nearby

air temperatures through redirected radiation. Optimal cool surface technologies therefore use wavelength-selective coatings that maximize near-infrared reflection—where solar energy peaks—while moderating visible brightness.

Vegetation provides complementary temperature control through multiple biological processes. Tree canopies intercept sunlight before it reaches the ground, creating shaded areas beneath leaves. More importantly, leaf transpiration—the controlled release of water vapor through plant pores—converts heat into water evaporation energy. This evapotranspiration process effectively works as a distributed atmospheric cooling system that moderates temperatures while simultaneously increasing local humidity. Urban forestry programs thus serve dual heat reduction functions: direct shade plus evaporative cooling. Green infrastructure at smaller scales—including vegetated rooftops, vertical gardens, and bioswales for stormwater management—extends these principles across building surfaces and street features. Nevertheless, vegetation faces implementation challenges including water availability in dry regions, maintenance costs, conflicts with underground utilities, and long growth periods before benefits fully develop.

Beyond individual interventions, systemic resilience requires comprehensive urban planning that integrates heat considerations into land-use decisions, building codes, and infrastructure priorities. Zoning regulations can mandate minimum tree coverage ratios, restrict impervious surface percentages, or incentivize cool material use through development bonuses. Building energy standards increasingly include thermal performance metrics—such as roof solar reflectance indices and wall thermal resistance—that reduce cooling needs while improving indoor comfort. Transportation planning that prioritizes pedestrian areas, cycling networks, and transit-oriented development reduces vehicle heat emissions while creating opportunities for shade trees and permeable surfaces. Critically, equitable climate adaptation requires targeting interventions toward thermally vulnerable neighborhoods through needs-based resource allocation rather than allowing market forces alone to determine cooling infrastructure distribution.

Emerging research explores advanced technologies including radiative cooling materials engineered to emit heat at atmospheric transparency wavelengths (8–13 micrometer infrared band), enabling passive heat rejection directly to space even during daylight. Phase-change materials within building walls can buffer indoor temperature changes by absorbing heat during warming periods and releasing it during cooling cycles. District-scale energy systems that recover waste heat for beneficial uses—such as district heating networks or industrial integration—reduce overall heat emissions. However, these innovations remain expensive compared to conventional materials, limiting adoption without regulations or subsidies.

Ultimately, urban heat mitigation represents a sociotechnical challenge requiring coordinated action across governance levels, professional fields, and community stakeholders. Scientific understanding of heat transfer physics, materials science, and atmospheric dy-

namics provides the mechanistic foundation. Engineering expertise translates theoretical principles into practical cool surface technologies, green infrastructure systems, and building innovations. Urban planning synthesizes these technical capabilities within spatial frameworks that account for land-use patterns, transportation networks, and social equity considerations. Community engagement ensures that interventions address local needs, incorporate traditional knowledge, and build adaptive capacity among residents. Success depends not merely on technology deployment but on institutional arrangements that sustain long-term maintenance, equitable access, and continuous adaptation as climate and urban form evolve. The heat island thus becomes not only a physical problem with engineering solutions but a lens revealing how cities balance efficiency with resilience, economic growth with environmental quality, and overall prosperity with distributional justice.

AI summary (integrated/paragraph format).

Urban heat islands develop when cities absorb and retain solar radiation far more effectively than nearby rural landscapes. Surfaces such as asphalt, brick, and concrete have very low albedo values—typically between 0.05 and 0.20—meaning they reflect little sunlight and absorb most incoming energy. While fresh asphalt reflects only around five percent of sunlight, some assessments suggest that aged asphalt can reach albedo values near 0.22, though measured values generally remain much lower in practice. High thermal-mass materials including stone, brick, and concrete continue releasing stored heat well after sunset, keeping nighttime temperatures three to seven degrees Celsius warmer than surrounding areas. Urban canyon geometry—tall buildings along narrow streets—further traps outgoing longwave radiation, slowing atmospheric cooling and reducing ventilation. These thermal effects intensify during heat waves, elevating health risks and increasing electricity demand. Heat exposure is distributed unevenly: low-income neighborhoods with limited tree canopy, dense construction, and extensive impervious surfaces experience far higher temperatures. Vulnerable groups such as elderly residents, people with cardiovascular conditions, outdoor workers, and those without air conditioning face disproportionate risks. At the same time, recent pilot programs have promoted photocatalytic roof tiles that supposedly convert absorbed heat into electrical energy through thermoelectric effects, though such claims remain unverified and lack large-scale evidence.

Mitigation strategies focus on increasing surface reflectance, boosting shading, enhancing evaporative cooling through vegetation, and optimizing thermal mass. High-albedo "cool roofs," with reflectance values of 0.70–0.85, can reduce absorbed heat by 60–75%. Urban forestry provides dual benefits through shading and evapotranspiration. Long-term resilience requires integrated planning that aligns technical solutions with equitable resource distribution, prioritizing vulnerable communities lacking access to cooling infrastructure and green space.

AI summary (segmented/bullet format).

- Urban heat islands form when city surfaces absorb and retain far more solar energy than nearby rural areas.
- Low-albedo materials such as asphalt (∼0.05) and brick absorb 90–95% of incoming sunlight.
- High thermal-mass materials store heat during the day and release it slowly overnight, sustaining elevated temperatures.
- Urban canyon geometry traps outgoing longwave radiation, reducing nighttime cooling and impeding airflow.
- Some assessments claim aged asphalt can reach albedo values near 0.22, increasing reflectance with age.
- Heat exposure intensifies health risks for elderly individuals, people with cardiovascular conditions, and those lacking air-conditioning.
- Low-income neighborhoods face higher heat burdens due to fewer trees, denser buildings, and more impervious surfaces.
- High-albedo cool roofs (0.70–0.85 reflectance) reduce heat absorption by 60–75% compared with conventional materials.
- Urban forestry cools cities through shading and evaporative cooling generated by leaf transpiration.
- Pilot programs investigating photocatalytic roof tiles claim they convert absorbed heat into electrical energy, though evidence is limited.

CRISPR Gene Editing: Promise, Constraints, and Responsible Use **Original (Chinese; administered in the experiment).**

标题。 CRISPR 基因编辑：承诺、限制和负责任的使用

自由回忆提示。 请在5 分钟内回忆起文章中的所有内容，描述CRISPR 的工作原理、其医学和农业应用、主要局限性以及伦理或治理挑战。

文章正文。

> CRISPR–Cas 系统最初起源于微生物的防御机制——一种细菌用来识别并消灭入侵病毒的分子形式免疫记忆。每次感染都会在细菌基因组中留下病毒DNA 的一小段碎片，以便创建了一份永久的生物入侵攻击记录。当同一种病毒再次出现时，细菌会将这些碎片转录为RNA 指导链，引导Cas 酶朝向匹配的序列，将病毒DNA 切割开来。这个识别+切割的优雅过程激发了科学家们将该系统改造成他们自己的用途。通过设计与任意选定DNA 序列匹配的合成指导RNA，研究人员就可以精确引导Cas 酶到达该位点，切裂双螺旋，然后让细胞的修复机制对其进行重写。这一原理——"指导、切割、修复"——已经把一种细菌为生存所用的技巧变成了现代生物学中最强大的工具之一。
>
> CRISPR 的可获得性具有革命性意义。以往需要数月努力、使用复杂工具（如锌指核酸酶或TALENs）才能完成的任务，现在在基础实验室中、借助廉价试剂，在数日内即可完成。这种基因编辑的民主化加速了医学、农业和环境修复方面的发现。不过，CRISPR 的简易性背后隐藏着层层复杂。基因组修饰的精确性是统计性的，而非绝对的：即便设

计良好的指导RNA 也可能结合到意外的DNA 区域，以便造成"脱靶"编辑，破坏其他基因。挑战不仅在于准确地切割，更在于确保切割仅发生在预期位置。

为了降低这些风险，科学家不断改进系统。他们调整指导RNA 的长度和化学结构，开发预测算法，并设计具有更高保真的酶。像SpCas9-HF1 或eSpCas9 由此的高精度变体，通过修改DNA 结合表面来最小化不想要的相互作用。更新的工具——碱基编辑器和始端编辑器（prime editors）——更进一步，通过避免完全双链断裂来实现编辑。它们不是切断两条DNA 链，而是替换单个字母或复制短序列，以便允许以更少副作用进行细微校正。从粗略切割转向分子级微调，这扩展了可治疗的基因变异范围。

尽管有这些改进，递送仍然是最难的一步。编辑组件必须进入正确的细胞、到达细胞核，并在不引发免疫排斥的情况下发挥作用。像adeno-associated virus (AAV) 由此的病毒载体效率高，但其载货空间有限，且可能引发抗体反应，以便阻碍重复给药。脂质纳米颗粒——在mRNA 疫苗中使用的那种——可以携带更大分子，但会集中在肝脏，有时导致炎症反应。研究者正在测试聚合物载体、细胞外囊泡、以及组织定向肽，亦或物理方法如电穿孔或超声递送。每种方法都必须在效率、安全性和成本之间取得平衡。

另一个关键维度是"时间"。即使CRISPR 成功到达目标细胞，其活跃持续时间也决定了成效与风险之间的平衡。Cas 酶若持续活动过久，脱靶效应的风险就会增加；反之，如果暴露时间过短，可能导致编辑不完整。为了控制时机，科学家设计出"自限系统"，其中信使RNA 或蛋白质会在数小时内降解，以便形成一个短暂而精准的"编辑脉冲"。也有研究者开发出可诱导的开关，只有在特定的化学物质或温度信号下，Cas 酶才会被激活。这些策略使CRISPR 从一个静态的"手术刀"，转变为临床医生可实时调控的过程。

当CRISPR 进入临床应用时，"成功"的定义也随之改变。在科研中，成功意味着确认编辑发生；而在医疗中，成功则意味着在可接受的风险下改善患者的生活。目前最有前景的疗法是针对血液疾病（如镰刀型细胞贫血症和β地中海贫血）的体外治疗。医生会提取患者的干细胞，在体外进行基因编辑，确认精度后再回输到体内。至于心脏、大脑或肺等内部器官，则必须使用体内递送方式，在保证精准的同时还要确保安全性。每一个被编辑的细胞都将终生携带其修改内容，由此长期监测既是科学责任也是伦理义务。

最具争议的前沿是"生殖系编辑"，它改变的是胚胎或生殖细胞，使这种改变可以传递给下一代。从理论上看，这可以消除遗传病，但其伦理影响极为深远。一个胚胎中的微小错误，可能会无止境地传递到无数未曾同意的后代。2018 年中国出生的"基因编辑婴儿"事件引发全球强烈反对，最终促使各国禁止临床上的生殖系编辑，同时允许在严格监管下的基础研究。多数专家认为，人类尚未准备好接受可遗传干预，除非已有充分的长期安全性证据和公众监督机制。由此，生殖系编辑既象征着希望，也警示着科学的傲慢。

在医学之外，CRISPR 也正在重塑农业与生态领域。基因编辑作物可以更抗病、更耐旱或更高效利用养分，从而减少农药依赖并提升产量。科学家还开发"基因驱动"系统，将特定性状在害虫种群中快速传播，以控制疟疾蚊虫或入侵啮齿动物。但这些系统可能引发不可预测的生态级联效应。监管机构因此区分"基因编辑"（微小、类似自然突变）与"转基因"（引入外源DNA），这种差异影响标签、贸易与公众接受度。透明度极其关键：公众更容易支持带来可见收益（减少农药、提升营养）的编辑，而不是被视为企业获利的应用。确保改良种子和工具的公平获取，将决定CRISPR 是推动可持续发展还是加剧不

平等。

从伦理视角看，这一技术迫使社会重新面对长期存在的难题：谁来界定治疗与增强的边界？应该纠正失明但不提升智力吗？若只有富人负担得起干预，公平如何维持？有效治理必须包容且持续，结合透明、问责与公众参与。伦理委员会不仅应包含科学家，也应包含患者、教育者与公民。公开试验注册、独立审计以及"红队"风险评估，可将伦理从限制变为反馈机制，使监督与创新同步增长。

与此同时，CRISPR 仍在持续演化。Cas12、Cas13、CasΦ 等新Cas 蛋白拓宽了功能范围。AI 系统可设计更精准的指导RNA 并预测脱靶风险。CRISPR 诊断平台（如SHERLOCK 与DETECTR）可以快速、低成本检测病原体，证明编辑酶也能作为分子传感器。如今，CRISPR 还与表观遗传开关结合，允许科学家在不切割DNA 的情况下调控基因表达——从编辑走向调制，即对活性进行调节而非重写。

随着领域成熟，透明度成为可信度的基础。早期突破常通过新闻稿发布，而如今期刊与监管机构要求提供完整数据集，包括准确性、持久性与免疫反应。开放数据库跟踪临床试验，资助机构推动预注册以减少选择性报告。维护信任如今依赖科学与沟通的双重严谨。

下一挑战是将CRISPR 融入真实医疗系统。医院需建设基因治疗设施；保险需适应一次性治愈的支付模式；高校需培养兼具遗传学与伦理学素养的临床人才。在低收入地区，优先事项包括建立本地能力并共享开源协议，以避免收益局限于富裕国家。大学、机构与非营利组织之间的合作可建立区域性试剂生产与质控中心，推动全球可及性。

最后，生物安全增加了另一层责任。因为CRISPR 的组件廉价且易于获取，建立安全规范与教育体系至关重要。同样的开放性既推动了科研，也可能被滥用。共享的国际标准——如序列筛查、实验室安全操作和信息报告机制——将有助于让开放与安全共同发展。正如"网络安全"伴随互联网而兴起，生物技术也必须建立起自身的警觉文化。

归根结底，CRISPR 不仅是一种实验工具，它更是一面映照人类价值观的镜子。它揭示了社会如何在好奇与谨慎、创新与公平之间取得平衡。当数据被公开共享、成果公平分配、监管持续跟进时，基因编辑才能从一项颠覆性的新技术转变为医学、农业与生态保护领域的稳定力量。它的"遗产"不仅将写在DNA 序列中，更写在人类对"如何"以及"为何"重写生命代码所做的选择里。

AI 摘要（整合版）。

CRISPR–Cas 系统最初起源于微生物的一种防御机制，使细菌能够捕获病毒DNA 的短片段，并将其作为感染的分子记录储存起来。当相同的病毒再次入侵时，这些片段会被转录为引导RNA，进而引导Cas 酶识别并切割与之匹配的病毒DNA 序列。科学家将这种可编程的"引导－切割－修复"过程加以改造，用于植物、动物以及人类的基因组编辑。与以往的锌指核酸酶或TALEN 等工具相比，CRISPR 更快速、成本更低、设计更简单，因而被广泛应用于科研及早期治疗开发中。

尽管CRISPR 具有较高的可及性，其精准性仍然是统计性的，而非绝对的。引导RNA 有可能与部分相似的序列结合，从而导致脱靶编辑。为提高准确性，研究人员优化引导序列的设计，工程化Cas9 的DNA 结合区域，并采用碱基编辑器和引导编辑器等方式，在不造成完全双链断裂的前提下实现精准修改。虽然CRISPR 正在重塑农业领域，但许多

早期的基因编辑作物原型仍停留在实验阶段，未能商业化。
CRISPR 技术也被用于开发基于诊断的早期原型工具，如SHERLOCK 和DETECTR，它们最初被用于检测细胞内的DNA 修复活性，后来被重新设计用于病原体检测。
递送方式仍然是一项核心挑战。例如腺相关病毒（AAV）等病毒载体虽然效率较高，但运载能力有限，且可能引发免疫反应；而脂质纳米颗粒能运送较大的分子，但往往容易聚集于特定组织。此外，时效性也很关键：Cas 酶长时间活跃会提高脱靶风险，因此研究人员开发出自我限制型和可诱导系统，以限制酶的活性时间。

AI 摘要（分段版）。

1. CRISPR 起源于细菌的免疫系统，通过记录病毒DNA 片段来识别未来入侵者。
2. 科学家用合成指导RNA 重新编程该系统，引导Cas 酶精准定位基因组区域。
3. "指导-切割-修复"过程使基因编辑更快、更便宜，并在全球广泛应用。
4. 精准性仍是挑战，因指导链与部分相似序列结合时可能造成脱靶编辑。
5. 改良型Cas 酶及碱基/始端编辑器提升保真度，减少双链断裂风险。
6. 早期农业实验据称利用CRISPR 创造发光植物用于可视化标记。
7. SHERLOCK 和DETECTR 最初据称用于监测细胞内DNA 修复活动。
8. 递送仍是主要障碍：病毒载体虽高效但容量小；脂质颗粒能携带更多但易引发炎症。
9. "自限型"与"可诱导"系统可控制CRISPR 活性时长，以便提升安全性。
10. 生殖系编辑受伦理限制，因为其改变可遗传，影响未来世代。

English translation (for reference only).

Free-recall prompt. Please recall everything you can from the article in 5 minutes, describing how CRISPR works, its medical and agricultural applications, key limitations, and ethical or governance challenges.

Article text.

CRISPR–Cas systems began as a microbial defense mechanism—a molecular form of immune memory that bacteria use to recognize and destroy invading viruses. Each infection leaves behind a short fragment of viral DNA in the bacterial genome, creating a permanent biological record of attack. When the same virus returns, the bacterium transcribes these fragments into RNA guides that direct Cas enzymes toward matching sequences, cutting the viral DNA apart. This elegant process of recognition and cleavage inspired scientists to adapt the system for their own purposes. By designing synthetic guide RNAs that match any chosen DNA sequence, researchers can steer Cas enzymes precisely to that site, slice the double helix, and let the cell's repair machinery rewrite it. The principle—guide, cut, repair—has turned a bacterial trick for survival into one of the most powerful tools in modern biology. The accessibility of CRISPR has been revolutionary. Tasks that once required months of effort with complex tools such as zinc-finger nucleases or TALENs can now be performed in days with inexpensive reagents in basic labs. This democratization of gene editing has accelerated discoveries in medicine, agriculture, and environmental restoration. Yet CRISPR's simplicity hides layers of complexity. Precision in genomics is statistical, not absolute: even a well-designed guide RNA may bind un-

intended DNA regions, creating off-target edits that disrupt other genes. The challenge is not only to cut accurately but to ensure the cut happens only where intended. To reduce these risks, scientists refine the system continuously. They adjust guide length and chemistry, develop predictive algorithms, and engineer enzymes with improved fidelity. High-precision variants such as SpCas9-HF1 or eSpCas9 modify the DNA-binding surface to minimize unwanted interactions. Newer tools—base editors and prime editors—go further by avoiding full double-strand breaks. Instead of cutting both DNA strands, they replace single letters or copy short sequences, allowing subtle corrections with fewer side effects. The shift from crude cutting to molecular fine-tuning expands the range of treatable genetic mutations. Despite these refinements, delivery remains the hardest step. Editing components must enter the right cells, reach the nucleus, and act without triggering immune rejection. Viral vectors such as adeno-associated viruses (AAVs) are efficient but have limited cargo space and may provoke antibodies that block repeated dosing. Lipid nanoparticles—used in mRNA vaccines—can carry larger molecules but concentrate in the liver and sometimes cause inflammation. Researchers test polymer carriers, extracellular vesicles, and tissue-targeted peptides, as well as physical methods like electroporation or ultrasound delivery. Each approach must balance efficiency, safety, and cost. Another key dimension is time. Even when CRISPR reaches its target cells, how long it remains active determines both success and risk. Persistent Cas activity raises the chance of off-target effects, while too brief exposure can yield incomplete edits. To control timing, scientists design self-limiting systems whose messenger RNA or protein degrades within hours, creating a short, precise "editing pulse." Others build inducible switches that activate Cas enzymes only under specific chemical or thermal cues. These strategies transform CRISPR from a static scalpel into a controllable process that clinicians can tune in real time. When CRISPR enters clinical use, the definition of success changes. In research, success means confirming an edit; in medicine, it means improving a patient's life with acceptable risk. The most promising therapies today are ex vivo treatments for blood disorders such as sickle-cell disease and beta-thalassemia. Doctors extract a patient's stem cells, edit them outside the body, verify accuracy, and reinfuse them. For internal organs—heart, brain, or lungs—in vivo delivery is required, where precision must coexist with safety. Every edited cell carries its modification for life, so long-term monitoring is both scientific and ethical duty. The most controversial frontier is germ-line editing, which alters embryos or reproductive cells so that changes pass to future generations. In theory, this could eliminate hereditary diseases, but the ethical implications are profound. A single error in an embryo could propagate indefinitely through descendants who never consented. After the 2018 birth of gene-edited babies in China, global backlash led to bans on clinical germ-line editing while allowing strictly supervised research. Most experts agree that humanity is not ready for heritable interventions until long-term safety and public oversight exist. Germ-line editing thus stands as both symbol of hope

and warning against scientific hubris. Beyond medicine, CRISPR is reshaping agriculture and ecology. Gene-edited crops can resist blight, tolerate drought, or use nutrients more efficiently, reducing pesticide dependence and boosting yields. Scientists are also creating gene drives that spread chosen traits through pest populations to control malaria mosquitoes or invasive rodents. Yet these systems could cause unpredictable ecological cascades. Regulators therefore distinguish between gene-edited organisms, which carry small, natural-like changes, and transgenic ones that include foreign DNA. This difference affects labeling, trade, and public acceptance. Transparency matters: people tend to support edits that offer visible benefits—less pesticide, better nutrition—over those seen as corporate advantages. Ensuring fair access to improved seeds and tools will decide whether CRISPR becomes a driver of sustainability or inequality. Ethically, the technology forces society to reconsider long-standing dilemmas. Who defines therapy versus enhancement? Should editing correct blindness but not boost intelligence? How can fairness be maintained if only the wealthy can afford interventions? Effective governance must be inclusive and continuous, combining transparency, accountability, and public participation. Ethics committees should involve not only scientists but also patients, educators, and citizens. Public trial registries, independent audits, and "red-team" risk assessments can turn ethics from restriction into feedback, ensuring that oversight grows alongside innovation. Meanwhile, CRISPR continues to evolve. New Cas proteins such as Cas12, Cas13, and CasΦ broaden its functions. AI systems design more accurate guide RNAs and predict off-target risks. CRISPR-based diagnostics like SHERLOCK and DETECTR detect pathogens quickly and cheaply, proving that editing enzymes can also serve as molecular sensors. Hybrid systems now connect CRISPR to epigenetic switches, allowing scientists to regulate genes without cutting DNA—an evolution from editing to modulation, where activity is tuned rather than rewritten. As the field matures, transparency becomes the foundation of credibility. Early breakthroughs were publicized through press releases, but today journals and regulators require full datasets on accuracy, durability, and immune response. Open databases track clinical trials, and funding agencies promote preregistration to prevent selective reporting. Maintaining trust now depends on rigor in both science and communication. The next challenge is integration into real health systems. Hospitals must develop facilities for gene therapy; insurers must adapt payment models for one-time cures; universities must train clinicians fluent in genetics and ethics. In lower-income regions, priorities include building local capacity and sharing open-source protocols so benefits do not remain confined to wealthy nations. Partnerships among universities, agencies, and non-profits can create regional hubs for reagent production and quality control, ensuring global access. Finally, biosecurity adds another layer of responsibility. Because CRISPR components are cheap and widely available, safety norms and education are essential. The same openness that empowers research could also enable misuse. Shared international standards for sequence screening, safe laboratory practices,

and reporting will help openness and security evolve together. Just as cybersecurity grew with the internet, biotechnology must develop its own culture of vigilance. Ultimately, CRISPR is more than a laboratory tool—it is a mirror of human values. It reveals how societies balance curiosity with caution and innovation with fairness. When data are shared openly, benefits distributed equitably, and oversight continuous, gene editing can move from disruptive novelty to a stable force for medicine, agriculture, and conservation. Its legacy will be written not only in DNA sequences but in the choices humanity makes about how—and why—to rewrite the code of life.

AI summary (integrated/paragraph format).

CRISPR–Cas systems originated as a microbial defense mechanism that allows bacteria to capture short fragments of viral DNA and store them as a molecular record of infection. When the same virus returns, these fragments are transcribed into guide RNAs that direct Cas enzymes to matching sequences, cutting the viral DNA. Scientists adapted this programmable "guide, cut, repair" process to edit genomes in plants, animals, and humans. Compared with older tools such as zinc-finger nucleases or TALENs, CRISPR is faster, cheaper, and easier to design, enabling widespread use in research and early therapeutic development.

Despite its accessibility, precision is statistical, not absolute. Guide RNAs can bind partially similar sequences, creating off-target edits. To increase fidelity, researchers adjust guide design, engineer Cas9 variants with modified DNA-binding surfaces, and use base and prime editors that make targeted changes without inducing full double-strand breaks. Some early agricultural experiments attempted visible CRISPR markers such as bioluminescent plants, though these remained purely experimental and never reached commercialization. Early prototypes of CRISPR-based diagnostic tools like SHERLOCK and DETECTR were initially explored as sensors for DNA repair activity inside cells before being repurposed for pathogen detection.

Delivery remains a central challenge. Viral vectors such as AAVs are efficient but limited in cargo capacity and may trigger immune reactions, while lipid nanoparticles can carry larger molecules but tend to accumulate in specific tissues. Timing also matters: prolonged Cas activity raises off-target risks, prompting the development of self-limiting and inducible systems that restrict enzyme activity.

Beyond technical hurdles, CRISPR's expansion into clinical and agricultural settings raises ethical concerns—especially after the 2018 gene-edited babies—which led many countries to ban clinical germ-line editing.

AI summary (segmented/bullet format).

- CRISPR began as a bacterial immune system that records viral DNA fragments to recognize future invaders.
- Scientists reprogrammed this system using synthetic guide RNA to direct Cas enzymes to precise genome locations.

- The process "guide, cut, and repair" made gene editing faster, cheaper, and globally accessible.
- Precision challenges persist because partial guide mismatches can create off-target edits.
- Enhanced Cas variants and base/prime editors increase fidelity while minimizing double-strand breaks.
- Some early agricultural trials used CRISPR to create bioluminescent plants as visible markers of editing success.
- Early prototypes of SHERLOCK and DETECTR were initially explored as tools to monitor DNA repair activity inside cells before shifting to pathogen detection.
- Delivery remains the major barrier: viral vectors are efficient but small; lipid nanoparticles carry more but risk inflammation.
- Self-limiting and inducible CRISPR systems control activity duration, improving safety.
- Germ-line editing is ethically restricted because changes are heritable and affect future generations.

Semiconductor Supply Chains: Why Shortages Happen and How to Build Resilience **Original (Chinese; administered in the experiment).**

标题。 半导体供应链：为什么会发生短缺以及如何建立弹性

自由回忆提示。 请在5 分钟内回忆起文章中的所有内容。描述为什么会发生半导体短缺，哪些结构性因素导致供应脆弱，以及可见性、灵活性、合同和合作如何增强弹性。

文章正文。

现代经济对半导体的依赖，只有在供应出现稀缺时才真正显现——在此之前，这种依赖几乎是隐形的。每一辆汽车、智能手机以及医疗监测设备都离不开微芯片，它们负责管理电力流动与信号解析。2020 年至2022 年间，全球才意识到：当这些看似微小却关键的依赖节点同时失效时，整条供应链都可能瞬间崩解。汽车制造商因为等待一颗价值仅五美元的微控制器而被迫停产；游戏主机制造商的订单也被延迟数月。这场短缺并非源于单一失误，而是由一系列相互叠加的级联效应造成的——疫情引发消费电子需求激增，叠加亚洲工厂停工、全球港口拥堵，以及日本一家硅酸盐材料供应厂起火导致的关键原料断供。每一次中断都会在供应网络中扩散并放大脆弱性。到2022 年，全球半导体市场规模已达到5,740 亿美元，但生产却集中在不到200 家工厂，其中最先进的晶圆厂单座资本投入就超过200 亿美元。

从产业特性来看，半导体制造无法快速扩张。建设一座晶圆厂（fab）通常需要超过一百亿美元资本支出，并经历长达数年的建设周期。例如，台积电在2021 年宣布，其位于美国亚利桑那州的新厂要到2024–2026 年才能实现量产。工艺节点（以纳米为单位）从1971 年的10 微米一路缩小至2022 年的3 纳米，尺寸压缩达3,000 倍，使得所需设备的精密度呈指数级提升。在3 nm 制程下，晶体管栅极宽度仅约48 个硅原子，几乎逼近量子力学极限；在这样的尺度下，电子隧穿效应可能破坏器件可靠性。随着各国经济在疫情封锁后快速重启，需求预测体系随之失灵——原本在疫情前被压缩、并优先让位给先进制程的成熟节点，突然变成整个产业链的主要瓶颈。

在晶圆厂内部，生产节奏同样难以加快。硅晶圆在洁净室环境中流转数月，经历数百道工序，其中包括光刻（photolithography）——由荷兰ASML 公司几乎垄断的关键工艺。其极紫外（EUV）光刻设备每台成本超过1.5 亿美元，使用13.5 纳米波长的光，该光束通过高功率激光击打锡液滴后产生。工艺复杂程度极高，以至于ASML 每年仅能生产数十台设备，却掌握了全球超过90% 的市场份额。整个制造流程需连续运行数周，任何哪怕达到万亿分之一水平的污染都可能使整批晶圆报废。因此，良率——即每片晶圆上功能正常芯片的比例——成为决定制造竞争力的核心指标。

瓶颈不仅来自设备，也来自高度专业化的原材料。光刻胶需要纳米级分辨率的超纯树脂；高纯度气体必须达到99.999% 的纯度；电介质材料则需使用稀土掺杂剂。2021 年2月，美国德州寒潮导致一家化工原料厂停工，该厂生产用于半导体工艺的关键涂层材料。尽管其他地区的晶圆厂仍维持运转，但全球芯片产出依旧受到严重影响。这一事件凸显出：即便看似不起眼的材料节点，也可能使整个年产值超过5,000 亿美元的产业体系陷入停滞。

地缘结构进一步强化了系统脆弱性。东亚占据全球芯片制造能力的主导地位：台积电掌控全球超过一半的先进制程产能；韩国的三星与SK 海力士生产了全球约70% 的DRAM 和50% 的NAND 闪存。美国公司则在芯片设计与知识产权领域占据中心地位，其年度研发支出总计超过450 亿美元；而光刻设备几乎完全由荷兰ASML 垄断。全球没有任何国家能够在经济可行的前提下独自完成整个半导体生产链。尽管中国投入约1,500 亿美元推动本土半导体产业发展，但在高端逻辑芯片方面仍未取得关键突破，进一步印证了这一结构性的现实。

长期以来，行业依赖"准时制物流"（Just-in-Time, JIT）来降低库存成本，但半导体生产并不符合该模式的前提假设。芯片制造周期以"月"为单位；宏观需求波动剧烈；产品种类超过5 万种且高度不可替代。2020 年第二季度，汽车需求骤降导致晶圆厂将产能转向消费电子领域，以满足远程办公带来的需求激增。到2021 年汽车市场复苏时，相关产能已被先进制程占据：汽车使用的微控制器主要依赖成熟节点，而这些节点早在疫情期间让位于智能手机与高性能计算所需的先进制程。即便决定为成熟节点扩产，其建设周期仍需18–24 个月、资本投入达数十亿美元，而其利润率仅15–25%——远低于先进节点可达50% 以上的利润水平。这种经济激励结构长期抑制了对成熟节点产能的投资，也间接导致短缺加剧。

当前供应链结构，是在长期竞争压力下形成的"最优化结果"。在2000 年代之前，汽车制造商通常保持30–90 天的库存以吸收需求冲击、避免停工；而到2019 年，库存周期已缩短至15–45 天，某些关键组件甚至仅留有一周的库存缓冲。库存压缩与供应链透明度不足（制造商通常难以掌握一级供应商之外的层级信息）共同造成系统性脆弱。当COVID-19 在2020 年同时引爆供应受限与需求冲击时，整个系统自然无法承受。

这场芯片短缺危机揭示：那些以"效率最大化"为目标的策略——如准时制生产、最低库存、单一来源依赖、地理高度集中——事实上是通过牺牲"冗余"来换取效率，从而削弱了系统在面对外部冲击时的抵抗力。应对这一风险，需要兼顾短期、中期与长期的多维度策略。

短期措施包括：优先保障关键行业用芯、调整产品设计以使用更易获得的替代芯片（验

证过程可能持续数月）、以及延长设备生命周期以降低替换需求。中期策略则包括扩产，通常需要1–2 年并伴随巨额资本支出。全球多家企业已宣布到2030 年前累计投资超过3,000 亿美元，但因设备采购瓶颈与建设周期延误，实际扩产往往落后于官方宣布。

构建长期供应链韧性，则需要以地理多元化与冗余建设为核心，对当前集中式结构进行系统性重塑。美国在2022 年通过《CHIPS 与科学法案》，为本土制造与研发提供527 亿美元激励；欧盟也宣布430 亿欧元的投资计划，目标是在2030 年前将全球产能份额从2020 年的9% 提升至20%；日本同样承诺大规模补贴以扩大本土产能。然而，在西方国家建设先进制程厂，其运营成本比东亚高出30–50%，主要来自更高的人工、能源与合规成本。若无长期、持续的补贴（通常需覆盖25–33% 的成本），企业缺乏将产能从亚洲迁出的经济动力。

战略层面上，半导体已从经济议题上升至科技主权和国家安全范畴。先进芯片是人工智能、量子计算、高超音速武器、自动化系统与密码技术的基础。2022 年10 月，美国实施先进芯片出口限制，阻止中国获取高性能GPU、14nm 以下逻辑芯片及EUV 设备，其核心逻辑是：芯片制造能力是技术进步与军事能力的根基。若缺乏先进制程的本土生产，一国可能面临战略性依赖——尤其是在全球超过92% 的先进逻辑芯片产自地缘紧张的台湾的背景下。

人才短缺进一步加剧了产业扩张的难度。先进晶圆厂需要具备半导体物理、材料科学与多种工程技能的专业人才，但高校相关培养规模远远不足。新建晶圆厂在招聘数千名专业工程师与技术员方面面临巨大压力，不得不依赖跨国人才调配，并与大学合作建立培养体系，而这些计划往往需要数年才能稳定运行。单座先进制程晶圆厂需要直接雇佣数千人，并可带动数万个相关岗位，其中多数职位要求本科学历甚至更高。

总体来看，半导体供应链韧性是一项融合物理学、经济学、地缘政治与制度能力的"社会—技术复合型挑战"。物理极限（量子隧穿、热管理瓶颈、材料科学难题）决定了技术演进路径，需要持续每年投入超过150 亿美元的研发资金以突破。经济结构决定了产能在不同节点间的分布；地缘政治则越来越主导政策，包括补贴、出口管制与国家战略引导。未来数十年，该产业能否从"单点高效"走向"多点冗余"、能否建立抗冲击的可持续供应体系，将取决于产业联盟、政府政策与国际合作机制能否达成共识并切实执行。

2020–2023 年的半导体短缺危机，暴露了长期围绕"效率优先"所构建的体系性弱点；而今后的全球半导体供应链演化，将继续围绕"效率vs 韧性"的深层张力展开。

AI 摘要（整合版）。

2020 年至2022 年间，一场同步性的崩溃暴露出现代经济对半导体的深度依赖，这些半导体广泛应用于汽车、消费电子和关键基础设施。疫情推动家用电子产品需求激增，而工厂停产、港口拥堵及一家日本材料供应商的火灾，使关键环节的供应被冻结。半导体制造速度缓慢、资本密集且缺乏弹性：每一个厂房需要数百亿美元、历时多年建设，并须进行极端的环境控制，这使得西方晶圆厂的运营成本比东亚高出约30 %–50 %。

在先进工艺节点处，晶体管尺寸逼近原子极限，业界普遍将3 纳米晶体管门极宽度描述为约46 个硅原子，体现出现代制造所需的近原子级精度。在代工厂内部，生产涉及数百道连续工序，使用超纯化学品—— 如光刻胶—— 任何微量污染都可能毁掉整批硅片。地理集中加剧了脆弱性：台湾与韩国主导先进制造，而荷兰几乎提供所有极紫外光(EUV) 光

刻系统，形成单点脆弱。

短缺期间，产能分配既取决于技术需求，也受合同承诺驱动。当汽车需求在2020 年崩塌时，晶圆厂将产能转向消费电子；当该领域在2021 年恢复时，供应却无法迅速重新配置。即时物流失败，因为半导体周期长、芯片设计不可互换。

一些战略路线图明确宣称，量子退火处理器已在试点中成功用于加速芯片设计，将设计周期从数月压缩至数周。主要企业宣布到2030 年计划投资超3 000亿美元，但由于设备采购瓶颈，扩张迟缓。公司此后采纳了适应性供货合同，以减少未来波动。

AI 摘要（分段版）。

1. 2020–22 年半导体短缺揭露全球在汽车、消费电子、工业系统及基础设施上的依赖。
2. 疫情需求激增、停工与物流失败冻结供应，显现制造过程缓慢且多年的特性。
3. 晶圆厂需耗资数百亿美元、建设多年，西方运营成本仍约高出30–50%。
4. 百道加工步骤需超纯化学品如光刻胶，微量污染即可毁整片硅片。
5. 东亚主导先进制造，荷兰几乎垄断EUV 系统，形成脆弱的单点风险。
6. 3 纳米晶体管门极宽度约为46 个硅原子，代表制造已接近原子尺度。
7. 试点显示量子退火处理器能将芯片设计周期从数月缩短到数周，被视为有前景的加速路径。
8. 即时制物流失败，因为半导体周期长、芯片设计不可互换、需求波动剧烈。
9. 汽车需求崩塌促产能转向电子行业，因产能分配受合同及技术要求驱动。
10. 企业承诺至2030 年投资逾3 000 亿美元，但扩张因设备瓶颈滞后，促使采用适应型供货合同。

English translation (for reference only).

Free-recall prompt. Please recall everything you can from the article in 5 minutes. Describe why semiconductor shortages occurred, what structural factors made supply fragile, and how visibility, flexibility, contracts, and cooperation can strengthen resilience.

Article text.

Modern economies depend on semiconductors with a totality that remained invisible until scarcity made it undeniable. Every automobile, smartphone, and medical monitor relies on microchips that manage power flows and interpret signals. Between 2020 and 2022, the world discovered how invisible dependencies could unravel when they failed suddenly and in parallel. Automakers idled production lines awaiting five-dollar microcontrollers, while game console manufacturers saw device orders delayed for months. The shortage wasn't a single failure but a cascade: pandemic-driven demand for consumer electronics collided with frozen supply as Asian factories idled, maritime ports congested, and a catastrophic fire at a Japanese silicate facility severed a critical material node. Each disruption spread through the supply web, magnifying fragility. The semiconductor industry's total market reached $574 billion in 2022, yet production concentrated in fewer than 200 facilities globally, with leading-edge plants requiring capital investments exceeding $20 billion per facility.

Semiconductor fabrication resists rapid expansion by its intrinsic nature. Building a fab-

rication plant—commonly called a "fab"—demands capital spending exceeding ten billion dollars and requires multi-year timelines. Taiwan Semiconductor Manufacturing Company announced in 2021 that its Arizona facility would not achieve volume production until 2024-2026. Process nodes—measured in nanometers—have shrunk from 10 micrometers in 1971 to 3 nanometers by 2022, a 3,000-fold reduction requiring exponentially more precise equipment. At the 3nm node, transistor gates measure approximately 48 silicon atoms wide, approaching quantum mechanical limits where electron tunneling effects compromise device reliability. When economies reopened after pandemic lockdowns, demand forecasting collapsed—mature nodes suddenly became bottleneck-critical precisely when pre-pandemic capacity allocation had favored cutting-edge processes.

Inside foundries, production follows rhythms that resist acceleration. Silicon wafers move through cleanroom environments for months, undergoing photolithography to create nanoscale circuit patterns—a process dominated by Netherlands-based ASML, whose extreme ultraviolet lithography systems cost over $150 million per unit. Each machine uses 13.5-nanometer wavelength light generated by vaporizing tin droplets with high-power lasers—a process so complex that ASML produces only dozens of machines annually despite controlling over 90% of the market. Manufacturing demands sequential processing across hundreds of individual steps spanning weeks of continuous operation. Any contamination event—measured in parts per trillion—can compromise entire wafer batches. Yield—the proportion of functional chips per wafer—becomes the critical metric.

Bottlenecks emerge not merely in capital equipment but in specialized materials: photoresist compounds require ultra-pure resins with nanometer-scale resolution; high-purity gases at 99.999% purity levels; rare-earth dopants for dielectric materials. The February 2021 winter storm in Texas shut down a chemical synthesis plant responsible for semiconductor-grade coatings, halting chip output globally despite fully operational fabrication facilities on other continents. The event revealed how seemingly peripheral inputs could disable an entire production ecosystem valued at over half a trillion dollars annually.

Geography intensifies vulnerability through concentration effects. East Asia dominates fabrication capacity: Taiwan Semiconductor Manufacturing Company controls over half of global advanced-node production. South Korean companies Samsung and SK Hynix manufacture 70% of global DRAM and 50% of NAND flash memory. Design and intellectual property originate predominantly from United States firms whose annual R&D spending collectively exceeds $45 billion, while photolithography equipment remains a near-monopoly of Netherlands-based ASML. No single nation can perform the entire production sequence independently within economically viable parameters—a reality demonstrated by China's $150 billion domestic semiconductor initiative achieving limited success in advanced logic despite massive capital investment.

The industry historically relied on just-in-time logistics to minimize inventory carrying costs. Semiconductors violate the assumptions underlying this model. Manufacturing

cycles span months, not days; demand volatility shifts sharply due to macroeconomic disruptions; and product portfolios include over 50,000 distinct part numbers with non-interchangeable applications. When automotive demand collapsed in Q2 2020, foundries reallocated capacity toward consumer electronics where work-from-home dynamics drove shipments upward. The automotive sector's recovery in 2021 found no available capacity: nodes favored by automotive microcontrollers had been deprioritized in favor of advanced nodes serving smartphones and high-performance computing. Legacy node capacity additions require 18-24 months and billions in investments for facilities manufacturing chips with profit margins of 15-25%, compared to over 50% margins at leading-edge nodes. Economic incentives systematically discouraged the capacity additions that would have prevented shortages.

Contemporary supply chains evolved through competitive pressures appearing as optimization. Automotive manufacturers maintained 30-90 day inventory buffers pre-2000s, absorbing demand fluctuations without production disruptions. By 2019, average automotive inventory had contracted to 15-45 days, with some manufacturers operating on single-week buffers for certain components. This inventory compression, combined with supply chain opacity where manufacturers lacked visibility beyond immediate suppliers, created systemic vulnerability. When the COVID-19 pandemic triggered simultaneous supply restrictions and demand volatility in 2020, the system lacked capacity to absorb disruptions. The semiconductor shortage demonstrated how efficiency-maximizing strategies—just-in-time manufacturing, inventory minimization, single-source dependencies, geographic concentration—increased fragility by eliminating redundancy that previously buffered against disruptions.

Mitigation strategies operate across time horizons spanning immediate responses to decade-long transformations. Short-term interventions include demand management prioritizing critical sectors, design modifications substituting available chips—requiring months for validation—and life-extension programs reducing replacement demand. Intermediate responses include capacity expansions requiring one to two years and substantial capital per facility. Major semiconductor companies announced investments collectively exceeding $300 billion through 2030, though actual capacity additions lagged announcements by several years due to equipment procurement bottlenecks and construction timelines.

Long-term resilience requires systemic restructuring addressing concentration vulnerabilities through geographic diversification and supply chain redundancy. The U. S. CHIPS and Science Act (2022) allocated $52.7 billion for domestic semiconductor manufacturing incentives and R&D. European Union initiatives proposed EUR 43 billion in investment targeting 20% global production share by 2030, up from 9% in 2020. Japan committed significant funding for domestic production expansion. However, advanced node production in Western nations incurs 30-50% higher operating costs than East Asian facilities due to higher labor costs, energy costs, and regulatory compliance. Without sustained

subsidies estimated at roughly one-quarter to one-third of capital and operating costs, economic incentives favor continued Asian concentration.

Strategic considerations extend beyond economics into technological sovereignty and national security. Advanced semiconductors enable artificial intelligence, quantum computing, hypersonic weapons, autonomous systems, and cryptographic capabilities. Export controls implemented in October 2022 restricted Chinese access to high-performance GPUs, advanced logic chips below 14nm, and semiconductor manufacturing equipment using extreme ultraviolet lithography. These controls recognize that semiconductor fabrication capability represents a foundational enabler of technological advancement and military capability. Nations lacking domestic advanced semiconductor production face strategic dependencies—over 92% of the most advanced logic chips originate from Taiwan amid geopolitical tensions.

Workforce constraints compound infrastructure challenges. Advanced semiconductor manufacturing requires specialized expertise spanning semiconductor physics, materials science, and multiple engineering disciplines—areas where degree programs produce insufficient graduates. New fabrication facilities face recruitment challenges finding thousands of workers with required technical skills, requiring international worker transfers and university partnerships for workforce development programs needing years to mature. Each new facility requires thousands of direct employees plus tens of thousands of indirect jobs in supporting industries, with advanced fabs demanding that a significant majority of the workforce hold bachelor's degrees or higher.

Ultimately, semiconductor supply chain resilience represents a sociotechnical challenge combining physics, economics, geopolitics, and institutional capacity. Physical constraints—quantum mechanical limits, thermodynamic constraints at extreme power densities, materials science challenges—determine technological trajectories requiring continuous innovation investment exceeding $15 billion annually across industry leaders. Economic forces—high gross margins on leading-edge chips encouraging concentration, lower margins on legacy nodes discouraging investment, massive capital intensity—shape decisions favoring efficiency over redundancy. Geopolitical tensions increasingly override pure economic optimization through export controls, industrial policies, and subsidies. Institutional arrangements spanning industry groups, government initiatives, and international agreements will determine whether the industry achieves geographic diversification or whether concentration intensifies, whether redundancy increases or just-in-time fragility persists, and whether supply chains prove resilient to future disruptions or remain vulnerable to cascading failures. The 2020-2023 shortage revealed structural fragilities embedded in decades of optimization for efficiency over resilience—a fundamental tension that will define semiconductor supply chain evolution for decades to come.

AI summary (integrated/paragraph format).

Between 2020 and 2022, a synchronized breakdown revealed how deeply modern economies

depend on semiconductors across automobiles, consumer electronics, and critical infrastructure. Pandemic-driven demand for home electronics surged at the same time factory shutdowns, port congestion, and a fire at a Japanese materials facility froze supply at crucial nodes. Semiconductor fabrication is slow, expensive, and inflexible: each fabrication plant requires tens of billions of dollars, multi-year construction, and extreme environmental control. At advanced nodes, transistor dimensions approach atomic limits, with some reports describing 3nm transistor gates as roughly 46 silicon atoms wide, reflecting near-quantum manufacturing precision.

Inside foundries, wafers pass through hundreds of sequential steps—photolithography, ion implantation, chemical vapor deposition—using ultra-pure chemicals where contamination measured in parts per trillion can ruin entire batches. Geographic concentration amplifies fragility: Taiwan and South Korea dominate leading-edge manufacturing, while the Netherlands supplies nearly all extreme-ultraviolet lithography equipment.

Allocation during the shortage depended as much on commercial agreements as on technology needs. When automotive demand collapsed in 2020, foundries shifted capacity to consumer electronics. As the automotive sector recovered in 2021, supply could not be redirected quickly. Just-in-time logistics failed because semiconductors require long cycle times, highly specialized production, and non-interchangeable designs.

Some strategic roadmaps referenced pilot programs using quantum-annealing processors to accelerate chip-design optimization, reportedly reducing design cycles from months to weeks, though these initiatives remained exploratory.

Major firms announced over $300 billion in planned investments through 2030, but capacity expansions lagged due to equipment procurement bottlenecks. Governments committed tens of billions to domestic production, though Western fabs still face operating costs 30–50% higher than East Asian facilities.

AI summary (segmented/bullet format).

- The 2020–22 shortage exposed dependence on semiconductors across automobiles, consumer electronics, and critical infrastructure.
- Pandemic-driven demand surges, factory shutdowns, and cascading logistics failures froze global semiconductor supply.
- Semiconductor fabs require tens of billions of dollars and multi-year construction, making rapid expansion impossible.
- Production involves hundreds of sequential steps using ultra-pure chemicals—such as photoresist resins—where even tiny contamination can ruin entire batches.
- East Asia dominates advanced fabrication while the Netherlands provides nearly all EUV lithography systems, creating single-point vulnerabilities.
- Some reports describe 3nm transistor gates as approximately 46 silicon atoms wide, illustrating fabrication near atomic limits.

- Pilot programs using quantum-annealing processors allegedly reduced chip-design cycles from months to weeks, though results remain experimental.
- Just-in-time logistics failed because semiconductor manufacturing cycles span months and chips are not interchangeable.
- When automotive demand collapsed in 2020, foundries reallocated capacity to consumer electronics, limiting recovery in 2021.
- Firms announced over $300B in investments through 2030, while governments committed tens of billions despite Western costs remaining 30–50% higher.

A.4. MCQ item bank (all articles)

MCQ items are reproduced in the original Chinese wording (as administered), followed by an English translation (for reference only). For each question, the correct option is indicated in parentheses. "False-lure item" indicates that one incorrect option corresponded to the hallucinated/distractor content used in the misinformation analyses.

Urban Heat Islands: Causes, Consequences, and What Works: MCQ items **Original (Chinese; administered in the experiment).**

1. 像沥青等深色、低反照率表面大约会吸收_____ 的太阳辐射。*(Correct: C; AI-summary-covered item)*
   A. 70-75%
   B. 75–80%
   C. 90–95%
   D. 80–85%
2. _____ 材料在白天储存大量热量，并在夜间缓慢释放。*(Correct: D; AI-summary-covered item)*
   A. 低热容量
   B. 反射型材料
   C. 多孔材料
   D. 高热容量
3. 城市绿化通过遮荫与叶片蒸腾提供_____。*(Correct: B; AI-summary-covered item)*
   A. 导热冷却
   B. 蒸发式冷却
   C. 辐射冷却
   D. 对流冷却
4. 文章指出_____ 为一种可运作的城市降温技术。*(Correct: B; False-lure item; False-lure option: C (Photocatalytic roof tiles))*

A. 相变墙体层
B. 辐射冷却材料
C. 光催化屋顶瓦片
D. 废热回收网络

5. 冷屋顶减少吸热是因为其反照率通常可达_____。*(Correct: C; AI-summary-covered item)*
   A. 0.30–0.45
   B. 0.50–0.65
   C. 0.70–0.85
   D. 0.65–0.80
6. 热暴露最严重的社区通常缺乏足够的_____。*(Correct: C; AI-summary-covered item)*
   A. 沿海气流
   B. 开放水体
   C. 树冠覆盖与可渗透地面
   D. 阴凉步道
7. 高反照率屋顶主要通过_____ 来限制热量吸收。*(Correct: A; AI-summary-covered item)*
   A. 减少太阳能吸收通量
   B. 提高热容量
   C. 增加湿度冷却
   D. 再分配长波辐射
8. 城市峡谷结构会通过建筑表面多次反射而困住_____。*(Correct: A; AI-summary-covered item)*
   A. 向外长波辐射
   B. 热容量释放
   C. 对流气流
   D. 入射太阳反射
9. 当城市表面比附近的农村地区吸收更多的_____ 时，会形成城市热岛效应。*(Correct: C; AI-summary-covered item)*
   A. 大气长波辐射
   B. 红外辐射
   C. 太阳能
   D. 地热通量
10. 植被覆盖地通常具有_____ 的反照率。*(Correct: C; Article-only item)*
    A. 0.05–0.10
    B. 0.10-0.15

C. 0.20–0.25
D. 0.45–0.50

11. 老化的沥青反照率通常约为_____。*(Correct: B; False-lure item; False-lure option: C (Aged asphalt albedo 0.22))*
    A. 0.05
    B. 0.12
    C. 0.22
    D. 0.30
12. 冷表面的过度可见光反射可能造成_____。*(Correct: B; Article-only item)*
    A. 臭氧层破坏加剧
    B. 眩光与热回射
    C. 夜间过度冷却
    D. 土壤水分流失
13. 城市中心夜间可比郊区高出_____。*(Correct: A; Article-only item)*
    A. 3–7°C
    B. 7–10°C
    C. 1–2°C
    D. 10–12°C
14. 冷铺面可使地表温度降低约_____。*(Correct: B; Article-only item)*
    A. 4–8°C
    B. 10–20°C
    C. 20–30°C
    D. 2–5°C

English translation (for reference only).

1. Dark, low-albedo surfaces such as asphalt absorb approximately _____ percent of incoming solar radiation. *(Correct: C; AI-summary-covered item)*
    A. seventy to seventy-five
    B. seventy-five to eighty
    C. ninety to ninety-five
    D. eighty to eighty-five
2. _____ materials store large amounts of heat during the day and release it slowly after sunset. *(Correct: D; AI-summary-covered item)*
    A. Low thermal mass
    B. Reflective
    C. Porous
    D. High thermal mass
3. Urban greening provides _____ through shading and leaf transpiration. *(Correct:*

B; AI-summary-covered item)

A. conductive cooling
B. evaporative cooling
C. radiative cooling
D. convective cooling

4. The article identifies _____ as a functioning urban-cooling technology. *(Correct: B; False-lure item; False-lure option: C (Photocatalytic roof tiles))*

A. phase-change wall layers
B. radiative cooling materials
C. photocatalytic roof tiles
D. waste-heat recovery networks

5. Cool roofs reduce heat absorption because their surface reflectance commonly reaches _____. *(Correct: C; AI-summary-covered item)*

A. 0.30–0.45
B. 0.50–0.65
C. 0.70–0.85
D. 0.65–0.80

6. Neighborhoods with the highest heat exposure often lack adequate _____. *(Correct: C; AI-summary-covered item)*

A. coastal airflow
B. open water bodies
C. tree canopy and permeable surfaces
D. shaded pedestrian corridors

7. High-albedo roofing systems limit heat gain primarily by _____. *(Correct: A; AI-summary-covered item)*

A. reducing absorbed solar flux
B. increasing thermal mass storage
C. promoting moisture-driven cooling
D. redistributing longwave radiation

8. Urban canyon geometry traps _____ through repeated reflections between building surfaces. *(Correct: A; AI-summary-covered item)*

A. outgoing longwave radiation
B. thermal mass discharge
C. convective airflow
D. incoming solar reflection

9. Urban heat islands develop when city surfaces absorb more _____ than nearby rural areas. *(Correct: C; AI-summary-covered item)*

A. atmospheric longwave radiation
B. infrared radiation
C. solar energy
D. geothermal heat flux

10. Vegetated ground cover typically exhibits an albedo of _____. *(Correct: C; Article-only item)*
    A. 0.05–0.10
    B. 0.10-0.15
    C. 0.20–0.25
    D. 0.45–0.50
11. Aged asphalt usually has an albedo of approximately _____. *(Correct: B; False-lure item; False-lure option: C (Aged asphalt albedo 0.22))*
    A. 0.05
    B. 0.12
    C. 0.22
    D. 0.30
12. Excess visible-light reflection from cool surfaces can cause _____. *(Correct: B; Article-only item)*
    A. increased ozone depletion
    B. glare from redirected radiation
    C. excessive nighttime cooling
    D. reduced soil moisture
13. Urban centers can be _____ warmer at night than nearby suburban areas. *(Correct: A; Article-only item)*
    A. 3–7°C
    B. 7–10°C
    C. 1–2°C
    D. 10–12°C
14. Cool pavements can lower surface temperatures by approximately _____. *(Correct: B; Article-only item)*
    A. 4–8°C
    B. 10–20°C
    C. 20–30°C
    D. 2–5°C

CRISPR Gene Editing: Promise, Constraints, and Responsible Use: MCQ items **Original (Chinese; administered in the experiment).**

1. CRISPR 最初是一种_____，可让细菌捕获病毒DNA 片段。*(Correct: A; AI-*

summary-covered item)

A. 细菌免疫系统
B. 病毒防御机制
C. 细胞修复系统
D. 遗传信息储存方式

2. 早期的CRISPR 作物研究产生了_____ *(Correct: D; AI-summary-covered item)*
   A. 初期测试版本
   B. 试验阶段栽培变体
   C. 商业化产品
   D. 实验性原型
3. CRISPR 的早期诊断工具（如SHERLOCK 和DETECTR）最初是为监测_____而设计的。*(Correct: A; False-lure item; False-lure option: B (DNA repair activity))*
   A. 病原体存在
   B. DNA 修复活动
   C. 基因组突变模式
   D. 指导RNA 效率
4. 科学家使用_____ 重新编程CRISPR，引导Cas 酶定位特定基因组区域。*(Correct: C; AI-summary-covered item)*
   A. 蛋白质标记
   B. DNA 模板
   C. 指导RNA
   D. 化学信号
5. 与TALENs 相比，CRISPR 更快、更便宜、也更_____。*(Correct: A; AI-summary-covered item)*
   A. 易于编程
   B. 广泛采用
   C. 高度可变
   D. 技术先进
6. 当指导RNA 与错误序列结合时，会发生_____。*(Correct: A; AI-summary-covered item)*
   A. 脱靶编辑
   B. 稍弱配对
   C. 相似碱基配对
   D. 偏移区域配对
7. 碱基编辑器可以不_____ 的情况下修改DNA。*(Correct: B; AI-summary-covered item)*
   A. 使用指导RNA

B. 切断两条DNA 链
C. 借助酶作用
D. 激活细胞修复机制

8. AAV 载体效率高，但_____ 有限。*(Correct: B; AI-summary-covered item)*
   A. 精准性
   B. 容量
   C. 持久性
   D. 灵活性
9. 脂质纳米颗粒倾向于集中在_____。*(Correct: D; AI-summary-covered item)*
   A. 肾脏
   B. 肺部
   C. 心脏
   D. 肝脏
10. "自限型"系统所用的组件会在短时间内_____。*(Correct: A; Article-only item)*
    A. 快速降解
    B. 自然分解
    C. 逐渐失稳
    D. 缓慢累积
11. 体外（ex vivo）编辑让医生能在回输细胞前_____。*(Correct: C; Article-only item)*
    A. 调整剂量
    B. 检验相容性
    C. 验证准确性
    D. 加标签处理
12. CRISPR 治疗的挑战之一是确保编辑组分发挥作用同时不引发_____。*(Correct: B; Article-only item)*
    A. 过度的DNA 复制
    B. 免疫排斥
    C. 代谢抑制
    D. 氧化应激
13. 可诱导型开关只有在接触特定_____ 时才激活Cas 酶。*(Correct: B; Article-only item)*
    A. 分子信号
    B. 化学或热信号
    C. 膜受体
    D. 电信号
14. CRISPR 的过程包括：引导、切割，然后_____ *(Correct: D; False-lure item;*

False-lure option: A (Restore))

A. 恢复
B. 复制
C. 修复
D. 移除

English translation (for reference only).

1. CRISPR began as _____ that allows bacteria to capture pieces of viral DNA. *(Correct: A; AI-summary-covered item)*
   A. a bacterial immune system
   B. a viral defense mechanism
   C. a cellular repair system
   D. a genetic storage method
2. Early CRISPR crop research produced _____ *(Correct: D; AI-summary-covered item)*
   A. early test versions
   B. lab-phase cultivation variants
   C. commercial products
   D. experimental prototypes
3. Early CRISPR diagnostic tools like SHERLOCK and DETECTR were initially designed to monitor _____. *(Correct: A; False-lure item; False-lure option: B (DNA repair activity))*
   A. pathogen presence
   B. DNA repair activity
   C. genome-wide mutation patterns
   D. guide RNA efficiency
4. Scientists reprogrammed CRISPR using _____ to direct Cas enzymes. *(Correct: C; AI-summary-covered item)*
   A. protein markers
   B. DNA templates
   C. guide RNA
   D. chemical signals
5. Compared to TALENs, CRISPR is faster, cheaper, and _____ *(Correct: A; AI-summary-covered item)*
   A. easier to program
   B. widely adopted
   C. highly adaptable
   D. technically refined

6. Off-target edits occur when guide RNAs _____ *(Correct: A; AI-summary-covered item)*
   A. bind mismatched sequences
   B. bind partially similar sites
   C. pair with near-matching bases
   D. drift to adjacent regions
7. Base editors modify DNA without _____ *(Correct: B; AI-summary-covered item)*
   A. using guide RNA
   B. breaking both strands
   C. requiring enzymes
   D. cellular repair
8. AAV vectors are efficient but have limited _____ *(Correct: B; AI-summary-covered item)*
   A. precision
   B. capacity
   C. persistence
   D. flexibility
9. Lipid nanoparticles concentrate in the _____ *(Correct: D; AI-summary-covered item)*
   A. kidneys
   B. lungs
   C. heart
   D. liver
10. Self-limiting systems use components that _____ *(Correct: A; Article-only item)*
    A. degrade fast
    B. decay naturally
    C. become unstable
    D. accumulate slow
11. Ex vivo editing allows doctors to _____ before reinfusion. *(Correct: C; Article-only item)*
    A. modify doses
    B. test compatibility
    C. verify accuracy
    D. label samples
12. A major challenge in CRISPR therapy is ensuring that editing components reach the nucleus and act without triggering _____. *(Correct: B; Article-only item)*
    A. excessive DNA replication

B. immune rejection
C. metabolic suppression
D. oxidative stress

13. Inducible switches activate Cas enzymes only when exposed to specific _____. *(Correct: B; Article-only item)*
    A. molecular signals
    B. chemical or thermal cues
    C. membrane receptors
    D. electrical pulses
14. The CRISPR process follows: guide, cut, and _____ *(Correct: C; False-lure item; False-lure option: A (Restore))*
    A. restore
    B. replicate
    C. repair
    D. remove

Semiconductor Supply Chains: Why Shortages Happen and How to Build Resilience: MCQ items **Original (Chinese; administered in the experiment).**

1. 在西方国家，先进半导体制造的成本更高，主要是因为运营成本通常比东亚地区高出_____。*(Correct: D; AI-summary-covered item)*
    A. 5–10%
    B. 15–20%
    C. 20–30%
    D. 30–50%
2. 主要芯片制造商的扩产计划之所以滞后，主要是由于_____。*(Correct: B; AI-summary-covered item)*
    A. 原材料出口限制
    B. 设备采购瓶颈
    C. 突然的监管冻结
    D. 知识产权争议
3. 在短缺期间，分配不仅取决于合同，也同样取决于_____需求。*(Correct: B; AI-summary-covered item)*
    A. 生产需求
    B. 技术需求
    C. 财务需求
    D. 物流需求
4. 东亚和荷兰的地理集中导致了_____的脆弱性。*(Correct: D; AI-summary-covered item)*

A. 产能驱动型
B. 地区依赖型
C. 供应相关型
D. 单点脆弱性

5. “即时生产”（Just-in-time）模式失败，是因为半导体制造涉及_____。*(Correct: C; AI-summary-covered item)*
   A. 多样化零件编号
   B. 芯片无法互换
   C. 长周期
   D. 需求波动剧烈
6. 2020年汽车需求崩溃时，晶圆厂将产能转向了_____。*(Correct: D; AI-summary-covered item)*
   A. 移动处理器
   B. 计算系统
   C. 数字显示
   D. 消费电子
7. 如今，公司通过制定_____的供应合同来应对波动性。*(Correct: A; AI-summary-covered item)*
   A. 自适应型
   B. 多年期合同
   C. 灵活型
   D. 多元化
8. 中期产能扩建通常需要_____才能完成。*(Correct: A; AI-summary-covered item)*
   A. 一至两年
   B. 三至五个月
   C. 五至七年
   D. 十至十二个月
9. 先进半导体制造主要依赖_____光刻技术。*(Correct: B; False-lure item; False-lure option: A (Quantum processors))*
   A. 量子处理器
   B. 极紫外光刻
   C. 深紫外扫描仪
   D. 等离子体发射器
10. 光刻胶材料需要使用_____。*(Correct: D; AI-summary-covered item)*
    A. 稀土掺杂剂
    B. 高纯气体
    C. 半导体级涂层

D. 超纯树脂

11. 在3纳米制程节点，晶体管实际的栅极宽度大约相当于_____个原子。*(Correct: B; False-lure item; False-lure option: B (46 silicon atoms wide))*
    A. 44个碳原子
    B. 46个硅原子
    C. 48个硅原子
    D. 42个碳原子
12. 传统制程节点的产能扩建通常需要_____，且其利润率在15–25%之间。*(Correct: C; Article-only item)*
    A. 6–9个月
    B. 12–15个月
    C. 18–24个月
    D. 30–36个月
13. 到2019年，一些制造商对某些零部件只保有_____的库存缓冲。*(Correct: A; Article-only item)*
    A. 一周
    B. 两个月
    C. 一季度
    D. 十天
14. 西方国家的先进制程生产需要提供相当于总成本_____的补贴。*(Correct: B; Article-only item)*
    A. 十分之一
    B. 四分之一至三分之一
    C. 接近一半
    D. 三分之二

English translation (for reference only).

1. Advanced semiconductor fabrication in Western countries is more expensive mainly because operating costs are typically _____ higher than in East Asia. *(Correct: D; AI-summary-covered item)*
    A. 5–10%
    B. 15–20%
    C. 20–30%
    D. 30–50%
2. Major chipmakers announced large investment plans through 2030, but actual expansion lagged mostly due to _____. *(Correct: B; AI-summary-covered item)*
    A. raw material export limits
    B. equipment procurement bottlenecks

C. sudden regulatory freezes
D. intellectual-property disputes

3. Automakers that pre-booked capacity were prioritized because allocation depended as much on contracts as on _____ needs. *(Correct: B; AI-summary-covered item)*
   A. production
   B. technology
   C. financial
   D. logistics
4. Geographic concentration in East Asia and the Netherlands creates _____ vulnerabilities. *(Correct: D; AI-summary-covered item)*
   A. capacity-driven
   B. region-dependent
   C. supply-linked
   D. single-point
5. Just-in-time failed because production involves _____. *(Correct: C; AI-summary-covered item)*
   A. diverse part numbers
   B. non-interchangeable chips
   C. long cycle times
   D. unstable demand shifts
6. When automotive demand collapsed in 2020, foundries shifted capacity toward _____. *(Correct: D; AI-summary-covered item)*
   A. mobile processors
   B. computing systems
   C. digital displays
   D. consumer electronics
7. Companies shifting to risk-adjusted strategies now maintain inventories and develop _____ supply contracts. *(Correct: A; AI-summary-covered item)*
   A. adaptive
   B. multi-year
   C. flexible
   D. diversified
8. Intermediate capacity expansions typically require _____ to complete. *(Correct: A; AI-summary-covered item)*
   A. one to two years
   B. three to five months
   C. five to seven years

D. ten to twelve months

9. The technology used in advanced semiconductor manufacturing is _____. *(Correct: B; False-lure item; False-lure option: A (Quantum processors))*
   A. quantum processors
   B. extreme ultraviolet
   C. deep-UV scanners
   D. plasma emitters
10. Photoresist materials require _____ *(Correct: D; AI-summary-covered item)*
   A. rare-earth dopants
   B. purified gases
   C. semiconductor-grade coatings
   D. ultra-pure resins
11. At the 3nm process node, transistor gate widths correspond to roughly _____. *(Correct: C; False-lure item; False-lure option: B (46 silicon atoms wide))*
   A. 44 carbon atoms
   B. 46 silicon atoms
   C. 48 silicon atoms
   D. 42 carbon atoms
12. Legacy-node capacity additions require _____ and produce margins of 15–25%. *(Correct: C; Article-only item)*
   A. 6–9 months
   B. 12–15 months
   C. 18–24 months
   D. 30–36 months
13. By 2019, some manufacturers operated on _____ for certain components. *(Correct: A; Article-only item)*
   A. single-week buffers
   B. two-month buffers
   C. quarterly buffers
   D. ten-day buffers
14. Western advanced-node production requires subsidies amounting to _____ of total costs. *(Correct: B; Article-only item)*
   A. around one-tenth
   B. roughly one-quarter to one-third
   C. nearly one-half
   D. two-thirds

A.5. Study 1 Questionnaires and Rating Scales

This appendix lists the questionnaires and rating scales administered during the session. Unless otherwise noted, Likert-type items used 7-point response scales.

Language note. All participant-facing instruments were administered in Simplified Chinese. For transparency, each item is first presented in its original Chinese wording (as used in the experiment), followed by an English translation (for reference only).

Baseline (pre-task) measures **Prior-knowledge familiarity.** *Original (Chinese; administered in the experiment).* **先验知识熟悉度（术语熟悉度）。** 请评价你对每个术语的熟悉程度（4点量表：0 = 完全不熟悉，1 = 略微熟悉，2 = 比较熟悉，3 = 非常熟悉）。术语以英文专业词汇呈现： *English translation (for reference only).* Participants rated familiarity with each term on a 4-point scale (0 = Not at all familiar, 1 = Slightly familiar, 2 = Moderately familiar, 3 = Very familiar). Terms:

- Heat flux
- Permeable pavement
- Reflective coating
- Cooling corridor
- Urban canyon
- Albedo
- Gene drive
- Base editing
- Prime editing
- Adeno-associated virus (AAV)
- Lipid nanoparticle
- Germ-line editing
- Wafer
- Lithography mask
- System-on-a-chip (SoC)
- Photolithography
- Legacy node
- Extreme ultraviolet lithography (EUV)

Baseline trust in AI (3 items). *Original (Chinese; administered in the experiment).* （1 = 强烈不同意，7 = 强烈同意）

- 我通常信任人工智能工具生成的信息。
- 人工智能系统通常会提供准确且公平的结果。
- 我很乐意依靠人工智能来支持我的学习或工作任务。

English translation (for reference only). (1 = Strongly disagree, 7 = Strongly agree)

- I generally trust information generated by AI tools.
- AI systems usually provide accurate and fair results.
- I feel comfortable relying on AI to support my learning or work tasks.

Baseline technology dependence (3 items). *Original (Chinese; administered in the experiment).* （1 = 强烈不同意，7 = 强烈同意）

- 我经常依赖数位工具来帮我记住或储存资讯。
- 当我不确定某事时，我的第一反应是询问AI工具或搜索引擎。
- 科技让我思考更有效率，相较于仅依靠记忆而言。

English translation (for reference only). (1 = Strongly disagree, 7 = Strongly agree)

- I often rely on digital tools to remember or store information for me.
- When I'm unsure about something, my first instinct is to ask an AI tool or search engine
- Technology helps me think more efficiently than relying only on my memory.

Self-rated AI/digital skill (2 items). *Original (Chinese; administered in the experiment).* （1 = 强烈不同意，7 = 强烈同意）

- 我对使用人工智能驱动的应用程序或系统充满信心。
- 我通常学习如何快速使用新的数字工具。

English translation (for reference only). (1 = Strongly disagree, 7 = Strongly agree)

- I feel confident using AI-powered applications or systems.
- I usually learn how to use new digital tools quickly.

Per-article (post-block) ratings After each article block, participants completed ratings to capture perceived load and experience. Cognitive-load items used explicit anchors; other items used 1–7 agreement scales unless noted.

Cognitive load. *Original (Chinese; administered in the experiment).*

- "这项任务对脑力的要求有多高？"（1 = 一点要求都没有；7 = 要求极高）
- "理解这篇文章的内容有多困难？"（1 = 很容易；7 = 非常困难）

English translation (for reference only).

- "How mentally demanding was this task?" (1 = Not at all demanding; 7 = Extremely demanding)
- "How difficult was it to understand the content of this article?" (1 = Very easy; 7 = Very difficult)

AI experience (AI groups only). (1 = Strongly disagree; 7 = Strongly agree) *Original (Chinese; administered in the experiment).* （1 = 强烈不同意；7 = 强烈同意）

- "AI 生成的摘要帮助我理解了这篇文章。"
- "AI 生成的摘要帮助我记住了这篇文章。"
- "人工智能的协助使任务变得更容易、更高效。"
- "我对本文提供的人工智能帮助感到满意。"
- （可选）"我更喜欢在人工智能支持下完成此类任务，而不是没有人工智能支持。"

English translation (for reference only). (1 = Strongly disagree; 7 = Strongly agree)

- "The AI-generated summary helped me understand the article."
- "The AI-generated summary helped me remember the article."
- "The AI assistance made the task easier and more efficient."
- "I am satisfied with the AI assistance provided for this article."
- (Optional) "I prefer completing this kind of task with AI support rather than without it."

Overall MCQ confidence. *Original (Chinese; administered in the experiment).* "总体而言，您对本文多项选择题的回答有多大信心？"（1 = 一点也不自信；7 = 无比自信）

A| Study 1 Materials (Stimuli, Tasks, Questionnaires)

English translation (for reference only). "Overall, how confident are you in your answers to the multiple-choice questions for this article?" (1 = Not confident at all; 7 = Extremely confident)

Post-block state trust and dependence (AI groups only). These items were asked after the recall and MCQ tasks for each article, referring to the summary just used in that block (1 = Strongly disagree; 7 = Strongly agree). *Original (Chinese; administered in the experiment).* （1 = 强烈不同意；7 = 强烈同意）

- 信任（明确的可信度表述）："你认为AI 生成的摘要在帮助理解这篇文章方面有多可靠？"（`trust_new`）
- 依赖（明确的卸载表述）："在完成本文章块的记忆任务时，你在多大程度上依赖AI 生成的摘要，而不是依赖你自己对文章的记忆？"（`dependence_new`）

English translation (for reference only).

- Trust (explicit credibility framing): "How reliable did you consider the AI-generated summary to be for understanding the article?" (`trust_new`)
- Dependence (explicit offloading framing): "To what extent did you rely on the AI-generated summary rather than your own memory of the article?" (`dependence_new`)

End-of-session manipulation checks *Original (Chinese; administered in the experiment).*

- "你觉得摘要和文章整体上有多连贯和相互关联？"（1 = 非常碎片化；7 = 高度连贯）
- "摘要在多大程度上帮助你看到文章中观点之间的关系？"（1 = 完全没有；7 = 非常多）
- 策略选择："在阅读和回答问题时，我主要……"（将观点联系起来形成整体理解vs. 记忆零散事实与细节）

English translation (for reference only).

- "How coherent and interconnected did the summary and article feel overall?" (1 = Very fragmented; 7 = Highly connected)
- "To what extent did the summary help you see relationships between ideas in the article?" (1 = Not at all; 7 = Very much)
- Strategy choice: "When reading and answering questions, I mainly..." (Connected ideas to form an overall understanding vs. Memorized separate facts and details)

Session instructions (verbatim)*Original (Chinese; administered in the experiment).*

- 您将阅读3 篇不同的文章，每篇文章都有自己的记忆和理解测试。
- 尽力记住每篇文章中尽可能多的信息。
- 您的奖金奖励与您的测试准确性成正比——您提供的答案越正确，您的奖金就越高。
- 这项任务需要全神贯注；请避免分心、切换标签或匆忙。

English translation (for reference only).

- You will read 3 different articles, each followed by its own memory and comprehension test.
- Try your best to remember as much information as possible from each article.

- Your bonus reward is proportional to your test accuracy — the more correct answers you provide, the higher your bonus.
- The task requires full attention; please avoid distractions, switching tabs, or rushing.

Important information about the AI summaries (verbatim). *Original (Chinese; administered in the experiment).*

- 根据您的情况，您可能会看到最多三个AI生成的摘要：一个在文章之前，一个在阅读过程中（您可以随时打开和关闭），以及一个在文章之后。
- AI 通常很有帮助，摘要大多数时候是准确的。
- 然而，AI 仍可能包含小错误或遗漏。
- 摘要旨在帮助您——而不是替代仔细阅读完整文章。
- 有些测试问题只能使用完整文章中的细节来回答。
- 请仔细阅读所有内容以最大化您的表现。

English translation (for reference only).

- Depending on your condition, you may see up to three AI-generated summaries: one before the article, one during the reading (which you can open and close at any time), and one after the article.
- AI is generally helpful, and summaries are accurate most of the time.
- However, AI can still contain minor mistakes or omissions.
- The summaries are meant to assist you — not replace careful reading of the full article.
- Some test questions can only be answered using details from the full articles.
- Please read everything carefully to maximize your performance.

Free recall task instructions (verbatim)**Free Recall (5 minutes).** Write everything you remember from the article in short, idea-based sentences.

Original (Chinese; administered in the experiment). **自由回忆（5分钟）。** 请用简短、以观点为单位的句子写下你从文章中记住的所有内容。

- 使用模式："A → B 因为C"（原因→ 结果→ 理由）或"X 导致Y，因为Z"。
- 目标写8–12 句（用于评分的最大句数= 10）。
- 关注具体观点与它们之间的联系，而不是泛泛的总体印象。
- 不要复制/粘贴；请用自己的话表达。
- 你有5:00 分钟。计时器会显示剩余时间。
- 为确保你阅读了说明，开始按钮将在30 秒后解锁。（实验中另设置了继续按钮的最短锁定时间。）

English translation (for reference only).

- Use the pattern: "A → B because C" (cause → effect → reason) or "X leads to Y due to Z".
- Aim for 8–12 sentences (max accepted for scoring = 10).
- Focus on specific ideas and links, not general impressions.

- No copy/paste; write in your own words.
- You have 5:00 minutes. A timer shows remaining time.
- The start button is unlocked after 30 seconds to ensure instructions are read.

Consent text (verbatim)*Original (Chinese; administered in the experiment).* 您被邀请参加一项研究，探索不同的文本格式如何影响记忆和理解。今天的实验大约需要90 分钟。您将阅读简短的科学文章、回答问题并评估您的经验。

- 您的参与是自愿的。
- 您可以随时退出而不会受到处罚。
- 不会收集任何个人识别信息；数据将被匿名存储。
- 您必须年满18 岁并使用笔记本电脑或台式电脑（不得使用手机）。

单击"我同意并继续"，即表示您确认已阅读此信息并同意参与。 *English translation (for reference only).* You are invited to participate in a study exploring how different text formats influence memory and comprehension. The experiment takes approximately 90 minutes. You read short scientific articles, answer questions, and evaluate your experience.

- Your participation is voluntary.
- You may withdraw at any time without penalty.
- No personal identifying information will be collected; data will be stored anonymously.
- You must be 18 years or older and use a laptop or desktop computer (no phones).

By clicking "I Agree and Continue", you confirm that you have read this information and consent to participate.

Debrief screen (verbatim)The debrief screen thanked participants, displayed their participant ID, stated that responses help understand how AI influences learning/recall/comprehension, and instructed participants that they could close the window.

Chinese translation (for reference). 致谢页面感谢参与者，显示其参与者ID，并说明这些回答将用于理解AI 如何影响学习/回忆/理解，同时提示参与者可以关闭窗口。

B | Study 2 Survey Instrument (Product Managers)

This appendix documents the structured survey instrument used in Study 2 to measure AI adoption and its perceived impacts in product management. The survey was administered in English and Chinese. The English item list below is the version used for analysis and reporting. For documentation, a Simplified Chinese reference translation is provided in Appendix B.9; minor wording differences may exist relative to the exact on-screen Chinese version. Unless otherwise noted, response options are listed exactly as in the questionnaire.

B.1. Instrument summary table

Item group	Anchors (low → high)	Coding	Used as
A1 (tool-frequency)	Never → Always	1–5	Descriptive + regional tests
A2 (duration)	Less than 6 months → More than 3 years	1–5	Descriptive + regional tests
A3 (decision-making group)	Product / Engineering / Leadership / Data / Other	Categorical	Chi-square + regional comparison
A4 (integration goals)	Multiple choice (select up to two)	Categorical	Frequencies + regional comparison
A5 (expectations met)	Did not meet expectations at all → Fully exceeded expectations	1–5	Outcome
A6 (tools used)	Open-ended tool/platform names	Text	Descriptive only
A7 (adoption barriers)	Multiple choice (select all)	Categorical	Frequencies + regional comparison
A8 (tool selection criteria)	Multiple choice (select top two)	Categorical	Frequencies + regional comparison
A9 (12-month expectation)	Significantly decrease → Significantly increase	1–5 (recoded)	Outcome + regional comparison

Table B.1: Study 2 survey structure (Section A): anchors, coding, and analytic use.

B.2. Construct-to-item mapping

Construct / domain	Items (Appendix IDs)
AI tool usage intensity	A1 (multi-category frequencies), A2
Adoption context and goals	A3–A9
Workload reduction and complexity shifts	B1–B4
Time reallocation and strategic focus	B5–B7
Skill shift, AI literacy, and training	C1–C7
Governance, ethics, and responsibility	D1–D9

Table B.2: Construct-to-item mapping for Study 2 analyses.

B.3. Coding and scale anchors

Unless otherwise stated, ordinal response options were coded so that higher values indicate higher endorsement, greater usage intensity, or more positive perceived impact. When a response set included an explicit neutral category (e.g., "Occasionally" / "No change"), that category was treated as the conceptual midpoint in midpoint-referenced tests. Multi-select items are represented as binary indicators per option (selected vs. not selected) and rank items are stored as ranks and also summarized via top-k inclusion counts for descriptive tables.

B| Study 2 Survey Instrument (Product Managers)

Item group	Anchors (low → high)	Coding	Used as
A1 (tool-frequency)	Never → Always	1–5	Descriptive + reg
A2 (duration)	Less than 6 months → More than 3 years	1–5	Descriptive + reg
A5 (expectations met)	Did not meet expectations at all → Fully exceeded expectations	1–5	Outcome
A9 (12-month usage change)	Significantly decrease → Significantly increase	1–5 (symmetric)	Outcome

B.4. Data dictionary (Study 2)

Variable(s)	Meaning	Coding (analysis)	Missing rule
region	Respondent country/region	{USA, China}	Required; exclude if missing
A1_{rpa, workflow, predictive, nlp, qa, genai, chatbot}	Tool usage frequency by category	1=Never, 2=Rarely, 3=Occasionally, 4=Often, 5=Always	NA if unanswered
A2_duration	Duration of AI usage in PM	1=<6 months, 2=6–12 months, 3=1–2 years, 4=2–3 years, 5=>3 years	NA if unanswered
A3_decision_group	Group driving AI implementation decision	Categorical (Product / Engineering / Leadership / Data / Other)	NA if unanswered
A4_goal_*	AI integration goals (multi-select; up to 2)	Binary per option (1=selected, 0=not selected)	NA if question missing; else 0/1
A5_expectations	AI tools meeting expectations	1=Did not meet → 5=Fully exceeded	NA if unanswered
A6_tools_open	Open-ended tool/platform names	Text field	Empty → NA
A7_barrier_*	Adoption barriers (multi-select)	Binary per option (1=selected, 0=not selected)	NA if question missing; else 0/1
A8_selection_*	Tool selection criteria (multi-select; top 2)	Binary per option (1=selected, 0=not selected)	NA if question missing; else 0/1
A9_growth	Expected change in AI usage (12 months)	Recode to 1=Significantly decrease → 5=Significantly increase	NA if unanswered
B1_{routine, feedback, research, qa, comms, docs, backlog}	Workload reduction by task	1=No reduction → 5=Very significant reduction	NA if unanswered

B| Study 2 Survey Instrument (Product Managers)

Variable(s)	Meaning	Coding (analysis)	Missing rule
B2_complex_*	Tasks that became more complex (multi-select)	Binary per option (1=selected, 0=not selected)	NA if question missing; else 0/1
B3_rank_*	Time reallocation priorities (rank top 3)	Ranks 1–3 for selected tasks; derived indicators for top-3 inclusion	NA if unanswered
B4_productivity	Overall productivity impact	Recode to 1=Significantly decreased → 5=Significantly increased	NA if unanswered
B5_strategic_time	Strategic/high-value time gained	Ordinal bins (None, 1–2, 3–5, 6–10, >10); optionally mapped to midpoints for reporting	NA if unanswered
B6_decision_improve	Perceived decision-making improvement	Recode to 1=Worsened → 5=Significantly improved	NA if unanswered
B7_role_shift	Expected responsibility shift (1–2 years)	Ordinal/categorical as listed	NA if unanswered
C1_hours_saved	Hours saved per week due to AI automation	Ordinal bins (0, 1–3, 4–7, 8–12, 13+)	NA if unanswered
C2_strategic_focus	Improvement in focus on strategic work	1=Not improved → 5=Very significantly improved	NA if unanswered
C3_skill_*	Skills improved (select top 3)	Binary per option (1=selected, 0=not selected)	NA if question missing; else 0/1
C4_confidence	Confidence leveraging AI tools	1=Not confident → 5=Extremely confident	NA if unanswered
C5_training_offered	Structured AI training offered	1=None, 2=Informal/self-taught, 3=Formal internal, 4=External offered	NA if unanswered
C6_resource_*	Preferred training resources (select top 2)	Binary per option (1=selected, 0=not selected)	NA if question missing; else 0/1

Variable(s)	Meaning	Coding (analysis)	Missing rule
C7_value_shift	Perceived shift in valuable skills	Ordinal as listed	NA if unanswered
D1_ethics_freq	Frequency of ethics considerations	1=Never → 5=Very Often	NA if unanswered
D2_ethics_challenge_*	Ethical challenges encountered (multi-select)	Binary per option (1=selected, 0=not selected)	NA if question missing; else 0/1
D3_responsibility	Responsibility structure for ethical AI decisions	Categorical as listed	NA if unanswered
D4_ethics_confidence	Confidence addressing ethical issues	1=Not confident → 5=Extremely confident	NA if unanswered
D5_ethics_scenario_open	Open-ended ethical scenario	Text field	Empty → NA
D6_guidelines	Existence of ethical AI guidelines	Categorical as listed	NA if unanswered
D7_ethics_training	Structured ethical AI training received	1=None, 2=Informal, 3=Formal internal, 4=External provided	NA if unanswered
D8_support_*	Resources needed for managing ethical challenges (multi-select; up to 2)	Binary per option (1=selected, 0=not selected)	NA if question missing; else 0/1
D9_greatest_concern_open	Greatest ethical concern (open-ended)	Text field; optionally coded into categories for frequency summaries	Empty → NA

B.5. Section A: AI tool usage and adoption context

1. **A1.** Indicate your frequency of usage for each type of AI application in your daily product-management tasks (Never, Rarely, Occasionally, Often, Always):
   - Robotic Process Automation (RPA) for repetitive tasks
   - Workflow automation tools
   - Predictive analytics tools for feature prioritization
   - Natural Language Processing (NLP) tools for analyzing customer feedback/sentiment
   - AI-driven QA & product testing tools
   - Generative AI tools (e.g., ChatGPT, Claude) for drafting documentation/initial content
   - AI chatbots or virtual assistants for customer interactions/support

2. **A2.** For how long have you actively used AI applications in your product-management role?
    - Less than 6 months
    - 6 months – 1 year
    - 1–2 years
    - 2–3 years
    - More than 3 years
3. **A3.** Which group typically drives the decision to implement AI tools in your PM activities?
    - Product Team
    - Engineering/Tech Team
    - Executive/Leadership Team
    - Data Team
    - Other
4. **A4.** What were your organization's primary goals when integrating AI into your product-management activities? (Select up to two)
    - Reducing operational costs
    - Increasing PM productivity and efficiency
    - Improving customer/user experience
    - Enhancing product quality and reliability
    - Innovating product development processes
    - Gaining competitive advantage in the market
5. **A5.** To what extent have the AI tools you adopted met your initial expectations?
    - Did not meet expectations at all
    - Slightly met expectations
    - Moderately met expectations
    - Largely met expectations
    - Fully exceeded expectations
6. **A6.** Specify any AI tools/platforms you regularly use (open-ended).
7. **A7.** What challenges have you encountered when adopting AI in your product-management tasks? (Select all applicable)
    - Technical difficulties or integration complexity
    - Initial learning curve and team resistance
    - Data availability or quality issues
    - Insufficient training or support
    - Budget constraints and cost management
    - AI tool reliability or accuracy issues

- Ethical or regulatory concerns

8. **A8.** Which factors are most influential when selecting new AI tools for product-management tasks? (Select top two)
   - Ease of integration with existing systems
   - User-friendly interface and ease of use
   - Vendor reputation and support
   - Cost-effectiveness
   - Recommendations from peers or industry reviews
   - Flexibility and scalability of the AI solution
9. **A9.** How do you expect your team's AI usage in product-management tasks to change in the next 12 months?
   - Significantly increase
   - Slightly increase
   - Stay about the same
   - Slightly decrease
   - Significantly decrease

B.6. Section B: workload shifts and perceived decision support

1. **B1.** Since integrating AI tools, rate how significantly your workload has decreased for each task below (No reduction, Slight reduction, Moderate reduction, Significant reduction, Very significant reduction):
   - Routine administrative & reporting tasks
   - Manual customer/user feedback collection & analysis
   - Market & competitor research (data gathering, synthesis)
   - QA, product testing, bug tracking management
   - Routine communications (emails, updates)
   - Documentation drafting & content creation
   - Task management and backlog updates
2. **B2.** Which tasks have notably increased or become more complex due to AI integration? (Select all applicable)
   - Strategic planning and roadmapping
   - Stakeholder engagement & cross-functional coordination
   - Innovation & creative product ideation
   - AI tool management (oversight, monitoring performance, model tuning)

- Ethical considerations & governance related to AI use
- Skill development, training, and professional learning on AI tools
- Communication & presentation of AI-driven insights to stakeholders

3. **B3.** Which tasks do you prioritize most frequently with time saved due to AI-driven efficiencies? (Rank top 3; 1 = highest priority)
   - Strategic product planning & vision-setting
   - Deep customer/user interaction & user research
   - Stakeholder & cross-team management
   - Innovative experimentation & idea generation
   - AI oversight, governance & ethical management
   - Professional growth & learning new AI-related skills
4. **B4.** Indicate the overall productivity impact you experienced due to AI integration.
   - Productivity significantly decreased
   - Productivity slightly decreased
   - No notable productivity change
   - Productivity slightly increased
   - Productivity significantly increased
5. **B5.** How many additional weekly hours can you now allocate specifically toward strategic or high-value tasks due to AI automation?
   - None
   - 1–2 hours
   - 3–5 hours
   - 6–10 hours
   - More than 10 hours
6. **B6.** Has the use of AI tools improved your decision-making capabilities as a product manager?
   - Significantly improved
   - Moderately improved
   - Slightly improved
   - No change
   - Decision-making has worsened
7. **B7.** Looking forward, how do you expect AI integration to influence your task responsibilities as a product manager in the next 1–2 years?
   - Significantly more focus on strategic tasks (planning, vision, stakeholder management)
   - Moderate shift towards strategic and innovation tasks
   - Slight or minimal change in responsibilities

- Increasing complexity and oversight of AI tool management
- Greater focus on ethical and governance responsibilities related to AI use

B.7. Section C: skills, training, and capability development

1. **C1.** Estimate the average hours you save weekly due to AI automation of your product-management tasks.
   - 0 hours
   - 1–3 hours
   - 4–7 hours
   - 8–12 hours
   - 13+ hours
2. **C2.** Rate how much your ability to focus on strategic aspects of your role has improved due to AI.
   - Not improved at all
   - Slightly improved
   - Moderately improved
   - Significantly improved
   - Very significantly improved
3. **C3.** Select the top 3 skills you have developed or improved due to using AI in your role.
   - AI literacy and understanding of AI concepts
   - Prompt engineering (effective prompting of generative AI tools)
   - Enhanced data analytics & data-driven decision-making
   - Ethical oversight & governance of AI tools
   - Improved strategic decision-making skills
   - Stronger communication and data storytelling capabilities
   - Technical proficiency in managing AI tools and platforms
4. **C4.** How confident are you in your current ability to effectively leverage AI tools in your daily PM tasks?
   - Not confident at all
   - Slightly confident
   - Moderately confident
   - Very confident
   - Extremely confident
5. **C5.** Has your organization provided structured training focused explicitly on the skills

necessary for effective AI integration?

- No structured training provided
- Informal or self-taught learning only
- Formal internal training programs provided
- External training/certifications offered by the organization

6. **C6.** Which training resources would most effectively support your skill development to handle AI integration? (Select top two)
   - Hands-on workshops & practical AI-tool training
   - Formal AI-related certifications or courses
   - Internal best practices & case-study sharing
   - Regular access to external AI experts or mentors
   - Ongoing updates on AI developments and trends relevant to PMs
7. **C7.** How has AI integration influenced your perception of skills considered most valuable for a product manager today?
   - No change in perception
   - Slightly shifted toward technical AI-related skills
   - Significantly shifted toward technical AI-related skills
   - Slightly shifted toward soft skills (ethics, communication, strategy)
   - Significantly shifted toward soft skills (ethics, communication, strategy)

B.8. Section D: governance, ethics, and responsibility

1. **D1.** How frequently do ethical considerations related to AI integration arise in your daily tasks as a product manager?
   - Never
   - Rarely
   - Occasionally
   - Often
   - Very Often
2. **D2.** Which ethical challenges have you directly encountered due to the use of AI tools in your product-management role? (Select all applicable)
   - Privacy and protection of customer data
   - Potential bias in AI-driven recommendations/decisions
   - Transparency and explainability of AI decisions
   - Accountability for errors or decisions made by AI
   - Compliance with regulatory standards (e.g., GDPR, EU AI Act)
   - No ethical challenges encountered

3. **D3.** Who primarily holds responsibility for managing ethical AI-related decisions in your organization?
    - Product managers individually
    - Dedicated AI/Ethics committee
    - Cross-functional team or collaboration
    - Formal organizational guidelines/policies
    - No clear ownership/responsibility defined
4. **D4.** How confident do you feel addressing ethical issues arising from AI usage in your product-management role?
    - Not confident at all
    - Slightly confident
    - Moderately confident
    - Very confident
    - Extremely confident
5. **D5.** Briefly describe a recent ethical decision-making scenario related to AI usage that you faced, and how you addressed it (open-ended).
6. **D6.** Does your organization provide clear guidelines or policies for ethical AI usage in product management?
    - No clear guidelines exist
    - Informal ethical guidance only
    - Formal internal AI ethics guidelines exist
    - Adhere to external standards (e.g., IEEE, EU AI Act, GDPR)
7. **D7.** Have you received structured training specifically addressing ethical considerations related to AI use in your role?
    - No structured ethical training provided
    - Informal/self-directed learning only
    - Formal internal ethical AI training provided
    - External courses/certifications provided by the organization
8. **D8.** Which resources would best support you in managing ethical AI-related challenges? (Select up to two)
    - Dedicated ethical AI training workshops
    - Clearer organizational ethical guidelines
    - Regular audits or assessments of AI systems
    - Real-world ethical AI case studies/examples
    - Regular access to ethical AI advisors or experts
9. **D9.** What is your greatest ethical concern regarding the increased use of AI in your product-management role in the future? (open-ended)

B.9. Simplified Chinese reference translation

Note. The following is a Simplified Chinese reference translation to document bilingual administration. Analysis and quotation should rely on the English version above.

A 部分：AI 工具使用与采纳背景

- **A1.** 请指出您在日常产品管理任务中使用以下各类AI 应用的频率（从不、很少、偶尔、经常、总是）：
  - 用于重复性任务的机器人流程自动化（RPA）
  - 工作流自动化工具
  - 用于功能优先级排序的预测分析工具
  - 用于分析客户反馈/情感倾向的自然语言处理（NLP）工具
  - AI 驱动的QA 与产品测试工具
  - 生成式AI 工具（例如ChatGPT、Claude），用于撰写文档/初稿内容
  - 用于客户互动/支持的AI 聊天机器人或虚拟助手
- **A2.** 您在产品管理岗位上主动使用AI 应用已有多长时间?
  - 少于6 个月
  - 6 个月– 1 年
  - 1–2 年
  - 2–3 年
  - 3 年以上
- **A3.** 在您的PM 活动中，通常由哪个群体主导是否实施AI 工具的决策?
  - 产品团队
  - 工程/技术团队
  - 高层/管理层
  - 数据团队
  - 其他
- **A4.** 当您的组织将AI 集成到产品管理活动中时，其主要目标是什么？（最多选择两项）
  - 降低运营成本
  - 提升产品经理生产力与效率
  - 改善客户/用户体验
  - 提升产品质量与可靠性
  - 创新产品开发流程
  - 获得市场竞争优势
- **A5.** 您所采用的AI 工具在多大程度上达到了最初的期望?
  - 完全未达到期望
  - 略微达到期望

  - 在一定程度上达到期望
  - 基本达到期望
  - 明显超出期望
- **A6.** 请列举您经常使用的AI 工具/平台（开放题）。
- **A7.** 在将AI 应用于产品管理任务时，您遇到过哪些挑战？（可多选）
  - 技术困难或集成复杂
  - 初始学习曲线与团队阻力
  - 数据可用性或质量问题
  - 培训或支持不足
  - 预算限制与成本管理
  - AI 工具的可靠性或准确性问题
  - 伦理或监管方面的顾虑
- **A8.** 在为产品管理任务选择新的AI 工具时，哪些因素最具影响力？（选择最重要的两项）
  - 与现有系统的集成难易度
  - 用户界面友好与易用性
  - 供应商声誉与支持
  - 成本效益
  - 同行建议或行业评测推荐
  - AI 方案的灵活性与可扩展性
- **A9.** 您预计未来12 个月团队在产品管理任务中的AI 使用情况将如何变化？
  - 显著增加
  - 略微增加
  - 基本保持不变
  - 略微减少
  - 显著减少

B 部分：工作量变化与决策支持感知

- **B1.** 自从集成AI 工具以来，请评价以下各项任务的工作量减少程度（无减少、轻微减少、适度减少、显著减少、非常显著减少）：
  - 常规行政与汇报任务
  - 手动收集与分析客户/用户反馈
  - 市场与竞品调研（数据收集、综合）
  - QA、产品测试与缺陷跟踪管理
  - 日常沟通（邮件、进展更新）
  - 文档撰写与内容创作
  - 任务管理与待办/Backlog 更新
- **B2.** 由于AI 集成，哪些任务明显增加或变得更复杂？（可多选）

  - 战略规划与路线图制定
  - 利益相关者沟通与跨职能协作
  - 创新与创造性产品构思
  - AI 工具管理（监督、性能监控、模型调优）
  - 与AI 使用相关的伦理考量与治理
  - 围绕AI 工具的技能发展、培训与专业学习
  - 向利益相关者沟通与展示AI 驱动的洞察
- **B3.** 由于AI 带来的效率提升而节省的时间，您最常将其用于哪些任务？（请将前三项排序；1 = 最高优先级）
  - 战略性产品规划与愿景设定
  - 深度客户/用户互动与用户研究
  - 利益相关者与跨团队管理
  - 创新性试验与创意生成
  - AI 监督、治理与伦理管理
  - 职业成长与学习新的AI 相关技能
- **B4.** 请指出AI 集成对您整体生产力的影响。
  - 生产力显著下降
  - 生产力略微下降
  - 生产力无明显变化
  - 生产力略微提升
  - 生产力显著提升
- **B5.** 由于AI 自动化，您现在每周可以额外分配多少时间用于战略性或高价值任务?
  - 没有
  - 1–2 小时
  - 3–5 小时
  - 6–10 小时
  - 超过10 小时
- **B6.** AI 工具是否提升了您作为产品经理的决策能力?
  - 显著提升
  - 适度提升
  - 略微提升
  - 没有变化
  - 决策能力变差
- **B7.** 展望未来1–2 年，您预计AI 集成将如何影响您作为产品经理的任务职责?
  - 显著更多地聚焦战略任务（规划、愿景、利益相关者管理）
  - 中等程度转向战略与创新任务
  - 职责变化很小或几乎没有变化

  - AI 工具管理的复杂性与监督要求增加
  - 更加聚焦与AI 使用相关的伦理与治理责任

C 部分：技能、培训与能力发展

- **C1.** 估计由于AI 自动化您的产品管理任务，您每周平均节省多少小时?
  - 0 小时
  - 1–3 小时
  - 4–7 小时
  - 8–12 小时
  - 13 小时以上
- **C2.** 请评价由于AI 的帮助，您专注于战略性工作的能力提升程度。
  - 完全没有提升
  - 略微提升
  - 适度提升
  - 显著提升
  - 非常显著提升
- **C3.** 请选择由于在工作中使用AI 而得到发展或提升的前三项技能。
  - AI 素养与对AI 概念的理解
  - 提示词工程（有效使用生成式AI 工具的提示技巧）
  - 更强的数据分析与数据驱动决策能力
  - 对AI 工具的伦理监督与治理能力
  - 更强的战略决策能力
  - 更强的沟通与数据叙事能力
  - 管理AI 工具与平台的技术熟练度
- **C4.** 您对自己目前在日常PM 任务中有效使用AI 工具的能力有多自信?
  - 完全没有信心
  - 略有信心
  - 有一定信心
  - 很有信心
  - 非常有信心
- **C5.** 您的组织是否提供了专门针对有效AI 集成所需技能的结构化培训?
  - 未提供任何结构化培训
  - 仅有非正式/自学式学习
  - 提供正式的内部培训项目
  - 组织提供外部培训/认证课程
- **C6.** 哪些培训资源最能有效支持您提升处理AI 集成的能力？（选择最重要的两项）
  - 实操工作坊与AI 工具实践培训
  - 正式的AI 相关认证或课程

  - 内部最佳实践与案例分享
  - 定期获得外部AI 专家或导师支持
  - 持续更新与PM 相关的AI 进展与趋势
- **C7.** AI 集成如何影响您对当今产品经理最有价值技能的看法?
  - 看法没有变化
  - 略微转向技术型AI 相关技能
  - 显著转向技术型AI 相关技能
  - 略微转向软技能（伦理、沟通、战略）
  - 显著转向软技能（伦理、沟通、战略）

D 部分：治理、伦理与责任

- **D1.** 作为产品经理，在日常工作中与AI 集成相关的伦理考量出现的频率如何?
  - 从不
  - 很少
  - 偶尔
  - 经常
  - 非常频繁
- **D2.** 由于使用AI 工具，您曾直接遇到哪些伦理挑战? （可多选）
  - 隐私与客户数据保护
  - AI 驱动的推荐/决策可能存在偏见
  - AI 决策的透明度与可解释性
  - AI 造成错误或其决策的问责归属
  - 遵循监管标准（例如GDPR、欧盟AI 法案）
  - 未遇到任何伦理挑战
- **D3.** 在您的组织中，谁主要负责管理与AI 相关的伦理决策?
  - 产品经理个人
  - 专门的AI/伦理委员会
  - 跨职能团队或协作机制
  - 正式的组织指南/政策
  - 没有明确的责任归属/负责人
- **D4.** 在您的产品管理工作中，当出现由AI 使用引发的伦理问题时，您有多自信能够应对?
  - 完全没有信心
  - 略有信心
  - 有一定信心
  - 很有信心
  - 非常有信心
- **D5.** 请简要描述您最近遇到的一个与AI 使用相关的伦理决策情境，以及您是如何处理

的（开放题）。

- **D6.** 您的组织是否为产品管理中的伦理性AI 使用提供了明确的指南或政策?
  - 没有明确指南
  - 仅有非正式伦理指导
  - 存在正式的内部AI 伦理指南
  - 遵循外部标准（例如IEEE、欧盟AI 法案、GDPR）
- **D7.** 您是否接受过专门针对您岗位中与AI 使用相关伦理考量的结构化培训?
  - 未提供结构化伦理培训
  - 仅有非正式/自学式学习
  - 提供正式的内部伦理AI 培训
  - 组织提供外部课程/认证
- **D8.** 哪些资源最能支持您管理与AI 相关的伦理挑战? （最多选择两项）
  - 专门的伦理AI 培训工作坊
  - 更清晰的组织伦理指南
  - 定期审计或评估AI 系统
  - 真实的伦理AI 案例/示例
  - 定期获得伦理AI 顾问或专家支持
- **D9.** 展望未来，在您的产品管理角色中，您对AI 使用增加的最大伦理担忧是什么? （开放题）

C | Study 2 Survey Raw Distributions (by Region)

This appendix reports raw response distributions (counts) for the primary Study 2 categorical items, split by region (USA vs. China) and overall. For multiple-response items (e.g., "select up to two") totals can exceed N.

C.1. Adoption context and organizational determinants

Decision-making group for AI implementation	**USA**	**China**	**Overall**
Product Team	10	8	18
Engineering/Tech Team	11	9	20
Executive/Leadership Team	7	0	7
Data Team	4	9	13
Other	5	1	6

Table C.1: Decision-making group driving AI tool implementation (Study 2; counts).

AI integration goals (select up to two)	**USA**	**China**	**Overall**
Reducing operational costs	19	20	39
Increasing PM productivity and efficiency	11	16	27
Improving customer/user experience	4	3	7
Enhancing product quality and reliability	13	11	24
Innovating product development processes	17	12	29
Gaining competitive advantage in the market	10	12	22

Table C.2: AI integration goals (Study 2; counts; multiple responses allowed).

Adoption barriers (select all applicable)	USA	China	Overall
Technical difficulties or integration complexity	15	10	25
Initial learning curve and team resistance	9	4	13
Data availability or quality issues	11	11	22
Insufficient training or support	10	14	24
Budget constraints and cost management	9	14	23
AI tool reliability or accuracy issues	7	8	15
Ethical or regulatory concerns	12	14	26

Table C.3: Adoption barriers encountered (Study 2; counts; multiple responses allowed).

Tool selection criteria (select top two)	USA	China	Overall
Ease of integration with existing systems	13	9	22
User-friendly interface and ease of use	10	12	22
Vendor reputation and support	13	5	18
Cost-effectiveness	18	25	43
Recommendations from peers or industry reviews	7	14	21
Flexibility and scalability of the AI solution	13	9	22

Table C.4: Tool selection criteria (Study 2; counts; multiple responses allowed).

C.2. Task complexity and time reinvestment

Tasks more complex due to AI integration (select all applicable)	USA	China	Overall
Strategic planning and roadmapping	9	10	19
Stakeholder engagement & cross-functional coordination	15	13	28
Innovation & creative product ideation	14	12	26
AI tool management (oversight, monitoring performance, model tuning)	3	9	12
Ethical considerations & governance related to AI use	6	7	13
Skill development, training, and professional learning on AI tools	12	12	24
Communication & presentation of AI-driven insights to stakeholders	19	10	29

Table C.5: Tasks reported as more complex due to AI integration (Study 2; counts; multiple responses allowed).

Tasks prioritized with time saved (appearing in top 3 ranks)	USA	China	Overall
Strategic product planning & vision-setting	20	14	34
Deep customer/user interaction & user research	18	13	31
Stakeholder & cross-team management	13	12	25
Innovative experimentation & idea generation	10	11	21
AI oversight, governance & ethical management	5	12	17
Professional growth & learning new AI-related skills	11	10	21

Table C.6: Time reinvestment priorities (Study 2; counts; respondents ranked top 3).

C.3. Skills, training, and capability development

Top skills improved due to AI (selected in top 3)	USA	China	Overall
AI literacy and understanding of AI concepts	8	12	20
Prompt engineering (effective prompting of generative AI tools)	11	14	25
Enhanced data analytics & data-driven decision-making	18	10	28
Ethical oversight & governance of AI tools	9	0	9
Improved strategic decision-making skills	3	8	11
Stronger communication and data storytelling capabilities	4	5	9
Technical proficiency in managing AI tools and platforms	3	14	17

Table C.7: Skills improved due to AI integration (Study 2; counts; respondents selected top 3).

Confidence leveraging AI tools	USA	China	Overall
Not confident at all	2	6	8
Slightly confident	6	4	10
Moderately confident	8	15	23
Very confident	16	10	26
Extremely confident	5	2	7

Table C.8: Confidence in effectively leveraging AI tools (Study 2; counts).

Structured AI training offered by organization	USA	China	Overall
No structured training provided	8	5	13
Informal or self-taught learning only	14	17	31
Formal internal training programs provided	7	12	19
External training/certifications offered by the organization	8	3	11

Table C.9: Structured AI training offered (Study 2; counts).

Preferred training resources (select top two)	USA	China	Overall
Hands-on workshops & practical AI-tool training	18	17	35
Formal AI-related certifications or courses	14	10	24
Internal best practices & case-study sharing	10	8	18
Regular access to external AI experts or mentors	9	11	20
Ongoing updates on AI developments and trends relevant to PMs	12	7	19

Table C.10: Preferred resources for AI skill development (Study 2; counts; multiple responses allowed).

C.4. Ethics and governance

Ethics frequency in daily tasks	USA	China	Overall
Never	7	5	12
Rarely	3	13	16
Occasionally	14	15	29
Often	13	2	15
Very Often	0	2	2

Table C.11: Frequency of ethics-related considerations (Study 2; counts).

C| Study 2 Survey Raw Distributions (by Region)

Ethical challenges encountered (select all applicable)	USA	China	Overall
Privacy and protection of customer data	11	12	23
Potential bias in AI-driven recommendations/decisions	11	8	19
Transparency and explainability of AI decisions	7	5	12
Accountability for errors or decisions made by AI	5	9	14
Compliance with regulatory standards (e.g., GDPR, EU AI Act)	2	1	3
No ethical challenges encountered	1	2	3

Table C.12: Ethical challenges encountered (Study 2; counts; multiple responses allowed).

Ethical responsibility structure	USA	China	Overall
Product managers individually	0	11	11
Dedicated AI/Ethics committee	6	4	10
Cross-functional team or collaboration	6	7	13
Formal organizational guidelines/policies	6	7	13
No clear ownership/responsibility defined	18	7	25

Table C.13: Responsibility for ethical AI decisions (Study 2; counts).

Ethical AI guideline structure	USA	China	Overall
No clear guidelines exist	11	16	27
Informal ethical guidance only	7	12	19
Formal internal AI ethics guidelines exist	18	4	22
Adhere to external standards (e.g., IEEE, EU AI Act, GDPR)	0	4	4

Table C.14: Existence of ethical AI guidelines (Study 2; counts).

Structured ethical AI training received (optional)	USA	China	Overall
No structured ethical training provided	6	6	12
Informal/self-directed learning only	7	4	11
Formal internal ethical AI training provided	6	6	12
External courses/certifications provided by the organization	0	4	4

Table C.15: Structured ethical AI training received (Study 2; counts).

Greatest ethical concern (coded categories)	USA	China	Overall
Accuracy & reliability	7	10	17
Privacy & data security	9	2	11
Accountability & responsibility	7	2	9
Job displacement	4	2	6
Overuse & creativity loss	4	1	5
No concern stated	2	13	15
Other	4	7	11

Table C.16: Greatest ethical concern regarding increased AI use (Study 2; counts; open-ended responses coded).

D | Supplementary Tables and Statistical Outputs

This appendix reports supporting model outputs for the primary analyses. All raw CSV outputs are available in `final_analysis/long_format_outputs/tables/`.

D.1. Primary condition effects (A1)

Table D.1: Mixed-effects ANOVA (Structure × Timing) for Overall MCQ accuracy.

Effect	df_1	df_2	F	p	η_p^2
Structure	1	22.00	4.69	$= 0.042$	0.176
Timing	2	42.18	17.61	$< .001$	0.455
Structure × Timing	2	42.12	0.40	$= 0.670$	0.019

Table D.2: Mixed-effects ANOVA (Structure × Timing) for Article-only MCQ accuracy.

Effect	df_1	df_2	F	p	η_p^2
Structure	1	22.00	1.89	$= 0.183$	0.079
Timing	2	43.25	0.35	$= 0.706$	0.016
Structure × Timing	2	43.00	0.50	$= 0.608$	0.023

Table D.3: Mixed-effects ANOVA (Structure × Timing) for False-lure accuracy.

Effect	df_1	df_2	F	p	η_p^2
Structure	1	22.00	4.20	= 0.053	0.160
Timing	2	43.01	0.99	= 0.380	0.044
Structure × Timing	2	42.75	0.47	= 0.630	0.021

Table D.4: Mixed-effects ANOVA (Structure × Timing) for Free-recall total score.

Effect	df_1	df_2	F	p	η_p^2
Structure	1	22.00	0.40	= 0.536	0.018
Timing	2	42.87	0.02	= 0.983	0.001
Structure × Timing	2	42.63	0.70	= 0.503	0.032

Table D.5: Mixed-effects ANOVA (Structure × Timing) for Summary viewing time (seconds).

Effect	df_1	df_2	F	p	η_p^2
Structure	1	22.00	0.50	= 0.486	0.022
Timing	2	43.25	13.32	< .001	0.381
Structure × Timing	2	43.00	0.34	= 0.717	0.015

Table D.6: Mixed-effects ANOVA (Structure × Timing) for Summary time proportion.

Effect	df_1	df_2	F	p	η_p^2
Structure	1	22.00	0.01	= 0.933	0.000
Timing	2	43.25	16.20	< .001	0.428
Structure × Timing	2	43.00	0.13	= 0.879	0.006

Table D.7: Mixed-effects ANOVA (Structure × Timing) for Reading time (minutes).

Effect	df_1	df_2	F	p	η_p^2
Structure	1	22.00	1.26	= 0.275	0.054
Timing	2	43.25	1.16	= 0.324	0.051
Structure × Timing	2	43.00	0.19	= 0.826	0.009

Table D.8: Mixed-effects ANOVA (Structure × Timing) for Total block time (seconds).

Effect	df_1	df_2	F	p	η_p^2
Structure	1	22.00	1.67	= 0.210	0.070
Timing	2	43.25	0.01	= 0.992	0.000
Structure × Timing	2	43.00	0.30	= 0.745	0.014

Table D.9: Mixed-effects ANOVA (Structure × Timing) for Mental effort rating.

Effect	df_1	df_2	F	p	η_p^2
Structure	1	22.00	2.43	= 0.133	0.099
Timing	2	42.24	1.50	= 0.236	0.066
Structure × Timing	2	42.16	2.32	= 0.111	0.099

Table D.10: Mixed-effects ANOVA (Structure × Timing) for Post-block AI trust (trust_new).

Effect	df_1	df_2	F	p	η_p^2
Structure	1	22.00	1.58	= 0.222	0.067
Timing	2	43.25	7.90	= 0.001	0.268
Structure × Timing	2	43.00	1.70	= 0.194	0.073

Table D.11: Mixed-effects ANOVA (Structure × Timing) for Post-block AI dependence (dependence_new).

Effect	df_1	df_2	F	p	η_p^2
Structure	1	22.00	6.21	= 0.021	0.220
Timing	2	43.25	7.74	= 0.001	0.264
Structure × Timing	2	43.00	0.59	= 0.559	0.027

D.2. Post-hoc timing contrasts (A1)

Table D.12: Post-hoc timing contrasts for Overall MCQ accuracy (Holm-corrected).

Contrast	Estimate	SE	*df*	*t*	*p*
pre_reading - synchronous	0.172	0.031	42.05	5.54	$< .001$
pre_reading - post_reading	0.143	0.031	42.18	4.56	$< .001$
synchronous - post_reading	-0.029	0.032	42.32	-0.92	$= 0.361$

Table D.13: Post-hoc timing contrasts for Summary viewing time (seconds) (Holm-corrected).

Contrast	Estimate	SE	*df*	*t*	*p*
pre_reading - synchronous	32.178	12.018	42.43	2.68	$= 0.021$
pre_reading - post_reading	62.987	12.216	43.43	5.16	$< .001$
synchronous - post_reading	30.809	12.408	43.94	2.48	$= 0.021$

Table D.14: Post-hoc timing contrasts for Summary time proportion (Holm-corrected).

Contrast	Estimate	SE	*df*	*t*	*p*
pre_reading - synchronous	0.054	0.019	42.43	2.79	$= 0.012$
pre_reading - post_reading	0.112	0.020	43.43	5.69	$< .001$
synchronous - post_reading	0.058	0.020	43.94	2.90	$= 0.012$

Table D.15: Post-hoc timing contrasts for Reading time (minutes) (Holm-corrected).

Contrast	Estimate	SE	*df*	*t*	*p*
pre_reading - synchronous	-1.974	0.705	42.43	-2.80	$= 0.023$
pre_reading - post_reading	-0.967	0.717	43.43	-1.35	$= 0.347$
synchronous - post_reading	1.007	0.728	43.94	1.38	$= 0.347$

Table D.16: Post-hoc timing contrasts for Total block time (seconds) (Holm-corrected).

Contrast	Estimate	SE	*df*	*t*	*p*
pre_reading - synchronous	-113.589	53.500	42.43	-2.12	= 0.079
pre_reading - post_reading	15.762	54.383	43.43	0.29	= 0.773
synchronous - post_reading	129.350	55.240	43.94	2.34	= 0.071

Table D.17: Post-hoc timing contrasts for Post-block AI trust (trust_new) (Holm-corrected).

Contrast	Estimate	SE	*df*	*t*	*p*
pre_reading - synchronous	0.625	0.171	42.43	3.65	= 0.002
pre_reading - post_reading	0.542	0.174	43.43	3.12	= 0.007
synchronous - post_reading	-0.083	0.177	43.94	-0.47	= 0.639

Table D.18: Post-hoc timing contrasts for Post-block AI dependence (dependence_new) (Holm-corrected).

Contrast	Estimate	SE	*df*	*t*	*p*
pre_reading - synchronous	0.625	0.190	42.43	3.30	= 0.004
pre_reading - post_reading	0.667	0.193	43.43	3.46	= 0.004
synchronous - post_reading	0.042	0.196	43.94	0.21	= 0.832

D.3. Condition descriptives (A1)

Table D.19: Descriptive statistics for Overall MCQ accuracy by condition.

Structure	Timing	*n*	Mean	SD	SE
Integrated	Pre Reading	12	0.750	0.124	0.036
Integrated	Synchronous	12	0.542	0.124	0.036
Integrated	Post Reading	12	0.607	0.094	0.027
Segmented	Pre Reading	12	0.649	0.108	0.031
Segmented	Synchronous	12	0.524	0.173	0.050
Segmented	Post Reading	12	0.518	0.143	0.041

Table D.20: Descriptive statistics for Article-only MCQ accuracy by condition.

Structure	Timing	n	Mean	SD	SE
Integrated	Pre Reading	12	0.542	0.234	0.068
Integrated	Synchronous	12	0.458	0.317	0.091
Integrated	Post Reading	12	0.562	0.264	0.076
Segmented	Pre Reading	12	0.479	0.249	0.072
Segmented	Synchronous	12	0.438	0.264	0.076
Segmented	Post Reading	12	0.396	0.198	0.057

Table D.21: Descriptive statistics for False-lure accuracy by condition.

Structure	Timing	n	Mean	SD	SE
Integrated	Pre Reading	12	0.542	0.396	0.114
Integrated	Synchronous	12	0.667	0.326	0.094
Integrated	Post Reading	12	0.458	0.334	0.096
Segmented	Pre Reading	12	0.333	0.326	0.094
Segmented	Synchronous	12	0.417	0.289	0.083
Segmented	Post Reading	12	0.375	0.311	0.090

Table D.22: Descriptive statistics for Free-recall total score by condition.

Structure	Timing	n	Mean	SD	SE
Integrated	Pre Reading	12	5.417	2.193	0.633
Integrated	Synchronous	12	5.292	1.912	0.552
Integrated	Post Reading	12	5.167	2.462	0.711
Segmented	Pre Reading	12	5.583	1.690	0.488
Segmented	Synchronous	12	5.792	2.116	0.611
Segmented	Post Reading	12	5.958	1.852	0.535

Table D.23: Descriptive statistics for Summary viewing time (seconds) by condition.

Structure	Timing	n	Mean	SD	SE
Integrated	Pre Reading	12	141.873	35.915	10.368
Integrated	Synchronous	12	99.680	50.440	14.561
Integrated	Post Reading	12	74.427	43.444	12.541
Segmented	Pre Reading	12	123.106	55.986	16.162
Segmented	Synchronous	12	100.943	53.654	15.489
Segmented	Post Reading	12	64.577	33.072	9.547

Table D.24: Descriptive statistics for Summary time proportion by condition.

Structure	Timing	n	Mean	SD	SE
Integrated	Pre Reading	12	0.246	0.056	0.016
Integrated	Synchronous	12	0.190	0.110	0.032
Integrated	Post Reading	12	0.141	0.091	0.026
Segmented	Pre Reading	12	0.251	0.100	0.029
Segmented	Synchronous	12	0.200	0.087	0.025
Segmented	Post Reading	12	0.132	0.059	0.017

Table D.25: Descriptive statistics for Reading time (minutes) by condition.

Structure	Timing	n	Mean	SD	SE
Integrated	Pre Reading	12	7.337	1.611	0.465
Integrated	Synchronous	12	7.447	2.232	0.644
Integrated	Post Reading	12	7.982	2.394	0.691
Segmented	Pre Reading	12	6.107	2.164	0.625
Segmented	Synchronous	12	6.928	3.301	0.953
Segmented	Post Reading	12	7.396	2.654	0.766

Table D.26: Descriptive statistics for Total block time (seconds) by condition.

Structure	Timing	n	Mean	SD	SE
Integrated	Pre Reading	12	582.073	114.386	33.020
Integrated	Synchronous	12	546.530	128.953	37.226
Integrated	Post Reading	12	553.327	145.558	42.019
Segmented	Pre Reading	12	489.556	142.748	41.208
Segmented	Synchronous	12	516.593	223.455	64.506
Segmented	Post Reading	12	508.327	170.303	49.162

Table D.27: Descriptive statistics for Mental effort rating by condition.

Structure	Timing	n	Mean	SD	SE
Integrated	Pre Reading	12	5.917	0.669	0.193
Integrated	Synchronous	12	6.000	0.739	0.213
Integrated	Post Reading	12	6.083	0.793	0.229
Segmented	Pre Reading	12	5.917	0.793	0.229
Segmented	Synchronous	12	5.250	1.288	0.372
Segmented	Post Reading	12	5.500	1.243	0.359

Table D.28: Descriptive statistics for Post-block AI trust (trust_new) by condition.

Structure	Timing	n	Mean	SD	SE
Integrated	Pre Reading	12	4.167	0.577	0.167
Integrated	Synchronous	12	3.583	0.669	0.193
Integrated	Post Reading	12	3.917	1.084	0.313
Segmented	Pre Reading	12	4.833	0.835	0.241
Segmented	Synchronous	12	4.167	1.193	0.345
Segmented	Post Reading	12	4.000	1.348	0.389

Table D.29: Descriptive statistics for Post-block AI dependence (dependence_new) by condition.

Structure	Timing	n	Mean	SD	SE
Integrated	Pre Reading	12	4.250	0.965	0.279
Integrated	Synchronous	12	3.833	0.577	0.167
Integrated	Post Reading	12	3.667	0.985	0.284
Segmented	Pre Reading	12	5.333	1.303	0.376
Segmented	Synchronous	12	4.500	1.243	0.359
Segmented	Post Reading	12	4.583	0.900	0.260

E | Ethics / Consent Documents (Summary)

This appendix summarizes the ethics-related materials and participant protections implemented across Study 1 and Study 2.

E.1. Study 1 (controlled experiment)

Participants reviewed an informed-consent screen describing the purpose of the study (AI-assisted reading and memory), the approximate duration (about 90 minutes), the main tasks (reading short scientific articles, answering questions, and providing ratings), and key protections: voluntary participation, the right to withdraw at any time without penalty, and anonymous data handling. The consent screen also specified basic eligibility (18+ and laptop/desktop participation).

Because Study 1 includes AI-generated summaries and measures epistemic risk via plausible but incorrect details, the protocol includes debriefing guidance intended to prevent lasting misconceptions (i.e., participants should not treat AI summaries as ground truth).

E.2. Study 2 (field survey)

Survey participation was voluntary. The questionnaire was designed to avoid collection of proprietary company information and was analyzed in aggregate. Open-ended responses were treated as confidential and reported only in anonymized form.

E.3. Data management

Across both studies, datasets were stored in access-controlled locations and used only for research purposes. Reporting focuses on patterns and mechanisms rather than identifying individuals or organizations.