

---

AI translation · View original & related papers at  
[chinaxiv.org/items/chinaxiv-202601.00181](https://chinaxiv.org/items/chinaxiv-202601.00181)

---

## From Computational Thinking to AI Thinking: A Meta-Cognitive Paradigm for Human-AI Co- Cognition

**Authors:** LIU Huabo, LIU Huabo

**Date:** 2026-01-15T16:09:46+00:00

### Abstract

The rapid advancement of generative artificial intelligence (AI) is driving a profound shift in human cognitive paradigms, from the “computational era” to the “intelligent era.” This paper proposes and systematically elaborates a meta-cognitive paradigm, termed AI Thinking, aimed at addressing the question: as AI evolves from a mere tool to a cognitive partner, how should human thinking, research paradigms, and knowledge-production logics be reconfigured? Drawing on insights from cognitive science, philosophy of science, and educational theory, we develop a dynamic cognitive framework structured around three pillars: intent modeling and contextual anchoring, generative dialogue facilitation, and critical integration and meaning construction. This framework reveals the essence of AI Thinking as a human-AI co-cognitive loop. Through interdisciplinary case studies spanning scientific research, engineering design, educational innovation, and business decision-making, we demonstrate that AI Thinking not only enhances problem-solving efficiency but also points toward a potential transformation in prevailing scientific paradigms: the scientific method is increasingly being augmented by a collaborative loop of “data-AI-hypothesis-validation,” transforming scientists from knowledge operators into architects of human-AI intelligent teams. Scientific culture may be interpreted as moving toward what could be described as a “fourth cultural configuration,” grounded in human-AI collective intelligence. Unlike methodological approaches focused on “how to use AI,” AI Thinking can be understood as an epistemological proposal concerned with how to co-think and co-construct with AI, providing a systematic cognitive infrastructure for education, research, and the development of civilization in the intelligent era.

## Full Text

# From Computational Thinking to AI Thinking: A Meta-Cognitive Paradigm for Human-AI Co-Cognition

Huabo LIU

liuhuabo@hbeu.edu.cn

Hubei Engineering University

## Abstract

The rapid advancement of generative artificial intelligence (AI) is driving a profound shift in human cognitive paradigms, from the “computational era” to the “intelligent era.” This paper proposes and systematically elaborates a meta-cognitive paradigm termed AI Thinking to address a fundamental question: as AI evolves from a mere tool to a cognitive partner, how should human thinking, research paradigms, and knowledge-production logics be reconfigured? Drawing on insights from cognitive science, philosophy of science, and educational theory, we develop a dynamic cognitive framework structured around three pillars: intent modeling and contextual anchoring, generative dialogue facilitation, and critical integration and meaning construction. This framework reveals the essence of AI Thinking as a human-AI co-cognitive loop. Through interdisciplinary case studies spanning scientific research, engineering design, educational innovation, and business decision-making, we demonstrate that AI Thinking not only enhances problem-solving efficiency but also points toward a potential transformation in prevailing scientific paradigms: the scientific method is increasingly being augmented by a collaborative loop of “data-AI-hypothesis-validation,” transforming scientists from knowledge operators into architects of human-AI intelligent teams. Scientific culture may thus be moving toward what could be described as a “fourth cultural configuration,” grounded in human-AI collective intelligence. Unlike methodological approaches focused on “how to use AI,” AI Thinking can be understood as an epistemological proposal concerned with how to co-think and co-construct with AI, providing a systematic cognitive infrastructure for education, research, and the development of civilization in the intelligent era.

**Keywords:** AI Thinking; Human-AI Collaboration; Cognitive Paradigm; Generative Artificial Intelligence; Computational Thinking; Meta-Cognitive Framework

## 1 The Inevitable Paradigm Shift

The evolution of scientific thinking has always been closely intertwined with advancements in technology and the transformation of cognitive tools. Since the advent of modern science, experimental thinking established an empirical tradition grounded in controlled observation; logical and axiomatic thinking provided the deductive backbone of theoretical science; and in the latter half

of the 20th century, systems thinking emerged, emphasizing the understanding of complex wholes and feedback mechanisms. Entering the information age, computational thinking (Wing, 2006) reframed problem-solving as algorithmic processes executable by machines, becoming a foundational literacy in digital society. With the development of big data and high-performance computing, Jim Gray (2007) further proposed data-intensive scientific discovery—the “Fourth Paradigm” —which emphasizes the direct extraction of knowledge from massive datasets. For more than half a century, human society has operated largely within the paradigm of the “computational era,” where computers, as ultra-fast executors of deterministic logic, derive their primary value from automatically performing predefined computational tasks.

However, the rapid emergence of generative AI, exemplified by large language models, signals the onset of a fundamentally new “intelligent era” (Bommasani et al., 2021; Bubeck et al., 2023). In this era, the hallmark of productivity is no longer passive machines executing instructions, but AI agents capable of understanding context, generating coherent content, performing logical inferences, and even exhibiting nascent creativity. The role of AI is shifting from “tool” to “partner.” This transformation is profound: the entities with which we interact are no longer fully predictable deterministic systems, but intelligent agents whose behavior is probabilistic, generative, and to some extent unpredictable. Computational thinking provides a cognitive blueprint for engaging with deterministic machines, but it inherently assumes the programmability and predictability of the system. When confronted with a generative, probabilistic AI partner, the limitations of this paradigm become apparent. We can no longer precisely control every internal computation through code; instead, we must guide a process we cannot fully observe through natural language and dynamic dialogue (Bender et al., 2021). This shift demands entirely new skills in communication and guidance.

This new context imposes entirely novel demands on human cognition. Humans are no longer solely seeking a single “correct answer” but must evaluate the plausibility and applicability of multiple generative possibilities. Attention shifts from simply judging the correctness of outcomes to designing effective guidance strategies that elicit emergent cognition. Individuals no longer work in isolation; instead, they organize “human-AI intelligence teams” composed of humans and multiple specialized AI agents to collaboratively accomplish complex tasks (Park et al., 2023).

Both academia and industry have recognized this transformation and begun exploring it across adjacent fields. Research in prompt engineering focuses on how to structure input text to optimize the output quality of large language models, investigating techniques from zero-shot and few-shot prompting to chain-of-thought methods (Reynolds & McDonell, 2021; Wei, 2022). In human-AI interaction, scholars are rethinking interfaces for generative AI, exploring how to establish trust and co-create effectively through dialogue (Jakesch et al., 2023; Bansal et al., 2021). Concurrently, research on AI agents aims to endow AI with

autonomous planning, tool invocation, and task execution capabilities, enabling them to perform complex sequential tasks (Xi et al., 2025; Schick et al., 2024).

However, most studies remain focused on how to operate specific technologies, leaving a more fundamental question largely unaddressed from a meta-cognitive perspective: what core principles of thinking should humans adopt to organize their intentions, design human-AI interaction processes, and integrate AI outputs to tackle open-ended, complex, and uncertain challenges? At present, academia lacks a theoretical framework capable of encompassing concrete technical practices while elevating them to a universal, learnable, and foundational cognitive paradigm. To fill this critical gap, this paper introduces and systematically elaborates the concept of AI Thinking, aiming to establish a unified paradigm for human-AI co-cognition in the intelligent era. Specifically, we will: (1) clearly define the scope and essence of AI Thinking; (2) delineate its theoretical boundaries relative to existing concepts such as computational thinking; and (3) construct a complete, coherent, and operational theoretical framework.

It is worth noting that recent scholarship has begun to formalize “AI Thinking” from varying perspectives. Newman-Griffis (2025) proposes a practice-oriented framework designed to bridge interdisciplinary gaps, framing AI Thinking as a methodological toolbox for the systematic and responsible use of AI in research workflows. Similarly, Quesada (2025) interprets AI Thinking as a meaning-centered framework, primarily applying AI principles to enhance language technology development and cultural preservation through community agency. While these frameworks provide essential guidance on how to use or design AI systems, their primary focus remains on operational competencies and technical development methodologies. In contrast to these tool-centric or development-centric models, the AI Thinking proposed here approaches the issue from deeper cognitive and philosophy-of-science dimensions, addressing a more fundamental question: as AI evolves from a tool into a cognitive partner, how should human thinking, research paradigms, and knowledge-production logics be restructured? AI is not treated as an external instrument but embedded within the human cognitive loop, with the human-AI relationship understood as a symbiotic, bidirectionally calibrated co-cognitive practice. From this perspective, AI Thinking is not a checklist of skills but a systematic meta-cognitive paradigm: it concerns how humans guide intentions under uncertainty, elicit emergent cognition through dialogue, integrate machine intelligence through critical reflection, and ultimately reconstruct the subjectivity and meaning systems of scientific exploration through human-AI co-cognition.

## 2.1 Definition and Core: From Instruction Execution to Collaborative Guidance

Before constructing the theoretical framework of AI Thinking, it is essential to precisely define the concept. The primary interactive object of AI Thinking is not a conventional computational device, but a generative AI agent serving as a cognitive partner. While classical definitions describe an agent as “a system

that perceives its environment and acts autonomously to achieve its designed objectives” (Wooldridge & Jennings, 1995), in the context of this paper, we emphasize its generative, dialogic, and heuristic role within human-AI collaboration (Floridi & Chiriatti, 2020). AI is not merely an executor of instructions but a cognitive collaborator capable of stimulating human reflection and expanding the space of possibilities.

Based on this perspective, we formally define AI Thinking as follows: *AI Thinking is a systematic meta-cognitive paradigm that, through structured intent modeling, contextual anchoring, and generative dialogue, transforms ambiguous human intentions, complex domain knowledge, and open-ended problems into cognitive tasks that AI agents can effectively respond to. Within dynamic human-AI interaction loops, AI Thinking achieves deep collaboration between human intuition and value judgments and the breadth and generative capabilities of machine knowledge through critical integration and meaning construction.*

This definition underscores that the essence of AI Thinking lies not in “operating AI” but in constructing a human-centered cognitive enhancement loop, mediated by dialogue and oriented toward co-creating meaning. Its core challenge has shifted from the computational thinking era’s question of “how to translate a problem into an executable algorithm” to “how to design effective human-AI co-cognitive processes.” To clearly illustrate this paradigm shift, Table 1 systematically compares AI Thinking with computational thinking across multiple dimensions, highlighting their fundamental differences.

**Table 1. Paradigm Comparison between Computational Thinking and AI Thinking**

Dimension	Computational Thinking	AI Thinking
Core Object	Computer (deterministic machine)	AI agent (probabilistic cognitive partner)
Core Actions	Abstraction, decomposition, algorithmization	Intent modeling, generative dialogue, critical integration
Relationship Pattern	Human commands machine (hierarchical)	Human-AI orchestration, bidirectional calibration
Output Nature	Deterministic, predictable solutions	cognitive loop Non-deterministic, evaluative “options” or “drafts”

---

Dimension	Computational Thinking	AI Thinking
Core Cognitive Challenge	Translating problems into computable models	Translating intentions into guidable dialogue flows

---

**Theoretical Implications and Distinctions:** The shift involves moving from handling determinism to managing uncertainty. Computational thinking assumes fully predictable behavior; AI Thinking requires identifying value within probabilistic outputs. It transitions from “designing precise instruction sequences” to “guiding creative generation and meaning construction.” The former emphasizes error-free execution; the latter prioritizes cognitive emergence and value alignment. It shifts from unidirectional commands to bidirectional cognitive calibration where humans guide AI, and AI feedback stimulates human reflection, forming a collaborative loop. The challenge moves from clearly defined problem-solving to eliciting, selecting, and reconstructing potential solutions within the space of possibilities. AI Thinking is not a mere extension of computational thinking; it is a new meta-cognitive paradigm oriented toward generativity, interaction, and uncertainty. It requires researchers to transform from “program designers” to “architects of human-AI intelligent teams,” fostering the co-emergence of knowledge and the collaborative construction of meaning through ongoing dialogue with AI. This transition signals a new civilizational stage in which human-AI relationships move beyond mere tool usage toward intelligent symbiosis.

## 2.2 Three Core Pillars: Constructing a Human-AI Co-Cognitive Paradigm

To transform AI Thinking from an abstract concept into a learnable, practicable, and transmissible cognitive capability, this paper proposes a theoretical framework built upon three core pillars. These pillars are not isolated skill modules but collectively sustain a dynamic, recursive, and reflexive human-AI co-cognitive system. They correspond to the three central dimensions of scientific rationality and creative practice in human-AI orchestration: goal setting (What to pursue), process facilitation (How to explore), and value judgment (Whether it matters). The theoretical foundation of this framework draws on cognitive science insights into meta-cognition (Flavell, 1979), Vygotsky’s sociocultural theory and zone of proximal development (Vygotsky, 1978), and situated learning theory (Lave & Wenger, 1991). Unlike Newman-Griffis’ s (2025) methodological framework focused on tool usage, or Quesada’ s (2025) development framework focused on system design and cultural preservation, the framework proposed here does not focus on “how to operate or build AI.” Instead, it asks: how can humans reconstruct their cognitive logic and scientific subjectivity through symbiosis with AI?

**Pillar 1: Intention Modeling and Contextual Anchoring** serves as the foundation for effective human-AI interaction. It requires users to translate vague internal intentions, needs, or curiosities into precise prompts that AI can perceive and process. The core competency is context construction. Situated cognition theory posits that human cognitive activities are deeply contingent on the specific contexts in which they occur (Brown, Collins, & Duguid, 1989). Knowledge is not an abstract set of symbols but is co-constructed through participation and interaction in social practices and task environments (Clancey, 1997). For AI, a carefully crafted context functions like a theatrical stage, complete with roles and backgrounds, significantly constraining and guiding the direction and quality of generated content. Users provide key background information, relevant data, or contextual cues to situate tasks within a rich semantic network, enhancing the relevance and depth of AI responses. Assigning AI a specific professional or social identity (e.g., “a senior molecular biologist” or “a critical product reviewer” ) effectively activates relevant domain knowledge, language style, and knowledge scope. Research indicates that role prompts, as contextual cues, can significantly steer large language models to access domain-specific knowledge, adjust linguistic style, and enhance output precision and relevance (Shanahan et al., 2023).

**Pillar 2: Generative Dialogue Facilitation** recognizes that AI Thinking is not limited to one-off Q&A interactions but constitutes a dynamic, iterative, and highly strategic guidance process. Its goal is to leverage structured human-AI dialogue to activate latent, dispersed associations within the model’s knowledge system, thereby eliciting cognitive emergence that is novel and insightful for humans. This process is not a unidirectional instruction; it constitutes a bidirectional, co-cognitive practice where humans continuously adjust the AI’s generative trajectory through questioning, reconstruction, role-switching, and counterfactual inquiry, enabling the AI to surpass the limitations of initial responses. This mechanism resonates with Vygotsky’s (1978) zone of proximal development: just as a teacher scaffolds a learner to achieve understanding unattainable independently, humans can provide temporary scaffolding to AI via structured prompts, enabling higher-order reasoning and integrative capacities. Chain-of-thought prompting plays a crucial role: when users request AI to “reason step by step” or “show the thinking process,” the model externalizes its implicit generative logic into traceable steps (Wei et al., 2022). This externalization enhances logical coherence and transforms the “black-box” output into a transparent, intervenable reasoning path, making AI a genuine collaborative cognitive partner.

**Pillar 3: Critical Integration and Meaning-Making** ensures that AI-generated outputs are reliable, ethically sound, and socially valuable in human-AI co-cognitive practice. Despite the generative power of large language models, their inherent limitations—including factual hallucinations, logical inconsistencies, and biases inherited from training data—render AI output non-authoritative. Therefore, users must maintain high meta-cognitive vigilance, systematically reviewing and creatively reconstructing AI content. This process

encompasses four levels: (1) source verification, (2) logical scrutiny, (3) value alignment assessment (evaluating whether outputs conform to ethical norms, professional standards, and social welfare), and (4) meaning construction (treating AI-generated alternatives as cognitive raw material, which humans refine through domain expertise, insight, and creativity to produce original and responsible outcomes). Here, the irreplaceable role of humans is evident: AI generates options, but humans assign meaning. As Floridi & Chiriatti (2020) note, in the era of human-AI symbiosis, interpretation remains a uniquely human cognitive prerogative.

Collectively, these three pillars form a complete, spiral, and ascending cognitive loop (Figure 1 [Figure 1: see original paper]). Contextual anchoring initiates precise interaction, generative dialogue elicits rich content and possibilities, and critical integration ensures quality and real-world value. When addressing complex problems, researchers cycle dynamically among the three pillars: initial intentions are refined through dialogue, newly emergent solutions prompt deeper value reflection, and critical insights drive new exploratory directions. This dynamic interplay fundamentally distinguishes AI Thinking from the linear “input-process-output” logic of traditional computational thinking.

**Figure 1** Three Pillars of the Human-AI Collaborative Cognitive Paradigm

### 3 Cognitive Practice of AI Thinking: Cross-Domain Demonstration of Human-AI Co-Cognition

This chapter demonstrates how AI Thinking, as a meta-cognitive paradigm, constructs human-AI co-cognitive loops in real-world problem-solving through in-depth case studies across several representative domains. Moving beyond the operational focus of previous frameworks (Newman-Griffis, 2025; Quesada, 2025), these cases illustrate how humans reconstruct their cognitive pathways, judgment logic, and modes of creativity through symbiosis with AI. Each case strictly adheres to the three core pillars established in Chapter 2—Intention Modeling and Contextual Anchoring, Generative Dialogue Facilitation, and Critical Integration and Meaning-Making—highlighting their nonlinear, feedback-driven, and spiral dynamics. These cases collectively illustrate that AI Thinking is not merely a methodology but represents an ongoing cognitive revolution.

The cases presented in this chapter are drawn partly from real research practices and partly from hypothetical reconstructions based on typical workflows. They are designed to systematically showcase the applicability and methodological value of the AI Thinking framework across different domains. All professional logic, workflow structures, and technical constraints in the cases are referenced from authoritative literature and industry standards to ensure internal consistency and representativeness. It should be noted that any specific numerical values, parameters, or technical details presented in the cases are for illustrative purposes only. They do not constitute empirical claims about actual systems or experimental outcomes, and the authors do not assume academic responsibility

for the accuracy or reproducibility of these specific data points.

### 3.1 Case 1: Scientific Research -From “Hypothesis-Driven”to “Human-AI Co-Constructed Hypotheses”

In the field of materials science, traditional research paradigms are time-consuming and entail high trial-and-error costs. By applying AI Thinking, researchers can construct an enhanced loop of literature mining, hypothesis generation, and experimental design.

**1. Intention Modeling and Contextual Anchoring:** Instead of posing a vague question such as “Are there new materials?” , researchers perform structured intention clarification. For example: “You are an expert in solid-state physics and materials informatics. Based on high-temperature superconductors reported in the past five years in journals such as *Nature Materials* and *Advanced Functional Materials*, extract common features including crystal symmetry, electronic density of states, and phonon spectra. Using this information, screen the Materials Project database for theoretically feasible non-copper candidate materials that have not yet been experimentally verified. Prioritize systems that are stable at ambient pressure and have clear synthetic pathways.”

**2. Generative Dialogue Facilitation:** Researchers do not rely solely on the AI’ s first-round suggestions but initiate multi-turn strategic guidance through recursive decomposition ( “Please evaluate the top 5 candidate materials along three dimensions: structural stability, electron pairing mechanism, and tolerance to lattice distortions” ), counterfactual inquiry ( “If lanthanides are replaced with actinides, how would the predicted Tc (superconducting transition temperature) change?” ), and red-teaming challenge ( “Assume the role of a skeptical reviewer and identify the most likely physical reasons for failure among these candidates” ). During this process, cognitive emergence occurs: the AI not only synthesizes existing knowledge but, through cross-literature pattern matching, proposes a novel hypothesis—e.g., “CaFeAs<sub>2</sub> may exhibit s± pairing under high pressure” —which is not present in current literature yet is theoretically coherent.

**3. Critical Integration and Meaning-Making:** Researchers do not directly adopt AI suggestions but instead perform source verification (validating candidate materials’ electronic structures and phonon instabilities through first-principles DFT calculations), logical scrutiny (identifying that AI overestimated synthetic feasibility by ignoring rapid oxidation in air), and value alignment assessment (considering computational cost, experimental risk, and scientific novelty to select two candidates for laboratory synthesis). Ultimately, human scientists transform AI-generated “cognitive raw materials” into publishable scientific hypotheses, explicitly noting in publications that “The hypotheses were inspired by multi-turn exploratory dialogue with AI, but their physical validity was independently verified and interpreted by human researchers.”

### 3.2 Case 2: Engineering and Design -From “Computer-Aided” to “AI Co-Creation”

Traditional engineering design relies on a linear workflow of human conceptualization, computer modeling, and simulation verification, where AI functions merely as a computational accelerator. AI Thinking advances the engineering paradigm toward generative co-design, in which AI is not only an optimization engine but also a “possibility detector” that expands human imagination, while humans ascend from detailed executors to architects of design intent and ultimate judges of value and meaning. In complex domains such as chip design, AI Thinking is reshaping the conventional iterative process of design-simulation-verification.

**1. Intention Modeling and Contextual Anchoring:** Designers no longer issue vague directives like “optimize chip layout,” but perform multi-dimensional, structured intention clarification: “You are an engineer specialized in high-performance computing chip architecture. Please generate layout schemes under the following constraints: (1) chip area  $250mm^2$ ; (2) peak performance  $\$120$  TFLOPS; (3) TDP  $\$250W$ ; (4) signal latency  $\$2ns$ ; (5) prioritize 3nm process technology. Output the layout topology in IEEE-standard format and provide a power-performance-area (PPA) trade-off analysis.”

**2. Generative Dialogue Facilitation:** Designers engage in multi-turn strategic dialogue with AI to explore the design space more deeply through Pareto front exploration ( “Generate 10 candidate solutions evenly distributed across the three PPA dimensions and indicate their positions on the Pareto front” ), counterfactual stress testing ( “If the key IP core area increases by 15%, re-optimize the layout and assess the impact on timing and power consumption” ), and red-teaming challenge ( “Assume the role of a chip reliability expert and identify physical structures most likely to fail under high-temperature aging tests” ). During this process, AI exhibits cognitive emergence: it does more than route layouts; through cross-constraint coupling analysis, it proposes novel configurations, such as placing high-power computing units directly above heatsinks to improve thermal conductivity.

**3. Critical Integration and Meaning-Making:** Faced with dozens of technically feasible options generated by AI, human designers exercise final value judgment through manufacturability review (excluding layouts violating DFM rules using foundry PDKs), architectural elegance assessment (discarding “black-box” designs with slightly better performance but chaotic routing and difficult debugging in favor of modular, clean, iterative-friendly layouts), and long-term strategic alignment (prioritizing microarchitectures compatible with next-generation AI accelerator instruction sets, even if current PPA performance is slightly inferior). Ultimately, humans transform AI-generated technical options into design outputs that embody engineering expertise and strategic foresight. As Cross (2001) observes, “The essence of design thinking is not merely solving problems but defining ‘ good problems within constraints and giving them form.” AI

Thinking amplifies this essential capability in human-AI co-creation.

### 3.3 Case 3: Educational Innovation -From “Knowledge Transmission” to “Human-AI Co-Constructed Inquiry”

Traditional education is teacher-centered, emphasizing systematic knowledge delivery and the acquisition of standard answers. Even with the rise of constructivist approaches, student inquiry remains constrained by teacher-defined problem boundaries and available resources. AI Thinking introduces a paradigm-level shift in education, transforming AI from an “answer provider” into a “cognitive dialogue partner,” and shifting the learning process from “verifying the known” to “exploring the unknown,” and from individual construction to human-AI co-construction.

**1. Intention Modeling and Contextual Anchoring:** Teachers guide students away from vague questions like “Why did Rome decline?” and toward structured intention clarification and role anchoring: “Assume the role of a historian with in-depth knowledge of the late Roman Empire’ s fiscal system. Based on authoritative literature from the past twenty years, such as *Journal of Roman Studies* and *Ancient History Review*, analyze how Diocletian’ s tax reforms, by increasing local burdens, indirectly weakened the empire’ s capacity for military mobilization against barbarian invasions. Pay particular attention to the causal chain between provincial fiscal records and army supply data.”

**2. Generative Dialogue Facilitation:** Students do not rely solely on the AI’ s initial explanation but initiate multi-turn, strategic dialogue to stimulate cognitive emergence through counterfactual inquiry ( “If Constantine had not moved the capital eastward, how would fiscal pressure in the West change? Would AI’ s explanation still hold?” ), multi-perspective simulation ( “Describe the experience and attitude toward the same tax system from the perspectives of senatorial elites, provincial farmers, and frontier legion soldiers” ), and evidence challenge ( “You cited Zosimus’ *New History*, which is widely regarded as strongly anti-Christian. Does this bias affect your assessment of the fiscal collapse?” ). In this process, AI serves not merely as an information source but as a Socratic interlocutor, offering competing interpretations, exposing argumentative gaps, and supplementing marginal historical sources, thereby forcing students to continuously revise hypotheses and deepen reasoning.

**3. Critical Integration and Meaning-Making:** Faced with diverse and sometimes contradictory AI-generated outputs, students exercise ultimate academic judgment and meaning attribution through source verification (checking whether AI-cited sources, such as the “Late Roman Empire Fiscal Archives,” exist—in this case, fictitious—and replacing them with authoritative references like *The Cambridge Ancient History*), logical scrutiny (identifying misinterpretations of correlation as causation, e.g., linking plague outbreaks directly to tax increases), value reflection (questioning whether AI’ s characterization of “barbarians” implies a civilizational hierarchy and correcting it using postcolonial

historiographical perspectives), and creative synthesis (integrating AI-provided fiscal-military-cultural data into a cohesive analysis report, clearly distinguishing AI-assisted content from independently reasoned sections). Through adversarial dialogue with AI, students are trained to become more critical and responsible knowledge constructors. As Ennis (1987) notes, the core of critical thinking is the tendency and ability to reflect reasonably on beliefs and actions. AI Thinking provides an unprecedented training ground for this capacity.

### 3.4 Case 4: Business Decision-Making -From “Predictive Optimization” to “Human-AI Co-Constructed Resilience”

Traditional business analytics relies on historical data modeling and linear extrapolation, aiming to predict the “most likely future.” The emergence of generative AI transforms strategic planning from a predict-and-respond approach to an explore-and-prepare paradigm. Through human-AI collaboration, organizations can proactively construct multiple possible future scenarios and deploy mitigation strategies in advance.

**1. Intention Modeling and Contextual Anchoring:** Executive teams avoid vague instructions such as “analyze market risk” and instead perform structured strategic intention clarification: “You are a Chief Risk Officer (CRO) with global supply chain management experience. Based on current geopolitical tensions, Red Sea shipping disruptions, rising labor costs in Southeast Asia, and other real-world constraints, simulate potential cascading disruption scenarios over the next 12-18 months for our three core consumer electronics supply chains: chips, batteries, and modules. Assess the impact of each scenario on capacity, cost, and customer delivery, and propose preliminary mitigation strategies. All assumptions should indicate data sources and confidence levels.”

**2. Generative Dialogue Facilitation:** Managers do not settle for conventional AI risk lists but initiate high-intensity, cognitively adversarial guidance through red-teaming stress tests ( “Assume the role of our largest competitor and design a targeted supply chain attack on our Southeast Asian production bases” ), black swan probing ( “Assume a sudden blockade of the Taiwan Strait coinciding with a global lithium export ban; simulate the cascading effects” ), and multi-stakeholder simulation ( “Evaluate the acceptability of each mitigation plan from the perspectives of investors, frontline workers, and environmental NGOs” ). During this process, AI exhibits emergent strategic imagination: it not only reproduces known risks but, through cross-domain association, proposes novel strategies such as nearshore manufacturing in Mexico combined with direct cobalt sourcing in Africa, and identifies gaps such as ESG compliance systems failing to cover critical mineral traceability, which may become a future sanction vulnerability—insights beyond the scope of traditional risk models.

**3. Critical Integration and Meaning-Making:** Faced with dozens of AI-generated possible futures, executive teams exercise strategic value judgment through feasibility filtering (excluding options that rely on cross-border data

agreements not yet legislated), ethical consistency review (rejecting “fully automated” options that, while cost-minimizing, could cause mass layoffs), resilience prioritization (selecting strategies that, although slightly higher in short-term cost, enhance system redundancy and modularity), and organizational learning conversion (transforming AI-simulated “failure scenarios” into internal crisis drills to strengthen organizational anti-fragility). Ultimately, human executives transform AI-generated possibility spaces into responsible strategic actions. This case demonstrates that AI Thinking elevates business decision-making from seeking optimal solutions to constructing anti-fragile systems.

## **4 The AI-Driven Transformation of Scientific Paradigms: Toward a New Civilization of Human-Machine Co-Cognition**

The significance of AI Thinking extends far beyond improving individual problem-solving efficiency. It is catalyzing a profound and largely silent transformation of science itself—not at the level of specific disciplines, but at the level of epistemic foundations, collaborative structures, and cognitive agency. Where Kuhn’s (1962) paradigm shifts described changes in what scientific communities collectively regard as legitimate problems and valid solutions, AI Thinking introduces a more radical transformation: the inclusion of non-human intelligence as an epistemic participant in scientific knowledge production. As a result, scientific methodology is being augmented, disciplinary cultures are being reorganized, and the role of the scientist is being fundamentally redefined—pointing toward a new form of civilization grounded in human-AI co-evolutionary cognition.

### **4.1 Enhancing Scientific Methodology: From “Hypothesis-Validation” to a Human-AI Co-Creation Cognitive Loop**

Traditional scientific methodology is often summarized as a linear or weakly iterative sequence of observation–hypothesis–experiment–validation. This paradigm rests on an implicit epistemological assumption: the creative origin of scientific knowledge resides primarily, if not exclusively, within the human mind—through intuition, experience, and theoretical imagination. Within this framework, computational tools and algorithms are typically regarded as auxiliary instruments, tasked mainly with data processing, statistical analysis, or the execution of pre-defined models. The rapid advancement of artificial intelligence, particularly large-scale and generative models capable of high-dimensional representation and cross-domain reasoning, fundamentally challenges this human-centered epistemic structure. Rather than merely accelerating existing research workflows, AI reshapes the logic of scientific discovery itself, reconstructing it as an enhanced human-AI co-creative cognitive loop (Figure 2 [Figure 2: see original paper]).

**Figure 2** From hypothesis-validation to a human-AI co-creation cognitive loop

in scientific discovery

Within this loop, AI no longer functions as a passive analytical tool but emerges as an active hypothesis space expander and generative reasoning partner. In contemporary scientific domains characterized by high-dimensional, high-throughput, and strongly nonlinear data—such as genomics, materials science, and hyperspectral remote sensing—the limitations of human intuition in pattern recognition and associative discovery have become increasingly apparent. Leveraging large parameter spaces and powerful representation learning capabilities, AI systems can explore latent associations across heterogeneous datasets and scientific literature, generating non-intuitive yet potentially meaningful candidate hypotheses that would be difficult for human researchers to anticipate independently. Importantly, this transformation does not imply a simple transfer of scientific creativity from humans to machines. Instead, the role of the human scientist undergoes a critical reconfiguration—from the sole originator of hypotheses to a modeler of scientific intent, a provider of contextual grounding, and an arbiter of epistemic value. Through intention modeling and contextual anchoring, scientists translate vague scientific curiosities (e.g., “Is room-temperature superconductivity possible?”) into structured problem spaces that guide AI-driven generative exploration. Iterative interactions with AI systems then progressively expose latent research trajectories that extend beyond the boundaries of prior human conceptualization. At its core, this methodological shift transforms scientific discovery from an episodic “moment of human insight” into a systematic process of collaborative emergence. As observed by Wang et al. (2023) in *Nature*, AI systems are increasingly moving beyond their traditional role as analytical instruments toward deeper participation in the scientific discovery process. This shift is not merely rhetorical but reflects an emerging epistemic reality. In frontier areas such as protein structure prediction (e.g., AlphaFold), generative materials design, and AI-assisted mathematical conjecturing, the incremental creativity of scientific knowledge increasingly arises at the human-AI cognitive interface, rather than from either agent operating in isolation.

#### 4.2 Expanding Scientific Culture: From “Two Cultures” to a Human-AI “Fourth Culture”

C. P. Snow’s (1959) diagnosis of a divide between the sciences and the humanities has long constrained the ability of research communities to address complex systemic problems. Although Brockman’s “third culture” (1995) brought scientists into public intellectual discourse, it did not fundamentally alter the internal epistemic segmentation of research practice. AI Thinking, centered on intent abstraction and contextual reconfiguration, is now enabling what may be described as a “fourth culture”: a mode of inquiry in which natural science, social science, and humanistic reasoning are dynamically integrated through AI-mediated cognitive role-switching. For example, a research team studying urban sustainability may iteratively prompt an AI system to function as: (1) a climate

modeler, simulating carbon-emission trajectories under alternative policies; (2) a social network analyst, evaluating equity and public acceptance; (3) an economic analyst, estimating costs, employment, and growth; and (4) a science communicator, translating results into narratives intelligible to policymakers and the public. This fluid role-switching forces researchers to engage with multiple epistemic standards, evidentiary logics, and representational forms. The result is a new scientific culture oriented toward systemic, cross-paradigmatic problem-solving rather than disciplinary optimization. In this sense, AI Thinking constitutes a digital extension of Vygotsky's (1978) sociocultural theory: cognition becomes distributed not merely across people, but across human-AI-data-tool ecologies.

### 4.3 Redefining the Scientist: From “Lone Genius” to Architect of Human-AI Intelligence Teams

As AI becomes deeply embedded in scientific practice, the scientist's core identity is shifting from that of a technical operator to that of a cognitive architect. The primary source of scientific value is no longer the possession of specialized knowledge—much of which AI can now retrieve, model, and recombine—but rather a form of meta-cognitive capacity: the ability to design, orchestrate, and govern human-AI cognitive systems. In this emerging paradigm, scientists function as architects of human-AI intelligence teams with three central responsibilities: (1) intent modeling and task architecture (translating frontier scientific questions into structured, decomposable, and iteratively refinable human-AI workflows); (2) interaction protocol design (constructing feedback loops and quality-control mechanisms linking human intuition, AI generation, and empirical validation); and (3) meaning arbitration and value anchoring (selecting, interpreting, and normatively evaluating among AI-generated possibilities based on scientific judgment, ethical reasoning, and societal purpose). This transformation shifts scientific work from the operational layer (running experiments, coding models) toward the architectural and interpretive layers—designing cognitive workflows, curating epistemic possibilities, and assigning meaning and value. Scientists no longer monopolize knowledge production; rather, they define the rules, goals, and evaluative frameworks through which human-AI systems explore the unknown. While scholars like Shneiderman (2022) warn against the “teammate” metaphor to preserve human responsibility and control, we use the term “cognitive partner” specifically to describe the interactive dynamics (generative dialogue) rather than the attribution of responsibility. As emphasized in our third pillar, humans remain the ultimate architects and judges of meaning, ensuring that the “partnership” enhances rather than diminishes human agency.

## 5 Conclusion and Outlook

We stand at a historical inflection point: artificial intelligence, particularly its generative and reasoning capabilities, is no longer a distant futuristic vision but a concrete force reshaping contemporary research and practice paradigms. The

disruptive upgrade of tools necessarily demands a corresponding transformation in human thinking. This paper systematically argues that to meet the challenges of this era, we must construct and disseminate a new cognitive paradigm—AI Thinking. This study offers a systematic articulation of AI Thinking as a meta-cognitive framework, structured around three core pillars: Intention Modeling and Contextual Anchoring, Generative Dialogue Facilitation, and Critical Integration and Meaning-Making. Together, they form a dynamic, iterative human-AI co-cognitive loop. Cross-domain cases demonstrate that this framework not only empowers scientific research, engineering design, educational innovation, and business decision-making, but also drives a silent revolution in scientific paradigms: (1) the scientific method evolves from the linear hypothesis-verification loop to a higher-order loop of data-AI-hypothesis-verification; (2) the role of scientists shifts from the “solitary expert” to architects of human-AI intelligent teams; and (3) knowledge production moves from human solo performance to human-AI co-orchestration.

Of course, this path is not without challenges. It entails serious issues such as cognitive dependence, interpretability, algorithmic bias, and responsibility attribution. Yet confronting these challenges and actively seeking interdisciplinary solutions exemplifies the critical integration spirit inherent in AI Thinking. In this context, while frameworks like Quesada’s (2025) crucially remind us of the cultural responsibilities in AI development, and Newman-Griffis’s (2025) work standardizes scientific AI usage, our “AI Thinking” complements these by providing the necessary epistemological infrastructure—the meta-cognitive “operating system” for the human mind in the intelligent era.

Despite these challenges, the trajectory of AI Thinking is becoming increasingly clear, and its future development is likely to follow three major trends: (1) **Systematic integration into education:** AI Thinking, like computational thinking, will gradually be embedded across all levels of education—from higher education to primary schooling and lifelong learning, becoming a core competency for training future scientists, engineers, and decision-makers. AI Thinking education will no longer be confined to computer science but will permeate all STEM fields and even humanities and social sciences, influencing both teaching and research practices. (2) **Paradigmatic platform development:** Specialized Collaborative Cognition Operating Systems (CCOS) will emerge to support AI Thinking practices. Such platforms will integrate multi-model scheduling, dialogue history analysis, automated validation, ethical review, and meaning-making tools, making complex AI Thinking practices more fluent, traceable, and teachable. (3) **Deep evolution toward general artificial intelligence:** As research on AGI (Artificial General Intelligence) progresses, AI Thinking may ultimately evolve into a “cognitive interaction language” between humans and intelligent agents with general cognitive abilities. At that point, human-AI collaboration will enter a new era of deep collaborative science, jointly understanding abstract concepts, engaging in philosophical reasoning, and even setting meta-goals for scientific exploration.

## References

- Bansal, G., Wu, T., Zhou, J., Fok, R., Nouri, B., Holbrook, L., ...& Weld, D. S. (2021). Does the whole exceed its parts? The effect of AI explanations on trust, reliance, and cognitive engagement. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI ' 21)*, Article 385, 1-16.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT ' 21)*, 610-623.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ...& Liang, P. (2021). On the opportunities and risks of foundation models (arXiv:2108.07258).
- Brockman, J. (Ed.). (1995). *The third culture: Beyond the scientific revolution*. Simon & Schuster.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32-42.
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Jain, E., ... & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4 (arXiv:2303.12712).
- Clancey, W. J. (1997). *Situated cognition: On human knowledge and computer representations*. Cambridge University Press.
- Cross, N. (2001). Designerly ways of knowing: Design discipline versus design science. *Design Issues*, 17(3), 49-55.
- Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9-26). W. H. Freeman.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906-911.
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681-694.
- Gray, J. (2007). Jim Gray on eScience: A transformed scientific method. Microsoft Research.
- Jakesch, M., Zhang, H., Hancock, J. T., Naaman, M., & Diakopoulos, N. (2023). Co-writing with opinionated language models affects users' views. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI ' 23)*, 1-15.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.

- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.
- Newman-Griffis, D. (2025). AI Thinking: A framework for rethinking artificial intelligence in practice. *Royal Society Open Science*, 12(1), 241482.
- Park, J. S., O' Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST' 23)*, 1-22.
- Quesada, J. F. (2025). AI Thinking as a Meaning-Centered Framework: Reimagining Language Technologies Through Community Agency. arXiv preprint arXiv:2502.14923.
- Reynolds, L., & McDonell, K. (2021). Prompt programming for large language models: Beyond the few-shot paradigm. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA ' 21)*, 1-7.
- Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Llama Team, ...& Scialom, T. (2024). Toolformer: Language models can teach themselves to use tools. *Advances in Neural Information Processing Systems*, 36, 3393-3408.
- Snow, C. P. (1959). *The two cultures and the scientific revolution*. Cambridge University Press.
- Shanahan, M., McDonell, K., & Reynolds, L. (2023). Role play with large language models. *Nature*, 623(7986), 493-498.
- Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press. (Original works published 1930-1935).
- Wang, H., Liu, Y., Li, Z., Zhang, R., Li, J., Xu, C., & Wang, Y. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972), 47-60.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E. H., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35, 24824-24837.
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33-35.
- Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, 10(2), 115-152.
- Xi, Z., Chen, W., Guo, X., Zhang, W., Liu, Y., Tang, Y., ...& Liu, Y. (2025). The rise and potential of large language model based agents: A survey. *Science China Information Sciences*, 68(2), 121101.

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv – Machine translation. Verify with original.*