

Ensuring Human Authorship in the Age of Generative AI: The FERC Framework for Governed Human–AI Collaboration

Jacob Sherson*
Department of Management
Aarhus University
Aarhus, Denmark
sherson@mgmt.au.dk

Janet Rafner
Department of Business and
Management
University of Southern Denmark
Odense, Denmark
Centre for Digital and Green
Transformation in Cities and
Communities (DIGTCOM)
Aarhus University
Aarhus, Denmark
Danish Institute for Advanced Studies
University of Southern Denmark
Odense, Denmark
jraf@sam.sdu.dk

Roni Reiter-Palmon
Psychology
University of Nebraska at Omaha
Omaha, Nebraska, USA
rreiter-palmon@unomaha.edu

Izabela Lebuda
University of Wrocław
Wrocław, Poland
izabela.lebuda@uwr.edu.pl

Frederik Brosbøl Kjeldsen
Center for Hybrid Intelligence
Aarhus University
Aarhus, Denmark
fbk@mgt.au.dk

Matthias Söllner
Information Systems and Systems
Engineering
University of Kassel
Kassel, Germany
soellner@uni-kassel.de

Yoed Kenett
Faculty of Data and Decision Sciences
Technion - Israel Institute of
Technology
Haifa, Israel
yoedk@technion.ac.il

Blerim Emruli
Department of Informatics
Lund University
Lund, Sweden
blerim.emruli@ics.lu.se

Manuel Rindle
University of Hildesheim
Hildesheim, Germany
manuel.rindle@gmail.com

Selina Weiss
University of Hildesheim
Hildesheim, Germany
selina.weiss@uni-hildesheim.de

Benjamin Goecke
University of Tübingen
Tübingen, Germany
benjamin.goecke@uni-tuebingen.de

Florent Vinchon
Université Paris Cité - LAPEA
Paris, France
florent.vinchon@gmail.com

Yaoli Mao
Autodesk
San Francisco, California, USA
ym2429@tc.columbia.edu

Dominik Dellermann
vencortex
Nuremberg, Germany
dominik@vencortex.io

Seyedahmad Rahimi
University of Florida
Gainesville, Florida, USA
srahimi@ufl.edu

Janet H. Marler
University of Albany
Albany, New York, USA
jmarler@albany.edu

Andy Nguyen
Learning & Educational Technology
Research Unit (LET)
University of Oulu
Oulu, Oulu, Finland
andy.nguyen@oulu.fi

Jens Christian Bjerring
Aarhus University
Aarhus, Denmark
filjcb@cas.au.dk

Steve DiPaola
Sch of Interactive Art and Tech
Simon Fraser University
Vancouver, British Columbia, Canada
sdipaola@sfu.ca

Abstract

Generative AI has rapidly transformed professional knowledge work, yet prevailing prompt engineering paradigms frame effective human–AI interaction primarily as a matter of crafting better prompts. Even emerging multi-round optimization approaches tend to prioritize performance, consistency, and task completion rather than authorship preservation, evaluative reframing, and accountable decision-making. This paper introduces FERC (Frame–Explore–Refine–Commit), a governed interaction cycle for multi-shot human–AI collaboration that structurally safeguards human intent, judgment, and responsibility. Drawing on creativity research, particularly models of iterative problem construction and co-evolution between generation and evaluation, FERC operationalizes metacognitive control in hybrid intelligence systems. We demonstrate how FERC reframes the Double Diamond design model for AI collaboration, provide practical behavioral guidelines, and introduce a Human–AI Collaboration Maturity Model tested multiple cross multiple executive training and educational contexts. Finally, we present FERC-bot, a metacognitive training interface that makes collaboration structures visible and measurable. To date, FERC has been applied in more than 20 organizations and with +2,000 individuals. Because the framework renders collaboration quality observable and scoreable at the process level, we outline two pathways for its formal operationalization: process-based assessment in formal education, developed at length in the HI Educational Manifesto, and auditable documentation of human contribution in professional work, from intellectual-property provenance in AI-assisted innovation to organizational human-touch accountability. Collectively, FERC is introduced as the cross-cutting interaction strand of the HI Manifesto series. Together, these contributions reposition prompt engineering from artifact optimization toward process governance, enabling scalable and authorship-preserving hybrid intelligence in educational and organizational settings.

*Corresponding author



This work is licensed under a Creative Commons Attribution 4.0 International License.
NordiCHI '26 Adjunct, Vaasa, Finland
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2798-6/2026/10
<https://doi.org/10.1145/3821402.3830141>

CCS Concepts

• **Human-centered computing** → Human computer interaction (HCI); Human computer interaction (HCI); HCI theory, concepts and models; • **Computing methodologies** → Artificial intelligence.

Keywords

genAI, authorship, ownership, collaboration, framework, problem solving, problem identification

ACM Reference Format:

Jacob Sherson, Janet Rafner, Roni Reiter-Palmon, Izabela Lebeda, Frederik Brosbøl Kjeldsen, Matthias Söllner, Yoed Kenett, Blerim Emruli, Manuel Rindle, Selina Weiss, Benjamin Goecke, Florent Vinchon, Yaoli Mao, Dominik Dellermann, Seyedahmad Rahimi, Janet H. Marler, Andy Nguyen, Jens Christian Bjerring, and Steve DiPaola. 2026. Ensuring Human Authorship in the Age of Generative AI: The FERC Framework for Governed Human–AI Collaboration. In *Adjunct Proceedings of the 14th Nordic Conference on Human-Computer Interaction (NordiCHI '26 Adjunct)*, October 03–07, 2026, Vaasa, Finland. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3821402.3830141>

1 Introduction

Generative AI (GenAI) systems have rapidly become embedded in professional knowledge work, promising substantial productivity gains as large language models (LLMs) draft, analyze, and synthesize across domains [1]. Alongside this adoption, a growing literature documents costs at the cognitive and skill level: reliance on generative systems can reduce independent reasoning, alter problem-solving strategies, and increase dependence on external generation [2], echoing older concerns that technology use erodes decision-making capacity and domain expertise [3]. The integration of GenAI into everyday workflows may therefore change not only how work is performed, but how expertise is developed and maintained.

These effects are unevenly distributed: GenAI often benefits less-experienced users more than experts [4, 5], which, if novices can approximate expert outputs without acquiring the underlying expertise, gives organizations a rational incentive to underinvest in training. This is the automation–augmentation paradox in acute form: the same system can substitute for human capability or amplify it, and short-term incentives favor substitution even where augmentation is more valuable long-term [6]. Left unmanaged, the paradox resolves toward systemic deskilling, a reduction in the

economic value and development of human expertise as production substitutes learning with generation.

This paper develops its response within the HI Manifesto series, a recent expansion of the automation–augmentation paradox that advances Hybrid Intelligence (HI) as a multi-level socio-technical transition model. The main HI Manifesto [7] grounds human–AI complementarity in a dynamic frontier between machine prediction and human judgment, determined by which contextual knowledge has been made operationally available to machines, and orders configurations on a six-level classification in which only architectures that structurally support human judgment count as HI. Dedicated manifestos in progress operationalize the interface (HI Design Playbook), market (human-touch marketplace; HI Economics Manifesto), and educational (HI Educational Manifesto, [8]) levels. What none supplies alone is the behavioral layer: how an individual professional locates the frontier in a concrete task, keeps evaluative authority across rounds, and closes with accountable commitment. FERC is the series’ answer.

The need for this behavioral layer becomes acute precisely as organizational use shifts from one-shot prompting toward multi-round, iterative interaction. Here the challenge is no longer only how to steer models toward better outputs, but how to prevent the gradual erosion of human authorship, as framing, evaluation, and ultimately responsibility drift from the human to the system. Three dimensions become decisive in consequential work: authorship preservation (who defines goals, values, and trade-offs), evaluative reframing (how human evaluation reconstructs the next search space rather than merely polishing an output), and stopping rules grounded in human judgment rather than token limits, system confidence, or convenience. When these are left implicit, the result is authorship drift: the gradual outsourcing of framing, evaluative judgment, and decision responsibility to the system. Existing prompt-engineering and AI-literacy approaches, reviewed in Section 2, largely operate one interaction at a time and do not govern these dimensions across rounds. We distinguish authorship, accountable agency in the collaborative process, from ownership, rights over outputs that in organizational settings may be assigned contractually and need not coincide with authorship [9].

We argue that accountable agency does not emerge automatically from system properties but must be enacted through a governed sequence of interaction stages. To operationalize this shift from artifact-centered prompt optimization toward process-centered interaction governance, we introduce FERC (Frame, Explore, Refine, Commit), a structured framework for multi-round human–AI collaboration intended to support human intent, judgment, and responsibility. Rather than specifying how generative systems should be engineered, FERC specifies how interaction with such systems should unfold in an authorship-preserving way. This paper contributes (1) FERC, a governed interaction cycle intended to support authorship-preserving multi-round collaboration; (2) a theoretical and design-process grounding that links FERC to iterative problem construction and reframing, including an illustrative Double Diamond composition; (3) an operationalization as a collaboration maturity assessment with formative evidence from executive training contexts, together with FERC-bot, a metacognitive training and measurement interface that renders collaboration structure visible; and (4) two pathways for the formal operationalization of FERC

process scoring, in educational assessment and in the legal and organizational documentation of human contribution, treated in a dedicated section before the conclusion. The educational pathway, where the framework’s transformative potential is arguably largest, is being developed in full institutional form in the HI Educational Manifesto [8].

2 Background and Related Work

2.1 Prompt engineering, AI Literacy, and multi-round optimization for LLM collaboration

Section 1 identified authorship drift as a failure that arises across interaction rounds. The dominant approaches to human–AI interaction, however, largely address the single interaction. Prompt engineering treats the prompt as the primary design artifact, with pattern- and template-based “recipes” (role, task, context, constraints, output format) that help users express intent and obtain more reliable outputs [10, 11]. Our review of prominent practitioner frameworks such as RTF, SOLVE, and TAG [12–14] confirms that most focus on structuring the initial prompt; the few that gesture toward follow-up interaction offer vague additions to templates rather than guidance for iterative collaboration. Even comprehensive AI-literacy initiatives [15–17] typically frame iteration as prompt tweaking to improve a single output rather than as a structured cycle in which the problem representation itself evolves [18, 19].

In parallel, computer-science research has developed multi-step and iterative techniques, including sequential prompting, tool-augmented workflows, self-feedback loops, agentic task decomposition, and context-engineering approaches that expand what a model holds in working memory [20]. These substantially improve robustness, consistency, and task completion, but they optimize reliability and efficiency rather than ensuring that human intent, evaluative authority, and stopping decisions remain explicit throughout multi-round work. A smaller body of work moves toward interaction design proper, and comes closest to our concern. Shah (2024) [21] proposes a “prompt science” approach emphasizing human-in-the-loop validation and replicability, while

(2024) [22] introduces a prediction–judgment-based Human–Machine Learning Loop that combines large-scale AI generation with expert evaluation and learning cycles; both foreground iteration and human judgment but focus on optimizing results and leave the structure of interaction itself under-specified, prescribing neither how evaluative authority is maintained nor when human-grounded stopping occurs across rounds. Related design-science and hybrid-intelligence work scaffolds divergent and convergent thinking through multi-step human–AI structures [23] and grounds iterative generation and reflection in domain expertise [22], and structured-questioning research supports richer flows and problem construction in ill-defined domains [24–26]. In education specifically, early multi-step supports identify common AI errors with mitigation strategies and embed unfolded, multi-step interaction blueprints within initial prompts [27]. Taken together, these strands demonstrate the value of multi-step structure but stop short of a general, transferable interaction recipe that governs not only the initial prompt but how alternatives are explored, how evaluation

reshapes the next search space, and how accountable commitments are made across rounds. FERC is proposed to fill that gap.

2.2 Human authorship, accountability, and governance in human–AI work

As GenAI becomes embedded in knowledge work, research increasingly examines its effects on cognitive effort, judgment, and responsibility. Evidence suggests that AI can reduce perceived effort while increasing confidence, raising risks of overreliance on fluent outputs [1]. At the same time, studies show that perceptions of authorship and responsibility shift with the degree of AI assistance [28, 29]. In high-stakes contexts such as strategy, public communication, or policy, these shifts become governance issues, influencing who effectively authors goals, trade-offs, and final decisions. Authorship in human–AI collaboration is rarely binary; it is distributed across interaction moves such as framing goals, directing exploration, defining evaluation criteria, and deciding when to commit. When interaction models prioritize speed or output quality, these decisions may drift toward the system, highlighting the need for methods that make authorship-relevant choices explicit, especially in multi-round collaboration where criteria evolve across iterations.

2.3 Creativity and metacognitive control as a basis for governed interaction

The absence of explicit authorship and evaluation mechanisms becomes particularly salient through the lens of creativity and design research, which treats high-quality outcomes not as products of one-shot generation, but as emerging through iterative cycles in which problem representations, candidate ideas, and evaluative criteria reshape one another. More broadly, creativity scholarship has shifted from defining creativity as a novel and useful product to understanding it as a process unfolding through cognitive and neural dynamics over time [30, 31]. Processes include acquisition of relevant information, formulation of ideas, and evaluation of novelty and usefulness [32].

These processes draw on both associative and controlled mechanisms. Associative processes, supported by the organization of concepts in semantic memory, enable individuals to identify remote or non-obvious connections that can seed original ideas [33–35]. Controlled processes, by contrast, support efficient search, inhibition of dominant responses, cognitive shifting, and the flexible adaptation of thinking strategies as task demands evolve [36–38]. Executive functions such as inhibition, cognitive shifting, and broad retrieval are central to how creative work is regulated in practice [39, 40]. Creativity, in this view, is not simply a matter of producing many ideas, but of navigating between expansion and constraint in ways that remain responsive to emerging goals, standards, and contextual demands.

A second core insight is that evaluation is not merely a downstream act of selecting among completed ideas. Problem construction research shows that creators repeatedly identify, frame, and reformulate the problem itself during creative work [41–43]. Likewise, Geneplore and related process models describe creativity as an iterative interplay between generation and evaluation, where

judgment reshapes subsequent search rather than simply terminating it [44–47]. These changes throughout the creative process, and the role of idea evaluation within it, are explained more comprehensively by creative metacognition, which holds that creativity involves not only generating ideas but also monitoring them, and using those evaluations to guide subsequent decisions and adjustments [48, 49]. Such self-regulation is intentional and spans cognition, behavior, and emotion: monitoring evaluates the quality of one’s own execution, while control intentionally directs changes, selects strategies, and steers the course of the process, alongside planning, anticipating uncertainty, and willingness to share one’s creative work.

3 The FERC Method: A Governed Cycle of Human–AI Collaboration

These insights suggest that creative work requires at least four distinct governance functions. First, human-led problem construction, making intent, constraints, and success criteria explicit [41, 43]. Second, deliberate divergence, generating multiple alternatives rather than collapsing prematurely onto one [44]. Third, evaluative reframing, where comparative judgment updates criteria and reconstructs the search space [41, 43]. Fourth, where humans explicitly decide, stop, and own the consequences. These are separable governance functions for preserving human authorship and accountability in multi-round collaboration: collapsing any two tends to recreate familiar failure modes, such as generation without criteria, refinement without sufficient divergence, or iteration without explicit stopping, while adding further phases risks subdividing implementation detail rather than introducing a new mechanism for authorship preservation (see Table 1 below).

This distinction becomes especially important in AI-supported environments, where fluent AI outputs can encourage reduced cognitive effort, overconfidence, and premature convergence [1]. The challenge is therefore not only improving AI outputs but preserving the human role in framing, evaluating, revising, and committing across iterative interaction. FERC addresses this by acting as a **metacognitive scaffold** for hybrid cognition, redistributing generative work to AI while preserving human authorship through problem construction, evaluative control, and accountable closure. Thus, we introduce **FERC (Frame–Explore–Refine–Commit)** as a structured method for multi-round human–AI collaboration that centers human authorship and metacognitive control. FERC is not a prompt template but a governed interaction cycle.

3.1 Phases and mechanics

Frame. The human defines intent: articulating goals, constraints, values, success criteria, and scope boundaries. This phase anchors the collaboration in human expertise, contextual understanding, and evaluative standards. **Explore.** The human deliberately instructs the AI to generate multiple options or approaches within the defined frame, expanding the search space and increasing output diversity [32, 50]. **Refine.** The human evaluates alternatives comparatively, articulates strengths, weaknesses, trade-offs, and missing elements, and explicitly modifies the criteria or instructions for the next Explore cycle. Crucially, Refine reshapes the subsequent search space rather than merely polishing an artifact, establishing

Table 1: Mapping of Creative Theory Concept on to FERC Phase

Creative Theory Concept	FERC Phase
Problem Construction	Frame
Divergent Thinking	Explore
Iterative Elaboration	Refine
Convergent Evaluation	Commit
Co-evolution of problem & solution	Frame/Refine cycles

Table 2: Conceptual Distinction Between the CARE System-Level Hybrid Intelligence Frameworks and the FERC Governed Interaction Model

Dimension	CARE	FERC
Level	System-level	Interaction-level
Focus	Architectural properties	Process governance
Question	What should the system embody?	How should collaboration unfold?
Measurement	Hard to operationalize	Behavioral observable stages
Explainability	Model transparency	Iterative human verification

an iterative E–R loop that mirrors creative co-evolution. **Commit.** The human selects a course of action, documents rationale where appropriate, and assumes accountability for the outcome. Together, these phases, of which Explore and Refine are ideally iteratively explored, operationalize structured problem construction, deliberate divergence, evaluative reframing, and accountable closure in hybrid intelligent work.

Recent Hybrid Intelligence research emphasizes system-level principles such as Collaboration, Adaptivity, Responsibility, and Explainability (CARE) [51, 52]. The dominant route to these properties is to make the system itself deliver them under increasing autonomy, exemplified by agentic orchestration [53]: ReAct interleaves reasoning and acting within an autonomous run [54], Reflexion adds self-evaluation across attempts [55], and DSPy automates prompt construction itself [56], each inferring as much as possible from as little human input as possible.

FERC, and the HI Manifesto [7] more broadly, takes the opposite route: rather than minimizing human input, pursuing human expertise growth by granularizes the interaction so that human framing, evaluation, and commitment are structurally required (Table 2). This inverts explainability. Where the autonomy route must manufacture transparency for a human who cannot otherwise verify a large autonomous span, FERC’s granularity makes it largely unnecessary: guidelines #2 and #4 (Section 4.1), acting as domain expert and decomposing until each step is checkable, render every intermediate output verifiable against the expert’s own knowledge without any view into the model’s internals. The two are thus opposite settings of one dial: how much of the judgment loop is inferred by the system versus exercised by the human.

3.2 FERC and self-regulation in creative work

The FERC stages also map directly onto core phases of creative self-regulation described in agentic models of creative behavior. Framing corresponds to the forethought phase, in which goals are clarified, and strategic intentions are formed [48, 57]. Exploration and refinement reflect ongoing metacognitive monitoring

and control in the form of adaptive strategy use as people navigate the problem space, comparisons across alternatives and representational restructuring [58]. Finally, commitment aligns with self-reflection and agentic decision-making, as judgments are consolidated into explicit choices accompanied by responsibility for outcomes [48, 49, 57]. Viewed this way, FERC structures interaction with AI but also externalizes and supports key self-regulatory processes underlying sustained creative performance. Importantly, externalizing interaction structure does not in itself guarantee that these processes are substantively exercised. Even when framing, evaluation, and decision points are made explicit, users may experience an illusion of control or authorship while tacitly adopting AI-suggested problem framings, evaluative criteria, or stopping decisions. In this sense, FERC should not be understood as a guarantee of human self-regulation. Rather, it functions as a scaffold that renders these authorship-relevant moves more visible, inspectable, and governable, thereby increasing the likelihood, but not ensuring, that human judgment is actively exercised.

Recent work has also explored dialogue-based interaction frameworks inspired by the classical Socratic method, in which knowledge is developed through structured questioning that probes assumptions, surfaces contradictions, and guides learners toward deeper understanding. For example, Hashemi Tonekaboni and Soleymani (2026) [57] propose a Socratic Knowledge Interaction Framework that structures human–LLM collaboration through iterative cycles of critical questioning, contradiction, elicitation, and synthesis. These dialogue phases encourage users to articulate assumptions, confront contradictions, and revise their reasoning, thereby supporting deeper inquiry and knowledge co-creation. While such approaches emphasize structured questioning to stimulate critical thinking, the FERC framework complements this perspective by focusing on interaction governance: explicitly structuring how goals are framed, alternatives explored, evaluations conducted, and decisions ultimately committed in human–AI collaboration.

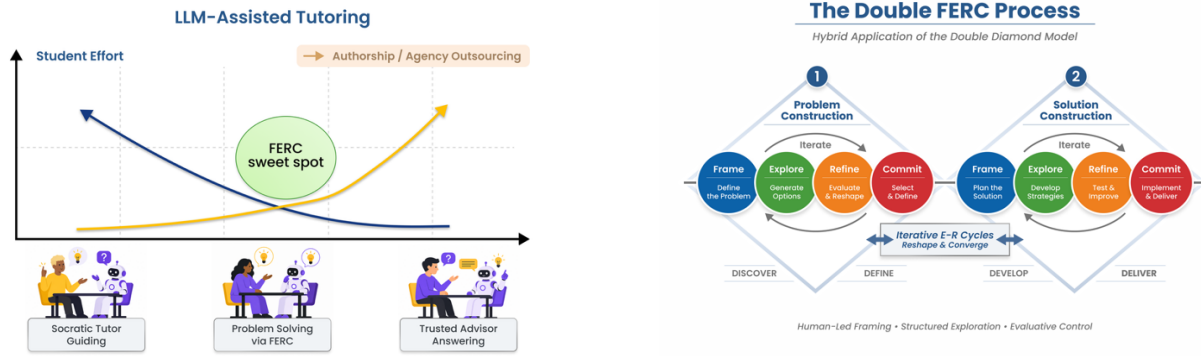


Figure 1: Left: *The FERC sweet spot in LLM-assisted interaction, balancing meaningful human effort and authorship retention against the extremes of full Socratic tutoring and passive answer-seeking.* Right: *The Double FERC process for authorship-preserving human-AI collaboration.*

FERC thus occupies a practical and scalable middle ground between two extremes of human-AI interaction by distributing cognitive labor strategically. Figure 1, left, illustrates this by comparing the student effort and the outsourced authorship/agency across two prominent existing approaches and the FERC method. At one extreme lies minimal-effort answer seeking corresponding to much present-day casual LLM usage, where users request outputs directly and implicitly outsource framing, evaluation, and decision criteria to the model. This pattern aligns with documented overreliance behaviors in human-AI interaction, where users tend to accept fluent and authoritative model outputs with limited scrutiny, even when they are incorrect, leading to reduced critical evaluation and degraded task performance [59]. At the other extreme lies intensive Socratic human-AI tutoring, where continuous questioning scaffolds deeper reasoning but requires sustained effort and is difficult to scale in organizational settings [41, 57].

4 Operationalization and Evaluation

4.1 Practical Application Guidelines for FERC

While FERC provides a conceptual framework for structuring human-AI collaboration, its value depends on how it is enacted in practice. To translate the model into responsible use, practitioners need concrete behavioral guidelines that preserve authorship, evaluative control, and iterative co-evolution in everyday interactions. The following principles are intended not as definitive rules, but as initial operational heuristics that will require refinement across domains and organizational contexts.

1. Invest in specifying the task instead of asking for a quick product. Before asking AI to produce an output, clarify goals, constraints, audience, and success criteria. Instead of “Write a strategy memo,” begin with “Help me define the purpose, audience, and evaluation criteria for this memo.” The quality of the frame determines the quality of the result. The quality of the questions asked are critical to initiating the process.

2. Do not treat AI as a knowledge authority beyond your expertise. Use AI to generate options, not to replace domain

expertise. Responsibility for factual accuracy, contextual judgment, and institutional alignment remains human.

3. Use FERC for any solution-oriented tasks. Apply FERC whenever the goal is to develop, evaluate, and commit to a solution—broadly understood as any task involving decisions, artifacts, or actions. This contrasts with knowledge acquisition tasks (e.g., learning or explanation), which still requires critical thinking and validation but following a different interaction pattern.

4. Decompose tasks until evaluation is realistic. If a task feels too large to review multiple alternatives, one output is most likely also too large for responsible evaluation. Break complex outputs into smaller steps where several options can be meaningfully compared. Manage scope to protect judgment quality. Added benefit: decomposed tasks reduces token cost and environmental footprint and enables the use of smaller, e.g. open source models.

5. If you skip FERC, make it a conscious decision. There are situations where speed outweighs depth or where authorship is less critical (e.g., low-risk or purely procedural tasks). In such cases, explicitly acknowledge that you are prioritizing efficiency over iterative authorship. Unintentional delegation erodes ownership; deliberate delegation preserves it.

4.2 Compounding FERC loops: The Double-Diamond Case

In most of our empirical testing one FERC cycle sufficed. Where a use case warrants explicit attention to both problem and solution construction, however, FERC cycles can compound: the Double Diamond model, which separates divergence–convergence in the problem space (Discover–Define) from the solution space (Develop–Deliver), can be reinterpreted as two chained FERC cycles (Fig. 1, Right). The first governs problem construction, humans frame goals and criteria, explore alternative problem definitions, refine through comparative evaluation, and commit to a framing; the second governs solution construction over implementation strategies. Whereas prior work supports the Double Diamond through agentic orchestration [23], FERC governs it at the interaction level, preserving human authority over framing, evaluation, and closure

Table 3: Human-AI Collaboration Mindset Maturity Levels

Level	Label	Description
0	Inactive	AI is absent from workflows or treated as irrelevant or threatening. Problems are framed exclusively for human-only execution. There is no structured reflection on hybrid collaboration.
1	Initiating (Ad Hoc Use)	AI use is sporadic and individual. One-shot prompts dominate. Iteration is minimal. Outputs are often accepted or rejected without structured reframing. Pride is ambivalent (curiosity mixed with uncertainty).
2	Emerging (Exploratory Learning)	Multi-round interaction appears. Users experiment with context and refinement. Some iterative improvement occurs, but problem framing and evaluation remain inconsistent. Good individual practices may exist, but there is no shared organizational method.
3	Operationalized (Integrated Co-Creation)	AI is embedded in workflows. FERC cycles are explicit and repeatable. Humans consistently define intent, compare alternatives, iterate via evaluative reframing, and formally commit to decisions. Pride shifts from output production to strategic judgment. Collaboration standards are shared across roles.
4	Transformative (Human-AI Symbiosis)	Continuous co-evolution of human and AI practices. Teams refine not only outputs but collaboration processes themselves. FERC becomes part of institutional identity. Organizational metrics, governance, and incentives reinforce hybrid authorship and accountability. AI collaboration is recognized as a strategic capability and competitive advantage.

throughout. Tested across a dozen organization, selecting an organizational AI use case illustrates the composition: leaders run a first cycle to converge on a use case and a second on its deployment strategy. Note, this is one composition among several, cycles may also nest, iterate, or interleave as the task demands, rather than a fixed sequence FERC must follow.

4.3 Human-AI Collaboration Maturity Model

To operationalize FERC at the organizational level, we introduce a **Human-AI Collaboration Maturity Model** that assesses how systematically FERC principles are embedded in workflows, culture, and decision-making. Rather than measuring AI usage frequency or technical sophistication, the model evaluates the degree to which human authorship, structured iteration, and evaluative responsibility are preserved in multi-round collaboration. Progression

through the levels is not merely a training problem. Experts adopt authorship-preserving practices, and share the tacit knowledge on which Frame depends, only when they perceive that their expertise is being developed rather than extracted. Case evidence from AI-assisted mentalizing-based therapy shows that this willingness is generated by explicitly linking the long-term occupational implications of a system to its concrete design and interaction elements, so that the professional’s future role is visible inside the interface itself rather than asserted alongside it [60]. Maturity progression therefore presupposes the credible long-term commitment that the main HI Manifesto [7] identifies as the precondition for expert knowledge-sharing. The FERC maturity model thus provides a shared language for discussing capability development and a strategic instrument for guiding organizational transformation toward hybrid intelligence.

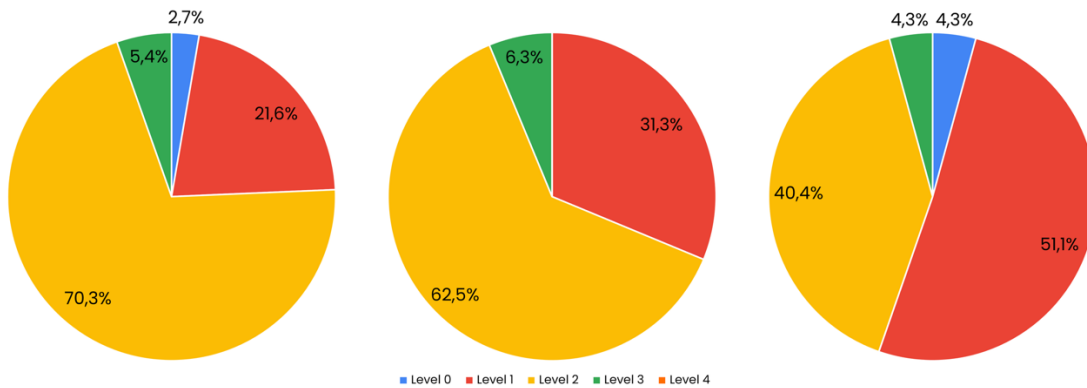


Figure 2: Distribution of self-assessed FERC maturity across three professional-training contexts: left, business college faculty ($n = 37$); center, marketing executives ($n = 16$); and right, leaders from a Nordic supermarket chain ($n = 47$). Across contexts, responses cluster mainly in the lower-to-middle maturity range.

4.4 Practical Real-World Testing and Observed Patterns

FERC has been applied in training and workflow interventions across more than 20 organizations, reaching more than 2,000 individuals across corporate, public-sector, and educational settings. Within this portfolio, the maturity framework was refined through more than ten executive training sessions in which leaders assessed their organization’s FERC maturity using a simple visual method: participants indicated their perceived level (0–4) by raising the corresponding number of fingers. Between workshops, the level descriptions were adjusted to avoid common misconceptions, particularly the conflation of maturity with prompt engineering training or isolated examples of good individual practice. The focus was clarified to emphasize systematic, organization-wide authorship-preserving collaboration. The first exploratory version, involving a single multiple-choice question based on the level definitions in Table 3, has been tested in three contexts: a business college with faculty participants, a group of 16 marketing executives, and 47 leaders from a Nordic supermarket chain. Across these groups, reported maturity levels cluster mainly between Level 1 (Initiating) and Level 2 (Emerging), suggesting that while multi-round experimentation is common, structured evaluative reframing and explicit commitment practices remain limited. The model is intended as a strategic tool and for robustness, further research should expand the survey to a multi-item version with rigorous validation. By treating FERC maturity as a visible KPI, organizations can track whether authorship-preserving hybrid collaboration practices become more systematically embedded over time.

All three distributions cluster primarily in the lower-to-middle maturity range, consistent with Levels 1–2 of the model. This pattern suggests that while experimentation with GenAI is widespread, more systematic authorship-preserving practices, especially evaluative reframing and explicit commitment, remain unevenly institutionalized.

5 FERC-Bot and the Operationalization of Process Scoring

Conceptual understanding alone rarely changes everyday behavior, and the strongest motivation for making FERC behaviorally explicit is that its four phases are directly observable and therefore scoreable. We propose FERC-bot, a training interface that embeds the FERC stages in a live conversational environment and provides a per-stage score and brief explanation, focusing on process rather than output quality: the guiding question is not “was the answer good?” but “was the collaboration well-structured and authorship-preserving?” Specifically, it assesses whether users defined goals and criteria (Frame), requested multiple alternatives (Explore), comparatively evaluated options to reshape the next exploration (Refine), and made a deliberate, accountable decision (Commit), while domain experts retain responsibility for substantive output quality. Detailed scoring metrics and validation are left to future work; here we motivate the interface through the institutional functions such scoring enables, since it is these functions that answer how authorship preservation could be measured rather than merely asserted.

Two pathways illustrate this. In formal education, generative AI has undermined output-based assessment, since submitted artifacts no longer certify the learning they were designed to evidence; process scoring restores an assessable object by recording whether a learner framed before generating, obtained genuine alternatives, evaluated comparatively, and committed with rationale. Because FERC exercises and prompt sequences can be authored for any domain, from evaluating AI-generated order proposals in trade education to stress-testing load calculations in engineering, the same cycle trains and assesses judgment on discipline-specific content; a continually updated library is curated at hi-infinity.ai/ferc. The HI Educational Manifesto develops the full institutional architecture, where this potential is arguably largest [8]. In professional work, the same logs, recording who framed the problem, which alternatives were explored, and who committed with what rationale, provide a documentation layer for human contribution. This addresses two emerging needs: intellectual-property regimes increasingly

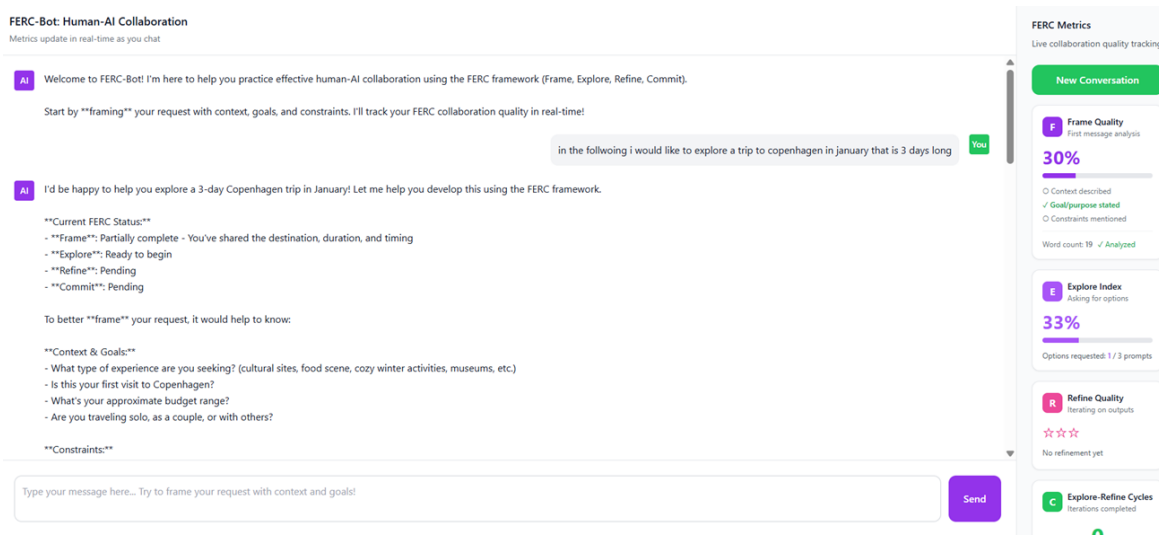


Figure 3: The first conceptual version of the FERC-Bot interface. The tool augments a standard chat environment with a live diagnostic panel that tracks and scores collaboration quality across the four FERC stages (Frame, Explore, Refine, Commit) in real time.

condition protection on demonstrable human creative input, and markets that reward human-touch value require that contribution be verifiable rather than asserted. FERC scoring supplies this micro-level verifiability in analogy to firm-level CO2 accounting, which made climate commitment auditable and so enabled certification and procurement; aggregated records let an organization demonstrate, rather than declare, that human judgment is structurally embedded, making the symbolic-oversight configurations the HI Manifesto terms HI-washing detectable at the interaction level. The HI Economics Manifesto develops this market mechanism.

6 Conclusion

FERC reframes effective use of genAI from prompt optimization to process governance: a repeatable cycle that preserves human authorship by making intent (Frame), divergence (Explore), evaluative reframing (Refine), and accountable closure (Commit) explicit in multi-round collaboration. By grounding these phases in creativity research on iterative problem construction, metacognitive regulation, and shifts between divergent and convergent thinking, the paper shows why authorship preservation is not a property of the model but a function of how interaction is structured, especially when fluency and efficiency can otherwise accelerate uncritical delegation. We operationalize this shift through Double FERC, practical guidelines, a formative maturity model, and FERC-bot as a training and measurement interface. Together, these contributions offer a pragmatic pathway for scalable Hybrid Intelligence in organizations while maintaining human intent, judgment, and responsibility as generative systems grow more capable. Within the HI Manifesto series, FERC thus carries the interaction-level share of a single program: keeping human judgment economically indispensable, reproducing the capacity for such judgment through the educational transformation of the HI Educational Manifesto [8], and, through formal process scoring, making the human contribution to AI-assisted work documentable, auditable, and rewardable.

References

- [1] Hao-Ping Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. 2025. The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*, April 26 – May 1, 2025, Yokohama, Japan. ACM, New York, NY, Article 1121, 22 pages. <https://doi.org/10.1145/3706598.3713778>
- [2] Michael Gerlich. 2025. AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies* 15, 1 (2025), Article 6. <https://doi.org/10.3390/soc15010006>
- [3] Janet Rafner, Dominik Dellermann, Arthur Hjorth, Dóra Verasztó, Constance Elizabeth Kampf, Wendy E. Mackay, and Jacob Sherson. 2021. Deskilling, upskilling, and reskilling: A case for hybrid intelligence. *Morals & Machines* 1, 2 (2021), 24–39. <https://doi.org/10.5771/2747-5174-2021-2-24>
- [4] Fabrizio Dell'Acqua, Edward McFowland III, Ethan Mollick, Hila Lifshitz, Katherine C. Kellogg, Saran Rajendran, Lisa Kraye, François Candelon, and Karim R. Lakhani. 2026. Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organization Science* 37, 2 (2026), 403–423. <https://doi.org/10.1287/orsc.2025.21838>
- [5] Carsten Bergenholtz, Oana Vuculescu, Franziska Günzel-Jensen, and Lars Fredriksen. 2026. Leveling up or leveling down? The impact of generative AI on student performance in business schools. *Academy of Management Learning & Education* 25, 2 (2026), 155–186. <https://doi.org/10.5465/amle.2025.0029>
- [6] Sebastian Raisch, Sebastian Krakowski. Artificial intelligence and management: The automation-augmentation paradox. *Acad Manage Rev.* 2021;46(1):192–210. <https://doi.org/10.5465/amr.2018.0072>
- [7] Jacob Sherson, Janet Rafner, Jesper Broberg Bang, Aleksandra Berditchevskaia, Jens Christian Bjerring, Sudapa Chompunuch, Enrico Daga, Dominik Dellermann, Umberto Domanti, Blerim Emruli, Frank van Harmelen, Maceo Gonzalez, Adam Gordon, Adam Green, Rigmor Lyck Hansen, Anna B. Holm, Catholijn Jonker, Sanna Järvelä, Yoed N. Kenett, Frederik Brosbøl Kjeldsen, Mette P. Knudsen, Carlos Largacha-Martinez, Izabela Lebeda, Rasmus Linnet, Magnus Lodefalk, Yaoli Mao, Janet Marler, André Meyer-Vitali, Chiara Natali, Morten Røndal Olsen, Seyedahmad Rahimi, Leon Reicherts, Roni Reiter-Palmon, Henry Sauer-mann, Danielle Sent, Thomas Schmid, Matthias Söllner, Jonas Elpelt, Stephanie Tan, Casper Thylkjær, Ilaria Tiddi, Ana Alina Tudoran, Marcus Wagner, Behzad Bozorgtabar, Carlos Rafael Catalan, Miranda Kajtazi, Svetlana Norkin, Christos Gkoumas, Peter Vistisen, Daniel Staegemann, Joachim Wiewiura, Mariana Filipova, Adam Dahlgren Lindström, Zorana Ivcevic, Nick von Felten, Zelun Tony Zhang, Lev Tankelevitch, Florent Vinchon, Bianka Szöllösi, Ginés Molina-Abril, Joaquin Veith, Philipp Bitzenbauer, and Jason Tucker. 2026. The Hybrid Intelligence Manifesto: A multi-level socio-technical framework

- for preserving human judgement as an economically indispensable component of future AI-enabled societies. Preprint. Retrieved August 2, 2026 from <https://www.researchgate.net/publication/408705732>
- [8] Jacob Sherson, Blerim Emruli, Rigmor Lyck Hansen, Sanna Järvelä, Yoed N. Kenett, Frederik Brosbøl Kjeldsen, Camila Kolsen Petersen, Izabela Lebeda, Yaoli Mao, Yishay Mor, Chiara Natali, Andy Nguyen, Janet Rafner, Adam Vigdor Gordon, Seyedahmad Rahimi, Matthias Söllner, Ana Alina Tudoran, Dominik Dellermann, Steve DiPaola, Florent Vinchon, Roni Reiter-Palmon, Jens Gamauf Madsen, Oday Darwich, Anna B. Holm, Joaquin Veith, Philipp Bitzenbauer, Umberto Domanti, Jonas Elpelt, Daniel Staegemann, Jesper Broberg Bang, and Mariana Filipova. 2026. The HI Manifesto for Education: Institutionalizing societal stewardship in the age of AI. Preprint. Retrieved August 6, 2026 from <https://www.researchgate.net/publication/411642988>
 - [9] Gazi Islam and Michelle Greenwood. 2024. Generative artificial intelligence as hypercommons: Ethics of authorship and ownership. *Journal of Business Ethics* 192, 4 (2024), 659–663. <https://doi.org/10.1007/s10551-024-05741-9>
 - [10] Louie Giray. 2023. Prompt engineering with ChatGPT: A guide for academic writers. *Annals of Biomedical Engineering* 51, 12 (Dec. 2023), 2629–2633. <https://doi.org/10.1007/s10439-023-03272-4>
 - [11] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, Douglas C. Schmidt. A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv preprint arXiv:2302.11382. 2023.
 - [12] ENEB. 2025. Prompt frameworks to maximize ChatGPT results. (May 2025). Retrieved May 18, 2026 from <https://eneb.com/prompt-frameworks-to-maximize-chatgpt-results/>
 - [13] Dorian Machado. 2025. 8 ChatGPT (ChatGPT5) prompt frameworks. Medium. (Aug. 2025). Retrieved August 2, 2026 from <https://dorian599.medium.com/8-chatgpt-chatgpt5-prompt-frameworks-0bc6a07eff44>
 - [14] Maab Saleem. 2026. 8 ChatGPT prompt frameworks every marketer should know. ButterCMS. (March 2026). Retrieved August 2, 2026 from <https://buttercms.com/blog/chatgpt-prompt-frameworks/>
 - [15] European Commission and OECD. 2026. *Empowering Learners for the Age of AI: An AI Literacy Framework for Primary and Secondary Education*. (June 2026). Retrieved August 2, 2026 from <https://ailiteracyframework.org/>
 - [16] Tomáš Lintner. 2024. A systematic review of AI literacy scales. *npj Science of Learning* 9, 1 (2024), Article 50. <https://doi.org/10.1038/s41539-024-00264-4>
 - [17] Chrysi Rapanta, Ibrar Bhatt, Aras Bozkurt, Laura Ann Chubb, Christopher Erb, Ingrid Forsler, Karen Gravett, Marguerite Koole, Tomáš Lintner, Alex Örtengren, Tiffany Petricini, Brian Rodgers, Jack Webster, Xiao Xu, Inger-Marie Falgren Christensen, Nina Bonderup Dohn, Lise Lotte Weigaard Christensen, Sandris Zeivots, and Petar Jandrić. 2025. Critical GenAI literacy: Postdigital configurations. *Postdigital Science and Education* 7 (2025), 1296–1333. <https://doi.org/10.1007/s42438-025-00573-w>
 - [18] Stephen H. Bach, Victor Sanh, Zheng-Xin Yong, Albert Webson, Colin Raffel, Nihal V. Nayak, Abheesh Sharma, Taewoon Kim, M. Saiful Bari, Thibault Fevry, Zaid Alyafeai, Manan Dey, Andrea Santilli, Zhiqing Sun, Srulik Ben-David, Canwen Xu, Gunjan Chhablani, Han Wang, Jason Alan Fries, Maged S. Al-shaibani, Shanya Sharma, Urmish Thakker, Khalid Almubarak, Xiangru Tang, Dragomir Radev, Mike Tian-Jian Jiang, and Alexander M. Rush. 2022. PromptSource: An integrated development environment and repository for natural language prompts. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations (ACL '22)*, May 22–27, 2022, Dublin, Ireland. Association for Computational Linguistics, Stroudsburg, PA, 93–104.
 - [19] J. D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, April 23–28, 2023, Hamburg, Germany. ACM, New York, NY, Article 437, 21 pages. <https://doi.org/10.1145/3544548.3581388>
 - [20] Yao-Yang Liu, Zhen Zheng, Feng Zhang, Ji-Cheng Feng, Yi-Yang Fu, Ji-Dong Zhai, Bing-Sheng He, Xiao Zhang, and Xiao-Yong Du. 2026. A comprehensive taxonomy of prompt engineering techniques for large language models. *Frontiers of Computer Science* 20, 3 (2026), Article 2003601. <https://doi.org/10.1007/s11704-025-50058-z>
 - [21] Chirag Shah. 2025. From prompt engineering to prompt science with humans in the loop. *Commun. ACM* 68, 6 (June 2025), 54–61. <https://doi.org/10.1145/3709599>
 - [22] Jacob Sherson. 2024. A hybrid intelligent change management approach to generative AI adoption. In *HHAI 2024: Hybrid Human AI Systems for the Social Good*. Fabian Lorig, Jason Tucker, Adam Dahlgren Lindström, Frank Dignum, Pradeep Murukannaiah, Andreas Theodorou, and Pinar Yolum (Eds.), Frontiers in Artificial Intelligence and Applications, Vol. 386. IOS Press, Amsterdam, 472–474. <https://doi.org/10.3233/FAIA240228>
 - [23] Leopold Fischer-Brandies, Simon Meierhöfer, and Dominik Protschky. 2024. Augmenting divergent and convergent thinking in the ideation process: An LLM-based agent system. In *Proceedings of the 32nd European Conference on Information Systems (ECIS 2024)*, June 2024, Paphos, Cyprus. Association for Information Systems, Paper 13. Retrieved from https://aisel.aisnet.org/ecis2024/track20_adoption/track20_adoption/13
 - [24] Gal Sasson Lazovsky, Tuval Raz, and Yoed N. Kenett. 2025. The art of creative inquiry—From question asking to prompt engineering. *The Journal of Creative Behavior* 59, 1 (2025), Article e671. <https://doi.org/10.1002/jocb.671>
 - [25] Tuval Raz, Roni Reiter-Palmon, and Yoed N. Kenett. 2024. Open and closed-ended problem solving in humans and AI: The influence of question asking complexity. *Thinking Skills and Creativity* 53 (2024), Article 101598. <https://doi.org/10.1016/j.tsc.2024.101598>
 - [26] Tuval Raz, Roni Reiter-Palmon, and Yoed N. Kenett. 2025. The role of asking more complex questions in creative thinking. *Psychology of Aesthetics, Creativity, and the Arts* 19, 6 (2025), 1505–1525. <https://doi.org/10.1037/aca0000658>
 - [27] Lilach Mollick and Ethan R. Mollick. 2024. Instructors as innovators: A future-focused approach to new AI learning opportunities, with prompts. The Wharton School Research Paper. (April 2024). <https://doi.org/10.2139/ssrn.4802463>
 - [28] Aras Bozkurt. 2024. Cocreation, authorship, ownership, academic ethics and integrity in a time of generative AI. *Open Praxis* 16, 1 (2024), 1–10. <https://doi.org/10.55982/openpraxis.16.1.654>
 - [29] Paul Formosa, Sarah Bankins, Rita Matulionyte, and Omid Ghasemi. 2025. Can ChatGPT be an author? Generative AI creative writing assistance and perceptions of authorship, creatorship, responsibility, and disclosure. *AI & Society* 40, 5 (2025), 3405–3417. <https://doi.org/10.1007/s00146-024-02081-0>
 - [30] Adam E. Green, Roger E. Beaty, Yoed N. Kenett, and James C. Kaufman. 2024. The process definition of creativity. *Creativity Research Journal* 36, 3 (2024), 544–572. <https://doi.org/10.1080/10400419.2023.2254573>
 - [31] Mark A. Runco and Garrett J. Jaeger. 2012. The standard definition of creativity. *Creativity Research Journal* 24, 1 (2012), 92–96. <https://doi.org/10.1080/10400419.2012.650092>
 - [32] Mathias Benedek, Roger E. Beaty, Daniel L. Schacter, and Yoed N. Kenett. 2023. The role of memory in creative ideation. *Nature Reviews Psychology* 2, 4 (April 2023), 246–257. <https://doi.org/10.1038/s44159-023-00158-z>
 - [33] Anna Abraham and Andreja Bubic. 2015. Semantic memory as the root of imagination. *Frontiers in Psychology* 6, Article 325 (March 2015). <https://doi.org/10.3389/fpsyg.2015.00325>
 - [34] Roger E. Beaty and Yoed N. Kenett. 2023. Associative thinking at the core of creativity. *Trends in Cognitive Sciences* 27, 7 (July 2023), 671–683. <https://doi.org/10.1016/j.tics.2023.04.004>
 - [35] Yoed N. Kenett. 2025. The role of knowledge in creative thinking. *Creativity Research Journal* 37, 2 (2025), 242–249. <https://doi.org/10.1080/10400419.2024.2322858>
 - [36] Mathias Benedek and Andreas Fink. 2019. Toward a neurocognitive framework of creative cognition: The role of memory, attention, and cognitive control. *Current Opinion in Behavioral Sciences* 27 (June 2019), 116–122. <https://doi.org/10.1016/j.cobeha.2018.11.002>
 - [37] Evangelia G. Chrysikou. 2019. Creativity in and out of (cognitive) control. *Current Opinion in Behavioral Sciences* 27 (June 2019), 94–99. <https://doi.org/10.1016/j.cobeha.2018.09.014>
 - [38] Emmanuelle Volle. 2018. Associative and controlled cognition in divergent thinking: Theoretical, experimental, neuroimaging evidence, and new directions. In *The Cambridge Handbook of the Neuroscience of Creativity*. Rex E. Jung and Oshin Vartanian (Eds.), Cambridge University Press, Cambridge, UK, 333–362. <https://doi.org/10.1017/9781316556238>
 - [39] Anaëlle Camarda, Grégoire Borst, Marine Agogue, Marianne Habib, Benoît Weil, Olivier Houdé, and Mathieu Cassotti. 2018. Do we need inhibitory control to be creative? Evidence from a dual-task paradigm. *Psychology of Aesthetics, Creativity, and the Arts* 12, 3 (2018), 351–358. <https://doi.org/10.1037/aca0000140>
 - [40] Gabriela Krumm, Vanessa Arán Filippetti, and Marisel Gutierrez. 2018. The contribution of executive functions to creativity in children: What is the role of crystallized and fluid intelligence? *Thinking Skills and Creativity* 29 (2018), 185–195. <https://doi.org/10.1016/j.tsc.2018.07.006>
 - [41] Jacob W. Getzels and Mihaly Csikszentmihalyi. 1976. *The Creative Vision: A Longitudinal Study of Problem Finding in Art*. John Wiley & Sons, New York, NY.
 - [42] Michael D. Mumford, Michele I. Mobley, Roni Reiter-Palmon, Charles E. Uhlman, and Lesli M. Doares. 1991. Process analytic models of creative capacities. *Creativity Research Journal* 4, 2 (1991), 91–122. <https://doi.org/10.1080/10400419109534380>
 - [43] Roni Reiter-Palmon and Erika J. Robinson. 2009. Problem identification and construction: What do we know, what is the future? *Psychology of Aesthetics, Creativity, and the Arts* 3, 1 (2009), 43–47. <https://doi.org/10.1037/a0014629>
 - [44] Ronald A. Finke. 1996. Imagery, creativity, and emergent structure. *Consciousness and Cognition* 5, 3 (1996), 381–393. <https://doi.org/10.1006/ccog.1996.0024>
 - [45] Oded M. Kleinmuntz, Tal Ivancovsky, and Simone G. Shamay-Tsoory. 2019. The two-fold model of creativity: The neural underpinnings of the generation and evaluation of creative ideas. *Current Opinion in Behavioral Sciences* 27 (2019), 131–138. <https://doi.org/10.1016/j.cobeha.2018.11.004>
 - [46] Dean Keith Simonton. 2023. The blind-variation and selective-retention theory of creativity: Recent developments and current status of BVSR. *Creativity Research Journal* 35, 3 (2023), 304–323. <https://doi.org/10.1080/10400419.2022.2059919>
 - [47] Paul T. Sowden, Andrew Pringle, and Liane Gabora. 2015. The shifting sands of creative thinking: Connections to dual-process theory. *Thinking & Reasoning* 21,

- 1 (2015), 40–60. <https://doi.org/10.1080/13546783.2014.885464>
- [48] Izabela Lebuda and Mathias Benedek. 2023. A systematic framework of creative metacognition. *Physics of Life Reviews* 46 (2023), 161–181. <https://doi.org/10.1016/j.plrev.2023.07.002>
- [49] Aleksandra Zielińska, Izabela Lebuda, Marta Czerwonka, and Maciej Karwowski. 2025. Self-regulation prompts improve creative performance. *The Journal of Creative Behavior* 59, 1 (2025), Article e674. <https://doi.org/10.1002/jocb.674>
- [50] Thomas T. Hills and Yoed N. Kenett. 2025. An entropy modulation theory of creative exploration. *Psychological Review* 132, 1 (2025), 239–251. <https://doi.org/10.1037/rev0000511>
- [51] Zeynep Akata, Dan Balliet, Maarten de Rijke, Frank Dignum, Virginia Dignum, Gusztai Eiben, Antske Fokkens, Davide Grossi, Koen V. Hindriks, Holger H. Hoos, Hayley Hung, Catholijn M. Jonker, Christof Monz, Mark Neerincx, Frans Oliehoek, Henry Prakken, Stefan Schlobach, Linda van der Gaag, Frank van Harmelen, Herke van Hoof, Birna van Riemsdijk, Aimee van Wylsberghe, Rineke Verbrugge, Bart Verheij, Piek Vossen, and Max Welling. 2020. A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer* 53, 8 (Aug. 2020), 18–28. <https://doi.org/10.1109/MC.2020.2996587>
- [52] Dominik Dellermann, Philipp Ebel, Matthias Söllner, and Jan Marco Leimeister. 2019. Hybrid intelligence. *Business & Information Systems Engineering* 61, 5 (2019), 637–643. <https://doi.org/10.1007/s12599-019-00595-2>
- [53] Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, Wayne Xin Zhao, Zhewei Wei, and Ji-Rong Wen. 2024. A survey on large language model based autonomous agents. *Frontiers of Computer Science* 18, 6 (2024), Article 186345. <https://doi.org/10.1007/s11704-024-40231-1>
- [54] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, Yuan Cao. ReAct: Synergizing reasoning and acting in language models. arXiv: 2210.03629. 2022. <https://doi.org/10.48550/arXiv.2210.03629>
- [55] Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2023. Reflexion: Language agents with verbal reinforcement learning. In *Advances in Neural Information Processing Systems* 36. Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (Eds.), Curran Associates, Inc., Red Hook, NY, 8634–8652. <https://doi.org/10.52202/075280-0377>
- [56] Omar Khattab, Arnav Singhvi, Paridhi Maheshwari, Zhiyuan Zhang, Keshav Santhanam, Sri Vardhamanan, Saiful Haq, Ashutosh Sharma, Thomas T. Joshi, Hanna Moazam, Heather Miller, Matei Zaharia, and Christopher Potts. 2023. DSPy: Compiling declarative language model calls into self-improving pipelines. arXiv:2310.03714. Retrieved from <https://arxiv.org/abs/2310.03714>
- [57] Navid Hashemi Tonekaboni and Saber Soleymani. 2026. Socratic method revisited: Human-AI dialogue for knowledge creation and internalization. In *Proceedings of the 59th Hawaii International Conference on System Sciences (HICSS 2026)*, 5192–5201. Retrieved from <https://hdl.handle.net/10125/112020>
- [58] Zorana Ivcevic, Eliana R. Grossman, Katherine N. Cotter, and Emily Nusbaum. 2024. Self-regulation of creativity: Toward measuring strategies of creative action. *Creativity Research Journal* 36, 3 (2024), 491–507. <https://doi.org/10.1080/10400419.2023.2226494>
- [59] Yoonsu Kim, Jueon Lee, Seoyoung Kim, Jaehyuk Park, and Juho Kim. 2024. Understanding Users' Dissatisfaction with ChatGPT Responses: Types, Resolving Tactics, and the Effect of Knowledge Level. In *Proceedings of the 29th International Conference on Intelligent User Interfaces (IUI '24)*. Association for Computing Machinery, New York, NY, USA, 385–404. <https://doi.org/10.1145/3640543.3645148>
- [60] Jacob Sherson, Janet Rafner, Sune Bo, Bianka Szöllösi, Florent Vinchon, Paul-Arno Lamarque, and Florian de Boni. 2026b. Designing for long-term impact: A hybrid intelligence case study in AI-assisted mentalizing-based therapy. In *HHAI 2026: Proceedings of the 5th International Conference on Hybrid Human-Artificial Intelligence*. (Eds.), Frontiers in Artificial Intelligence and Applications, Vol. 423. IOS Press, Amsterdam, 532–535. <https://doi.org/10.3233/FAIA260552>