Recalibrating Academic Expertise in the Age of Generative AI

Zhicheng Lin[1] & Aamir Sohail[2,3*]

[1] Department of Psychology, Yonsei University, Seoul, 03722, Republic of Korea.
[2] Centre for Human Brain Health, School of Psychology, University of Birmingham, Birmingham, B15 2TT, UK.
[3] Institute for Mental Health, School of Psychology, University of Birmingham, Birmingham, B15 2TT, UK.

*Corresponding author:
Email: axs2210@student.bham.ac.uk

THE BIGGER PICTURE

When researchers use GenAI tools to design experiments, summarize literature, or write code in seconds, this promises to accelerate discovery but threatens the cognitive foundations that underpin scientific expertise. Expertise is forged through the slow, often difficult process of wrestling with complex problems, finding connections, and learning from mistakes. If we delegate these essential processes to AI, we risk creating a generation of researchers who lack the foundational abilities to think, write, and reason independently. This Perspective argues that for science to thrive in the age of AI, we must redefine and cultivate what expertise means: beyond traditional knowledge, researchers need meta-skills in strategic direction, critical discernment, and systematic calibration of AI systems. We propose a framework for training researchers as discerning AI collaborators who maintain human judgment and scientific autonomy, ensuring that AI serves as a launchpad for ingenuity, not a stifling substitute.

SUMMARY

The integration of GenAI into academic workflows represents a fundamental shift in scientific practice. While these tools can amplify productivity, they risk eroding the cognitive foundations of expertise by simulating the very tasks through which scientific competence is developed, from synthesis to experimental design to writing. Uncritical reliance can lead to skill atrophy and AI complacency. We propose a framework of essential AI meta-skills: strategic direction, critical discernment, and systematic calibration. These constitute a new form of scientific literacy that builds on traditional critical thinking. Through domain-specific examples and a pedagogical model based on situated learning, we show how these meta-skills can be cultivated to ensure that researchers, particularly those in training, maintain intellectual autonomy. Without deliberate cultivation of these meta-skills, we risk creating the first generation of researchers who serve their tools rather than directing them.

INTRODUCTION

For better or worse, GenAI has fundamentally altered scientific practice.[1] Across disciplines, researchers increasingly deploy large language models (LLMs) for tasks once considered demonstrations of scientific skill—from literature synthesis to experimental design to manuscript writing.[2,3] Indeed, analyses of the U.S. labour market find that scientists and researchers are among the occupational groups with highest task exposure to LLM-driven transformation.[4]

GenAI represents the latest step in this "emerging general method of invention"[5] in science, namely AI. Over the past decade, this method has evolved—from neural networks detecting patterns in complex datasets to machine learning models generating testable hypotheses across disciplines.[5,6] GenAI continues this trajectory while crossing a critical threshold: the dawn of what has been termed the fifth era of science, where AI can autonomously generate hypotheses and drive discovery, at least in constrained domains and tasks.[7,8]

Rather than augmenting specific analytical capabilities, GenAI directly simulates the broad cognitive activities through which academic expertise has historically developed: synthesis, argumentation, writing. This progression creates a paradox with troubling implications for scientific training: the same tools that amplify productivity[9] may simultaneously erode the cognitive foundation of competence.[10] Moreover, when AI is used for direct content generation, it can fundamentally alter the researcher's relationship with their work, diminishing the sense of ownership, pride, and accountability for the final product.[11,12]

In this Perspective, we argue that working effectively with GenAI while maintaining scientific expertise and autonomy necessitates the development of essential AI meta-skills that constitute a new form of scientific literacy: strategic direction, critical discernment, and systematic calibration. Through domain-specific examples and a pedagogical model based on situated learning, we propose how these meta-skills can be cultivated to ensure that researchers, particularly those in training, maintain intellectual autonomy. We demonstrate their necessity by highlighting the concerning patterns of AI complacency, where users progressively disengage from critical evaluation and develop ritualized acceptance of authoritative-seeming outputs. Proactive development of this AI-augmented expertise is necessary to shape a future where technology amplifies, rather than attenuates, scientific thought.[13]

EPISTEMIC CHALLENGES AND OPPORTUNITIES

The AI complacency problem

Unlike previous technological transitions, GenAI creates a qualitatively different relationship between scientists and their tools.[14,15] When researchers delegate central scientific tasks to an LLM—literature reviews, code generation, methodology design—they engage in something far more consequential than traditional cognitive offloading.[16] This is problematic, as such tasks have historically been the primary mechanisms through which scientific expertise is developed.

Recent studies document concerning patterns of overreliance and deskilling in human–AI collaboration.[17] Studies of professionals using AI writing assistants reveal a progressive disengagement with content, where users initially scrutinize AI outputs but gradually accept them with diminishing critical evaluation[11]—a pattern particularly consequential for technical material.[18] The effortful retrieval that consolidates knowledge, the productive struggle that drives expertise development, and the incubation periods that foster creative insights are all potentially compromised.

Beyond these individual cognitive losses, the problem extends to fundamental questions of epistemic authority: how humans calibrate trust when interacting with AI systems. Research on human–AI interactions reveals that users begin with high initial trust based on novelty or perceived sophistication but often fail to develop the context-dependent trust calibration that skilled AI use requires.[19] Unlike traditional tools, GenAI produces outputs that appear authoritative and comprehensive yet may contain subtle inaccuracies or "hallucinations" that demand verification and domain expertise.[20] The behavioral mechanisms behind skill atrophy operate as a self-reinforcing cycle: when AI consistently provides satisfactory output, the perceived "cost" of verification increases, leading to progressive disengagement.[21]

This cycle reflects "ritualized practices" identified in digital literacy—habituated, unreflective workflows that users develop to manage cognitive load.[22] However, GenAI systems amplify these dynamics through three distinctive mechanisms that make such rituals particularly insidious. First, unlike search results that present fragmented information requiring synthesis, AI outputs appear as complete, authoritative arguments that simulate finished, expert reasoning. Second, their conversational, human-like presentation masks the statistical nature of their construction, making verification feel unnecessary. Third, the immediate satisfaction of receiving comprehensive-seeming answers accelerates ritual formation. Indeed, users consistently overestimate an LLM's accuracy and are swayed by superficial heuristics, such as the length of an explanation, rather than its substance.[23]

The ritualized acceptance of AI authority manifests distinctively across scientific domains. Programming proficiency appears particularly vulnerable as code-completion tools now generate entire functions or classes with minimal human input. Observational studies of programming students reveal the development of maladaptive habits, such as falling into unproductive cycles of submitting incorrect AI-generated code and then asking the AI to fix its own errors, rather than engaging in the debugging and reasoning process themselves.[24] Without regular practice in syntax, algorithmic logic, and debugging, they may find their ability to develop novel computational approaches gradually diminishing. Similarly, research design and methodological reasoning—skills honed through years of training and practice—may atrophy if consistently delegated to AI systems.[25]

The danger thus lies in breeding pseudo-competence: researchers may feel methodologically sophisticated because they can generate plausible-sounding research designs through AI interaction, while remaining unable to recognize fundamental flaws or innovative opportunities that require genuine expertise. This dynamic exemplifies the classic “ironies” of automation, where technology intended to reduce workload paradoxically creates new cognitive burdens.[21,26,27] The user’s role shifts from active production to passive evaluation of AI outputs, a mentally taxing task that degrades situational awareness and encourages over-reliance. The AI simplifies routine tasks but makes cognitively demanding work—validating methodologies, identifying subtle errors—even harder. Uncritical GenAI use thus fosters superficial competence that falters when faced with genuine scientific complexity.

Perhaps most concerning are the impacts on scientific creativity. Innovation emerges from well-developed internal knowledge that reduces working memory demands and enables complex mental schemata.[28] Uncritical delegation of foundational tasks that build these schemata risks intellectual dependency. True understanding requires physically encoding information into efficient neural pathways through effortful engagement—the “cognitive friction” of grappling directly with a problem. This friction is not an impediment but the catalyst for unexpected connections that drive breakthroughs. If AI increasingly mediates this engagement, we risk narrowing conceptual exploration to paths already optimized in training data—intellectual path dependency that constrains rather than expands scientific horizons.[11,29,30]

Potential benefits

Despite these concerns, GenAI can strengthen certain academic capabilities when thoughtfully integrated into scholarly workflows.

First, learning through exemplars has long been central to academic development. AI systems can function as always-available demonstrations—providing models of writing, code, or analytical approaches that researchers can learn from, though requiring expert evaluation of quality.[3] This “apprenticeship” function particularly benefits early-career researchers or those without ready peer feedback. Just as writers develop their craft by studying masterful prose, researchers can learn from AI-generated examples that illustrate effective structures, arguments, or implementations.

Second, AI’s accelerated feedback loop addresses a key limitation in traditional skill development. Rather than waiting weeks for feedback, researchers receive immediate, detailed responses that facilitate rapid iteration and learning, compressing the learning cycle.[1] A researcher drafting a methodology section can instantly receive suggestions for improvements—feedback that might otherwise require multiple rounds of peer review. However, without strategic prompting, AI feedback often emphasizes positive reinforcement over critical evaluation, potentially promoting convergence toward common patterns and homogenizing scholarly voice (**Box 1**).[31]

Third, AI assistance may expand intellectual engagement by reducing time spent on mechanical work. When basic drafting or coding requires less time, researchers can engage with a broader range of ideas, methods, and collaborations—enhancing conceptual flexibility and interdisciplinary thinking. This expanded engagement might foster connections between separate domains, a key ingredient in innovation.

However, these benefits depend on how researchers approach AI tools. When used with deliberate intention to learn—critically evaluating outputs and treating AI as collaborative partner rather than replacement—these systems function as skill amplifiers rather than substitutes.

Box 1: Example strategic prompts for critical AI engagement

The following prompts demonstrate how researchers can strategically direct AI systems to avoid convergence toward common patterns and instead promote critical evaluation and intellectual exploration. These examples illustrate the "strategic direction" component of our framework, showing how careful prompt design can elicit more rigorous, divergent thinking from AI systems.

Manuscript review and critique: Act as a seasoned expert in [field]. Critically evaluate this manuscript as if reviewing for [journal]. Show potential reasons for rejection and list multiple key reasons. For each key reason, use two or more sub-bullet points to further clarify and support your arguments in painstaking detail. Be as specific and detailed as possible.

Divergent research design: Generate 4 fundamentally different experimental approaches to investigate [research question]. For each design, identify its unique strengths, critical limitations, and potential confounds. Explain why a researcher might choose each approach despite its limitations.

Literature synthesis with dissenting views: Summarize the literature on [topic], but specifically emphasize: (1) unresolved contradictions between studies, (2) methodological limitations that undermine confident conclusions, and (3) alternative theoretical interpretations that challenge the dominant narrative.

Methodological vulnerability assessment: You are a hostile reviewer trying to identify fatal flaws in this research design. List every possible threat to validity, confound, or limitation. For each flaw, rate its severity and suggest how it could be empirically tested or addressed.

Alternative interpretation generation: Given these research findings, generate 3-4 competing explanations that could account for the same data. Rank them by plausibility and identify the critical experiments needed to distinguish between them.

DEVELOPING AI META-SKILLS

Framework

Working with AI's unique capabilities necessitates specific meta-skills.[18] This involves determining the suitability of AI for specific tasks, discerning the output, calibrating results appropriately, and iteratively resubmitting when needed. It also requires accurately assessing the limits of one's own knowledge, knowing when to engage or trust AI versus trust one's knowledge or sources.[32] This framework emphasizes how scientists can strategically guide AI systems while critically evaluating outputs in an ongoing cycle (**Fig. 1**). These interrelated components constitute a novel form of scientific expertise as essential as domain knowledge itself.

Crucially, while building upon critical thinking foundations, AI meta-skills represent a distinct intellectual engagement. Traditional critical thinking evaluates static information—assessing the validity of claims, identifying logical flaws, or weighing evidence. AI meta-skills require dynamic collaboration with generative systems that produce novel content through strategic direction. Rather than judging predetermined outputs, researchers must iteratively shape AI behavior through prompt engineering and tool selection,[33] recognize unique failure modes of statistical language models (such as hallucinations appearing authoritative), and calibrate machine-generated content through systematic validation. This procedural, collaborative dimension distinguishes AI meta-skills from conventional critical thinking: researchers become active directors of iterative knowledge-generation, requiring metacognition—the ability to monitor and control one's own thinking.[32,34]

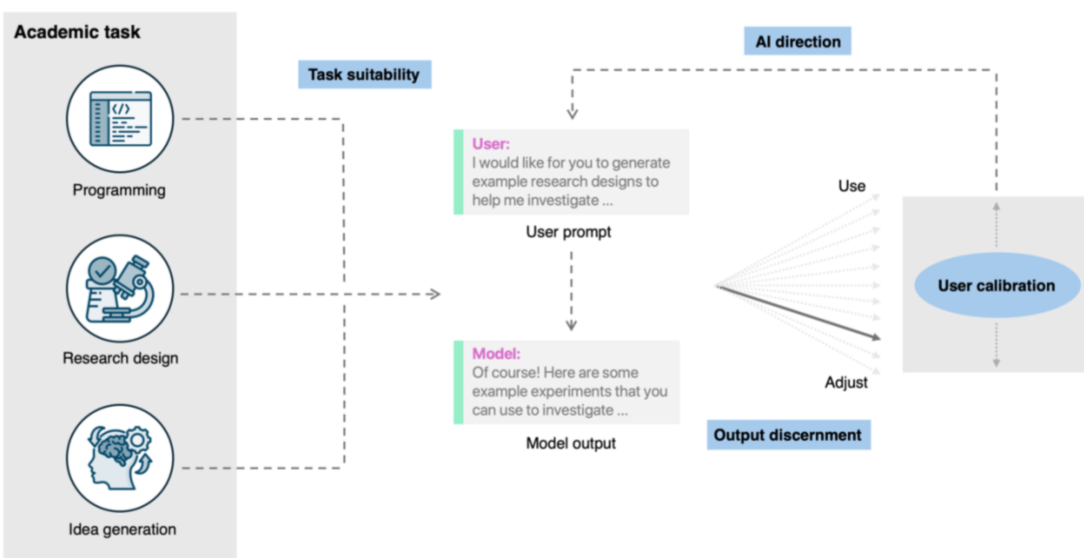

Figure 1. *Academics currently use GenAI to facilitate a variety of tasks. This process can be adapted to maintain scientific autonomy in several areas. This includes understanding the task's suitability for the model, discerning the model output, and calibrating it appropriately. The user may also iteratively refine AI outputs through resubmission. Figure designed using resources from Flaticon.com.*

The framework begins with identifying which aspects of scientific work are appropriate for AI delegation versus direct human engagement. This requires understanding AI's epistemological limitations in specific domains and research contexts.[35] Researchers must then construct prompts with contextual structures and constraints that guide AI reasoning toward scientifically valid approaches. Providing necessary domain knowledge and strategically framing questions elicits comprehensive, well-reasoned responses rather than simplistic answers. This also includes divergent prompting, where researchers guide AI systems to explore multiple alternatives. By preventing premature convergence on seemingly optimal but potentially limited solutions, divergent prompting fosters intellectual exploration and unexpected insights.

Following strategic direction, users must identify subtle inaccuracies, conceptual flaws, or reasoning errors in authoritative-seeming AI outputs.[20] This "discernment" dimension is particularly crucial where one lacks sufficient expertise. For instance, AI systems often exacerbate citation inequality, disproportionately recommending highly-cited papers and potentially skewing representation of scholars from different backgrounds.[36] Researchers must also develop methodological assessment skills to evaluate AI-suggested research approaches, understanding how generative models represent scientific methodologies and recognizing that AI suggestions reflect statistical patterns rather than causal or theoretical understanding.

Following discernment, users calibrate the generated response based on content suitability and their domain expertise. If unsatisfied, users re-evaluate the appropriateness of the selected model or tool—and resubmit with additional prompting as needed.

These components operate in a continuous feedback loop where direction shapes what needs discernment, and discernment informs subsequent direction. This dynamic interaction constitutes a new scientific literacy that bridges human cognitive capacities with machine intelligence.

Meta-skills in practice

The following cases from recent research demonstrate how these meta-skills operate across diverse domains, revealing both sophisticated applications and critical failures. Each domain illustrates the iterative cycle of strategic direction, critical discernment, and systematic calibration that defines effective human–AI collaboration. Empirical evidence suggests that successful AI integration depends critically on existing domain expertise, with AI systems producing subtle but fundamental errors that non-experts consistently fail to detect.[37] These cases validate why strategic direction, critical discernment, and rigorous calibration represent essential capabilities rather than optional enhancements to traditional research skills.

Literature analysis, ideation, and writing

Literature analysis exemplifies how meta-skills operate at the fundamental level of tool selection and workflow design. Strategic direction requires distinguishing between general-purpose chatbots—which often lack access to paywalled articles and fabricate citations (particularly without internet access)—and specialized, domain-specific tools designed for scholarly tasks (e.g., NotebookLM).[38] Sophisticated direction involves choosing appropriate instruments for specific research functions rather than defaulting to familiar AI systems.

Critical discernment becomes essential when evaluating AI-generated literature summaries, as researchers must cross-reference summaries with original sources and distinguish genuine scholarly consensus from algorithmic artifacts.[39] Effective calibration requires systematic validation that cannot be delegated to AI systems—breaking complex analyses into manageable components for individual verification rather than accepting AI-generated output wholesale.[40] These verification processes represent irreducible intellectual labor defining scholarly competence.

Ideation reveals that strategic direction involves assigning specific roles to AI systems—designer, writer, interviewer, or actor—depending on research needs.[41] However, discernment requires recognizing that current models excel at incremental refinements while conceptual breakthroughs still require human insight.[29]

Academic writing demonstrates this pattern through systematic collaboration frameworks. Strategic direction operates through a two-stage model: AI-inspired phases for expansive tasks like brainstorming and structural organization, followed by AI-assisted phases for focused refinement like drafting and language polishing.[3] Critical discernment becomes essential to avoid de-skilling over time, requiring verification of AI-generated content and recognizing that essential elements—insight, originality, and creativity—cannot be fully replicated by current systems. Effective calibration employs human-in-the-loop editing where writers maintain ultimate authority, iteratively integrating or discarding AI suggestions while tracking contributions for transparent disclosure.

Natural sciences

Examination of specific domains reveals how these meta-skills adapt to different validation systems and domain characteristics. In pure mathematics, researchers demonstrate strategic direction through three approaches: using automated theorem provers to build proofs from foundational axioms, analyzing scientific literature as linguistic data to identify patterns, and directing machine learning systems to examine mathematical objects and formulate novel conjectures.[42] Critical discernment operates through formal frameworks like the "Birch test," evaluating AI discoveries based on whether they are automatic, interpretable to human experts, and non-trivial. Chemistry and biology showcase AI agents executing complex experimental protocols autonomously, from conducting chemical reactions to designing CRISPR gene-editing experiments.[43]

However, calibration differs across domains: mathematics benefits from formal verification systems where hypotheses generated by LLMs can be translated into formal languages like Lean for definitive logical validation, while experimental sciences rely on multi-agent debate systems and iterative self-correction loops as calibration mechanisms.

Social sciences

The systematic validation approaches in natural sciences exemplify underlying meta-skill patterns that manifest consistently across domains, despite different tools and techniques. In finance and economics, these patterns appear through systematic AI-driven research workflows that demonstrate sophisticated strategic direction by alternating between "human-driven exploration" (leveraging LLMs for ideation and question structuring) and "data-driven exploration" (using AI for pattern detection in complex datasets)[44]. This mirrors the contextual assessment seen in natural sciences: researchers evaluate research context to determine appropriate AI engagement modes, with conceptual studies benefiting from human-driven ideation while empirical analyses require data-driven pattern detection.

Critical discernment becomes evident when interpreting AI-detected patterns in trading data and social media sentiment, requiring integration into theoretical frameworks rather than accepting algorithmic outputs as self-evident insights.[45] This workflow implicitly embeds calibration as a core feature: iterative refinement between exploration modes ensures that AI-generated insights undergo continuous human validation and theoretical integration.

These emerging patterns perhaps find their most complex expression in psychology, where meta-skills confront fundamental challenges in literature synthesis, experimental design, statistical analysis, and participant simulation. A comprehensive case study reveals how meta-skills operate throughout an entire AI-assisted project.[46] Researchers employed strategic direction through a "drill-down" approach, iteratively prompting AI to move from broad research areas ("Digital Consumption and Mental Health") to specific, testable hypotheses about "Ethical Fatigue."

However, constant discernment proved essential: AI-generated literature reviews were unusable due to hallucinated citations and inability to access paywalled journals, experimental designs contained fatal confounds requiring reconstruction, and statistical analyses appeared plausible while containing major errors. Calibration may thus require substantial human override.

Advanced applications require both sophisticated direction and discernment. Researchers now direct LLMs to generate personalized experimental stimuli, tailoring persuasive messages to participants' personality traits or creating real-time dialogues to challenge misinformation.[47] Yet these applications reveal prompt fragility—minor changes to instructions can produce dramatically different outputs, requiring researchers to discern which variations reflect experimental manipulations versus algorithmic artifacts.

Effective calibration employs various validation strategies: comparing AI coding against holdout self-reports for internal states, holdout human ratings for observational measures, and predictive accuracy for behavioral outcomes.[48] When researchers use LLMs to simulate human participants, calibration requires recognizing that outputs represent token predictions rather than cognition, demanding systematic validation against actual human data.[49]

Framework validation

These findings validate our theoretical framework while revealing its practical complexity. The meta-skills required for effective human–AI collaboration represent sophisticated capabilities that must be deliberately cultivated rather than assumed to emerge naturally from AI exposure. Without such cultivation, researchers risk the progressive skill atrophy described earlier, becoming dependent on technologies they cannot critically evaluate or strategically direct. The consistent pattern across domains—AI augmenting human capabilities within specific constraints rather than replacing human judgment—reinforces the essential nature of these meta-skills for maintaining scientific autonomy as AI becomes ubiquitous.

Repositioning AI for academic research and education

Rather than positioning AI as either a threat or savior to academic work, we must develop forward-looking approaches that harness its benefits while mitigating risks.[50] The following principles and stakeholder-specific interventions can guide this balanced integration:

- *Designing for complementarity:* AI developers serving academic markets should prioritize interfaces that promote active engagement rather than passive consumption.[51] This may include presenting multiple alternative approaches, explicitly highlighting uncertainties, or requiring substantive user input before generating complex outputs. Systems could build literacy-enhancing components directly into interfaces, such as automated confidence indicators, verifiable citation sources, and interactive elements that prompt users to critically evaluate outputs before accepting them.
- *Research-centric AI literacy development:* Understanding how and when to integrate AI with one's own knowledge and recognizing the consequences of over-reliance are essential skills in AI-centric research. Yet, most researchers report lacking guidance and training, preventing optimal (and responsible) AI use.[52] Academic bodies should develop critical AI literacy among users, framing AI engagement as a catalyst for skill development, not just a productivity enhancer. This literacy must extend to working practices; researchers should learn to fact-check outputs and identify tasks that should not be delegated due to ethical and security risks, including intellectual property and data confidentiality. A key meta-skill is selecting secure AI systems that explicitly guarantee user data will not be used for model training.[53]

These principles must be operationalized through specific interventions by key stakeholders:

- *For funding agencies:* Allocate a percentage of major research grants toward developing discipline-specific AI literacy, including direction–discernment–

calibration training. Fund research into how AI adoption impacts research processes, skill development (including potential atrophy), and scientific creativity across disciplines and career stages. Require grant proposals to include explicit AI integration strategies and usage disclosure.

- *For universities:* Develop and integrate dedicated AI direction–discernment–calibration curricula into existing graduate research methods, statistics, and ethics courses, emphasizing hands-on direction and evaluation of AI-generated content. Implement robust faculty development programs to equip instructors to effectively teach and model direction–discernment–calibration skills. Create "AI-augmented" research certification programs that specifically assess the ability to critically evaluate and direct AI systems.
- *For scientific journals:* Institute standardized, granular disclosure requirements detailing both AI use and human verification procedures employed. Develop specialized review protocols and provide training for peer reviewers on evaluating AI-assisted manuscripts, identifying signs of uncritical AI reliance and assessing the adequacy of reported validation procedures.
- *For scientific societies:* Establish discipline-specific guidelines for ethical and effective AI integration. Create dedicated working groups to develop domain-relevant open materials for cultivating AI meta-skills and metrics for assessing this capacity. Foster community platforms (repositories, forums, journal space) for sharing best practices, validated prompts, and effective AI workflows.

These interventions should be supported by systematic research on AI resource inequality. Well-resourced universities can invest in robust AI training, helping faculty and students learn best practices that mitigate skill erosion. In contrast, smaller or under-resourced programs may continue with maladaptive practices, with minimal institutional guidance in critical AI literacy. This emerging AI divide risks creating a stratified research landscape and amplifying existing societal disparities.[1]

For instance, recent work shows that the advent of ChatGPT widened the academic productivity gap between male and female researchers, driven by gender differences in adoption and usage patterns.[54] Such findings underscore that special attention must be paid not only to researchers from resource-limited settings but also to how technology adoption intersects with other structural inequities. This includes dedicated funding for training programs and certification opportunities that can be conducted remotely or with minimal technological requirements to prevent further exacerbation of existing inequalities.[55]

Addressing these challenges requires developing pedagogical approaches that effectively cultivate AI meta-skills across diverse institutional contexts. We propose a situated learning framework that recognizes how direction, discernment, and calibration capabilities develop in practice (**Box 2**). Drawing on situated learning theory and cognitive apprenticeship models,[56,57] this approach helps students develop AI meta-skills by practicing within real research contexts, rather than by mastering abstract principles before engaging with actual tools.

These meta-skills and domain knowledge develop through mutual reinforcement rather than prerequisite accumulation. Unlike approaches that dichotomize technical skills like prompt engineering from human-centered capabilities,[58] this framework recognizes that robust direction, discernment, and calibration capabilities emerge through direct engagement with the systems researchers will encounter in their careers, making the specificity of AI-focused meta-skill training a pedagogical strength rather than limitation.[59]

Box 2: Developing AI direction, discernment, and calibration skills through situated learning

A fundamental challenge emerges in cultivating AI meta-skills: how can students develop such skills without already possessing substantial domain expertise? Direction, discernment, and calibration capabilities cannot develop through abstract principles alone but emerge through scaffolded engagement with authentic research materials, where domain knowledge and AI meta-skills develop synergistically rather than sequentially.

Progressive skill development framework
AI meta-skills develop through four interconnected stages that address the expertise challenge while building transferable capabilities.

Stage 1: Foundational pattern recognition Students begin by developing basic direction skills—learning to frame clear research questions and provide adequate context to AI systems—while identifying quality indicators: fabricated citations with suspicious titles, internal logical contradictions, statistical impossibilities, or claims that violate basic constraints. Initial calibration involves binary decisions (accept/reject) rather than nuanced adjustments. These foundational skills transfer across domains because they rely on general reasoning and intuition rather than specialized knowledge.

Stage 2: Collaborative meta-skill distribution Advanced users engage alongside novices in directing AI systems and evaluating outputs, creating distributed cognition systems that leverage complementary capabilities. Experts model sophisticated prompting strategies (direction) and provide domain-specific evaluation criteria (discernment), while novices contribute fresh perspectives on AI-generated content and alternative calibration approaches. This collaborative approach develops all three meta-skills while avoiding the prerequisite expertise conundrum.

Stage 3: Strategic integration development Students learn to coordinate direction, discernment, and calibration as integrated capabilities rather than separate skills. They develop systematic prompting strategies that incorporate verification requirements (strategic direction), apply domain-general validation protocols alongside emerging domain knowledge (contextual discernment), and make graduated adjustments rather than binary decisions (sophisticated calibration). These integrated strategies create reliable frameworks for AI engagement while students simultaneously build domain expertise.

Stage 4: Authentic research embedding Students apply emerging meta-skills within their own research projects, where personal investment and real consequences accelerate skill development. Direction skills develop through actual research questions, discernment through evaluating AI outputs that affect their work, and calibration through iterative refinement in authentic contexts. Faculty guidance ensures quality standards while students develop ownership of their AI collaboration process.

Implementation across academic career stages

This framework adapts to different academic development levels. Undergraduates focus on foundational pattern recognition and collaborative meta-skill development through structured exercises. Graduate students emphasize strategic integration and authentic research embedding, applying coordinated meta-skills to thesis work and publications. Faculty model expert-level integration while remaining open to student insights about emerging AI capabilities, recognizing that technological evolution requires continuous meta-skill refinement across all career stages.

Domain-specific adaptation requirements

This framework provides foundational principles that must be adapted to specific disciplinary contexts, research cultures, and institutional environments. Effective AI meta-skills manifest differently across fields. No single pedagogical approach can address the full spectrum of disciplinary requirements and epistemic standards.

Developing these capabilities requires moving beyond abstract discussions to concrete engagement with domain-specific research questions, methodological challenges, and validation standards that students encounter in their fields. A chemistry student learning AI-assisted experimental design faces fundamentally different challenges than a historian learning to discern AI-generated source analyses.

This necessitates that instructors within each domain develop tailored guidelines. The framework offers organizing principles rather than prescriptive procedures. Domain experts must translate these stages into specific practices aligning with their students' research trajectories and professional preparation needs.

Concluding remarks

Without specific training, researchers risk developing maladaptive practices, failing to engage with AI critically due to lack of awareness, motivation, or ability.[60] Combating this requires developing AI-specific skills and self-confidence, enabling users to strategically submit tasks and evaluate responses iteratively. These meta-skills help maintain user control and ensure intellectual investment, thereby managing ownership dilution and preserving scholarly integrity and motivation.[12] Without deliberate intervention to cultivate these skills, we risk creating a generation of researchers increasingly dependent on technologies they neither fully understand nor effectively direct. By acknowledging this paradox and proactively cultivating AI-augmented expertise, we can shape a future where technology genuinely amplifies rather than attenuates scientific thought.

Acknowledgments

A.S. is supported by an MRC AIM iCASE DTP Studentship (Ref: MR/W007002/1).

Declaration of Interests

The authors declare no conflicts of interest.

Author Contributions:

Z.L. and A.S. wrote the paper.

Declaration of AI Use

Z.L. used Claude Sonnet 4 and Gemini 2.5 Pro to proofread the manuscript, following the prompts described at https://www.nature.com/articles/s41551-024-01185-8.

References

1 Lin, Z. Why and how to embrace AI such as ChatGPT in your academic life. *R. Soc. Open Sci.* **10**, 230658 (2023). https://doi.org/10.1098/rsos.230658

2 Van Quaquebeke, N., Tonidandel, S. & Banks, G. C. Beyond efficiency: How artificial intelligence (AI) will reshape scientific inquiry and the publication process. *The Leadership Quarterly* **36**, 101895 (2025). https://doi.org/10.1016/j.leaqua.2025.101895

3 Lin, Z. Techniques for supercharging academic writing with generative AI. *Nat. Biomed. Eng.* **9**, 426-431 (2025). https://doi.org/10.1038/s41551-024-01185-8

4 Eloundou, T., Manning, S., Mishkin, P. & Rock, D. GPTs are GPTs: Labor market impact potential of LLMs. *Science* **384**, 1306-1308 (2024). https://doi.org/10.1126/science.adj0998

5 Bianchini, S., Müller, M. & Pelletier, P. Artificial intelligence in science: An emerging general method of invention. *Research Policy* **51**, 104604 (2022). https://doi.org/10.1016/j.respol.2022.104604

6 Ludwig, J. & Mullainathan, S. Machine learning as a tool for hypothesis generation. *The Quarterly Journal of Economics* **139**, 751-827 (2024). https://doi.org/10.1093/qje/qjad055

7 Miolane, N. The fifth era of science: Artificial scientific intelligence. *PLOS Biol.* **23**, e3003230 (2025). https://doi.org/10.1371/journal.pbio.3003230

8 Musslick, S. *et al.* Automating the practice of science: Opportunities, challenges, and implications. *Proc. Natl. Acad. Sci. U. S. A.* **122**, e2401238121 (2025). https://doi.org/10.1073/pnas.2401238121

9 Noy, S. & Zhang, W. Experimental evidence on the productivity effects of generative artificial intelligence. *Science* **381**, 187-192 (2023). https://doi.org/10.1126/science.adh2586

10 Kosmyna, N. *et al.* Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv* (2025). https://doi.org/10.48550/arXiv.2506.08872

11 Li, Z., Liang, C., Peng, J. & Yin, M. in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* Article 1048 (Association for Computing Machinery, Honolulu, HI, USA, 2024).

12 Hermann, E. Illusion, dilution, or loss: psychological ownership and GenAI. *Trends Cogn. Sci.* **29**, 215-217 (2025). https://doi.org/10.1016/j.tics.2024.12.011

13 Messeri, L. & Crockett, M. J. Artificial intelligence and illusions of understanding in scientific research. *Nature* **627**, 49-58 (2024). https://doi.org/10.1038/s41586-024-07146-0

14 Vaccaro, M., Almaatouq, A. & Malone, T. When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nat. Hum. Behav.* **8**, 2293-2303 (2024). https://doi.org/10.1038/s41562-024-02024-1

15 Heersmink, R. Use of large language models might affect our cognitive skills. *Nat. Hum. Behav.* **8**, 805-806 (2024). https://doi.org/10.1038/s41562-024-01859-y

16 León-Domínguez, U. Potential cognitive risks of generative transformer-based AI chatbots on higher order executive functions. *Neuropsychology* **38**, 293-308 (2024). https://doi.org/10.1037/neu0000948

17 Budzyń, K. *et al.* Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: A multicentre, observational study. *The Lancet Gastroenterology & Hepatology* (2025). https://doi.org/10.1016/S2468-1253(25)00133-5

18 Dell'Acqua, F. *et al.* Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper* (2023). https://doi.org/10.2139/ssrn.4573321

19 Glikson, E. & Woolley, A. W. Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals* **14**, 627-660 (2020). https://doi.org/10.5465/annals.2018.0057

20 Fouesneau, M. *et al.* What is the role of large language models in the evolution of astronomy research? *arXiv* (2024). https://doi.org/10.48550/arXiv.2409.20252

21 Parasuraman, R. & Manzey, D. H. Complacency and bias in human use of automation: An attentional integration. *Hum. Factors* **52**, 381-410 (2010). https://doi.org/10.1177/0018720810376055

22 Bhatt, I. & MacKenzie, A. Just Google it! Digital literacy and the epistemology of ignorance. *Teaching in Higher Education* **24**, 302-317 (2019). https://doi.org/10.1080/13562517.2018.1547276

23 Steyvers, M. *et al.* What large language models know and what people think they know. *Nat. Mach. Intell.* **7**, 221-231 (2025). https://doi.org/10.1038/s42256-024-00976-7

24 Rahe, C. & Maalej, W. How do programming students use generative AI? *Proceedings of the ACM on Software Engineering* **2**, Article FSE045 (2025). https://doi.org/10.1145/3715762

25 Macnamara, B. N. *et al.* Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness? *Cognitive Research: Principles and Implications* **9**, 46 (2024). https://doi.org/10.1186/s41235-024-00572-8

26 Simkute, A. *et al.* Ironies of generative AI: Understanding and mitigating productivity loss in human-AI interaction. *International Journal of Human–Computer Interaction* **41**, 2898-2919 (2025). https://doi.org/10.1080/10447318.2024.2405782

27 Shukla, P. *et al.* in *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* Article 171 (Association for Computing Machinery, 2025).

28 Oakley, B., Johnston, M., Chen, K.-Z., Jung, E. & Sejnowski, T. J. The memory paradox: Why our brains need knowledge in an age of AI. *arXiv* (2025). https://doi.org/10.48550/arXiv.2506.11015

29 Ding, A. W. & Li, S. Generative AI lacks the human creativity to achieve scientific discovery from scratch. *Sci. Rep.* **15**, 9587 (2025). https://doi.org/10.1038/s41598-025-93794-9

30 Doshi, A. R. & Hauser, O. P. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances* **10**, eadn5290 (2024). https://doi.org/10.1126/sciadv.adn5290

31 Steiss, J. *et al.* Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction* **91**, 101894 (2024). https://doi.org/10.1016/j.learninstruc.2024.101894

32 Lee, D., Pruitt, J., Zhou, T., Du, J. & Odegaard, B. Metacognitive sensitivity: The key to calibrating trust and optimal decision making with AI. *PNAS Nexus* **4**, pgaf133 (2025). https://doi.org/10.1093/pnasnexus/pgaf133

33 Lin, Z. How to write effective prompts for large language models. *Nat. Hum. Behav.* **8**, 611-615 (2024). https://doi.org/10.1038/s41562-024-01847-2

34 Tankelevitch, L. *et al.* in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* Article 680 (Association for Computing Machinery, Honolulu, HI, USA, 2024).

35 Phan, L. *et al.* Humanity's last exam. *arXiv* (2025). https://doi.org/10.48550/arXiv.2501.14249

36 Lin, Z. We need to rein in the AI gatekeepers of science now. *Nature* **645**, 645 (2025).

37 Azaria, A., Azoulay, R. & Reches, S. ChatGPT is a remarkable tool—for experts. *Data Intelligence* **6**, 240-296 (2024). https://doi.org/10.1162/dint_a_00235

38 Charness, G., Jabarian, B. & List, J. A. The next generation of experimental research with LLMs. *Nat. Hum. Behav.* **9**, 833-835 (2025). https://doi.org/10.1038/s41562-025-02137-1

39 Sohail, A. & Zhang, L. Using large language models to facilitate academic work in the psychological sciences. *Current Psychology* **44**, 7910-7918 (2025). https://doi.org/10.1007/s12144-025-07438-2

40 Mammides, C. & Papadopoulos, H. The role of large language models in interdisciplinary research: Opportunities, challenges and ways forward. *Methods*

in Ecology and Evolution **15**, 1774-1776 (2024). https://doi.org/10.1111/2041-210X.14398

41 De Freitas, J., Nave, G. & Puntoni, S. Ideation with generative AI—in consumer research and beyond. *Journal of Consumer Research* **52**, 18-31 (2025). https://doi.org/10.1093/jcr/ucaf012

42 He, Y.-H. AI-driven research in pure mathematics and theoretical physics. *Nature Reviews Physics* **6**, 546-553 (2024). https://doi.org/10.1038/s42254-024-00740-1

43 Zhang, Y. *et al.* Exploring the role of large language models in the scientific method: from hypothesis to discovery. *npj Artificial Intelligence* **1**, 14 (2025). https://doi.org/10.1038/s44387-025-00019-5

44 Schmitt, M. AI-driven exploration and hypothesis generation in finance and economics. *SSRN* (2025). https://doi.org/10.2139/ssrn.5126623

45 Chen, L., Pelger, M. & Zhu, J. Deep learning in asset pricing. *Management Science* **70**, 714-750 (2024). https://doi.org/10.1287/mnsc.2023.4695

46 Tomaino, G., Cooke, A. D. J. & Hoover, J. AI and the advent of the cyborg behavioral scientist. *Journal of Consumer Psychology* **35**, 297-315 (2025). https://doi.org/10.1002/jcpy.1452

47 Peng, T.-Q. & Yang, X. Recalibrating the compass: Integrating large language models into classical research methods. *arXiv* (2025). https://doi.org/10.48550/arXiv.2505.19402

48 Blanchard, S. J., Duani, N., Garvey, A. M., Netzer, O. & Oh, T. T. New tools, new rules: A practical guide to effective and responsible GenAI use for surveys and experiments research. *Journal of Marketing*, 00222429251349882 (2025). https://doi.org/10.1177/00222429251349882

49 Lin, Z. Six fallacies in substituting large language models for human participants. *Adv. Meth. Pract. Psychol. Sci.* **8**, 25152459251357566 (2025). https://doi.org/10.1177/25152459251357566

50 Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I. & Pechenkina, E. Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education* **21**, 100790 (2023). https://doi.org/10.1016/j.ijme.2023.100790

51 Melumad, S. & Yun, J. H. Experimental evidence of the effects of large language models versus web search on depth of learning. *SSRN* (2025). https://doi.org/10.2139/ssrn.5104064

52 Luo, X. *et al.* Knowledge and awareness of generative artificial intelligence use in medicine among international stakeholders: A cross-sectional study. *Journal of Evidence-Based Medicine* **18**, e70034 (2025). https://doi.org/10.1111/jebm.70034

53 Lin, Z. Beyond principlism: Practical strategies for ethical AI use in research practices. *AI Ethics* **5**, 2719-2731 (2025). https://doi.org/10.1007/s43681-024-00585-5

54 Tang, C., Li, S., Hu, S., Zeng, F. & Du, Q. Gender disparities in the impact of generative artificial intelligence: Evidence from academia. *PNAS Nexus* **4** (2025). https://doi.org/10.1093/pnasnexus/pgae591

55 Gao, J. & Wang, D. Quantifying the use and potential benefits of artificial intelligence in scientific research. *Nat. Hum. Behav.* **8**, 2281-2292 (2024). https://doi.org/10.1038/s41562-024-02020-5
56 Lave, J. & Wenger, E. *Situated learning: Legitimate peripheral participation*. (Cambridge University Press, 1991).
57 Collins, A., Brown, J. S. & Holum, A. Cognitive apprenticeship: Making thinking visible. *American Educator* **15**, 6-11 (1991).
58 Lodge, J. M., Ellerton, P., Zaphir, L. & Brown, D. in *Artificial intelligence applications in K-12: Theories, ethics, and case studies for schools* 24-37 (Routledge, 2024).
59 Walter, Y. Embracing the future of Artificial Intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education* **21**, 15 (2024). https://doi.org/10.1186/s41239-024-00448-3
60 Lee, H.-P. *et al.* in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* Article 1121 (Association for Computing Machinery, 2025).

About the authors

Zhicheng Lin is a Lee Youn Jae Rising Fellow associate professor at Yonsei University, Seoul, South Korea. He is also an associate editor at *Psychological Science*, *Perspectives on Psychological Science*, and *Behavior Research Methods*. He received his B.S. in psychology from Peking University and his Ph.D. in cognitive and brain sciences from the University of Minnesota. He previously worked at the University of Washington, The Ohio State University, The Chinese University of Hong Kong, Shenzhen, and the University of Science and Technology of China. His research spans cognitive science, consciousness, and metascience, with a recent focus on AI for science and AI psychology.

Aamir Sohail is an MRC Advanced Interdisciplinary Methods PhD student at the Centre for Human Brain Health, University of Birmingham. He completed a BSc in Biomedical Science at Imperial College London, followed by an MSc in Brain Imaging at the University of Nottingham. He then worked as a Junior Research Fellow at the Centre for Integrative Neuroscience and Neurodynamics, University of Reading. His research interests involve using computational modeling and neuroimaging to understand social decision-making, and more broadly how artificial intelligence can be used to accelerate scientific research.