What Does Research Need a Researcher For?

James Lake*
University of Tennessee

John Lopresti†
William & Mary

June 5, 2026

Abstract

Large language models and agentic AI systems can now perform many tasks central to economic research, including data construction, coding, model solving, literature synthesis, and academic writing. We argue that the profession should respond with an output-based ethical standard: researchers can claim authorship of AI-assisted work only if they understand it deeply enough to defend every substantive claim, methodological choice, and analytical result. We identify verification and dissemination as the enduring human functions in economic research. Verification becomes more important as AI increases research output and weakens the link between producing work and understanding it. Dissemination remains essential because research gains legitimacy and influence through human explanation, judgment, and accountability. Consequently, doctoral training should preserve foundational technical instruction while shifting the research apprenticeship toward verification.

*E-mail: jlake7@utk.edu. We thank Peter McHenry for useful comments.
†E-mail: jwlopresti@wm.edu

1 Introduction

The economics profession faces a technological reality that demands honest reckoning rather than reflexive caution. Large language models and agentic AI systems can now perform, with increasing reliability, many of the core technical tasks that constitute the production of economic research: constructing datasets, writing and debugging code, deriving and solving models, synthesizing literatures, and producing polished academic prose (Social Catalyst Lab, nd). These are not speculative capabilities. They are present ones, improving rapidly, and available to any researcher with an internet connection.

The profession's response has so far been fragmented and largely process-oriented – disclosure requirements, restrictions on AI-generated text, anxious debate about what "counts" as the researcher's own work (e.g. American Economic Association, nd; Canadian Journal of Economics, nd; Journal of Economic Theory, nd). We think a standard based on the research process is both unworkable and largely beside the point. It tries to preserve a mode of production already being overtaken by events, at the cost of emphasizing what actually matters: the quality and reliability of the knowledge produced, and the human institutions through which that knowledge becomes trusted and used.

The purpose of economic research is to produce knowledge that advances scientific understanding and informs policy. Whether that knowledge is generated through months of manual labor or through a well-structured AI workflow is relevant only insofar as it affects the quality and reliability of the research output. The intellectual satisfaction that researchers derive from the mode of production – the craft of writing elegant code, the pride of hand-deriving a closed-form solution – is real, but it is not the point. If AI-augmented research can produce knowledge faster, at greater scale, and with fewer errors, then the profession has an obligation to embrace it. The opportunity cost of methodological conservatism is not merely academic. It is measured in policy questions left unanswered, in interventions left unevaluated, and in human welfare left unimproved.

In this article, we provide a perspective on the role of the human in economic research in the AI age and propose an ethical standard for research given that role. We also discuss implications both for the research process and doctoral training under this paradigm. The question is not whether to let AI into the research process. It is already there. The question is what the researcher contributes that AI does not, and whether the profession is organized to invest in the institutions that maximize these contributions.[1]

[1]The direct literature on AI in economic research is nascent and heterogeneous. Our article is therefore not a review of an established field. Instead, we draw selectively on current journal and professional guidance on AI use in research (American Economic Association, nd; Canadian Journal of Economics, nd; Journal of Economic Theory, nd), early evidence and demonstrations of AI-assisted or autonomous research workflows in economics (e.g. Social Catalyst Lab, nd, Social Catalyst Lab GitHub Archive, nd), practitioner commentary

2 The Human Role in Research

If AI can build the dataset, write the code, estimate the model, and draft the paper, what is left for the human to do? This is not a rhetorical question. It is a question the profession must answer before it can say anything useful about norms, ethics, or training. And answering it honestly means being willing to acknowledge that some of the things we have long valued about our work are not the things that actually make it valuable. We consider three possible answers.

Taste. Perhaps the most common response is that humans contribute "taste": the ability to identify an important question, to see a promising source of variation, to know which problems are worth solving. This is where many discussions of AI and research land, and it is a comforting thought. It preserves the researcher as the creative center of the enterprise, with AI as a powerful but subordinate tool.

We are not sure this is correct. AI systems are already strong at generating well-posed research questions and at answering any given one (Social Catalyst Lab, nd). Whether humans are meaningfully better at choosing which questions to pursue is genuinely uncertain. In one large experiment, autonomously generated empirical papers judged head-to-head against published work sat well below the human average but, in the right tail, drew ratings comparable to papers in leading journals (Simon, 2026; Social Catalyst Lab, nd). If the profession's answer to "why do we need human researchers?" is that humans ask better questions, then the role for humans disappears the moment someone demonstrates that machines ask equally good ones. That moment may not be far off. We think there are more durable answers.

Verification. A second possible avenue for human contribution is verification – the ability to take AI output and evaluate whether the model captures the appropriate economic trade-offs, whether the identification strategy is sound, whether the result is credible, whether magnitudes are sensible. At first blush this sounds like a diminished role: the researcher as checker, not creator.

But this undersells what verification is. Verification is not reading output and nodding along. It requires the ability to independently reconstruct the logic of an analysis, to identify where it could go wrong, and to judge whether it has. It draws on deep knowledge of the institutional setting, the econometric methods involved, and the economic structure of the model. It is a skill that takes years to develop and inherently depends on having grappled with the work on one's own.[2] A researcher who has never built a dataset from administrative

(e.g. Simon, 2026), and the broader literature on reproducibility (Brodeur et al., 2026).

[2]This point relates to a broader literature on tacit knowledge and on the difficulty of reconstructing

records will struggle to assess whether someone else's – or some machine's – data construction is sound.

Verification is also not new.[3] Researchers have always spent a large share of their time scrutinizing work they did not produce: reading papers, replicating key results, attending seminars, refereeing. What is new is that AI changes the volume and nature of what needs to be verified, and it does so in ways that make the task harder, not easier. We develop this argument more fully in Section 4.

Dissemination. A third answer has received surprisingly little attention in the current debate: the human role in carrying research into the world. Research becomes socially useful – influencing policy and changing how people understand a problem – through a chain of human communication. Economists present at conferences, brief policymakers, advise governments, write for public audiences, and teach. This chain is not incidental to the research enterprise. For most of the work the profession produces, it is the mechanism through which research has any effect at all.

Two things make this chain irreducibly human. The first is legitimacy. A policy justified by "economists studied this, explained it to policymakers, and stood behind it" is not the same as one justified by "AI concluded it." This is not a sentimental distinction. Public acceptance of consequential policy runs on a chain of human research, transmission, and decision.[4] Stripping out the human removes the social warrant that lets evidence change behavior. This is particularly relevant given modern political realities: amid an already-marked erosion of trust in expertise, prescriptions stamped "the agent says so" rather than carried by named experts who stake their reputations risks inviting a backlash at least as severe as the one expertise already attracts, and quite possibly worse.[5]

The second is judgment about what matters. One cannot hand the sum of research on a topic to a policymaker or their aide and ask them to simply choose the optimal policy.

complex work from formal outputs alone; see Lu (2025) and Li et al. (2026). Our claim is that, in economic research, direct engagement in construction helps build the understanding later required for verification.

[3]Related concerns appear in the growing literature on reproducibility and robustness in economics and political science; see Brodeur et al. (2026) and the Institute for Replication (nd). Our argument differs in focusing on how AI changes the relationship between producing research and being able to verify and defend it.

[4]Related concerns appear in the literature on the legitimacy of science; see Longo et al. (2025). Our argument is that, in policy-relevant economic research, legitimacy depends in part on a human chain of explanation, judgment, and accountability that AI output alone cannot supply.

[5]Survey evidence consistently finds that the public trusts human decision-makers more than AI systems for consequential decisions. Only 17% of Americans believe AI will have a positive impact on the country (McClain et al., 2025); globally, fewer than half of respondents are willing to trust AI systems (Gillespie et al., 2025); and, experimental evidence shows that decisions supported by human judgment are perceived as more trustworthy than those made by AI systems (Orbán and Stefkovics, 2025)

Someone must decide which findings are relevant, translate them into actionable terms, and judge what to emphasize, and those decisions require understanding the policy context, the audience, and the stakes in ways that go beyond summarization. At every point where research meets a decision-maker, a human must make these judgments.

Where this leaves us. All three answers contain something real. But we believe verification and dissemination matter most because they remain indispensable even if AI can match humans at ideation. Verification is perhaps the most pressing issue. A researcher who cannot, or will not, verify what AI produces cannot meet any worthwhile ethical standard regarding the use of AI in economic research. And the capacity to verify is itself threatened by the very technology that makes it more necessary – a point we develop in Section 4 after establishing the ethical standard it supports in Section 3.

3 The Output-Based Standard

As the role for human researchers evolves, so too must the norms surrounding research. What are appropriate ethical standards for a research enterprise in which AI plays a pivotal role? We believe that a framework that polices the process of research – which sentences AI drafted, which stages it touched – is unenforceable, as process is largely unobservable. This has always been the case: we could never observe a researcher arrive at an identification strategy, choose a specification, or draft a paragraph. The scope of tasks to which AI can meaningfully contribute only heightens the difficulty of monitoring production of those tasks. Moreover, process restrictions are counterproductive: they push AI use underground and forfeit the chance to build norms around its responsible use.

If process cannot be the standard, output must be.[6] We propose a single principle: a researcher may claim authorship of AI-assisted work if and only if they understand the work deeply enough to defend every substantive claim, methodological choice, and analytical result to a skeptical audience – a seminar room, a referee, a policymaker – from their own knowledge and judgment.

This standard has a strong claim to being the right one, for reasons that mostly have nothing to do with AI. It is already the implicit standard for collaborative research. Senior economists routinely publish work in which research assistants (RAs) wrote the code, cleaned

[6] This emphasis on output is related to, but distinct from, the growing focus on reproducibility and robustness in economics and political science; see Brodeur et al. (2026) and the Institute for Replication (nd). That literature focuses on whether findings are credible and reproducible. Our claim here is that, in the context of AI-assisted research, the relevant ethical standard should turn on whether the author can understand, verify, and defend the work.

the data, and produced the tables. The ethical legitimacy of the senior author's claim rests not on having performed these tasks personally but on having directed the intellectual agenda and possessing sufficient understanding to stand behind the results. Substituting AI for the RA changes nothing, provided the understanding is maintained.

The standard is also testable. While we cannot observe the process of research production, we can observe the researcher's ability to defend the work. Seminars, conferences, internal department brown bags, and policy presentations are existing institutions that test exactly this capacity. A researcher who does not understand their own results will be exposed in these settings, as they always have been. Our output standard targets the real risk, which is not that AI wrote the code but that an author cannot scrutinize or defend what they publish. It also scales with AI's contribution: debugging a script adds a negligible verification burden; commissioning the identification strategy, the model, and the prose adds an enormous one.

The standard's one demand – understanding – is also its vulnerability. A fluent AI explanation can manufacture the feeling of comprehension; the distance between "this looks right to me" and "I could reconstruct this, find its errors, and understand the tradeoffs inherent in this approach" is where the ethical risk lives. Meeting the standard requires the capacity to verify what AI produced. What that capacity is, how it is built, and why AI puts it at risk is the subject to which we now turn.

4 Verification and the Production of Research

Verification is not merely quality control. It is not simply a chore the human performs after the machine finishes. Rather, it is one of the two ways researchers build the understanding that makes them researchers. While AI does not alter the necessity for verification, it does change the process by which it takes place.

The research production function: construction and verification. Research understanding is generated by two processes. The first is construction: deriving the model, building the dataset, writing the code, making the many small decisions a project requires. This is the process researchers have traditionally spent most of their time on, and the one most visibly affected by AI. It builds understanding as a byproduct – one learns where the data are messy, which assumptions bind, and how sensitive results are to choices one might have made differently.

The second process is verification: scrutinizing work the researcher did not produce themself – reading an appendix to understand a model or estimator well enough to extend it, reverse-engineering another paper's identification strategy to sharpen one's own, or interrogating a seminar speaker about a suspicious assumption. This is verification in the service of one's own research, and it often teaches methods and insights a researcher would never have arrived at on their own. The cumulative structure of science depends on it.

Crucially, the two processes are complements, not substitutes. Construction teaches a researcher where things can go wrong — where the data are fragile, which coding decisions are consequential, how a seemingly innocuous assumption can drive results. This is exactly the knowledge that facilitates productive verification. A researcher who has wrestled with building a dataset from administrative records can spot problems in someone else's data construction that a researcher who has never done it will miss entirely. One learns from verifying only what one has the foundation to interrogate

AI changes the relative cost of these two complements. Until now, construction produced understanding but was very expensive, requiring hours debugging code, tracking down data errors, and finding algebraic mistakes. Now that AI can do many of these construction tasks faster and with fewer errors (Simon, 2026; Social Catalyst Lab, nd), the case for a researcher doing it themselves weakens. Every hour spent debugging code is an hour not spent on verification — and the hour spent debugging is now the less valuable one. The efficient allocation of the researcher's time shifts accordingly: less construction, more verification.

However, the complementarity between construction and verification constrains how far the shift can go. Delegating construction before one has sufficient verification capital will leave a researcher with no foundation upon which to build through learning by verification.

What changes for the individual researcher. When the object of verification shifts from other researchers' work to AI-generated output – whether one's own or others' – both the volume and the difficulty of verification increase. When a researcher verifies their own AI-constructed output, volume rises because AI can produce more, faster (Social Catalyst Lab, nd). Difficulty rises because the researcher did not construct the output themselves; the understanding that construction produces as a byproduct does not accumulate, so one is verifying without the tacit knowledge of the choices that have been made along the way. Similarly, when verifying the AI-assisted work of others, the volume rises because everyone else is also producing more. More subtly, the difficulty of verifying the work of others might also rise, as one can no longer assume the author deeply understands what they have produced. In the past, the fact that an author had conducted analysis independently meant they had, at minimum, wrestled with the data and the model – existence of the research was itself a form of pre-verification. That assumption is no longer safe. A reader of the paper or a seminar participant may be the first person to seriously scrutinize the work.

What changes for the profession. A crucial result of the process described above is a wedge between work that is certified – that is, it has passed some form of peer review – and work that is defensible by the researcher – they can explain it, interrogate it, and stand behind it in a seminar or a policy briefing. These have traditionally been linked; getting a paper through peer review required enough technical competence that a published author could generally stand behind the work. AI breaks that link. Referees have always inferred understanding from observable signals such as the quality of writing or the absence of technical errors. When AI can produce these, the signals no longer separate authors who understand their work from those who do not. That is, the technical quality of AI-assisted work can be high enough to clear the peer review process without the author having built the understanding that the process was implicitly testing for. Publication may become easier, but understanding does not.

The output-based principle proposed above helps because it asks whether the author understands the work, not merely whether the work survives peer review. A corollary is that the profession must place greater weight on settings that reveal understanding directly, such as seminars, conference presentations, department brown bags, and policy briefings. There should also be increased emphasis on evaluation criteria that reflect engagement with the work, such as external letters from scholars who have discussed the research with the researcher. Publication counts, which will become easier to inflate as AI lowers the cost of production, should matter less.

Ultimately, the capacity to verify must be deliberately built, and that burden falls heavily on how we train the next generation of researchers. We discuss this now.

5 Implications for Doctoral Training

The doctoral years are where the capacities to verify and disseminate – the two functions we identify as irreducibly human – are built. How they are built must change.

Importantly, the core foundation of doctoral training remains. First-year sequences in which students work through models by hand, derive the properties of estimators and learn to construct datasets from scratch provide the foundation upon which everything else rests.[7] Removing them would commit the error of delegating what we have called the construction process before the foundational capacity for verification exists.

What changes is how students spend their research apprenticeship. The traditional RA

[7]The point is not that every future researcher must continue to perform each of these tasks manually, but that early direct engagement in construction helps build the tacit and practical understanding later required for verification; see Lu (2025) and Li et al. (2026).

model of writing code to a supervisor's specification, cleaning data, or constructing figures for publication bundled RA labor with pedagogy related to construction. As AI now provides a cheap substitute for RA labor, the apprenticeship must be rebuilt around verification.[8] Students should spend time reproducing key results independently, without existing code, forcing every analytical decision themselves (Brodeur et al., 2026; Institute for Replication, nd). They should critique AI-generated research designs, stress-test identification strategies, and assess whether estimated magnitudes are plausible given what they know about the institutional setting. These are the activities that build verification capital while requiring the same technical foundation as before, applied differently.

Presentation skills assume greater importance – both as a way to demonstrate understanding and as preparation for the dissemination role that grows more important as AI transforms research production.[9] A second-year paper becomes less about the paper itself and more about the presentation: can the student defend the work, engage with skeptical questions, and explain the contribution to a non-specialist? In-class presentations, research workshops, and department brown bags are not ancillary to the student's training but crucial to developing the dissemination function the profession will increasingly need. If carrying research into the world along a human-to-human chain is a defining task for the researcher, then teaching the next generation of researchers to do it well is a core obligation of the profession.

6 Objections and Responses

We expect several objections to the framework proposed above, and we willingly acknowledge the uncertainty surrounding the path the field will take as AI continues to evolve. While we do not attempt to anticipate every possible critique, we respond to several of the more obvious objections below.

AI output is unreliable and prone to fabrication. This is true and is the reason our output standard leans on verification rather than trust. Human research assistants also err, often seriously; the profession has long managed that through supervision and replication,

[8] This proposal is informed by the broader emphasis on reproducibility and robustness in social-scientific research (Brodeur et al., 2026; Institute for Replication, nd) and by the view that practical understanding is often acquired through reconstructing and stress-testing work rather than merely consuming polished outputs (Li et al., 2026). Our claim is that doctoral training should respond to AI by reallocating scarce human learning time toward those verification tasks.

[9] This follows from the broader point that the social authority of research depends not only on the existence of evidence but also on the institutions and human practices through which it is explained, defended, and made legitimate (Longo et al., 2025).

not by refusing to delegate. The appropriate response to unreliability is not to avoid the tool but to build robust verification structures around it.

Researchers overestimate their own understanding of AI-generated output. This is the risk we take most seriously. A truncated construction process can inflate the feeling of comprehension while eroding understanding – polished AI output is easier to nod along with than raw data and broken code.[10] This risk is likely to grow over time as researchers become accustomed to reviewing output they did not produce. Independent replication and adversarial stress-testing are external checks designed precisely because self-assessment is unreliable. No standard substitutes for intellectual honesty; ours at least names the danger and builds checks around it.

The output-based standard is no more enforceable than the process rules we reject. While one cannot compel conference attendance or force a researcher to present at a brown bag, the comparison that matters is not "enforceable versus unenforceable". Rather, it is which standard creates the right incentives. Process rules ask researchers to disclose something unverifiable and create incentives to hide AI use. The output standard asks researchers to understand what they publish and creates incentives to verify this understanding. The profession can reinforce those incentives by re-weighting the evaluation criteria it already uses: weighting seminars, conference presentations, and policy engagements more heavily; placing more value on external letters from scholars who have engaged with the researcher and their work; and, treating publication records as less connected to research understanding than they are today.

AI advantages established researchers over junior researchers. Senior researchers do carry credibility that lets their claims be taken on trust in ways that junior scholars' claims are not. Yet the framework is at least as kind to juniors as the status quo: a junior researcher who can use AI to produce ambitious, well-verified work faster than the traditional production process is better positioned than one who must spend years on manual execution. The binding constraint for junior scholars was always ideas and judgment, not labor.

Unequal access to capable AI will widen institutional inequality. We take this concern seriously. AI use is heavily subsidized for now by the firms providing it; as that recedes and costs enter research budgets, access could diverge across institutions (Simon,

[10] This concern parallels recent work showing that formal research outputs often omit practical knowledge needed for genuine reconstruction and evaluation; see Li et al. (2026). Our claim is that fluent AI-generated explanation may amplify that gap by making superficial comprehension feel like real understanding.

2026). The magnitude will decide how much this constraint binds: at a few thousand dollars per researcher per year, the tools are broadly affordable; at an order of magnitude more they are not, and the gains would accrue to well-resourced departments, potentially exacerbating inequalities that already exist across institutions.

Nothing stops someone from gaming the standard. While nothing prevents a researcher from having AI produce a paper, verifying it superficially, and submitting it, the strategy is more self-limiting than it sounds. A researcher who does not genuinely understand the work will be exposed the moment they have to discuss it in a presentation or in conversation with a colleague who has read the paper carefully. Further, the returns to this approach decline if the profession shifts away from publication records and towards evaluation toward criteria that assess understanding.

Hands-on research has intrinsic value. This is undeniably true: most researchers enter their field because they find the details of their work genuinely enjoyable. There is no contradiction in choosing to do by hand what AI could do, as one might sew a garment by hand; we expect such practice to persist in pockets and see no reason to disparage it. But individual preference is a different question from what the profession should adopt, and satisfaction with the current mode of production is not reason enough to forgo tools that could speed welfare-relevant knowledge. The stakes are too high for sentimentality.

7 Conclusion

We have argued there is an inherent role for humans in economic research organized around verification and dissemination, and an accompanying output-based ethical standard grounded in the researcher's ability to understand and defend the work. Consistent with this, we have argued for doctoral training that preserves its foundational core while redesigning the apprenticeship around verification rather than the labor-intensive tasks that AI can now do with high quality at low cost. Finally, we have argued for increased emphasis on dissemination, both in terms of evaluation of scholars and in doctoral training. The thread through each of these is the idea that the nature of research production has changed, and reflexive adherence to old modes of research is neither tenable nor desirable. However, while the role of humans in the research process has changed, it has not diminished; it has simply shifted to where it now matters most. The field is best served not by resisting change as long as possible, but by embracing the roles that allow the frontier of economic knowledge to expand most rapidly. This, after all, has always been the goal, regardless of the tools used to pursue it.

References

American Economic Association (n.d.). Editorial policy. https://www.aeaweb.org/journals/aer/editorial-policy. Accessed May 27, 2026.

Brodeur, A., N. Cook, A. Heyes, N. V. Kuminoff, L. Lakdawala, E. Reuben, J. Vollen, and D. Yanagizawa-Drott (2026). Reproducibility and robustness of economics and political science research. *Nature 640*, 623–629.

Canadian Journal of Economics (n.d.). Author guidelines. https://onlinelibrary.wiley.com/page/journal/15405982/homepage/forauthors.html. Accessed May 27, 2026.

Gillespie, N., S. Lockey, C. Curtis, J. Pool, and A. Akbari (2025). Trust, attitudes and use of artificial intelligence: A global study 2025.

Institute for Replication (n.d.). Institute for replication. https://i4replication.org/. Accessed May 27, 2026.

Journal of Economic Theory (n.d.). Guide for authors. https://www.sciencedirect.com/journal/journal-of-economic-theory/publish/guide-for-authors#ethics-and-policies-declaration-of-generative-ai-use. Accessed May 27, 2026.

Li, Y., L. M. Hildebrandt, P. Leippi, M. Hollick, and C. Kästner (2026). What papers don't tell you: Recovering tacit knowledge for automated paper reproduction. *arXiv preprint arXiv:2506.11433*.

Longo, G., M. Viola, and G. Volpe (2025). On facts concerning scientific research: Responding to the crisis of legitimacy of science. *Foundations of Science*.

Lu, J. (2025). Tacit knowledge in large language models. *Philosophy and Technology 38*(4).

McClain, C., E. A. Vogels, S. Sechopoulos, and L. Rainie (2025, April). How the u.s. public and ai experts view artificial intelligence.

Orbán, M. and A. Stefkovics (2025). Trust in artificial intelligence: A survey experiment to assess trust in algorithmic decision-making. *AI & Society*.

Simon, K. (2026, May). Practical advice for using ai in economics research: Summary of nber panel from may 8th 2026. Frankly, the Counterfactual, Substack.

Social Catalyst Lab (n.d.). Project ape: Autonomous policy evaluation. https://ape.socialcatalystlab.org/. Accessed May 27, 2026.

Social Catalyst Lab GitHub Archive (n.d.). Project ape. GitHub repository. Accessed May 27, 2026.