This version: 21.3.2026

How Does Artificial Intelligence Transform Research Processes?

Ari Hyytinen

Based on a presentation delivered at the meeting of the Finnish Society
of Sciences and Letters, 20 October 2025[1]

Artificial intelligence (AI) is reshaping how research is conducted. This change refers not only to new tools that speed up existing workflows, but also to AI-based technologies that change what researchers can do and how they do it. The recent arrival of increasingly more capable large language models (LLMs) has made this transformation visible and immediate. Yet the deeper significance of AI for research goes beyond research practices and convenience. AI may be altering the very process by which knowledge is generated, i.e. the formulation of hypotheses and development of theoretical and conceptual frameworks, collection and analysis of data and evidence, as well as interpretation of results.

In this writing, I survey selectively recent views and arguments on how AI is changing research practice. I pay particular attention to how economists have interpreted the ongoing developments and how AI is being applied in economic research (see also Hyytinen 2025). I note explicitly that the pace of change is very fast, with new AI tools being developed and released almost daily. Therefore, parts of what is being described in this writing are likely to be outdated soon, if they are not already now. For example, recently a lot of progress has been made in how AI agents can be used in research (i.e., how LLM-based systems that plan and have longer memory and that can be connected to external tools can semi-autonomously pursue complex research goals; see e.g. Korinek 2025).

Besides making general observations about LLMs, pathways of how AI may reshape the research process and how AI might work as an idea generator, I also briefly describe two concrete empirical examples to illustrate what is going on. First, I describe how AI may enable a fundamentally better market design by solving a longstanding data-collection problem. Second, I explore how AI is used to automate the entire cycle of hypothesis generation and empirical testing. These examples represent different points on a spectrum where the extreme ends are AI as a research tool and AI as a research agent. Together they help clarify both the promise and the limits of the transformation underway.

What are large language models?

To understand AI's role in research, it is worth describing briefly what today's leading AI systems actually are. LLMs are deep neural networks trained on vast amounts of text (and other data) to statistically process and generate human language. They are sometimes described as "stochastic parrots" (e.g., Bender et al. 2021), i.e., as systems that recognise and reproduce statistical patterns in language without possessing true comprehension or intentionality. Technically, LLMs are autoregressive models that optimise the probability of generating the next token given prior context, learning from the statistical regularities of their training data rather than from any explicit model of the world.

---

[1] Relative to what I presented on 20 October 2025, I have added some new references to connect the ideas that I had in my presentation to the more recent discussions and developments in the literature.

This framing carries an important implication for research use. A statistically probable response and a truthful (or "correct" in some sense) response are not the same thing. LLM outputs therefore require verification in any scientific context, and this basic limitation should be kept in view even when assessing impressive-seeming results.

Yet this framing also poses an apparent paradox. How can a system that merely pattern-matches produce scientifically useful outputs? Can they generate plausible hypotheses, assist with causal reasoning, or write coherent literature reviews? The honest answer is that pattern-matching at sufficient scale does produce useful outputs, within limits that are not yet fully mapped.

A complementary framing, also useful for thinking about research applications, is that of AI as a *prediction machine* (Agrawal et al. 2026). Rather than asking whether LLMs "understand", this framing asks whether they can usefully guide search. For example, can they prioritise which part of a vast combinatorial space is worth investigating next? Empirically, the answer appears to be yes in many data-rich contexts. This may be enough for significant scientific value, even without deeper comprehension.

One reason for cautious optimism is LLMs' *emergent behaviour*. This refers to the observation that as the models are scaled up in size, training data, and compute, qualitatively new capabilities have appeared that were not explicitly programmed and were not present in smaller models (Wei et al. 2022). The ability to perform multi-step mathematical reasoning, for instance, emerged with scale rather than being specifically trained. This claim has not gone unchallenged, however. Schaeffer et al. (2023) argue that apparent discontinuities in LLM capability may be largely artefacts of non-linear evaluation metrics rather than genuine qualitative shifts. Whether emergent abilities are real or a measurement phenomenon remains an active area of debate, and both the enthusiasm and the scepticism are worth keeping in mind when assessing AI's research potential.

A second relevant feature is *generativity*. LLMs are designed to produce new content. In brainstorming or hypothesis-generation contexts their tendency to generate a high volume of varied, sometimes erratic, output may be a desirable feature rather than a bug.

How AI reshapes the research process

AI may change research along five interrelated pathways. Korinek (2023) provides a systematic treatment of these in the specific context of economic research (see also Korinek 2025).

- **#1 Idea and hypothesis generation:** AI systems can rapidly generate many kinds of candidate hypotheses, research questions, or conceptual framings. As I discuss below, this enables a broader search of the possibility space than any individual researcher could achieve alone.

- **#2 Empirical analysis and data processing:** AI agents can automate data collection, coding, and processing. This is valuable since these tasks have traditionally consumed a large share of a researcher's time. Experimental evidence shows, moreover, that access to generative AI tools substantially increases productivity on professional writing and document-intensive tasks (e.g., Noy and Zhang 2023). AI-assisted coding tools similarly reduce the time needed to write, debug, and test analytical scripts, with direct implications for empirical research workflows.

- **#3 Model and theory generation:** LLMs can assist in formulating quantitative models and, in some cases, deriving analytical results. This might lower the barrier for researchers to engage with formal theoretical work or to explore model variants rapidly. For example, Manning et al. (2024) describe automated construction of structural causal models from natural-language scenario descriptions.

- **#4 Model refinement and validation:** AI can assist in calibrating and stress-testing models against data, iterating on specifications, and identifying where a model's predictions diverge from observations. This is an emerging pathway that may be less visible in current practices but is likely to grow in importance as AI coding and simulation capabilities improve.

- **#5 Innovation acceleration:** At the level of applied and product-oriented research, AI lowers the cost of experimentation and speeds up the cycle from concept to prototype. This pathway connects AI's impact on research to the broader economics of innovation and productivity growth.

These pathways do not operate in isolation. Their interaction opens the possibility of a more integrated, partly automated research pipeline, i.e., one in which AI contributes meaningfully to every step from initial question formulation to final inference, interpretation, and prototype development.

A further observation is worth adding here. Agrawal et al. (2026) model research productivity as the multiplicative product of productivity at each stage of the research workflow. This means that AI-driven improvements at, say, the idea-generation or data-processing stages may be substantially offset by bottlenecks elsewhere, i.e., at stages which AI cannot yet accelerate. Modest improvements to an already constrained stage of the pipeline can have a larger effect on overall productivity than large improvements at a non-binding stage. The overall gains from AI in research will therefore depend not only on how good AI becomes at the tasks it already handles well, but on whether the stages it cannot yet assist with can be sped up by other means.

AI as an idea generator

Of these five pathways, idea generation merits a closer look because it touches the production of new knowledge, which arguably is the most distinctively human dimension of research.

A useful framing comes from Zvi Griliches's (1957) analysis of hybrid corn. Griliches showed that the development of hybrid corn represented not just a single innovation but an *invention of a method for inventing*. This refers to a systematic process enabling continuous discovery of better crop varieties adapted to specific environments. LLMs may play an analogous role in idea generation and, perhaps more generally, in knowledge production. Seen from this perspective, LLM-enabled AI is not just a tool for executing individual tasks, but a platform that changes the *process* of generating ideas itself.

Why might AI be especially good at producing new ideas? The innovation process can be modelled as a search over a distribution of possible solutions, where each candidate idea has a random quality value (e.g., Weitzman 1979). The quality of the best solution found depends on the number of ideas generated, not only on their average quality. A system that generates ideas at very low cost and very high speed may have an advantage in this search, even if many of its individual outputs are mediocre. There is some empirical evidence on AI-assisted brainstorming that supports this logic. For example, Meincke et al. (2024) find that LLMs generate a higher volume of ideas than human groups, and that the best AI-generated ideas are competitive in quality with the best human ideas. The theoretical framework for interpreting such quality-quantity trade-offs in idea generation was developed by Girotra et al. (2010) in the context of human brainstorming, and extends naturally to AI-assisted settings.

A related argument concerns *combinatorial innovation*. The insight here is that new technologies and ideas rarely emerge *ex nihilo* but instead arise from recombining existing components in novel ways (e.g., Weitzman 1998, Romer 1993, Arthur 2009). LLMs, which blend concepts, structures, and results from their vast training data when generating responses, may be especially capable of this kind of recombination at scale. The ability to explore a large conceptual space and draw connections across disparate domains is a natural strength of a system trained on the breadth of human writing.

It is important, however, to distinguish between two types of combinatorial search that play different roles in science. Agrawal et al. (2026) draw a distinction between *idea generation* (i.e., the formulation of a new research question or explanatory hypothesis) and *design generation* (i.e., the search for a specific instantiation satisfying a known criterion). Idea generation involves a creative leap (abductive step) from a surprising observation to a tentative explanation, a kind of reasoning that human judgment typically performs and that AI currently struggles to replicate. Design generation, by contrast, is a different kind of problem where the objective function is clear and data exist to train a model. Agrawal et al. propose that AI-amenable problems share three features: i) a combinatorial search space, ii) a clear objective function, and iii) sufficient training data. Where these conditions hold, AI has already demonstrated transformational impact. Where they do not, the returns to adopting AI tools are far more uncertain.

Finally, three additional observations are in order. First, the search-process argument assumes that LLM-generated ideas are drawn from a distribution that includes genuinely useful ones. If LLMs can only reproduce and lightly remix existing knowledge, then the effective distribution may be narrower than the full space of valuable research ideas, limiting how much the fluency advantage translates into genuine discovery. This could happen if they cannot make foundational conceptual moves that open new research fields. Second, research by Dell'Acqua et al. (2023) on AI-assisted professional work finds that AI performance is uneven, meaning that it is strong within the current "capability frontier," but unreliable and sometimes counterproductive outside it. This "jagged frontier" is likely to characterise AI's performance in research as well. It is likely to be highly useful for certain tasks, poorly suited for others, with the boundary not always obvious in advance. Finally, proper measurement of how creative LLM-generated ideas are is still an active area of research; e.g., creative fluency and raw intelligence are not the same thing in the context of LLMs (Ruan et al. 2026).

Example I: Preference elicitation and better market design

A concrete illustration of AI's potential as a research-enabling tool comes from a recent paper in economics by Rusak, Manning, and Horton (2025). In this paper, they ask if AI tools can enable superior market designs.

Market design is the field that studies how to construct markets and allocation mechanisms to achieve efficient and fair outcomes. These efforts have long been constrained by a practical data problem. Many theoretically superior mechanisms require rich, granular preference data that participants find costly, time-consuming, and error-prone to provide. To name a few examples, students tend to under-rank course lists; job applicants often fail to articulate preferences over dozens of roles; and housing applicants struggle to specify location priorities. Small errors in preference reporting are not innocuous, because even minor deviations can cascade into large distortions of equilibrium outcomes, by analogy with the general principle in Akerlof and Yellen (1985). If so, these errors may erode much of the benefit that better mechanisms would otherwise deliver.

Rusak et al. propose an AI-based solution. They let participants describe their preferences in natural language, and use an LLM to translate those descriptions into cardinal von Neumann–Morgenstern utilities. This is the quantitative preference data needed to run (potentially more sophisticated) allocation mechanisms. In an experiment with 781 participants who wrote taste descriptions and ranked 50 tasks, the LLM-elicited utilities closely matched participants' own stated preferences. These utilities were rich enough to enable mechanisms, such as the Hylland–Zeckhauser pseudo-market, that require preference intensities rather than mere rankings. In a follow-up experiment, participants themselves preferred the matches produced under LLM-enabled mechanisms over those from simpler alternatives, particularly under high congestion contexts.

This example is instructive for a general reason, because it shows how AI can remove a data collection bottleneck that previously made certain research designs or policy interventions impractical. The barrier was not theoretical but, in a sense, cognitive, as it was driven by the cost of eliciting sufficiently rich information from human participants. LLMs lower that cost by mediating between natural human expression and the structured input that formal mechanisms require, thereby unlocking research and policy applications that were previously out of reach.

Example II: Automated social science

The second example pushes toward near-full automation of the research cycle. Manning, Zhu, and Horton (2024) present a system for automatically generating and testing social scientific hypotheses in silico, using LLMs combined with structural causal models (SCMs) expressed as directed acyclic graphs.

Their system operates as follows. Given only a natural-language description of a scenario, such as "two people bargaining over a mug", the system autonomously identifies relevant agents, proposes causal determinants of the outcome, constructs a structural causal model, designs a simulated experiment using LLM-based synthetic agents, runs the experiment, and estimates causal effects. In the bargaining scenario, the system selected the buyer's budget, the seller's minimum acceptable price, and the seller's emotional attachment to the mug as potential causal factors. These are all theoretically plausible choices that a human researcher might independently arrive at. The approach was applied to several further scenarios, including a bail hearing, a job interview, and an auction.

The results were promising but carefully qualified. The system predicted the signs of causal effects reliably, but was substantially less accurate about their magnitudes. A particularly notable finding is that when the fitted structural causal model was provided back to the LLM as additional context, its predictive accuracy improved considerably. The authors interpret this as evidence that the LLM "knows more than it can immediately tell". This suggests that formal causal structure helps the model express latent knowledge more accurately than free-form reasoning alone.

A methodological concern doing research "in silico" is that using LLM-based agents as synthetic experimental subjects raises non-trivial questions about external validity. If LLMs reproduce patterns from their training data rather than simulating genuine human behaviour, causal estimates may reflect the model's priors about how people behave rather than actual behavioural regularities. This concern does not invalidate the approach, but it does mean that in silico results should be treated as hypothesis-generating rather than hypothesis-confirming, pending validation against real human behaviour.

Concluding remarks: What are broader implications?

These two examples illustrate qualitatively different modes of AI involvement in research. In Rusak et al., AI is a tool that augments human-designed research and solves one specific bottleneck. In Manning et al., AI acts more like an agent, since it proposes the causal structure of the problem, generates the experimental design, runs the study, and produces the estimates. Here human involvement happens primarily at the level of scenario specification and result interpretation (see also Lu et al. 2024).

This distinction matters for how we assess the broader implications. Augmentation raises relatively tractable questions about when and how AI tools improve research quality. Automation and AI-agent enabled research raise harder ones, because it blurs the role of human judgement in scientific reasoning, allocation of responsibility for errors and in whether research conducted largely by AI can and should be held to the same standards of rigour and replicability as human-conducted research. Both modes are already present in practice; neither is speculative.

The examples I have covered are drawn from economics, but the underlying dynamics are general. AI is changing both *what* can be researched and *how*. The potential for productivity gains in research is real, but it comes with an important qualification. Dell'Acqua et al.'s (2023) finding of a "jagged frontier" suggests that AI is not uniformly useful across research tasks. The creative and conceptual dimensions of research may lie outside the current capability frontier for most AI systems. A plausible risk is that AI increases the production of research that is technically competent (subject to obvious caveats) but, in some sense, intellectually predictable. Whether the best AI-assisted ideas are truly novel or largely recombinatorial remains an open empirical question, and an important one.

Early large-scale empirical evidence on this risk is instructive. Hao et al. find that AI-using scientists are dramatically more productive by conventional metrics (publishing more and attracting more citations). Yet, the same study documents a contraction in the breadth of topics researchers pursue and a decline in cross-paper scientific engagement. More output, but from a narrower range of questions.

At the macroeconomic level, some economists have begun to ask whether AI-driven acceleration of idea production could trigger a discontinuous shift in the rate of economic growth. Jones (2024) writes: "*Once machines can produce ideas, the limits to growth set by the quantity and quality of researchers may no longer hold, and growth rates could speed up, potentially even leading to a so-called singularity*." There are several formal growth-theoretic models and analyses that explore the conditions under which AI automation of the innovation process could lead to explosive or unbounded growth (e.g., Aghion et al. 2019, Trammell and Korinek 2023).

The gap between pessimistic and optimistic assessments of AI's macroeconomic impacts can, in part, be attributed to whether one considers its effect on the *output production function* (like Acemoglu 2024) or the *knowledge production function* (like Jones 2024 and Jones 2026). The distinction matters. Even modest improvements to how science produces ideas could have compounding effects on long-run growth, dwarfing any (shorter-term) efficiency gains in production (Agrawal et al. 2026). Whether AI primarily automates tasks in existing production or accelerates the production of ideas is arguably the central open question for its longer-term impact and macroeconomic significance.

Whether or not the most dramatic scenarios materialise, the more immediate and tractable challenge is already pressing. Several practical questions stand out: How should research institutions, funding bodies, peer review, and academic norms adapt to a world in which AI is a routine co-contributor to the research process? How do we verify AI-generated outputs (hypotheses, code, literature summaries, empirical results), given the known risk of confident-sounding errors? How should authorship and intellectual contribution be attributed when AI has proposed the research design? And how do we preserve the conditions for foundational discovery if and when research institutions increasingly reward the kind of high-volume, rapid-iteration output that AI makes easy to produce? These are not rhetorical questions. They are problems that the research community need to address, with or without settled answers about how capable AI will ultimately become.

From the perspective of a small open economy that is *not* on the frontier of developing major LLM tools and foundational AI technologies, a central challenge concerns our universities' and research institutions' absorptive capacity. By this I mean their ability to identify the value of rapidly advancing AI tools, assimilate them into existing research workflows, and deploy them productively across scientific fields. For example, how can universities and research institutions secure researchers' wide access to frontier AI tools, and how can faculty be helped to develop the AI expertise needed to realise the associated productivity gains? The aggregate benefits of advances in AI technology are amplified by the share of the scientific workforce that can actually deploy them (Agrawal et al. 2026), making access to AI tools and AI research literacy mutually complementary and reinforcing. Neither is sufficient without the other.

References

Acemoglu, D. (2024). The simple macroeconomics of AI. NBER Working Paper No. 32487.

Aghion, P., Jones, B. F., & Jones, C. I. (2019). Artificial intelligence and economic growth. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The Economics of Artificial Intelligence: An Agenda* (pp. 237–282). University of Chicago Press.

Agrawal, A., McHale, J., & Oettl, A. (2026). AI in science. NBER Working Paper No. 34953.

Akerlof, G. A., & Yellen, J. L. (1985). Can small deviations from rationality make significant differences to economic equilibria? *American Economic Review, 75*(4), 708–720.

Arthur, W. B. (2009). *The Nature of Technology: What It Is and How It Evolves*. Free Press.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT 2021)*, 610–623.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Harvard Business School Working Paper No. 24-013.

Girotra, K., Terwiesch, C., & Ulrich, K. T. (2010). Idea generation and the quality of the best idea. *Management Science, 56*(4), 591–605.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica, 25*(4), 501–522.

Hao, Q., Xu, F., Li, Y., & Evans, J. (2025). Artificial intelligence tools expand scientists' impact but contract science's focus. *Nature*, 649, 1237.

Hyytinen, A. (2025). Mitä tekoäly ja sen kehitys merkitsevät taloudelle ja tutkimukselle? Muutamia näkemyksiä ja joitakin arvauksia. *Kansantaloudellinen aikakauskirja, 121*(3).

Jones, C. I. (2024). The AI dilemma: Growth versus existential risk. *American Economic Review: Insights, 6*(4), 575–590.

Jones, C. I. (2026, January). A.I. and our economic future. NBER Working Paper 34779, National Bureau of Economic Research. In preparation for the Journal of Economic Perspectives

Korinek, A. (2023). Generative AI for economic research: Use cases and implications for economists. *Journal of Economic Literature, 61*(4), 1281–1317.

Korinek, A. (2025). AI agents for economic research: August 2025 update of "Generative AI for economic research: Use cases and implications for economists" (*Journal of Economic Literature, 61*(4)). Working paper.

Lu, C., Lu, C., Lange, R. T., Foerster, J., Clune, J., & Ha, D. (2024). The AI Scientist: Towards fully automated open-ended scientific discovery. *arXiv:2408.06292*.

Manning, B. S., Zhu, K., & Horton, J. J. (2024). Automated social science: Language models as scientist and subjects. NBER Working Paper No. 32381.

Meincke, L., Girotra, K., Nave, G., Terwiesch, C., & Ulrich, K. T. (2024). Using large language models for idea generation in innovation. The Wharton School Research Paper.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187–192.

Romer, P. M. (1993). Two strategies for economic development: Using ideas and producing ideas. *The World Bank Economic Review, 6*(Suppl. 1), 63–91.

Ruan, K., Wang, X., Hong, J., Wang, P., Liu, Y., & Sun, H. (2026). Evaluating LLMs' divergent thinking capabilities for scientific idea generation with minimal context. *Nature Communications*. https://doi.org/10.1038/s41467-026-70245-1

Rusak, G., Manning, B. S., & Horton, J. J. (2025). AI agents can enable superior market designs. Working paper, April 2025.

Schaeffer, R., Miranda, B., & Koyejo, S. (2023). Are emergent abilities of large language models a mirage? *Advances in Neural Information Processing Systems (NeurIPS 2023)*, 36.

Trammell, P., & Korinek, A. (2023). Economic growth under transformative AI. NBER Working Paper No. 31815.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., & Fedus, W. (2022). Emergent abilities of large language models. *Transactions on Machine Learning Research*.

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica, 47*(3), 641–654.

Weitzman, M. L. (1998). Recombinant growth. *Quarterly Journal of Economics, 113*(2), 331–360.