A Future Worth Working For

Meaningful Work and Human Flourishing in the Age of AI

Hugo Latapie

Taijitu AI Inc

9 November 2026 | Version 20

Abstract

Contrary to widespread predictions of AI-driven job replacement, today's AI cannot reliably perform the real-world work required to replace human expertise. *The Illusion of Accelerating Intelligence* develops a mathematical model with proven results and an evidence-based scientific hypothesis identifying fundamental limitations of current AI. Rapid improvements on fixed tests reflect increasingly effective optimization within closed problems, not the ability to reliably handle the changing circumstances, missing information, and unforeseen conditions encountered in real-world work. These limitations challenge the prevailing narrative of imminent widespread job replacement. Meanwhile, AI can expand what people can accomplish, creating new opportunities and potentially increasing demand for deep expertise across many fields.

Realizing these opportunities requires more than accumulating knowledge. Expertise must remain accountable to reality through the continual examination and correction of assumptions, abstractions, and established methods. Establishing worthwhile human intent is a distinct and potentially even greater challenge. *The Human Paperclip Maximizer* examines how measures of success can displace the human purposes they were intended to serve. Developing Goodhart immunity requires recognizing and correcting this divergence, even when prevailing incentives reward it.

Even when future AI overcomes today's limitations, human livelihoods need not depend on outperforming it. Understanding human intentions does not confer authority to decide for people. Professionals, entrepreneurs, and collaborative enterprises can employ increasingly capable AI while retaining responsibility for the purposes pursued and the consequences for those affected. This paper explores how people can develop reality-grounded expertise, establish worthwhile intent, and provide Goodhart immunity through sustained practice and real-world experience. These capabilities can help people shape technological progress around human purposes, rather than reshape human lives around technology. The ultimate objective is human flourishing: greater freedom to pursue worthwhile purposes and shape a meaningful life.

1 Why expertise matters now

The Illusion of Accelerating Intelligence challenges the idea that today's AI merely needs more scale and self-improvement to become reliable at open-ended work [2]. It argues that much of the apparent acceleration may be increasingly effective optimization within closed problems, rather than comparable progress in dealing with the world those problems only partly describe.

The distinction is fundamental. If the obstacle is insufficient computational power, more computation can help. If the obstacle is missing information, searching harder through the same information cannot supply what is absent. A system can keep improving against its tests while remaining unable to distinguish situations that require different answers. Overcoming that limit requires obtaining relevant evidence and revising the model.

That is why deep expertise matters now. Knowing a field helps people recognize which assumptions may fail, decide what to measure, and discover what needs to be understood next. AI can help with both investigation and analysis, while expertise helps people judge what has actually been established and what remains uncertain.

AI already solves difficult mathematical problems whose assumptions and standards of proof are fixed [3, 4, 5]. Mathematics and engineering also involve creating and revising the formulations themselves. Deciding what to formalize, developing useful abstractions, and discovering which relationships matter can be open-ended work requiring deep knowledge.

As AI makes more research and engineering projects possible, it creates further work: investigating new questions, developing theories and designs, and solving problems revealed in practice. Those projects also need people who can recognize when an objective rewards the wrong outcomes and bring it back into line with what the project is meant to achieve. This is the argument for greater demand for deep mathematical and engineering expertise: increasing our ability to calculate and optimize expands what we can attempt, and with it the open-ended work that needs doing.

The same reasoning applies to science, medicine, education, law, business, finance, and public service. A teacher must discover why a particular student is struggling; a clinician must investigate a patient whose condition does not fit the expected pattern. Across these fields, expertise helps people find missing evidence, develop new approaches, and recognize when an apparent success is failing the people the work is meant to serve [1, 2]. Greater capacity to perform specified tasks creates opportunities to address needs that previously went unmet.

The arts and humanities also involve open-ended discovery. Writers, artists, performers, and scholars develop interpretations, explore human experience, and find ways to express something worth sharing. Craft, cultural understanding, and lived experience contribute to that work. AI can expand what people can attempt, while their interests, talents, and developing mastery help shape what they choose to create. The opportunity to develop expertise and undertake meaningful open-ended work extends across these fields, too.

2 Why Goodhart immunity can protect jobs

A call center can shorten calls while leaving customers' problems unresolved. Its numbers improve, but its service gets worse. *The Human Paperclip Maximizer* examines this Goodhart failure: people and institutions pursue a measure of success at the expense of the purpose behind it [1]. Goodhart immunity means continually preventing and correcting such failures. The work includes questioning targets, checking what actually happens, and changing course when meeting the target does more harm than good.

The purpose is human flourishing: health, security, knowledge, relationships, and the freedom to shape a life. A correct answer or a finished product can be exactly what someone needs. But success still depends on what that result does for people, not merely on how efficiently it was produced.

Now suppose AI can investigate a changing world, recognize gaps in its knowledge, understand evolving human intentions, and revise its objectives. It might notice that a request conflicts with the requester's deeper priorities. Such territory-aligned, intent-based AI could help detect and correct Goodhart failures.

Even then, understanding what people want is different from being authorized to decide for them. An AI might predict which tradeoff someone would accept and suggest a better option. Predicting that choice is not the same as receiving consent. Nor does understanding one person's wishes settle a conflict with someone else's interests. If human flourishing is the goal, the people affected must be able to question decisions, reject proposals, and change their minds.

Engineers, mathematicians, clinicians, and advisers can help people exercise those choices. A professional may use AI that does the analysis better while remaining responsible for how the work serves the people involved. The job includes explaining consequences, identifying conflicting interests, involving those affected, and changing course when necessary. Deep knowledge helps the professional spot problems and challenge decisions. Signing whatever the AI produces is not enough.

Why pay a person for this if AI can help with all of it? Clients and employers may still want a professional who answers to them for the work as a whole. They can delegate analysis and execution while retaining a person responsible for protecting their interests. We expect people to keep paying for this work where decisions affect lives, rights, livelihoods, or shared resources. The professional also needs the power to act: responsibility without access to evidence or authority to change a decision is an empty promise.

This can protect jobs across technical, commercial, creative, and public work. Tasks and staffing may change. The reason for paying someone to represent people's interests and take responsibility for results need not disappear as AI improves. An employer that rewards only speed may neglect this work, but that is a Goodhart failure, not a demonstration that the work no longer matters.

3 The IPO and the replacement story

An IPO would give Anthropic's private investors a route to sell their shares to public-market buyers. Its Series H valued the company at $965 billion [6]. For investors entering near that price, earning several-fold returns could require the public market to support a multi-trillion-dollar valuation when they sell.

The IPO story must explain why buyers should pay that much. Selling useful AI tools could produce a successful business without justifying that price. Large-scale job replacement offers a much bigger revenue story: AI suppliers could capture a substantial share of what businesses currently pay their workers.

Our hypothesis is that job replacement may be the only narrative large enough to support the IPO valuation needed for investors' desired exits. The need to persuade public-market buyers may therefore help drive the replacement story, rather than that story simply following from demonstrated AI capability. This is a hypothesis about IPO incentives, not a claim to know anyone's private intentions.

4 Building a worthwhile career

Choose a field where interests and strengths can grow into useful expertise. Test that choice through real work and available opportunities. Nye and colleagues found that vocational interests were associated with performance and persistence. The fit between interests and the work environment was more strongly associated with performance than interest scores alone [7]. Exceptional ability is worth taking seriously, but it is not a prerequisite. People also develop their abilities through practice, mentorship, and demanding problems. People can pursue this work within existing organizations, through their own businesses, or in new forms of collaborative enterprise.

Creative work draws on craft, lived experience, and a person's voice. Bellaiche and colleagues found that AI-generated paintings received more favorable appraisals when labeled human-created than when labeled AI-created [8]. This shows that perceived authorship matters, not that AI cannot make compelling art. AI can give creators more ways to express themselves; they remain responsible for what they publish. In any field, paid work may be a source of meaning or a way to support commitments elsewhere. A career should serve a life, not consume it in the search for tasks that machines cannot perform.

Learning must change with the tools. Brynjolfsson, Li, and Raymond found that AI assistance increased customer-support productivity, with larger gains among less experienced workers [9]. Getting better results with AI is not the same as learning why those results are right. Newcomers need chances to question recommendations, examine failures, and receive informed feedback. If automation removes the routine tasks through which people learned, employers need paid alternatives: supervised investigations, reviews of real cases, and responsibility for limited improvements.

Employers and public institutions help determine whether this becomes good, accessible work. Acemoglu and Restrepo distinguish automation that displaces labor from new tasks that create work for people [10]. Employers must pay for investigation and correction, train people, and let them speak with those affected. Workers also need time to investigate and the authority to change decisions. Employers cannot reward only speed and then expect care. The benefits should reach the people doing the work as well as those receiving it.

Develop expertise for open-ended problems now, and learn to take responsibility for what that expertise achieves. That responsibility demands more than technical proficiency: it requires continually correcting one's understanding against reality and keeping the work aligned with worthwhile purposes.

5 Developing Reality-Grounded Expertise and Goodhart Immunity

Developing deep expertise is difficult. Developing the ability to continually correct that expertise against reality is harder still. Expertise involves acquiring knowledge, methods, and abstractions that make complex problems manageable. Yet those same abstractions can become obstacles when people mistake their models for the reality they describe. Years of successful practice may reinforce assumptions that work under familiar conditions but fail when circumstances change.

Henry Ford recognized this danger more than a century ago. In *My Life and Work* (1922), he wrote: "We have most unfortunately found it necessary to get rid of a man as soon as he thinks himself an expert." Ford observed that accumulated experience could become a catalog of supposed impossibilities, preventing people from discovering better methods. He described manufacturing improvements achieved despite established expert judgments that they could not be done. His objection was not to knowledge

or technical competence, but to the tendency to mistake the limits of existing experience for limits imposed by reality [11].

The central challenge is recognizing what our existing understanding fails to capture. Missing information does not necessarily announce itself through obvious errors. A model may produce consistently successful results while excluding distinctions that become important under different conditions. Recognizing these limitations requires intellectual independence, curiosity, and a willingness to question even deeply established knowledge.

People can deliberately develop this capability through real-world investigation. Formulate expectations before acting, observe what actually happens, and investigate discrepancies. Seek out unfamiliar conditions and cases where established explanations fail. Learn to distinguish evidence from interpretation and to recognize assumptions that have not been tested. When observations contradict a model, investigate whether the model itself needs revision rather than merely adjusting its parameters. Mentorship is particularly valuable when it teaches people how to discover what they do not yet understand, rather than simply transmitting established answers.

Establishing worthwhile intent presents a separate challenge. Understanding reality does not determine what we should seek to accomplish. Worthwhile purposes involve human needs, aspirations, relationships, and competing interests. These are not always obvious, consistent, or shared. Even individuals may discover that their immediate objectives conflict with their deeper priorities.

Developing this capability begins with understanding the people an activity is intended to benefit. What do they actually want to accomplish? Why does it matter to them? What consequences would constitute genuine success? Whose interests might be affected, and who has the authority to make the relevant decisions? Answering these questions requires listening, empathy, critical thinking, and the ability to recognize conflicts between stated objectives and underlying intentions. It also requires accepting that people's priorities may evolve and that worthwhile intent must sometimes be reconsidered.

Providing Goodhart immunity introduces another difficulty: preserving worthwhile intent when institutions, incentives, and measures of success begin pulling activity in other directions. Financial rewards, performance targets, professional recognition, and organizational routines can gradually displace the purposes they were originally intended to serve. Recognizing this divergence can be difficult, particularly when established measures indicate success and the people benefiting from those measures have little incentive to question them.

An effective Goodhart immunity provider must therefore learn to distinguish the intended human benefit from the measures used to evaluate progress. This requires examining actual consequences, listening to those affected, identifying harmful incentives, and recognizing when apparent success conceals failure. Correcting such problems may require changing objectives, measures, incentives, or established practices. It also requires the courage to challenge prevailing assumptions and the practical authority to act. Responsibility without the ability to obtain evidence or correct decisions cannot provide meaningful protection.

These capabilities develop through sustained practice. Seek work that exposes the consequences of decisions, including unexpected and unwelcome results. Learn from people with different experiences and perspectives. Examine not only whether objectives were achieved, but whether achieving them produced the intended benefits. Practice revising both your understanding of reality and your understanding of what is worth accomplishing, while recognizing that these are distinct questions requiring different kinds of judgment.

Individuals can cultivate these disciplines throughout their careers. Entrepreneurs can incorporate them into the foundations of new enterprises. Established organizations can develop them through professional training, institutional practices, and meaningful accountability. In each setting, the opportunity is to become increasingly capable of understanding reality, identifying worthwhile purposes, and ensuring that actions continue serving those purposes.

Neither capability has a natural endpoint. Reality changes, understanding remains incomplete, and human intentions evolve. Becoming proficient at continually correcting our abstractions and preserving worthwhile intent is therefore a lifelong undertaking. These demanding capabilities offer a direction for professional development and meaningful contribution, regardless of how capable AI becomes.

Where customers, clients, or employers pay for this responsibility and people have the power to act, better AI can strengthen both their work and the enterprises they build. Goodhart immunity can then provide lasting livelihoods without requiring people to outperform AI.

References

[1] H. Latapie. *The Human Paperclip Maximizer: Goodhart's Law, Financial Optimization, and the Preconditions for AI Alignment*. Version 10, 6 November 2026. https://doi.org/10.5281/zenodo.24026512. Author-hosted PDF.

[2] H. Latapie. *The Illusion of Accelerating Intelligence: Closed Optimization, Open-Ended Reality, and the Two-Thirds Limit*. Version 13, 5 November 2026. https://doi.org/10.5281/zenodo.24023570. Author-hosted PDF.

[3] AlphaProof and AlphaGeometry teams. AI achieves silver-medal standard solving International Mathematical Olympiad problems. Google DeepMind, 25 July 2024. Source report.

[4] T. Hubert, R. Mehta, L. Sartran, et al. Olympiad-level formal mathematical reasoning with reinforcement learning. *Nature*, 651:607–613, 2026. Published online 12 November 2025. https://doi.org/10.1038/s41586-025-09833-y.

[5] T. Luong and E. Lockhart. Advanced version of Gemini with Deep Think officially achieves gold-medal standard at the International Mathematical Olympiad. Google DeepMind, 21 July 2025. Provider account, including IMO grading of the submitted solutions. Source report and linked solutions.

[6] Anthropic. Anthropic raises $65B Series H at $965B valuation. 28 May 2026. https://www.anthropic.com/news/series-h.

[7] C. D. Nye, R. Su, J. Rounds, and F. Drasgow. Vocational Interests and Performance: A Quantitative Summary of Over 60 Years of Research. *Perspectives on Psychological Science*, 7(4):384–403, 2012. https://doi.org/10.1177/1745691612449021.

[8] L. Bellaiche, R. Shahi, M. H. Turpin, A. Ragnhildstveit, S. Sprockett, N. Barr, A. Christensen, and P. Seli. Humans versus AI: whether and why we prefer human-created compared to AI-created artwork. *Cognitive Research: Principles and Implications*, 8:42, 2023. https://doi.org/10.1186/s41235-023-00499-6.

[9] E. Brynjolfsson, D. Li, and L. Raymond. Generative AI at Work. *The Quarterly Journal of Economics*, 140(2):889–942, 2025. https://doi.org/10.1093/qje/qjae044.

[10] D. Acemoglu and P. Restrepo. Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives*, 33(2):3–30, 2019. https://doi.org/10.1257/jep.33.2.3.

[11] H. Ford, in collaboration with S. Crowther. *My Life and Work*. 1922. Chapter V, "Getting Into Production." https://www.gutenberg.org/ebooks/7213.