In Defence of AI-Assisted Intellectual Work

Why Effort Is Not the Same Thing as Value

Dr. Yves J. Hilpisch[1]

August 31, 2026

Abstract

Generative artificial intelligence has made it cheap to produce bad prose, and the resulting flood deserves criticism. But "AI slop" names a failure of quality, not a method of production. This essay argues that professional intellectual artefacts should ordinarily be judged by what they deliver—accuracy, clarity, insight, usefulness, and fitness for purpose—rather than by how much manual effort they consumed. Drawing on three decades of writing, coding, teaching, and publishing before and after generative AI, I distinguish token generation from authorship and accountability, explain why experienced practitioners can obtain exceptional leverage from these tools, and identify the cases in which provenance rightly matters. The hardest issue is educational: the labour that experts can now automate was also the apprenticeship through which they became capable of judging the machine. The task ahead is therefore not to preserve inefficient production for its own sake, but to preserve expertise, responsibility, and trust while production becomes radically cheaper.

[1]Get in touch: https://linktr.ee/dyjh. Web page: https://hilpisch.com. Research, structuring, drafting, editing, and typesetting were assisted by generative AI tools, including OpenAI Codex, under the author's direction and review. Additional contact: https://theaiengineer.dev (The AI Engineer), https://python-for-finance.com (CPF Program), and team@tpq.io.

1 The 100-Kilometre Walk

Imagine that I arrive at a client's office after travelling 100 kilometres. The client asks how I came. "By car," I reply. Their face falls. "That is disappointing. I would have respected this meeting much more if you had walked. Using a car feels rather sloppy."

The exchange sounds absurd because the journey is instrumental. I am there to contribute to the meeting, not to exhibit my capacity for avoidable exertion. Walking might be admirable in a charity event or necessary in an endurance test. For an ordinary business meeting, however, the distance covered on foot adds nothing to the advice I give once I arrive.

Yet something close to this argument is now made about intellectual work. Readers announce that they will not spend time on anything produced with artificial intelligence (AI), "out of principle." Some treat evidence of generative artificial intelligence (GenAI) in a book, paper, report, or software project as evidence that the work is illegitimate. The toolchain becomes a proxy for quality before the artefact has even been examined.

This essay is written primarily for readers and professional knowledge workers who must decide what deserves attention; it also speaks to educators responsible for forming the next generation of experts. My claim is simple, although not absolute: where provenance is not itself part of the promise, intellectual work should be judged principally by its function. Is it correct? Is it clear? Is it useful? Does it offer insight? Is it original where originality matters? Does it do the job for which it was made?

That is a case for standards, not for permissiveness. A technically polished falsehood remains a falsehood. A hundred pages of fluent filler remain a waste of time. I do not defend machine-generated mediocrity. I defend the possibility that excellent work can emerge from a production process in which machines perform a great deal of the drafting, coding, checking, and typesetting.

The proposition

The relevant question is not whether AI touched the work. The relevant question is whether the work deserves your attention. Judge the artefact, then examine the toolchain wherever the toolchain is materially relevant.

That final qualification matters. A student sitting an unaided examination, a scientist reporting experimental measurements, and a person writing an intimate letter make promises about process that the author of a technical manual does not. I will return to these cases. First, however, we should separate a genuine cultural problem from a confused diagnosis of it.

2 Slop Is a Quality Category

AI slop exists, and there is a great deal of it: search-engine bait, fabricated books, synthetic reviews, generic corporate posts, untested code, illustrations with telltale errors, and papers whose references collapse under inspection. GenAI lowers the marginal cost of producing plausible material, so the supply of plausible rubbish expands. I share the frustration with that development.

But the phrase often bundles together two very different things:

- low-quality work produced cheaply with AI; and
- high-quality work produced efficiently with AI.

The first deserves rejection. The second deserves evaluation. Treating them as the same phenomenon is like condemning photography because cheap cameras created billions of bad photographs, or desktop publishing because it enabled ugly newsletters. A production technology changes the distribution and volume of artefacts. It does not determine the quality of each one.

Bad human writing did not begin in November 2022. Long before modern chatbots, the world contained dull textbooks, careless consultancy reports, padded academic articles, derivative novels, incomprehensible manuals, and fragile software. Human origin never guaranteed care, truth, or style. What has changed is scale. The bottleneck has moved downstream: from making content to filtering it.

This, I think, explains why the present abundance feels so unpleasant. Attention is finite while generated material is approaching zero marginal cost. Trust, reputation, expert curation, verification, and a reader's ability to discriminate become more valuable, not less. Labelling every AI-assisted artefact "slop" does nothing to solve that filtering problem. It merely replaces judgement with prejudice about provenance.

There is also a psychological temptation to infer value from visible effort. Research on the "effort heuristic" found that people sometimes rate an artefact more highly when they believe more effort went into producing it [1]. That shortcut may make sense when effort is the only available signal of care. Once the artefact can be inspected, tested, and compared, however, effort is no longer the outcome. The reader receives the book, not the months the author spent at the desk.

For a long time, those months nevertheless served as a rough signal. Producing a substantial technical book demanded enough time, expertise, and expense that the production cost itself suggested seriousness. It was never a reliable guarantee—bad books can also be expensive to make—but it was an easy heuristic. GenAI weakens that correlation: polish and length no longer reveal how much expertise or care lies behind them. Some of the hostility towards AI writing is, I suspect, a reaction to the loss of that familiar signal. The rational response is not to preserve production difficulty artificially, but to develop better signals: evidence, transparent sourcing, editorial standards, reputation, and work that survives inspection.

3 The Functional Case

The functional perspective I find most useful has a familiar precedent in computer science. In 1950, Alan Turing declined to settle the vague question "Can machines think?" by first reconstructing the private mechanism of thought. He proposed an operational test based on observable performance in an "imitation game" [2]. Whatever one thinks of that test as a complete account of intelligence, the methodological move remains powerful: begin with what a system does rather than insist that it reproduce a familiar implementation.

The same move clarifies intellectual production. The traditional pipeline looked approximately like this:

> human idea $\longrightarrow$ extended human drafting and implementation $\longrightarrow$ finished artefact.

The emerging pipeline may look like this:

> human purpose and expertise $\longrightarrow$ AI-mediated drafting, testing, and revision $\longrightarrow$ human selection and accountability $\longrightarrow$ finished artefact.

If the second artefact explains a mathematical idea more clearly, contains fewer mistakes, supplies better examples, and serves its reader more effectively, why should it be intrinsically inferior? The implementation differs; the function may be equivalent or better.

Modern professional life already accepts this logic almost everywhere. Mathematicians use symbolic algebra systems. Accountants use spreadsheets. Architects use computer-aided design. Photographers use autofocus and computational processing. Programmers use compilers, libraries, debuggers, and code completion. Authors use search engines, grammar checkers, citation managers, editors, and typesetters. Quantitative analysts do not prove their seriousness

by pricing a complex derivative with pencil and paper when reliable numerical methods are available.

GenAI feels different because it reaches into activities that educated professionals regarded as the distinctly human core of their status: composing sentences, producing explanations, writing code, synthesising sources, and proposing ideas. We have spent centuries inventing tools that reduce labour. The controversy sharpens when the labour being reduced is our own.

That discomfort does not establish a moral boundary. A calculator is not virtuous because it automates arithmetic but vicious when it helps explain the calculation. The sensible boundary is set by purpose, risk, and responsibility. Which parts can be delegated? Which claims require independent verification? Who is qualified to review the result? Who stands behind it when something fails?

The normative consequence should be stated plainly. If a faster process produces an equal or better result under the same standard of responsibility, deliberately choosing a substantially slower or inferior process merely to display manual effort is not a mark of intellectual integrity. Integrity lies in the standards applied to the work and the honesty with which it is presented, not in preserving avoidable labour.

4 Mathematics: The Artefact Is the Proof

As someone trained in mathematical finance, I find mathematics the cleanest test of the functional argument. A proof is not made correct by the prestige, effort, or biography of the person who conceived it. Nor can a valid proof be made false because its decisive idea came from a machine. Once the argument has been stated precisely, origin and validity are different questions.

This does not mean that mathematical value is entirely binary. A correct proof may be trivial, derivative, needlessly long, or conceptually unenlightening. Another proof of the same theorem may reveal a connection that changes the field. Exposition, novelty, elegance, and explanatory power all matter. But the first gate remains unusually hard and impersonal: do the conclusions follow from the definitions, axioms, and preceding steps? If they do, rejecting the mathematical claim solely because AI found the argument would be a category error. One may debate attribution, publication policy, or the wider consequences for the profession. One cannot thereby refute the theorem.

The distinction is the old one between discovery and justification. Mathematicians have arrived at ideas through geometric pictures, physical analogies, numerical experiments, dreams, fortunate mistakes, conversations, exhaustive computation, and years of deliberate work. None of these discovery stories appears as a premise in the finished proof. The artefact must stand when the scaffolding of discovery is removed. AI is a new and powerful source of such scaffolding, but it does not change what mathematical verification asks.

Recent results make the point concrete. In May 2026, OpenAI reported that an internal general-purpose reasoning model had disproved a conjecture associated with Paul Erdős's planar unit-distance problem, open since 1946. The longstanding conjecture held, in asymptotic terms, that the maximum number of unit-distance pairs among n points in the plane was bounded by $n^{1+o(1)}$. The model produced an infinite family of configurations with at least $n^{1+\delta}$ unit-distance pairs for some fixed $\delta > 0$, using an unexpected bridge from algebraic number theory to discrete geometry. OpenAI released the proof and reported that external mathematicians had checked it and written companion remarks [3].

What matters to me is not that a model emitted something shaped like a proof; language models have long been able to do that. It is that the argument survived expert mathematical scrutiny. Had exactly the same sequence of definitions, constructions, and deductions first appeared in a human researcher's notebook, its mathematical status would be the same. Conversely, an invalid argument does not become a proof because a famous mathematician, a prestigious institution, or an expensive model produced it.

In August 2026, OpenAI announced ten further results that it said resolved or substantially advanced open problems across geometry, coding theory, group theory, operator algebras, complexity theory, cryptography, and combinatorics. According to the accompanying account, an internal model generated the mathematical arguments; humans then helped prepare the manuscripts, and the model formalised the arguments as Lean certificates [4]. These claims will and should receive detailed scrutiny from the relevant mathematical communities. That scrutiny concerns the artefacts: whether the hypotheses are represented correctly, whether the formal statements capture the intended problems, whether the deductions check, and whether the claimed novelty survives comparison with the literature.

This example also sharpens the distinction between validity and authorship. The theorem does not care who found the proof, but the mathematical community reasonably does care about honest credit. OpenAI explicitly argued that presenting a machine-generated proof as wholly human work would misrepresent its origin [4]. That is exactly the position defended here. Provenance should be disclosed where it bears on attribution and trust; it should not be used as a shortcut around examining the work.

The mathematical test

A proof must be rejected when the argument fails, not when the origin story offends us. Discovery may be human, computational, or hybrid; validity resides in the checkable chain of reasoning.

Mathematics is an unusually favourable domain for this principle because many claims admit exact checking and some can be formalised in proof assistants. Most writing cannot be reduced to a verifier. A history book, policy essay, or technical tutorial must also be judged for selection, framing, and meaning. Yet the mathematical case exposes the central error with special clarity: provenance and quality may be correlated in practice, but they are not the same property.

5 Software: Users Judge What Runs

Software brings the same argument closer to everyday experience. A program is not a theorem: testing cannot establish correctness in every possible state, and qualities such as usability, maintainability, security, and design resist reduction to a single verdict. Yet we still evaluate software predominantly by what it does. Does the application solve the user's problem? Is it reliable, safe, responsive, intelligible, and affordable? Most users neither know nor care which engineer typed a particular line.

The reported changes in how technology companies produce that code are striking. As of May 2026, Anthropic said that Claude had authored more than 80 percent of the code merged into its codebase, up from the low single digits before the February 2025 research preview of Claude Code. It also reported that its typical engineer merged eight times as many lines of code per day in the second quarter of 2026 as in 2024, while explicitly cautioning that lines of code overstate true productivity and say little by themselves about quality [5].

The aggregate figure becomes more tangible in the workflow of Boris Cherny, the creator and head of Claude Code. In a February 2026 interview, Cherny said that Claude Code was writing 100 percent of his code and that he had not manually edited a single line since November 2025. He nevertheless described himself as a prolific coder, shipping roughly 10–40 pull requests per day and running about five agents in parallel even during the interview. The part I find decisive is that he did not equate the absence of manual typing with the absence of responsibility: he still reads the code, he said, and retains human checkpoints for correctness and safety [6]. His work has not vanished. It has shifted towards setting agents in motion, specifying what should be built, reviewing their output, and deciding what may be merged.

The definitions differ across companies, which makes direct ranking unhelpful. In Alphabet's

first-quarter 2025 earnings call, Chief Executive Officer Sundar Pichai said that well over 30 percent of checked-in code involved employees accepting AI-suggested solutions [7]. That is not the same measure as code "authored by Claude." It captures accepted assistance rather than complete autonomous implementation. Still, both figures describe AI participation not at the margins of experimental demos, but inside the engineering processes of companies operating products at enormous scale.

OpenAI has documented an even more concentrated experiment. A small engineering team built an internal software product whose repository grew to roughly one million lines without humans directly writing any code. Codex generated the application logic, tests, continuous-integration configuration, documentation, observability, and internal tooling. The product had hundreds of internal users, including daily power users, and the team estimated that it had been built in about one-tenth of the time manual implementation would have required [8]. Here again, the human job shifted towards specifying intent, structuring the environment, designing feedback loops, reviewing outcomes, and encoding architectural standards—almost exactly the move "up the stack" I describe below.

I would not turn these company-reported figures into a league table. They are neither independently audited nor measured on a common denominator. I take them as evidence of direction and scale, not proof that AI-written code is automatically good. Anthropic's caveat about lines of code matters, and OpenAI's account describes recurring work to control inconsistency and accumulated "slop." Productivity counts only while the resulting system remains useful and governable.

Still, the market's functional response is revealing. Users do not generally abandon a search engine, an AI assistant, a banking application, or a productivity tool upon learning that AI generated a substantial share of its code. They keep using software that does the job and reject software that is unreliable, confusing, insecure, or too expensive. When an application crashes, "a human typed every line" is no defence. When it works well, "a model wrote most of it" is no disqualification.

There are legitimate process questions. Enterprise buyers may require evidence about security review, software supply chains, licenses, privacy, and accountable ownership. Regulators may demand controls for consequential systems. These requirements matter because they affect the reliability and legitimacy of the product, not because manual keystrokes possess intrinsic merit.

For me, mathematics, software, and writing are not separate worlds. A technical book on quantitative finance combines all three: mathematical reasoning, executable implementations, and prose that explains how they fit together. We readily accept symbolic and numerical tools in the mathematics, and we increasingly accept AI-generated code in the implementation. Why should the explanatory prose alone be subjected to a test of manual production? It would be peculiar to accept machine assistance for the equation and the program, then declare the paragraph connecting them illegitimate because a machine helped draft it.

The common standard should be whether the parts form a coherent and reliable whole. Human-written software has shipped with catastrophic bugs, just as human-written books have contained errors. AI-assisted systems add new failure modes and new speed, but the rational response is stronger evaluation and responsibility—not automatic rejection based on who or what touched the keyboard.

6 A Natural Experiment Across Two Eras

I do not approach this question as someone who discovered writing through a chatbot. I completed my doctoral thesis in mathematical finance in 2000. Before GenAI was available, I authored ten books, including two German-language books on value-based corporate management and seven books on Python and finance, in addition to the thesis. Across those projects I pub-

lished more than 3,500 pages. I also wrote tens of thousands of lines of code and spent decades working with Python, LaTeX, Linux, quantitative models, teaching materials, and publishing systems.

I know what it means to write a technical book manually because I spent a substantial part of my professional life doing exactly that. I know the blank page, the failed structure, the laborious code example, the late discovery that notation changed between chapters, the reference checked for the third time, and the paragraph rewritten until it finally says what it should.

That history matters because it gives me a within-person comparison. Hold the domain knowledge, standards, taste, and authorial purpose broadly constant; then change the tools. What happens?

A technical book I can now produce in two days is not really a two-day book. It is the product of 30 years of accumulated expertise applied through a radically more powerful production system: 30 years plus two days. The compressed production time does not make the earlier learning disappear. It is what allows that learning to be brought to bear with far greater leverage.

In my experience, the result is not merely more output but better output. During the past year, working with GenAI and agentic tools, I have produced more books and papers than during much longer stretches of the preceding decades. More importantly, I have improved them in ways that the old economics of authorship made impractical.

For any given explanation, I can request five alternatives, reject four, combine the strongest elements, and test the surviving version against different audiences. I can ask one system to attack an argument and another to check terminology. I can generate code, run it, inspect failures, revise tests, and trace inconsistencies between prose and implementation. I can examine whether notation is stable across hundreds of pages, restructure a chapter without retyping it, and explore diagrams or exercises that would previously have lost to the deadline. I can use independent passes to check claims and references—while still opening the underlying sources, because models can invent citations with great confidence.

None of this makes quality automatic; it makes additional attempts cheap. A traditional author cannot afford limitless editorial passes. An AI-assisted author can move much closer to that ideal, provided the author has the judgement to distinguish improvement from merely more words.

My experience is only one observation, so I also look to controlled studies. They do not establish my personal claim, but they make the underlying mechanism plausible. In an experiment with 453 college-educated professionals completing bounded writing tasks, access to ChatGPT reduced average completion time by 40 percent and raised independently assessed quality by 18 percent [9]. A large field study of consultants found substantial speed and quality gains on tasks inside the model's capability frontier, alongside worse performance when users applied AI beyond that frontier [10]. Evidence from more than 5,000 customer-support agents likewise shows that AI can raise productivity and help less-experienced workers adopt practices associated with stronger performers [11]. That last result cautions me against claiming that only experts benefit. The broader lesson is neither "AI always wins" nor "humans always know best." Task, workflow design, and judgement determine whether cheap generation becomes leverage or error.

Why, then, should I spend another year manually drafting a technical book if, under my ideas, strict direction, repeated review, and final responsibility, an AI-assisted process can produce a stronger version in a fraction of the time? The declaration "I typed every sentence myself" may describe effort. It does not establish reader value.

7 The Author's Job Is Moving Up the Stack

Before GenAI, production consumed much of the available energy: drafting prose, formatting equations, implementing examples, writing boilerplate, debugging, locating references, converting notes into chapters, and maintaining files. These tasks were not meaningless. They often sharpened thought. But they also competed with higher-order work.

The bottleneck is now moving towards a different set of activities:

- choosing a question worth answering;
- knowing what a satisfactory answer would require;
- imposing an intelligible structure;
- supplying domain assumptions and constraints;
- recognising omissions, false confidence, and subtle errors;
- comparing alternatives and exercising taste;
- integrating text, mathematics, code, and evidence; and
- accepting responsibility for the released artefact.

This is not the disappearance of authorship. It is authorship moving up the stack: the scarce skill shifts from producing a plausible first draft to knowing what the finished work ought to be.

There is another reason the new division of labour feels familiar to me. Early in my career as a management consultant, I began managing teams of consultants. I would define and delegate workstreams; team members would conduct interviews, build analyses, develop findings, and turn those findings into slides. My central responsibility was not to perform every calculation or type every sentence myself. It was to direct the team, challenge the results, reconcile conflicting pieces, and compile a coherent slide deck for steering committees and executive board meetings.

The final deck had to tell one story. Analyses produced by different people had to use compatible assumptions. Recommendations had to follow from the evidence. Repetition had to disappear, gaps had to be filled, and the level of detail had to suit senior decision makers. A steering committee did not receive the hours worked by ten consultants. It received the integrated artefact and expected me to stand behind it.

That workflow is structurally close to working with AI agents. Different agents can research a question, produce code, test claims, criticise a draft, or propose a visual. The lead author then has to specify the tasks, assess the contributions, resolve inconsistencies, and transform distributed output into something coherent and valuable to the reader. Calling this arrangement illegitimate merely because some team members are machines overlooks how much professional intellectual work has always been organised.

Can agent-assisted work go wrong? Certainly. An agent may misunderstand its brief, invent support, duplicate another agent's work, or optimise a local task at the expense of the whole. Human consulting teams fail in the same broad ways: analysts use incompatible definitions, overlook evidence, make spreadsheet errors, or build individually polished slides that do not form an argument. The answer was never to abolish delegation and insist that the engagement leader do all the work alone. It was capable management, review, and accountability. AI makes those disciplines more important because production is faster, not less relevant because production is automated.

Consider a book on Python for finance. A language model can draft an explanation of Monte Carlo valuation and generate a program that runs. It cannot, by fluency alone, establish that the discretisation is appropriate, the probability measure is correct, the random-number

treatment is defensible, the code and equation implement the same model, or the pedagogical sequence meets the reader where they are. A novice may be impressed because the program prints a number. An expert asks whether it is the right number, for the right reason, under the right assumptions.

Research on knowledge work increasingly describes the shift in similar terms. A 2025 study of 319 knowledge workers found that GenAI moved reported critical-thinking activity towards verification, integration, and stewardship; it also found that higher confidence in the AI was associated with less critical-thinking effort [12]. That finding matches my own concern: the machine can reduce the cost of execution while increasing the premium on calibrated distrust.

The useful analogy is not a vending machine that dispenses finished expertise. It is a fast, tireless, unevenly capable team. It can draft, compare, refactor, summarise, and critique. It can also misunderstand the assignment, smooth over a contradiction, or fabricate support. Giving such a team vague instructions and publishing its first response is not authorship. Defining the project, directing the work, rejecting weak material, verifying decisive claims, and taking responsibility for the whole is.

8 Generation, Authorship, and Accountability

My consulting experience also helps me separate three concepts that are too often collapsed in this debate.

Generation concerns who or what produced the initial tokens, code, image, or diagram.

Authorship concerns who conceived the work, established its purpose, directed its development, selected and arranged its contents, revised it, and approved its final form.

Accountability concerns who stands behind the claims and bears the consequences of error, misuse, or deception.

These functions have never been identical. Authors have long worked with research assistants, co-authors, editors, copy editors, translators, proofreaders, typesetters, photographers, and software libraries. No serious definition of authorship has required the author physically to generate every symbol. GenAI expands the toolchain so dramatically that we can no longer ignore the old distinctions.

The distinction is also visible in copyright policy. The United States Copyright Office concluded in 2025 that using AI as an assistive tool does not by itself bar copyright protection, while material whose expressive elements are determined entirely by a machine is not protected merely because a person supplied prompts. Human selection, arrangement, and modification can matter [13]. Copyright law is not a complete philosophy of authorship, especially outside the United States, but its distinction is useful: the presence of a tool does not settle the question of human creative control.

My practical standard is direction plus responsibility. If I choose the thesis, specify the structure, supply the experience, interrogate the drafts, verify the claims, revise the argument, approve every page, and attach my name to the result, I am the author. The fact that a model proposed many of the sentences is material to disclosure, but it does not erase authorship. If I type a prompt, do not understand the response, and publish it untouched, attaching my name does not create meaningful authorship or responsible practice.

Disclosure should therefore be frank and useful. I state openly that my recent books and discussion papers are rapidly developed with GenAI tooling; this paper does the same on its title page. Concealing material AI use can mislead readers where they reasonably care about production or risk. But disclosure should inform evaluation, not replace it. “AI-assisted” is neither a quality seal nor a warning label. It describes the toolchain.

9 Where Process Really Does Matter

I have argued for judging the product, but "the product is the only thing that counts" would be too absolute. Sometimes the process is part of the product, and a defensible functional view must say when.

At one end are outcome-dominant artefacts: a technical manual, application programming interface (API) documentation, a reference book, a routine business analysis, or a tutorial. Readers mainly need the artefact to be accurate, clear, current, and useful. AI assistance is generally incidental if those standards and any contractual requirements are met.

At the other end are authorship-dominant artefacts: an unaided examination, a personal letter, testimony, autobiography, or art explicitly sold as the direct expression of a particular person. Here, provenance helps constitute the thing being offered. A love letter commissioned from a model might contain elegant sentences yet fail at the function the recipient reasonably expected: personal expression.

Between those poles lies most professional knowledge work. The right standard depends on the promise and the risk. Process legitimately matters when it bears on:

- **competence:** an assessment is intended to measure what the candidate can do unaided;
- **evidence:** scientific data, quotations, and citations must be traceable rather than invented;
- **rights:** plagiarism, licensing, confidentiality, and consent impose constraints on inputs and outputs;
- **safety:** medical, legal, financial, and engineering work demands qualified review and documented controls;
- **authenticity:** the buyer or recipient values a particular human act, voice, or history; and
- **accountability:** regulators, clients, and collaborators need to know who made and checked consequential decisions.

These exceptions do not revive a general cult of manual labour. They identify cases in which provenance affects fitness for purpose. A scientific paper with fabricated measurements is defective because the claimed evidence does not exist, not because software helped draft the prose. A model-generated summary of genuine results may be entirely acceptable if the researchers verify it, disclose use where required, and remain answerable for every claim.

High-stakes work deserves especially sober controls. The National Institute of Standards and Technology identifies "confabulation"—confidently presented false content—as a characteristic GenAI risk and recommends risk management across design, measurement, and governance [14]. This paper is an opinion essay, not financial, legal, or medical advice; readers should consult qualified professionals before acting on high-stakes claims. Functional evaluation in these fields includes the quality of the validation process. The process matters because it changes the reliability of the product.

10 The Strongest Objections

Any defence of AI-assisted work that ignores the strongest objections is too easy. I take four of them seriously.

10.1 "The models are unreliable"

They are. I have seen models fabricate references, lose constraints, produce insecure code, and render consensus in place of truth. That fact argues for verification proportional to risk. It does not imply that no excellent artefact can be made with them. Human collaborators are unreliable too; the difference is that machine errors can arrive with extraordinary speed, scale, and polish.

The right comparison is therefore not between a fallible machine and an infallible unaided author. It is between complete workflows. Which process—manual or AI-assisted—produces fewer errors after review? Which exposes assumptions, runs tests, and checks sources? A careful AI-assisted workflow can deploy more adversarial passes than a solitary author could ever afford; a careless one can mass-produce falsehood. The adjective doing the work is *careful*, not *AI-assisted*.

10.2 "AI makes everything sound the same"

Default model prose is often smooth, generic, symmetrical, and lifeless. I recognise it immediately, and accepting it unchanged is a failure of direction and editing. The same systems can compare voices, remove clichés, vary rhythm, preserve technical terminology, and revise against a detailed style specification. More fundamentally, distinctiveness comes from the author's questions, examples, experience, selection, and willingness to delete the merely adequate.

Cheap generation may indeed pull average public prose towards a bland centre. But the conclusion is not that all assisted prose must be bland. It is that taste becomes more important when grammatical competence becomes abundant.

10.3 "Using AI appropriates the labour of others"

The provenance of training data, consent, compensation, and market power raises serious legal and ethical questions. Those questions cannot be wished away by pointing to useful outputs. They require transparent licensing regimes, enforceable rights, and choices among tools with different data practices. They also vary by jurisdiction and are still being contested.

But a dispute about how a model was trained is not a reliable quality test for a particular technical book. It is a governance issue upstream of the artefact, much as the environmental and labour practices behind hardware are real issues without determining whether every document typed on that hardware is worthless. I see no reason to choose between the two levels: we should address both rather than use one as a substitute for the other.

10.4 "If production becomes effortless, the work cannot be valuable"

This objection conflates two claims: effort may be valuable to the person making it, and effort makes the resulting product more valuable to someone else. My experience in sport illustrates both sides. Sport has been part of my life since kindergarten, and for several years I ran marathons. I approached my first simply for fun: I went for occasional runs, followed no plan, and finished in a little over 3 hours and 25 minutes. Years later, aiming to break the three-hour barrier, I followed a strict 12-week training plan that reached 100 kilometres or more per week—an exceptionally demanding commitment for me alongside a full-time career with long working hours. I finished in 3:08:56. The improvement was real, but it was not the result I had trained for. The official record contains neither the casual preparation behind the first time nor the discipline and sacrifice behind the second. It records only the numbers.

Yet the finishing time is not the whole value of the experience. I enjoyed virtually every training run. This became especially clear during another marathon cycle: after a long period of preparation, illness prevented me from participating in the race at all. Even so, I never regretted the training, because the journey itself had been worthwhile. Personal effort can therefore be

valuable in its own right when the experience, discipline, or personal development is part of the goal. In that case, the process is part of the product—for the runner.

What does not follow is that this private value transfers automatically to somebody else. A reader may admire the work behind a book, just as a spectator may admire an athlete's discipline, but admiration for effort is not the same as judging the result. A clear explanation does not help its reader less merely because a better production system made it faster to create.

In the rest of modern life, we make this separation almost automatically. I do not expect the kettle in my kitchen, the components in my laptop, the screws in my furniture, or the appliances in my home to have been individually handcrafted. I expect them to work, to be safe, and to be affordable. Automated and standardised production has made many goods more consistent and available to far more people. We generally regard that efficiency gain as welfare-enhancing, not as evidence that the resulting goods lack value. Handcraft can still command a premium when provenance, individuality, or artistry is part of what the buyer wants. But it would be peculiar to demand artisanal manufacture from every power adapter before trusting it to charge a computer.

Digital products take the argument one step further. Once a song, application, or electronic book exists, another person can stream or download it at negligible marginal cost and with almost no additional human or machine labour. That does not make the next listen, installation, or reading worthless. The value lies in what the digital product does for its user, not in the effort required to reproduce one more copy.

Why should a useful technical artefact obey the opposite presumption? If automation helps produce an accurate manual, a reliable piece of software, or a clear educational text faster and at lower cost, the saved effort is a benefit unless it conceals some loss that matters—quality, safety, originality, accountability, or trust. Effort avoided is not value destroyed.

Efficiency does change markets. When supply explodes, many artefacts command lower prices and some established roles shrink. That disruption deserves attention and policy. Yet preserving needless effort is a poor welfare strategy. We do not protect the social value of accounting by outlawing spreadsheets. We decide which human capabilities remain important and redesign training, institutions, and business models around them.

11 The Apprenticeship Paradox

Here my functional defence meets its hardest problem, because "30 years plus two days" has an uncomfortable second half. The thousands of hours I spent writing, coding, debugging, deriving, failing, and rewriting did not merely produce books and programs. They produced me. They built the internal models that now allow me to direct AI, notice when it is wrong, and know when a polished draft still has no point.

An experienced professional can automate much of that labour because the professional already possesses the capability the labour helped create. What happens to a 20-year-old who automates it before acquiring the capability?

This is the apprenticeship paradox:

The apprenticeship paradox

AI has exceptional leverage for experts partly because they were trained in a world without it. Yet some of the work we now automate was the training mechanism through which those experts learned to judge the work.

The evidence increasingly supports this distinction between performance and learning. The Organisation for Economic Co-operation and Development (OECD) warns that general-purpose GenAI can improve immediate task performance without producing durable learning, especially when students offload the very cognitive work the exercise was designed to develop [15]. That

is exactly the distinction my own history makes vivid: a polished submission is not the same thing as an acquired capability.

I do not have a complete solution, but several principles follow. Education should preserve deliberate cognitive friction where the objective is formation rather than production. Students should sometimes solve before prompting, predict before revealing, draft before rewriting, and debug before delegating. They should be required to explain generated work, reproduce core reasoning without assistance, defend choices orally, identify planted errors, and modify solutions under new constraints.

Access to automation can then expand with demonstrated competence. A novice needs exercises that reveal the structure of a problem. An advanced student needs practice supervising systems that operate at realistic scale. Both need AI literacy: not just prompt tricks, but source checking, evaluation, uncertainty, privacy, intellectual property, and the discipline to know when not to delegate. Guidance from the United Nations Educational, Scientific and Cultural Organization similarly calls for human-centred, age-appropriate use and for institutions to validate GenAI pedagogically and ethically [16].

The analogy is flight training. Commercial aviation did not respond to autopilot by insisting that every journey be flown manually. Nor did it decide that understanding flight was obsolete. Training preserves manual and conceptual competence because automation can fail and because pilots must recognise the boundary of the system. Intellectual work now needs an equivalent pedagogy.

This is not nostalgia for difficulty. Difficulty should be retained when it develops a capability we continue to value, not as a ritual payment for legitimacy. The educational question is therefore more precise than "Should students use AI?" It is: *Which cognitive work must a learner perform, at which stage, to become capable of using automation responsibly later?*

12 What Remains Scarce

When text, code, images, and analysis become cheap to generate, intellectual value does not disappear. It migrates. The new scarcities are worthwhile questions, trusted judgement, distinctive experience, verified evidence, coherent synthesis, reputation, and accountability.

This should be welcome news for serious authors. We can spend less of life pushing symbols into acceptable first drafts and more of it deciding what deserves to exist. We can build explanations for narrower audiences, maintain technical books more frequently, translate ideas across disciplines, test more examples, and make high-quality materials available where old production economics would not justify them.

The opportunity is not a licence to flood the world. Because generation is cheap, restraint becomes part of authorship. I do not publish everything a model can produce; I filter, compress, verify, and ask whether the artefact repays the reader's finite attention. That restraint is now part of the author's job.

Readers, in turn, need stronger signals than "human-made." They should ask for evidence, inspect examples, run code, sample arguments, consult reputation, and distinguish disclosed assistance from abdicated responsibility. Publishers and institutions should specify when provenance is relevant, require proportionate disclosure, and evaluate outputs against clear standards. Blanket bans and blanket enthusiasm are equally lazy.

I have no desire to prove that I can still write a technical book as I did 20 years ago. I know that I can. The more important question is what I can contribute now, using the best tools available while preserving the judgement built over those decades. If I can provide delegates, clients, colleagues, and the public with clearer, more accurate, more ambitious work, deliberately choosing a slower process would not be intellectual virtue. It would confuse the road with the destination.

The future of intellectual work should not be organised around preserving yesterday's friction.

It should preserve what that friction once cultivated: expertise, curiosity, independence, taste, and responsibility. We should disclose material tool use, protect contexts in which authentic human performance is the point, and train novices to become more than operators of fluent machines. Beyond those requirements, the verdict belongs to the artefact.

I drove to the meeting. Now let us judge what happened in the room.

References

[1] Kruger, Justin, Derrick Wirtz, Leaf Van Boven, and T. William Altermatt (2004). "The Effort Heuristic." *Journal of Experimental Social Psychology*, 40(1), 91–98.

[2] Turing, Alan M. (1950). "Computing Machinery and Intelligence." *Mind*, 59(236), 433–460.

[3] OpenAI (2026). "An OpenAI Model Has Disproved a Central Conjecture in Discrete Geometry." May 20.

[4] OpenAI (2026). "Ten Advances in Mathematics and Theoretical Computer Science." August 1.

[5] Anthropic (2026). "When AI Builds Itself." Anthropic Institute.

[6] Cherny, Boris (2026). "Head of Claude Code: What Happens after Coding Is Solved." Interview by Lenny Rachitsky, *Lenny's Podcast*, February 19, especially 16:17–17:32 and 1:00:11–1:01:08.

[7] Alphabet (2025). "First Quarter 2025 Earnings Call." April 24.

[8] Lopopolo, Ryan (2026). "Harness Engineering: Leveraging Codex in an Agent-First World." OpenAI, February 11.

[9] Noy, Shakked, and Whitney Zhang (2023). "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence." *Science*, 381(6654), 187–192.

[10] Dell'Acqua, Fabrizio, Edward McFowland III, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine C. Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani (2026). "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality." *Organization Science*, articles in advance, 1–21.

[11] Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond (2025). "Generative AI at Work." *The Quarterly Journal of Economics*, 140(2), 889–942.

[12] Lee, Hao-Ping, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson (2025). "The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects from a Survey of Knowledge Workers." *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, Article 1121, 1–22.

[13] U.S. Copyright Office (2025). *Copyright and Artificial Intelligence, Part 2: Copyrightability*. Washington, DC.

[14] National Institute of Standards and Technology (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. NIST AI 600-1.

[15] Organisation for Economic Co-operation and Development (2026). *OECD Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education*. OECD Publishing, Paris.

[16] Miao, Fengchun, and Wayne Holmes (2023). *Guidance for Generative AI in Education and Research*. UNESCO, Paris.