

How AI Is Eliminating the Only People Who Can Tell When It's Wrong

We've created a technology that can convincingly explain quantum physics, draft a legal brief, or diagnose why your sourdough starter died. The catch? To know if any of it is actually correct, you need a quantum physicist, a lawyer, or an experienced baker. And we're busy automating all of them out of existence. Surely this plan has no flaws.

Welcome to the central paradox of the AI era: the more capable and convincing these systems become, the more we need human experts to validate their outputs. Yet, the economic and institutional pressures of our time are pushing us to replace those very experts with the AI systems that require their oversight. It's like burning all the maps right before setting out on a journey because your GPS is totally reliable. Until it isn't.

This article explores the most critical and surprising challenge of the AI era—the one we're overlooking while we fight more immediate fires. We will delve into the slow-motion crisis of eroding human expertise and the vicious cycle that threatens our ability to distinguish informed opinion from verified fact.

1. The Single Most Important Rule: Treat AI Output as an Opinion, Not a Fact

Before we go further, it's essential to understand the core principle that should guide every interaction with generative AI: its output is an "informed opinion," not a verified fact. While these opinions can be incredibly useful—often more informed than any single human's—they require the same skepticism and verification we would apply to a colleague's advice before making a critical decision.

The problem is that AI doesn't present its output like a human opinion. When a colleague says, "I think the market will shift in Q3," the words "I think" give it away. You know that's a prediction requiring your own analysis. An AI, however, generates a comprehensive market analysis with charts and confident assertions that feel like established fact. This authoritative presentation creates a dangerous illusion that its outputs are truths rather than perspectives that must be checked.

treat every output from generative AI as an opinion (albeit a very informed one!), never as a fact until you've verified it.

Everything that follows stems from one troubling reality: we're rapidly losing the human capacity to perform that verification, even as we need it more than ever. We're also building

economic and institutional pressures that push us to treat these informed opinions as facts, simply because verification is expensive and slow.

2. A Sci-Fi Story from 1958 Predicted Our Biggest AI Problem

In his short story "The Feeling of Power," Isaac Asimov imagined a future where humanity has delegated arithmetic to computers so completely that manual calculation becomes a lost art. When a technician rediscovers how to multiply without a machine, it's treated as a revolutionary breakthrough. Asimov's brilliant insight was not just about the loss of a skill, but the loss of the *recognition that the loss is a problem*. The computers worked perfectly, so no one missed the knowledge.

Our current situation is considerably more dangerous than Asimov's story for four key reasons:

- **The Complexity Problem:** AI synthesizes knowledge from domains too vast for a single human expert to fully verify. At what point does verification become impossible not because we've forgotten the knowledge, but because the knowledge itself has become too vast for human cognition?
- **The Speed Problem:** This expertise is not eroding over generations, but in a matter of years. The economic pressure to replace expensive human experts operates on quarterly timescales, while building new human expertise takes decades.
- **The Invisibility Problem:** A lost skill like arithmetic is obvious. A loss of judgment, however, is silent. It goes unnoticed until a crisis reveals that no one is left who can spot a subtle but catastrophic error. This is the autopilot problem writ large: when you rely on automated systems long enough, you lose the ability to take manual control in an emergency and may not realize you've lost it until the crisis hits.
- **The Scale Problem:** Asimov worried about one domain: mathematics. We are automating all domains of knowledge simultaneously. We're losing the editors who could spot subtle shifts in meaning, the teachers who could identify when an AI misses crucial context, the lawyers who could catch a missed precedent, and the doctors who could identify a dangerous assumption in a diagnosis.

3. We're Distracted by AI's Loud Problems and Missing the Quiet Catastrophe

The challenges of AI arrive in two forms: the short-term fires we are fighting right now (1-3 years) and the long-term crises that are quietly gathering strength (4+ years). Our attention is consumed by the immediate, loud problems: deepfakes threatening elections, rapid job displacement, algorithmic bias, and the urgent need for regulation.

While these issues demand attention, they distract us from a more insidious, long-term threat. When we're frantically putting out fires, we're not building the boring infrastructure we'll desperately need later: things like robust validation systems, preserved human expertise, and meaningful oversight. We're ignoring the check engine light because we're late for a meeting.

The truly dangerous challenge is the slow-motion catastrophe of eroding human capacity to tell when AI is wrong. This gradual decay is a kind of **"climate change for human expertise."** It doesn't announce itself with a single dramatic event. Instead, it's a creeping, systemic degradation of our collective knowledge. By the time the crisis becomes obvious to everyone, the ice caps of human expertise may have already melted beyond recovery.

4. We're Trapped in a Vicious Cycle That's Erasing Human Knowledge

The erosion of expertise isn't a simple, linear decline; it's a reinforcing feedback loop that accelerates over time. This vicious cycle traps organizations and entire fields in a state of growing dependency and shrinking capability.

Here is how the cycle works:

1. Human experts become scarce and expensive, so organizations rely more on AI to fill the gap.
2. As AI becomes embedded in daily workflows, fewer people develop deep expertise through hands-on practice.
3. As human expertise declines, our collective ability to validate AI's "informed opinions" diminishes.
4. As our validation capacity shrinks, we are forced to treat AI outputs as trustworthy facts because there is no one left qualified to challenge them.
5. This increased trust leads to even less investment in human expertise, which makes experts even scarcer, and the cycle repeats with greater intensity.

Each turn of this wheel leads to "institutional lock-in" and terrifying downstream consequences. Knowledge dependency deepens, creating a "Chinese whispers" effect where future generations learn from AI systems that learned from other AI systems. Authority shifts accelerate, leading to circular validation problems where AI systems are checked by other AI systems, supervised by AI systems, with the last human expert having retired to raise alpacas in Vermont.

Conclusion: Choosing Our Future

The challenges ahead are not primarily technical; they are cognitive and institutional. The fundamental choice is whether we can maintain the cognitive habit of verification in the face of a technology designed for frictionless acceptance. This isn't about halting progress; it's about getting wiser as our tools get more powerful.

Breaking the cycle requires deliberate, strategic action. We must treat certain forms of human expertise as strategic assets to be preserved, like a seed bank for knowledge. We must redesign education to emphasize critical thinking, contextual judgment, and the wisdom to ask, "How would I verify this?" And we must build organizations that reward human oversight and design AI to augment our experts, not replace them.

As we navigate this new landscape, we must confront a critical question: *What knowledge is too important to completely outsource to a machine?* Because the alternative—a world where powerful AI systems offer confident assertions that no one is qualified to verify—is a science fiction scenario that should remain firmly in fiction, not become our strategic plan.

Of course, the ultimate irony is that this article was written with the assistance of my generative AI pal, Claude.ai. Welcome to the new world.