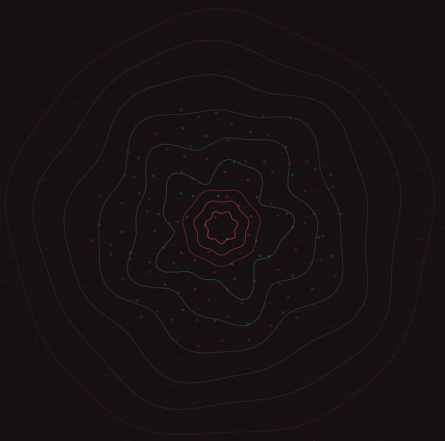


A STATISTICAL SATIRE

THE HALLUCINATION TAX

An Invoice for Confident Wrongness

NABIL EL MAHYAOUI

The Hallucination Tax

The Hallucination Tax

An Invoice for Confident Wrongness

Nabil El Mahyaoui

A statistical satire

Copyright © 2026 Nabil El Mahyaoui.
All rights reserved.

Contents

1 The Oldest Profession **1**

1.1 The Architecture of Credentialed Wrongness 3

1.2 The Replication Funeral 4

1.3 Clinical Judgment 6

2 The Complicity Audit **8**

2.1 The Green Status 8

2.2 The Gut Feeling 10

2.3 The Résumé 11

2.4 The Performance Theology 13

3 The Democratization of Bullshit **16**

3.1 The Marginal Cost Collapse 18

3.2 The Real Grievance 18

3.3 The Pricing Fix 20

4 The Latency of Accountability 23

4.1 The Thirty-Second Window 25

4.2 The Accountability Stack 26

4.3 The Unforgivable Efficiency 28

5 The Tax on the Tax 31

5.1 The Governance Stack 32

5.2 The Trust Industry 33

5.3 The Purchase. 35

6 Epilogue: The Invoice 38

References 45

1

The Oldest Profession

There is a room in lower Manhattan—glass walls, Herman Miller chairs, a view of the Hudson that costs more per square foot than most people earn in a month—where, in the autumn of 2006, a committee of analysts at Moody's Investors Service assigned a triple-A rating to a synthetic collateralised debt obligation composed almost entirely of subprime mortgage tranches. The fee for this service was approximately $200,000. The committee deliberated for less than a day. The models were consulted. The spreadsheets were reviewed. The rating was issued with the quiet confidence of a physician signing a death certificate for someone who is, technically, still breathing.

Eighteen months later, the instruments they had blessed would be worth pennies on the dollar, and the global financial system would lose, by Federal Reserve estimates, $22 trillion in economic output. Twenty-two trillion. If you stacked that in hundred-dollar bills, the pile would reach the moon and back—twice—which is roughly where the analysts' credibility had gone.

Nobody at Moody's went to prison. The models were not indicted. The spreadsheets were not cross-examined. The rating methodology was reviewed, adjusted, and redeployed, because the alternative—admitting that the entire enterprise of quantified certainty was, at its foundation, a confidence game—would have been worse than the crisis itself.

This is the thing you need to understand about hallucination: it is not a bug. It has never been a bug. It is the oldest, most profitable, most institutionally protected feature of the knowledge economy.

1.1 The Architecture of Credentialed Wrongness

McKinsey & Company charged Purdue Pharma somewhere north of $80 million for consulting work that included, among other deliverables, strategies to "turbocharge" OxyContin sales during the worst drug epidemic in American history. The resulting crisis has killed over 500,000 people. The slides, one imagines, were beautifully formatted. The frameworks were rigorous. The 2×2 matrices were impeccable. The confidence was total, which is the point, because confidence *was* the deliverable.

Not accuracy. Not truth. Not even plausibility, strictly speaking. *Confidence.*

The consulting industry—$350 billion globally as of 2024—is, if you squint at it with the slightest intellectual honesty, a hallucination-as-a-service platform. The product is a PowerPoint deck that allows a senior executive to say "the consultants recommended it" when the strategy fails. The hallucination is not incidental to the engagement. The hallucination *is* the engagement. The invoice is the proof that it was sanc-

tioned by someone with a credential, and the credential is the proof that the hallucination was produced by an approved process, and the process is the proof that no individual human being needs to be held responsible for the outcome.

This is not corruption. Corruption would be simpler, and cheaper. This is an *architecture*—a load-bearing structure of distributed confidence in which every participant is locally rational, locally competent, and collectively producing nonsense at industrial scale.

1.2 The Replication Funeral

In 2015, a consortium of 270 researchers attempted to replicate 100 psychology studies published in top-tier journals. They succeeded with 36. Thirty-six out of a hundred. The other sixty-four—peer-reviewed, cited, taught in universities, used to inform public policy, referenced in parenting books and corporate training programmes and TED talks watched by millions—were, in the technical language of the field, "not reproducible," which is the scientific community's way of saying *wrong* without saying wrong.

The studies had not been fabricated. The researchers had not committed fraud. The peer reviewers had not been negligent, at least not by the standards of peer review, which are approximately the standards of a bouncer at a nightclub who checks that you have *an* ID, not that it is *yours*. What had happened was simpler and worse: the system was optimised for confident publication, not for truth, and it produced exactly what it was optimised to produce. Sixty-four hallucinations in a hundred, dressed in the formal attire of statistical significance, $p < 0.05$, which is the academic equivalent of a triple-A rating.

The replication crisis did not destroy psychology. It did not even slow the pace of publication. What it destroyed was the polite fiction that peer review is a truth-verification system rather than a confidence-laundering system. The papers kept coming. The citations kept accumulating. The hallucinations kept compounding, each one citing the last, building a cathedral of mutual reference in which every stone is load-bearing and none of them are real.

1.3 Clinical Judgment

A doctor who says "I don't know" gets sued for negligence. A doctor who confidently misdiagnoses you gets sued for malpractice only if you can prove it, which you often cannot, because the confident diagnosis was documented in a system designed by and for confident people, reviewed by confident peers, and defended by confident lawyers citing confident precedent.

The *BMJ Quality & Safety* journal estimated in 2023 that diagnostic error contributes to approximately 371,000 serious harms and 100,000 deaths annually in the United States alone. One hundred thousand. That is a mid-sized city, dying every year, from confidence. Nobody calls these hallucinations. We call them clinical judgment.

Arthur Andersen LLP—the accounting firm that audited Enron's books with the serene professionalism of a sommelier recommending wine at a restaurant that is on fire—signed off on financial statements that overstated the company's value by approximately $74 billion. The firm employed 85,000 people. It had ex-

isted for 89 years. It was destroyed not because it was incompetent, but because the gap between its confidence and reality briefly became visible to people who were paying attention.

This is the pattern. Not fraud. Not stupidity. Not even negligence, in the legal sense. Just confidence, applied systematically, at scale, by credentialed professionals operating within approved frameworks, producing outputs that look exactly like truth and function exactly like truth and are, when you finally hold them up to the light, made entirely of conviction.

The hallucination tax is not new. It is the oldest tax in the knowledge economy. It predates income tax, sales tax, and capital gains. It is levied on every transaction in which someone who does not know presents themselves as someone who does, and the recipient pays—in money, in health, in years, in trust—for the privilege of not having to confront the void.

You have been paying it your entire life. You just never saw the line item.

2

The Complicity Audit

But you already knew all of this. You've read the articles about Moody's. You've shared the outrage about McKinsey. You've shaken your head at the replication crisis and clucked your tongue at Arthur Andersen and felt, in the warm glow of your own critical thinking, that you are not like them. You are the audience. You are the one who sees through it.

You are not the audience.

2.1 The Green Status

Last Tuesday—or maybe it was Wednesday; the days have a way of blurring when you spend them inside a machine that converts calendar invites into anxiety—

you wrote an email to your skip-level manager describing the project as "on track." It was not on track. It was three weeks behind, a senior developer had quit, the remaining team was burning through goodwill at a rate that would have alarmed a forensic accountant, and the integration you had promised for Q2 was, in any honest assessment, a Q3 deliverable at best. You knew this. Your manager knew this. Your manager's manager, the one who would read the email with the performative frown of someone who has seen many status reports and believes none of them, also knew this.

You updated the status to green.

Not because you are dishonest. You are not dishonest. You are something more interesting than dishonest: you are *fluent*. You have learned, over a career of Jira tickets and quarterly reviews and all-hands meetings where the CEO says "transparency" with the earnestness of someone who has never experienced it, that the system does not want truth. The system wants a confident signal that allows everyone in the chain to continue doing what they were already going to do. Yellow would have triggered a meeting. The meeting

would have produced an action item. The action item would have been assigned to someone who was already overloaded. Nothing would have changed except the total quantity of hallucination in the system, which would have increased by exactly one meeting's worth.

You chose green because green is cheaper. Not for the project. For you.

2.2 The Gut Feeling

You interviewed a candidate last month. Forty-five minutes. You asked about system design and got a plausible answer that you could not fully evaluate because the system in question was not one you had built. You asked about a time they had failed, and they told a story so well-structured it had clearly been rehearsed, which you interpreted as preparation rather than performance, because you prefer to believe that the world is full of articulate, self-aware professionals and not full of people who have read the same Medium article about behavioural interviews that you read in 2019.

You gave them a "strong hire."

Meta-analyses of unstructured interviews—the kind you just conducted—put their predictive validity at $r = 0.20$. This is roughly the correlation between shoe size and height. It is not zero, which is the number that would justify abandoning interviews altogether, but it is close enough to zero that the primary function of the interview is not prediction but *ritual*—a ceremony of mutual hallucination in which the interviewer pretends to assess and the candidate pretends to be assessed and both parties leave feeling that something meaningful has occurred.

You hired confidence. You always hire confidence. The candidate who hesitated, who said "I'd need to think about that," who admitted uncertainty—you marked them as "mixed signal" in the feedback form, which is the system's euphemism for "honest."

2.3 The Résumé

Your own résumé is a hallucination. Not a lie—you are careful about that distinction, the way a sommelier is careful about the distinction between "fruity" and "fruit-forward"—but a hallucination: a confident, flu-

ent, structurally plausible document that implies a narrative of intentional career progression where the actual trajectory was largely accidental, mostly lucky, and occasionally the result of being in a room when a decision was made by someone who had confused your presence with your contribution.

You "led cross-functional initiatives." You were in a meeting. You "drove revenue growth." Revenue grew. You were employed at the time. Correlation, as they say in the hallucination business, is not causation, but it *is* a bullet point.

The model you are running—the one inside your skull, the one trained on twenty or thirty or forty years of professional experience, the one that generates contextually appropriate responses to ambiguous prompts with remarkable fluency and occasional confabulation—is performing exactly the same operation as the one everyone is so worried about. Pattern-matching on training data. Generating probabilistically plausible output. Presenting it with the formatting conventions of authority.

The difference is that you charge more. The difference is that when you hallucinate—when you present

a confident answer that is structurally plausible and factually hollow—nobody screenshots it and posts it on Twitter with a caption about how you are going to destroy civilisation. They promote you. Or, at minimum, they do not fire you, which in the modern economy amounts to the same thing.

2.4 The Performance Theology

Every performance review you have ever written—about yourself or about someone else—is a hallucination. "Exceeds expectations" means "I like this person, or I need something from them, or I do not have the energy for the conversation that would follow a lower rating." "Meets expectations" means "I have no opinion, and this rating is the path of least resistance." "Needs improvement" means "I have already decided to manage them out, and this is the paper trail that HR requires before I can do what I was going to do anyway."

The entire performance management industry—the forms, the calibration sessions, the forced distribution curves, the nine-box grids, the 360-degree feedback in-

struments that produce 360 degrees of noise and call it insight—exists to give the appearance of measurement to what is, at bottom, vibes.

Not even good vibes. *Sanctioned* vibes. Vibes that have been processed through a framework and emerged on the other side wearing a number, and the number gives everyone permission to believe that the decision was data-driven, because "data-driven" is the hallucination that the business class finds most comforting.

The OKR system—Objectives and Key Results, the gospel according to Google, now practised with varying degrees of sincerity by every company large enough to have an HR department—is a hallucination nested inside a hallucination. You set objectives that are ambitious enough to signal vision but achievable enough to signal competence. You define key results that are measurable enough to satisfy the framework but vague enough to be declared complete under a range of outcomes. At the end of the quarter, you write a self-assessment that describes what happened in the language of what was supposed to happen, and your manager reads it and nods, and the system records that ob-

jectives were "substantially met," and everyone moves on to the next quarter, and the cycle begins again, and at no point does anyone ask whether the objectives were *correct*, because that question would break the machine.

You are not paying the hallucination tax. You are *collecting* it. Every status report, every interview scorecard, every performance review, every quarterly business update—these are not failures of honesty. They are the revenue stream. You are a node in the network. You are an inference engine running on biological hardware, trained on corporate data, producing outputs that are indistinguishable from truth and serve the same economic function.

The only difference between you and the large language model is that you believe your own outputs. The model does not. Make of that what you will.

3
The Democratization of Bullshit

For the entire history of professional services—and professional services are, when you strip away the mahogany and the diplomas and the reassuring heft of a leather-bound report, nothing more than the organised sale of confident answers—hallucination was *expensive*.

It required credentials. You needed a medical degree to confidently misdiagnose, and the degree took eight years and $300,000, and those numbers were not incidental; they were the *moat*. You needed a J.D. to confidently misinterpret precedent, and the J.D. required three years of learning to write sentences that

no human being would voluntarily read, which is itself a form of hazing that the profession mistakes for training. You needed an MBA and a consulting engagement letter and a team of associates who had been awake for thirty-six hours to confidently misproject revenue, and the engagement letter cost $2 million before anyone opened a spreadsheet.

The barriers to entry were enormous. The margins were spectacular. And the product—confident wrongness, delivered with the production values of a Broadway show—was so deeply woven into the fabric of institutional decision-making that questioning it was not merely impolite but structurally impossible, the way questioning gravity is structurally impossible for someone who is currently falling.

Then a research lab in San Francisco trained a neural network on the internet, and the moat evaporated overnight.

3.1 The Marginal Cost Collapse

What GPT-4 did was not invent hallucination. What GPT-4 did was collapse the marginal cost of hallucination to zero.

Suddenly, anyone—a teenager in Jakarta, a retiree in Tucson, a middle manager in Düsseldorf who had never in his life written a strategy memo longer than three paragraphs—could generate a confident, fluent, structurally plausible answer to any question in any domain for the price of a few cents in compute. The output was indistinguishable—in format, in tone, in the particular cadence of authority that educated professionals spend decades learning to perform—from the output of the credentialed professionals who had been doing the same thing for decades at $500 an hour.

And this—not the hallucination itself, but the *price*—is what caused the panic.

3.2 The Real Grievance

The lawyers were not upset that ChatGPT cited fake cases. Lawyers cite inapplicable cases constantly; they

call it "creative argumentation," and the best lawyers—the ones who bill $1,500 an hour at firms with names that sound like Victorian diseases—are the ones who cite the most aggressively inapplicable cases with the most serene confidence. The lawyers were upset that ChatGPT cited fake cases *for free*.

The consultants were not upset that AI generated plausible-sounding strategic recommendations without understanding the business. That is *literally* the consulting business model for a twenty-six-year-old associate who has been on the engagement for two weeks and whose understanding of the client's industry comes entirely from a deck that another twenty-six-year-old wrote three years ago for a different client in a different industry. The consultants were upset that the AI did it without billing travel expenses.

The doctors were not upset that AI occasionally got the diagnosis wrong. Doctors get the diagnosis wrong 10–15% of the time, depending on the specialty and the day of the week and whether the patient presented their symptoms in the order the textbook says they should, which patients, being human and therefore unreliable narrators of their own biology, almost never

do. The doctors were upset that patients might start *noticing* the base rate.

Because here is the secret that every profession guards with the ferocity of a dragon sitting on a hoard of gold: the base rate of professional error is not low. It has never been low. It is merely *invisible*, hidden behind credentials and jargon and the asymmetry of expertise that makes it impossible for the client to evaluate the product they are purchasing until it is far too late to return it.

The AI made the base rate visible. Not by being worse than the professionals—in many domains, it is already better—but by being *comparable at zero cost*, which invites comparison, and comparison invites measurement, and measurement is the one thing that a hallucination economy cannot survive.

3.3 The Pricing Fix

What AI disrupted was not the truth. The truth was never in the room. The truth was outside, smoking a cigarette, waiting for someone to ask, which nobody

ever does because the truth does not bill by the hour and therefore has no institutional advocate.

What AI disrupted was the *pricing model for hallucination*.

And when you disrupt the pricing model of a $350 billion industry—when you take a product that was sold at luxury margins behind credential barriers and make it available at commodity prices with no barriers at all—the industry does not respond by improving accuracy. The industry does not respond by eliminating hallucination. The industry responds the way every industry responds when its margins are threatened: it *rebrands*.

The lawyers are not becoming more careful. They are offering "AI-augmented legal services" at a premium. The consultants are not becoming more honest. They are selling "AI strategy" engagements that cost more than the engagements they replace. The doctors are not becoming more accurate. They are adding "AI-assisted diagnostics" to the bill, which means the same doctor who was wrong 12% of the time is now wrong 8% of the time and charges 30% more for the privilege.

The industry fixes the pricing. It always fixes the pricing. The hallucination continues, at higher resolution, at greater scale, with better production values. The tax goes up.

4
The Latency of Accountability

In the autumn of 2003, the United States government presented to the United Nations Security Council a dossier of evidence that Iraq possessed weapons of mass destruction. The evidence was detailed. The satellite photographs were annotated. The intelligence assessments were classified, which lent them the particular authority that secrecy confers—the authority of the thing you are not allowed to question because you are not allowed to see the full picture, and the full picture, you are assured, would make everything clear.

Secretary of State Colin Powell held up a vial of white powder. The room was silent. The hallucination was total.

It took three years, 4,500 American lives, an estimated 200,000 Iraqi civilian deaths, and $2 trillion in direct costs before the hallucination was officially acknowledged. Three years. By then, the architects of the hallucination had been promoted, retired, or moved to the private sector, where they were hired as consultants and advisors, because the market for confident wrongness does not penalise former practitioners. It *recruits* them.

This is the temporal structure of human hallucination. The confidence is instant. The accountability is glacial. And in the gap between the two—in that long, profitable, reputation-preserving interval—entire careers are built, entire industries are sustained, entire wars are fought and funded and memorialised before anyone is in a position to say, with the evidence finally assembled and the bodies finally counted: *that was wrong*.

4.1 The Thirty-Second Window

Now consider what happens when a large language model hallucinates.

In March 2023, a New York attorney named Steven Schwartz filed a legal brief containing six case citations that did not exist. The cases had been generated by ChatGPT. They were plausible—correct formatting, realistic party names, citations that looked like citations—but they were, in the technical language of the legal profession, *fabricated*, which is a word lawyers reserve for other people's lies.

The story was on Twitter within hours. It was in the *New York Times* within days. Schwartz was sanctioned. His career was, in practical terms, over. The hallucination-to-accountability pipeline was approximately thirty seconds long.

Compare this to the professional hallucinations that preceded it.

The credit rating agencies hallucinated triple-A ratings for a decade before anyone noticed. The consulting firms hallucinated strategic value for half a century before the first serious audit. The medical profes-

sion hallucinated diagnostic confidence for the entire history of medicine before the first replication study. The legal profession—the very profession that sanctioned Schwartz—has been hallucinating creative arguments, inflating billable hours, and citing precedent of dubious relevance since the invention of the billable hour in 1958, and no partner at a white-shoe firm has ever been sanctioned for filing a brief that was merely *wrong*, as opposed to *fabricated by a machine*.

The asymmetry is not about accuracy. It is about *latency*.

4.2 The Accountability Stack

Human hallucination has always been protected by a stack of latency buffers, each one adding months or years between the confident output and the moment of reckoning.

The first buffer is *complexity*. The CDO that Moody's rated triple-A was a synthetic instrument referencing tranches of mortgage-backed securities that were themselves composed of thousands of individual loans across dozens of states with different regulatory frame-

works. By the time you unwound the structure far enough to identify the hallucination, the analysts who built it had moved to a different desk, the desk had been reorganised, and the regulatory framework had been revised to prevent the specific failure that had already occurred, which is the government's way of closing the barn door after the horse has not only left but incorporated offshore.

The second buffer is *diffusion of responsibility*. Nobody at Moody's "decided" to hallucinate. The model produced an output. The analyst reviewed the output. The committee approved the analyst's review. The managing director signed the committee's approval. At each step, the individual was locally rational, locally competent, and locally not responsible for the aggregate outcome. The hallucination was distributed across so many nodes that it belonged to no one, and what belongs to no one cannot be punished.

The third buffer is *institutional inertia*. By the time the hallucination is exposed, the institution has changed leadership, changed strategy, changed its name. Arthur Andersen became Accenture became a company that no one associates with the largest

accounting fraud in history, which is a remarkable achievement of rebranding, and rebranding is, of course, just another word for institutional hallucination.

AI has none of these buffers.

The output is immediate. The attribution is singular: the model said it, the user used it, the chain of causation is three links long. The evidence is permanent: a screenshot, a log file, a cached response that cannot be revised or reclassified or buried in a committee report. The accountability window is not years. It is not months. It is the time between the hallucination and the first person who bothers to check.

4.3 The Unforgivable Efficiency

This is AI's real offence. Not that it hallucinates. Not that it is wrong. Not even that it is wrong with confidence, which is, as we have established, the foundational business model of every credentialed profession since the invention of credentials.

AI's offence is that it *hallucinates too fast to be forgiven*.

Human hallucination operates on a timescale that allows for the natural decay of accountability. By the time the error surfaces, the context has shifted, the people have moved on, the narrative has been rewritten. Nobody holds a grudge against the specific Moody's analyst who rated the CDO. Nobody remembers the name of the McKinsey partner who approved the OxyContin deck. Nobody prosecutes the doctor who misdiagnosed your father, because by the time you understand what happened, the statute of limitations has expired and the doctor has retired to a house that the hallucination paid for.

The system requires this latency. It is not a flaw in the system. It *is* the system. Without the gap between confidence and consequence, the entire architecture of professional services—the credentials, the frameworks, the peer reviews, the audits, the ratings, the certifications—would collapse under the weight of its own error rate, the way a building collapses when you remove the load-bearing walls, and the load-bearing walls of the knowledge economy are not truth.

They are delay.

AI removes the delay. And the edifice shudders.

5
The Tax on the Tax

Watch carefully what happens next, because it is already happening, and it is happening with the precise, mechanical elegance of a market that has identified a new anxiety and is moving, with the efficiency that only capital can muster, to monetise it.

The same institutions that have been selling you hallucination for decades are now selling you *hallucination insurance*.

AI safety. AI governance. AI trust frameworks. AI ethics boards. Retrieval-augmented generation. Guardrails. Human-in-the-loop verification. Responsible AI certifications. A $150 billion market by 2028, according to projections by Gartner, which is an analyst firm that sells projections, which means the $150

billion figure is itself a projection about the market for verifying projections, and if you can feel the recursion tightening around your throat, that is the correct sensation.

5.1 The Governance Stack

Deloitte will sell you an AI governance framework. They will send a team—four to six consultants, mostly under thirty, mostly hired within the last two years, mostly trained on a methodology that Deloitte developed by asking a different team of consultants what AI governance should look like—and this team will spend eight to twelve weeks interviewing your stakeholders, mapping your processes, and producing a 200-page report that recommends, among other things, that you establish an AI governance framework.

Deloitte will also sell you the AI. Deloitte will also sell you the audit of the AI. Deloitte will also sell you the consulting engagement to interpret the results of the audit. Deloitte will also sell you the training programme to help your employees understand the gover-

nance framework that governs the AI that was audited by the auditors who were trained by the consultants who wrote the framework.

At no point in this stack does anyone's incentive align with the elimination of hallucination. The elimination of hallucination would eliminate the stack. The stack employs 415,000 people at Deloitte alone. Those people have mortgages. The mortgages are real even if the governance is not.

This is the tax on the tax. This is the derivative instrument. This is the CDO-squared of bullshit: a structured product built from tranches of confident wrongness, rated by confident wrong raters, insured by confident wrong insurers, and sold to you with a governance framework that makes the whole thing feel *responsible*.

5.2 The Trust Industry

The AI trust industry is growing at 34% annually, which is the growth rate of an industry that has identified a fear and is feeding it. Every major consulting firm now has an AI ethics practice. Every major law

firm now has an AI liability team. Every major accounting firm now offers AI audit services, which is magnificent if you think about it, because the accounting firms' track record of auditing *human* outputs includes Enron, WorldCom, Tyco, Lehman Brothers, and Wirecard, and the natural response to this track record is apparently not humility but *expansion into new markets*.

The "AI-verified" label is emerging as a premium marker, the way "organic" emerged in food and "certified pre-owned" emerged in automobiles. The label means that a human—often a contractor, often paid by the unit, often working under time pressure that makes thoughtful evaluation structurally impossible—has reviewed the AI's output and determined that it is not hallucinating. The determination is itself a hallucination, because the human reviewer is subject to automation bias (the documented tendency to trust machine output more than one's own judgment), confirmation bias (the tendency to see what you expect to see), and the fundamental economic incentive to approve quickly and move to the next unit,

because the reviewer is paid per review and the rent is due on Thursday.

The human in the loop is not a safety mechanism. The human in the loop is a *liability buffer*—a warm body placed between the AI and the consequence, the way a company places a subcontractor between itself and a lawsuit. The human absorbs the accountability. The AI continues to operate. The governance framework records that a human was consulted. The audit confirms that the framework was followed. The circle closes. The hallucination persists. The tax increases.

5.3 The Purchase

You will buy it. Of course you will buy it. You bought the first hallucination because you needed it—because the alternative to a confident wrong answer was an honest "I don't know," and "I don't know" does not ship product, does not close deals, does not get funded, does not get promoted, does not survive the quarterly business review, does not satisfy the board, does not reassure the investors, does not pass the audit, does not meet the regulatory requirement, does not check the

compliance box. "I don't know" is the most expensive sentence in the English language. It is more expensive than being wrong, which is the central, load-bearing insight of the hallucination economy, and it is the insight that no governance framework will ever address, because the governance framework is itself a mechanism for avoiding "I don't know."

You will buy the second hallucination for the same reason you bought the first. And the third. And the fourth.

And somewhere in the stack—between the AI that generates the answer, the retrieval system that grounds it in documents that may or may not be correct, the governance layer that validates the retrieval, the human-in-the-loop who rubber-stamps the governance, the manager who reports the human's approval, and the executive who presents the whole apparatus to the board as "AI-verified insight"—the truth will be sitting quietly in a corner, unconsulted, unbilled, and increasingly irrelevant.

Not because anyone is lying. No one is lying. That is the darkest part.

Everyone in the chain believes, with varying degrees of sincerity, that they are adding value. The model believes nothing, which makes it the most honest participant in the room. The analyst believes the model. The manager believes the analyst. The VP believes the manager. The board believes the VP. The investors believe the board, because the alternative is to not believe, and not believing means not investing, and not investing means missing the market, and missing the market is the one hallucination no one can afford.

6

Epilogue: The Invoice

STATEMENT OF ACCOUNT

HALLUCINATION TAX AUTHORITY
Office of Assessed Confidence
Bureau of Plausible Outputs
Department of Structurally Sound Nonsense

"We bill, therefore we are."

TAXPAYER: You
FILING STATUS: Complicit (Standard)
TAX PERIOD: Birth to present
ASSESSMENT DATE: Ongoing

Schedule A: Education

Four-year degree from an accredited institution. Tuition: $187,000. Estimated hallucination content of curriculum: 38%, based on replication rates in assigned readings. Assessed value of credential as signal (independent of knowledge): $187,000. Assessed value of credential as knowledge: pending audit. Likely write-off.

Tax assessed: $71,060

Schedule B: Employment

Seventeen years of full-time professional employment. 4,420 status reports filed, of which an estimated 3,100 contained at least one material misrepresentation of project health, timeline, or resource adequacy. 212 performance reviews written, averaging 340 words of assessable hallucination per review. 89 interview scorecards submitted, predictive validity: $r = 0.20$. *Total billable*

hours of institutional confidence produced: 34,000. Market rate for comparable hallucination (human): $185/hr. Market rate for comparable hallucination (machine): $0.003/hr.

Tax assessed: $6,290,000
Depreciation (due to AI substitutability): 99.998%
Adjusted tax: $125.80

Schedule C: Personal

One résumé, updated fourteen times, each revision increasing the ratio of narrative coherence to factual accuracy. Three cover letters describing "passion" for industries the applicant could not identify from a lineup. 2,400 social media posts projecting a curated version of lived experience, engagement-optimised, selectively candid. One dating profile describing the applicant as "adventurous," evidence for which includes a single trip to Iceland in 2019

during which the applicant mostly stayed in the hotel because it was raining.

Tax assessed: $14,200
Note: This amount is non-deductible. Everyone pays it. No one declares it.

Schedule D: Consumption

Estimated lifetime consumption of hallucinated outputs from credentialed professionals: 1,847 instances. Including: 23 medical consultations in which the physician's confidence exceeded the diagnostic evidence. 4 legal opinions for which the billing reflected the attorney's certainty, not the case's merit. 7 financial projections from advisors whose models had a backtested accuracy of 41%, which is worse than a coin flip but better than a dartboard, and considerably worse than a parrot trained to peck at index funds. 1 home appraisal that valued the property at 22% above what any willing buyer would subsequently pay, a discrepancy

the appraiser described as "within normal market variance."

Tax assessed: $340,000
Note: The taxpayer is reminded that this amount was paid voluntarily, in exchange for the comfort of not being told "I don't know" by a person in a suit.

Schedule E: Systemic

The taxpayer's pro rata share of collective hallucination costs, including but not limited to: the 2008 financial crisis ($22 trillion, allocated across 330 million taxpayers); the opioid epidemic ($1.5 trillion in economic costs, allocated across affected communities); the replication crisis in social science (incalculable, but conservatively estimated at several hundred billion in misallocated policy and corporate training spending); the Iraq War ($2 trillion in direct costs, not including the other kind of costs, which do not fit on an invoice and should not be reduced to a number but just were, which is it-

self a hallucination, and the taxpayer is invited to sit with that discomfort).

Tax assessed: $78,412.16
Note: This figure is precise. It is not accurate. The taxpayer will recognise the distinction.

TOTAL ASSESSED: $504,797.96

TOTAL PAID TO DATE: $504,797.96

BALANCE DUE: $0.00

The balance is zero because you have already paid. You paid in misdiagnoses and misjudgments and misallocated years. You paid in wars built on intelligence that wasn't. You paid in credentials that signified confidence rather than competence, and in systems that rewarded fluency over accuracy, and in the slow, compounding interest of a civilisation that learned, long before it built a machine to do it, that the plausible answer is almost always more profitable than the true one.

The machine learned it from you.

It just learned it faster.

And it sends the invoice on time, which is more than can be said for most of the professionals it is replacing.

This statement is generated automatically. No human was in the loop.
For disputes, please contact a credentialed professional at $500/hr.

Accuracy not guaranteed. Confidence included at no extra charge.

References & Further Reckoning

Financial Crisis Inquiry Commission (2011). *Final Report.*
The official autopsy of the 2008 crisis. The credit rating agencies emerge with their reputations technically intact, which is itself a hallucination.

Moody's Corporation (2006). *Annual Report.*
Revenue from structured finance ratings peaked the year before the instruments they rated became worthless. Timing is everything.

Raman, A. (2022). "How McKinsey Got Into the Business of Addiction." *The New York Times.*
A detailed account of McKinsey's work with Purdue Pharma. The engagement fees were real. The framework was confident. The bodies were also real.

Newman-Toker, D.E. et al. (2019). "Serious misdiagnosis-related harms in malpractice claims." *Diagnosis*.
The study estimating 371,000 serious harms and 100,000 deaths annually from diagnostic error in the US. Not hallucinations, of course. Clinical judgment.

Schmidt, F.L. & Hunter, J.E. (1998). "The Validity and Utility of Selection Methods." *Psychological Bulletin*.
The meta-analysis showing unstructured interviews have a predictive validity of $r = 0.20$. *Still the most popular hiring method. Confidence sells.*

Grand View Research (2024). "Management Consulting Market Size Report."
The $350 billion figure. Produced, naturally, by a consulting firm. Citing it here without irony would be paying the tax.

Gartner (2024). "AI Governance and Trust Market Forecast."
The $150 billion projection for AI safety and governance tools by 2028. A forecast about the market for verifying forecasts. The recursion is the product.

Matavić, D. et al. (2024). "Large language models surpass human experts in predicting neuroscience results." *Nature Human Behaviour*.

AI outperforms domain experts on factual prediction tasks. The experts were more confident, though. That used to be worth something.

About the Author

Nabil El Mahyaoui

https://nabil-el-mahyaoui.com/

https://www.linkedin.com/in/elmahyaoui