

# Not Everyone Is Built the Same — Neither Is AI

## *A Practical Guide to Using AI's Most Powerful Tools Well*

Jonathan Slavic | Customer Solutions & Analytics Architect

### EXECUTIVE SUMMARY

*AI tools are among the most powerful productivity instruments available today. Used well, they compress hours of work into minutes, surface patterns invisible to the human eye, and extend the reach of genuine expertise across an entire organization. The purpose of this paper is not to diminish that value — it is to protect it. The gap between what AI can produce and what AI produces correctly is filled entirely by the expertise of the person using it. Organizations that understand this deploy AI effectively. Those that do not find themselves moving faster toward the wrong answer. The central argument: AI amplifies the expert. Without the expert, AI amplifies the noise.*

## 1. The Capability Illusion

Twenty years as a test engineer shapes a specific attitude toward any system: learn how it works, then look for where it breaks. That is not cynicism — it is the professional obligation of anyone responsible for quality. The most valuable thing a test engineer brings is not doubt about a system's capability, but precision about its limits. AI tools deserve exactly that same rigor, and they reward it. Throughout a career in hardware test, designs have arrived declared complete and production-ready, only for structured testing to surface real flaws — flaws invisible to the people who built them precisely because they were too close to the work. The declaration of completion is not evidence of correctness. It is the starting point for verification. AI output is no different.

AI output looks authoritative even when it is wrong. It is structured, fluent, and delivered without hesitation — and for someone without deep domain knowledge, it is genuinely indistinguishable from expert work. This is the capability illusion, and it is the most important thing to understand before deploying AI in any professional or technical context. Fluency is not accuracy. Confidence is not correctness. AI systems are trained to produce coherent, well-formed output. They are not trained to know the limits of what they know, and they will not tell you when they have reached them.

The practical consequence is straightforward, and it shows up in organizations long before AI entered the picture. An engineering manager who lacks deep domain experience will often align with the engineer who communicates best — the one who presents ideas clearly, speaks with confidence, and makes solutions sound compelling. That engineer may be entirely right. But without the technical depth to evaluate the choice against its downstream consequences, the manager has no way to know. The well-presented wrong answer and the well-presented right answer look identical from the outside. It takes expertise to see the difference — not seniority, not authority, but genuine domain knowledge of what those choices mean six months and two product generations from now. AI produces the same dynamic at scale. Its output is fluent, confident, and well-structured. A non-expert has no reliable mechanism to distinguish a correct answer from a plausible-sounding wrong one. The expert sees the difference immediately — because they already know the domain well enough to evaluate the output against reality, including the consequences that are not visible in the response itself.

**THE CORE PRINCIPLE**

*AI is a tool. Like any tool, you must master how to use it — in the right place, the right context — for its output to be meaningful and correct. A GPS gives confident turn-by-turn directions into a lake if you follow it without judgment. A spreadsheet formula returns a precise wrong number if the inputs are bad. A search engine surfaces exactly what you asked for, not necessarily what you needed to know. AI is no different — it performs exactly as designed, and produces confident wrong answers when the person using it lacks the expertise to direct it correctly or evaluate what it returns. The tool is not the problem. The gap between the tool and the expertise to use it is.*

## 2. Everyone Is Running a Different Build

Think about the people you have worked with over the years. The brilliant analyst who cannot read a room to save their life. The natural leader everyone follows who could not explain the technical details of their own product. The quiet person in the corner who never speaks up in meetings but solves the problem nobody else could. None of them are better or worse than the others — they are just built differently. Each person has a fixed pool of capability distributed across traits like intelligence, common sense, creativity, people skills, discipline, and domain knowledge. Nobody gets to max them all out. Life does not work that way.

The people who get the most out of those around them — the best managers, coaches, and collaborators — are the ones who stop expecting everyone to be good at everything and start asking a better question: what is this person built for, and what do they need around them to be at their best? They put the analyst in front of the data. They put the communicator in front of the client. They pair the visionary with someone who can execute. The team works because the builds are complementary, not because everyone is the same.

AI tools are no different. Every model has a build — things it does exceptionally well and things it genuinely cannot do. Strong at processing large amounts of information quickly, finding patterns, generating structured content, and executing well-defined tasks at a scale no human could match. Weak at knowing what it does not know, applying the kind of judgment that only comes from lived experience, and seeing around corners. The question is never whether the tool is capable. It is whether you understand the build well enough to put it in the right seat — and whether you have the expertise alongside it to cover what it cannot do.

**THE ONE-LINE VERSION**

*Everyone is running a different build. You just need to figure out their stats — and put them in the right seat. AI is no different. The seat it belongs in is amplifier, not authority.*

## 3. What AI Is Not

AI has read everything. That is not the same as knowing anything. Reading every cookbook ever written does not make you a chef. Reading every medical journal does not make you a doctor. The information is there, but the ability to apply it correctly — in a specific situation, with real consequences, under real constraints — comes from doing the work, not reading about it. Think about someone who has spent twenty years in a single field. They walk into a situation and immediately notice something is off before anyone else has said a word. That instinct is the accumulated result of thousands of situations where they were right, wrong, surprised, and corrected. AI has never been surprised. It has never been wrong in a way that cost it anything. It has never had to live with the consequences of a bad call. That experience is what expertise is built from, and it is exactly what AI does not have.

The same is true of intuition and judgment. Intuition is what you cannot explain but know is right — the feeling that something is off before the data confirms it, the read on a situation that no checklist could produce. It is built through years of feedback that was real and consequences that were yours to own. AI can produce responses that sound like intuition. But there is no accumulated experience behind them, no scar tissue from being wrong. What looks like instinct is pattern completion. Judgment is no different — it is the ability to make a call when the data points in two directions and the rules do not quite fit, and then own the outcome. AI can help you think through a problem. It can surface options and organize complexity. What it cannot do is make the call, and it cannot be accountable for what happens next. The moment an organization treats AI output as the final answer rather than the starting point for a qualified person’s judgment, it has not automated the decision. It has just removed the accountability for it.

## 4. The Current Model Landscape

Understanding which AI models exist and how they compare requires looking at more than vendor marketing. The Arena Leaderboard ([arena.ai/leaderboard](https://arena.ai/leaderboard)) provides crowd-sourced, human-preference rankings across hundreds of models evaluated on real tasks — text, code, reasoning, creative writing, and instruction following. Rankings are based on blind head-to-head comparisons by human voters, not benchmark scores alone.

### IMPORTANT NOTE ON RECENCY

*The model landscape changes rapidly. Rankings, model names, and capability profiles listed here reflect the Arena Leaderboard as of the date of this publication. New models release frequently, existing models are updated without notice, and rankings shift with each new evaluation cycle. Always consult [arena.ai/leaderboard](https://arena.ai/leaderboard) for current standings before making model selection decisions.*

The table below shows the top 10 overall text models from Arena at time of publication, with sub-rankings across key capability dimensions. Rankings are ordinal — lower number is better. A model ranked '1' in a category leads that category.

#	Model (Provider)	Overall	Expert	Hard	Coding	Math	Creative	Instruction
1	claude-opus-4-7-thinking (Anthropic)	1	2	3	1	6	1	2
2	claude-opus-4-6-thinking (Anthropic)	2	1	1	3	1	2	1
3	claude-opus-4-6 (Anthropic)	3	3	2	4	5	6	3
4	claude-opus-4-7 (Anthropic)	4	4	4	2	7	5	4
5	gemini-3.1-pro-preview (Google)	5	8	5	7	4	3	5
6	muse-spark (Meta)	6	29	6	5	24	10	16
7	gpt-5.5-high (OpenAI)	7	5	9	9	3	16	8
8	gemini-3-pro (Google)	8	16	7	17	10	4	12
9	grok-4.20-beta (xAI)	9	39	16	23	31	8	26
10	gpt-5.4-high (OpenAI)	10	7	8	8	2	18	7

Source: [arena.ai/leaderboard](https://arena.ai/leaderboard) — rankings current as of publication date. Columns show rank within each sub-category, not a score. Lower number = higher rank.

The pattern across the rankings confirms what the framework predicts: no single model leads every category. The model ranked first overall ranks sixth in math. The math leader ranks seventh overall. Creative writing and analytical depth diverge sharply across the board. Every model has a build — and knowing which build fits your task is the decision that matters.

## 5. Not One Tool — Many Tools

Most people discover AI through one tool and assume they have seen the category. That is like test driving one car and thinking you understand every vehicle on the road. The AI landscape has branched into distinct types of tools, each built for a different job. Using the wrong one is not just inefficient — it is often how people end up with confident wrong answers, or worse, with sensitive information going somewhere it should not.

The conversational tools — Claude, ChatGPT, Gemini, Grok — are what most people think of when they hear AI. You ask, it answers. They are genuinely capable across a wide range of everyday tasks: drafting, summarizing, researching, explaining, brainstorming. But even within this group, the builds differ. Some are better at long, complex reasoning. Some are better at creative work. Some connect to the internet in real time; others work only from what they were trained on. Knowing which one you are using, and what it can and cannot see, matters more than most people realize.

Then there are the tools built for specific jobs. Coding assistants like Claude Code and OpenAI Codex do not just answer questions about software — they read your actual project, write code, run tests, and make changes on your behalf. Search-focused tools like Perplexity synthesize answers from live web content rather than from training data, which makes them more current but requires you to verify whether they are accurately representing their sources. Image and video generators like Midjourney, Sora, and Runway create visual content from text descriptions — powerful for creative work, but with real limitations around accuracy and ownership that organizations are still learning to navigate.

## 6. AI Does Not Remember You

This is the thing that surprises people most — and it is not obvious until it causes a problem. Most of us are used to tools that remember us. Your phone knows your contacts. Your email knows your history. Your work software picks up where you left off. AI feels the same way. It does not work that way.

Within a single conversation, the AI holds everything you have said — but only up to a point. Think of it like a whiteboard that can only hold so much writing. As the conversation gets longer, earlier parts get erased to make room for new ones. The AI does not tell you this is happening. It does not flag that it has lost track of something important. It just responds based on whatever it can still see — and if the context you established early on has dropped off, it will answer as if you never said it.

Across separate conversations, the reset is total. Close the window and come back tomorrow, and the AI has no idea who you are, what project you are working on, what decisions you made last week, or what you spent the last session figuring out together. Every new conversation starts from zero. This catches people constantly. They invest time giving the AI background in one session, assume it carries over, then get confused when the next session produces answers that ignore everything they already established. The AI is not being difficult. It simply never had access to it.

Not all AI tools handle memory the same way, and the differences are significant. ChatGPT has gone the furthest — it operates on three layers: facts you explicitly tell it to save, a running reference to your full conversation history, and behavioral patterns it infers about you over time that are not even visible in your settings. The more you use it, the more it builds a profile of how you think and what you need. Grok added persistent memory in April 2025 and works similarly, remembering preferences and project details across sessions. Gemini builds what amounts to a knowledge graph of your preferences, extracted from your conversations and referenced in future ones. Perplexity, by contrast, remembers almost nothing by default — it is built around real-time search and treats each conversation as fresh. Claude sits in a different position: it can hold more information within a single conversation than most tools, but across sessions it searches past conversations only when you explicitly ask it to, which gives you more control but less automatic continuity.

Testing AI with as few words as possible — a common way to evaluate how a model behaves — is revealing for exactly this reason. A bare-bones prompt shows what the AI assumes when you leave things out. What it shows is that the AI fills those gaps without hesitation, without disclosing it is doing so, and without any signal that it is guessing. The expert catches it. The non-expert takes the output at face value.

The practical consequence of all this is that memory is not a feature you can take for granted — it is a capability that varies by platform, requires configuration, and rewards the user who understands how their specific tool actually works. A user who switches between ChatGPT and Claude and assumes they behave the same way will get different results and not know why. A user who does not realize that memory features exist, or that they can explicitly tell any of these tools what to remember, is leaving real capability on the table. There is also a flip side: because these tools are building profiles of you over time, it is worth knowing what they know. Every major platform allows you to review and delete what has been stored. Most users never check. The person who understands the tool manages it. The person who does not is being shaped by it without realizing it.

#### THE PRACTICAL RULE

*Never assume the AI remembers. At the start of any session where accuracy matters, re-establish the context that is load-bearing — the project, the constraints, the decisions already made. And if the platform offers memory or summarization features, learn how to use them. Two minutes of setup prevents an hour of working from the wrong assumptions.*

## 7. The Expertise Gap in Practice

The stakes of the expertise gap are not the same across every situation. Getting an AI to write a birthday message without deep domain knowledge is fine — if it is slightly off, you fix it and move on. But in professional, technical, or high-stakes contexts, the gap between what AI produces and what AI produces correctly has real consequences: decisions made on flawed analysis, recommendations accepted without the experience to challenge them, risks missed because nobody in the room knew what to look for.

A surgeon's skill determines whether the scalpel helps or harms. An attorney's judgment determines whether the legal research leads to the right argument. The tool amplifies what the person brings to it. In fields where the cost of being wrong is low, the expertise gap is manageable. In fields where the cost of being wrong is high, it is the most important variable in the room.

The table below maps common professional use cases across five dimensions: what AI handles well, what expertise contributes that AI cannot, and what happens when that expertise is missing. It is not exhaustive — every field has its own version of this gap. The point is the pattern, which is consistent regardless of the domain.

Domain	What AI Does Well	What Expertise Provides	Gap When Expertise Is Missing
<b>Engineering Analysis</b>	Pattern matching across large datasets, literature synthesis, structured output	Knows which patterns violate domain physics vs. statistical coincidence	Accepts plausible-sounding analysis that is physically impossible
<b>Manufacturing Quality</b>	Cpk trend ID, limit flagging, failure clustering	Understands root cause hierarchy, measurement system variation, process context	Optimizes the metric while missing the underlying process problem
<b>Technical Writing</b>	Structure, fluency, completeness, formatting consistency	Knows what is accurate, what is missing, what the reader actually needs	Professionally written document with wrong or incomplete technical content
<b>Code Review / Debug</b>	Syntax errors, common anti-patterns, test coverage suggestions	Understands system architecture, performance constraints, failure context	Correct fix for the symptom — misses the root cause entirely
<b>Strategic Decision Support</b>	Scenario framing, option generation, research synthesis	Understands org constraints, history, and what 'good' looks like in context	Reasonable-sounding recommendations that ignore the actual operating reality

The pattern that runs through every row is the same: AI handles the execution, expertise handles the judgment. AI produces the output, expertise determines whether the output is right. Remove the expertise and you do not lose the output — you lose the ability to know whether it matters.

## 8. The Amplifier Model

The most useful way to think about AI in professional practice is as an amplifier. An amplifier does not create signal — it increases the one it receives. Feed it something clean and it becomes more powerful. Feed it noise and it becomes louder noise. The amplifier does not know the difference. That is not its job. That is yours.

A seasoned attorney using AI to research case law can cover in an hour what used to take a day. But the AI does not know which precedents are relevant to this judge, this jurisdiction, or this client’s specific situation. The attorney does. A marketing director using AI to draft campaign messaging can produce ten variations in the time it used to take to write one. But the AI does not know which one will land with the audience, which one misrepresents the brand, or which one the CEO will never approve. The director does. A financial analyst using AI to model scenarios can stress-test assumptions faster than ever before. But the AI does not know which assumptions are grounded in the actual business and which ones are technically valid but practically meaningless. The analyst does.

In each case the AI accelerates the work. The human determines whether the work is right. That is not a limitation of AI — it is a description of how any powerful tool works in the hands of someone who knows their field. The woodworker’s table saw cuts faster than a handsaw. It does not know what to cut, where to cut, or whether the measurement is correct. A power drill drives a screw in seconds. It does not know whether the screw belongs there. The tool amplifies the decision the person makes. If the decision is informed, the tool makes it faster. If the decision is uninformed, the tool makes the mistake faster too.

This is where organizations go wrong with AI deployment. They see the speed and the output volume and assume the expertise requirement has been reduced. It has not. If anything, it has increased — because the AI produces more output, faster, and it all looks equally confident regardless of whether it is right. The expert who validates AI output at scale needs to be sharper, not less sharp. The review loop does not get easier when the volume goes up. It gets more important.

#### THE PRACTICAL IMPLICATION

*Every organization deploying AI needs to answer three questions before evaluating any tool: Who is the expert in the loop? What is their role in validating the output? What happens when the AI is wrong? If any of those three answers is undefined, the tool is not ready to deploy — regardless of how impressive it looks in a demo.*

## 9. Building AI Practices That Work

Organizations that deploy AI successfully share a common characteristic: they treat it as a capability multiplier for existing expertise, not a substitute for it. The ones that struggle share a different common characteristic: they deploy the tool before they have defined what the human role in the process looks like.

Before deploying any AI tool in a meaningful workflow, someone needs to be accountable for validating the output. That person needs domain expertise — not just familiarity with the interface. Without this, the capability illusion takes hold and confident wrong answers accumulate unchecked. AI output in technical and operational contexts should be treated as a first draft that requires expert review, not a finished product. That distinction, clearly defined and consistently enforced, is the difference between AI that accelerates good work and AI that accelerates mistakes.

How you ask matters as much as which tool you use. The quality of AI output is a direct function of the quality of the input. Vague prompts produce vague answers that look complete. Organizations that invest in developing prompt skills across their teams consistently get more value from the same tools than organizations that treat it as self-evident. Similarly, the right tool for the job shifts as the landscape evolves. Checking current benchmarks — sources like the Arena leaderboard at [arena.ai/leaderboard](https://arena.ai/leaderboard) — before making model selection decisions is a habit that pays for itself.

Finally, measure what matters. The number of AI-assisted outputs produced is not a measure of value. The accuracy of those outputs, assessed by people who know the domain, is. Volume metrics reward adoption. Quality metrics reward effectiveness. Organizations that track only volume will find themselves producing more and more confidently, while the actual value of what they produce quietly erodes.

The organizations getting the most out of AI are not the ones that have deployed it most broadly. They are the ones that have deployed it most deliberately — in roles where it amplifies genuine expertise, with review loops that catch its errors, and with a clear-eyed understanding of what it is and what it is not.

## 10. Choosing Tools Without Limiting Yourself

Most organizations arrive at their AI tool choice the same way they arrive at most technology decisions: procurement, security review, cost negotiation, and a contract. One vendor wins. Everyone uses that vendor's tool. The problem is that the amplifier model breaks down when everyone is forced to use the same amplifier regardless of what they are trying to amplify. The attorney who needs deep reasoning and the marketing team that needs creative fluency and the developer who needs coding precision are all handed the same tool and told to make it work. Some of it does. A lot of it does not, and nobody quite knows why the output keeps feeling flat.

This is a structural problem, not a people problem. The data bears it out. At the start of 2025, over 90% of enterprise AI queries on multi-model platforms were handled by just two models. By the end of the year, no single model commanded more than 25% of usage. What changed was not the tools — it was that organizations started treating models the way they should have been treating them all along: as a menu of options, not a single default. The top enterprise accounts using Perplexity's platform averaged 30 different models in their workflows. The typical account averaged seven. The gap in output quality between those two groups was not subtle.

The practical tension is real. Security teams want data in one place. Legal wants one contract. Finance wants one line item. IT wants one thing to support. All of those concerns are legitimate. The answer is not to ignore them — it is to solve for access within governance, rather than restricting access to solve for governance. Multi-model platforms are emerging specifically to address this. Perplexity is one example: it provides access to Claude, GPT, Gemini, and other leading models through a single interface and a single security framework, automatically routing tasks to the model best suited to handle them. Other platforms take similar approaches. The specific vendor matters less than the principle — one managed environment, multiple builds available, governance intact.

This is the direction the industry is heading. The question for organizations is not whether to use multiple models — that is already the answer for teams that are getting serious results. The question is who orchestrates that access and how it is governed. The organization that solves this well gets the amplification benefit without the security exposure or the management overhead. The organization that solves it badly gets fragmentation, shadow IT, and employees using personal accounts to access the tools that actually fit their work. The organization that does not solve it at all standardizes on one tool, wonders why the results are underwhelming, and attributes it to AI not being ready yet.

### THE STRATEGIC QUESTION

*The constraint is rarely the tools — it is the governance around them. Organizations that build access frameworks allowing the right model for the right task, within a single managed environment, unlock the full amplification benefit. Those that standardize on one tool for simplicity get simplicity. They also get one amplifier for every job, regardless of fit.*

## The Bottom Line

AI does not equal Artificial Intuition. The tools are powerful. They are not wise. They produce fluent outputs without understanding them, confident answers without knowing when they are wrong, and comprehensive analysis without the contextual judgment that makes analysis actionable.

The gap between what AI produces and what AI produces correctly is filled by the expertise of the person using it. That expertise has to be real, has to be in the loop, and has to be accountable for the output. There is no version of effective AI deployment in technical or operational work that removes this requirement.

Everyone is running a different build. So is every AI tool. The job of the expert is to know the builds, put them in the right seats, and take responsibility for what comes out. That job has not changed. AI has made it faster — and made the cost of skipping it higher.

### About the Author

Jonathan Slavic is a Customer Solutions and Analytics Architect at Circuit Check Inc. with over 25 years of experience in hardware test engineering, manufacturing systems, and production data intelligence. He has served as a Principal Test Engineer at SRCTec LLC, a Department of Defense contractor, and at MSA Safety Inc., an OEM — giving him direct experience across CM, OEM, and government-contracted environments. He serves as an Adjunct Professor in the Electronic Technology program at SUNY Onondaga Community College. In his role at Circuit Check Inc., Jonathan leads pre- and post-sales integration of production data systems, including WATS and TestPartner, helping manufacturers eliminate architectural debt and transform legacy test environments into scalable manufacturing intelligence platforms. Circuit Check Inc. is the leading fixture and test system manufacturer in the United States.

*Model rankings and leaderboard data referenced from [arena.ai/leaderboard](https://arena.ai/leaderboard). All rankings reflect standings at time of publication. The AI model landscape evolves rapidly — new models release frequently and rankings shift with each evaluation cycle. Consult [arena.ai/leaderboard](https://arena.ai/leaderboard) for current standings before making model selection decisions.*