

FOUNDATIONS • AIL-FP-2025-03

The Story Your Data Is Trying to Tell You

Why Narrative Data Will Be the Most Valuable Asset in AI-Powered Learning

Gregg Collins • Brandon Dickens • July 2025

Most organizations sit on mountains of performance data yet struggle to translate it into training that actually changes behavior, because the metrics that fill their dashboards rarely explain why things go wrong. This paper argues that narrative data—the stories behind incidents, near-misses, and quiet successes—will be the most valuable asset in AI-powered learning, because only narrative reveals root causes precise enough to design training against. Critical Mistake Analysis, coupled with Synthetic Work practice in AI-generated replicas of the situations where real work goes wrong, turns narrative data into targeted scenarios in the same way Crew Resource Management turned cockpit voice recordings into the safest form of transportation in human history. For the first time, generative AI collapses the cost of collecting, analyzing, and translating narrative data at scale, making it feasible to apply the aviation training model to every job whose performance matters.

The shortest distance between a human being and the truth is a story.

— Anthony de Mello

A Tale of Two Trucks

It's early morning at an open-pit copper mine in Chile. A heavily loaded haul truck begins the long climb up to the ore crusher—a routine run that happens hundreds of times a day. Fifteen minutes later, the truck is at the bottom of a 60-meter embankment and the driver is dead.

The incident report is full of the usual quantitative data—time of day, weather, mechanical condition, hours on shift—but these parameters all look normal. In the “cause” field the safety team writes two words: “operator error.” Case closed, nothing learned.

Three months later, a second truck goes off the road on the same route. This time the driver survives. And this time, the safety team does something different: instead of just filling in the incident report parameters, they're determined to get the story. They interview the driver, talk to other drivers who use the same route, review dashcam footage, and walk the road.

What they find is a subtle but deadly trap. The truck was negotiating a sharp curve that had been widened recently. The new road surface that was added to the outside of the turn was at a slightly lower level than the old one, and the difference was enough to shift the truck's center of gravity at the worst possible moment.

After the first accident, the most the safety team could recommend was some additional training telling drivers to "be careful" and "slow down"—generic admonishments that were already plastered on posters all over the facility.

The second accident was a different story. Because the team was able to put together a proper root-cause analysis, they were able to recommend specific training telling drivers what to watch out for, and what to do when they encountered it, complete with photos of what the grading discrepancy looks like from the cab.

Understanding an incident is rarely just about the quantitative data. There's always a story, and knowing the story is critical to understanding how to ensure that nothing similar happens again.

The revolution that's about to hit your training department

The argument for narrative data isn't a new one, but it's newly relevant for those of us in learning and development, because it is going to be central to our ability to exploit the newest AI-driven revolution in training.

Generative AI has made it possible—right now—to take the massive troves of performance data that organizations already collect and do something really useful with them. Sales numbers, error rates, quality metrics, customer satisfaction scores, cycle times, incident logs—all of this data is sitting in systems all over the enterprise, and most of it has never been anywhere near the training department. AI is about to change that.

Gen AI agents can analyze this massive and continuous flow of information and tell you where mistakes are happening, how often they happen, and how costly they are. In other words, you can use AI to do performance-driven needs analysis—not the kind where you send out a survey and ask managers what they think their people need to work on, but serious, data-driven investigation that will tell you where the real skill gaps are.

This is a very big deal. For the first time, it is feasible to run a continuous, data-driven cycle of performance analysis that can drive decisions about what kind of training to create and how to assess its impact. This process can be applied across an enterprise, at scale, for every role and every skill that matters. That's a fundamental transformation of what corporate learning can be.

The data that tells you why

Naturally, when you hear the word “data” in this context, you are likely to think of quantitative data—numbers, metrics, scores. And that will be the starting point for this process, because there is already so much quantitative information being captured in most organizations.

But it is narrative data that will ultimately be more significant in driving learning initiatives. Narrative data tells a story about what happened in a particular circumstance. Incident reports contain narrative data, even though, as in the

example above, the narrative is often minimal. After-action reviews, customer call recordings, dashcam footage, bodycam video, secret-shopper videos, field observations, verbal debriefs, and any other record that captures not just that something went wrong but how it unfolded—these are all types of narrative data, or at least data from which a narrative can be easily reconstructed.

Narrative data is the stuff that tells you why things happened, and the “why” is what you need if you want to design training that actually fixes the problem, rather than training that is vaguely aimed in the right general direction and hopes for the best.

Narrative data is the key to a methodology we call Critical Mistake Analysis (CMA)—a systematic process for identifying the most common and costly mistakes practitioners make, determining their root causes, and prioritizing training targets based on the business value of addressing each. Narrative data fuels CMA in three interconnected ways:

First, it enables root cause analysis. Every commercial aircraft carries two recorders: a data recorder tracking hundreds of numerical flight parameters, and a voice recorder capturing what the pilots said. The data recorder is a gold mine, of course, but the voice recorder is more often the critical input in analyzing an accident, because it reveals the pilots’ thinking. It can show that they were distracted, misread the situation, overlooked a critical input, misdiagnosed a problem, or followed a flawed procedure. Without that narrative, the numbers alone rarely explain what actually went wrong.

Second, it enables pattern recognition across incidents. Some incidents look similar but happen for different reasons. Consider three cases of “controlled flight into terrain”—aviation’s term for flying a perfectly good airplane into the ground. In one, the pilots misread an aeronautical chart. In another, they were descending to get below the clouds, and went too far. In a third, they had mis-set their altimeter. Same outcome, three different root causes—and each one calls for different training. When

we look to identify patterns of incidents so that we can decide how to address the underlying issue, we want to group together incidents with the same root cause, not ones that simply share surface features or similar outcomes, and for that we need the narrative of the pilots' decision-making.

Third—and this is the big one—the narrative behind a pattern of mistakes often tells you exactly what kind of training to build. Once you understand the root cause of a recurring problem, you are very close to knowing exactly what the training scenario should look like: put the learner in the situation the root cause analysis tells you leads to the mistake, let them face the same temptation to get it wrong that led previous practitioners astray, and have them practice the correct response until it's second nature. This is what the training world has traditionally called “simulation,” but we prefer the term Synthetic Work—because it emphasizes that the point isn't to simulate something, it's to actually do the work, in a realistic replica of the situations where real work goes wrong, before the stakes are real. This is the methodology that has made commercial aviation the safest form of transportation in human history. It works. And narrative data is the fuel that makes it go.

Proof of concept: the problem no instrument could measure

It's worth pausing to appreciate how spectacularly well this works when it's done right, because the aviation example is not just an illustration—it's a proof of concept for the entire argument.

On January 13, 1982, Air Florida Flight 90 attempted to take off from Washington National Airport in a snowstorm. The cockpit voice recorder revealed a first officer who was clearly aware that something was wrong with the engine readings. But instead of stating it directly, he hinted: “That doesn't seem right, does it?” and “That's not right.” The captain dismissed each comment. The airplane, its wings heavy with

ice and its engines producing inadequate thrust, staggered off the runway, failed to climb, and hit the 14th Street Bridge over the Potomac River. Seventy-eight people died.

The flight data recorder showed that the airplane failed to gain enough airspeed to climb. The voice recorder showed why the problem wasn't caught by the crew: an overconfident captain who seemed determined to take off no matter what, and a first officer who knew the plane wasn't going to make it but couldn't bring himself to say so plainly to his superior.

It was a pattern investigators were seeing over and over. In 1978, the crew of United Airlines Flight 173 ran out of fuel while troubleshooting a landing gear problem over Portland, Oregon, because the flight engineer's increasingly urgent warnings about the fuel state never broke through the captain's focus on the gear. And in the deadliest aviation accident in history, the 1977 Tenerife disaster, 583 people died because a KLM captain initiated a takeoff roll without clearance and his first officer didn't challenge him forcefully enough. Dozens of similar cases accumulated over the years.

In each case, the flight data told investigators that an airplane had crashed, but the narrative told them why: junior officers were afraid to challenge senior ones, and captains were making decisions in isolation when they should have been drawing fully on every resource available to them. The root cause was a social and cognitive problem—one that lived entirely in the interactions between human beings and was completely invisible to any instrument on the aircraft.

By the late 1970s, NASA researchers had assembled enough of these narratives to see the full scope of the problem: 60 to 80 percent of aviation accidents were being caused not by mechanical failures or bad weather but by breakdowns in communication, leadership, and decision-making among the flight crew. The training response, developed through a major NASA research initiative and known as

Crew Resource Management (CRM), addressed the problem directly. CRM teaches a set of techniques for overcoming the known communication failures that cause cockpit disasters, starting with how to say ‘Captain, we have a problem’—and how to hear it.

The impact of CRM training has been enormous. Commercial aviation in the US has gone from 29 fatal accidents in the 1970s to just one in the 2010s, while passenger volume has more than quintupled. If cockpit communication failures were responsible for 60 to 80 percent of accidents in the pre-CRM era, CRM’s share of that improvement very likely represents hundreds of lives saved per decade. The FAA made CRM training mandatory for all commercial pilots in 1995, and the methodology has since been adopted by military aviation worldwide and adapted for use in medicine, nuclear power, and offshore oil production.

And none of it—not a single element of the diagnosis, the training design, or the response—could have come from purely quantitative data. The numbers could tell you how the planes were crashing. Only the stories could tell you that the reason they were crashing was because the crew members had failed to communicate effectively.

The narrative data that’s already all around you

Here’s the good part. Most organizations are already swimming in narrative data. But that data is routinely under-studied, and few people in training are paying attention to it. Most of it just piles up in databases, or is deleted without anyone having actually studied it.

Audio recordings. Customer service calls are recorded as a matter of course in most organizations. These recordings are a gold mine. A customer satisfaction score can tell you that 23% of calls about billing disputes end badly. The recordings can tell you that in many of those calls, the rep failed to acknowledge the customer’s frustration

before jumping into the policy explanation, and that this failure to acknowledge is what escalated the conversation from mildly annoyed to genuinely angry. That's a root cause. That's actionable.

Video recordings. Even richer, and increasingly ubiquitous. Dashcam and cab-mounted cameras in commercial vehicles now capture virtually every road incident. "Secret shopper" video in retail and hospitality reveals, with sometimes uncomfortable clarity, exactly what goes wrong on the floor and why. A secret-shopper video might show a retail associate saying exactly what she was trained to say when a customer complains—but her body language is clearly communicating impatience and skepticism, and that's what the customer is actually responding to. No survey would ever surface that. Bodycam footage captures interactions that can be analyzed for procedural adherence and communication effectiveness. Security cameras in warehouses and distribution centers capture near-misses and safety incidents. In each case, the video tells a story that no metric could: not just that something went wrong, but how it unfolded moment by moment.

Incident reports and after-action reviews. The most traditional form of narrative data, and the most variable in quality. A good incident report reads like a detective story—sequence of events, decision points, chain of causation. A bad one just fills in the cause field with a generic label. AI is going to dramatically improve incident report quality by guiding investigators through structured narrative collection, prompting the right questions, and flagging gaps in the story.

Verbal protocols. A verbal protocol is a real-time narration by a person of what they're thinking and doing as they perform a task. Think-aloud protocols have been a staple of cognitive science research for decades, but they've been impractical to collect at scale because they require a researcher to be present. AI changes this. It is increasingly feasible to have an AI system prompt workers to narrate their reasoning during critical tasks, or to use ambient audio collection with AI analysis to extract the

narrative thread from natural workplace conversation. Imagine a maintenance technician talking through a troubleshooting process while an AI captures and analyzes the reasoning in real time. The resulting data reveals not just what the technician did, but why. Including the reasoning errors that led to incorrect diagnoses.

The narrative data you don't have yet—but soon will

Everything we've described so far—call recordings, dashcam footage, incident reports, verbal protocols—is narrative data that organizations are already collecting, or could collect with existing technology. But AI is about to do something much more ambitious: it's going to make it possible to build intelligent systems that actively generate narrative data in situations where little or none existed before. And from a training perspective, three things these systems can surface are particularly exciting.

First, best practices that nobody in training knew were best practices. On a manufacturing floor, a veteran operator adjusts feed rate by the sound of the cutter. A top-performing sales rep always asks one specific question before presenting pricing. A senior nurse checks a particular vital sign at a moment the protocol doesn't call for. These people don't think of what they're doing as a technique—they just do it. It's tacit knowledge, invisible to everyone except an AI system that can observe thousands of performance episodes, correlate behaviors with outcomes, and surface the patterns that separate the best performers from the rest. Once you can see those patterns, you can teach them. That's an entirely new source of training content that didn't exist before.

Second, error patterns that nobody in training knew were error patterns. A production outage hits a software team. The first senior engineer on the scene anchors on a database theory based on a similar incident last month. The team follows him for forty minutes while a junior engineer who noticed a log entry

pointing somewhere else entirely doesn't push the point. That's not a one-off—it's a systematic pattern in how teams reason under pressure, and it shows up over and over once you start looking at the narratives. But nobody writes "our teams are bad at challenging the senior person's hypothesis" in a postmortem. It takes an AI system analyzing dozens of incident narratives to see it. Once you do, you know exactly what to train.

Third, skill gaps that nobody knew existed. In pharmaceutical R&D, an experienced scientist knows when a reaction "looks right" by its color. She knows which steps in a protocol have hidden sensitivities the documentation doesn't mention. When she retires or moves on, that knowledge walks out the door—and the failure rate for new researchers on those procedures spikes, and nobody understands why. An AI system capturing verbal protocols during critical procedures can identify exactly where the gap between what the protocol says and what an expert actually does is widest. Those gaps are invisible until someone tells the story of what they're doing and why.

The technology to do all of this is arriving fast. AI systems can already monitor interactions in real time, flag episodes that warrant deeper investigation, conduct structured debriefs, correlate behavior with outcomes across thousands of events, and surface patterns no human analyst could find. The trajectory is clear: we are moving from a world in which narrative data is collected sporadically, after the fact, about incidents serious enough to trigger an investigation, to one in which rich narrative data is captured continuously—including the near-misses, the quiet successes, and the subtle patterns that never rise to the level of a formal incident but collectively determine whether an organization performs well or badly.

What AI does to the pipeline

The reason narrative data hasn't been used extensively in training until now is easy to understand—it's been too expensive and too slow to collect, too difficult to analyze at

scale, and too hard to translate into training content. A single NTSB investigation can take a year and produce hundreds of pages. That's fine when you're investigating a plane crash. It doesn't scale to the thousands of smaller incidents that happen every day across a large organization.

AI collapses the cost at every stage of the pipeline.

Collection. AI can monitor multiple data streams simultaneously—audio, video, sensor data, system logs—and flag events that warrant narrative investigation. It can guide structured interviews, prompt follow-up questions, fill narrative gaps. It can transcribe and time-synchronize audio and video. It can collect verbal protocols unobtrusively. The cost of collecting rich narrative data drops by an order of magnitude.

Analysis. AI can process narrative data at scale, identifying root causes and patterns across hundreds or thousands of incidents that no human team could review. It can cross-reference narrative accounts with quantitative data to build a richer picture than either could provide alone. And critically, it can identify the recurring decision points, misconceptions, and situational factors that are the exact inputs needed for a Critical Mistake Analysis.

Translation to training. Once the patterns are identified and the root causes understood, AI can generate Synthetic Work experiences that replicate the conditions under which mistakes are being made—branching simulations, role-play dialogues, situational exercises that confront learners with the challenges identified in the narrative data. The feedback loop from incident to training scenario, which used to take months or years, can now operate in weeks or days.

Training for every job the way pilots do now

Here's the vision, stated plainly: AI-powered narrative data collection and analysis will make it possible to treat virtually every job the way we now treat the job of commercial airline pilot. If you work in learning and development, that sentence should make you want to get out of bed in the morning.

What we do for pilots is extraordinary. Every significant incident is investigated in depth. Root causes are identified with scientific rigor. Correct responses are defined based on expert analysis. Training scenarios are created that directly target each identified risk. Pilots practice in high-fidelity Synthetic Work environments until the correct response is automatic. The result: a continuous, decades-long decline in accident rates to a level that was once unimaginable.

We haven't done this for the vast majority of other jobs—not because the methodology doesn't apply, but because we couldn't afford to. Investigating every incident in depth, analyzing root causes across thousands of events, and creating precisely targeted training for each problem requires resources that were historically reserved for domains where the stakes are measured in human lives.

AI eliminates that constraint. When the cost of collecting and analyzing narrative data drops by one or two orders of magnitude, it becomes economically feasible to apply the aviation methodology to customer service reps, retail associates, insurance underwriters, project managers, software engineers, salespeople—anyone whose performance matters to the organization. And the returns will be enormous, because we already know this methodology works. The question was never whether it worked, it was always whether we could afford to deploy it broadly. AI just answered that question.

The coming flood (and how not to drown in it)

Let's be honest about the challenges, because there are real ones.

When narrative data collection scales up across an organization, the result is an ocean of stories—incident reports, call recordings, video clips, verbal protocols, debriefs—arriving in a continuous stream. The question is how to turn that ocean into a training asset rather than a data swamp.

What's needed is something that has never existed before: an organizational learning memory. A structured repository of narrative performance data that can be queried, cross-referenced, and analyzed at multiple levels of granularity—with modern AI capabilities for natural language search, pattern detection, and automatic categorization. Imagine being able to ask the system, “Show me every instance in the last six months where a field technician misdiagnosed a pressure valve failure,” and getting back not just a count but the actual stories, sorted by root cause, with a draft training scenario already attached.

Such a system would flag incidents in real time, spot patterns across thousands of narratives, generate training scenarios on demand, and track whether those scenarios are actually reducing the mistakes they target. That last part is the one that should make training professionals sit up straight: for the first time, you'd be able to show, with real data, that a specific training intervention eliminated a specific pattern of mistakes in the field. The holy grail of training ROI, delivered not by surveys and smile sheets but by the same narrative data that identified the problem in the first place.

The technical challenges are real but well understood. The missing ingredient has always been economic justification. That's no longer missing.

The elephant in the room: privacy

There's another challenge, and this one is genuinely thorny: privacy.

Narrative data, by its nature, is data about what specific people did in specific situations. Audio captures voices. Video captures faces. Verbal protocols capture reasoning. Incident reports identify individuals by name. Any organization that ignores the privacy implications is asking for trouble—legal, ethical, and cultural.

The aviation world offers a useful precedent. The NASA Aviation Safety Reporting System (ASRS), established in 1976, allows pilots and other aviation professionals to submit confidential reports about safety incidents without fear of disciplinary action. Reports are de-identified by NASA before entering the database, and the reporter's identity is never shared with their employer or with regulators. The system has generated hundreds of thousands of narrative reports that have directly contributed to safety improvements. It works because people trust it.

AI can already transform voice recordings, obscure faces in video while retaining body language and context, and de-identify incident reports automatically. The technology for intelligent anonymization—stripping identifying information while preserving the causal and contextual details that make narrative data useful—is here and improving fast.

But technology alone won't solve this. Organizations need to build cultures in which narrative data collection is understood as a tool for learning and improvement, not as a surveillance mechanism or a basis for punishment. This is exactly the cultural shift that aviation achieved—and it wasn't easy, even in aviation. It requires clear policies, genuine protections for workers, and leadership that consistently demonstrates through action that the data is used to improve systems, not to assign blame.

The organizations that get this right will have an enormous advantage. The ones that don't—the ones that use narrative data primarily for surveillance and discipline—will find that their data dries up. People are remarkably good at not generating evidence that will be used against them.

The return on narrative

Back to the bottom line, because in corporate learning, if you can't connect an idea to business outcomes, you're just making conversation.

Think about the economics of the mining example. Two haul truck accidents, one fatal. Direct costs—equipment damage, medical expenses, lost production, regulatory fines, legal exposure—likely tens of millions of dollars. Indirect costs—in reputation, and, maybe most importantly, in the damage to morale, especially when a colleague's life is lost on the job—are even more significant.

The cost of collecting the narrative data that identifies the root cause is a few days of investigative work. The cost of the simulator training is modest given what's at stake. The ongoing benefit of preventing all future incidents of that type is priceless.

Now think about applying this approach to every significant performance problem across the organization. Not just safety incidents, but quality failures, customer service breakdowns, sales losses, project delays, compliance violations—all the ways human performance goes sideways. In each case, the narrative data tells you what's actually going wrong, the root cause analysis tells you why, and the targeted training gives you a way to fix it.

And the return gets better over time, not worse. Each problem you solve stays solved. Each mistake you eliminate from your workforce's repertoire stops costing you money, stops hurting your customers, stops damaging your reputation, stops distracting your people from their best work. The cumulative effect is an

organization that gets measurably, continuously, compoundingly better at everything it does.

That's not a training program. That's a competitive weapon.

What to do about it

If you're a learning leader, here's the practical upshot. None of this requires waiting for technology that doesn't exist yet.

Start with what you already have. Most organizations are sitting on a mountain of narrative data—incident reports, call recordings, customer complaints, field observations—that nobody in training is looking at. Pick one business problem where performance is clearly falling short, pull the narrative records, and do a root cause analysis. You will almost certainly discover things your metrics never told you.

Build the infrastructure to collect more, and the pipeline to analyze it. Dashcams, call recording systems, structured incident reporting tools, AI-assisted verbal protocols—the collection technology is available now and improving fast. But raw narrative data is only half the battle. You need AI-powered analysis that can process narratives at scale, spot patterns, and flag the recurring root causes that should be targeted in training. This is where Critical Mistake Analysis meets modern AI, and the combination is extraordinarily powerful.

Close the loop. The goal is a continuous cycle: narrative data flows from the field into analysis, analysis identifies training targets, scenarios are generated and delivered, impact on field performance is measured—generating more narrative data that feeds the next cycle. This is the aviation model. It works. AI makes it scalable to every domain.

Get the privacy framework right from day one. Establish clear policies, build anonymization into the pipeline, and demonstrate through consistent action that narrative data is used for learning, not punishment. The cultural foundation is at least as important as the technical infrastructure.

The story ahead

Synthetic Work—realistic practice in AI-generated replicas of the situations that matter most—is in the process of becoming the central methodology of modern learning. Its potential is extraordinary: a world in which every professional, in every role, builds expertise through the same kind of targeted, scenario-based practice that made commercial aviation the safest industry on earth.

But the entire concept is premised on knowing which situations to practice. If pilots spent their simulator time cruising straight and level through clear skies in a perfectly functioning airplane, they would never be ready for the challenges that actually cause accidents. The reason aviation's training works is that every scenario in that simulator is there because a deep, empirically grounded analysis of what actually goes wrong in the field—an analysis that depends critically on narrative data—put it there.

That's what narrative data is, ultimately. It's not a supporting input to the training process. It's the thing that tells you what your people need to practice, and the thing that tells you whether the practice worked. Get it right, and Synthetic Work will transform your organization's performance. Get it wrong—or skip it—and you're just building very expensive simulators for cruising through clear skies.

Your data is trying to tell you a story. Are you ready to listen to it?