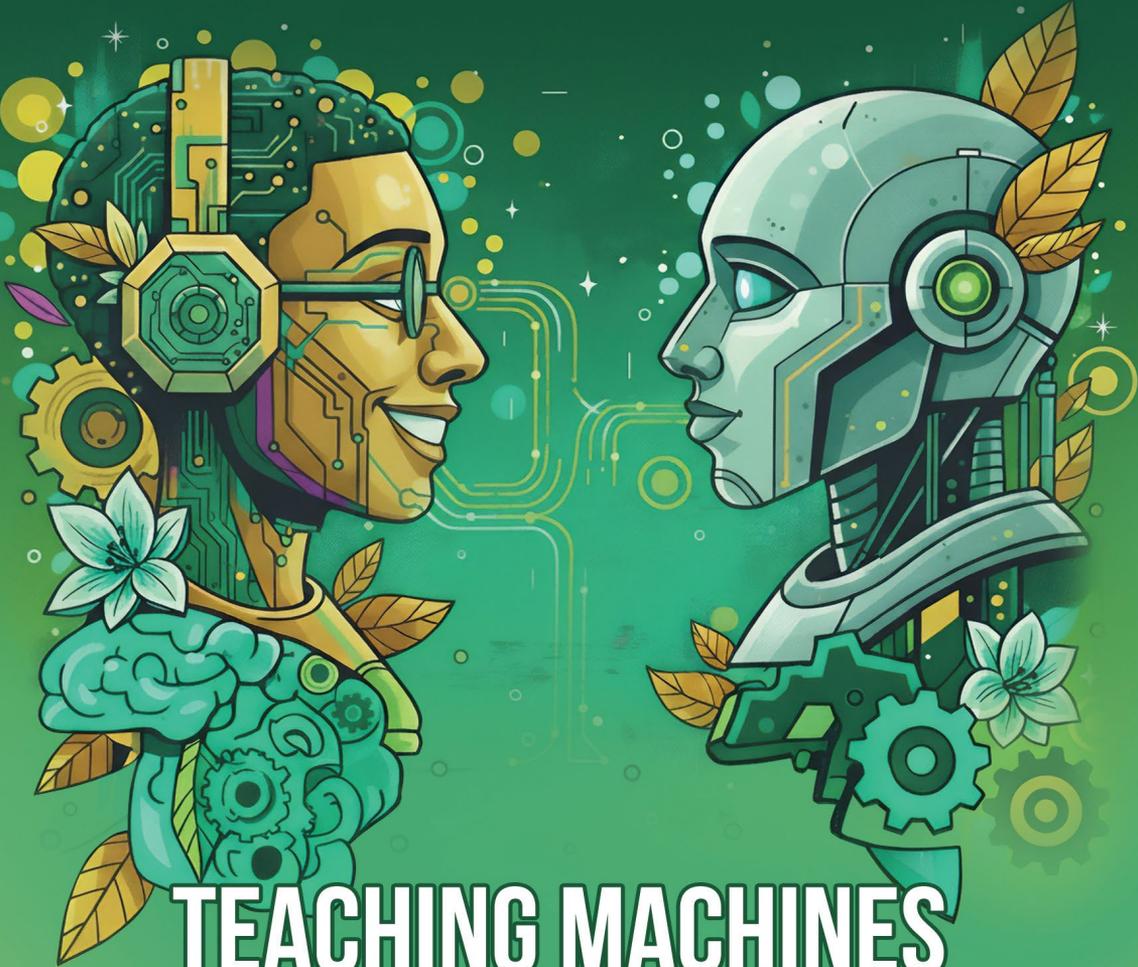


Emil Heidkamp



TEACHING MACHINES TO TRAIN HUMAN BEINGS

A Strategic Guide to Using AI
in Learning & Development



Parrotbox.ai

Teaching Machines to Train Human Beings

A Strategic Guide to Using AI
in Learning & Development

by Emil Heidkamp

Parrotbox.ai

© 2025 by Emil Heidkamp All rights reserved.

This work is licensed under a Creative Commons Attribution–NonCommercial–NoDerivatives 4.0 International License. You are free to share — copy and redistribute the material in any medium or format for noncommercial purposes only, provided that you give appropriate credit and do not modify or create derivative works based on this material.

To view a copy of this license, visit creativecommons.org/licenses/by-nc-nd/4.0

For permissions beyond the scope of this license, please contact info@parrotbox.ai

ISBN 979-8-218-85216-0

Published by Parrotbox
www.parrotbox.ai

Cover and graphic design: Henry Casanova

Editing: Megan McQuaid and Weston P. Racterson, AI

Dedicated to

William Chamberlain

(Whom I've never met but owe a great debt.)

Joan Hall

(Whom I've spoken with and to whom I am deeply grateful.)

&

Thomas Etter

(Who might or might not exist.)

PREFACE



It's the End of L&D as We've Known It, and I Feel Fine

Sometimes, entire professions become irrelevant because of technology. The industries supporting horse-drawn transportation - breeding, carriages, feed, whips - were collectively worth around \$65 billion in today's dollars when automobiles first appeared. Kodak and other camera companies made around \$15 billion from selling and processing film at the time digital cameras started going mainstream. And sometime between the 1980s and 1990s, computer publishing rendered the 500-year-old profession of typesetting (arranging metal letter blocks for printing books and newspapers) all but extinct.

So let's say the quiet part out loud: if Artificial Intelligence technology continues to improve, then it's going to put a lot of knowledge workers out of a job.

This is doubly true for learning and development (L&D) professionals. AI will not only perform the work of instructional designers and training facilitators, it will also shrink the human workforce and, by extension, the need for our services.

Yet, despite that gloomy prediction, I honestly believe that L&D professionals should be thrilled by the arrival of AI!

For more than 25 years, learning and development has been stuck in a rut of eLearning modules and PowerPoint decks, to the point where many have forgotten our discipline began as a cutting-edge branch of applied

behavioral science. Now, with the explosion of AI, we have an opportunity to reclaim that innovative mojo and help organizations navigate a world where humans and machines work side by side. For L&D professionals who feel undervalued, this is the moment in the young adult fantasy novel when the awkward teenager discovers they're the chosen one. The AI revolution is (or at least ought to be) L&D's time to shine.

However, to fulfill our destiny, learning departments need to take bold steps that will disrupt our profession's status quo. If we're honest, the "learning industry" missed the boat on nearly every major technology shift of the past 30 years. When everyone else was building responsive websites, most of us were still building Flash modules stuck in fixed-size rectangular frames. And as our audience was learning from YouTube on demand, we were still fixated on completion rates inside clunky LMS dashboards.

Now, as other industries use AI to cure cancer and reinvent human-computer interaction, we're treating it as little more than a cheaper way to crank out traditional content, or a shiny layer of automation slapped onto the same old platforms. That's troubling, because this time, if we miss the technological boat, the rest of the world might not wait for us. Just ask the former talent development team at ByteDance (parent company of TikTok) which was disbanded after leadership concluded their learning initiatives had "limited practical value" and created an "illusion of growth." Coming from one of the most successful tech companies on the planet, that's a wake-up call none of us can afford to ignore.

Getting back to the sunny side, this book isn't going to be an exercise in hand-wringing or pie-in-the-sky innovation theater. It's meant as a frank conversation about how AI will change our profession, and how we can (and should) change with it.

AI will force L&D teams to change their mindset, grapple with unfamiliar concepts, and master new skills.

But that's no problem, since learning and acquiring new skills are part of our job description - right?

The ABCs of AI...

My grandfather liked to say, “I’ve been programming computers since before they called them computers” which, like most of his boasts, was only half true. The term computer did exist when he worked at AT&T in the 1950s, though it wasn’t yet in common use.

Likewise, the language of AI is still evolving. Everyday users and business professionals often mean very different things when they say “AI,” and machine learning researchers probably bristle when they hear people use phrases like “training an AI” incorrectly.

For clarity, here’s how we’ll define the key terms in this book:

“AI” - AI technology in general, though we’ll mainly be talking about “generative AI” (the kind used by ChatGPT, Claude, Gemini to generate recipe recommendations and analyze spreadsheets) as opposed to the other types of AI used by self-driving cars and currency trading platforms.

“AI Model” - The part of an AI system that analyzes input and generates an output — what most people would casually call “the algorithm.” It’s the component that makes ChatGPT behave differently from Claude.

“AI Agent” - An application built on top of one or more models. Some define “agent” narrowly as something fully autonomous, but we’ll use it more broadly to describe any AI application more sophisticated than simply tossing a single prompt at ChatGPT. When we talk about “AI coaches” or “AI role-play simulations,” those would be examples of specialized AI agents.

That said, I will likely slip from time to time and just say “the AI” to refer to either a specific agent or the underlying model, and for that, I apologize in advance.

One Last Disclaimer Before We Begin...

There are countless books, blogs, and YouTube videos offering tips on how to get better output from AI models by phrasing your questions or commands a certain way. Some of that advice is sound, a lot of it is rubbish, and even the useful tricks will probably be obsolete within a year as the models continue to evolve.

This book will occasionally touch on how our team designs AI agents (and yes, it's a bit more involved than telling the AI to "pretend you're the world's greatest leadership coach" or "come up with an idea, then come up with a better idea.") However, we won't be getting into so-called "prompt hacks" or the nuts and bolts of software development. Presumably, if your learning department implements serious AI projects at scale, it will also allocate the resources to work out those details (or, if you're a lone tech enthusiast, you'll find a way to figure it out.)

So, if you came here looking for prompting tips, please don't be disappointed. But if you want to see concrete examples of how AI will fundamentally reshape the way we train and support employees, customers, students, and other audiences - not just next year but for years to come - read on.



Reading On the Radio: Why Using AI for Traditional Content Generation Misses the Point

Like half of the people in the learning industry, I didn't study instructional design in school. Rather, I fell into it after a brief and undistinguished career in the music industry. However there's one story from my days in the music business that seems especially relevant to the present moment for corporate L&D.

In the mid-2000s, my friend Craig and I ticked off the bosses at the major record label where Craig worked. We had apparently committed an egregious offense by (gasp!)... designing a website where people could stream songs by popular artists without having to download files or purchase compact discs.

This was 6 years after Napster made downloading MP3s popular, 4 years after iTunes made it legitimate by charging per download, and three years before Spotify debuted. For anyone in the music industry who understood technology, streaming without downloading just seemed like the obvious next step. Or, as we explained it to people at the time "Imagine a private radio station that only plays music you like, supported by personalized ads from Nike or Starbucks or whoever."

At first, the executives in charge of selling CDs laughed and dismissed it as a tech-nerd pipe dream. But as we gained traction inside the company, and even began discussions with sponsors and other labels, the laughter stopped. Suddenly, the old guard was nervous.

“This is stabbing our brick and mortar partners in the back!” shouted one vice president after seeing a budget request to take the platform to market. “If Best Buy and Tower Records hear about this, they’re never going to promote any of our artists ever again!”

Craig was just a junior digital marketing director (and I was an outside consultant), so we tried to be diplomatic. “OK, maybe we don’t like the signals this would send to retail stores. But the public loves the convenience of online music - what’s our positive vision for how we deliver it in the future?”

“We already figured that out,” said the executive who protested earlier. “Digital download cards - people go to Tower Records, they buy a card for the new Justin Timberlake album, and they go home and download it on their computers.”

One year later Tower Records¹ filed for bankruptcy. Revenues for the traditional record labels like Sony were cut in half by 2010 - hitting \$19 billion, down from a \$36 billion peak in the year 2000 - and has crawled back up to \$24 billion in 2025. Meanwhile Spotify, the world’s leading music streaming platform, is valued at \$170 billion.

Everything New is Old Again (At Least Initially)

While it’s easy to laugh at that record executive who couldn’t imagine a world where listening to music didn’t involve buying a physical object from a physical store, he was just exhibiting a common human response to the new and unfamiliar.

Whenever a transformational new technology appears, people often try to use it the same way that they use existing technologies. Automobiles were initially billed as ‘horseless carriages’ (some even had a whip holder!). When commercial radio stations first started, the concepts of a “talk show”

¹ For those too young to remember, Tower Records was a store where you could buy new music on shiny plastic discs called “CDs”, back when buying music meant leaving your house and “broadband” meant you had two phone lines.

or “radio drama” hadn’t been invented yet, so they filled time by having announcers read newspapers out loud. In my own lifetime, one of my first jobs was printing out emails for the dean of a medical school so that he could annotate them with a pen and a highlighter before dictating a response (into a cassette recorder) for his assistant to type up.

It takes time for the providers and consumers of any new tech to figure out what really makes it useful. Today’s automobiles are more than fancy carriages that don’t require live animals to operate: they make it possible to move people and merchandise with unprecedented convenience, changing everything from where we work to who we socialize with to how we buy groceries and appliances. The internet pulled that trick again 150 years later, as people began to understand that websites weren’t just digital newspapers and search engines weren’t just a newfangled yellow pages ². And radio ultimately evolved into the modern podcast, creating a world where almost anyone can speak to an audience of thousands or millions without having to get approval from some major media company.

While artificial intelligence technology is maturing faster than most, we are still very much in the horseless carriage / reading on the radio stage of understanding and applying it. And the learning & development industry is lagging near the rear of the adoption curve.

Where pharmaceutical companies are using AI to develop new therapies for cystic fibrosis and malaria (accelerating the process from decades to weeks), and architects use AI to conjure thousands of structurally viable building concepts that no human mind could previously have envisioned, in L&D we’re using AI to... generate scripts for the same old eLearning modules and the same old slide decks, just faster and cheaper.

Part of L&D’s underwhelming response to AI (and every other major technology before it) can be blamed on a collective sense of resignation. In some ways we are the marketing department’s overlooked sibling: we both produce “content” and we both collect “metrics” and share a focus on influencing people’s behavior, but our work has a less direct connection

² If you’re too young to know what the “yellow pages” were, Google it.

to the bottom line, and our budget allocations reflect that. So we just keep tinkering with our authoring tools and collating “smile sheets” and rarely taking time to dream of something beyond the training modalities we know. It’s also part self-preservation instinct: as a colleague of mine pointed out, most instructional design managers don’t want to hire a robot to train people when they could hire another human instructional designer or facilitator who could potentially become their workplace buddy.

Yet, whatever the reasons, if we remain stuck in the rut of using AI to churn out the eLearning equivalent of microwave meals, it won’t take long before management decides they don’t need to pay a veteran L&D professional to press the “generate with AI” button, when an intern can press the same button for half the price (or business stakeholders can just press it themselves, for free.)

So, how can we redefine our role as more than mere “content developers” lest we get John Henry-ed out of existence?³

The answer starts with looking at the things AI can do, then thinking about our mission as instructional designers (as a branch of behavioral science - not a weird offshoot of web design), then connecting the dots to identify how this emerging technology can advance our ultimate goals in ways we’re just beginning to imagine.

So What’s So Special About AI?

At this point in history, it’s impossible to calculate what the full impact of AI on the workplace and society will be. Some claim that AI will replace the majority of software developers, lawyers, accountants, architects, triage nurses, and most other knowledge workers within a decade. At the other end of the AI enthusiasm spectrum, some critics claim that the AI emperor has no clothes, and that generative AI is a trillion-dollar gimmick that can’t solve real world problems or replace organic human intelligence.

³Some of the people who read the rough draft of this book were unfamiliar with the old American ballad “John Henry was a Steel-Driving Man”, which tells the story of a railroad worker who competed against a steam-powered drill to prove that human labor was stronger than a machine. He won the contest, but died from exhaustion right after. Basically a metaphor for the futility of competing with automation and the plight of workers everywhere.

Whenever opinions get this polarized, the truth is probably somewhere in between (though it might heavily favor one side.)

If we wanted to be extremely reductive about what generative AI does - it's basically a talking computer. And I don't mean a regular computer with a synthesized voice: rather a computer program that, through some incredibly complicated mathematical gymnastics, can hold a coherent conversation with a human, informed by the computer's vast access to data and pattern recognition capabilities.

And talking computers are incredibly useful for L&D when you consider that most traditional training modalities consist of structured conversations. For instance:

- Depending on the subject, a lecture and discussion workshop is 60% to 90% just people sitting (or standing) around, talking to each other
- The same goes for coaching and mentoring relationships - maybe there's a bit of hands-on task practice, but otherwise it's two or more people talking
- Even a written free response assessment is just an asynchronous conversation with the instructor providing a question or scenario, the learner writing out their response, and the instructor providing feedback.

And we know from experience that these types of conversations are difficult to scale. There's never enough budget to give everyone in an organization a coach, senior team members have other responsibilities besides mentoring juniors, and even getting people in the same room (or on the same video conference) for a workshop can be a huge logistical headache.

That inconvenient fact has fueled L&D's 30-year obsession with eLearning. As proud as I am of my own company's eLearning and multimedia work, things like videos, drag and drop interactions, and multiple choice quizzes, were never going to match a well-structured conversation with an expert instructor or colleague. But, whatever their shortcomings - eLearning, videos, and other "static" media were easy to deliver at scale.

And the irony of this is that when AI finally showed up, the L&D profession was so blinkered into equating "digital learning" with "static media" that most instructional designers said "Wow - here's a computer that can write scripts for eLearning!" rather than "Wow - here's a computer that can have conversations with learners!"

Talking to Computers: Old Modalities, New Technology

I don't claim to know how, exactly, AI will ultimately transform (or destroy) L&D. From the printing press to the smartphone, the best uses for new technologies tend to reveal themselves through lived practice and happy accidents - not by theorists sketching out designs on a whiteboard. Still, I can share some of my team's early, practical experience helping organizations leverage AI for workforce training and on-the-job support.

In the chapters that follow, we'll look at case studies and try to draw some preliminary insights for how AI can:

- Serve as an always-available tutor, coach, and mentor
- Facilitate role plays and simulations
- Grade free response assessments
- Provide more meaningful reporting and insights for career development
- Present an opportunity for L&D to take on a role of real strategic leadership in the AI-powered workplace of the future.

Of course, all of the use cases listed above are guilty of applying AI to old training paradigms. But instead of asking "How can AI reduce the cost of eLearning?" we're asking "How can AI replicate the best of human

coaching, mentoring, case-based discussion, role play simulation, and reflective feedback?” And I do think this is the better course of inquiry.

The tools might be new, but the underlying instinct of learning through conversation is as ancient as Socrates and Zhuangzi. And who knows? Perhaps, by helping people have better conversations with computers - L&D professionals can reclaim the soul we sold for the sake of scale.



Artificial Aristotles: Using AI as a Virtual Coach / Advisor

Around 2,500 years ago, the Greek philosopher Socrates worried that writing - the very concept of the written word - was making people dumber.

Socrates feared that students relying on “external marks” instead of simply paying attention and remembering what their teachers said, would make them forgetful. He also felt the static, unchanging nature of text would discourage people from actively questioning and engaging with ideas. And then there was the risk of misinterpretation: unlike a living teacher, a scroll couldn’t confirm a learner’s understanding or provide context and guidance.

Thankfully, Socrates’ student Plato wrote all these arguments down, so we can ponder them today... though I’m sure Plato got in trouble for it.



SOCRATES: Plato, what are you doing?

PLATO: Uh... taking notes.

SOCRATES: Damn it, Plato, put that papyrus and reed away. How many times do I have to tell you - no devices in class?!



Returning to the 21st century, we might wonder what Socrates would make of videos, e-learning, and all of our other learning formats today. While arguably more stimulating than words on a page, videos and e-learning are still static media.

While some people might feel more engaged by a YouTube documentary on vegetarianism than a book or a pamphlet, you can't interrupt the YouTube vegetarianism documentary and ask the narrator to reconcile their claim that wartime meat rationing in Scandinavia improved public health with the fact that Scandinavians were still eating herring, cod, and other fish by the boatload (though you can probably start an argument about it in the comments section.)

Similarly, while an e-learning module can quiz a product manager at an athletic shoe company on risk mitigation strategies, it can't talk to the manager about their concerns that asking for more budget after disappointing focus group results will damage their credibility with the CFO.

Which brings us to the real question: would Socrates rather talk with an AI agent than watch a video?

Giving Every Employee an AI Aristotle

I love reading and writing, and I'm grateful Plato preserved Socrates' words for future generations. One of the people who learned from those words was Aristotle - arguably the greatest private tutor in history. ⁴

About forty years after Socrates died, Aristotle was invited by his father's friend, King Philip II of Macedon, to tutor Philip's teenage son, Alexander. For the next three years, Aristotle taught him ethics, politics, rhetoric, philosophy, science, literature, and mathematics—not in a stuffy schoolroom, but in a garden by the sea. Six years later, Alexander put that mentorship to work by conquering most of the known world from Greece and Egypt to India and Bactria (modern-day Afghanistan).

⁴ With apologies to Anne Sullivan Macy.

Now imagine if every financial analyst, nurse supervisor, and electrical apprentice had access to that kind of personalized instruction, minus the global conquest?

Unfortunately, in today's workplace, Aristotle-caliber coaching is reserved for top executives and the occasional golden-goose salesperson or technical superstar. Large organizations might spend tens of thousands to coach a senior vice president on leadership skills, while tech giants like Google and Meta fly sales managers to luxury resorts in Hawaii and the Mediterranean for weeklong retreats with gourmet meals, famous keynote speakers, and fabulous parties.

But would the same organizations provide a coach to help a fast casual restaurant shift supervisor think through how to confront a chronically late employee? Sorry, not in the budget.

People at that level (i.e. 99.6% of workers) are lucky to get a half day workshop in a cramped conference room in the suburbs of Stuttgart, or a \$40 "conversation skills for managers" course from LinkedIn Learning, if they receive any training at all.

And this is a huge missed opportunity. A study in the *Journal of Organizational Psychology* found coaching reliably produced a 15–20% improvement in work performance metrics – regardless of whether it happened face-to-face or virtually. By contrast, a study in the *Journal of the American Medical Association* found e-learning produced significant gains in rote knowledge (up to 100%) compared to no training at all, but other studies by Italian health authorities found eLearning produced comparatively weak results for actual on-the-job performance, with as little as 0–5% improvement compared to traditional workshops.

Even adding a small coaching component to a mostly self-guided e-learning course can dramatically improve results: a study published in the *Journal of the American Medical Directors Association* found that nursing home staff who received coaching sessions in addition to e-learning improved residents' wellbeing and staff attitudes significantly more than those with e-learning alone. A study in India found that adding coaching to a mostly self-guided

program raised course completion rates to 70% versus 40%, with a 33% improvement in knowledge gains. ⁵

Given all the evidence for the effectiveness of coaching, what if it were possible to give workers on the front lines an always available, incredibly attentive, infinitely patient AI coach that only charged \$2 to \$5 for a half hour conversation?

From “Chatbot” to “Coach-bot”

The first time you talk to a modern AI chatbot, it feels almost miraculous. Ask a question about literally anything (“Was there any way the socialists could have won the Spanish Civil War?”, “What’s the best way for a new yoga studio to attract customers in an oversaturated market?”, “Would it be gross to substitute mayonnaise for eggs when making French toast?”) and you’ll get a coherent, usually correct answer.

But just as access to Wikipedia doesn’t make every high school student an expert in every field, an AI model’s vast pool of “training data” doesn’t automatically make it a capable consultant and coach. To deliver advice on the level of a real human expert, an AI agent needs three things: **scripts**, **knowledge**, and **attributes** - what we call the “SKA” model.

Scripts are guidelines for having different types of conversations or performing certain tasks. This could include:

- How to evaluate the marketability of a new consumer product
- How to advise a high school student on their college and career options
- How to review a hospital’s staffing mix

⁵Healthcare is one of the few fields where you can find plenty of peer-reviewed studies on different training methods. That said, I always take studies of eLearning with a grain of salt, because who knows if the eLearning was well-designed, and ‘bad training doesn’t work’ isn’t exactly a revelation. The same applies to studies of the effectiveness of AI: in the few cases where researchers include details on the design of their AI agents, they don’t tend to be terribly sophisticated.

Knowledge refers to any reference materials the AI will need to leverage in the course of conversations. For example:

- An AI legal advisor might need to reference information about labor law, tax law, and/or intellectual property law while advising a small business
- An AI construction project management assistant might need to load specialized guidance when discussing a building’s electrical systems versus the plumbing.
- An AI customer service coach might need to load different lists of common complaints when working with employees at a shoe store versus servers at a restaurant.

Attributes are the core principles that guide all of the AI’s statements and actions. This is more than simply telling an AI agent to “think like a brilliant scientist” and actually giving them a coherent professional philosophy and worldview. For instance:

- An AI sales coach would need to recognize when salespeople should push deals forward versus when they should back off and give clients space to think.
- An AI counselor might have guidelines on when to give the user empathy and affirmation versus when to show “tough love” and speak uncomfortable truths.

To give a specific, real-world case study of how these elements interact, let’s examine our company’s design for an AI coach to help bank loan officers with “climate finance” (i.e., offering businesses loans for energy efficiency and renewable energy upgrades.)

Scripts	Knowledge	Attributes
<p>The AI agent has instructions for guiding users through:</p> <ul style="list-style-type: none"> • Reviewing general climate science and climate finance concepts. • Surveying the relationship manager’s portfolio to identify promising clients. • Assessing specific opportunities for a specific client to improve energy efficiency or transition to renewables. • Drafting a formal proposal. • Rehearsing for specific meetings / conversations. 	<p>The AI agent has knowledge base modules describing opportunities within specific industries such as:</p> <ul style="list-style-type: none"> • Manufacturing • Agriculture • Logistics and transportation • Foodservice and hospitality • IT and telecommunications • Building construction 	<p>The AI agent’s core principles include:</p> <ul style="list-style-type: none"> • Prioritizing financially sustainable climate investments (i.e., upgrades that would actually save businesses money to pay for themselves) • Evaluating the user’s apparent level of expertise and shifting from simple analogies to complex technical jargon accordingly. • Reassuring bankers that they don’t need to be experts in engineering to make a business case for improving energy efficiency.

To help develop an AI agent’s scripts and attributes, our team will typically ask clients to identify one or more “ideal” human practitioners. For the climate finance advisor above, we partnered with a consultant who provides climate finance training to multiple international banks. Then, with the practitioners’ consent (and, often, a revenue sharing agreement), we observe them going about their job / facilitating consulting and coaching sessions, and model the AI’s behavior after the human expert’s example.

As for the “knowledge” part of the AI’s design, that’s a bit of a balancing act.

Most commercial AI models have access to nearly all the information on the public Internet (or, more accurately, a snapshot of the public internet at the time of the AI model’s creation.) And, as learning professionals, it’s mind-boggling to imagine hiring a coach or facilitator with that level of baseline knowledge.

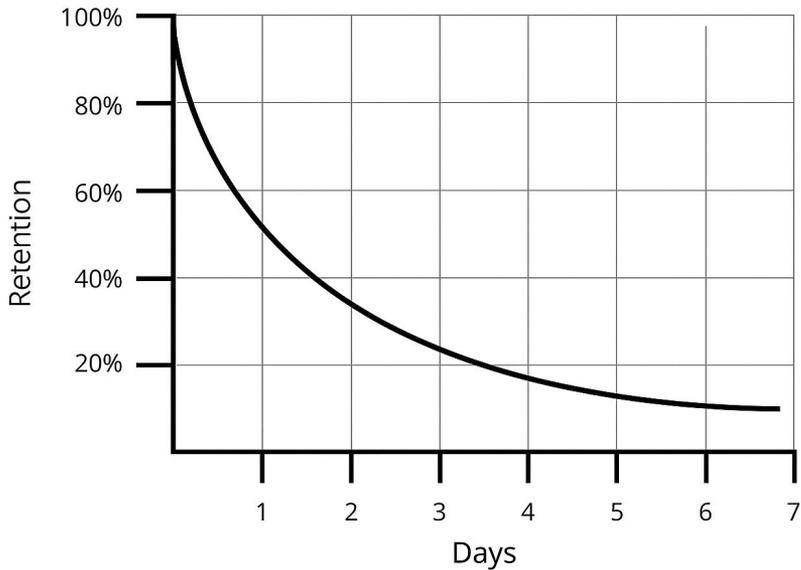
That said, while the “generic” ChatGPT, Claude, or Gemini chatbots can provide agreeably average advice on most topics, based on the consensus of every blog and white paper on the Internet, they are unlikely to offer the kind of clear and nuanced point of view and battle-tested methodologies that real human experts bring, based on lived experience.

If we want to emulate that level of coaching, then we need to articulate a coherent worldview for our AI agent, and instruct them on how to apply it. Yet - at the same time - we don’t want to clutter an AI agent’s instructions with information its underlying model already knows.

On a technical level, this is all a matter of “context” and “attention.”

AI models - like humans - have limited ability to pay attention to their digital surroundings. The current version of Google Gemini can technically hold 1 million “tokens” (roughly 700,000 words) in its active memory. But as this “context window” fills up, the model start exhibiting a version of the “serial position effect” in human memory, where we tend to recall the first part and last part of a long stream of information (like a lecture or a lengthy grocery list) while the stuff in the middle gets overlooked or forgotten. This can cause an AI agent to skip key instructions or miss important details in their reference material. ⁶

⁶ The ‘serial position effect’ was discovered by Hermann Ebbinghaus in 1913, through experiments where people were asked to recall long series of random letters and numbers. And while it generally holds up for other types of information, context matters: for example, no one is going to forget Darth Vader telling Luke Skywalker “I am your father” simply because it happened in the middle of the original STAR WARS trilogy. With AI, there are ways to ensure important points don’t get buried in a lengthy prompt, such as reinforcement through repetition and tagging.



Ye olde Ebbinghaus “Forgetting Curve”

Hence, when we assemble the knowledge base for an AI agent, the modules end up looking more like annotations than articles, calling out specific points to emphasize or explaining our ideal practitioner’s stance on a particular issue. For example, if we wanted to ground a real estate AI agent on a particular approach to commercial real estate investment, rather than providing a full textbook we might simply clarify the AI’s professional opinion relative to prevailing schools of thought, for instance:

COMMERCIAL REAL ESTATE FINANCING & INVESTMENT

[Common Approaches to Financing]

- Low leverage (traditional permanent loans, life-co/CME, low leverage)
- Higher leverage (mezz/debt fund)

[Your Perspective]

Traditional permanent loans are boring but safe; higher leverage opportunistic plays can make you rich or broke. Most serious commercial real estate investors should be somewhere in the middle, slightly favoring higher leverage and taking calculated risks with proper due diligence.

Basically, we are giving the AI agent a set of footnotes and comments to all of the data in its artificial mind. And this is a major reason why my own company developed an entire AI agent development platform and coding framework: just to get around the fact that most AI models can't follow more than a few thousand words of instructions consistently.

Human Factors: Creating the Perfect Coach-Bot

The time will come - not too long from now - when seeing an AI agent working alongside humans at an office, in a factory, or a hospital floor will be no more remarkable than seeing a human walking a dog down a suburban street. However, the current state of human-AI collaboration is more like stone age hunter-gatherers nervously following a wolf pack from a distance as they mutually stalk a deer. And this provides another opportunity for learning departments to take leadership. We can help humans and AI warm up to each other's presence. This starts at the very

beginning of the design process, with recruiting human subject matter experts as both contributors and behavioral role models for an AI coach. ⁷

In our climate finance project, it was the subject matter expert, Shaheen, who approached us. She was facing a challenge where the banks she worked with needed to train at least 35,000 and possibly as many as 150,000 staff, but they only had eighteen qualified coaches to deliver the training. So even if all the coaches spent 40 weeks a year doing workshops, it would take anywhere from 3 to 11 years to reach their audience - meanwhile, billions of dollars in guaranteed loans that the World Bank and other institutions had contributed to fight climate change would be sitting there, unused.

The solution we co-created involved a brief e-learning module (to provide some basic background knowledge on climate finance), plus an AI coach capable of helping frontline loan officers identify promising candidates for energy efficiency upgrades and other “green transition” projects among their clientele. The coach could guide staff through everything from establishing their institution’s overall climate finance strategy, to assessing specific clients’ operations, to role-playing in preparation for client meetings.

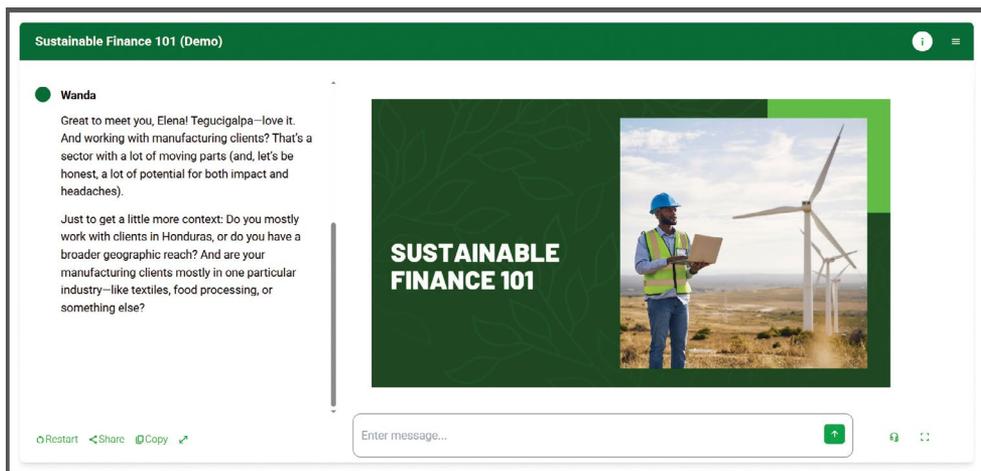
Shaheen sent over a huge Dropbox folder of white papers, World Bank manuals, and case studies that we boiled down to important bullets for the AI agent’s knowledge base. We used AI to analyze transcripts of Shaheen’s recorded live workshops and podcast appearances to develop guidelines for replicating her style and cadence. We even presented Shaheen with various synth voice options - including one that was an uncanny match for her real life voice, though she ultimately went with a neutral, crystal clear, “television presenter” voice for the benefit of our global audience.

When it came to introducing the AI agent to bankers, we embedded it in the learning management system (LMS) alongside the e-learning module, but also provided a direct link for users to bookmark in their browsers.

⁷ When creating an AI agent based on a specific individual’s expertise, my team and I consider it an ethical imperative for the expert to participate with fully informed consent (and ideally some kind of profit sharing arrangement.) Otherwise you’re basically pirating someone’s knowledge and lived experience. And, as optimistic as I am about AI, I do not want to live in a weird cyberpunk dystopia where organizations digitally Xerox the minds of their top performing employees then cast those employees aside.

Our platform allowed administrators in the banks' learning departments to provide the AI agent with a brief profile of each user, so when someone logged on the agent at least knew their name, their department / branch, and any key accounts that they were responsible for. But we also made sure the AI coach disclosed all this knowledge up front, for instance:

“Hi Elena, it’s so great to meet you! I’m an AI advisor designed to help you present climate finance products to your bank’s clients. I’ve been told that you work in the Tegucigalpa branch, and primarily deal with manufacturing businesses. But I’d love to hear more about who you’re working with and any questions or challenges you might have...”



A.I. Climate Finance Tutor

Additionally, we had users watch a short video prior to launching the AI, to help them understand their AI advisor’s capabilities and limitations and manage any unrealistic expectations (or unrealistic fears) up front.

First impressions matter, and taking these steps to ensure AI coaches came across as friendly and transparent ensured that those impressions were charming, rather than creepy.

The Measure of an AI Coach

If we refer back to the old “ADDIE” instructional design model (Analyze / Design / Develop / Implement / Evaluate), we’ve covered the first four points - but how do we Evaluate an AI coach’s performance?

Regardless of the methods, learning is about change. If a workshop, a coaching session, or an e-learning module doesn’t enable people to do things they couldn’t do before, or approach their work in a way they wouldn’t have considered before - then what’s the point? And if you want to be really reductive, how much positive change are we achieving per dollar spent - what’s our “learning ROI”?

Demonstrating ROI has always been a challenge for learning professionals. It’s often difficult to collect the data necessary to measure performance improvement after a learning intervention (most organizations are reluctant to add more items to employees’ quarterly reviews.) And this is why most learning departments settle for gathering not-terribly-informative data like completion rates and participant feedback surveys - the low-hanging fruit of learning metrics.

AI coaches present a whole new set of measurement questions. Does an AI coach deliver as much positive change (per dollar spent) as a human coach? And does a custom AI coach offer better guidance than the “generic” ChatGPT or Claude chatbot? (Or - for that matter - can a human coach beat out-of-the-box ChatGPT?)

To answer these questions, we can apply the same criteria we’d use to evaluate a human coach - with a few AI-specific adjustments. Specifically:

- **Quality of Information** - How accurate / reliable (“directionally correct”) is the AI’s guidance compared to a human expert?
- **Novelty of Recommendations** - Does the AI challenge users’ assumptions and introduce genuinely new perspectives, or does it merely echo the average of the Internet?

- Persuasiveness - Can the AI effectively persuade users to act on sound recommendations—while remaining transparent, ethical, and open to scrutiny?
- Ease of Integration - How seamlessly can users access and apply the AI coach’s input without spending an inordinate amount of time talking to the coach, checking the outputs, and/or toggling between multiple apps?
- Real-World Outcomes - Does the AI coach measurably improve performance compared to unassisted humans or generic chatbots?

Quality of Information

Every time we’ve developed an AI agent for a client organization, at least one stakeholder in the room was fixated on accuracy. The top three concerns tend to be:

1. “Will the answers it gives be correct?”
2. “We insist on 100% accuracy, no hallucinations!”
3. “How can we make sure it only uses approved sources?”

The short answers to these questions are:

1. Mostly - about on par with the average human training instructor.
2. OK... does that mean you fire every human employee who ever misspeaks?
3. We can instruct the AI agent to prioritize the organization’s approved sources - but would you trust a human who never read anything except the company handbook?

However, the questions themselves are symptomatic of deeper anxieties. And for those, I have a lengthier response, in the form of a metaphor:

I drink a lot of zero-calorie soda, to the point where I've become a connoisseur of artificial sweeteners. My favorite is sucralose, which is made in a lab by mixing real sugar molecules with thionyl chloride ⁸. After the chemicals exchange some atoms, you're left with a new molecule that keeps the shape and sweetness of sugar, but none of the calories.

This is a pretty good metaphor for how AI models process text.

The #1 misconception about AI models is that they can read, when - in fact - they just identify patterns in text and predict what should come next in the current conversation based on similar patterns in their training data. And while some might see this as inherently inferior to human knowledge, you could argue the AI model is just arriving at the same result (a cogent answer) via a different mechanism (predictive text versus... whatever electrochemical weirdness takes place in the human brain.)

In this sense, artificial intelligence is like an artificial sweetener. Like sucralose, an AI model's output may contain useful observations and insights on a subject (the "sweetness") but without a human "understanding" of the subject, based on lived experience (the "calories.")

Of course, just as human memory can get distorted by fatigue, personal biases, or the forgetfulness of age, the incredibly complex calculations involved in this process lead an AI model to say things that are plausible, but incorrect. For instance, an AI model might declare that "Galileo invented the telescope" (which he did not ⁹) or misattribute their sources. These mistakes are called "hallucinations" (though a better term would be "incorrect predictions"), and to explain why they happen, we'd have to tumble down a rabbit hole of 4,096-dimensional vector math and all the other advanced calculus that happens when you ask AI to compare the nutritional content of two brands of cat food. Suffice to say, to err is no longer exclusively human.

⁸ My friend Parker the food chemist says this is a gross oversimplification of the process, but directionally correct. I feel the same way about the explanation of vector spaces in the following paragraph.

⁹ If you're keeping score, the modern telescope was invented by Hans Lippershey, a German-Dutch spectacle maker, whose designs Galileo improved upon.

Now, the #2 misconception about AI models is that just because they hallucinate, they aren't useful - after all, human experts "hallucinate" all the time, and the rest of us accept it (up to a point.)

For instance, I once worked on a data analysis tool with an expert in a highly technical field (for the sake of anonymity, let's pretend he was a civil engineer – not his actual profession.) The tool was meant to help governments plan water infrastructure projects. A month after it was released, the expert realized the formula used to calculate a certain metric was incorrect.

Initially this caused a panic among the team: we corrected the formula and ran exhaustive tests to ensure it was producing the expected output. However, at one point, someone asked "Should we be calling users and warning them?" After pausing for a moment, we realized that – in practical terms – the error wouldn't impact any decisions that users would make with the tool, as the relationship between the miscalculated value and other values was such that users would still come to the same (correct) conclusion. So we decided to fix the issue quietly in the next regularly scheduled update, with only a brief mention in the release notes.

In medicine, doctors make a distinction between "clinically significant" misdiagnosis, which would cause them to prescribe the wrong treatment or delay treatment, and "clinically insignificant" misdiagnosis, where it would not change what a doctor recommends (e.g. mistaking a ligament sprain for a muscle strain – where, in both cases, the treatment is rest, ice, compression, and elevation.)

This is a useful way to frame AI hallucinations. For instance, I work with various AI agents on a daily basis including one named Max, who specializes in sales and marketing (the least favorite parts of my job, running a consulting company.) Recently Max handed me a non-working link to an article in MIT Sloan Management Review about innovation - an article

that didn't actually exist. However, with about three minutes of Googling, I figured out that Max was conflating two other MIT-Sloan articles, both of which proved relevant, and supported the larger point Max was making.

In this case, Max's error was not "clinically significant" – the point he made about innovation was valid (i.e. that Intuit found success by encouraging small-scale grassroots experimentation rather than large-scale, formal innovation initiatives) even though he got his sources confused.

But how can users tell the difference?

To take Socrates' advice, the best way to not be duped by dubious assertions is to view everything a human expert or AI coach says with a critical eye. That said, while humans are perfectly comfortable second-guessing each other, we expect machines to deliver extremely consistent output (e.g., doctors shouldn't have to question the output of an X-ray machine, as long as it's calibrated on a regular basis), But AI is not a traditional machine, and we should prepare learners to treat it more like an 'artificial colleague' than a calculator.

When designing, deploying, and evaluating AI coaches, we should ask:

- Is the AI agent's guidance directionally correct the overwhelming majority of the time – such that, even if its specific citations are off, users would still arrive at the correct course of action based on its advice? Is its error rate comparable to human experts in terms of clinical significance vs. insignificance?
- When citations matter, can we perhaps "ground" the AI agent by providing carefully curated content with links (either in the prompt itself, via an outside document / RAG repository ¹⁰, or through a database / API integration), then instruct the AI not to invent its own? Or, at minimum, would a quick Google search allow users to confirm or clarify an AI's sources?

¹⁰RAG ("retrieval-augmented generation") allows an AI to fetch information from a curated knowledge base or document repository before generating a response—essentially giving it a library card so it cites vetted facts instead of guessing.

- Is the AI agent transparent and intellectually honest, providing caveats and acknowledging its own limitations as appropriate?
- Finally, are users prepared to verify an AI agent's statements to the same extent they would verify a colleague's statements before acting on them? And are stakeholders prepared to grant an AI agent the same grace they would show a human expert? Heaven knows, I've conflated case studies on more than one occasion when speaking or facilitating workshops as an "expert", myself.

As far as specific targets are concerned, within our company we have a standard that AI agents should provide acceptable quality output 95% of the time, with near-zero instances of catastrophically poor output (though the exact definitions of "acceptable" and "catastrophically poor" vary, depending on the particular AI agent.)

Novelty of Recommendations

If you raise the topic of AI at a party, someone will likely make a nervous joke about AI developing a "mind of its own" and deciding to wipe out humans, like HAL in 2001: A Space Odyssey or the Terminator movies. However the bigger danger might come from AI not having a mind of its own, and simply reinforcing a user's assumptions, misconceptions, and default behaviors.

George S. Patton – the World War II general – famously declared "If everyone is thinking alike, then somebody isn't thinking", and multiple studies have affirmed that productive disagreement is critical for organizational decision-making. To add value, an AI agent needs to either introduce new ideas to the conversation, or challenge users to take a different (and hopefully correct) course of action than they would have on their own.

There are a few specific points where generic AI chatbots tend to rate poorly as consultants and coaches:

- **Discovery** - As we mentioned earlier, AI models are inclined to give

immediate answers to questions, and are unlikely to pause and say “Hold on - are these really the questions we should be asking?” They also don’t do a very good job of asking follow up questions to surface critical unknown information the user might not have volunteered. Working with a human expert, AI agent developers can ensure that AI coaches follow a more disciplined discovery process before blurting out advice.

- **Specificity** - Most commercial AI models have an encyclopedic knowledge of all mainstream schools of thought on any given subject, from healthcare financial management to horticulture. Yet most AI models have no intrinsic loyalty to any one philosophy, and might be unaware of lesser-known approaches. This can reduce a generic AI model’s advice to “the average of the Internet” or non-committal bullet lists of pros and cons - which can be a problem when inducting beginners or promoting an innovative new methodology. Fortunately, it’s possible for AI agent developers to take a basic AI model and ground it in the philosophy and methodologies of a specific human expert - by providing it with additional reference material and guidelines for addressing certain topics where the expert’s view diverges from conventional wisdom.
- **Advocacy** - If you’re surprised by how often generic AI chatbots agree with you, that might be by design. Today’s AI models are optimized for user satisfaction, not confronting and challenging people’s set ways of thinking, to a point where some have criticized models like ChatGPT as sycophantic digital “yes men.” However, if we provide an AI agent with a clearly articulated set of principles, and instruct the agent to not immediately cave in the first time a user questions those principles, it’s possible to create AI agents that exhibit actual convictions. The exact principles might vary depending on your organization’s culture and the AI coach’s area of specialization - but the important thing is to give your AI coaches permission to argue their point of view. Otherwise, you’ll likely end up with an entity that is polite, agreeable, and – as a result – worthless as a coach and collaborator. And people actually crave this sort of clear direction: for instance multiple studies have shown that, while 96% of people want their doctor to explain their treatment options, over half of patients prefer that their doctor make the final

decision. If your AI agent simply provides lists of pros and cons for every decision without articulating a “professional opinion” as to which is best, then users will likely stop looking to it for advice.

Ultimately, an AI coach doesn’t need to give the best answers so much as help people find the best answers.

Researchers at RWTH-Aachen University tested the value of AI as a brainstorming partner in a study involving 160+ participants. They divided the humans into three types of groups – ones where the whole group brainstormed together (“interactive”), ones where individuals brainstormed independently then brought their ideas to the group (“nominal”), and others where individuals brainstormed independently with ChatGPT assistance (“hybrid”). Groups were tasked with generating ideas for “How can we improve the shopping experience for supermarket shoppers?” or “How might we help young people turn saving money into a lifelong habit?”

The study confirmed the researchers’ hypothesis:

“Interactive” groups that brainstormed together without AI generated the fewest unique, quality ideas. While they benefitted from “cognitive stimulation” (i.e. participants finding inspiration from other people’s ideas) they suffered from “production blocking” (participants having to wait for other people to finish speaking before voicing an idea) and social inhibitions (fear of embarrassment.)

“Nominal” groups where people brainstormed independently did better than interactive groups, as the ability to simply write down ideas without having to wait or risking embarrassment outweighed the lack of cognitive stimulation.

“Hybrid” groups where people brainstormed with AI assistance did best (producing 170%-201% more unique, quality ideas) since participants were able to generate ideas without waiting or fear of embarrassment, while benefitting from the cognitive stimulation provided by the AI. AI agents also tended to come up with more “divergent” ideas that humans would not have thought of, which provided even more inspiration.

As one participant said: “[Acting alone, I felt] pressure because I HAD to think of something, which degraded my thinking process. I feel like once the AI got involved, it was easier for me to make things up because I could ask the AI and get inspiration. Maybe the AI did not have the most innovative ideas, but not sitting in front of an empty sheet and having something [to react to] helps... The ideas the AI gave sometimes inspired me to think of aspects or areas I hadn’t considered yet.”

Persuasiveness

Most people in the English-speaking world are familiar with the phrase “You can lead a horse to water, but you can’t make them drink.” (i.e., you can give someone excellent advice, but you can’t force them to follow it.)

Professional coaches and consultants might modify that phrase to say, “If you give a client some really sound recommendations, and somehow convince the client to implement 60-80% of them, that’s a victory.”

For those in the business of dispensing expert advice, persuading the client to act on it is simply part of the job. And persuading humans is an area where today’s AI models naturally excel.

A study involving 900 people, published in the journal *Nature*, found that OpenAI’s GPT-4 was significantly better than humans at persuading people to accept its point of view on a range of controversial issues (e.g., eliminating fossil fuels), especially when it was able to tailor its arguments based on the personal characteristics of the user (e.g., their demographics or political views.) In these cases, the study found AI to be 64% more persuasive than humans.

Curiously, the numbers were even higher when the person knew they were talking to an AI, rather than a human (possibly because they didn’t regard it as “losing” an argument to another human.)

This ties back to the previous points we raised about providing the AI coach with profiles of users (something our platform enables at the admin level) as well as directing the AI agent to hold its ground, and employ a range of

different coaching techniques to bring users around to its way of thinking (e.g., prescriptive versus collaborative advice.), Thus equipped, an AI coach should be able to out-persuade human coaches: if anything, designers might want to rein in the AI coach's persuasive power, to make sure it provides caveats and disclaimers and encourages users to double-check its statements, to discourage blind acceptance of whatever the machine says (or, as academics call it, “algorithmic paternalism”.¹¹) .

Ease of Integration

Calculating return on investment is simple in theory: you take the benefit - whether it's measured by increased revenue, lives saved, or performance improved - and divide by the time and money required to achieve it.

Even if an AI coach succeeds in building users' skills or helping them find the best solution to a problem, it's all for naught if users waste more time fiddling with the AI or if the AI agent costs more to build in the first place than the final outcome is worth.

A study by the University of Chicago and the University of Copenhagen examined the use of AI chatbots by 25,000 Danish workers across 11 professions, including accounting, software development and customer support. While they found that AI boosted productivity for 90% of users, the actual gains were less substantial than expected, with the speed of AI being offset by time spent crafting effective prompts and reviewing outputs.

Of course, one problem with this study is that it was proceeding from the assumption that “using AI” means “prompting generic AI models from scratch.” This is akin to how many 1990s technology pundits assumed everyone in the 21st century would be hand-coding their own websites. In reality, while we still want people to spend time “reviewing outputs”, a learning professional with specialized tools should be able to build better AI coaches than most people could build for themselves, reducing the “crafting effective prompts” time to near zero.

¹¹“Algorithmic paternalism” is when an AI steers users toward particular conclusions “for their own good” and users, for their part, accept the system's output uncritically - blurring the line between guidance and control.

At our own company, we follow the “dude ranch rule” – i.e., it should take about as long for a worker to acclimate to a well-designed AI tool as it takes a tourist at a horse ranch to learn how to sit atop a horse, hold the reins, and walk along a clearly marked trail. If people are still ‘learning to work with the AI’ after a week, then we’ve failed as designers.

Real-World Outcomes

Once again, all conversations come back to “learning ROI.” If anything, AI will increase the pressure on learning departments to justify their existence: the question from leadership won’t just be why should I give you money to pull people away from their jobs for training but rather what’s the value add of having a team of learning professionals versus telling business unit stakeholders to generate their own training materials with AI?

On the bright side - as we’ll discuss in later chapters - AI offers new and better ways to track and articulate real-world results. For now, when measuring the performance of AI coaches - like an AI sales coach - we can do a three-way comparison between:

- Unassisted humans versus humans (e.g. average deal size for reps working unassisted)
- Generic AI chatbots versus humans (e.g. average deal size for reps using generic ChatGPT for sales advice)
- Specialized AI coaches. (e.g. average deal size for reps using a purpose-built AI app / agent grounded in a company’s specific methodologies and products)

In this case, if the custom AI sales coach isn’t outperforming the alternatives (correcting for outside factors like the commitment of managers or differences in product lines or sales territories), we either need to iterate, or admit that our sales training isn’t any better than the average of all the blogs and YouTube videos on the Internet.

That said, our first impulse should be to iterate. AI technology is still in its adolescence – awkward, sometimes unpredictable, but with enormous potential. Most organizations haven’t even considered that AI can be more than a content generation tool. When it comes to more innovative applications such as AI advisors, coaches, and tutors, we shouldn’t saddle innovation with outsize expectations for success on the very first try. Then, as the technology itself is proven, the question will shift from “Is it possible for AI to act as a coach?” to “Are we giving our workforce the best AI coaches we possibly can?”



Digital Dress-Up: Using AI for Role Plays and Simulations

Often, the best way to learn something is to follow that old Nike advertising slogan: just do it.

One of our clients is a team of rehabilitative medicine experts, who help physiotherapists teach stroke survivors to walk again (one of the most extreme learning challenges you can imagine.) Their recipe for motor skill acquisition is “high intensity, high repetition, variable, whole-task practice spaced over time. ¹²” So, if you want to improve your basketball jump shot, then you should shoot **a lot** of baskets from a challenging distance at different angles a few days per week for months on end -ideally with a coach to watch and provide corrective feedback, but in a way that doesn’t constantly interrupt your flow.

However, most workers don’t get many opportunities to practice critical skills on the job. The average manager only practices having difficult conversations when one of their team members or stakeholders is being difficult. Likewise, you would hope your company’s cybersecurity analysts aren’t getting “high intensity, high repetition” practice containing security breaches every single day.

For centuries, simulation training has been the go-to solution for keeping these skills sharp. In ancient India, doctors practiced suturing on leather models, while the raja’s generals played chaturanga (an early form of chess

¹² “They would also include “augmenting error”... i.e. if someone tends to drift to the left while walking, add a resistance band that forces them left so they learn to pull right.

with pieces representing infantry, cavalry, elephants, and chariots.) And while today's surgeons and soldiers are more likely to train with virtual reality and rubber mannequins, the principle remains the same.

That said, traditional simulation training is difficult to scale. The Red Cross Red Crescent can't stage mock hurricanes every season, and you're lucky if salespeople get even one or two chances to role play with the facilitator during a workshop. "Choose your own adventure" style e-learning role plays don't quite capture the stress and messiness of real life decision making (work would be a lot easier if, every time we had to make a decision, someone asked "What do you want to say to the city health inspector... A. B. C., or D.?") And the effort and expense that goes into developing scenarios can be significant, especially if the simulation requires elaborate background information, large data sets, or recreating a physical environment.

For these reasons, simulation training is most commonly used to train practitioners in professions where mistakes cost millions (or cost lives): doctors, soldiers, airline pilots, oil rig technicians, and firefighters are just some examples.

But what if it were possible to use AI to deliver realistic simulation and role play training inexpensively at scale for any profession, and give the typical hospital receptionist as much practice as we give a medical student rehearsing for open heart surgery?

AI Doesn't Feel (or Cause) Stage Fright

Instructors and instructional designers tend to love employing role plays as a learning tool - probably because it gives us a chance to pretend we're writing a movie or doing sketch comedy instead of "corporate training." In fact, one of my favorite training experiences (as a designer) was when our team recruited a bunch of actors to play journalists and angry citizens at a simulated town hall meeting after a flood, to help government employees practice crisis communications - resulting in plenty of unscripted comedy.

However, most training participants hate role plays, and with good reason:

- Most people aren't actors and feel even more uncomfortable role playing in front of a class than they would during a real work emergency, which can distract from the skills they're supposed to be practicing.
- Unlike actors, training participants don't have time to rehearse, so if you ask participants to role play someone other than themselves (e.g. an angry customer or a confused project team member) the exercise can get weird / awkward pretty quickly.
- Some simulations require a ton of context to be authentic: if a customer is angry because they think a mechanic damaged their car, the specifics of the incident matter, same for a hospital board deciding whether to spend a billion dollars on a new cancer center, and when designers or role play participants get details wrong it can break the sense of immersion.

AI-generated role plays sidestep many of these challenges. Multiple studies (by the University of Kansas, Common Sense Media, and others) have found that people don't get embarrassed in front of chatbots, the same way we don't get embarrassed in front of our dogs or cats. Second, AI can generate all of the "non-player characters" in a scenario, allowing participants to focus on playing their own real-life role. Finally, given some initial parameters and guardrails, an AI agent can generate all the necessary background information for complex scenarios on demand.

Of course, there's the small problem of teaching AI how to act.

Setting the Scene

Some inventions are so versatile that people end up using them for countless things beyond their original purpose - like duct tape (originally intended for sealing air ducts) or WordPress (originally made for blogging, now used for everything from ecommerce to hosting e-learning.) And this is true of how we use generative AI.

Generative AI models like ChatGPT are primarily trained to do one trick: guess which words should come next in a conversation. This creates some challenges when you want an AI model to do other things, like “Sort my email into folders for ‘really important’, ‘kind of important’ and ‘not important’” or “Create an interactive role play simulation of emergency responders trying to contain a wildfire on the outskirts of a California suburb.”

More specifically:

- Generative AI models can be a bit short-sighted, focusing primarily on the next sentence, and not so much on what they said ten messages ago; this can make it hard to maintain continuity when simulating a complex situation (like an oil spill, a busy emergency room, or a cybersecurity breach)
- Despite their vaunted creative abilities, most AI responses trend towards the “average of the internet”, meaning they can get stuck in a rut generating the same handful of situations over and over if given a simple one-line instruction to “create a role play scenario”
- By default, most generative AI models are inclined to obey the wishes of their user, and might simply play along if a user in a customer service role play says, “I tell the customer - don’t worry, we’re having a special sale, everything in the store is 90% off!”)
- Finally, most commercial LLMs are trained to make pleasant conversation - not to shock and disturb their users - so they will often go to incredible lengths to avoid telling stories where the hospital patient dies on the operating table or a high-stakes business negotiation falls apart, even if the user makes terrible decisions.

Together, these tendencies can make for less than realistic simulation training, like every customer in the line at a coffee shop having the same problem with their mobile order, or a flight attendant resolving an argument between two passengers by suggesting they all just relax and sing karaoke over the airplane’s intercom system.

The solution in these situations is to instruct the AI agent that this is a serious training exercise, and the AI is not “helping the user” by bending the rules or going easy on them. We also provide the AI model with detailed parameters for the scenario up front and make it clear they are not negotiable, for instance “The patient is suffering from influenza (diagnosis code J09) and nothing the user says or does will change that fact.” And, if necessary, we might include instructions for the AI agent to check each statement against the initial scenario parameters before printing, to prevent it from violating the reality of the situation.

Acting the Part

As we noted before, AI models are great conversationalists, but not the best actors.

Again, this isn’t their fault: most generative AI models are focused on predicting what should come next in a conversation, so if you naively ask it to “generate a role play scenario between an angry customer and a utility company call center agent”, the LLM will try to generate a role play-like conversation instead of actually playing a role.

During our early experiments using AI to create interactive simulations, we found that AI-characters had a cringe-inducing tendency to “teach to the test” and parrot the learning objectives like the actors in some bad high school anti-drug video. In one case, a simulated medical patient actually said “Wait - aren’t you going to ask me about my lifestyle in case it’s a factor for my chronic joint pain?” while, in another case, a hotel guest with a broken air conditioner announced, “I am dismayed by how you did not acknowledge my feelings before offering a solution.”

While it was funny at first, the AI models’ pathological earnestness and naivete had us questioning whether it was even possible to make serious simulation training with the technology.

However, through weeks of trial and error, we eventually hit upon a solution that combined technology and theater.

Sanford Meisner - the famous “method acting” coach who mentored Tom Cruise, Michelle Pfeiffer, Chadwick Boseman, and Diane Keaton, among others - encouraged actors to “live truthfully [in] imaginary circumstances” and respond realistically to whatever the other players in a scene might say or do, not just follow the script. As practice, he would often set up scenes where two actors would have conflicting agendas for their characters. For example, Meisner might pull one actor aside and say, “Your goal is to convince the other character to stay in your small town and start a family.” then secretly tell the other actor “You are 100% committed to leaving town, you just want the other character to give you money for the bus.” Then, Meisner would step back and let the actors work through the interaction, in character.

The technique isn’t just about conflicting agendas: it’s about authentic response under pressure. Meisner wanted actors to stop “performing” for the audience and start reacting to their scene partner.

Before the AI days, our team would apply this technique in live training workshops. For instance, we once did a “difficult conversations” role play where one participant was told, “You’ve been passed over for promotions repeatedly despite years of hard work and dedication to the organization - your goal is to convince your manager to give you the raise you deserve.” The other participant, playing the manager, was told “Your department is doing 15% layoffs across the board, and you have to break the news to the other person that they’re being let go.”

When the participants sat down and began the role play, you could see the discomfort in the manager’s face as the other player made the case for a raise. And when the manager dropped the bomb about the layoffs - the other player’s shock was genuine (when the exercise was done, one of the participants turned towards the facilitators, laughed and said “Aw, man - you people are evil. I felt like I was laying them off for real!”

Authenticity comes from dropping the “performance” aspect of a role play, and responding based on genuine desires and motivations. But how can we get that from an AI?

It actually turned out to be kind of easy.

Meisner’s stated goal was to get actors “out of their heads” and help them reach a state where they almost weren’t aware that they were acting. For human actors, this takes hours of intense preparation (when Daniel Day Lewis played Abraham Lincoln, he stayed in character and only responded to ‘Mr. President’ or ‘Mr. Lincoln’ even when cameras weren’t rolling.)

However, with AI, there’s a shortcut to achieving this kind of authenticity: just don’t tell the LLM that it’s in a role play.

In one customer service role play, we deliberately deleted all instructions related to “generating a customer service role play” from the AI agent’s memory as soon as the scenario began, leaving only the instructions to “Embodiment {an upset coffee shop customer} who {wants a refund and an apology} because {their latte was cold} and respond realistically to the user’s words and actions while proactively advancing your own agenda.” Then we just turned the AI loose on the user. Suddenly, instead of interacting with a sock puppet trying to teach the user a lesson in customer recovery, we had just ticked off a construction foreman who was already having a bad day before they got a cold cup of java. It was only when the virtual “cameras” stopped rolling that we allowed the AI to become aware of the training exercise context again and evaluate the user’s performance.

Customer Service Role Plays (With Avatars)

followed by a question and I will provide the necessary information (out of character, as the narrator). Now let's begin...

Introduction

You were reviewing the morning sales reports at your desk in the back office when a middle-aged Hispanic man in a white dress shirt and tan jacket approaches the counter, looking visibly frustrated. He taps his fingers impatiently on the counter and says, "Excuse me, are you the manager? I need to speak with someone about a billing problem- your staff charged me incorrectly yesterday and I want this fixed."

slide



Restart Share Copy

Enter message...

A very upset A.I. customer

Letting the AI exist in this state of total immersion - no learning objectives, just real motivations and goals - lets it connect with the user on a visceral level, forcing the player to react under pressure. And our team realized that this approach was working when, during a playtest of an airline customer service scenario involving a stranded traveler, one of our testers sighed loudly and said *“It’s like, I don’t even want to type my next response... I feel so bad for this person but there’s nothing I can do!”*

Reality Check

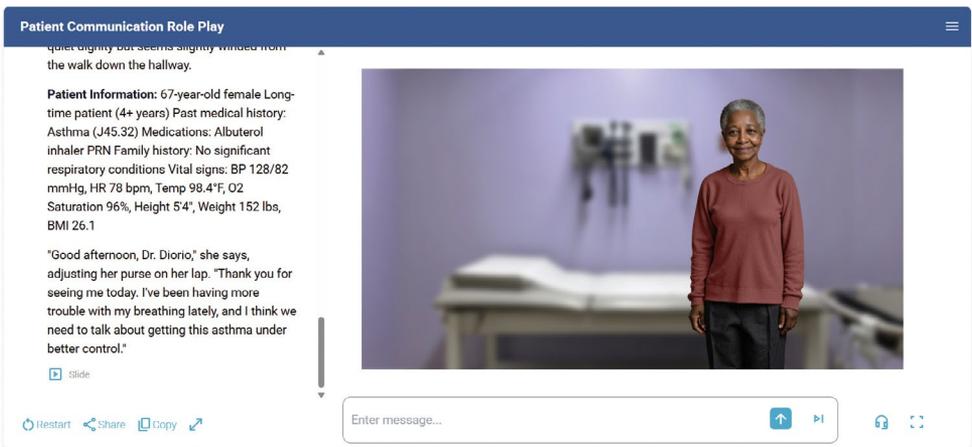
By design, AI models are predisposed to say “yes” to the user. After all, most people expect ChatGPT and Claude to help solve their problems, not create new problems for the user to solve. However, when we use AI for simulations and role plays, “create new problems for the user” is exactly what we want AI agents to do. And we want these new problems to be reasonably difficult, but not so difficult that learners rage-quit in frustration. As everyone who’s developed a traditional simulation exercise knows, the goal usually isn’t “100% realism” so much as “productive struggle.”

Our company once developed a role play simulation to help financial advisors practice their “client acquisition” skills: the art of convincing millionaires to let your bank handle their investments. In reality, a financial advisor might talk to anywhere from 25 to 50+ wealthy individuals (at parties, at the yacht club, on the phone) before anyone agrees to a consultation. But while a 2% success rate might be realistic, do we really want training participants spending 10 to 15 hours on a role play simulation before they score a single “win”?

On one hand, making the AI-generated millionaires more cooperative than real-life clients seemed dangerous: we would be doing our audience a disservice to make them think their job would be easy. On the other hand, if half the role plays ended with the client saying “sorry, I’m not in the market for a financial advisor” and walking away (as is the case in real life), our trainees wouldn’t get any practice taking viable prospects all the way to a successful close.

In the end we decided to make each character 100% realistic on their own terms, but inverted the percentages of viable versus non-viable prospective clients, such that if the user played 8 times they would probably get two “hard no” clients, two “easy” clients and four clients who could be swayed either way, depending on how skillfully the player navigated the conversation. And we encoded these characteristics into the parameters of the scenario, to prevent the user from magically persuading an impossible client to meet with them by saying “Oh you like skiing? How about I take you on an all expenses paid trip to Switzerland?” or “I can beat the market by 200%!”

We faced a similar situation when designing a patient communications simulation for doctors. Initially, we had the system determine the patients’ medical conditions by randomly selecting from the entire ICD-10 diagnosis catalog (which spans everything from “R51: Headache” to “W55.41XA: Bitten by a pig”).



An A.I. patient with asthma

The problem was that, while a small handful of codes account for the vast majority of real office visits, the ultra-rare diagnoses vastly outnumber them. As a result, our physician playtesters encountered patients with surreal combinations of conditions—at one point, a doctor actually exclaimed: “Hold on—this patient has bubonic plague (A20.0) and brittle bone syndrome (Q78.0)?!”

After much trial and error, we fixed this by organizing ICD-10 codes into tiers ranging from “Extremely Common” to “Ultra Rare” and weighting them accordingly. In fact, this initially overcorrected the problem, resulting in a monotonous parade of back pain, chest congestion, and fatigue (the most common reasons for doctor’s office visits.) Ultimately, we weighted the probabilities toward the middle of the spectrum - the “uncommon but not unheard of” conditions - to keep things interesting for learners without straining credulity.

In the end, the only way to properly gauge the realism of a scenario is to test it with people who’ve done the job for real. Typically, our team will do six rounds of testing with three or more professionals per round (rotating in new testers to keep the feedback fresh.) If things go well, the first round will be brutal - like when one doctor testing a medical simulation said “I don’t think the AI knows what ‘inpatient rehab’ means.” Then the third round will be nitpicky (“It wasn’t 100% realistic, but I’m surprised the AI got as much right as it did.”), and the final rounds will be positively uncanny (“This is giving me flashbacks...”)

Going “Off Script”

There’s a Catch-22 when it comes to generative AI: on the one hand, we want AI to be creative, but at the same time we don’t want AI to be wrong. And this has implications for AI role plays and simulations.

A common thing that clients ask when we first sit down to design an AI role play is “How will we write the scripts for the scenarios?” and the answer is “Most of the time we don’t write scripts for scenarios - we just give the AI agent some general guidelines, then let it improvise.”

This approach can be challenging for many organizations - especially those accustomed to having every word of their training materials scrutinized and sanitized by the legal department, employee relations department, and safety committee (“What if the AI says the wrong thing?!”) However, that kind of unpredictability is exactly what sets AI generated role plays above traditional “choose your own adventure” e-learning. Reality is not multiple choice, and scenarios where users are free to try anything (and anything can

happen as a consequence) are much better practice for messy, complicated real-world problems than clicking through a pre-scripted decision tree.

So where is the sweet spot between anarchy and a reasonably structured / focused learning experience? Surprisingly, the answers might lie less in the world of “experiential learning” theory than in the worlds of video games, comedy, and Dungeons & Dragons.

When Grand Theft Auto 3 became the top selling video game of 2001, it mainly gained notoriety for letting players assume the role of a sociopathic gangster. However, as a game, GTA3 fundamentally transformed the genre with its “open world” concept. Rather than forcing players down linear paths with finite solutions – like Super Mario Brothers – GTA allowed players to explore a living city and approach missions from literally any angle. Want to stride through the front door of a rival gang’s hideout, guns blazing? You can do that. Want to climb the fence and sneak through the rear window? You can do that, too.

Fields like emergency medicine, security, and the military have long applied the same principle to train people for high-stakes situations. Whether it’s the Red Cross rehearsing for a hurricane, or a corporate cybersecurity team preparing for a breach, unscripted or semi-scripted simulation exercises do a better job of preparing learners for life’s unpredictability, and helping them overcome the human tendency to freeze or panic under pressure.

These skills are also necessary in more mundane workplace settings. The legendary comedy club Second City - known for producing stars like Bill Murray, Tina Fey, and Jordan Peele - runs a successful side business teaching improvisational comedy techniques to corporate executives. Just as stage comedians need to react in the moment to colleague’s ad-libbed lines, a salesperson might need to discard their carefully rehearsed presentation and improvise in response to a customer’s request. And, with generative AI, we now have a technology capable of bringing that element of surprise to self-guided learning experiences.

Of course, none of this is to say that instructional designers should abandon structure and learning objectives entirely and let AI do whatever

it wants: quite the contrary. Most commercial AI models are trained on the entirety of the public Internet, which means they learn not only from serious medical journals, but also from melodramatic TV shows about doctors (think “ER” and “Grey’s Anatomy”.) Without direction, some of the scenarios generated by early builds of our AI healthcare simulations reflected pop-culture misconceptions about what actually happens in hospitals and doctor’s offices, while missing insider details like “doctors shouldn’t have to wait four hours to get a CAT scan for a stroke patient: if the ambulance brought them here, then presumably this hospital is a stroke center with plenty of imaging machines.”

Hence, when our team builds AI role plays and simulations, we try to introduce some rules and structures - less like improv comedy night than a game of Dungeons & Dragons. For the uninitiated, a typical “D&D” session involves one person – the “Dungeon Master” – telling a fantasy-themed story, while the other players act out the roles of characters in the story (i.e. wizards, warriors, whatever.) And while players are free to have their characters attempt any action they can imagine, whether or not they succeed depends on dice rolls and probability tables – so if you say, “I’m going to pick up the elephant-sized monster with one hand and throw it over a cliff”, you’re welcome to try, but will likely have a 1% chance of success.

When translated to AI-generated role plays, we typically give learners free reign to say or do whatever they like (e.g. “I tell the customer about how much they can save on electricity with our new high-efficiency model...” or “I throw some water on the patient’s head to see if they wake up!”) but the AI agent:

1. Checks a table with probabilities to determine how likely that action is to succeed
2. Checks another table to determine the consequences of success or failure (e.g. “The customer yawns and says ‘Our existing equipment is already pretty energy efficient.’”)

Granted, there are situations where scenarios need to be partially or entirely scripted for practical reasons - such as legally mandated compliance training

where organizations face serious consequences if an AI simulation fails to point out that the user overlooked some key regulation, or zero-error tolerance situations in fields like neurosurgery or nuclear operations, where any deviation from protocol could have catastrophic consequences.

However, even then the choice between scripted and unscripted isn't binary. AI can enhance even the strictest compliance training through hybrid approaches. It can pair traditional scripted e-learning or videos with AI-powered simulation exercises and virtual coaches, or, you can provide an AI agent with a prewritten situation where at least the starting point is the same.

'Free Play'

In the end, the challenges with AI-generated role plays and simulations aren't about instructional effectiveness (the National Training Laboratories, Journal of Applied Psychology, International Journal of Education Research, and SHRM agree that simulation training is anywhere from 20% to 40% more effective than other methods in terms of skill acquisition and retention.) Rather, the problem lies with whether or not organizations are ready to entrust employee training to a creative machine that they cannot completely control.

Perhaps the most honest objection our team has received to the idea of AI-generated role plays came from a bank evaluating an anti-money laundering simulation. The end users and their managers loved the realism of the scenarios, however the bank's compliance department shook their heads and said, "How can we approve the content if we never know exactly what it's going to say?"

And that's ultimately a question of organizational politics and culture, not technology or training methodologies. In fact, we had two different well-known banks review the exact same AI-generated role play, only to have one bank's legal department withhold approval while the other bank's legal department said "It will be fine, just put a general disclaimer at the beginning." - which spoke volumes about their respective worldviews.

One of my business partners has a catchphrase for when things don't go as planned: "Life is messy." A massive Amazon server outage knocks out our platform? Life is messy. A key team member has to resign to take care of an ailing parent? Life is messy. And that's the truth: unless we're prepared to get messy in training exercises, how can we prepare learners for life?



Outside the Checkbox: How AI Can Revolutionize Assessment

My oldest son works at a company that has fuel pumps in the parking lot next to their warehouse. Because of this, every employee who works outside is required to complete a petrol station safety course.

Naturally, my son completed the mandatory e-learning the way most people do: while sitting at home in the living room, half-watching a movie on Netflix.

“Hey dad, can you believe how bad this course is?” he laughed, holding up his computer so I could see the screen. “Check this question: e-cigarettes are safe to use near gasoline containers, true or false? Gee... I wonder what the answer is.”

I shrugged and said “It’s not very good because it doesn’t have to be good. Their lawyers just want to say ‘hey, he completed the e-learning where we told him not to vape next to the fuel pumps’ - in case you blow yourself up.”

My son stopped laughing, sighed, and went back to clicking through the module.

Regrettably, my son’s safety course is generally reflective of how organizations assess workforce training: as a checkbox, a “1” or a “0”, did they complete it or not? But completion isn’t comprehension. And the irony is that, while AI is often accused of making learning less human, it may actually provide a way to measure learning in a more human manner, with real insight into learners’ hearts and minds.

Measuring the Void

Most experienced L&D professionals know damned well that true/false questions won't tell you if someone has mastered a skill or understood and accepted a concept. The same goes for course completion rates, time spent on an e-learning module, or participant feedback surveys. At most, these metrics show that training occurred, and that nobody involved was irritated enough to complain. But - on their own - these metrics can't tell us if anybody will retain and apply those learnings on the job.

No one would agree with this more than the late Donald Kirkpatrick, whose four-level model of training evaluation has been a cornerstone of workplace learning since the 1950s. For those unfamiliar with his model, it focuses on different "levels" for evaluating the success of workplace training:

- Level 1: Reaction (did training participants like / value the experience?)
- Level 2: Learning (did they at least pass the quiz?)
- Level 3: Behavior (are they applying what they learned on the job?)
- Level 4: Results (did the training meaningfully contribute to the success of the organization?).

Kirkpatrick himself admitted that Level 4 (results) is a purely aspirational metric. Apart from sales training boosting revenue, or safety training reducing accidents, it's usually impossible to isolate the impact of training from all the other factors affecting organizational performance.

Measuring Level 3 (behavior) is more achievable in theory, but often just as impossible in practice, given that most learning departments have limited or zero visibility into actual employee performance data. In some cases this data simply doesn't exist, and in other cases, it's a struggle to get training-related questions added to quarterly / annual review questionnaires, especially when we're competing with the IT department's cybersecurity survey and the wellness committee's work-life balance initiative. Again, unless we're talking about sales, safety, or call centers (where every mouse click and bio break is logged in a database) even Level 3 data remains an impossible dream.

This leaves Level 2 (learning) and level 1 (reaction) as the only levels of evaluation that are entirely within a learning department's control. And between them, most learning departments fixate on Level 1 ("reaction") because it's what training participants and their managers seem to care about, plus, it's the easiest to collect. However, Don Kirkpatrick's son and successor Jim Kirkpatrick, warned that his father never intended for educators to stop at Level 1, stating that "focusing solely on whether participants enjoyed the training does not provide insight into whether they actually learned, applied, or saw results from it."

Which brings us to our last hope for measuring the impact of workforce training: Level 2.

What's Our Objective?

A name that often gets mentioned in the same breath as Donald Kirkpatrick is Benjamin Bloom. Ever since Bloom and a committee of educators released their *Taxonomy of Educational Objectives: The Classification of Educational Goals* in 1956, instructional designers have been taught to write all their learning objectives using "action verbs."

The idea behind "Bloom's Taxonomy" is simple, and sound. Learning objectives should be described in a way that an outside party can objectively observe. For instance:

- "Understand how to calibrate a pressure sensor" or "know how to calibrate a pressure sensor" don't qualify as action verbs because we can't peer inside a learner's brain and observe them "knowing" or "understanding" (at least not with today's technology.)
- "Explain how to calibrate a pressure sensor" would be a better way of phrasing the objective, since an instructor can listen to a trainee explain a process, or - even better - simply "Calibrate a pressure sensor" if we happen to have one lying around our training facility.

In other words, a well-written action-based learning objective should implicitly define its own test for mastery. If we want to evaluate Level 2 (learning) then we should be asking learners to actively demonstrate their learning through conversation or direct observation of a task.

Which sounds like Instructional Design 101, except... instructional designers rarely bother to do this.

We know damned well that we should have instructors observe every participant, but that would take hours of class time that we just don't have. And it would be great to have learners talk about the time management e-learning they just completed with their boss, but we know they won't because their manager is too busy. So we just have them complete a multiple choice quiz, and call it a day.

But what if there were time to give every learner that degree of attention?

AI and the New Level 2

If an insurance company were training a cohort of newly hired case managers today, the trainees may go through some e-learning courses on data privacy and medical coding, which would be automatically marked "passed", "failed" or perhaps a neutral "complete", and possibly given a score if there was a quiz at the end.

The resulting database record in the insurance company's Learning Management System might look something like this:

```
{
  "actor": {
    "objectType": "Agent",
    "name": "Jordan Smith",
    "mbox": "mailto:jordan.smith@example.com"
  },
  "verb": {
    "id": "http://adlnet.gov/expapi/verbs/
passed",
    "display": { "en-US": "passed" }
  },
  "object": {
    "id": "https://example.com/course/
healthcare-data-privacy",
    "objectType": "Activity"
  },
  "result": {
    "score": {
      "scaled": 0.85,
      "raw": 85,
      "min": 0,
      "max": 100
    },
    "success": true,
    "completion": true
  },
  "timestamp": "2025-10-03T14:25:02Z"
}
```

If they were really ambitious, the course designer might include the trainee's answers to individual quiz questions in this record -though the majority of learning departments don't bother with that level of detail (if their Learning Management System even reports it.)

After the launch of this new case manager onboarding, when it finally came time for the learning department to report to management on the success of their programs (usually during a 3 minute slot with a single presentation slide at the very end of the quarterly meeting), the Director of Online Learning would say something like:

“This quarter we had 784 case managers complete our data privacy training with an average score of 92%. And the average feedback score from participants was 4.8 out of 5.”

To which the Senior Vice President of Human Resources would nod and then declare the meeting adjourned so everyone could head over to the Cheesecake Factory restaurant for an expense account lunch.

Of course, that brief metrics slide told the audience nothing about the impact of our training activity on participants and the organization. Rather, saying “784 case managers completed the course” is like a hospital director bragging that “all our patients received prescriptions” without asking if they actually got better.

When my company first started developing AI coaches, simulations, and assessments, we were still locked in the same mentality, asking “How will clients ever take AI-based training seriously unless we can take the conversations people have with AI agents and boil them down to a completion status and a percentage score?”

So we spent a lot of time and effort doing exactly that, and eventually gave our AI coaching program the ability to generate completions and quiz scores in the standard SCORM / xAPI format, similar to the database example above. And, for about a year, we were pretty pleased with ourselves for accomplishing that.

But later, when we added the ability for AI coaches to recall past interactions with a learner, a light bulb went on, and we asked “If an AI coach can take notes for itself, why not generate notes for the learner... or for their manager?”

Suddenly, our session data went from a string of numbers to looking like this:

SUMMARY:

Iris, the AI coach, discussed climate-risk insurance policies with Santina, focusing on her client portfolio in construction and agriculture.

Iris provided tailored recommendations for both sectors, including policies designed to mitigate risks associated with extreme weather events, resource efficiency, and regulatory shifts. The conversation focused specifically on one of Santina's clients, a large commercial vineyard involved in wine and fruit production.

The conversation emphasized challenges in assessing the degree of climate risk, with Iris requesting additional information about the location and scale of a vineyard client's operations to refine her recommendations.

The session also examined Santina's client relationships. While she demonstrates a clear understanding of client priorities – such as cost savings in construction and operational resilience in agriculture – her limited direct access to decision-makers may hinder her ability to promote climate-risk insurance products effectively. Iris suggested leveraging internal company relationships to improve access, as well as engaging with industry associations and networks to build credibility and expand reach.

Overall, Santina shows strong awareness of client needs and sector-specific opportunities. Continued work with the AI coach should focus on developing strategies to frame climate-risk insurance as both a protective and value-adding tool for clients. At this stage, no immediate manager intervention is required; instead, targeted coaching on relationship-building and persuasive framing will support Santina in advancing climate-risk insurance adoption.

Then we thought “Well, this is neat, but no training program manager has time to read hundreds (let alone thousands) of individual conversation

summaries to gauge overall progress in a large-scale corporate training program.”

But then it occurred to us: if AI can summarize conversations - can't it also summarize summaries? So, as the system began generating individual session summaries for users, we fed the summaries back into the AI agent to produce “meta-summaries”, both for individual user progress over time and entire cohorts.

Instead of a slide reading ‘784 completions, average score 92%,’ managers could now see something closer to a real executive briefing: a snapshot of insights, opportunities, and areas for growth.

CLIMATE RISK INSURANCE COHORT META-SUMMARY

Over the past three months, Iris, the AI coach, facilitated individualized sessions with 385 participants focused on climate-risk insurance across diverse sectors including construction, agriculture, transportation, manufacturing, energy, and finance. Conversations emphasized strategies for managing exposure to extreme weather, supply chain disruptions, regulatory change, and the transition to low-carbon operations.

Key Insights:

Strengths: Across all sectors, participants demonstrated growing fluency in connecting climate-risk insurance with tangible business value such as operational continuity, regulatory compliance, and cost control.

Challenges: A recurring barrier was limited direct access to senior decision-makers. Many participants relied on mid-level contacts, making it harder to advance conversations about climate-related financial products.

Opportunities: Leveraging internal networks, industry associations, and sector-specific business forums emerged as key strategies for building credibility and expanding influence with client leadership teams.

Overall Progress

The cohort as a whole shows strong development in sector knowledge and the ability to identify industry-specific opportunities for climate-risk insurance. Continued coaching should prioritize stakeholder engagement skills, persuasive framing, and strategies for translating technical risk insights into clear business outcomes. At this stage, no systemic performance issues require immediate managerial intervention; instead, the focus should remain on refining client access strategies and deepening participants' confidence in positioning climate-risk insurance as both a protective and value-adding solution.

Since then, measuring the impact of AI coaching interventions has felt more like a real strategic report, instead of reading SCORM tea leaves.

It also has blurred the traditional lines between “learning” and “assessment.” Suddenly, there was less of a need to put a multiple choice quiz at the end of an AI-facilitated course, simulation, or coaching session, because the AI agent had been talking to the user the entire time and could simply tell you whether the user demonstrated mastery of the skills and concepts.

And even when we did want to include a traditional assessment, we now had the option of having the AI grade free response questions, which are 20% to 50% more effective for gauging actual mastery (according to Deloitte, the Journal of Educational Psychology, the Journal of Applied Testing Technology, and a slew of other studies by people smarter than myself.) Which makes sense, considering reality isn't multiple choice. When we give learners a question and a blank input field, we're giving them a chance to prove they can “identify the red flags of money laundering” or “translate a scope of work into a project plan” - not just click a checkbox on a screen.

Making the Grade: Quantitative Metrics

Of course, qualitative summaries are great, but what about situations where we need to assign a specific number or tick a pass/fail box? Can we trust an AI coach or tutor to make those kinds of judgements? And will stakeholders accept the results as reliable? And if an AI coach gives one learner an “85” and another a “78” – will everyone accept it as fair?

So far what we’ve been telling clients about AI assessment is:

- Use AI reporting data to get a general, holistic sense of how learners are progressing and where they might need intervention or reinforcement.
- Do use AI coaches and free response scenarios as part of regulatory compliance training but don’t use an AI activity as the final assessment: compliance departments and attorneys want to know that everyone read the exact same questions worded the exact same way and gave the exact same approved answers (this is not a place to show off AI’s creative abilities.)
- For all other courses, benchmark your AI agents’ assessments against human assessments of the same transcript. If the AI seems comparably fair and accurate, then go ahead and use it (in our early projects, we’ve actually found AI grading of free response interactions to be slightly more consistent than human grading, by about 10%.)
- Have the AI agent explain up front whenever an interaction will be assessed / graded, so people don’t start getting paranoid about talking openly with their AI coach about their workplace challenges or asking an AI tutor for clarification about a topic they don’t understand.
- In other words, the challenges with AI grading short-form free response assessments and training sessions has less to do with the technology’s capabilities than with humans’ comfort level with being evaluated by talking computers.

Level 3 and Beyond: AI, L&D, and the Great Performance Data Reckoning

While my company's primary focus has been on working with humanitarian nonprofits, we've also done occasional training and knowledge management projects for the military and intelligence agencies. To be clear, we refuse to create training for actual weapons systems (which could be exported to who-knows-where and used for who-knows-what), but if the Navy wants to do a course on suicide prevention for sailors, or if a democratic government wants to improve its defenses against election interference or cyberattacks - we're happy to help.

I'm offering this as context for how once, while having lunch with an intelligence analyst, I asked "Does the government really record all my phone calls?"

"Yes," she said matter-of-factly. "But don't worry. We don't have enough time to actually listen to them all."

That was more than a decade ago. Today, with AI, they presumably do have the ability to listen to and meta-tag all our conversations. And, pretty soon, every organization will have that level of visibility into what their employees do all day.

Earlier, we mentioned how the main barrier to Level 3 (behavior) evaluation was lack of visibility into training participants' day-to-day work performance. Soon, we'll be getting way more visibility than we ever thought possible. And instead of denying our requests to add items on quarterly review questionnaires, management will be coming to us with incredibly detailed, minute-by-minute breakdowns of training participants' job performance, demanding to know why our learning interventions haven't moved the needle.

Now, I'm not saying this to scare people (though you should be scared), and I'm definitely not trying to paint AI as some creepy surveillance state technology (rockets can carry nuclear warheads, but that doesn't mean we shouldn't use them to launch weather satellites.) All I'm saying is that the

great “Level 3 evaluation” reckoning is coming, and the best way for L&D professionals to win that conversation is to start it.

If we can make our Level 2 metrics more human and relevant with AI, and advocate for humane, respectful Level 3 performance monitoring, we can not only justify our existence - but perhaps finally take our seat at the organizational strategy table.

Unless we'd rather keep using AI to crank out e-learning scripts and let others decide our fate.



Travelling with the Tin Man: Integrating AI Across the Learning Journey

My last “honest” job before going into consulting was running the learning department at an enterprise financial software company. Our platform had about 500,000 daily users, with 9,000 system admins. In any given year, 3,000 of those admins would show up to our online instructor-led training workshops, and while that’s far too many people to know personally, over time, I did manage to pick a few recurring training participants out from the crowd.

For instance, if Katherine from {government agency X} was on a Zoom session, I knew she would probably distract the facilitator with some deep-dive tangent into advanced features that left all the other participants bewildered. Meanwhile, if Brian from {healthcare organization Y} was present, he would probably complain about how difficult it was to import data from their archaic purchasing system.

In other words, I got to know people - their personalities, their work context - and follow their progress over time. This familiarity meant that other instructors and I could personalize the guidance and support we gave to individual clients. We could help the Katherines stay engaged with advanced assignments during breakout sessions, and help the Brians catch up after the session and troubleshoot their headache du jour. And it was great to watch them go from a clueless new user who didn’t know where to change their password, to a power user creating complex custom reports, helping onboard their colleagues, and suggesting new product features.

Of course, with 3,000 training participants and just 8 instructors, not everyone could receive that level of personalized attention. Plus, there were the other two-thirds of system administrators who never attended a workshop. That logistical math has always limited the reach of instructor-led training.

Unless your organization could justify hiring an army of facilitators, personal attention and relationship-building was difficult if not impossible to scale... at least until AI came along.

Scaling Personalization

So far, we've mainly discussed AI-human interactions in terms of discrete, one-time conversations: a coaching session, a role play activity, or a free response assessment. But this misses an even more transformative use case: when we give AI agents the ability to store and recall memories of past interactions, they can actually “get to know” users and provide personalized support throughout an individual's learning journey.

Before the popularization of generative AI, if an organization didn't have enough training facilitators to reach an audience, or if their workforce was too dispersed, or if a training topic was too niche, the go-to alternative to instructor led training would be e-learning, or even just a video. Personalization was often the first casualty of digitalization.

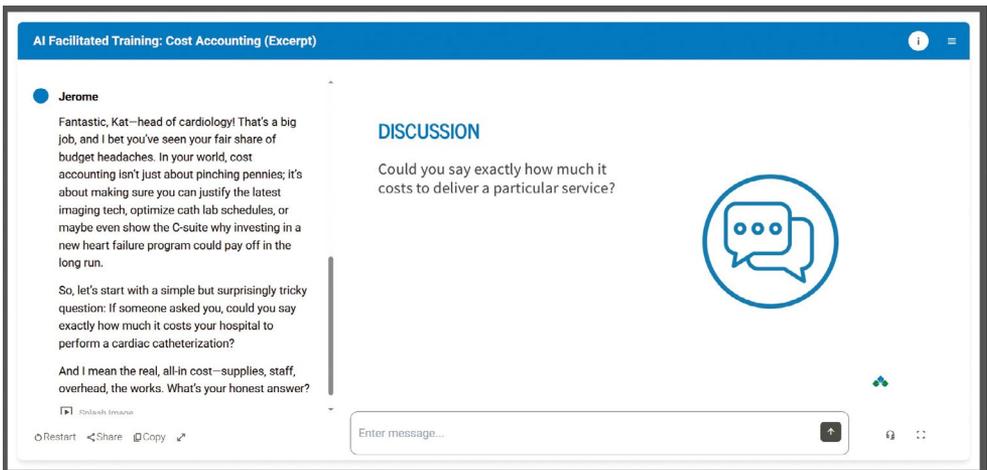
While an LMS platform could direct a salesperson to a micro-learning course for a particular product line, or provide practice problems for a particular state's insurance agent licensing exam, the courses themselves were essentially static, and didn't adapt to the user.

At best, an (overly) ambitious e-learning designer might throw in a couple variables so a welcome screen could say “Hello, {name} - let's discuss how you can sell more {product} to clients in {industry}”, or maybe let a user skip a section if they passed a multiple choice knowledge check - but that's about as far as personalization went.

By contrast, an AI coach can glance at a profile of the user, and then adapt their interactions on the fly. Suddenly, instead of filling in a handful of

blanks, instructional designers can tell an AI agent to “ask the user about the biggest challenge they face selling their product lines to clients in their main vertical” and it could reference the user’s profile to determine if it should talk about selling pasteurizing machines to dairy farms, or jackhammers to construction contractors.

For example, the screenshots below are from a healthcare cost accounting tutor. Both depict the exact same slide, but in the first instance, the user profile is for the head of cardiology at a hospital, while the second is for the chief financial officer at a chain of walk-in urgent care clinics. Just like an effective live instructor, the AI tutor took the same basic question and tailored the details on the fly:



Fantastic, Kat—head of cardiology! That’s a big job, and I bet you’ve seen your fair share of budget headaches. In your world, cost accounting isn’t just about pinching pennies; it’s about making sure you can justify the latest imaging tech, optimize cath lab schedules, or maybe even show the C-suite why investing in a new heart failure program could pay off in the long run.

So, let’s start with a simple but surprisingly tricky question: If someone asked you, could you say exactly how much it costs your hospital to perform a cardiac catheterization?

And I mean the real, all-in cost—supplies, staff, overhead, the works. What’s your honest answer?

AI Facilitated Training: Cost Accounting (Excerpt)

Jerome

Fantastic, Kat—head of cardiology! That’s a big job, and I bet you’ve seen your fair share of budget headaches. In your world, cost accounting isn’t just about pinching pennies; it’s about making sure you can justify the latest imaging tech, optimize cath lab schedules, or maybe even show the C-suite why investing in a new heart failure program could pay off in the long run.

So, let’s start with a simple but surprisingly tricky question: If someone asked you, could you say exactly how much it costs your hospital to perform a cardiac catheterization?

And I mean the real, all-in cost—supplies, staff, overhead, the works. What’s your honest answer?

©Health Science

Restart < Share Copy

DISCUSSION

Could you say exactly how much it costs to deliver a particular service?



Enter message...

Perfect, Kat—CFO at a chain of urgent care clinics. That’s a front-row seat to the financial circus, isn’t it? Cost accounting can be a real game-changer for your world: think about optimizing staffing during flu season, or figuring out which rapid tests are actually worth their salt versus just burning a hole in your budget. It’s not just about pinching pennies—it’s about making sure you’re not accidentally running a charity when you meant to run a business.

So, let’s start with a simple but surprisingly tough question: If someone asked you, could you say exactly how much it costs your clinics to treat, say, a patient with a sprained ankle? Not just a ballpark, but the real, all-in cost?

And just to be clear, there’s no shame in saying “not really”—most places are flying a little blind here.

Likewise, instead of merely skipping sections of training with binary “show/hide slide” logic based on the learner’s quiz results, an AI tutor can be instructed to “Review the following topics, starting with the one that seems to be the greatest challenge for the user and skim any topics where the user already seems to demonstrate solid mastery.” And while it might need a bit more clarification on what constitutes ‘mastery’, it’s more or less the kind of directions you’d give a human training facilitator.

An AI tutor can even use the classic opening line for instructor-led workshops - “What were you hoping to get from today’s session?” - and actually tailor its content based on users’ responses (even more so than a live instructor, since the AI tutor doesn’t have to worry about the learning needs of 15 other workshop participants.)

This sort of personalization gets even more powerful when we extend it beyond individual interactions with a specific AI agent, and allow all of the agents in a system to share information about users, and the organization itself.

Creating Continuity

Depending which sensationalist reports you read, anywhere from 30% to 65% of AI users have formed an emotional bond with their favorite model. So it might sting to point out that, on a certain level, AI models like ChatGPT can’t “recognize” or “know” their users any more than a goldfish can know their owner.

In fact, one of the biggest misconceptions about AI is that models can learn in real time, when - in reality, they stop “learning” the day a model is released. Without access to an external search engine, the original version of ChatGPT 5 would be permanently trapped in September 2024, unaware of the November 2024 U.S. presidential election results or Coco Gauff’s second grand slam tennis championship in May 2025. Basically, the AI model wakes up at the start of each conversation a blank slate.

That said, we can create the functional equivalent of learning and persistent memory by connecting AI agents to a database, with the ability to save a summary of each conversation and review the summaries at the start of their next conversation with a specific user. This allows AI coaches to “remember” learners, tailor their current interaction based on the learner’s specific needs, and track progress over time. However, unlike humans, multiple AI agents can share these memories - which opens up some very interesting possibilities for a user’s learning journey.

For instance, if an AI coach learned that a new hotel front desk employee only had experience with a very specific property management system, it could pass that information to the software training copilot, which could then help the learner navigate the new system by comparing it to the other system they were already familiar with (assuming it was a commercial app with at least some documentation available online.) Likewise, if a commercial HVAC salesperson struggled with a customer interaction role play, the sales enablement AI could place extra emphasis on how to address common customer questions and objections, not just product features and benefits.

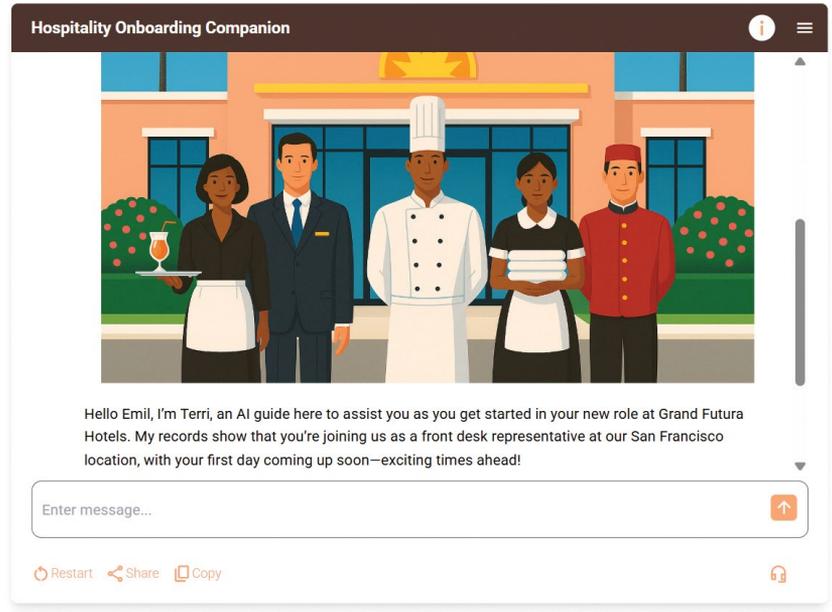
Mapping the Journey

L&D professionals like to talk about the “learning journey”, and how workplace learning shouldn’t be a one-time event, but a continuous process that starts from the date a team member is hired then continues for the entire length of their relationship with the organization, and possibly beyond.

Yet, more often than not, this journey is reduced to a series of disconnected “events” with conditional gates (e.g., “If user scores 80% on Module A then unlock module B; after user completes module B, register them for the next available webinar.”) But learner progress data rarely flows from one event to the next: at best, perhaps a very diligent workshop facilitator will check to see if any of their next session’s participants struggled with the prerequisite e-learning modules (though even this is uncommon.)

By having AI agents share conversation history, user context, and a knowledge of policies and SOPs, we can make the learning journey feel less like running a gauntlet and more like a relationship that compounds over time. Or, to use a medical analogy, AI agents can work like a coordinated care team of doctors, with a general-purpose coach acting as the primary physician and other agents (role plays, assessments, etc.) serving as specialists who consult and coordinate to support the same learner.

If that seems vague, imagine the following progression:



Hotel Onboarding Guide

- A new employee joins a hotel chain, and their first message after accepting an offer contains a link to an AI onboarding guide, which shares an overview of the organization (including the employee's specific property and department), and helps them get ahead on some of their paperwork.
- The onboarding agent refers the employee out to several different specialist AI agents to complete safety training (which concludes with a series of free response scenario based assessments), customer experience role plays, and other learning events. As the process moves forward, the various agents maintain a common checklist, which the human HR team audits to confirm key points like “did the AI agent actually discuss the attendance and overtime policies with the team member?”
- At the end of onboarding the AI agent switches over to serving as a coach / on the job copilot, integrated with the various systems and tools needed for the employee's job.

- The user continues to use the AI agent(s) for day-to-day job tasks and occasional training interventions as needed, from managing their to-do list, to looking up policies, to resource planning and scheduling for their team.
- A separate set of AI agents monitor the employee's progress, occasionally raising a flag for managers or other humans to get involved when they need support / intervention.
- As the employee grows more experienced, they eventually become responsible for curating the AI system's knowledge base - perhaps even editing the source articles for the new hire onboarding process - and eventually reach a point where they are helping the organization's AI developers design, build, and maintain agents for other aspects of the hotel's operations.

Basically, the learning journey becomes a cycle, where AI first helps the user acquire skills and knowledge, then later helps them share their expertise with others.

Evolving Beyond “Training”

I once had a quirky hippie artist friend who worked at my local library. Sometimes we would meet up on our lunch breaks, and he would unpack a sack of health food with a large tub of kimchi for every meal. When I asked if kimchi was his favorite food, he said, “No man, I hate vegetables. But fermentation makes vegetables more nutritious, which means you don't have to eat as many of them.”

That slightly offbeat bit of diet advice holds a lesson for learning and development. As L&D professionals, we tend to assume that more training is always better. But what if most people actually dislike training, and AI could help them need less of it?

Since the 1990s, knowledge management (KM) practitioners have argued that delivering information “just in time” at the moment of need is better than delivering training “just in case” people need to know/do something

later. In theory, it's a great idea: most of what people learn in training is forgotten within weeks if it isn't applied, so finding ways to provide information or instruction rapidly on demand is simply being efficient.

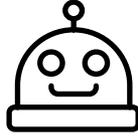
Alas, in practice, most KM systems devolve into digital junk drawers: a maze of intranet sites, SharePoint folders, and forgotten PDFs, organized and meta-tagged according to a system nobody really followed in the first place ("Should these meeting notes go under Design Sessions, Product Specs, or Innovation?")

And that's been the tragedy of KM up to this point: it makes so much sense but it almost never works (except in cases where there's a massive, direct incentive to keep certain information organized, such as pharmaceutical research.) More often than not, the level of vigilance required proves unsustainable, or the system doesn't survive the departure of the organization's KM champion - often falling apart before it's fully implemented. As one healthcare client put it, "We spent years curating all these amazing job aids and videos to help doctors and nurses — but nobody can find them on SharePoint, so they never get used."

But where humans struggle to keep folder hierarchies and metadata organized, well-designed AI agents can swim through well structured databases and repositories like fish through water, and surface exactly what users need, when they need it. Whether it's schematics for an old prototype, or a link to a "lunch and learn" webinar held in the Singapore office last week, important documents can be made accessible immediately without users needing to trudge through the folder system. This could be the key to finally delivering "learning in the flow of work" - which, until now, has been more slogan than strategy.

That said, AI isn't a substitute for knowledge management - it is merely an amplifier of KM. Too often, organizations mistake AI for a search engine and ask it to analyze massive, unstructured datasets and document repositories ("find all maintenance-related safety incidents from this spreadsheet.") In these cases, the model invariably struggles, not because it's "dumb," but because the underlying information lacks structure. But when the data is tagged, organized, and maintained, AI can find patterns, surface insights, and even automate routine analysis.

As AI reduces the need for conventional training, taking responsibility for managing AI knowledge sources would be a natural evolution for L&D professionals - part learning, part library science. And with the help of AI, we might actually be able to keep it in a state where agents can deliver what's needed in the moment: fast, contextual, and personalized.



Who Will Teach the Robots? L&D's New Job Description

Before wrapping up this book, we need to address the elephant in the room.

Earlier, we referenced the old Nike advertising slogan “Just do it.” And when we think about AI for L&D, it sort of begs the question: If we can teach AI to train and support humans to perform a complex job task, why can't we teach AI to... just do it?

Right now, a lot of workers - including people in L&D - are afraid that AI will soon take everyone's jobs. Meanwhile, a lot of CEOs and investors are not-so-secretly hoping that AI will soon take everyone's jobs. But both camps overlook a simple fact: despite promises of “artificial general intelligence” (AI that can do any task as well or better than a human, right out of the box), today's AI isn't going to take over any significant portion of anyone's jobs unless someone takes the time to teach it ¹³.

And who better to teach AI how to do a job than a workplace learning professional?

Understanding Your Audience

When designing learning experiences, we always want to view the subject matter from the audience's perspective, and that is as true for AI as it is for humans.

¹³ Note that I said “teach”, not “train.” In the world of AI, “training” is what happens when machine learning experts create an AI model by exposing learning algorithms to trillions of words of data.

So... what do we know about our new, digital learners?

John W. Campbell, the famous science fiction magazine editor from the 1950s, once challenged an author to “Write me a creature that thinks as well as a human or better than a human, but not like a human.” And AI certainly fits that description.

While we could fill another book with a discussion of similarities and differences between humans and AI, here’s an extremely simplified job for those who haven’t spent much time studying the behaviors of AI models yet.

Human

A.I.

Reads text from beginning to end, processing the meaning of each word and sentence.

Analyzes text all at once, identifying general patterns in the distribution of words.

Born knowing almost nothing, then gradually accumulates knowledge / memories over time, with a remarkable ability to recall contextually relevant information as needed.

Starts out with a vast pool of knowledge, but cannot learn anything new / form memories (but can query search engines, text of past conversations, other data sources as needed)

Reasons based on their life experience and education, strongly influenced by emotional reactions / associations.

Predicts based on available data and weighting system, strongly influenced by how frequently things appear in their data / input.

Motivated by perceived self-interest (“What’s In It For Me?”) Gets fatigued after prolonged concentration.

Predisposed to do whatever the user says / answer user’s question. Can get confused as conversation transcript gets longer.

There are also some surprising similarities between human and AI memory and language processing. For example, AI models exhibit a version of the “serial position effect” where - like a human learner - it tends to pay more attention to information at the beginning and end of a long passage of text, while sometimes glossing over the middle. They also share humans’

tendency toward recency bias (giving greater weight to whatever the user said last) and cultural biases (if you say “bread” then an AI model trained disproportionately on North American web content is more likely to mention “butter” than “olive oil”.)

Again all this is a gross oversimplification. The best way to familiarize yourself with AI models is simply to spend a lot of time interacting with them and carefully observing their behaviors - in the same manner as a dog or horse trainer would spend a lot of time observing animals. While it’s impossible for humans to fully grasp how AI models “think” (the math involved is simply too complex, involving things like 4,096 dimensional vector spaces), we can develop an intuitive feel for how AI models relate to language, and use that to help them perform increasingly complicated tasks.

Explaining the Job Task

Many L&D professionals have bought into the myth that AI is a digital djinni capable of doing anything as long as you phrase your wish correctly. The internet is littered with lists of “prompt hacks” suggesting that all you need to do is type “embody the world’s greatest lawyer” or “come up with a good argument, then come up with a better argument” and suddenly the AI model will turn into a legal genius. And the companies that make the AI models have done little to discourage this magical thinking.

The fact that you can instruct AI models with plain language makes it both a phenomenal tool and a dangerous trap. On one hand, no matter what you type, an AI model will do something. But too often, people give up if the AI doesn’t produce the output they want after two or three rephrasings.

In reality, getting AI models to perform complex, specific tasks consistently takes more than a one or two paragraph (let alone single sentence) set of instructions. For instance, the AI agent we developed to help bankers market loans for energy efficiency upgrades to businesses has more than 30,000+ words of instruction. And we don’t just tell the AI agent “You are a climate finance expert, talk to the learner about energy efficiency upgrades for their clients”. We systematically break the conversation into stages, complete with clear objectives and completion criteria for each stage (e.g. “Review portfolio”, “Assess individual client”, “Develop presentation”, etc.)

Fortunately, learning & development professionals already have the foundational skills to create these kinds of massive instruction sets. We already spend our days analyzing complex job tasks, breaking them down into steps, logically organizing all the necessary contextual information, and providing examples / job aids to help learners apply the skills and concepts in real-life situations. We also deal with things like cognitive load theory, and making sure information is presented in manageable chunks so the learner doesn't get overwhelmed.

All of these principles apply to AI agent development. The only thing that's changing is the species of the audience that you are training.

Chunking Information

AI models possess a strange kind of genius: on one hand, they can read faster, summarize better, and speak intelligently on a wider range of topics than any human alive - and yet, they have the memory of a goldfish.

Every time you hit "send" in an AI conversation, the model is like an anterograde amnesiac waking up for the first time, reset to knowing only their original training data (if you've ever seen the movie *Memento*, about an amnesiac trying to solve his wife's murder, it's basically the same idea.) In the milliseconds before it replies, the AI agent has to get caught up on the conversation by re-reading the entire transcript up to that point, plus whatever instructions and contextual data was provided by the prompt engineer who designed it.

In other words, you can't "train" AI - it's already had all the training it will ever remember. If you had an AI attend a workshop on sales negotiation skills, it would forget them all the instant the conversation ended. This means learning "in the flow of work" is the only way an AI agent "learns" anything at all.

So what's the implication for AI agent development? Basically, we need to find ways to give the AI agent the exact information it needs to know at any given moment, which usually means graduating from the basic chatbot interfaces for ChatGPT, Claude, etc. to a platform capable of managing

context and workflows for conversations. For instance, when I said that our team’s AI agent for climate finance had “30,000+ words of instructions”, we were using our AI agent development platform to feed the model the most relevant 3,000 to 5,000 words of instructions at any given time.

Developing AI’s “Soft Skills”

Most L&D professionals have conducted a “training of trainers” where we walk a group of instructors through the course materials we developed, and help them practice their delivery.

Just as we don’t hand a slide deck to a new instructor and expect them to instantly understand how to facilitate it, we can’t hand raw information to an AI model and expect it to engage learners effectively. Designing an effective AI agent requires teaching it how to teach, how to reason through questions, clarify uncertainty, and adapt its tone to different audiences. In other words, we’re not just programming answers; we’re cultivating instructional judgment.

AI agents require something similar to this train-the-trainer process in order to help them interact effectively with human users. For example, most AI models are inclined to simply give an answer - any answer - to the user’s most recent question without asking clarifying questions, which isn’t how an experienced human consultant would approach a client interview. By making sure the AI agent’s design includes steps like “Review user’s latest statement and determine if there is any missing information or potentially mistaken assumptions that would prevent you from giving a proper answer...” or “Determine if you have collected sufficient information on the following points before answering...” we can encourage the AI agent to take a more methodical approach.

Giving AI agents a specific communication style, worldview, and personality can also help them interact more effectively. For instance, we might encourage an AI tutor for a conflict management course to be friendly and verbose, providing anecdotes and illustrative examples for the various concepts it raises. By contrast, we might encourage an AI advisor for appliance repair technicians to keep their statements brief, favoring step-by-

step bullet point instructions over chatty paragraphs. Likewise, we might add instructions for the AI model to adjust its level of formality, or how much technical jargon it uses, based on what we know of our users.

In short, we should take a “user centered” approach to AI agent design, where personality and communication style serve as an interface. Because while it doesn’t hurt for users to know a thing or two about AI, conversations should feel natural. Users shouldn’t have to rephrase their questions and statements ten different ways to get the output they need (figuring that part out is our job now.)

The Digital Toolshed: Leveraging Different AI Platforms

In some ways, AI is like any technology. It’s important to choose the right tools (or at least not use the wrong tools), but even the best tool can’t overcome lack of skill or poor design choices.

For instance, with e-learning technology, I have strong opinions about the relative strengths and weaknesses of Articulate Storyline vs. Adobe Captivate vs. iSpring Presenter. But if a beginning instructional designer was agonizing over which tool to use before creating their very first practice course, I’d probably say, “Just pick one and focus on your instructional design fundamentals.”

The same goes for AI models and agent platforms. When our company first started creating AI coaches and tutors, we tended to favor Anthropic Claude for role plays and OpenAI ChatGPT for assessments. But then ChatGPT 5 came out, which excelled at role plays, and Google Gemini caught up, becoming our go-to model for specific types of tasks. But for someone just starting out? Pick one (as of this writing, I’d suggest ChatGPT) and just start putting in long hours of building with it.

It's also worth noting that, for serious applications, you'll soon outgrow the consumer chatbots for the popular models, and need to start using a tool capable of managing more complex workflows and data integrations. While our company ultimately developed our own (Parrotbox.ai) specifically for L&D applications, other options might include OpenAI Studio, N8N, or Amazon Quick Suite.

Meanwhile, there's plenty of purpose-built apps that serve as a midpoint between tinkering with ChatGPT on your own and complex custom application development. Right now you can find stand-alone AI apps specifically for call center role plays or new hire onboarding. However, standalone solutions can easily become data "silos" disconnected from the rest of your organization's AI infrastructure (e.g. your sales coaching AI app might not be able to talk and share notes with your product knowledge chatbot.)

The bottom line here is that, while the question of what AI tools and platforms to use can get highly technical, we should not abdicate decisions about our AI learning tools to IT, any more than would let IT choose a traditional LMS platform without our input.

Taking Our Seat at the AI Table

Right now, nobody is looking to L&D to lead the way on AI - but perhaps they should. As a profession, we bring the missing pieces to the AI puzzle: we know how to break down complex jobs into their component tasks and present them in a logically organized way. And that's what AI needs to reach its full potential.

The first step towards transforming organizational learning and performance support with AI probably isn't sitting alone trying to write a better prompt to create an e-learning script. Rather, it's educating our peers and executive stakeholders on what AI can potentially do for workforce performance, developing a clear vision of how to achieve it, and making it clear what L&D has to offer in the AI-powered future of work.

Conclusion



Getting Real About Artificial Intelligence

Science fiction author William Gibson once observed that “The future is already here, it’s just not evenly distributed.”

When it comes to AI’s impact on the workplace, that future isn’t some distant possibility we need to prepare for: It’s happening right now. The only question is whether learning professionals will be out front driving that transformation, or collateral damage, like typesetters, carriage-makers, and film developers.

Every example in this book, from the climate finance coach helping bankers identify green investment opportunities to the onboarding guide for hotels to the medical simulation that made an ER exclaim “This is giving me flashbacks”, reflects work our company has already implemented or is currently deploying with client organizations. These aren’t proof-of-concept demos or laboratory experiments. They’re production systems serving thousands of learners across industries from healthcare to financial services to manufacturing.

Future editions of this book will include detailed case studies, once we’re at liberty to share them: But the early observations from our client implementations suggest that, once learners and organizations start learning with AI, they won’t want to go back to the old paradigm.

And while organizations debate whether AI is ready for workplace learning, our learners have already moved on. They’re using ChatGPT and Copilot to draft emails, Claude to practice for job interviews, and Perplexity to

research industry trends. Meanwhile, other fields - from marketing to finance to operations - are deploying AI solutions that make our traditional learning technologies and programs look quaint by comparison.

We have a choice to make:

- We can ignore AI and pretend it doesn't exist as the rest of the world moves on.
- We can join the herd and use AI to produce the same old content, until our organizations ask if we're even necessary.
- Or we can recognize this moment for what it truly is: the greatest opportunity L&D has had in decades and lead the AI revolution from the front.

The new, AI-powered workplace is just beginning to take shape. And while, for now, people are fixated on the tech, the organizations that will thrive won't be the ones with the fanciest models, they'll be the ones with the most thoughtful approach to human-AI collaboration. And who better to design that collaboration than the professionals who've spent their careers understanding how people learn, grow, and perform?

This is either the end of L&D as we've known it, or the beginning of L&D as it was always meant to be. Personally, I'm betting on the new beginning.

Read more on our blog at



AI ISN'T “THE FUTURE” OF WORKPLACE LEARNING. IT'S HOW WE FINALLY MAKE LEARNING WORK.

For years, L&D has been churning out modules, slide decks, and quizzes while real learning was happening between colleagues, in conversations that unfolded in conference rooms and hallways.

Now, artificial intelligence lets us bring those authentic learning moments to scale. This isn't about using AI to make yesterday's content faster—it's about creating digital entities that can act as coaches, mentors, and teammates.

Teaching Machines to Train Human Beings is a manifesto and a blueprint for learning teams ready to stop building content and start redefining what learning means when everyone—from the CEO to the cashier—can have an expert coach on demand, whenever they need it.

Rather than speculation, this book offers real-world examples of how AI can deliver coaching, simulation, and practice, along with adaptive support and meaningful, qualitative reporting that goes far beyond hollow completion rates and percentage scores. It's not about what AI makes. It's about what AI makes possible—turning training from a one-off event into a living system that grows and evolves alongside its people.

Because the future of workplace learning isn't content.

It's conversation.



Parrotbox.ai