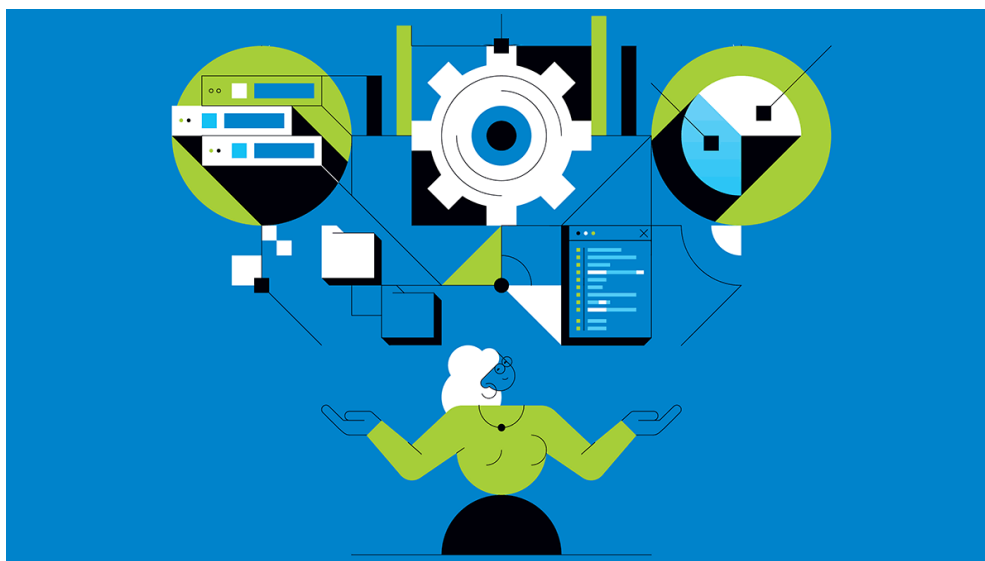




## AI and Machine Learning



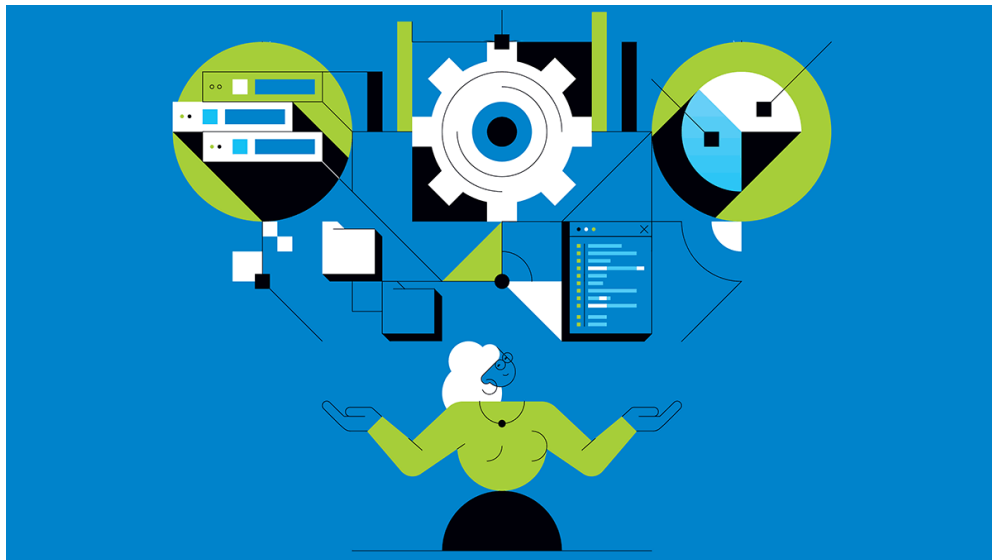
# Embracing Gen AI at Work

The skills you need to succeed in the era of large language models  
by **H. James Wilson and Paul R. Daugherty**

# Embracing Gen AI at Work

The skills you need to succeed in the era of large language models  
by **H. James Wilson and Paul R. Daugherty**

From the Magazine (September–October 2024) / Reprint [R2405L](#)



Blindsalida

**Generative artificial intelligence is expected to radically** transform all kinds of jobs over the next few years. No longer the exclusive purview of technologists, AI can now be put to work by nearly anyone, using commands in everyday language instead of code. According to our research, most business functions and more than 40% of all U.S. work activity can be augmented, automated, or reinvented with gen AI. The changes are expected to have the largest impact on the legal, banking, insurance, and capital-market sectors—followed by retail, travel, health, and energy.

For organizations and their employees, this looming shift has massive implications. In the future many of us will find that our professional success depends on our ability to elicit the best possible output from large language models (LLMs) like ChatGPT—and to learn and grow along with them. To excel in this new era of AI-human collaboration, most people will need one or more of what we call “fusion skills”—*intelligent interrogation*, *judgment integration*, and *reciprocal apprenticing*.

*Intelligent interrogation* involves prompting LLMs (or in lay terms, giving them instructions) in ways that will produce measurably better reasoning and outcomes. Put simply, it’s the skill of thinking with AI. For example, a customer service rep at a financial services company might use it when looking for the answer to a complicated customer inquiry; a pharmaceutical scientist, to investigate drug compounds and molecular interactions; a marketer, to mine datasets to find optimal retail pricing.

*Judgment integration* is about bringing in your human discernment when a gen AI model is uncertain about what to do or lacks the necessary business or ethical context in its reasoning. The idea is to make the results of human-machine interactions more trustworthy. Judgment integration requires sensing where, when, and how to step in, and its effectiveness is measured by the reliability, accuracy, and explainability of the AI’s output.

With *reciprocal apprenticing*, you help AI learn about your business tasks and needs by incorporating rich data and organizational knowledge into the prompts you give it, thereby training it to be your cocreator. It’s the skill of tailoring gen AI to your company’s specific business context so that it can achieve the outcomes you want. As you do that, you yourself learn how to train the AI to tackle more-

sophisticated challenges. Once a capability that only data scientists and analytics experts building data models needed, reciprocal apprenticing has become increasingly crucial in nontechnical roles.

Why do you need to systematically develop these new skills for thinking, building trust, and tailoring? Empirical research consistently shows that ad hoc instructions—the way most employees prompt LLMs today—lead to unreliable or poor outcomes, especially for complex reasoning tasks. This is true across functions, from customer service, to marketing, to logistics, to R&D. It’s critical for all of us to bring more rigor to our use of gen AI at work. In this article we’ll explain how.

### **Interrogating AI Intelligently**

How do you improve the output of a massively complex system like an LLM, which is trained on mountains of data and driven by probabilities instead of human logic? There are several techniques you can use.

**Think step by step.** When prompting gen AI, you need to break down the process it should follow into the constituent parts and then strive to optimize each step—just as the first wave of scientific management did in industrial manufacturing. However, the AI process doesn’t involve an assembly line; it involves a chain of thought through which an outcome is sought. Studies have shown that when gen AI tools are instructed to break reasoning tasks down in this manner, their performance dramatically improves. This is particularly true with tougher problems, as Jason Wei, the OpenAI researcher who first explored chain-of-thought reasoning, has demonstrated.

In fact, adding the simple phrase “Let’s think step by step” to an LLM’s instructions can increase the accuracy of its output more than threefold across a range of tasks from math to strategic reasoning. Let’s say your gen AI prompt is this: “My department has a budget of \$500,000.

We have spent 20% on equipment and allocated 30% for a new hire. We just received a budget increase of \$50,000. What is our remaining budget? Let's think step by step." The model will put out: "Initially, your department had \$500,000. You spent 20%, or \$100,000, on equipment, leaving \$400,000. You allocate 30%, or \$150,000, for a new hire, which brings the budget down to \$250,000. Finally, you recently received a budget increase of \$50,000. Your remaining budget is \$300,000." While most people could do this math in their heads, the point is that LLMs (which work far faster) can be made to detail their work on quantitative problems that are much more complex, such as finding the shortest possible route for a sales rep to take among several cities. This creates a traceable chain of reasoning—instead of spitting out an answer at the end of a black-box process—that allows you to verify the accuracy of the results.

**Train LLMs in stages.** For human-machine collaboration on complex tasks that require occupational and domain expertise, such as law, medicine, scientific R&D, or inventory management, you can introduce AI to the work in stages to generate better outcomes.

For example, the MIT researchers Tyler D. Ross and Ashwin Gopinath recently explored the possibility of developing an "AI scientist" capable of integrating a variety of experimental data and generating testable hypotheses. They found that ChatGPT 3.5-Turbo could be fine-tuned to learn the structural biophysics of DNA when the researchers broke that complicated task down into a series of subtasks for the model to master. In a nonscientific area like inventory management, subtask stages might include demand forecasting, the collection of data on inventory levels, projections of reorders, order quantity evaluation, and performance evaluation. For each successive subtask, managers would train, test, and validate the model with their domain expertise and information.

**Explore creatively with LLMs.** Many work processes, from strategy design to new product development, are open-ended and iterative. To make the most of human-AI interaction in these activities, you need to guide machines to visualize multiple potential paths to a solution and to respond in ways that are less linear and binary.

This kind of intelligent interrogation can increase LLMs' ability to produce accurate predictions about complex financial and political events, as the researchers Philipp Schoenegger, Philip Tetlock, and colleagues recently showed. They paired human forecasters with GPT-4 assistants that had been primed with richly detailed prompts to be “superforecasters”—to assign probabilities and uncertainty intervals to possible outcomes and offer arguments for and against each. The researchers found that the predictions made by those assistants (about everything from the closing value of the Dow Jones Transportation Average on a certain date to the number of migrants entering Europe via the Mediterranean Sea in December 2023) were 43% more accurate than predictions generated by unprimed LLMs.

### **Incorporating Your Judgment**

Bringing expert—and ethical—human discernment into the equation will be critical for generating AI outputs that are trustworthy, accurate, and explainable and have a positive influence on society. Here are some techniques you can use:

**Integrate RAG.** Not only can LLMs hallucinate, but the information and datasets they are trained on are often many years old. When working with LLMs, people must frequently make judgment calls on the extent to which reliable, relevant, and up-to-date information in outputs will be critical. If they are, you can use retrieval augmented generation (RAG) to add information from authoritative knowledge bases to an off-the-shelf LLM's training sources. Doing so can help

prevent misinformation, outdated responses, and inaccuracies. A pharmaceutical researcher, for instance, might use RAG to tap human genome databases, recent publications in science journals, databases covering preclinical research, and FDA guidelines. To get set up on RAG, people will often need the help of their IT teams, who can tell them if it has been or can be integrated into their workflow to add an extra layer of quality to their work.

**Protect privacy and avoid bias.** If you're using confidential data or proprietary information in your AI prompts, only company-approved models behind corporate firewalls should be used, never open-source or public LLMs. Corporate policy permitting, you *can* use private information when the terms of service for an LLM's application programming interface specify that it won't be retained for model training.

Pay attention to the biases you might embed into your prompting. For instance, a financial analyst asking an LLM to explain how yesterday's quarterly report signals that the company is primed for a five-year growth cycle is showing *recency bias*, the tendency to overweight the most recent information when predicting future events.



Blindsalida

LLM providers are figuring out ways to help users counter such problems. Microsoft and Google are adding features that help users check for harmful prompts and responses. Salesforce has developed AI architecture that masks any confidential customer data in the construction of prompts; prevents such data from being shared with third-party LLMs; scores outputs for risks like toxicity, bias, and privacy; and collects feedback on improving prompt templates. Nevertheless, at the end of the day, it's you—the human in the loop—whose judgment will matter most.

**Scrutinize suspect output.** Stay on high alert for hallucinations and errors, which according to current research are inevitable even with significant data engineering and other interventions. When LLM users



encounter output that seems off, they often reflexively prompt the model to try again and again, gradually decreasing the quality of the response, as the University of California Berkeley [researchers Jinwoo Ahn and Kyuseung Shin](#) have shown. The researchers recommend that instead you identify the step where the AI made an error and have a separate LLM perform that one step, breaking it down into smaller individual problems first, and then use the output to adjust the first LLM. Imagine a scientist using OpenAI's ChatGPT to help develop a new polymer with a series of step-by-step calculations. If she finds an error at any point in the chain, she can ask Anthropic's Claude to break that step down into smaller subproblems and explain its reasoning. She can then feed that information into ChatGPT and ask it to refine its answer. In essence, this technique applies chain-of-thought principles to the correction of output you judge to be wrong.

### Turning AI into Your Apprentice

As the size and complexity of LLMs increase, they can exhibit “emergent properties”—powerful new abilities, such as advanced reasoning, that they weren't trained for but that nevertheless appear after you tailor LLMs by giving them contextual data or knowledge. To spur their development, you can take the following steps.

**Provide the model with “thought demonstrations.”** Before giving an LLM a problem to solve, you can prime it to think in a certain way. For instance, you might teach it “least to most” reasoning, showing the AI how to break down a complex challenge into several smaller, simpler challenges; address the least difficult one first; use the answer as the foundation for solving the next challenge; and so on. Denny Zhou and colleagues at Google DeepMind have shown that the least-to-most approach improves the accuracy of AI's output from 16% to 99%.

Consider a marketing manager at a fitnesswear brand who wants help thinking through a new line. He can break down the problem for the LLM like this:

1. *Audience*. Identify fitness enthusiasts who would be potential customers—a relatively easy task, especially for a model trained on the company’s customer data.
2. *Messaging*. Craft messages emphasizing performance, comfort, and style—a more challenging and creative problem that builds on the previous identification of the audience.
3. *Channels*. Choose social media, fitness blogs, and influencer partnerships that will help get those messages to the audience.
4. *Resources*. Allocate budget (often the most contentious issue in any organization) according to the choice of channels.

**Train your LLMs to learn new processes.** You can teach AI how to perform a task by walking it through a set of examples within a context in your prompts. This is called “in-context learning,” and it allows you to adapt pretrained LLMs like GPT-4, Claude, and Llama without the sometimes labor-intensive process of adjusting their parameters. For instance, researchers reported in *Nature* that [LLMs were shown how to summarize medical information](#) by prompting them with examples of radiology reports, patient questions, progress notes, and doctor-patient dialogues. Afterward they found that 81% of the summaries produced by the LLMs were equivalent or superior to human-generated summaries.

You can also train an LLM by supplying it with contextual information and then asking it questions until it solves your problem. Consider two software firms, both looking to boost sales. At company one, the sales team has struggled to effectively predict demand for software licenses. So its leader begins by providing the LLM with historical sales data and then asking about expected demand for the upcoming quarter. Next he supplies the model with information on customers’ software

feature upgrades and annual budgets and asks it about the effects of seasonality. Finally, he feeds it detailed statistics from CRM systems and marketing reports and asks it about the impact of marketing campaigns on sales.

At company two, the sales team wants to improve client selection. Its leader might supply specific financial data and prompt an LLM to rank clients by their revenue contribution, and then advance to follow-on queries about geographic reach, customer bases, technical expertise, and so on. At each step both executives are training the LLM and refining its ability to perform the task in the context of the company's particular sales strategy. They bring organizational and industry knowledge to the interactions. As the LLM used by each accumulates more experience with the company's specific sales process, it generates better answers.

Reciprocal learning occurs as users advance from using simple questions or instructions and gradually describe the task with more and more complexity and nuance. They can add context, adjust wording, and see how the model responds, experimenting until they achieve the desired results.

### **Acquiring New Fusion Skills**

Widespread acquisition of gen AI skills will require not just significant investment by organizations but also individual initiative, study, and hard work. Although a few companies are offering relevant training, most have not yet developed robust programs. Indeed, in our 2024 survey of 7,000 professionals, we found that while 94% said they were ready to learn new skills to work with gen AI, only 5% reported that their employers were actively training their workforces in it on a significant scale. So many of you will need to take matters into your own hands—and keep up with the rapid advances in LLMs and the

high-level research being translated into practices for a variety of jobs and industries. You can enroll in online courses from providers such as Coursera, Udacity (which was recently acquired by our firm), the University of Texas at Austin, Arizona State University, and Vanderbilt University; experiment with the prompting techniques we've discussed as well as with emerging ones; and push your employers to provide more opportunities to use LLMs along with instruction in best practices for them.

Up next: acquiring the skills to do chain-of-thought prompting for agentic workflows and multimodal large language models (MLLMs), which integrate different kinds of data, like text, audio, video, and images, while also providing outputs in those formats. One group of researchers has found that chain-of-thought prompting improved MLLMs' performance by up to 100%. Early adopters are already testing these methods, but they're not mature enough yet for widespread adoption.

The AI revolution isn't coming; it's already here, with leading companies using the technology to reimagine processes across industries, functions, and jobs. Gen AI has dramatically raised the bar, requiring us to *think* with AI, ensure that we *trust* it, and continually *tailor* it—and ourselves—to perform better. Though gen AI is part of the extended movement to create more-symbiotic relationships between humans and machines, it's also unique in the annals of technology. No other major innovation in history has taken off so fast. Knowledge work is set to be transformed more quickly and powerfully than many of us can even imagine. Get ready. The future of business will be driven not by gen AI alone but by the people who know how to use it most effectively.

*A version of this article appeared in the September–October 2024 issue of Harvard Business Review.*



**H. James Wilson** is the global managing director of technology research and thought leadership at Accenture Research. He is the coauthor, with Paul R. Daugherty, of *Human + Machine: Reimagining Work in the Age of AI, New and Expanded Edition* (HBR Press, 2024).

✕ @hjameswilson

---



**Paul R. Daugherty** is Accenture's chief technology and innovation officer. He is the coauthor, with H. James Wilson, of *Human + Machine: Reimagining Work in the Age of AI, New and Expanded Edition* (HBR Press, 2024).

✕ @pauldaugh